On Approximability, Convergence, and Limits of CSP Problems

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von
Roland Markó
aus
Szeged, Ungarn

Bonn 2016

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter: Prof. Dr. Marek Karpinski
2. Gutachter: Prof. Dr. Anton Bovier

Tag der Promotion: 18. Juli 2016
Erscheinungsjahr: 2016

Abstract

This thesis studies dense constraint satisfaction problems (CSPs), and other related optimization and decision problems that can be phrased as questions regarding parameters or properties of combinatorial objects such as uniform hypergraphs. We concentrate on the information that can be derived from a very small substructure that is selected uniformly at random.

The research focused on theoretic foundations of the properties of very large combinatorial objects has received increasing attention in recent years, substantial part of this development was influenced by advances in the topic of approximation algorithms. Complex graphs and hypergraphs are in particular ubiquitous in this regard due to their application in modeling of real-world networks and community structures, respectively. The enormous size of such systems makes it practically impossible to capture them completely at any given moment, usually only a random substructure of bounded size is available for investigation.

The development of a self-contained theory of convergence and limit objects was initiated for dense graphs by Lovász-Szegedy, that was soon recognized to be a melting pot of several related areas, and poses fundamental connections to classical real analysis. The concept had appeared in a different form earlier in the work of Arora-Karger-Karpinski and Alon-de la Vega-Kannan-Karpinski dealing with dense instances of NP-hard optimization problems in terms of approximation.

In this thesis, we present a unified framework on the limits of CSPs in the sense of Lovász-Szegedy which depends only on the remarkable connection between graph sequences and exchangeable arrays established by Diaconis-Janson without recourse to the Frieze-Kannan type weakly regular partitions of graphs. In particular, we formulate and prove a representation theorem for compact colored r-uniform directed hypergraphs and apply this to rCSPs.

We investigate the sample complexity of testable r-graph parameters, and discuss a generalized version of ground state energies (GSE) and demonstrate that they are efficiently testable. The GSE is a term borrowed from statistical physics that stands for a generalized version of maximal multiway cut problems from complexity theory, and was studied in the dense graph setting by Borgs-Chayes-Lovász-Sós-Vesztergombi.

A notion related to testing CSPs that are defined on graphs, the nondeterministic property testing, was introduced by Lovász-Vesztergombi, which extended the graph property testing framework of Goldreich-Goldwasser-Ron in the dense graph model.

The novel characteristic requires the existence of a testable certificate property of edge colored graphs that connects to the original property through a certain general edge coloring operation. Lovász-Vesztergombi verified that any nondeterministically testable property is indeed testable in the original sense, the question regarding the relationship of the sample complexity of the original and the certificate property was left open.

In this thesis, we study the sample complexity of nondeterministically testable graph parameters and properties and improve existing bounds by several orders of magnitude. Further, we prove the equivalence of the notions of nondeterministic and deterministic parameter testing for uniform dense hypergraphs of arbitrary rank. This generalizes the result previously known only for the case of simple graphs. By a similar method we establish also the equivalence between nondeterministic and deterministic hypergraph property testing. We provide the first effective upper bound on the sample complexity of any nondeterministically testable r-uniform hypergraph parameter as a function of the sample complexity of its witness parameter for arbitrary r. The dependence is of the form of an exponential tower function with the height linear in r. Our argument depends crucially on the new upper bounds for the r-cut norm of sampled r-uniform hypergraphs. We employ also our approach for some other restricted classes of hypergraph parameters, and present some applications which give previously not known positive testability results.

Acknowledgements

First of all, I would like to express my deepest gratitude to my advisor, Professor Dr. Marek Karpinski, who introduced me to the topic of the thesis for his guidance, support and constant encouragement. I benefited immensely from his broad scientific expertise and his openness to new ideas and directions during our numerous discussion sessions. The amount of freedom I received from him contributed substantially to improve my abilities as a researcher.

In the course of the research that culminated in the current thesis I also profited greatly from the exchange with my other co-authors Krámli András and Ádám Timár, therefore I would like to thank them for their precious contribution.

I greatly appreciate the financial support and the excellent working conditions provided by the Hausdorff Center for Mathematics and the Computer Science Department of the University of Bonn.

Finally I would like to say thank you to my family and Zélie for their invaluable support.

Contents

CHAPTER 1

Introduction

The topic of Constraint Satisfaction Problems (commonly abbreviated as CSPs) is a central subject of research in the field of theoretical computer science, considerable effort has been invested into understanding their computational complexity. Instances of such decision and optimization problems can be described in a simplified way as a finite collection of variables taking values from a finite domain together with a family of boolean expressions called constraints defined over some subsets of these variables, together these comprise a CSP formula. The goal is to resolve whether all the constraints can be made true simultaneously, or maximize the number of satisfied constraints over all possible assignments of values to the variables.

As an example, consider a graph G that constitutes of a vertex set $V(G) = [n]$ and an edge set $E(G) \subset \binom{[n]}{2}$, the formula corresponding to the bipartition problem (the possibility of splitting the vertex set into two parts so that no edge runs between vertices falling into the same class) is $\wedge_{ij \in E(G)}(x_i \oplus x_j)$, where the variables are Boolean and the constraints are the sums in the clauses. In this case the graph G is bipartite if and only if each of the constraints is satisfied, that means that there exist an assignment of 0s and 1s to the variables that whenever $ij \in E(G)$, then x_i and x_j have to be different. Another common example is 3-SAT, that is a conjunction of disjunctions of the form $a \vee b \vee c$, where the symbols stand for Boolean variables x_i or their negations $\neg x_i$.

The task in the decision problem version is to determine whether there is an assignment of values to the variables so that all the constraints are satisfied, this problem class is known as CSP. In a broader sense, one can also ask for a quantitative statement, namely, if it is not possible to satisfy all of them, what is the maximal number of constraints that can be turn into a true statement at the same time, we refer to these counting problems as MAXCSP. In the above example this corresponds to finding a bipartition of the graph that cuts as many edges as possible, this is the so-called MAX-CUT problem. Complementary, one can examine the number of constraints that have to be satisfied by any assignment, the so-called MINCSP, and MIN-CUT, respectively. Further, the condition may be added that each constraint contains the same number of r variables, the general term for these problems is MAX-rCSP, or impose weights corresponding to the constraints, numerous other advanced generalizations appear in literature in several topics. We wish to clarify that the tasks of delivering only a verdict, and giving also the accompanying assignment are quite different but related endeavors, in the main part of this thesis we mostly deal with the (a priori easier) first, but for the time being we stick to the latter algorithmic problem in order to provide a more complete background. All these discussed problems are known to be NP-hard in general.

The common question one can ask in this setup in the complexity sense is to determine the range of resources (time, space, or measure of error possibilities in any sense imaginable) that are required in order to be able to solve these problem instances in general or under some additional restrictions. If one settles for some trade-off in the output a dramatical improvement regarding the above factors is possible. For counting problems, such as MAXCSP, approximation algorithms have been developed that are computationally feasible.

1.1 From approximation algorithms to sampling

The most frequent way to define the quality of the approximation algorithm $\mathcal{A}$ is in the case of maximization problems to relate the output value given by the algorithm $\mathcal{A}(I)$ to the optimal value $OPT(I)$ the following way: one has to determine the maximal value of $\mathcal{A}(I)/OPT(I)$ over all instances I, if this is larger than some $c \leq 1$, then the problem is c-approximable, and the largest such c is called the approximation ratio of the algorithm (sometimes the reciprocal is referred to under this name). The measurement of the quality in this sense is often called multiplicative approximation. For one of the most prominent problems, MAX-CUT, being given an arbitrary ordering of the vertices, the trivial stepwise greedy choice for the placing of the vertices provides an approximation ratio of 1/2. A significant breakthrough in this direction was the now classical approximation result by Goemans and Williamson [62] delivering a ratio of 0.878....

Polynomial time approximation schemes This concept raises the question whether an NP-hard problem is c-approximable in a computationally feasible way for every $c < 1$, with other words, does it have a so-called *polynomial time approximation scheme* (PTAS). Several NP-hard problems are known to permit PTAS, such as the euclidean traveling salesman problem or the minimal vertex cover on planar graphs. However, for the MAX-rCSPs we are dealing with there are genuinely such barriers so that beyond them the approximation is also hard, see [19].

A general multiplicative approximation scheme $\mathcal{A}$ is a parametrized family of approximation algorithms $(\mathcal{A}(\varepsilon,.))_{\varepsilon>0}$, where for each $\varepsilon > 0$ and instance I we have $\mathcal{A}(\varepsilon, I)/OPT(I) \geq 1 - \varepsilon$, for the PTAS case $\mathcal{A}(\varepsilon,.)$ runs in polynomial time in the size of the instance, but perhaps in exponential time in the inverse of the proximity parameter ε.

The conditions a PTAS has to fulfill can be modified by replacing $\mathcal{A}(\varepsilon, I)/OPT(I) \geq 1 - \varepsilon$ by the requirement $|OPT(I) - \mathcal{A}(\varepsilon, I)| \leq \varepsilon|I|$, where $|I|$ is the size of the instance, or the whole size of the data needed to describe it, which is problem-dependent. For example, in the case of optimization problems concerning dense graphs $|I|$ can be the square of the size of the vertex set, but for a parameter of bounded degree graphs it mostly means the cardinality of the vertex set. With this alteration we speak about additive approximation, this is also our main concern in this work.

There is no straightforward way to construct an additive PTAS from multiplicative one in general, or vice versa, but in case of dense MAX-rCSP (meaning that only instances are considered with at least ρn^r constraints with n being the number of variables) one can derive from an additive one a multiplicative PTAS. First note that $OPT(I) \geq (\rho/2^r)n^r$ as the expected value for a uniform random assignment is at least this large. This implies for the additive scheme $\mathcal{A}$ that $\mathcal{A}(\alpha, I)/OPT(I) \geq 1 - \frac{\alpha n^r}{OPT(I)} \geq 1 - 2^r\alpha/\rho$, which can be arbitrary close to 1 by setting α sufficiently small. The analogous reasoning cannot be carried out for MIN-rCSP even in the dense case.

The first systematic work delivered on the existence of PTAS for dense instances of optimization problems through additive PTAS was carried out by Arora, Karger, and Karpinski [18], it was shown using exhaustive sampling and smooth integer programs for several subfamilies of dense MAX-rCSP, such as MAX-CUT or MAX-r-SAT, that there exist additive PTAS, and a scheme that is suitable for general MIN-rCSP was presented. The algorithms suggested a required a sample size of $\Theta(\log n)$, giving also the algorithm for finding a near-optimal assignment, but no information about properties of random subproblems was obtained. In this thesis we focus almost entirely on the behavior of the subproblems.

Random sampling in approximation algorithms Research sparked by approximation schemes employing the sampling method motivates the concept of the *sample complexity* that is one of the central subjects of the current thesis. In short, the question is what can the observer of a randomly chosen subproblem tell about the solution of the original one with the knowledge of the mechanisms of the random process that generates the sample. The answer may and in general does depend on the amount of the perceived random data, the least size that meets our expectations is called sample complexity.

Subsequently to [18], Goldreich, Goldwasser, and Ron [66] introduced the framework of *property testing* for combinatorial structures, and derived from their testers the first *constant time approximation schemes* for partition problems on graphs drawing on ideas presented in [18], including setups that give additive PTASs for the value of MAX-2CSP. However, the method of [66] does not provide a feasible assignment, and this is certainly not possible in constant time.

A graph property is a subset of graphs invariant under isomorphism. Property testers in the most general sense are algorithms that can distinguish between the cases that an instance is having the property and that it is ε-far from any instance that has it, with sufficiently large probability. A property is testable if such an algorithm exists.

Accordingly, the tester in [66] for example for the property of a graph of size n having a cut with at least ρn^2 cut edges was designed as follows. First a uniform random vertex sample of size poly($1/\varepsilon$) is taken, and after making some calculations on the induced subgraph (perhaps without accessing the information regarding all of its adjacencies), and the tester provides a decision whether the graph has the property or no modification of at most εn^2 adjacencies would result a graph with the property. This was a major breakthrough in the area, since it was shown that the MAX-CUT value of G (more precisely, the density of the maximal cut) can be approximated well by observing a uniformly sampled subgraph whose size is independent of the order of the original graph, the approach is reminiscent of methods in mathematical statistics.

Using similar techniques Andersson and Engebretsen [17] obtained PTAS for MAX-rCSP with non-Boolean domain. Parallel to these developments Frieze and Kannan [59] proposed an algorithmically efficient version of Szemerédi's Regularity Lemma, which is later in literature often referred to as the Weak Regularity Lemma, in order to handle the MAX-2CSPs. The original version due to Szemerédi [115] has risen to the single most important tool in several branches of combinatorics. Informally it states that for every graph the vertex set can be partitioned into a bounded (as a function of an error parameter ε) number of classes so that for most pairs of classes the bipartite graph spanned between the two parts is random-like. It eluded practical algorithmic applications, since the upper bound regarding the number of classes of the partition provided was an exponential tower of height $1/\varepsilon$ in magnitude, this bound later turned out to be sharp. Actually, it was implicitly used to derive testing algorithms that sample subgraphs whose size only depends on $1/\varepsilon$ (but, as above, in a computationally infeasible way) for k-colorability by Rödl and Duke [103], and for triangle-freeness by Ruzsa and Szemerédi [111] a decade before the introduction of the concept of property testing.

The variant in [59] introduced the cut norm of matrices $\|A\|_\square = \max_{S,T\subset[n]} |A(S,T)|$, and reads as follows: For every square matrix A there exists a partition of rows and columns (the same for the two) such that the matrix B obtained from A by taking the average on the rectangles determined the by pairs of classes is close to A in the cut norm, that is $\|A - B\|_\square \le \varepsilon n^2$. The lemma admits partitions into parts whose number is only exponential in $1/\varepsilon^2$, therefore it is a suitable tool for designing PTAS. The authors of [59] applied the lemma to the MAX-CUT problem and the quadratic assignment problem among others combined with uniformly sampling exp(poly($1/\varepsilon$)) vertices, but did not establish any connection between the solution of the subproblem and of the original one.

A meta-algorithm called canonical testing was first employed for MAX-CUT by Alon, Fernandez de la Vega, Kannan, and Karpinski [14] for finding a suitably good value estimation that runs in constant time. The method itself is remarkably simple: One takes a random sample of size depending on ε, evaluates the optimum on the sample by brute force (analogously to the exhaustive sampling method), taking as output that value satisfies the permitted error condition in an additive sense, although only with high probability. The related approach for testing properties also turned out to be fruitful.

Inspired by the regularity approach of [59], Alon, Fernandez de la Vega, Kannan, and Karpinski [14] also improved on the speed of the PTAS for MAX-rCSP instances with Boolean domain, and refined the analysis regarding the optimal value on the random subproblems. Using the so-called cut decomposition method and linear programming duality they were able to obtain the upper bound $O(\varepsilon^{-4}\log(1/\varepsilon))$ for the sample complexity.

The original motivation for our research in the current thesis in a broader sense is formulated in the following question.

Question. *What decision and optimization problems admit such constant time approximation schemes as in [14]? What is the exact magnitude of the sample complexity in these cases?*

We will not be concerned with approximation algorithms that find an actual certificate, instead we will consider very large problems.

1.2 Sampling from large combinatorial structures

Any family of CSP formulas can be interpreted as a class of combinatorial objects: consider a hypergraph such that each variable of the formula corresponds to a vertex, whereas constraints can be thought of as edges with certain colors. This way, MAXCSP is a parameter of the hypergraph, and further, the satisfiability characteristic transforms to a property of edge colored directed hypergraphs. We demonstrated how this goes in the opposite direction at the start of the chapter with the maximal cut of graphs.

One consequence of the graph representation is the ability to formulate and study structural questions in the graph theory framework beyond the objective function of the optimization problem. This motivates the wide scope of the thesis, we cover topics beyond the complexity theory range and deal with recent developments in combinatorics.

1.2.1 Testing properties of graphs

The combinatorial representation brings us back to property and parameter testing for hypergraphs in the dense model with CSPs being a special case, the general question above reads in this setting as follows.

Question 1. *Which properties and parameters of edge colored hypergraphs are testable, what is the magnitude of the sample complexity?*

Whereas in the design of PTASs the goal is fast estimation of graph or hypergraph parameters and also giving a nearly optimal assignment, in the case of constructing related property testers one only attempts to distinguish between graphs having a given property, and the ones that are far from having it.

Rubinfeld and Sudan [109] studied testing first in the setup of functions, for instance, when the task is deciding about linearity. Subsequently, as also mentioned above, Goldreich, Goldwasser, and Ron [66] initiated the systematic research in the field for combinatorial objects by obtaining various results and establishing the connection to the topic of approximation algorithms. The first framework that in [66] was investigated was the dense graph model. The distance here is measured between two graphs by the minimal number of edges that have to be modified to arrive from one at an isomorphic copy of the other (modification is meant for exchanging an edge for a non-edge or vice versa), this is called edit distance. Queries of two types are permitted: one may demand to be given an uniformly chosen random vertex, or ask for the information whether an edge runs between two previously received vertices. The answers are assumed to be given by a probabilistic oracle machine, the construction method of such an oracle is not of interest here.

This model suits best the cases when most graphs that have the property are dense, but density alone does not imply the property. Other models subsequently defined deal with the bounded degree case or a mixture of these two, see Goldreich and Ron [64]. In the dense framework, that is the main subject of the current work, a tester for the graph property $\mathcal{P}$ is an algorithm $\mathcal{A}$ that for an n-vertex graph G and a proximity parameter $\varepsilon > 0$ runs with $q(n, \varepsilon)$ queries and (*i*) accepts G with probability at least 2/3, if $G \in \mathcal{P}$, and (*ii*) accepts G, if it is ε-far from $\mathcal{P}$ (that is, no modification of at most εn^2 edges results in an element of $\mathcal{P}$) with probability at most 1/3.

In our setup, we assert testability only in the situation, when the number of required queries does not depend on the size of the graph. Furthermore, a tester is called one-sided if it always accepts instances that satisfy the property, and non-adaptive, if the queries made are not dependent on the outcome of previously made ones. We restrict ourselves to obtaining a so-called canonical tester. The algorithm $\mathcal{A}$ is a canonical tester for $\mathcal{P}$, if there exists a function $k\colon \mathbb{R}^+ \to \mathbb{N}$ and another graph property $\mathcal{P}'$, such that

(i) $\mathbb{P}(\mathbb{G}(q, G) \in \mathcal{P}') \geq \frac{2}{3}$ for every $G \in \mathcal{P}$ and $q \in \mathbb{N}$, and

(ii) for every $\varepsilon > 0$ and G that is ε-far from $\mathcal{P}$ with at least $k(\varepsilon)$ vertices, and for every $q \geq k(\varepsilon)$ we have $\mathbb{P}(\mathbb{G}(q, G) \in \mathcal{P}') \leq \frac{1}{3}$,

where $\mathbb{G}(q, G)$ is the random induced subgraph of G on a uniformly chosen subset of $V(G)$ of cardinality q. The minimal function $k(.)$ that satisfies the above conditions is called *sample complexity* of the property.

Every tester can be turned into a canonical tester accepting at most a quadratic trade-off in the query complexity, see [65], also, canonical testers are non-adaptive. Designing PTAS from testers is possible for some threshold-like properties, in [66] it was noted for ρ-cut (the property that the density of the MAX-CUT is at least ρ) that their result does in fact yield an algorithm that finds a large cut of size at least MAX-CUT$(G) - \varepsilon n^2$ in sublinear time. However, there is no straightforward way to construct such an algorithm from general testers.

A major step towards understanding which properties are testable, and to what extent with regard to query complexity, was taken by Alon, Fischer, Newman, and Shapira [16]. The authors provided the first combinatorial characterization of graph testability using Szemerédi's Regularity Lemma. They proved that the property of having a particular fixed regular partition is testable, and introduced the notion of regular reducibility. A property has this characteristic if there is a finite number of regular partitions such that at least one applies to any graph that has the property. It was shown that regular reducibility is equivalent to testability. However, as a result of the application of the Regularity Lemma, upper bounds on the sample complexity were outperformed by previous efforts for known special cases. On the other hand, the characterization lead to the first example of natural graph properties that are not testable, namely being isomorphic to a copy of the Erdős-Rényi graph. Subsequently it was shown by Alon [3] that triangle-freeness is testable, but it requires strictly more than poly($1/\varepsilon$) queries.

Property testing serves as a major motivation for the study of limits of graph sequences. Several testability result can be proven or re-proven in a concise analytic fashion (however, non-effectively) using limit theory notions and theorems together with the fundamental concepts in real analysis, see [49] and [94]. We will cover this aspect of the limit theory in Chapter 3 in greater detail.

1.2.2 Emergence of graph limits

As we mentioned above, in this thesis CSPs are considered with an extremely large number of variables, therefore the corresponding hypergraphs are also huge. Imagine an evolving, ever-growing random hypergraph, that we can only access through sampling, what can be said about the driving model if the data we gain shows some patterns? Equivalently, we would like to make sense of some limit structure that carries the important features of the random model with the noise filtered out, perhaps a hypergraph on a countably infinite vertex set.

From the traditional perspective of computer scientists, the input data (G for example for MAX-CUT(G)) in its raw form is of central importance, and the goal is often to determine exact parameter values of graphs, problems are analyzed with regard to computational hardness (running time or memory space) depending on the size of the graph.

In numerous applications one has to deal with large random objects with completely unknown or parameterized underlying distribution, with ones whose data is corrupted by random noise, or with instances that are simply too large to store and analyze, therefore it is sensible to write off exactness in favor of resource optimization. At this point a statistical perspective enters the picture and approximate solutions are sufficient for the problems above. The recently emerged limit theory of discrete objects captures this aspect, see [89]. The concept has gained considerable attention and fueled a large amount of research output, and also open questions have been raised within this framework. One important advantage of the limit theory is that it encompasses the methodology of how to deal with the problems defined on very large graphs in a conceptual way corresponding to the theory of classical statistics and analysis.

Question 2. *How can we interpret extremely large structures so that we can tackle the first, qualitative part of Question 1?*

The idea to identify and study the limit space for relational structures originates from the fact that recent technological development has made it possible to capture large graph-like data sets emerging in the real-world, common examples are social networks or protein interaction networks, where pair-wise interaction between points is observable. These turn out to be too huge to be able process their whole data, may be heavily time-dependent, and are often driven by some random process. The central question is whether methods of mathematical analysis successful in other branches of natural sciences are applicable or can be adapted to these complex networks in order to extract meaningful statistics that are robust in some sense.

In a series of papers Lovász and Szegedy [91], Borgs, Chayes, Lovász, Sós, and Vesztergombi [30, 32], Borgs, Chayes, and Lovász [31] laid out the foundation of the limit theory in the case of simple dense graphs. This class seems to be the most straightforward choice as a starting point from the mathematical viewpoint and the results beautifully correspond with the framework of measure theory and functional analysis. On the other hand the vast majority of real-world complex networks are sub-dense, therefore other classes are of similarly high interest.

A simple graph $G = (V(G), E(G))$ is given by its vertex set $V(G)$ and edge set $E(G) \subset \binom{V(G)}{2}$, we use the abbreviation uv for sets $\{u, v\}$. The graph G can be represented by a $\{0, 1\}$-valued symmetric square matrix A_G of size $|V(G)|$ dubbed the adjacency matrix of G. Two graphs G and G' are said to be isomorph if there exists a bijection ϕ between $V(G)$ and $V(G')$ that preserves edges and non-edges, that is $uv \in E(G)$ if and only if $\phi(u)\phi(v) \in E(G')$. We speak about labeled graphs if we wish to stress the vertex set, and about unlabeled graphs if isomorphism classes are meant, the first will be considered as default in this work. The map $\phi : V(G) \to V(G')$, without the condition of being a bijection, is a graph-homomorphism when it preserves only the presence of edges, not necessarily their absence. Graph limits are defined in terms of the densities of homomorphisms of small graphs: a sequence $(G_n)_{n=1}^{\infty}$ is convergent if for any simple graph F the numerical sequences $(t^*(F, G_n))_{n=1}^{\infty}$ converge, where $t^*(F, G) = \hom(F, G)/|V(G)|^{|V(F)|}$ is the homomorphism densityof F in G, and $\hom(F, G)$ is the number of maps from $V(F)$ to $V(G)$ that are homomorphisms [91]. The measurable functions $W \colon [0, 1]^2 \to [0, 1]$ that are symmetric in the sense that $W(x, y) = W(y, x)$ for every pair $x, y \in [0, 1]$ are referred to as graphons, they can be interpreted as graphs with vertex sets that have continuum cardinality. The homomorphism densities can analogously be defined for graphons as the integral

$$t^*(F, W) = \int_{[0,1]^k} \prod_{ij \in E(F)} W(x_i, x_j) \mathrm{d}\lambda(x),$$

where $V(F) = [k]$. The most important result of [91] is that for every convergent graph sequence $(G_n)_{n=1}^{\infty}$ there exists a limit graphon W in the sense that for each simple F the sequence $t^*(F, G_n)$ tends to $t^*(F, W)$. Moreover, the converse is also true, that is, every graphon W serves as the limit object of some convergent graph sequence.

Another major contribution in [30] was the natural metrization of the topology described above, which was also extended to the limit space. The definition of the metric itself originates from the concept of the cut norm introduced by Frieze and Kannan [59]. Various other equivalent convergence criteria were enlisted in [32], we elaborate on this aspect in the main body of the thesis. Uniqueness of the limit W up to a certain measure-preserving equivalence was subsequently verified in [31].

Testability of graph parameters is a central concept around which the topics of this thesis are assembled, it is also the combinatorial analog to additive approximation for the optimal value of discrete optimization problems. A graph parameter f is testable, if for every margin of error $\varepsilon > 0$ the value $f(G)$ for a graph G can be additively ε-estimated by looking at a random induced subgraph of bounded size independent of the size of G. The estimation is allowed to fail with probability at most ε.

A characterization of this property in terms of graph limits was given in [30] mainly reformulating the definition in the light of the results outlined above. Namely, the testability of a parameter f is equivalent to the statement that for every convergent sequence $(G_n)_{n\geq 1}$ of simple graphs with $|V(G_n)| \to \infty$, the numerical sequence of the values $(f(G_n))_{n\geq 1}$ also converges. However, in [30] no attempt was made to determine the magnitude of the required sample size as a function of f and ε, the current work endeavors to shed some light on this issue.

Question 3. *Can we derive results for discrete problems by studying their counterparts in the limit space?*

The topic of this thesis is at the intersection of complexity theory,statistics, and graph theory. Each of the general problems we addressed in the above questions have been subject of ongoing and extensive study, partial answers are available, but there are still blind spots, see Section 1.4 for an overview. In the next section we will outline our attempts and results on more concrete problems in the context of the above questions. Their thorough discussion gives the bulk of the thesis.

1.3 Our contribution

Substantial parts of the thesis previously appeared in the preprints Karpinski and Markó [79, 80, 81, 82], and in the paper Krámli and Markó [86].

Nondeterministic hypergraph testing (to Question 1)

Lovász and Vesztergombi [97] introduced the framework of *nondeterministic graph property testing*. A simple graph property $\mathcal{P}$ has this characteristic if it can be certified by a property $\mathcal{Q}$ of edge-colored graphs that is testable in the sense that for any element $G \in \mathcal{P}$ there is a $\mathbf{G} \in \mathcal{Q}$ such that if we omit edges with certain colors and disregard the colors of the remaining edges then we obtain G, see Definition 6.1.2. It was shown by Lovász and Vesztergombi [97] and Gishboliner and Shapira [61] that nondeterministic testability is equivalent to testability in the graph case.

We study this question in the context of graph parameters and obtain an upper bound on the sample complexity of a nondeterministically testable parameter that is a 3-fold exponential of the sample complexity of its certifying witness parameter, Theorem 6.1.5. We generalize the result to the case of r-uniform hypergraphs (definitions are analogous), and obtain an upper bound on the sample complexity int terms of the witness as in the graph case that is an $l(r)$-fold iteration of the exponential function, where l is linear in r, Theorem 6.1.6. We further show that these results also hold for properties, Theorem 6.1.7. Our main tools are variants of Szemerédi's Regularity Lemma, Lemma 3.3.15 and Lemma 3.3.27.

We apply these results to some hypergraph parameters and properties, and show their testability, some of these are new results. We also consider more restrictive forms of nondeterministic testability, and obtain improvements in these cases over the general upper bounds. The results regarding graphs can also be found in the preprint [80], whereas the results in the hypergraph case appeared in the preprints [81, 82].

Limits of colored hypergraphs (to Question 2)

We give a characterization for the limit space of r-uniform edge colored hypergraphs for arbitrary r in the situation when edge colors are taken from a compact Polish space. We present two methods, both generalizations of previous approaches of Diaconis and Janson [43] using exchangeability principles, and of Elek and Szegedy [49] relying on non-standard analysis, respectively. We combine these methods with the concept of Lovász and Szegedy [93] employed to deal with sequences of graphs with compact edge colors. The first framework is more straight-forward and gives insight regarding random graph models, whereas the second one allows for more applications such as direct testability assertions and regularity lemmas.

We prove a suitable representation of the limit space following the program of [43] developed for simple graphs. The crucial ingredient is a deep result regarding the representation of *exchangeable r-arrays*. Our main result here, Theorem 2.2.10, states that the limit objects are r-graphons that take probability measure values. We apply the result to formulate the representation theorem for the limits of CSP formulas, Corollary 2.3.2. These contributions appeared also in the first part of the preprint [79].

Our motivation for the second approach to hypergraph limits comes from a variant of the ground state energy problem that is reminiscent of MAX-rCSP, where in the 3-graph case we optimize over vertex pair colorings rather than simple vertex colorings, see Definition 5.3.1. We refer to them as r-GSE.

The *ultralimit* method for combinatorial structures was first employed by Elek and Szegedy [49]. With the main link between discrete and continuous objects, the separable realization, it enabled proofs for qualitative results, such as the testability of hereditary properties or the Hypergraph Regularity Lemma that are rather direct consequences of measurability in the limit [49]. This structural connection is the main advantage over the exchangeability correspondence principle.

First we consider the case when the edges of the hypergraphs are colored with colors from a finite set. Using the analogous statement to the separable realization of

[49], Corollary 5.2.14, we manage to prove the existence and the uniqueness statement regarding the limit objects, Corollary 5.2.9 and Theorem 5.2.15, respectively. Building on this we are able verify that r-GSEs are testable, Theorem 5.3.4. This part appeared previously as a section of the preprint [81].

We generalize the limit representation for sequences of colored hypergraphs, where the colors are taken from a compact Polish space in Theorem 5.4.11. Here the limit space is similar to the previous case, the r-graphons are probability measure valued. We also formulate the condition of two limit objects being equivalent similar to the finite color case in Theorem 5.4.12.

An application of these results is a new proof of (a version of) the Hypergraph Regularity Lemma for r-graphs with compact colors, see Theorem 5.5.2.

Testability of the ground state energies (to Question 3)

We introduce a generalization of the ground state energies borrowed from statistical physics that were studied first in [32] regarding their testability in a qualitative way, see Definition 4.1.2. This new notion also encompasses MAX-rCSP.

We prove a quantitative bound for the ground state energies inspired by the approach of Alon, Fernandez de la Vega, Kannan, and Karpinski [14], Theorem 4.1.4. Our result reproves the sample complexity upper bound $O(\log(1/\varepsilon)(1/\varepsilon)^4)$ obtained in [14] for MAX-rCSPs, extends it for CSPs with non-Boolean domain, additionally, our bound regarding the error probability is tighter.

We further use this result to obtain similar conclusions for the ground state energy variant with an external magnetic field, Theorem 4.3.2, and the microcanonical version, Theorem 4.3.6. Another application is a related testability result for quadratic assignment problems, Corollary 4.3.9. These results previously appeared in the second part of the preprint [79]. Additionally, we prove testability for ground state energies of graphs with non-negative unbounded weights Theorem 4.4.3 under an L^p-bound condition.

We investigate the separation quality of ground state energies, more precisely under what conditions these energies determine a graph. In the course of this, we prove a convergence hierarchy of the lower threshold ground state energies, Theorem 4.5.11. This former result can also be found in the paper [86].

All the above results are proved first in the limit space, and as an application the discrete counterparts are deduced in the spirit of [32].

1.4 Related work

In this section we provide a review of results in the different areas we touch in the course of the thesis.

Further polynomial time approximation schemes

Inspired by the regularity approach of Frieze and Kannan [59], Alon, Fernandez de la Vega, Kannan, and Karpinski [14] improved on the speed of the PTAS for MAX-rCSP instances with boolean domain, and refined the analysis regarding the optimal value on the random subproblems. Using the so-called Cut Decomposition method and linear programming duality they were able to obtain the upper bound $O(\varepsilon^{-4}\log(1/\varepsilon))$ for the sample complexity. One key tool that is proved in [14] is that the cut norm (or its density, depending on the definition) of a large matrix is preserved under row/column sampling in the sense of graph parameter testing, the required sample size can be upper bounded by $O(\varepsilon^{-4}\log(1/\varepsilon))$. This auxiliary result also happens to be a an important ingredient in showing that the $\delta_\square$ distance of graphs defines the same topology as graph convergence in the context of graph limits, see below. The upper bound on the sample complexity of the cut norm was further improved to $O(\varepsilon^{-2}\log(1/\varepsilon))$ in a weaker sense by Rudelson and Vershynin [110] using the singular value decomposition (SVD) and decoupling: it holds that $\mathbb{E}\|A_S\|_\square \varepsilon^4 \leq C\|A\|_\square/n^2$ for some $C > 0$ large enough, where A is an arbitrary $n \times n$ matrix, and A_S is its randomly sampled $\varepsilon^{-2} \times \varepsilon^{-2}$ submatrix.

Continuing the research initiated in [14], Fernandez de la Vega, Kannan, and Karpinski [53] extended the sample complexity result to MAX-rCSP with additional global constraints, such as MAXBISECTION. In the same line of work Fernandez de la Vega and Karpinski [51, 50] obtained PTAS for slightly subdense instances with $\Omega(n^2/\log n)$ constraints, and for weighted families of graphs with unbounded edge-weights obeying a certain density condition, respectively. As a further contribution, Fernandez de la Vega, Kannan, Karpinski, and Vempala [52] developed PTAS for core-dense instances which can be subdense and contain quasi-metrics using tensor decomposition of arrays (a method resembling the singular value decomposition of matrices) and uniform sampling.

Departing from regularity approaches, Mathieu and Schudy [98] described a surprisingly simple PTAS for MAX-rCSP running in quadratic time consisting of exhaustive sampling and greedy steps, analysis was conducted by martingale methods using a fictitious cut construction that is updated at each step of the generating process, proving also the best sample complexity upper bound $O(\varepsilon^{-4})$ known to date. Their technique was then refined by Karpinski and Schudy [83] to produce a PTAS that runs in sublinear time, see also Yaroslavtsev [118].

All the above PTAS take advantage of sampling arguments and therefore by default provide a satisfactory output only with high probability. Nevertheless, all of them can be derandomized in a trivial way to obtain a feasible value estimation (except for [18], where a random walk on an expander has to be performed, but this is also a standard method) with a trade-off in the running time that remains still polynomial in the problem size.

We note that additive PTAS cannot be obtained when the factor determining the magnitude of the permitted deviation ($|V(G)|^2$ in case of MAX-CUT) is substantially is reduced. Ailon and Alon [1] proved inapproximability bounds for non-trivial families of rCSP (a trivial situation is for example, when the number of edges of graphs is given as the MAX-rCSP, then it can be calculated in polynomial time, so studying approximation hardness is redundant), even in the fully dense case, where each r-tuple carries at least one constraint, it is NP-hard in general to give an approximate solution within an additive error at most $n^{r-\delta}$ for any $\delta > 0$ in permitted time.

Approximation algorithms beyond PTAS

It is well-known that MAX-rCSP is contained in the class of NP-hard problems, therefore considerable effort has been made towards the design of approximation algorithms. For one of the most prominent problems, MAX-CUT, being given an arbitrary ordering of the vertices, the trivial stepwise greedy choice for the placing of the vertices provides an approximation ratio of 1/2.

A significant breakthrough in this direction was the now classical approximation result by Goemans and Williamson [62] delivering a ratio of 0.878... using the semidefinite programming (SDP) relaxation of $\max_{x\in\{-1,1\}^n} \frac{1}{2}\sum_{i,j} A_{ij}(1-x_ix_j)$ and random hyperplane rounding. This result is optimal under the Unique Games Conjecture, the technique employed in [62] was a highly influential contribution to the area.

Semidefinite programs are special convex optimization problems that are known to be solvable in polynomial time, see for example Grötschel, Lovász, and Schrijver [70], in the relaxation above one replaces the domain of variables from $\{-1,1\}$ to the n-dimensional unit sphere S^{n-1}. Clearly, the optimum of the relaxed version is at least as large as that of the original, therefore the task is to obtain an integer solution with only a small decrease in the value from the former. This approach has been adapted to other MAXCSPs, however, the approximation ratio gained through this method is highly problem-dependent and is connected to the shape of the objective function. The estimation of Charikar and Wirth [35] considering $\max_{x\in\{-1,1\}^n} x^TAx$ achieved a ratio $\Omega(1/\log n)$ that is decreasing in the size n of the problem using SDP relaxation and randomized rounding.

The cut norm of matrices with both positive and negative entries is a central subject in graph limit theory, it is used to measure closeness of two graphs by calculating the norm of the difference of their adjacency matrices. Alon and Naor [6] algorithmically approximated the cut-norm relying on Grothendieck's inequality, by the virtue of the cut-norm and the $\|.\|_{\infty\to 1}$ operator norm being equivalent they obtained an approximation through considering $\max_{x,y\in\{-1,1\}^n} x^TAy$, and relaxing this to $\max_{u,v\in(S^{n-1})^n}\sum_{ij} A_{ij}\langle u_i, v_j\rangle$. The optimal values of the two expressions are known to be at most a constant factor K away from each other, it is referred to as Grothendieck's constant in functional analysis, whose exact value is not known, however $K \leq 1.782...$ is true. In [6], three different rounding methods were presented: an approach using orthogonal arrays of strength 4, one with Gaussian rounding, and the original argument leading to the proof of Grothendieck's inequality. Subsequently, Alon, Makarychev, Makarychev, and Naor [15] introduced the Grothendieck constant K_G of an arbitrary graph G. It holds by definition that $K_G \max_{x\in\{-1,1\}^n}\sum_{ij\in E(G)} A_{ij}x_ix_j \geq \max_{u,v\in(S^{n-1})^n}\sum_{ij\in E(G)} A_{ij}\langle u_i, v_j\rangle$ for any real square matrix A, where the integer program appears in the study of spin glass models.

It was verified in [15] that the integrality gap is at most $O(\log(\theta(G^c)))$, where θ is the Lovász Theta Function. For a more complete picture we refer to the recent survey of Khot and Naor [84] on the application of SDP relaxations and Grothendieck-type inequalities in combinatorial optimization.

Recently, Raghavendra and Steurer [102] presented a rounding scheme originated from the SDP relaxation of MAXCSPs that achieves the optimal integrality gap for any CSP family in consideration. Building on this, Barak, Hardt, Holenstein, and Steurer [25] investigated the testability behavior of mathematical relaxations of combinatorial optimization problems in the fashion of property testing, and among other results presented a sample complexity result for MAX-rCSPs that applies to a more general class of problems, than the dense case. For dense instances poly$(1/\varepsilon)$ sample size is sufficient according to their analysis, for the bounded degree case at least $\Omega(n)$ size is required, testability also holds for BasicSDPs and BasicLPs that are relaxations of a special form. Negative results involving relaxations with higher levels of the Lasserre and Sherali-Adams hierarchies, respectively, were provided showing the limitations of the uniform sampling method.

As another approach, Drineas, Kannan, and Mahoney [46], designed an approximation algorithm for the MAX-CUT problem for the case of heterogeneous weighted graphs. The desired proximity magnitude is modified in [46] in comparison to additive PTASs, the tolerated deviation of the output from the optimal is at most $\varepsilon n \sqrt{|E(G)|}$ instead of εn^2. The sampling method was adapted to the heterogeneous setting: nodes are picked following a non-uniform distribution through the employment of the $C\tilde{U}R$ matrix decomposition.

More on limits of combinatorial structures, exchangeability methods

The developments described in Section 1.2.2 fueled a significant amount of research generalizing the limit theory of dense graph in any possible direction. For the dense case we have an almost complete picture now, questions that come up regarding limits are first posed in this context. What is most important for the current thesis is the parameter testing framework established in [32] and the connections revealed to statistical physics. Lovász and Szegedy [95] defined a distance on the space of vertices of a given graph instance measuring the similarity of vertices as part of the global network, in particular two vertices are close if they have a close number of common neighbors with a randomly chosen third vertex. This similarity distance has advantages in the design of PTASs for MAX-2CSPs. The same authors generalized the Regularity Lemma, both the weak and Szemerédi's version to an approximation result in Hilbert spaces in [92], translating its statement it into a compactness and a covering result with respect to the topology inside the graphs from the previously mentioned paper [95]. Among other results of algebraic nature Lovász and Szegedy [96] introduced the notion of finitely forcible graphons, that are exactly determined by a certain finite set of subgraph densities. Prior to this, Lovász and Sós [90] introduced a generalized quasi-randomness condition for graphs that is fulfilled by being infinitesimally close to a certain graphon in the cut distance. The graphon here is a step function whose steps are rectangles corresponding to a partition of $[0,1]$ akin to the driving models of stochastic block models, the notion itself is similar to regular reducibility in testing. In [90] it was shown that such graphons are finitely forcible. In [96], the authors gave necessary conditions and expanded the class by graphons that mostly resemble well-known extremal graphs.

Existence and uniqueness of limit objects, as well as the existence of a regularity lemma were subsequently extended to hypergraphs by Elek and Szegedy [49] via the ultralimit method using non-standard analysis. The limit objects of sequences of r-uniform hypergraphs turned out to be the measurable functions defined on the unit cube $[0,1]^{2^r-2}$ that take values in $[0,1]$. The coordinates of the domain $[0,1]^{2^r-2}$ correspond to non-trivial subsets of $[r]$, the functions serving as a limit object must be invariant under coordinate permutations induced by any permutation of the symmetric group S_r. Using similar techniques, Elek [48] described the limit space of metric spaces and analyzed the emergence of observables in the limit.

The case of bounded degree graph limits is quite different, less measure theoretical and more algebraic arguments are entering the picture. This line of research was initiated by Benjamini and Schramm [26], subsequent progress has been made by Elek [47]. In this case the similarity of graphs is measured locally, the condition of convergence is defined by the convergence of the probability distribution of rooted l-neighborhoods of a uniformly chosen random root vertex for each $l \in \mathbb{N}$, this is called local convergence. The limit objects can be represented by unimodular distributions on rooted countable graphs with bounded degree, or by graphings, that are bounded degree graphs whose vertex set is a Borel probability space, the edge set is a measurable set, that satisfies a measure preserving property analogous to symmetry in the case of dense simple graphs.

Another line of investigation studies the limits of dense graphs with unbounded weights. Note that this setup is analogous to the sub-dense case excluding bounded degree graphs. The original definition of convergence quickly falls apart as in general the subgraph densities are infinite in this region. Since the cut distance, with prior 1-norm normalization of the weights of the graphs, is still meaningful, the definition of convergence by the Cauchy property is justified. Research in this direction was initiated by Bollobás and Riordan [29], further progress has been made by Borgs, Chayes, Cohn, and Zhao [33, 34] expanding the considered class of graphs to those that are in L^p as functions for some $p \geq 1$, they showed the equivalence of the energy convergence to the metric convergence analogously to the results in [32].

Connected to the latter developments, limits with edge colors from sets with topological but without algebraic structure were considered, compact decorated graphs were explored by Lovász and Szegedy [93], whereas Banach space decorated graphs were analyzed by Kunszenti-Kovács, Lovász, and Szegedy [87]. In both cases the limit objects resemble graphons, for the first model they are graphons whose values are probability distributions on the compact space, in the second case values are taken from the dual of the Banach space.

The work by Diaconis and Janson [43] placed the limit theory of graphs in the context of previous efforts in probability theory, the authors shed some light on the correspondence between combinatorial aspects (that is, graph limits via weak regularity) and the probabilistic viewpoint of sampling: Graph and hypergraph limits provide an infinite random graph model that has the property of array exchangeability. An infinite sequence with random elements is said to be exchangeable if the distribution of the sequence is invariant under finite permutations of the elements. Similar definition applies for arrays, where the distribution has to be invariant under row and column permutations of finite range.

According to de Finetti's Theorem, the exchangeable $\{0,1\}$-valued sequences are mixtures of i.i.d.'s in the sense that first a random $p \in [0,1]$ is generated obeying some distribution, and subsequently independent Bernoulli(p) trials are executed for the entries. In the case of arrays the characterization is a bit more complicated. To obtain a typical exchangeable random infinite matrix, a $[0,1]$-valued symmetric function f on the unit square has to be fixed. First, one has to color the rows and columns by an exchangeable sequence with $[0,1]$ values, and afterwards independent coin tosses for every ij entry with success probability $f(x_i, x_j)$ have to be carried out, where x_i is the color of the ith row, and x_j of the jth column obtained in the preliminary stage of the generating process. This construction clearly produces an exchangeable matrix, and it is non-trivial to show that there is basically no other way to generate one. Similar level-wise independent coloring schemes apply for higher dimensional cases, the proof that this construction delivers every exchangeable random r-array was proved first by de Finetti [41] (for $\{0,1\}$ entries) and Hewitt and Savage [72] (in the case of general Polish space entries) for $r = 1$, independently by Aldous [2] and Hoover [74] for $r = 2$, and by Kallenberg [78] for $r \geq 3$.

The meta-correspondence between the distribution of exchangeable arrays and graph limits presented in [43] turned out to be quite fruitful, we enlist some classes where an analogous starting approach was taken in order to describe the limit objects.

- Diaconis, Holmes, and Janson [44] considered threshold graphs, one of the characterizations for them is that there are weights $\{w_i\}$ and a threshold t such that $ij \in E(G)$ if and only if $w_i + w_j \geq t$. The authors extracted the limit object using exchangeability principles, the limits can be represented as symmetric, increasing, $\{0,1\}$-valued functions on the unit square $[0,1]^2$.

- Diaconis, Holmes, and Janson [45] regarded interval graphs that are graphs, that can be equipped with a collection of intervals on the real line corresponding to the nodes in such a way that an edge is present if and only if the intersection of the intervals assigned to its vertices is nonempty. The limit space consists here of probability measures on the triangle $\{(a,b) \mid 0 \leq a \leq b \leq 1\}$.

- Partially ordered sets were analyzed by Janson [76] with respect to their limiting behavior, the limit objects can be represented by measurable kernels $W\colon \mathcal{S}\times\mathcal{S} \to [0,1]$ on an ordered probability space $(\mathcal{S}, \mathcal{F}, \mu, \prec)$ (that is, $\prec$ is a partial order satisfying that $\{(x,y) \mid x \prec y\}$ is $\mathcal{F} \times \mathcal{F}$-measurable), such that (i) $W(x,y) > 0$ implies $x \prec y$, and (ii) $W(x,y)W(y,z) > 0$ implies $W(x,z) = 1$. Later Hladky, Mathe, Patel, and Pikhurko [73] proved that $\mathcal{S}$ can always be chosen to be the uniform measure on $[0, 1]$ with the usual total ordering.

- Hoppen, Kohayakawa, Moreira, Ráth, and Menezes Sampaio [75] dealt with the nature of limits of permutation sequences, it turned out that the limits can be described as measurable functions $Z\colon [0, 1]^2 \to [0, 1]$, where $Z(x, .)$ is a distribution function for any $x \in [0, 1]$, and it holds for each $y \in [0, 1]$ that $\int_0^1 Z(x, y)\mathrm{d}x = y$.

As yet another analogy to classical probability theory Chatterjee and Varadhan [37] developed a large deviation (LD) result for dense random graphs, showing among other things that the conditional distribution of the Erdős-Rényi graphs does not resemble the unconditioned when it is imposed for it to have a large number of triangles, furthermore it was shown that there is a double phase transition dependent on the triangle density. Based on these concepts Chatterjee and Diaconis [36] studied the LD behavior of exponential graph models with distribution $p(G) \sim \exp(\sum_{i=1}^{k} \beta_i f_i(G))$, where the f_i's are graph parameters, and deduced that in many cases this setup leads to Erdős-Rényi graphs in the long run.

On a further note, in order to highlight the impact of the concept we mention that results by Olhede and Wolfe [100] and Orbanz and Roy [101] provide evidence that principles developed in graph limit theory are entering the toolbox of Bayesian statistics.

Most of the above are presented comprehensively in the recent survey textbook by Lovász [89] with all the necessary details and further background.

Property testing for dense graphs

We give an overview of the state of the art in property testing in the dense graph and hypergraph model. Roughly said, two main groups of graph properties emerged that are testable: the partition problem type, and the forbidden subgraph type. The k-colorability property can be regarded as being in the intersection of the two groups.

A partition problem is given by two non-negative k-vectors serving as lower and upper bounds for the desired class sizes, and two matrices of order k that contain the desired bounds for the edge densities between classes. A graph satisfies the partition problem property, if it has a k-partition of its vertex set satisfying the bounds in the description of the problem. This family contains several natural properties such as k-colorability, having a cut with at least ρn^2 edges, having a bisection into equal classes such that at least ρn^2 edges are crossing, or having a clique of size at least ρn. It was shown in [66] that all these properties are 2-sided testable, with a sample complexity at most $\tilde{O}(\frac{k^2}{\varepsilon})^{2k+8}$, with some problem specific approaches leading to somewhat better upper bounds.

Some special cases corresponding to MAX-rCSP were subsequently improved in [59], [17], [14], and [98] which we discussed in the part on PTAS earlier in this section. In the direction of partition problems testable with 1-sided testers we mention the results that deliver improvements for proper k-colorability [5], for satisfiability for [7] graphs, and the generalization of colorability for hypergraphs [40]. Recently, Sohler [114] proved the generalization for hypergraphs regarding satisfiability using a refined analysis on the sampling procedure building on the previous approaches to obtain the upper bound $O((1/\varepsilon)\log(1/\varepsilon))$ on the sample complexity, which is optimal up to the log-factor. For the general hypergraph partitioning problem given by density arrays instead of matrices Fischer, Matsliah, and Shapira [57] showed testability with at most polynomial sample complexity. The latter approach also incorporated the study of simultaneous partitioning of a finite number of hypergraphs perhaps of different rank making it suitable for testing regularity instances. That means a graph that satisfies the property has not only a vertex partition where the edge densities are as desired, but the partition is also ε-regular.

We turn to review the developments for properties corresponding to forbidden substructures. As also discussed above, Alon and Shapira [7] showed that the property of an rCSP (referred to by them as (r,d)-Function-SAT) being completely satisfiable is testable using a sample size of $O(\varepsilon^{-2})$, generalizing the result obtained for graph and hypergraph node colorability problems. A more general class of properties of graphs formulated as special $\exists\forall$-type first order logical formulas had been shown to be testable by Alon, Fischer, Krivelevich, and Szegedy [13]. The property studied there is for a tester equivalent to the existence of a vertex coloring such that the colored graph avoids some finite family of vertex colored subgraphs, and hereby encompasses triangle-freeness as well as k-colorability. The approach, as several others below, use the Regularity Lemma and give impractical bounds for the sample complexity.

It was shown subsequently by Alon and Shapira [11] that every monotone graph property is testable, followed by a complete characterization of properties admitting a one-sided tester [10] by the same authors. They showed that every hereditary property (closed under taking induced subgraphs) is testable, and the converse, that every property that is testable without false negatives is semi-hereditary. The same authors investigated properties obtained by forbidding a single subgraph, all these are testable as a consequence of the Graph Removal Lemma, but H-freeness is polynomially testable exactly when H is bipartite [8]. The case for forbidden induced subgraphs is quite different, these require super-polynomial tests except for the cases K_1,K_2,P_3(path with 3 edges), and C_4(cycle of length 4) [9]. The first two are trivially polynomially testable, the same was shown by Alon and Fox [4] for P_3, while the situation for C_4 remains open. Also, in Alon and Fox [4] it was demonstrated that perfectness is not polynomially testable. As a result regarding hypergraphs, it was established that hereditary uniform hypergraph properties are testable by Rödl and Schacht [104], this was later generalized to edge colored hypergraphs by Austin and Tao [24].

A complete combinatorial characterization of the class of testable properties was given by Alon, Fischer, Newman, and Shapira [16] in terms of graph regularity. As seen above, a regularity instance problem is a singleton partition problem with the additional requirement that a satisfying partition of a graph has also to be regular. The result of [16] can be roughly formulated as follows: If a property is testable, then it is equivalent to satisfying a regularity instance taken from a finite family, which characteristic is referred to as regular reducibility. Another characterization by Fischer and Newman [56] says that testability of any property is equivalent to the estimability (or testability) of the numerical parameter of the edit distance to the property.

We mention that in all these cases one can employ canonical testers with at most polynomial blow-up in the sample complexity, see Goldreich and Trevisan [65]. Further, some well-studied properties are trivially testable: planarity is not satisfied by any dense graph, having a prefect matching (or a triangle) can always be achieved by adding a very little number of edges. Finally we would like to point out that although most natural properties are testable, some are not: Fischer and Matsliah [55] showed that it is not possible to tell the difference by oblivious (size-independent) sampling between the cases that a graph is a disjoint union of two isomorphic ones, or is ε-far from being such.

One main open question in the area is whether there is any property with exponential sample complexity, since every positive result comes either with a polynomial or a tower-type upper bound. Also, a characterization for properties with polynomial sample complexity remains to be established. For a more complete picture we refer to the surveys by Goldreich [63], and by Rubinfeld and Shapira [108].

1.5 Organization of the thesis

The organization of the thesis is as follows. In Chapter 2 we formally define CSP formulas, give the connection to hypergraphs, and prove the representation theorem for colored hypergraph and CSP limits by means of exchangeability. Chapter 3 presents the concept of regular partitions of graphs, we prove some new version of the Regularity Lemma, afterwards we introduce graph parameter and property testing, and demonstrate some results regarding them via limit theory. Chapter 4 continues with the study of ground state energies in the limit space. Next, in Chapter 5, we establish the connection between colored hypergraph sequences and hypergraphons via the ultralimit method, and use this link to derive results on a generalized version of ground state energies, as well as a hypergraph version of the Regularity Lemma. In the subsequent Chapter 6 we describe the concept of nondeterministic testing, prove its equivalence to traditional testing for graphs and hypergraphs, which is followed by some variants and applications. We conclude the thesis with Chapter 7 summarizing our results, and giving an outlook on directions of further research.

CHAPTER 2

Limits of hypergraphs and CSPs via exchangeability

2.1 Introduction

In this chapter we develop a general framework for the CSP problems which depends only on the principles of the array exchangeability without a recourse to the weakly regular partitions used hitherto in the general graph and hypergraph settings. Those fundamental techniques and results were worked out in a series of papers by Borgs, Chayes, Lovász, Sós, Vesztergombi and Szegedy [30],[32],[91], and [94] for graphs including connections to statistical physics and complexity theory, and were subsequently extended to hypergraphs by Elek and Szegedy [49] via the ultralimit method. The central concept of the r-graph convergence is defined through convergence of sub-r-graph densities, or equivalently through weak convergence of probability measures on the induced sub-r-graph yielded by uniform vertex sampling. Our line of work particularly relies on ideas presented in [43] by Diaconis and Janson, where the authors shed some light on the correspondence between combinatorial aspects (that is, graph limits via weak regularity) and the probabilistic viewpoint of sampling: Graph limits provide a countably infinite random graph model that has the property of exchangeability. The precise definitions, references, and results will be given below, here we only formulate our main contribution informally: We prove a representation theorem for compact colored r-uniform directed hypergraph limits. This says that every limit object in this setup can be transformed into a measurable function on the $(2^r - 2)$-dimensional unit cube that takes values from the probability distributions on the compact color palette, see Theorem 2.2.10 below. This extends the result of Diaconis and Janson [43], and of Lovász and Szegedy [93]. As an application, the description of the limit space of rCSP formulas is presented subsequent to the aforementioned theorem.

2.1.1 Definitions and preliminaries

General notation We provide here some standard notation that will be used throughout the thesis. The sets $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Z}^+$, $\mathbb{R}$, and $\mathbb{R}^+$ stand for the natural numbers, integers, positive integers, real numbers, and positive real numbers, respectively. The set $[n]$ for some positive integer n is a short form for $\{1, 2, \dots, n\}$. The expression S_n is the symmetric group of degree n. The $\{0, 1\}$-valued function $\mathbb{1}_A$ is the indicator function of the subset A of the domain, $\mathbb{1}_a$ is the indicator of an element a of the domain. For a countable set S and an integer r the subset $\binom{S}{r}$ of the power set of S constitutes of the subsets of S of cardinality r, $\binom{n}{r}$ is the cardinality of $\binom{[n]}{r}$ for non-negative integers n and r, and as a common convention $\binom{0}{0} = 1$. The symbols $\mathbb{P}$ and $\mathbb{E}$ denote the probability of an event, and the expectation of a random variable, respectively. For a real measurable function f defined on the measure space $(\Omega, \mathcal{A}, \mu)$, and $p \geq 1$ the p-norm of f is $\|f\|_p = (\int_\Omega |f|^p \mathrm{d}\mu)^{1/p}$, and the supremum or the maximum norm of f is $\|f\|_\infty = \sup_{\omega \in \Omega} |f(\omega)|$. If we talk about a measurable space $(\Omega, \mathcal{A})$, then the space of continuous functionals on Ω is $C(\Omega)$, whereas $\mathcal{P}(\Omega)$ denotes the space of regular probability measures on Ω.

CSP formulas We will consider the objects called *rCSP formulas* that are used to define instances of the decision and optimization problems called rCSP and MAX-rCSP, respectively. In the current framework a formula consists of a variable set and a set of boolean or integer valued functions. Each of these functions is defined on a subset of the variables, and the sets of possible assignments of values to the variables are uniform. Additionally, it will be required that each of the functions, which we will call *constraints* in what follows, depend exactly on r of the variables.

For the treatment of an rCSP (of a MAX-rCSP) corresponding to a certain formula we are required to simultaneously evaluate all the constraints of the formula by assigning values to each of the variables in the variable set. If we deal with an rCSP optimization problem on some combinatorial structure, say on graphs, then the formula corresponding to a certain graph has to be constructed according to the optimization problem in question. The precise definitions will be provided next.

Let $r \geq 1$, K be a finite set, and f be a boolean-valued function $f\colon K^r \to \{0, 1\}$ on r variables (or equivalently $f \subseteq K^r$). We call f a *constraint-type on K in r variables*, $\mathcal{C} = \mathcal{C}(K, r)$ denotes the set of all such objects.

Definition 2.1.1 (rCSP formula)**.** *Let $V = \{x_1, x_2, \dots, x_n\}$ be the set of variables, $x_e = (x_{e_1}, \dots, x_{e_r}) \in V^r$ and f a constraint-type on K in r variables. We call an n-variable function $\omega = (f; x_e)\colon K^V \to \{0, 1\}$ with $\omega(l_1, \dots, l_n) = f(l_{e_1}, \dots, l_{e_r})$ a constraint on V in r variables determined by an r-vector of constrained variables and a constraint type.*

We call a collection F of constraints on $V(F) = \{x_1, x_2, \dots, x_n\}$ in r variables of type $\mathcal{C}(K, r)$ for some finite K an rCSP formula.

Two constraints $(f_1; x_{e_1})$ and $(f_2; x_{e_2})$ are said to be equivalent if they constrain the same r variables, and their evaluations coincide, that is, whenever there exists a $\pi \in S_r$

such that $e_1 = \pi(e_2)$ (here π permutes the entries of e_2) and $f_1 = \hat{\pi}(f_2)$, where $[\hat{\pi}(f)](l) = f(\pi(l))$. Two formulas F_1 and F_2 are equivalent if there is a bijection ϕ between their variable sets such that there is a one-to-one correspondence between the constraints of F_1 and F_2 such that the corresponding pairs $(f_1; x_{e_1}) \in F_1$ and $(f_2; x_{e_2}) \in F_2$ satisfy $(f_1; \phi(x_{e_1})) \equiv (f_2; x_{e_2})$.

In the above definition the set of states of the variables in $V(F)$ denoted by K is not specified for each formula, it will be considered as fixed similar to the dimension r whenever we study a family of rCSPs. We say that F is symmetric, if it contains only constraints with constraint-types which are invariant under the permutations of the constrained variables. When we relax the notion of the types to be real or $\mathcal{K}$-valued functions on K^r with $\mathcal{K}$ being a compact space, then we speak of weighted rCSP formulas.

The motivation for the name CSP formula is immediately clear from the notation used in Definition 2.1.1 if we consider constraints to be satisfied at some point in K^n, whenever they evaluate to 1 there. Most problems defined on these objects ask for parameters that are, in the language of real analysis, global or conditioned extreme values of the objective function given by an optimization problem and a formula. A common assumption is that equivalent formulas should get the same parameter value.

Definition 2.1.2 (MAX-rCSP)**.** *Let F be an rCSP formula over a finite domain K. Then the MAX-rCSP value of F is given by*

$$\text{MAX}-r\text{CSP}(F) = \max_{l \in K^{V(F)}} \sum_{\omega=(f;x_e)\in F} \omega(l), \tag{2.1}$$

and F is satisfiable, if $\text{MAX}-r\text{CSP}(F) = |\{\, \omega \mid \omega = (f; x_e) \in F \,\}|$.

Such problems are for example MAX-CUT, fragile MAX-rCSP, MAX-3-SAT, and Not-All-Equal-3-SAT, where only certain constraint types are allowed for instances, or MAX-BISECTION, where additionally only specific value assignments are permitted in the above maximization. In general, formulas can also be viewed as directed r-graphs, whose edges are colored with constraint types (perhaps with multiple types), and we will exploit this representation in our analysis.

Typically, we will not store and recourse to an rCSP formula F as it is given by its definition above, but we will only consider the r-array tuple $(F^z)_{z\in K^r}$, where

$$F^z(e) = \sum_{\phi\in S_r} \sum_{(f;x_{\phi(e)})\in F} f(z_{\phi(1)}, \dots, z_{\phi(r)}) \tag{2.2}$$

for each $e \in [n]^r$. The data set $(F^z)_{z\in K^r}$ is called the *evaluation representation of F*, or short $\mathfrak{eval}(F)$, we regard $\mathfrak{eval}(F)$ as a parallel colored (with colors from $[q]^r$) multi-r-graph, see below. We impose a boundedness criteria on CSPs that will apply throughout the chapter, that means we fix $d \geq 1$ for good, and require that $\|F^z\|_\infty \leq d$ for every $z \in K^r$ and CSP formula F with $\mathfrak{eval}(F) = (F^z)_{z\in K^r}$ in consideration. We note that for each $z \in K^r$,

$e \in [n]^r$ and $\phi \in S_r$ we have the symmetry $F^z(e) = F^{z_{\phi(1)},\dots,z_{\phi(r)}}(e_{\phi(1)},\dots,e_{\phi(r)})$, also, on the diagonal F^z is 0.

The main motivation for what follows in the current chapter originates from the aim to understand the long-range behavior of a randomly evolving rCSP formula together with the value of the corresponding MAX-rCSP by making sense of a limiting distribution. This task is equivalent to presenting a structural description of rCSP limits analogous to the graph limits of [91].

The convergence notion should agree with parameter estimation via sampling. In this setting we pick a set of variables of fixed size at random from the constrained set $V(F)$ of an rCSP formula F defined on a large number of variables, and ask for all the constraints in which the sampled variables are involved and no other, this is referred to as the induced subformula on the sample. Then we attempt to produce some quantitative statement about the parameter value of the original formula by relying only on the estimation of the corresponding value of the parameter on a subformula, see Definition 3.2.1 in Chapter 3 below, where we deal in-depth with testability.

Having formally introduced the notion of rCSP formulas and MAX-rCSP, we proceed to the outline of the necessary notation and to the analysis of the limit behavior regarding the colored hypergraph models that are used to encode these formulas.

Graphs and hypergraphs Simple r-uniform hypergraphs, r-graphs in short, on n vertices forming the family Π^r_n are subsets G of $\binom{V(G)}{r}$ with $|V(G)| = n$, where $V(G)$ is the vertex set of G and the size of such a G is n, and the elements of $\binom{[n]}{r}$ are r-edges, or simply edges. We will regard hypergraphs also as symmetric subsets of $K_r([n]) = [n]^r \setminus \mathrm{diag}([n]^r)$, or explicitly write A_G for the adjacency array of G. The rank of an r-uniform hypergraph is r.

For some arbitrary set $\mathcal{K}$, denote by $\Pi(\mathcal{K}) = \Pi^r(\mathcal{K})$ the set of all unlabeled $\mathcal{K}$-colored undirected r-uniform hypergraphs, in short $(\mathcal{K}, r)$-graphs, where we will suppress r in the notation, when it is clear which r is meant (alternatively, $\Pi(\mathcal{K})$ denotes the isomorphism classes of the node labeled respective objects). Also, let $\Pi^r_n(\mathcal{K})$ denote the subset of $\Pi^r(\mathcal{K})$ whose elements have vertex cardinality n. Let k be a positive integer, then $\Pi^r_n([k])$ (also denoted by $\Pi^{r,k}_n$ for simplicity) is the set of k-colored r-graphs of size n, their elements are partitions $G = (G^\alpha)_{\alpha\in[k]}$ of $\binom{[n]}{r}$ into k classes, we say that color α assigned to $e \in \binom{[n]}{r}$ (short $G(e) = \alpha$) whenever $e \in G^\alpha$. In this sense simple r-graphs are regarded as 2-colored. Additionally we have to introduce the special color ι for loop edges that are multisets of $[n]$ with cardinality r having at least one element that has a multiplicity at least 2. Let $\tilde\Pi(\mathcal{K})$, $\tilde\Pi^r(\mathcal{K})$, $\tilde\Pi^r_n(\mathcal{K})$, and $\tilde\Pi^r_n([k])$ denote the directed counterparts of the above families.

Let $\mathbb{G}(k, G)$ denote the random induced subgraph of G, that is an element of any of the above a families, on the set $T \subset V(G)$ that is chosen uniformly among the subsets of $V(G)$ of cardinality k. Note that $\mathbb{G}(k, G)$ belongs to the same family as G.

For a finite set S, let $\mathfrak{h}_0(S)$ and $\mathfrak{h}(S)$ denote the power set and the set of nonempty subsets of S, respectively, and $\mathfrak{h}(S, m)$ the set of nonempty subsets of S of cardinality at most m, also $\mathfrak{h}_0(S, m) = \mathfrak{h}(S, m) \cup \{\emptyset\}$. A $2^r - 1$-dimensional real vector $x_{\mathfrak{h}(S)}$ denotes

$(x_{T_1}, \dots, x_{T_{2^r-1}})$, where $T_1, \dots, T_{2^r-1}$ is a fixed ordering of the nonempty subsets of S with $T_{2^r-1} = S$, for a permutation π of the elements of S the vector $x_{\pi(\mathfrak{h}(S))}$ means $(x_{\pi^*(T_1)}, \dots, x_{\pi^*(T_{2^r-1})})$, where π^* is the action of π permuting the subsets of S. Similar conventions apply when x is indexed by other set families.

We say that an edge-$\mathcal{K}$-colored sequence $(G_n)_{n=1}^{\infty}$ converges if for every k the sequence of random graphs $(\mathbb{G}(k, G_n))_{n=1}^{\infty}$ converges in distribution, see more precisely Definition 2.2.3 below. Consider a graph parameter, and suppose we know a sample size q that is sufficient for an additive $\varepsilon/2$ estimation for its value through sampling. Then, without any further knowledge about the smoothness of the parameter we can assert, that the values for two graphs are ε-close if their sampled graphs of size q coincide with a probability larger than ε. This happens when they have approximately the same number of isomorphic copies of any graph of size q as subgraphs, with the appropriate normalization determined by the size of the two original graphs. This remark justifies the focus on counting substructures.

The main content of the current chapter starts with the general setting of edge-$\mathcal{K}$-colored r-uniform hypergraphs in Section 2.2 where we obtain a characterization of the limit space for compact $\mathcal{K}$ in Theorem 2.2.10. The rCSP setting will be considered as a special case in this topic whose limit characterization will be derived as Corollary 2.3.2 in Section 2.3. Some of the basic cases are already settled regarding the representation of the limits prior to our work. Without claiming to provide a complete list of previously established results we refer to Lovász and Szegedy [91], [93], [89] for the $r = 2$, general $\mathcal{K}$, undirected case, to Elek and Szegedy [49] for the general r, $\mathcal{K} = \{0, 1\}$, undirected case; and Diaconis and Janson [43] for $r = 2$, $\mathcal{K} = \{0, 1\}$, directed and undirected case. These three approaches are fundamentally different in their proof methodology (they rely on weak regularity, ultralimits, and exchangeability principles, respectively) and were respectively further generalized or applied by Zhao [119] to general r, by Aroskar [21] to the weighted case, and by Austin [23] to general r.

2.2 Limits of $\mathcal{K}$-colored r-uniform directed hypergraphs

Let $\mathcal{K}$ be a compact Polish space and $r \geq 1$ an integer. Recall that a space $\mathcal{K}$ is called Polish if it is a separable completely metrizable topological space. In what follows we will consider the limit space of $\mathcal{K}$-colored r-uniform directed hypergraphs, or with different words r-arrays with non-diagonal entries from $\mathcal{K}$, and the diagonal entries are occupied by a special element which also can be in $\mathcal{K}$, but in general this does not have to be the case.

2.2.1 Definition of convergence

Let C denote space $C(\mathcal{K})$ of continuous functionals on $\mathcal{K}$, and let $\mathcal{F} \subset C$ be a countable generating set with $\|f\|_\infty \leq 1$ for each $f \in \mathcal{F}$, that is, the linear subspace generated by

$\mathcal{F}$ is dense in C in the L^∞-norm. We define the homomorphism densities next.

Definition 2.2.1. *Let $\mathcal{K}$ be an arbitrary set or space, and $C(\mathcal{K})$ be the set of continuous functionals on $\mathcal{K}$. If for some $r \geq 1$ $F \in \Pi^r(C(\mathcal{K}))$ is a uniform directed graph with $V(F) = [k]$ and $G \in \Pi^r(\mathcal{K})$, then the homomorphism density of F in G is defined as*

$$t(F,G) = \frac{1}{|V(G)|^k} \sum_{\phi\colon [k]\to V(G)} \prod_{i_1,\dots,i_r=1}^{k} F(i_1,\dots,i_r)(G(\phi(i_1),\dots,\phi(i_r))). \tag{2.3}$$

The injective homomorphism density $t_{\mathrm{inj}}(F,G)$ is defined similarly, with the difference that the average of the products is taken over all injective ϕ maps (normalization changes accordingly).

In the special case when $\mathcal{K}$ is finite we can associate to the elements of $\Pi(\mathcal{K})$ functions in $\Pi(C(\mathcal{K}))$ through replacing the edge colors in $\mathcal{K}$ by the corresponding indicator functions. This way we have in accordance with the above definition that the density of $F \in \Pi^r_k(\mathcal{K})$ in $G \in \Pi^r(\mathcal{K})$ is

$$t(F,G) = \frac{1}{|V(G)|^k} \sum_{\phi\colon [k]\to V(G)} \prod_{e\in[k]^r} \mathbb{1}_{G^{F(e)}}(\phi(e)),$$

where the product is 1 when $G[\phi([k])] = F$ as vertex labeled graphs, otherwise it is 0. With other words, $t(F,G)$ gives the number of labeled induced isomorphic copies of F in G normalized by the number of all labeled induced subgraphs on k vertices.

The variant for uncolored r-graphs that was used in Lovász [89] and related works is the t^* density that counts the copies of F in G appearing as subgraphs, but not necessarily as induced ones. More precisely, for uncolored (or 2-colored) F and G it is defined as

$$t^*(F,G) = \frac{1}{|V(G)|^k} \sum_{\phi\colon [k]\to V(G)} \prod_{e\in F} \mathbb{1}_{G}(\phi(e)), \tag{2.4}$$

or alternatively as the probability that $F \subset G[\phi([k])]$ for an uniformly chosen map $\phi\colon [k] \to V(G)$. Its advantage over the induced density is that in the simple graph case the corresponding density in the limit has a more compact form.

Note that if $F, G \in \Pi^{r,q}$, then $t_{\mathrm{inj}}(F,G) = \mathbb{P}(\mathbb{G}(k,G) = F)$.

Let the map τ be defined as $\tau(G) = (t(F,G))_{F\in\Pi(\mathcal{F})} \in [0,1]^{\Pi(\mathcal{F})}$ for each $G \in \Pi(\mathcal{K})$. We set $\Pi(\mathcal{K})^* = \tau(\Pi(\mathcal{K})) \subset [0,1]^{\Pi(\mathcal{F})}$, and $\overline{\Pi(\mathcal{K})^*}$ to the closure of $\Pi(\mathcal{K})^*$. Also, let $\Pi(\mathcal{K})^+ = \{\, (\tau(G), 1/|V(G)|) \mid G \in \Pi(\mathcal{K}) \,\} \subset [0,1]^{\Pi(\mathcal{F})} \times [0,1]$, and let $\overline{\Pi(\mathcal{K})^+}$ be the closure of $\Pi(\mathcal{K})^+$. The function $\tau^+(G) = (\tau(G), 1/|V(G)|)$ will be useful for our purposes, because, opposed to τ, it is injective, which can be verified easily. For any $F \in \Pi(\mathcal{F})$ the function $t(F,.)$ on $\Pi(\mathcal{K})$ can be uniquely continuously extended to a function $t(F,.)$ on $\overline{\Pi(\mathcal{K})^+}$, this is due to the compactness of $[0,1]^{\Pi(\mathcal{F})} \times [0,1]$. For an element $\Gamma \in \overline{\Pi(\mathcal{K})^+} \setminus \Pi(\mathcal{K})^+$, let $t(F,\Gamma)$ for $F \in \Pi(\mathcal{F})$ denote the real number in $[0,1]$ that is the coordinate of Γ corresponding to F.

The functions $\tau_{\text{inj}}(G)$ and $\tau^+_{\text{inj}}(G)$, and the sets $\Pi_{\text{inj}}(\mathcal{K}) = \tau_{\text{inj}}(\Pi(\mathcal{K}))$ and $\Pi_{\text{inj}}(\mathcal{K})^+$ are defined analogously. It was shown in [91] that

$$|t_{\text{inj}}(F,G) - t(F,G)| \leq \frac{|V(F)|^2 \|F\|_\infty}{2|V(G)|} \tag{2.5}$$

for any pair $F \in \Pi(C)$ and $G \in \Pi(\mathcal{K})$.

The precise definition of convergence will be given right after the next theorem which is analogous to a result of [93].

Theorem 2.2.2. *Let $(G_n)_{n=1}^\infty$ be a random sequence in $\Pi(\mathcal{K})$ with $|V(G_n)|$ tending to infinity in probability. Then the following are equivalent.*

(1) The sequence $(\tau^+(G_n))_{n=1}^\infty$ converges in distribution in $\Pi(\mathcal{K})^+$.

(2) For every $F \in \Pi(\mathcal{F})$, the sequence $(t(F,G_n))_{n=1}^\infty$ converges in distribution.

(3) For every $F \in \Pi(C)$, the sequence $(t(F,G_n))_{n=1}^\infty$ converges in distribution.

(4) For every $k \geq 1$, the sequence $(\mathbb{G}(k,G_n))_{n=1}^\infty$ of random elements of $\Pi(\mathcal{K})$ converges in distribution.

If any of the above apply, then the respective limits in (2) *and* (3) *are $t(F,\Gamma)$ with Γ being a random element of $\overline{\Pi(\mathcal{K})^+}$ given by* (1)*, and also $\Gamma \in \overline{\Pi(\mathcal{K})^+} \setminus \Pi(\mathcal{K})^+$, almost surely.*

If $t(F,G_n)$ in (2) *and* (3) *is replaced by $t_{\text{inj}}(F,G_n)$, then the equivalence of the four statements still persists and the limits in* (2) *and* (3) *are $t(F,\Gamma)$.*

If every G_n is concentrated on some single element of $\Pi(\mathcal{K})$ (non-random case), then the equivalence holds with the sequences in (1), (2)*, and* (3) *being numerical instead of distributional, while* (4) *remains unchanged.*

Proof. The equivalence of (1) and (2) is immediate. The implication from (3) to (2) is also clear by definition.

For showing that (2) implies (3), we consider first an arbitrary $F \in \Pi(\langle\mathcal{F}\rangle)$, where $\langle\mathcal{F}\rangle$ is the linear space generated by $\mathcal{F}$. Then there exist $F^1,\dots,F^l \in \Pi(\mathcal{F})$ on the same vertex set as F, say $[k]$, and $\lambda_1,\dots,\lambda_l \in \mathbb{R}$ such that for any non-random $G \in \Pi(\mathcal{K})$ and $\phi\colon [k] \to V(G)$ it holds that $\prod_{i_1,\dots,i_r=1}^k F(i_1,\dots,i_r)(G(\phi(i_1),\dots,\phi(i_r))) = \sum_{j=1}^l \lambda_j \prod_{i_1,\dots,i_r=1}^k F^j(i_1,\dots,i_r)(G(\phi(i_1),\dots,\phi(i_r)))$. So therefore we can express $t(F,G) = \sum_{j=1}^l \lambda_j t(F^j,G)$. We return to the case when G_n is random. The weak convergence of $t(F,G_n)$ is equivalent to the convergence of each of its moments, its tth moment can be written by the linearity of the expectation as a linear combination of a finite number of mixed moments of the densities corresponding to $F^1,\dots,F^l \in \Pi(\mathcal{F})$. For an arbitrary vector of non-negative integers $\alpha = (\alpha_1,\dots,\alpha_l)$, let F^α be the element of $\Pi(\mathcal{F})$ that is the disjoint union α_1 copies of F^1, α_2 copies of F^2, and so on. It holds that $t(F^1,G_n)^{\alpha_1}\dots t(F^l,G_n)^{\alpha_l} = t(F^\alpha,G_n)$, and in particular the two random variables on the two sides are equal in expectation. Condition (2) implies that $\mathbb{E}[t(F^\alpha,G_n)]$ converges for each α, therefore the mixed moments of the $t(F^i,G_n)$ densities and the moments of $t(F, G_n)$ also do. This implies that $t(F, G_n)$ also converges in distribution for any $F \in \Pi(\langle F \rangle)$. Now let $F' \in \Pi(C)$ and $\varepsilon > 0$ be arbitrary, and $F \in \Pi(\langle F \rangle)$ on the same vertex set $[k]$ as F' be such that its entries are at most ε-far in L^∞ from the corresponding entries of F'. Then

$$
\begin{aligned}
&|t(F', G) - t(F, G)| \\
&= \Bigg| \frac{1}{|V(G)|^k} \sum_{\phi\colon [k] \to V(G)} \prod_{i_1, \dots, i_r = 1}^{k} F(i_1, \dots, i_r)(G(\phi(i_1), \dots, \phi(i_r))) \\
&\qquad\qquad - \prod_{i_1, \dots, i_r = 1}^{k} F'(i_1, \dots, i_r)(G(\phi(i_1), \dots, \phi(i_r))) \Bigg| \\
&= \frac{1}{|V(G)|^k} \sum_{\phi\colon [k] \to V(G)} \sum_{i_1, \dots, i_r = 1}^{k} \Bigg| \prod_{(j_1, \dots, j_r) < (i_1, \dots, i_r)} F(j_1, \dots, j_r)(G(\phi(j_1), \dots, \phi(j_r))) \Bigg| \\
&\qquad \Bigg| \prod_{(j_1, \dots, j_r) > (i_1, \dots, i_r)} F'(j_1, \dots, j_r)(G(\phi(j_1), \dots, \phi(j_r))) \Bigg| \\
&\qquad \Big| F(i_1, \dots, i_r)(G(\phi(i_1), \dots, \phi(i_r))) - F'(i_1, \dots, i_r)(G(\phi(i_1), \dots, \phi(i_r))) \Big| \\
&\le k^r \varepsilon \max\{(\|F'\|_\infty + \varepsilon)^{k^r - 1}, 1\}
\end{aligned}
$$

for any $G \in \Pi(\mathcal{K})$ (random or non-random), which implies (3), as $\varepsilon > 0$ was chosen arbitrarily.

We turn to show the equivalence of (3) and (4). Let $\Pi_k(\mathcal{K}) \subset \Pi(\mathcal{K})$ the set of elements of $\Pi(\mathcal{K})$ with vertex cardinality k. The sequence $(\mathbb{G}(k, G_n))_{n=1}^{\infty}$ converges in distribution exactly when for each continuous function $f \in C(\Pi_k(\mathcal{K}))$ on $\Pi_k(\mathcal{K})$ the expectation $\mathbb{E}[f(\mathbb{G}(k, G_n))]$ converges as $n \to \infty$. For each $F \in \Pi(C)$ and $\alpha \ge 1$, the function $t_{\mathrm{inj}}^{\alpha}(F, G)$ is continuous on $\Pi^{|V(F)|}(\mathcal{K})$ and $t_{\mathrm{inj}}(F, G) = t_{\mathrm{inj}}(F, \mathbb{G}(|V(F)|, G))$, so (3) follows from (4).

For showing the other direction, that (3) implies (4), let us fix $k \ge 1$. We claim that the linear function space $M = \langle t(F, .) | F \in \Pi(C) \rangle \subset C(\Pi_k(\mathcal{K}))$ is an algebra containing the constant function, and that it separates any two elements of $\Pi_k(\mathcal{K})$. It follows that $\langle t(F, .) | F \in \Pi(C) \rangle$ is L^∞-dense in $C(\Pi_k(\mathcal{K}))$ by the Stone-Weierstrass theorem, which implies by our assumptions that $\mathbb{E}[f(\mathbb{G}(k, G_n))]$ converges for any $f \in C(\Pi_k(\mathcal{K}))$, since we know that $\mathbb{E}[t_{\mathrm{inj}}(F, \mathbb{G}(k, G_n))] = \mathbb{E}[t_{\mathrm{inj}}(F, G_n)]$ whenever $|V(F)| \le k$. We will see in a moment that $t_{\mathrm{inj}}(F, .) \in M$, convergence of $\mathbb{E}[t_{\mathrm{inj}}(F, G_n)]$ follows from (2.5) and the requirement that $|V(G_n)|$ tends to infinity in probability.

Now we turn to show that our claim is indeed true. For two graphs $F_1, F_2 \in \Pi(C)$ we have $t(F_1, G) t(F_2, G) = t(F_1 F_2, G)$ for any $G \in \Pi_k(\mathcal{K})$, where the product $F_1 F_2$ denotes the disjoint union of the two C-colored graphs. Also, $t(F, G) = 1$ for the graph F on one node with a loop colored with the constant 1 function. Furthermore we have that

$\hom(F,G) = k^{|V(F)|}t(F,G) \in M$ for $|V(G)| = k$, so therefore

$$\mathrm{inj}(F,G) = \sum_{\mathcal{P} \text{ partition of } V(F)} (-1)^{|V(F)|-|\mathcal{P}|} \prod_{S\in\mathcal{P}} (|S|-1)! \hom(F/\mathcal{P},G) \quad \in M,$$

where $\mathrm{inj}(F,G) = t_{\mathrm{inj}}(F,G)k(k-1)\dots(k-|V(F)|+1)$ and $F/\mathcal{P} \in \Pi^{|\mathcal{P}|}(C)$ whose edges are colored by the product of the colors of F on the edges between the respective classes of $\mathcal{P}$. This equality is the consequence of the Mobius inversion formula, and that $\mathrm{inj}(F,G) = \sum_{\mathcal{P} \text{ partition of } V(F)} \hom(F/\mathcal{P},G)$. For G and F defined on the node set $[k]$ recall that

$$\mathrm{inj}(F,G) = \sum_{\phi\in S_k} \prod_{i_1,\dots,i_r=1}^{k} F(i_1,\dots,i_r)(G(\phi(i_1),\dots,\phi(i_r))). \tag{2.6}$$

Now fix $G_1, G_2 \in \Pi_k(\mathcal{K})$ and let $F \in \Pi_k(C)$ such that $\{F(i_1,\dots,i_r)(G_j(l_1,\dots,l_r))\}$ are algebraically independent elements of $\mathbb{R}$ (such an F exists, we require a finite number of algebraically independent reals, and can construct each entry of F by polynomial interpolation). If G_1 and G_2 are not isomorphic, than for any possible node-relabeling for G_2 there is at least one term in the difference $\mathrm{inj}(F,G_1) - \mathrm{inj}(F,G_2)$ written out in the form of (2.6) that does not get canceled out, so therefore $\mathrm{inj}(F,G_1) \neq \mathrm{inj}(F,G_2)$.

We examine the remaining statements of the theorem. Clearly, $\Gamma \notin \Pi(\mathcal{K})^+$, because $|V(G_n)| \to \infty$ in probability. The results for the case where the map in (1) and the densities in (2) and (3) are replaced by the injective version are yielded by (2.5), the proof of the non-random case carries through in a completely identical fashion. □

We are now ready to formulate the definition of convergence in $\Pi(\mathcal{K})$.

Definition 2.2.3. *If $(G_n)_{n=1}^{\infty}$ is a sequence in $\Pi(\mathcal{K})$ with $|V(G_n)| \to \infty$ and any of the conditions above of Theorem 2.2.2 hold, then we say that $(G_n)_{n=1}^{\infty}$ converges.*

We would like to add that, in the light of Theorem 2.2.2, the convergence notion is independent from the choice of the family $\mathcal{F}$.

The next lemma gives information about the limit behavior of the sequences where the vertex set cardinality is constant.

Lemma 2.2.4. *Let $(G_n)_{n=1}^{\infty}$ be a random sequence in $\Pi_k(\mathcal{K})$, and additionally be such that for every $F \in \Pi(\mathcal{F})$ the sequences $(t_{\mathrm{inj}}(F,G_n))_{n=1}^{\infty}$ converge in distribution. Then there exists a random $H \in \Pi_k(\mathcal{K})$, such that for every $F \in \Pi(\mathcal{F})$ we have $t(F,G_n) \to t(F,H)$ and $t_{\mathrm{inj}}(F,G_n) \to t_{\mathrm{inj}}(F,H)$ in distribution.*

Proof. We only sketch the proof. The distributional convergence of $(G_n)_{n=1}^{\infty}$ follows the same way as in the proof of Theorem 2.2.2, the part about condition (2) implying (3) together with the part stating that (3) implies (4). The existence of a random H satisfying the statement of the lemma is obtained by invoking the Riesz representation theorem for positive functionals. □

2.2.2 Exchangeable arrays

The correspondence analogous to the approach of Diaconis and Janson in [43] will be established next between the elements of the limit space $\overline{\Pi(\mathcal{K})^+}$ that is compact, and the extreme points of the space of random exchangeable infinite r-arrays with entries in $\mathcal{K}$. These are arrays, whose distribution is invariant under finite permutations of the underlying index set.

Definition 2.2.5 (Exchangeable r-array)**.** *Let $(H(e_1,\dots,e_r))_{1\le e_1,\dots,e_r<\infty}$ be an infinite r-array of random entries from a Polish space $\mathcal{K}$. We call the random array separately exchangeable if*

$$(H(e_1,\dots,e_r))_{1\le e_1,\dots,e_r<\infty}$$

has the same probability distribution as

$$(H(\rho_1(e_1),\dots,\rho_r(e_r)))_{1\le e_1,\dots,e_r<\infty}$$

for any $\rho_1,\dots,\rho_r \in S_{\mathbb{N}}$ collection of finite permutations, and jointly exchangeable (or simply exchangeable), if the former holds only for all $\rho_1 = \dots = \rho_r \in S_{\mathbb{N}}$.

It is clear that if we consider a measurable function $f\colon [0,1]^{\mathfrak{h}_0([r])} \to \mathcal{K}$, and independent random variables uniformly distributed on $[0,1]$ that are associated with each of the subsets of $\mathbb{N}$ of cardinality at most r, then by plugging in these random variables into f for every $e \in \mathbb{N}^r$ in the right way suggested by a fixed natural bijection $l_e\colon e \to [r]$, the result will be an exchangeable random r-array. The shorthand $\mathrm{Samp}(f)$ denotes this law of the infinite directed r-hypergraph model generated by f.

The next theorem, states that all exchangeable arrays with values in $\mathcal{K}$ arise from some f in the former way.

Theorem 2.2.6. *[78] Let $\mathcal{K}$ be a Polish space. Every $\mathcal{K}$-valued exchangeable r-array $(H(e))_{e\in\mathbb{N}^r}$ has law equal to* $\mathrm{Samp}(f)$ *for some measurable $f\colon [0,1]^{\mathfrak{h}_0([r])} \to \mathcal{K}$, that is, there exists a function f, so that if $(U_s)_{s\in\mathfrak{h}_0(\mathbb{N},r)}$ are independent uniform* $[0,1]$ *random variables, then*

$$H(e) = f(U_\emptyset, U_{\{e_1\}}, U_{\{e_2\}}, \dots, U_{\bar{e}\setminus\{e_r\}}, U_{\bar{e}}) \tag{2.7}$$

for every $e = (e_1,\dots,e_r) \in \mathbb{N}^r$, where $H(e)$ are the entries of the infinite r-array.

If H in the above theorem is invariant under permuting its coordinates, then the corresponding function f is invariant under the coordinate permutations that are induced by the set permuting S_r-actions.

Theorem 2.2.6 was first proved by de Finetti [41] (in the case $\mathcal{K} = \{0,1\}$) and by Hewitt and Savage [72] (in the case of general $\mathcal{K}$) for $r = 1$, independently by Aldous [2] and Hoover [74] for $r = 2$, and by Kallenberg [78] for arbitrary $r \ge 3$. For equivalent formulations, proofs and further connections to related areas see the recent survey of Austin [23].

In general, there are no symmetry assumptions on f, in the directed case $H(e)$ might differ from $H(e')$, even if e and e' share a common base set. In this case these two entries do not have the property of conditional independence over a σ-algebra given by some lower dimensional structures, that means for instance the independence over $\{ U_\alpha \mid \alpha \subsetneq e \}$ for an exchangeable r-array with law Samp(f) given by a function f as above.

With the aid of Theorem 2.2.6 we will provide a form of representation of the limit space $\overline{\Pi(\mathcal{K})^+}$ through the points of the space of random infinite exchangeable r-arrays. The correspondence will be established through a sequence of theorems analogous to the ones stated and proved in [43, Section 2 to 5], combined with the compactification argument regarding the limit space from [93], see also [89, Chapter 17.1] for a more accurate picture. The proofs in our case are mostly ported in a straightforward way, if not noted otherwise we direct the reader for the details to [43].

Let $\mathcal{L}_\infty = \mathcal{L}_\infty(\mathcal{K})$ denote the set of all node labeled countably infinite $\mathcal{K}$-colored r-uniform directed hypergraphs. Set the common vertex set of the elements of $\mathcal{L}_\infty$ to $\mathbb{N}$, and define the set of $[n]$-labeled $\mathcal{K}$-colored r-uniform directed hypergraphs as $\mathcal{L}_n = \mathcal{L}_n(\mathcal{K})$. Every $G \in \mathcal{L}_n$ can be viewed as an element of $\mathcal{L}_\infty$ simply by adding isolated vertices to G carrying the labels $\mathbb{N} \setminus [n]$ in the uncolored case, and the arbitrary but fixed color $c \in \mathcal{K}$ to edges incident to these vertices in the colored case, therefore we think about $\mathcal{L}_n$ as a subset of $\mathcal{L}_\infty$ (and also of $\mathcal{L}_m$ for every $m \geq n$). Conversely, if G is a (random) element of $\mathcal{L}_\infty$, then by restricting G to the vertices labeled by $[n]$, we get $G|_{[n]} \in \mathcal{L}_n$. If G is a labeled or unlabeled $\mathcal{K}$-colored r-uniform directed hypergraph (random or not) with vertex set of cardinality n, then let $\hat{G}$ stand for the random element of $\mathcal{L}_n$ (and also $\mathcal{L}_\infty$) which we obtain by first throwing away the labels of G (if there where any), and then apply a random labeling chosen uniformly from all possible ones with the label set $[n]$.

A random element of $\mathcal{L}_\infty$ is *exchangeable* analogously to Definition 2.2.5 if its distribution is invariant under any permutation of the vertex set $\mathbb{N}$ that only moves finitely many vertices, for example infinite hypergraphs whose edge-colors are independently identically distributed are exchangeable. An element of $\mathcal{L}_\infty$ can also be regarded as an infinite r-array whose diagonal elements are colored with a special element ι that is not contained in $\mathcal{K}$, therefore the corresponding r-arrays will be $\mathcal{K} \cup \{\iota\}$-colored.

The next theorem relates the elements of $\Pi(\mathcal{K})^+$ to exchangeable random elements of $\mathcal{L}_\infty$.

Theorem 2.2.7. *Let $(G_n)_{n=1}^\infty$ be a random sequence in $\Pi(\mathcal{K})$ with $|V(G_n)|$ tending to infinity in probability. Then the following are equivalent.*

(1) $\tau^+(G_n) \to \Gamma$ in distribution for a random $\Gamma \in \overline{\Pi(\mathcal{K})^+} \setminus \Pi(\mathcal{K})^+$.

(2) $\hat{G}_n \to H$ in distribution in $\mathcal{L}_\infty(\mathcal{K})$, where H is a random element of $\mathcal{L}_\infty(\mathcal{K})$.

If any of these hold true, then $\mathbb{E}t(F, \Gamma) = \mathbb{E}t_{\text{inj}}(F, H|_{[k]})$ for every $F \in \Pi_k(C)$, and also, H is exchangeable.

Proof. If $G \in \Pi(\mathcal{K})$ is deterministic and $F \in \Pi_k(\mathcal{F})$ with $|V(G)| \geq k$ then $Et_{\text{inj}}(F, \hat{G}|_{[k]}) = t_{\text{inj}}(F, G)$, where the expectation E is taken with respect to the random (re-)labeling $\hat{G}$

of G. For completeness we mention that for a labeled, finite G the quantity $t(F,G)$ is understood as $t(F,G')$ with G' being the unlabeled version of G, also, F in $t(F,G)$ is always regarded a priori as labeled, however the densities of isomorphic labeled graphs in any graph coincide. If we consider G to be random, then by the fact that $0 \leq t(F,G) \leq 1$ (as $\|F\|_\infty \leq 1$) we have that $|\mathbb{E}Et_{\text{inj}}(F,\hat{G}|_{[k]}) - \mathbb{E}t_{\text{inj}}(F,G)| \leq \mathbb{P}(|V(G)| < k)$ for $F \in \Pi_k(\mathcal{F})$.

Assume (1), then the above implies, together with $\mathbb{P}(|V(G_n)| < k) \to 0$ and (1), that $\mathbb{E}Et_{\text{inj}}(F,\hat{G}_n|_{[k]}) \to \mathbb{E}t(F,\Gamma)$ (see Theorem 2.2.2). This implies that $\hat{G}_n|_{[k]} \to H_k$ in distribution for some random $H_k \in \mathcal{L}_k$ with $\mathbb{E}t_{\text{inj}}(F,H_k) = \mathbb{E}t(F,\Gamma)$, see Lemma 2.2.4, furthermore, with appealing to the consistency of the H_k graphs in k, there exists a random $H \in \mathcal{L}_\infty$ such that $H|_{[k]} = H_k$ for each $k \geq 1$, so (1) yields (2).

Another consequence is that H is exchangeable: the exchangeability property is equivalent to the vertex permutation invariance of the distributions of $H|_{[k]}$ for each k. This is ensured by the fact that $H|_{[k]} = H_k$, and H_k is the weak limit of a vertex permutation invariant random sequence, for each k.

For the converse direction we perform the above steps in the reversed order using

$$|\mathbb{E}Et(F,\hat{G}_n|_{[k]}) - \mathbb{E}t(F,G_n) \leq \mathbb{P}(|V(G_n)| < k)$$

again in order to establish the convergence of $(\mathbb{E}t(F,G_n))_{n=1}^\infty$. Theorem 2.2.2 certifies now the existence of the suitable random $\Gamma \in \overline{\Pi(\mathcal{K})^+} \setminus \Pi(\mathcal{K})^+$, this shows that (2) implies (1).

□

We built up the framework in the preceding statements Theorem 2.2.2 and Theorem 2.2.7 in order to formulate the following theorem, which is the crucial ingredient to the desired representation of limits.

Theorem 2.2.8. *There is a one-to-one correspondence between random elements of* $\overline{\Pi(\mathcal{K})^+} \setminus \Pi(\mathcal{K})^+$ *and random exchangeable elements of* $\mathcal{L}_\infty$*. Furthermore, there is a one-to-one correspondence between elements of* $\overline{\Pi(\mathcal{K})^+} \setminus \Pi(\mathcal{K})^+$ *and extreme points of the set of random exchangeable elements of* $\mathcal{L}_\infty$*. The relation is established via the equalities* $\mathbb{E}t(F,\Gamma) = \mathbb{E}t_{\text{inj}}(F,H|_{[k]})$ *for every* $F \in \Pi_k(C)$ *for every* $k \geq 1$.

Proof. Let Γ a random element of $\overline{\Pi(\mathcal{K})^+} \setminus \Pi(\mathcal{K})^+$. Then by definition of $\Pi(\mathcal{K})^+$ there is a sequence $(G_n)_{n=1}^\infty$ in $\Pi(\mathcal{K})$ with $|V(G_n)| \to \infty$ in probability such that $\tau^+(G_n) \to \Gamma$ in distribution in $\Pi(\mathcal{K})^+$. By virtue of Theorem 2.2.7 there exists a random $H \in \mathcal{L}_\infty$ so that $\hat{G}_n \to H$ in distribution in $\mathcal{L}_\infty$, and H is exchangeable. The distribution of $H|_{[k]}$ is determined by the numbers $\mathbb{E}t_{\text{inj}}(F,H|_{[k]})$, see Theorem 2.2.2, Lemma 2.2.4, and the arguments therein, and these numbers are provided by the correspondence.

For the converse direction, let H be random exchangeable element of $\mathcal{L}_\infty$. Then let $G_n = H|_{[n]}$, we have $G_n \to H$ in distribution, and also $\hat{G}_n \to H$ in distribution by the vertex permutation invariance of G_n as a node labeled object. Again, we appeal to Theorem 2.2.7, so $\tau^+(G_n) \to \Gamma$ for a Γ random element of $\overline{\Pi(\mathcal{K})^+} \setminus \Pi(\mathcal{K})^+$, which is determined completely by the numbers $\mathbb{E}t(F,\Gamma)$ that are provided by the correspondence, see Theorem 2.2.7.

The second version of the relation between non-random Γ's and extreme points of exchangeable elements is proven similarly, the connection is given via $t(F,\Gamma) = \mathbb{E}t_{\text{inj}}(F, H|_{[k]})$ between the equivalent objects.

□

The characterization of the aforementioned extreme points in Theorem 2.2.8 was given [43] in the uncolored graph case, we state it next for our general setting, but refrain from giving the proof here, as it is completely identical to [43, Theorem 5.5.].

Theorem 2.2.9. *[43] The distribution of H that is an exchangeable random element of $\mathcal{L}_\infty$ is exactly in that case an extreme point of the set of exchangeable measures if the random objects $H|_{[k]}$ and $H|_{\{k+1,\dots\}}$ are probabilistically independent for any $k \geq 1$. In this case the representing function f from Theorem 2.2.6 does not depend on the variable corresponding to the empty set.*

2.2.3 Graphons as limit objects

In this subsection we encounter the first time the notion of a *graphon* (the term introduced in [30]) in the thesis, that is a centerpiece of the whole work. We start by describing the basic case of the definition corresponding to simple r-graphs, and then proceed to the more general setting.

Kernels, graphons and sampling Let the *r-kernel* space $\hat{\Xi}_0^r$ denote the space of the bounded measurable functions of the form $W\colon [0,1]^{\mathfrak{h}([r],r-1)} \to \mathbb{R}$, and the subspace Ξ_0^r of $\hat{\Xi}_0^r$ the symmetric r-kernels that are invariant under coordinate permutations π^* induced by some $\pi \in S_r$, that is $W(x_{\mathfrak{h}([r],r-1)}) = W(x_{\pi^*(\mathfrak{h}([r],r-1))})$ for each $\pi \in S_r$. We will refer to this invariance in the thesis both for r-kernels and for measurable subsets of $[0,1]^{\mathfrak{h}([r])}$ as *r-symmetry*. The kernels $W \in \Xi_I^r$ take their values in some interval I, for $I = [0,1]$ we call these special symmetric r-kernels *r-graphons*, and their set Ξ^r. In what follows, λ as a measure always denotes the usual Lebesgue measure in $\mathbb{R}^d$, where the dimension d is everywhere clear from the context.

Analogously to the graph case we define for a positive integer q the space of *q-colored r-graphons* by $\Xi^{r,q}$ whose elements are referred to as $W = (W^\alpha)_{\alpha\in[q]}$ with each of the W^α components being r-graphons. The special color ι that stands for the absence of colors has to be also employed in this setting as rectangles on the diagonal might correspond to loop edges, see below for the case when we represent a q-colored r-graph as a graphon. The corresponding r-graphon W^ι is $\{0,1\}$-valued. Furthermore, W has to satisfy $\sum_{\alpha\in[q]} W^\alpha(x) = 1 - W^\iota(x)$ everywhere on $[0,1]^{\mathfrak{h}([r],r-1)}$, so if $W^\iota \equiv 0$, then $\sum_{\alpha\in[q]} W^\alpha(x) = 1$. For $x \in [0,1]^{\mathfrak{h}([r])}$ the expression $W(x)$ denotes the color at x, we have $W(x) = \alpha$ whenever $\sum_{i=1}^{\alpha-1} W^i(x_{\mathfrak{h}([r],r-1)}) \leq x_{[r]} \leq \sum_{i=1}^{\alpha} W^i(x_{\mathfrak{h}([r],r-1)})$. The space of q-colored r-digraphons denoted by $\tilde{\Xi}^{r,q}$ is defined analogously with elements $W = (W^{(\alpha_{T_1},\dots,\alpha_{T_l})})_{\alpha_{T_1},\dots,\alpha_{T_l}\in[q]}$, where $T_1,\dots,T_l$ are the orderings of the elements of $[r]$. The symmetry assumption here is $W^{(\alpha_{T_1},\dots,\alpha_{T_l})}(x_{\pi(\mathfrak{h}([r],r-1))}) = W^{(\alpha_{\pi(T_1)},\dots,\alpha_{\pi(T_l)})}(x_{\mathfrak{h}([r],r-1)})$ for $\pi \in S_r$.

For $k \geq 1$ and $W \in \Xi^{r,q}$ the random q-colored r-graph $\mathbb{G}(k, W)$ is generated as follows. The vertex set of $\mathbb{G}(k, W)$ is $[k]$, first we have to pick uniformly a random point $(X_S)_{S \in \mathfrak{h}([k],r-1)} \in [0,1]^{\mathfrak{h}([k],r-1)}$, then conditioned on this choice we conduct independent trials to determine the color of each edge $e \in \binom{[k]}{r}$ with the distribution given by $\mathbb{P}_e(\mathbb{G}(k, W)(e) = \alpha) = W^\alpha(X_{\mathfrak{h}(e,r-1)})$ corresponding to e. Recall that ι is a special color which we want to avoid in most cases during the sampling process, therefore we will highlight the conditions that have to be imposed on the above random variables so that $\mathbb{G}(k, W) \in \Pi^{r,q}$. We also call $\mathbb{G}(k, W)$ a W-random graph, in the special case when W is constant this is the Erdős-Rényi random graph.

For $F \in \Pi_k^{r,q}$, the F-density of W is defined as $t(F, W) = \mathbb{P}(F = \mathbb{G}(k, W))$, which can be written following the above definition of the sampled random graph as

$$t(F, W) = \int_{[0,1]^{\mathfrak{h}([k],r-1)}} \prod_{e \in \binom{[k]}{r}} W^{F(e)}(x_{\mathfrak{h}(e,r-1)}) \mathrm{d}\lambda(x_{\mathfrak{h}([k],r-1)}). \tag{2.8}$$

If $W \in \Xi^r$, and $F \in \Pi^r$ is simple, then the above formula reduces to

$$t(F, W) = \int_{[0,1]^{\mathfrak{h}([k],r-1)}} \prod_{e \in E(F)} W(x_{\mathfrak{h}(e,r-1)}) \prod_{e \notin E(F)} (1 - W(x_{\mathfrak{h}(e,r-1)})) \mathrm{d}\lambda(x_{\mathfrak{h}([k],r-1)}), \tag{2.9}$$

also the non-induced version related to (2.4) is defined as $t^*(F, W) = \mathbb{P}(F \subset \mathbb{G}(k, W))$, or alternatively as

$$t^*(F, W) = \int_{[0,1]^{\mathfrak{h}([k],r-1)}} \prod_{e \in E(F)} W(x_{\mathfrak{h}(e,r-1)}) \mathrm{d}\lambda(x_{\mathfrak{h}([k],r-1)}). \tag{2.10}$$

We turn to definition of the most general type of limit object we will use in this thesis. Let $\mathcal{K}$ be a compact Polish space, and $W\colon [0,1]^{\mathfrak{h}([r])} \to \mathcal{K}$ be a measurable function, we will refer to such an object as a $(\mathcal{K}, r)$-*digraphon*, their set is denoted by $\tilde{\Xi}^r(\mathcal{K})$. Note that there are no symmetry assumptions in this general case, if additionally W is r-symmetric, then we speak about $(\mathcal{K}, r)$-*graphons*, their space is $\Xi^r(\mathcal{K})$. For $\mathcal{K} = \{0, 1\}$ the set $\mathcal{P}(\mathcal{K})$ can be identified with the $[0, 1]$ interval encoding the success probabilities of Bernoulli trials to get the common r-graphon form as a function $W\colon [0,1]^{2^r-2} \to [0,1]$ employed in [49].

The density of a $\mathcal{K}$-colored graph $F \in \tilde{\Pi}_k(C(\mathcal{K}))$ in the $(\mathcal{K}, r)$-digraphon W is defined analogously to (2.3) and (2.8) as

$$t(F, W) = \int_{[0,1]^{\mathfrak{h}([k],r)}} \prod_{e \in [k]^r} F(e)(W(x_{\mathfrak{h}(e,r)})) \mathrm{d}\lambda(x_{\mathfrak{h}([k],r)}). \tag{2.11}$$

For $k \geq 1$ and an undirected $W \in \Xi^r(\mathcal{K})$ the random $(\mathcal{K}, r)$-graph $\mathbb{G}(k, W)$ is defined on the vertex set $[k]$ by selecting a uniform random point $(X_S)_{S \in \mathfrak{h}([k],r)} \in [0,1]^{\mathfrak{h}([k],r)}$ that enables the assignment of the color $W(X_{\mathfrak{h}(e)})$ to each edge $e \in \binom{[k]}{r}$. For a directed W the sample point is as above, the color of the directed edge $e \in [k]^r$ is $W(X_{\mathfrak{h}(e)})$, but in this case the ordering of the power set of the base set $\overline{e}$ of e matters in contrast to the undirected situation and is given by e, as W is not necessarily r-symmetric.

Additionally we define the averaged sampled r-graph for $\mathcal{K} \subset \mathbb{R}$ denoted by $\mathbb{H}(k, W)$, it has vertex set $[k]$, and the weight of the edge $e \in \binom{[k]}{r}$ is the conditional expectation $\mathbb{E}[W(X_{\mathfrak{h}(e)}) \mid X_{\mathfrak{h}(e,1)}]$, and therefore the random r-graph is measurable with respect to $X_{\mathfrak{h}([k],1)}$. We will use the compact notation X_i for $X_{\{i\}}$ for the elements of the sample indexed by singleton sets.

We define the random exchangeable r-array H_W in $\mathcal{L}_\infty$ as the element that has law $\mathrm{Samp}(W)$ for the $(\mathcal{K}, r)$-digraphon W, as in Theorem 2.2.6. Furthermore, we define $\Gamma_W \in \overline{\Pi(\mathcal{K})^+} \setminus \Pi(\mathcal{K})^+$ to be the element associated to H_W through Theorem 2.2.7.

Now we are able to formulate the representation theorem for $\mathcal{K}$-colored r-uniform directed hypergraph limits using the representation of exchangeable arrays, see (Theorem 2.2.6). It is an immediate consequence of Theorem 2.2.7 and Theorem 2.2.8 above.

Theorem 2.2.10. *Let $(G_n)_{n=1}^\infty$ be a sequence in $\Pi(\mathcal{K})$ with $|V(G_n)| \to \infty$ such that for every $F \in \Pi(\mathcal{F})$ the sequence $t(F, G_n)$ converges. Then there exists a function $W\colon [0,1]^{\mathfrak{h}([r])} \to \mathcal{K}$ (that is $W \in \Xi^r(\mathcal{K})$) such that $t(F, G_n) \to t(F, \Gamma_W)$ for every $F \in \Pi(\mathcal{F})$. In the directed case when the sequence is in $\tilde{\Pi}(\mathcal{K})$, then the corresponding limit object W is in $\tilde{\Xi}^r(\mathcal{K})$.*

We mention that $t(F, \Gamma_W) = t(F, W)$ for every $F \in \Pi(C(\mathcal{K}))$ and $W \in \Xi^r(\mathcal{K})$. Alternatively we can also use the form $W\colon [0,1]^{\mathfrak{h}([r],r-1)} \to \mathcal{P}(\mathcal{K})$ for $(\mathcal{K}, r)$-graphons and digraphons in $\Xi^r(\mathcal{K})$ whose values are probability measures, this representation was applied in [93].

Directed graphs and graphons In previous works, for example in [43], the limit object of a sequence of simple directed graphs without loops was represented by a 4-tuple of 2-graphons $(W^{(0,0)}, W^{(1,0)}, W^{(0,1)}, W^{(1,1)})$ that satisfies $\sum_{i,j} W^{(i,j)}(x,y) = 1$ and $W^{(1,0)}(x,y) = W^{(0,1)}(y,x)$ for each $(x,y) \in [0,1]^2$. A generalization of this representation can be given in our case of the $\Pi(\mathcal{K})$ limits the following way. We only present here the case when $\mathcal{K}$ is a continuous space, the easier finite case can be dealt with analogously.

We have to fix a Borel probability measure μ on $\mathcal{K}$, we set this to be the uniform distribution if $\mathcal{K} \subset \mathbb{R}^d$ is a domain or $\mathcal{K}$ is finite. The limit space consists of collections of $(\mathbb{R}, r)$-kernels $W = (W^u)_{u \in \mathcal{U}}$, where $\mathcal{U}$ is the set of all functions $u\colon S_r \to \mathcal{K}$. Additionally, W has to satisfy $\int_{\mathcal{U}} W^u(x) \mathrm{d}\mu^{\otimes S_r}(u) = 1$ and $0 \le W^{\pi^*(u)}(x) = W^u(x_{\pi^*(\mathfrak{h}([r],r-1))})$ for each $\pi \in S_r$ and $x \in [0,1]^{\mathfrak{h}([r],r-1)}$. As before, the action π^* of π on $[0,1]^{\mathfrak{h}([r],r-1)}$ is the induced coordinate permutation by π, with the unit cubes coordinates indexed by non-trivial subsets of $[r]$. Without going into further details we state the connection between the limit form spelled out above and that in Theorem 2.2.10. It holds

$$\int_U W^u((x_{\mathfrak{h}([r],r-1)}) \mathrm{d}\mu^{\otimes S_r}(u) = \mathbb{P}[(W(x_{\pi^*(\mathfrak{h}([r],r-1))}, Y)_{\pi \in S_r}) \in U]$$

for every measurable $U \subset \mathcal{U}$ and $x \in [0,1]^{\mathfrak{h}([r],r-1)}$, where Y is uniform on $[0,1]$, and the W on the right-hand side is a $(\mathcal{K},r)$-digraphon, whereas on the left we have the corresponding representation as a (possibly infinite) collection of $(\mathbb{R},r)$-kernels.

Naive graphons, graphs as graphons In several applications, among them some that are presented in the current thesis, it is more convenient to use a naive form for the limit representation, from which the limit element in question is not decisively retrievable. The naive limit space consists of *naive* $(\mathcal{K},r)$*-graphons* $\overline{W}\colon [0,1]^r \to \mathcal{P}(\mathcal{K})$, where now the arguments of W are indexed with elements of $[r]$. From a proper r-graphon $W\colon [0,1]^{\mathfrak{h}([r])} \to \mathcal{K}$ we get its naive counterpart by averaging, that is the $\mathcal{K}$-valued random variable $\mathbb{E}[W(x_{\mathfrak{h}([r],1)}, U_{\mathfrak{h}([r],r-1)\setminus\mathfrak{h}([r],1)}, Y)|Y]$ has distribution $\overline{W}(x_1,\dots,x_r)$, where $(U_S)_{S\in\mathfrak{h}([r],r-1)\setminus\mathfrak{h}([r],1)}$ and Y are i.i.d. uniform on $[0,1]$.

On a further note we introduce *averaged naive* $(\mathcal{K},r)$*-graphons* for the case, when $\mathcal{K} \subset \mathbb{R}$, these are of the form $\tilde{W}\colon [0,1]^r \to \mathbb{R}$ and are given by complete averaging, that is $\mathbb{E}[W(x_1,\dots,x_r, U_{\mathfrak{h}([r])\setminus\mathfrak{h}([r],1)}, Y)] = \tilde{W}(x_1,\dots,x_r)$, where $(U_S)_{S\in\mathfrak{h}([r])\setminus\mathfrak{h}([r],1)}$ are i.i.d. uniform on $[0,1]$. A *naive r-kernel* is a real-valued, bounded function on $[0,1]^r$, or equivalently on $[0,1]^{\mathfrak{h}([r],1)}$.

We can associate to each $G \in \Pi_n^r(\mathcal{K})$ an element $W_G \in \Xi^r(\mathcal{K} \cup \{\iota\})$ by subdividing the unit r-cube $[0,1]^{\mathfrak{h}([r],1)}$ into n^r small cubes the natural way and defining the function $W'\colon [0,1]^{\mathfrak{h}([r],1)} \to \mathcal{K}$ that takes the value $G(i_1,\dots,i_r)$ on $[\frac{i_1-1}{n},\frac{i_1}{n}]\times\dots\times[\frac{i_r-1}{n},\frac{i_r}{n}]$ for distinct $i_1,\dots,i_r$, and the value ι on the remaining diagonal cubes, note that these functions are naive $(\mathcal{K},r)$-graphons. Then we set $W_G(x_{\mathfrak{h}([r],r-1)}) = W'(p_{\mathfrak{h}([r],1)}(x_{\mathfrak{h}([r],r-1)}))$, where $p_{\mathfrak{h}([r],1)}$ is the projection to the suitable coordinates. The special color ι here stands for the absence of colors has to be employed in this setting as rectangles on the diagonal correspond to loop edges. The corresponding r-graphon W^ι is $\{0,1\}$-valued. The sampled random r-graphs $\mathbb{G}(k,W_G)$ and $\mathbb{H}(k,W_G)$ from the naive r-graphons are defined analogously to the general case. If $\mathcal{K} \subset \mathbb{R}$, then note that $\mathbb{H}(k,W_G) = \mathbb{G}(k,W_G)$ for every G, because the colors of G are all point measures.

Note that $t(F,G) = t(F,W_G)$, and

$$|t_{\mathrm{inj}}(F,G) - t(F,W_G)| \le \frac{\binom{k}{2}}{n-\binom{k}{2}} \tag{2.12}$$

for each $F \in \Pi_k^{r,q}$, hence the representation as naive graphons is compatible in the sense that $\lim_{n\to\infty} t_{\mathrm{inj}}(F,G_n) = \lim_{n\to\infty} t(F,W_{G_n})$ for any sequence $(G_n)_{n=1}^\infty$ with $|V(G_n)|$ tending to infinity. This implies that $d_{\mathrm{tv}}(\mathbb{G}(k,G_n),\mathbb{G}(k,W_{G_n})) \to 0$ as n tends to infinity.

We remark that naive and averaged naive versions in the directed case are defined analogously.

2.3 Representation of rCSP formulas as hypergraphs, and their convergence

In this section we elaborate on how homomorphism and sampling is meant in the CSP context. Recall Definition 2.1.1 for the way how we perceive rCSP formulas.

Let F be an rCSP formula on the variable set $\{x_1, \dots, x_n\}$ over an arbitrary domain K, and let $F[x_{i_1}, \dots, x_{i_k}]$ be the induced subformula of F on the variable set $\{x_{i_1}, \dots, x_{i_k}\}$. Let $\mathbb{G}(k, F)$ denote the random induced subformula on k uniformly chosen variables from the elements of $V(F)$.

It is clear using the terminology of Definition 2.2.5 that the relation $\omega = (f; x_e) \in F[x_{i_1}, \dots, x_{i_k}]$ is equivalent to the relation

$$\phi(\omega) = (f_\phi, x_{\phi(e)}) \in F[x_{i_1}, \dots, x_{i_k}] \tag{2.13}$$

for permutations $\phi \in S_r$, where $f_\phi(l_1, \dots, l_r) = f(l_{\phi(1)}, \dots, l_{\phi(r)})$ and $\phi(e) = (e_{\phi(1)}, \dots, e_{\phi(r)})$. This emergence of symmetry will inherently be reflected in the limit space, we will demonstrate this shortly.

Special limits Let $K = [q]$. As we mentioned above, in general it is likely not to be fruitful to consider formula sequences as sequences of $C(K, r)$-colored r-graphs obeying certain symmetries due to constraint splitting. However, in the special case when each r-set of variables carries exactly one constraint we can derive a meaningful representation, MAX-CUT is an example. A direct consequence of Theorem 2.2.10 is the following.

Corollary 2.3.1. *Let $r \geq 1$, and K be a finite set, further, let $\mathcal{K} \subset C(K, r)$, so that $\mathcal{K}$ is permutation invariant. Let $(F_n)_{n=1}^\infty$ be a sequence of rCSP formulas with $|V(F_n)|$ tending to infinity, and each r-set of variables in each of the formulas carries exactly one constraint of type $\mathcal{K}$. If for every formula H obeying the same conditions the sequences $(t(H, F_n))_{n=1}^\infty$ converge as $(\mathcal{K}, r)$-graphs, then there exists a $(\mathcal{K}, r)$-digraphon $W\colon [0,1]^{\mathfrak{h}([r])} \to \mathcal{K}$ such that $t(H, F_n) \to t(H, W)$ as n tends to infinity for every H as above.*

Additionally, W satisfies for each $x \in [0,1]^{\mathfrak{h}([r])}$ and $\pi \in S_r$ that $W(x_{\pi(\mathfrak{h}([r]))}) = \hat{\pi}(W(x_{\mathfrak{h}([r])}))$, where $\hat{\pi}$ is the action of π on constraint types in $C(K, r)$ that permutes the rows and columns of the evaluation table according to π, that is $(\hat{\pi}(f))(l_1, \dots, l_r) = f(l_{\pi(1)}, \dots, l_{\pi(r)})$.

General limits via evaluation In the general case of rCSP formulas we regard them as their evaluation representation $\mathfrak{eval}$.

For $|K| = q$ we identify the set of rCSP formulas with the set of arrays whose entries are the sums of the evaluation tables of the constraints on r-tuples, that is F with $V(F) = [n]$ corresponds to a map $\mathfrak{eval}(F)\colon [n] \times \dots \times [n] \to \{0, 1, \dots, d\}^{([q]^r)}$ that obeys the symmetry condition given after (2.2). This will be the way throughout the thesis we look at these objects from here on. It seems that storing the whole structure of an rCSP formula does not provide any further insight, in fact splitting up constraints would produce non-identical formulas in a complete structure representation, which does not seem sensible.

We denote the set $\{0, 1, \ldots, d\}^{([q]^r)}$ by L for simplicity, which one could also interpret as the set of multisets whose base set is $[q]^r$ and whose elements have multiplicity at most d. This perspective allows us to treat rCSPs as directed r-uniform hypergraphs whose edges are colored by the aforementioned elements of L, and leads to a representation of rCSP limits that is derived from the general representation of the limit set of $\Pi(L)$. We will show in a moment that the definition of convergence in the previous subsection given by densities of functional-colored graphs is basically identical to the convergence via densities of sub-multi-hypergraphs in the current case.

The definition of convergence for a general sequence of rCSP formulas, or equivalently of elements of $\Pi(L)$, was given in Definition 2.2.3. We describe here the special case for parallel multicolored graphs, see also [93].

Consider the evaluation representation of the rCSP formulas now as r-graphs whose oriented edges are parallel multicolored by $[q]^r$. The map $\psi\colon \mathfrak{eval}(H) \to \mathfrak{eval}(F)$ is a homomorphism between two rCSP formulas H and F if it maps edges to edges of the same color from the color set $[q]^r$ and is consistent when restricted to be a mapping between vertex sets, $\psi'\colon V(H) \to V(F)$, for simple graphs instead of CSP formulas this is the multigraph homomorphism notion.

Let H be an rCSP formula, and let $\tilde{H}$ be the corresponding element in $C(L)$ on the same vertex set such that if the color on the fixed edge e of H is the q-sized r-array $(H^z(e))_{z\in[q]^r}$ with the entries being non-negative integers, then the color of $\tilde{H}$ at e is $\prod_{z\in[q]^r} x_z^{H^z(e)}$. More precisely, for an element $A \in L$ the value is given by

$$[\tilde{H}(e)](A) = \prod_{z\in[q]^r} A(z)^{H^z(e)}.$$

The linear space generated by the set

$$\tilde{L} = \left\{ \prod_{z\in[q]^r} x_z^{d_z} \mid 0 \le d_{z_1,\ldots,z_r} \le d \right\}$$

forms an L^∞-dense subset in $C(L)$, therefore Theorem 2.2.2 applies, and for a sequence $(G_n)_{n=1}^\infty$ requiring the convergence of $t(F, \mathfrak{eval}(G_n))$ for all $F = \tilde{H}$ with $H \in \Pi(L)$ provides one of the equivalent formulations of the convergence of rCSP formulas in the subformula density sense with respect to the evaluations.

The limit object will be given by Theorem 2.2.10 as the space of measurable functions $W\colon [0,1]^{\mathfrak{h}([r])} \to L$, where, as in the general case, the coordinates of the domain of W are indexed by the non-empty subsets of $[r]$. In our case, not every possible W having this form will serve as a limit of some sequence, the above mentioned symmetry in (2.13) of the finite objects is inherited in the limit.

We state now the general evaluation rCSP version of Theorem 2.2.10.

Corollary 2.3.2. *Let $(F_n)_{n=1}^{\infty}$ be a sequence of rCSP formulas that evaluate to at most d on all r-tuples with $|V(F_n)| \to \infty$ such that for every finite rCSP formula H obeying the same upper bound condition the sequence $(t(\tilde{H}, \mathfrak{eval}(F_n)))_{n=1}^{\infty}$ converges. Then there exists an (L, r)-graphon $W\colon [0,1]^{\mathfrak{h}([r])} \to L$ such that $t(\tilde{H}, \mathfrak{eval}(F_n)) \to t(\tilde{H}, W)$ for every H. Additionally, W satisfies for each $x \in [0,1]^{\mathfrak{h}([r])}$ and $\pi \in S_r$ that $W(x_{\pi(\mathfrak{h}([r]))}) = \hat{\pi}(W(x_{\mathfrak{h}([r])}))$, where $\hat{\pi}$ is as in Corollary 2.3.1 when elements of L are considered as maps from $[q]^r$ to non-negative integers.*

We should keep the notion of the naive and the averaged naive form of the representation of the limit object in mind, as discussed above towards the end of Section 2.2. They will be utilized in further chapters.

Exchangeable partition-indexed processes We conclude the section with a remark that is motivated by the array representation of *r*CSPs. The next form presented seems to be the least redundant in some aspect, since no additional symmetry conditions have to be fulfilled by the limit objects.

The most natural exchangeable infinite random object fitting the one-to-one correspondence of Theorem 2.2.7 with *r*CSP limits is the following process, that preserves every piece of information contained in the evaluation representation.

Definition 2.3.3. *Let $N_q^r = \{\mathcal{P} = (P_1, \dots, P_q) \mid$ the sets $P_i \subset \mathbb{N}$ are pairwise disjoint and $\sum_{i=1}^{q} |P_i| = r\}$ be the set of directed q-partitions of r-subsets of $\mathbb{N}$. We call the random process $(X_{\mathcal{P}})_{\mathcal{P}\in N_q^r}$ that takes values in some compact Polish space $\mathcal{K}$ a partition indexed process. The process $(X_{\mathcal{P}})_{\mathcal{P}\in N_q^r}$ has the exchangeability property if its distribution is invariant under the action induced by finite permutations of $\mathbb{N}$, i.e., $(X_{\mathcal{P}})_{\mathcal{P}\in N_q^r} \overset{d}{=} (X_{\rho^*(\mathcal{P})})_{\mathcal{P}\in N_q^r}$ for any $\rho \in \mathrm{Sym}_0(\mathbb{N})$.*

Unfortunately, the existence of a representation theorem for partition-indexed exchangeable processes analogous to Theorem 2.2.6 that offers additional insight over the directed colored *r*-array version is not established, and there is little hope in this direction. The reason for this is again the fact that there is no standard way of separating the generating process of the elements $X_{\mathcal{P}}$ and $X_{\mathcal{P}'}$ non-trivially in the case when $\mathcal{P}$ and $\mathcal{P}'$ have the same underlying base set of cardinality *r* but are different as partitions into two non-trivial random stages with the first being identical for the two variables and the second stage being conditionally independent over the outcome of the first stage.

CHAPTER 3

Graph parameter testability, norms, distances

3.1 Introduction

This chapter contains the formal definition of testability via random uniform sampling for dense combinatorial structures for both parameters and properties that is one of the main subjects of the thesis. We gave a taste of known results in the introductory Chapter 1, in the current chapter we aim to demonstrate how graph limit theory, that we already encountered in Chapter 2, can be employed to reprove some of those results, and to give new characterizations in analytical terms.

In order to do this we require an understanding of distances that are relevant for this undertaking, which connects us to regular partitions of graphs.

We will further review the concept of this regularity notion together with the norms and distances that arise in this context, some technical results obtained here will be utilized in the subsequent Chapter 6. Our new versions of the regularity lemma, both for graphs and uniform hypergraphs of higher rank might be relevant on their own right. Here we highlight how they fit into the hierarchy of previous versions of regularity lemmas in terms of the conditions imposed on suitable partitions on the local and the global scale.

With the required tools at hand, we turn to the aforementioned part that shows the relevance of limit theory techniques in testing and estimation. This part is followed by a discussion on efficient testability that connects us directly to the topic of Chapter 4, where we study extensively the efficiently testable graph and graphon parameters given by the ground state energies.

We conclude the chapter with the discussion of the aspect of the limit concept with regard to statistical physics, especially spin models. This connection had been established in [32], here we recall their results, briefly investigate the situation regarding uniform hypergraphs of higher order, and describe a set of relevant models.

3.2 Random sampling and testing

Recall the term $\mathbb{G}(k, G)$ for a uniformly sampled subgraph defined in Chapter 2. Originally, in [30], the testability of $(\mathcal{K}, r)$-graph parameters (which are real functions invariant under r-graph-isomorphisms) was defined as follows.

Definition 3.2.1. *A $(\mathcal{K}, r)$-graph parameter f is testable, if for every $\varepsilon > 0$ there exists a $k(\varepsilon) \in \mathbb{N}$ such that for every $k \geq k(\varepsilon)$ and simple $(\mathcal{K}, r)$-graph G on at least k vertices*

$$\mathbb{P}(|f(G) - f(\mathbb{G}(k, G))| > \varepsilon) < \varepsilon.$$

The minimal function k that satisfies this condition is the sample complexity of f and is denoted by q_f.

A $(\mathcal{K}, r)$-graphon parameter f is a measurable functional on the space of r-graphons that is invariant under graphon equivalence (meaning that the moments $t(F, W_i)$ for W_1 and W_2 all F coincide), for $r = 2$ this is the action induced by a measure preserving map ϕ from $[0, 1]$ to $[0, 1]$, that is, $f(W) = f(W^\phi)$, where $W^\phi(x, y) = W(\phi(x), \phi(y))$. For general r, see the uniqueness statements in Chapter 5 that formulate the r-graphon equivalence in terms of the existence of certain structure preserving maps on the graphon domain.

A graphon parameter g may give rise to a graph parameter f through $f(G) = g(W_G)$, the reverse is not possible in general as graph parameters are not required to be invariant under equitable blow-ups of their vertices. The testability in the graphon case is defined analogously to Definition 3.2.1, with the difference that we have to additionally require the existence of a graph parameter $\hat{f}$ in order to be able approximate $f(W)$ by the quantity $\hat{f}(\mathbb{G}(k, W))$.

A closely related notion to parameter testing is property testing. A simple graph property $\mathcal{P}$ is characterized by the subset of the set of simple graphs containing the graphs which have the property, in what follows $\mathcal{P}$ will be identified with this subset. Informally, $\mathcal{P}$ can be described as testable if we can distinguish between the cases that an instance has $\mathcal{P}$, and that it requires many edge modifications to reach a graph that has the property through the means of uniform vertex sampling with a certain confidence.

We have seen the most general form of the definition of testability for graph properties in Section 1.2.1. Here, we restrict ourselves to the canonical version, and employ it throughout the thesis. In this setting we reduce the scope of a possible tester by requiring it to completely inspect a sampled induced graph, and to reach its decision deterministically. This framework couples sample complexity (vertex cardinality of the sample) and query complexity (number of accesses to the adjacency matrix), thus it eliminates the need of algorithmic design on the sampled graph if we only aim for the sample complexity up to a polynomial factor.

The number of queries satisfying the formulation below can always be chosen to be at most quadratic in the more general non-canonical one, but as shown in [64] there exist properties, where this quadratic increase is truly a consequence of the restrictive definition, a tester feasible for the general definition can give a correct output with the same quality by only inspecting partially the representation of the sampled graph. Such a tester makes its edge queries in general sequentially and utilizes the information it has gained for the further choice of queries, and is called adaptive.

Definition 3.2.2. *$\mathcal{P}$ is testable, if there exists another graph property $\mathcal{P}'$, such that*

(a) $\mathbb{P}(\mathbb{G}(k,G) \in \mathcal{P}') \geq \frac{2}{3}$ for every $k \geq 1$ and $G \in \mathcal{P}$, and

(b) for every $\varepsilon > 0$ there is a $k(\varepsilon)$ such that for every $k \geq k(\varepsilon)$ and G with $d_1(G,\mathcal{P}) \geq \varepsilon$ we have that $\mathbb{P}(\mathbb{G}(k,G) \in \mathcal{P}') \leq \frac{1}{3}$,

where d_1 is the normalized edit distance between graphs. The minimal function $k(.)$ that satisfies this condition is the sample complexity of $\mathcal{P}$ and is denoted by $q_{\mathcal{P}}$. The definition of testable $(\mathcal{K}, r)$-graph properties is analogous.

As for parameters, also here the testability notion can be extended to graphon properties in a straightforward way.

Note that $\frac{1}{3}$ and $\frac{2}{3}$ in the definition can be replaced by arbitrary constants $0 < a < b < 1$, this change may alter the corresponding certificate $\mathcal{P}'$ and the function k, but not the characteristic of testability. It was mentioned in [89] that once the testability of $\mathcal{P}$ is established, the test property can be always chosen to be

$$\mathcal{P}' = \left\{ G \mid |V(G)| = 1, \text{ or } \delta_\square(G,\mathcal{P}) \leq \frac{20}{\sqrt{\log |V(G)|}} \right\}, \tag{3.1}$$

where $\delta_\square$ is the cut distance between graphs, see Definition 3.3.6 below.

One link, first established by Fischer and Newman [56], subsequently reproved with the graph limit machinery by Lovász and Szegedy [94], between the notions of qualitative (property) and quantitative (parameter) testing is presented below. These concepts may be extended to the infinitary space of graphons, where a similar notion of sampling is available.

Lemma 3.2.3. *[56][94] $\mathcal{P}$ is a testable graph property if and only if $d_1(.,\mathcal{P})$ is a testable graph parameter.*

Some properties can be described by imposing a condition on one or several parameters of the graph. We remark that in these cases in general we cannot point out in advance that one type of testability is harder than the other in the sense of the required sample size.

Remark 3.2.4. Consider the property of having a large clique of size at least αn in an n-vertex graph for a fixed $\alpha > 0$. This property falls into the category of partition problems: we are looking for a bipartition (V_1, V_2) of the vertex set with sizes $\delta_1 n$ and $\delta_2 n$, and density matrix $(\mu_{ij})_{1\leq i,j\leq 2}$, and the conditions of our property are $\delta_1 \geq \alpha$ together with $\mu_{11} = 1$. Therefore by [66] this property is testable, even with polynomial sample size. On the other hand, the parameter of the maximal clique number density is not testable. Suppose we are given a complete graph on $2n$ vertices minus a perfect matching, the clique number in this case is n, so the density is 1/2. Whenever we sample a small subgraph of size independent of n, we will see with high probability a complete graph, as no pairs will be selected, hence we have a maximal clique density of 1.

For other instances, property testing can be the harder task. For example, it is known that triangle-freeness is testable, albeit with a sample size that is at least exponential. In contrast, we will see below that the triangle density only requires a sample of size $O(\log(1/\varepsilon)(1/\varepsilon)^2)$ for an ε-test.

We conclude our remarks on this issue that testing both the property that the density of the densest $|V(G)|/2$-sized subgraph is at least α, and testing the corresponding parameter are equally hard to test.

We presented an overview of the current state of the art in the area of testing in the dense model in Chapter 1, here we will focus on their study via limit theory. One important ingredient of that approach is the $\delta_\square$-distance between graphs and graphons introduced in [30] that is related to regular partitions of graphs.

3.3 Regularity lemmas and related notions

We are going to present next the famous Regularity Lemma (RL in short) by Szemerédi [115], that has become the single most important tool in several branches of combinatorics where one studies large, dense structures. Informally the lemma states, that any simple graph can be vertex partitioned into a bounded number of parts such that most pairs of classes form random-like bipartite graphs, meaning that they resemble random graphs generated by independent trails for the adjacency choices with some common success probability.

Definition 3.3.1. *Let G be a simple graph with subsets A and B of its vertex set, we call the pair (A, B) ε-regular, if they are disjoint and for any $A' \subset A$ and $B' \subset B$ with $|A'| \geq \varepsilon|A|$ and $|B'| \geq \varepsilon|B|$ it holds that*

$$\left|\frac{e_G(A,B)}{|A||B|} - \frac{e_G(A',B')}{|A'||B'|}\right| \leq \varepsilon, \tag{3.2}$$

where $e_G(A, B)$ denotes the number of edges running between A and B in G.

Lemma 3.3.2 (Szemerédi's Regularity Lemma)**.** *[115] For any $\varepsilon > 0$ there exists a positive integer $M(\varepsilon) \geq \frac{1}{\varepsilon}$ such that the following holds. For any graph G with $|V(G)| \geq M(\varepsilon)$ there exists a partition of $V(G)$ into $m \leq M(\varepsilon)$ parts $(V_1, \ldots, V_m)$ such that*

(i) for any $i, j \in [m]$ we have $||V_i| - |V_j|| \leq 1$, and

(ii) at least $(1-\varepsilon)m^2$ of the pairs $(i, j) \in [m]^2$ satisfy that (V_i, V_j) is ε-regular.

Such a partition is called a ε-regular partition.

The main significance of the is lemma that it provides a uniform approximation of any graph by weighted template graphs obtained by putting the original edge densities onto the edges of the complete graph whose vertices are associated to the partition classes (also called reduced graphs), therefore their order is upper bounded by $M(\varepsilon)$ that only depends on the quality ε of the approximation. Unfortunately, the original proof ensures only that $M(\varepsilon)$ is at most an exponential tower of twos of height $O(\varepsilon^{-5})$, the bound is essentially tight by Gowers [67], who showed that there exist graphs that do not admit ε-regular partitions with the number of parts being at most a tower of height $\Omega(\varepsilon^{-1/16})$. This feature of the lemma prevents it from practical applications, however it serves as a central tool in extremal graph theory and other related areas, and further, several versions of varying strength required for specific problems were developed, see Komlós, Shokoufandeh, Simonovits, and Szemerédi [85]. Another important aspect is that although the statement of the lemma is true for sub-dense graphs (that have a sub-quadratic number of edges in the size of the vertex set), an appropriate partition can always be chosen to be the trivial one. However, there exist also sparse versions of the lemma, see Gerke and Steger [60] for a survey.

Lemma 3.3.2 was first used to obtain a proof for Roth's Theorem, that is now known as the first non-trivial case ($k = 3$) of Szemerédi's Theorem, stating that for any $k \geq 1$, every subset A of the positive integers of positive upper density contains infinitely many k-term arithmetic progressions. In the course of that work, the Triangle Removal Lemma was established, which informally states that if a graph G contains $o(|V(G)|^3)$ triangles, then it can be made triangle-free by removing $o(|V(G)|^2)$ of its edges. This is a not very complicated consequence of the Regularity Lemma: If after cleaning up the graph by removing edges with both endpoints inside the same class, between irregular pairs of classes, and between pairs that have low edge density we still see a triangle, then there must be many by the regularity of the non-empty pairs of the partition classes. The result is one of the starting points of the area of Graph Property Testing, also it has several generalizations that all fall under the name Removal Lemma, some of them corresponding to other versions of the Regularity Lemma, see Conlon and Fox [39] for a recent survey.

It is worth mentioning that the notion of an ε-regular pair in the definition of a ε-regular partition can be replaced by the term δ-locally regular pair with $\delta = \mathrm{poly}(\varepsilon)$. A disjoint pair $A, B \subset V(G)$ is δ-locally regular if $|e^4_G(A,B) - c_G(A,B)| \leq \delta|A|^4|B|^4$, where $c_G(A,B)$ is the number of labeled 4-cycles in the bipartite graph spanned by the edges of G running between A and B. This condition can be checked more easily than the original, and is relevant for algorithmic versions of the lemma, see Fischer, Matsliah, and Shapira [57].

Further direct consequences of almost every version of the RL are the accompanying Counting Lemmas. These state that some statistic (for example, a specific subgraph density) of the original graph can be well-estimated by only looking at the reduced graph. We will use this term also more general as a continuity assertion, when there is a distance measure between weighted graphs that encodes the regularity characteristic in some sense, see Lemma 3.3.7 as an example below.

We continue with a well-known computationally more efficient version of the RL by

Frieze and Kannan [59], also called the Weak Regularity Lemma (WRL). Conceptually it is very similar to the original one, it ensures the existence of a weakly regular partition with significantly less classes, the trade-off being that the weaker regularity here captures local modifications of the graph insufficiently for some applications such as Removal Lemmas.

Definition 3.3.3. *Let $\varepsilon > 0$, G be a simple graph and $\mathcal{P} = (P_1, \dots, P_t)$ be a partition of its vertex set. We call the partition weakly ε-regular if for any $S, T \subset V(G)$ we have*

$$\left| \frac{1}{|V(G)|^2} \left[e_G(S,T) - \sum_{i,j=1}^{t} \frac{e_G(P_i, P_j)}{|P_i||P_j|} |S \cap P_i||T \cap P_j| \right] \right| < \varepsilon. \tag{3.3}$$

Lemma 3.3.4 (Weak Regularity Lemma)**.** *[59] For any $\varepsilon > 0$ there exists a positive integer $\frac{1}{\varepsilon} \leq M(\varepsilon) \leq 2^{\frac{4}{\varepsilon^2}}$ such that the following holds. For any graph G there exists an weakly-ε regular partition of $V(G)$ into at most $M(\varepsilon)$ parts.*

The extension of the lemma requiring the partition to have classes of almost equal size or to refine a given partition is not overly involved, and it only increases the granted upper bound $M(\varepsilon)$ by a constant factor in the exponent. Similar to the situation for the original lemma it has been established by Conlon and Fox [38] that the upper bound on $M(\varepsilon)$ in the statement is essentially the best possible.

A different formulation of the Regularity Lemma that contains a wide range of other versions as special cases including the two discussed above is the general regularity lemma in Hilbert spaces, see Lovász and Szegedy [92]. The aforementioned version unifies the previous approaches in some sense.

Lemma 3.3.5. *[92] Let $\mathcal{K}_1, \mathcal{K}_2, \dots$ be arbitrary subsets of a Hilbert space $\mathcal{H}$. Then for every $\varepsilon > 0$ and $f \in \mathcal{H}$ there is an $m \leq \frac{1}{\varepsilon^2}$ and there are $f_l \in \mathcal{K}_l$ and $\gamma_l \in \mathbb{R}$ $(1 \leq l \leq m)$ such that for every $g \in \mathcal{K}_{m+1}$ we have*

$$|\langle g, f - \sum_{l=1}^{m} \gamma_l f_l \rangle| \leq \varepsilon \|f\| \|g\|. \tag{3.4}$$

As remarked in [92], one interpretation of the above lemma is related to orthogonal decomposition in a Hilbert space: if the $\mathcal{K}_i$ are orthogonal subspaces, then for each n one can find for each $f \in \mathcal{H}$ a $\hat{f} \in \oplus_{i \leq n} \mathcal{K}_i$ such that $f - \hat{f}$ is orthogonal to each of $\mathcal{K}_{m+1}, \mathcal{K}_{m+2}, \dots$. The above Lemma 3.3.5 says that if we relax the condition on the $\mathcal{K}_i$'s to be subspaces there still is an m such that the error of $f - \hat{f}$ is almost orthogonal to $\mathcal{K}_{m+1}$, for a linear combination $\hat{f}$ in the form as above.

3.3.1 Regularity for graphs

We proceed by enumerating the norms and distances that are relevant for the current work and are related to the graph limit theory and property testing. Later we also present the various analogous notions for r-uniform hypergraphs for arbitrary r. We call the partition of $[0,1]$ into n consecutive intervals of equal measure the canonical n-partition. If the canonical n-partition refines a partition, then we speak of an $\mathcal{I}_n$-partition, and an $\mathcal{I}_n$-set is the union of some classes of the canonical n-partition. Further, a measure-preserving map from $[0,1]$ to $[0,1]$ is referred to as an $\mathcal{I}_n$-permutation if it corresponds to a permutation of the classes of the canonical n-partition. Functions on $[0,1]$ and $[0,1]^2$ are called $\mathcal{I}_n$-functions if they are constant on the classes of the canonical n-partition and products of those, respectively.

Definition 3.3.6. *The cut norm of a real $n \times n$ matrix A is*

$$\|A\|_\square = \frac{1}{n^2} \max_{S,T \subset [n]} |A(S,T)|\,,$$

where $A(S,T) = \sum_{s \in S, t \in T} A(s,t)$.

The cut distance of two labeled simple graphs F and G on the same vertex set $[n]$ is

$$d_\square(F,G) = \|A_F - A_G\|_\square,$$

where A_F and A_G stand for the respective adjacency matrices. The cut norm of a 2-kernel W is

$$\|W\|_\square = \max_{S,T \subset [0,1]} \left| \int_{S \times T} W(x,y) \mathrm{d}x \mathrm{d}y \right|, \tag{3.5}$$

where maximum is taken over all pairs of measurable sets S and T. We speak of the n-cut norm of 2-kernels when the maximum in (3.5) is only taken over pairs of $\mathcal{I}_n$-sets, it is denoted by $\|W\|_\square^{\langle n \rangle}$. The cut norm of a $k \times k$-tuple of 2-dikernels $\mathbf{W} = (W^{(\alpha,\beta)})_{\alpha,\beta=1}^k$ is

$$\|\mathbf{W}\|_\square = \sum_{\alpha,\beta=1}^{k} \|W^{(\alpha,\beta)}\|_\square.$$

The cut distance of two 2-graphons W and U is

$$\delta_\square(W,U) = \inf_{\phi,\psi} \|W^\phi - U^\psi\|_\square,$$

where the infimum runs over all pairs of measure-preserving map from $[0,1]$ to $[0,1]$, and the graphon W^ϕ is defined as $W^\phi(x,y) = W(\phi(x),\phi(y))$. Similarly, for $(k,2)$-digraphons $\mathbf{W}$ and $\mathbf{U}$ we have

$$\delta_\square(\mathbf{W},\mathbf{U}) = \inf_{\phi,\psi} \|\mathbf{W}^\phi - \mathbf{U}^\psi\|_\square,$$

with the difference being component-wise. The cut distance for arbitrary unlabeled graphs F and G is

$$\delta_\square(F,G) = \delta_\square(W_F, W_G),$$

the definitions for the colored directed version is identical. Another variant is for the case when $V(F) = [m]$ and $V(G) = [n]$ such that m is a divisor of n. Then

$$\hat{\delta}_{\square}^{\langle n \rangle}(F, G) = \min_{\phi} d_{\square}(F[n/m], G^{\phi}),$$

where $F[t]$ is the t-fold blow up of F and the minimum goes over all node relabellings ϕ of G. In the case $n = m$ we omit the upper index and use $\hat{\delta}_{\square}$.

In fact, $\hat{\delta}_{\square}$ and $\delta_{\square}$ define only pseudometrics, graphs have distance zero whenever they have equitable blow-ups that are isomorphic. For graphons we introduce the term graphon equivalence for the case whenever the $\delta_{\square}$ distance is 0, but will refer to the above with a slight abuse of notation as proper distances, see Section 5.4 for details regarding the equivalence.

Observe that for two graphs F and G on the common node set $[n]$ the distance $d_{\square}(F, G) = \|W_F - W_G\|_{\square} = \|W_F - W_G\|_{\square}^{\langle n \rangle}$. Also note that in general for F and G with identical vertex cardinalities $\delta_{\square}(F, G)$ is not necessarily equal to $\delta_{\square}^{\langle n \rangle}(F, G)$, however in [30] it was demonstrated that $\delta_{\square}(F, G) \leq \hat{\delta}_{\square}(F, G) \leq 32(\delta_{\square}(F, G))^{1/67}$.

An important property of the distances introduced above is that subgraph densities are uniformly continuous in the topology defined by them, this is a certain type of counting lemma, these were mentioned above related to regularity lemmas.

Lemma 3.3.7. *[30] Let U and W be two graphons. Then for every simple graph F on q vertices we have*

$$|t(F, W) - t(F, U)| \leq \binom{q}{2} \delta_{\square}(U, W).$$

The connection to graph limits is given in the next theorem from [30].

Theorem 3.3.8. *[30] A graph sequence $(G_n)_{n \geq 1}$ (a $(k, 2)$-digraph sequence $(\mathbf{G}_n)_{n \geq 1}$, respectively) is convergent if and only if it is Cauchy in the $\delta_{\square}$ metric.*

A remarkable feature of the $\delta_{\square}$ distance is that the deviation of a sampled graph from the original graph or graphon can be upper bounded by a function that decreases logarithmically in the inverse of the sample size. Originally, this result was established to verify Theorem 3.3.8.

Lemma 3.3.9. *[30] Let $\varepsilon > 0$ and let U be a graphon with $0 \leq U \leq 1$. Then for $q \geq 2^{100/\varepsilon^2}$ we have*

$$\mathbb{P}\left(\delta_{\square}(U, \mathbb{G}(q, U)) \geq \varepsilon\right) \leq \exp\left(-4^{100/\varepsilon^2} \frac{\varepsilon^2}{50}\right). \tag{3.6}$$

One can easily deduce the RL version of Frieze and Kannan [59] from Lemma 3.3.5, we turn our attention to the continuous formulation in the graphon space. For a partition $\mathcal{P}$ of $[0, 1]$ and a 2-kernel W we obtain $W_{\mathcal{P}}$ from W by averaging on every rectangle given by product sets from $\mathcal{P}$.

Lemma 3.3.10 (Weak Regularity Lemma for 2-kernels). *[59], [92] For every $\varepsilon > 0$ and $W \in \hat{\Xi}_0^2$ there exists a partition $\mathcal{P} = (P_1, \dots, P_m)$ of $[0,1]$ into $m \leq 2^{\frac{8}{\varepsilon^2}}$ parts, such that*

$$\|W - W_{\mathcal{P}}\|_\square \leq \varepsilon \|W\|_2. \tag{3.7}$$

In the same way we get the version for k-colored graphons.

Lemma 3.3.11 (Weak Regularity Lemma for $(k,2)$-digraphons). *For every $\varepsilon > 0$ and k-colored digraphon $\mathbf{W}$ there exists a partition $\mathcal{P} = (P_1, \dots, P_m)$ of $[0,1]$ into $m \leq 2^{k^4 \frac{8}{\varepsilon^2}} = t'_k(\varepsilon)$ parts, such that*

$$d_\square(\mathbf{W}, \mathbf{W}_{\mathcal{P}}) = \sum_{\alpha,\beta=1}^{k} \|W^{(\alpha,\beta)} - (W^{(\alpha,\beta)})_{\mathcal{P}}\|_\square \leq \varepsilon. \tag{3.8}$$

When $\mathbf{W} = \mathbf{W}_\mathbf{G}$ for a k-colored digraph $\mathbf{G}$ with vertex cardinality n, then one can require in the above statement that $\mathcal{P}$ is an $\mathcal{I}_n$-partition.

We would like to elaborate on the last result: Appealing to Lemma 3.3.5 we set $\mathcal{H}$ to be the space of k^2-tuples of $L^2([0,1]^2)$ functions with the inner product being the sum of the component-wise L^2-products. Further, each $\mathcal{K}_i$ consists of all k^2-tuples of indicator functions of the form $\mathbb{1}_{S\times T}(x,y)$. Then for an arbitrary k-colored digraphon $\mathbf{W}$ we have $\|\mathbf{W}\|_2 \leq 1$ and for any element $\mathbf{U}$ of $\mathcal{K}_i$ we have $\|\mathbf{U}\|_2 \leq k$. It follows that there exists a $\mathbf{V}$ that is a weighted sum of at most $\frac{4k^2}{\varepsilon^2}$ elements of $\mathcal{K}_i$ and by this a proper step function with at most $2^{\frac{8k^4}{\varepsilon^2}}$ steps forming $\mathcal{P}$ such that $d_\square(\mathbf{W}, \mathbf{V}) \leq \varepsilon/2$. Since $\|.\|_\square$ is contractive with respect to averaging we have $d_\square(\mathbf{W}, \mathbf{W}_{\mathcal{P}}) \leq \varepsilon$.

The following 2-kernel norm shares some useful properties with the cut-norm. Most prominently it admits a regularity lemma that outputs a partition whose number of classes is considerably below the tower-type magnitude in the desired accuracy. On the other hand, it does not admit a straight-forward definition of a related distance by calculating the norm of the difference of two optimally overlayed objects as in Definition 3.3.6. This is the result of the general assumption that the partition $\mathcal{P}$ involved in the definition always belongs to one of the graphons whose deviation we wish to estimate. Therefore a relabeling of this graphon should also act on $\mathcal{P}$, hence symmetry fails. Its advantages in comparison to the cut norm will become clearer in Chapter 6.

Definition 3.3.12. *Let W be a 2-kernel and $\mathcal{P} = (P_1, \dots, P_t)$ a partition of $[0,1]$. Then the cut-$\mathcal{P}$-norm of W is*

$$\|W\|_{\square\mathcal{P}} = \max_{S_i,T_i \subset P_i} \sum_{i,j=1}^{t} \left| \int_{S_i \times T_j} W(x,y) \mathrm{d}x \mathrm{d}y \right|. \tag{3.9}$$

For two kernels U and W let $d_{W,\mathcal{P}}(U)$ denote the cut-$\mathcal{P}$-deviation of U with respect to W that *is defined by*

$$d_{W,\mathcal{P}}(U) = \inf_{\phi} \|U^{\phi} - W\|_{\square\mathcal{P}}, \tag{3.10}$$

where the infimum runs over all measure preserving maps from $[0,1]$ *to* $[0,1]$.

For $n \geq 1$, *a partition* $\mathcal{P}$ *of* $[n]$ *and a directed weighted graph* H *the cut-*$\mathcal{P}$*-norm of* H *on* $[n]$ *is defined as*

$$\|H\|_{\square\mathcal{P}} = \|W_H\|_{\square\mathcal{P}'}, \tag{3.11}$$

where $\mathcal{P}'$ *is the partition of* $[0,1]$ *induced by* $\mathcal{P}$ *and the map* $j \mapsto [\frac{j-1}{n}, \frac{j}{n})$.

The definition for the k-colored version is analogous.

Definition 3.3.13. *Let* $\mathbf{W} = (W^{(1,1)}, \dots, W^{(k,k)})$ *be a* $(k,2)$*-dikernel and* $\mathcal{P} = (P_1, \dots, P_t)$ *a partition of* $[0,1]$. *Then the cut-*$\mathcal{P}$*-norm of* $\mathbf{W}$ *is*

$$\|\mathbf{W}\|_{\square\mathcal{P}} = \sum_{\alpha,\beta=1}^{k} \|W^{(\alpha,\beta)}\|_{\square\mathcal{P}}. \tag{3.12}$$

For two $(k,2)$*-dikernels* $\mathbf{U}$ *and* $\mathbf{W}$ *let* $d_{\mathbf{W},\mathcal{P}}(\mathbf{U})$ *denote the cut-*$\mathcal{P}$*-deviation of* $\mathbf{U}$ *with respect to* $\mathbf{W}$ *that is defined by*

$$d_{\mathbf{W},\mathcal{P}}(\mathbf{U}) = \inf_{\phi} \|\mathbf{U}^{\phi} - \mathbf{W}\|_{\square\mathcal{P}} = \inf_{\phi} \sum_{\alpha,\beta=1}^{k} \|(U^{(\alpha,\beta)})^{\phi} - W^{(\alpha,\beta)}\|_{\square\mathcal{P}}, \tag{3.13}$$

where the infimum runs over all measure preserving maps from $[0,1]$ *to* $[0,1]$.

It is not hard to check that the cut-$\mathcal{P}$-norm is in fact a norm on the space where we identify two kernels when they differ only on a set of measure 0. From the definition it follows directly that for arbitrary kernels U and W, and any partition $\mathcal{P}$ we have $\|W\|_{\square} \leq \|W\|_{\square\mathcal{P}} \leq \|W\|_1$ and $\delta_{\square}(U,W) \leq d_{W,\mathcal{P}}(U) \leq \delta_1(U,W)$, the same is true for the k-colored directed version.

Remark 3.3.14. We present a different description of the cut-$\mathcal{P}$-norm of W and $\mathbf{W}$ respectively that will allow us to rely on results concerning the cut-norm of Definition 3.3.6 more directly. For a partition $\mathcal{P}$ with t classes and $A = (A_{j,l})_{j,l=1}^{t} \in \{-1,+1\}^{t \times t}$, let $W^A(x,y) = A_{j,l}W(x,y)$ ($\mathbf{W}^A$ is given by $(W^{(\alpha,\beta)})^A(x,y) = A_{j,l}W^{(\alpha,\beta)}(x,y)$ respectively) for $x \in P_j$ and $y \in P_l$. Then $\|W\|_{\square\mathcal{P}} = \max_A \|W^A\|_{\square}$ and $\|\mathbf{W}\|_{\square\mathcal{P}} = \max_A \|\mathbf{W}^A\|_{\square}$.

This newly introduced norm admits a uniform approximation in the following sense that is essential to conduct the proof of Theorem 6.1.5 in Chapter 6. We call a 2-kernel a step function if there are partitions $\mathcal{S} = (S_1, \dots, S_t)$ and $\mathcal{T} = (T_1, \dots T_t)$ of $[0,1]$ into the same number of classes such that the kernel is constant on $S_i \times T_j$ for each $i, j \in [t]$. The kernel is a proper step function if the two partitions above can be chosen to be the same. For a partition $\mathcal{P}$ the integer $t_{\mathcal{P}}$ denotes the number of its classes.

Lemma 3.3.15. *For every* $\varepsilon > 0$, $m_0\colon [0,1] \to \mathbb{N}$, $k \geq 1$ *and k-colored directed graphon* $\mathbf{W} = (W^{(\alpha,\beta)})_{\alpha,\beta\in[k]}$ *there exists a partition* $\mathcal{P} = (P_1, \dots, P_t)$ *of* $[0,1]$ *into* $t \leq \frac{(16m_0(\varepsilon))^{2^{\frac{k^4}{\varepsilon^2}}}}{4} = t_k(\varepsilon, m_0(\varepsilon))$ *parts, such that for any partition* $\mathcal{Q}$ *of* $[0,1]$ *into at most* $\max\{t, m_0(\varepsilon)\}$ *classes we have*

$$\|\mathbf{W} - \mathbf{W}_{\mathcal{P}}\|_{\square\mathcal{Q}} \leq \varepsilon. \tag{3.14}$$

If $\mathbf{W} = \mathbf{W}_{\mathbf{G}}$ *for some k-colored* $\mathbf{G}$ *with* $|V(\mathbf{G})| = n$, *then one can require that* $\mathcal{P}$ *is an* $\mathcal{I}_n$-*partition. If we want the parts to have equal measure (almost equal in the graph case), then the upper bound on the number of classes is modified to* $\frac{((2k)^{12} m_0(\varepsilon)/\varepsilon^4)^{2^{2k^4/\varepsilon^2}}}{(2k)^6/\varepsilon^2}$.

Proof. Fix an arbitrary $\varepsilon > 0$ and a $\mathbf{W} \in \tilde{\Xi}^{2,k}$, and set $m_0 = m_0(\varepsilon)$. We construct a sequence of partitions $\mathcal{R}_0, \mathcal{R}_1, \dots, \mathcal{R}_m$ such that $\mathcal{R}_0 = [0,1]$ and each $\mathcal{R}_{i+1}$ refines the preceding $\mathcal{R}_i$. The integer m is a priori undefined.

The construction is sequential in the sense that we assume that we have already constructed $\mathcal{R}_0, \mathcal{R}_1, \dots, \mathcal{R}_{i-1}$ before considering the ith step of the construction.

If for $i \geq 1$ there exists a partition $\mathcal{Q} = (Q_1, \dots, Q_{t_\mathcal{Q}})$ of $[0,1]$ into at most $\max\{t_{\mathcal{R}_{i-1}}, m_0\}$ classes such that

$$\|\mathbf{W} - \mathbf{W}_{\mathcal{R}_{i-1}}\|_{\square\mathcal{Q}} > \varepsilon, \tag{3.15}$$

then we proceed to the construction of $\mathcal{R}_i$. In the case of $i = 1$ we choose $\mathcal{Q}$ to have exactly m_0 parts of positive measure, this can be achieved since for any refinement $\mathcal{Q}'$ of $\mathcal{Q}$ we have $\|\mathbf{W} - \mathbf{W}_{\mathcal{R}_{i-1}}\|_{\square\mathcal{Q}'} \geq \|\mathbf{W} - \mathbf{W}_{\mathcal{R}_{i-1}}\|_{\square\mathcal{Q}}$. The inequality (3.15) implies that there are $\alpha_0, \beta_0 \in [k]$ and measurable sets S and T such that

$$\sum_{i,j=1}^{t_\mathcal{Q}} \left| \int_{(S\cap Q_i)\times(T\cap Q_j)} W^{(\alpha_0,\beta_0)}(x,y) - W^{(\alpha_0,\beta_0)}_{\mathcal{R}_{i-1}}(x,y)\mathrm{d}x\mathrm{d}y \right| > \varepsilon/k^2. \tag{3.16}$$

In this case we define $\mathcal{R}_i$ to be the coarsest common refinement of $\mathcal{R}_{i-1}$, $\mathcal{Q}$, and $\{S, T\}$ for some arbitrary choice of the latter partition and sets satisfying (3.16).

Set $S_j = S \cap Q_j$ and $T_j = T \cap Q_j$ for $j \in [t_\mathcal{Q}]$ and $U = W^{(\alpha_0,\beta_0)} - W^{(\alpha_0,\beta_0)}_{\mathcal{R}_{i-1}}$, further define the step function $V = \sum_{j,l\in[t_\mathcal{Q}]} \mathrm{sgn}(\int_{S_j\times T_l} U)\mathbb{1}_{S_j\times T_l}$

In this case

$$\begin{aligned}
\|\mathbf{W}_{\mathcal{R}_i}\|_2^2 - \|\mathbf{W}_{\mathcal{R}_{i-1}}\|_2^2 &= \sum_{\alpha,\beta=1}^{k} \langle W^{(\alpha,\beta)}_{\mathcal{R}_i}, W^{(\alpha,\beta)}_{\mathcal{R}_i}\rangle - \langle W^{(\alpha,\beta)}_{\mathcal{R}_{i-1}}, W^{(\alpha,\beta)}_{\mathcal{R}_{i-1}}\rangle \\
&= \sum_{\alpha,\beta=1}^{k} \langle W^{(\alpha,\beta)}_{\mathcal{R}_i} - W^{(\alpha,\beta)}_{\mathcal{R}_{i-1}}, W^{(\alpha,\beta)}_{\mathcal{R}_i} - W^{(\alpha,\beta)}_{\mathcal{R}_{i-1}}\rangle \\
&= \sum_{\alpha,\beta=1}^{k} \|W^{(\alpha,\beta)}_{\mathcal{R}_i} - W^{(\alpha,\beta)}_{\mathcal{R}_{i-1}}\|_2^2
\end{aligned} \tag{3.17}$$

$$\geq \|W^{(\alpha_0,\beta_0)}_{\mathcal{R}_i} - W^{(\alpha_0,\beta_0)}_{\mathcal{R}_{i-1}}\|_2^2$$

$$\geq \frac{1}{\|V\|_2^2} |\langle W^{(\alpha_0,\beta_0)}_{\mathcal{R}_i} - W^{(\alpha_0,\beta_0)}_{\mathcal{R}_{i-1}}, V\rangle|^2 \tag{3.18}$$

$$= \frac{1}{\|V\|_2^2} |\langle W^{(\alpha_0,\beta_0)} - W^{(\alpha_0,\beta_0)}_{\mathcal{R}_{i-1}}, V\rangle|^2 \tag{3.19}$$

$$\geq |\langle W^{(\alpha_0,\beta_0)} - W^{(\alpha_0,\beta_0)}_{\mathcal{R}_{i-1}}, V\rangle|^2 \tag{3.20}$$

$$> \varepsilon^2/k^4. \tag{3.21}$$

Here we used first in (3.17) that $\langle W^{(\alpha,\beta)}_{\mathcal{R}_{i-1}}, W^{(\alpha,\beta)}_{\mathcal{R}_i}\rangle = \langle W^{(\alpha,\beta)}_{\mathcal{R}_{i-1}}, W^{(\alpha,\beta)}_{\mathcal{R}_{i-1}}\rangle$, since $W^{(\alpha,\beta)}_{\mathcal{R}_{i-1}}$ is constant on $\mathcal{R}_{i-1}$ rectangles, and the integral of the two functions is equal on these rectangles. In (3.18) we used the Cauchy-Schwarz inequality, then in (3.19) the fact that $\langle W^{(\alpha_0,\beta_0)}_{\mathcal{R}_i}, V\rangle = \langle W^{(\alpha_0,\beta_0)}, V\rangle$, that is true by V being constant on $\mathcal{R}_i$ rectangles and $W^{(\alpha_0,\beta_0)}_{\mathcal{R}_i}$ and $W^{(\alpha_0,\beta_0)}$ having the same integral value on taken on $\mathcal{R}_i$ rectangles. We concluded the calculation in (3.20) by $\|V\|_2 \leq 1$ and in (3.21) using the condition (3.16).

If for some $i_0 \geq 0$ we have

$$\|\mathbf{W} - \mathbf{W}_{\mathcal{R}_{i_0}}\|_{\square Q} \leq \varepsilon \tag{3.22}$$

for every partition Q of $[0,1]$ into at most $\max\{t_{\mathcal{R}_{i_0}}, m_0(\varepsilon)\}$ classes, then we stop the process and set $\mathcal{P} = \mathcal{R}_{i_0}$ and $m = i_0$.

We have $\|\mathbf{W}_{\mathcal{R}_j}\|_2^2 \leq \|\mathbf{W}\|_2^2 \leq \|\mathbf{W}\|_1^2 \leq 1$ for each $j \geq 0$, and at each non-terminating step we showed $\|\mathbf{W}_{\mathcal{R}_i}\|_2^2 - \|\mathbf{W}_{\mathcal{R}_{i-1}}\|_2^2 > \varepsilon^2/k^4$. Therefore by

$$\|\mathbf{W}_{\mathcal{R}_j}\|_2^2 \geq \sum_{i=1}^{j} \|\mathbf{W}_{\mathcal{R}_i}\|_2^2 - \|\mathbf{W}_{\mathcal{R}_{i-1}}\|_2^2,$$

for each $j \geq 1$ we conclude that the process terminates definitely after a finite number of steps and $m \leq k^4/\varepsilon^2$. The partition $\mathcal{P}$ satisfies

$$\|\mathbf{W} - \mathbf{W}_{\mathcal{P}}\|_{\square Q} \leq \varepsilon$$

for each Q with $t_Q \leq \max\{t, m_0(\varepsilon)\}$ by the choice of m and the construction of the partition sequence, we are left to verify the upper bound on $t_{\mathcal{P}}$ in the statement of the lemma.

We know that $t_{\mathcal{R}_0} = 1$, and if the partition does not terminate before the first step, then we assume $m_0 \leq t_{\mathcal{R}_1}$. This lower bound does not affect generality, the partition Q_0 that certifies that $\mathcal{R}_0$ is not sufficient for the choice of the partition $\mathcal{P}$ in the statement of the lemma is selected to have $t_{Q_0} = m_0$. For this particular choice of Q_0 and $\mathcal{R}_1$ we can reformulate the condition that the terminating partition $\mathcal{R}_m$ has to fulfill as

$$\|\mathbf{W} - \mathbf{W}_{\mathcal{R}_m}\|_{\square Q} \leq \varepsilon$$

for every Q partition of $[0,1]$ into at most $t_{\mathcal{R}_m}$ classes, since $t_{\mathcal{R}_i} \geq m_0$ for every $i \geq 1$.

We set $s(0) = 1$, $s(1) = 4m_0$, and further define $s(i+1) = 4s(i)^2$ for each $i \geq 1$. We claim that for each $i \geq 0$ we have $t_{\mathcal{R}_i} \leq s(i)$, this can be easily verified by induction, since at each step $\mathcal{R}_{i+1}$ is the coarsest common refinement of two partitions with $t_{\mathcal{R}_i}$ classes and two additional sets.

Further, for each $i \geq 1$ we have now $\log 4s(i+1) = 2\log 4s(i)$, therefore $s(i) = \frac{(16m_0)^{2^{i-1}}}{4}$, and consequently $s(m) \leq \frac{(16m_0)^{2^{\frac{k^4}{\varepsilon^2}-1}}}{4}$.

The case regarding $\mathbf{W} = \mathbf{W}_\mathbf{G}$ for $\mathbf{G} \in \tilde{\Pi}_n^{2,k}$ follows completely identically, at each step of the construction of the partitions $\mathcal{R}_i$, the partition Q and the sets S and T can be chosen to be an $\mathcal{I}_n$-partition and sets, respectively. Hence, $\mathcal{P}$ is an $\mathcal{I}_n$-partition, the upper bound on $t_\mathcal{P}$ is identical to the one in the general case.

In a similar way we can achieve that $\mathcal{P}$ is an equiv-partition, or a $\mathcal{I}_n$-partition with classes of almost equal size in the graph case respectively. Fix $\varepsilon > 0$ and a $\mathbf{W} \in \Xi^{r,k}$. For this setup we define the partition sequence somewhat differently, in particular each element is an equiv-partition. Set $\mathcal{R}_0 = [0,1]$ and each $\mathcal{R}_{i+1}$ refines the preceding $\mathcal{R}_i$.

If for $i \geq 1$ there exists a partition $Q = (Q_1, \ldots, Q_{t_Q})$ of $[0,1]$ into at most $\max\{t_{\mathcal{R}_{i-1}}, m_0\}$ classes such that

$$\|\mathbf{W} - \mathbf{W}_{\mathcal{R}_{i-1}}\|_{\square Q} > \varepsilon, \tag{3.23}$$

then we proceed to the construction of $\mathcal{R}_i$, otherwise we stop, as above, and set $\mathcal{P} = \mathcal{R}_{i-1}$, and $m = i-1$. Assume that we are facing the first case. Let $\mathcal{R}'_{i-1}$ be the coarsest common refinement of $\mathcal{R}_{i-1}$, Q, and $\{S, T\}$, where the sets S and T certify (3.23) as above. Then $\|\mathbf{W}_{\mathcal{R}'_{i-1}}\|_2^2 - \|\mathbf{W}_{\mathcal{R}_{i-1}}\|_2^2 > \varepsilon^2/k^4$. Let $\mathcal{R}''_{i-1} = (R_1^2, \ldots, R_l^2)$ be the partition that is obtained from the classes of $\mathcal{R}'_{i-1} = (R_1^1, \ldots, R_l^1)$, such that the measure of each of the classes of $\mathcal{R}''_{i-1}$ is an integer multiple of $\varepsilon^2/(14k^6 t_{\mathcal{R}'_{i-1}})$ with $\lambda(R_i^1 \triangle R_i^2) \leq \varepsilon^2/(14k^6 t_{\mathcal{R}'_{i-1}})$ for each $i \in [l]$. (We disregard the technical difficulty of $1/\varepsilon^2$ not being an integer to facilitate readability.)

Claim 1. For any 2-kernel $W\colon [0,1]^2 \to \mathbb{R}$ and partitions $\mathcal{P} = (P_1, \ldots, P_l)$ and $\mathcal{S} = (S_1, \ldots, S_l)$ we have

$$\|W_\mathcal{P} - W_\mathcal{S}\|_1 \leq 7 \sum_{i=1}^{l} \lambda(P_i \triangle S_i). \tag{3.24}$$

To see this, let $T_i = P_i \cap S_i$, $N_i = P_i \setminus S_i$, and $M_i = S_i \setminus P_i$ for each $i \in [l]$, and let $\mathcal{T}_1 = (T_1, \ldots, T_l, N_1, \ldots, N_l)$ and $\mathcal{T}_2 = (T_1, \ldots, T_l, M_1, \ldots, M_l)$ be two partitions of $[0,1]$. Then

$$\|W_\mathcal{P} - W_{\mathcal{T}_1}\|_1 \leq \sum_{i,j=1}^{l} \left| \int_{T_i \times T_j} \frac{\int_{P_i \times P_j} W}{\lambda(P_i)\lambda(P_j)} - \frac{\int_{T_i \times T_j} W}{\lambda(T_i)\lambda(T_j)} \right| + 2 \sum_{i=1}^{l} \lambda(N_i)$$

$$\leq \sum_{i,j=1}^{l} \frac{1}{\lambda(P_i)\lambda(P_j)\lambda(T_i)\lambda(T_j)} \Bigg| \int_{T_i\times T_j} \lambda(T_i)\lambda(T_j) \left[\int_{N_i\times T_j} W + \int_{T_i\times N_j} W + \int_{N_i\times N_j} W \right]$$

$$- \left[\lambda(N_i)\lambda(T_j) + \lambda(T_i)\lambda(N_j) + \lambda(N_i)\lambda(N_j)\right] \int_{T_i\times T_j} W \Bigg| + 2\sum_{i=1}^{l} \lambda(N_i)$$

$$\leq \sum_{i,j=1}^{l} 2\|W\|_\infty \frac{\lambda^2(T_i)\lambda^2(T_j)\left[\lambda(N_i)\lambda(T_j) + \lambda(T_i)\lambda(N_j) + \lambda(N_i)\lambda(N_j)\right]}{\lambda(P_i)\lambda(P_j)\lambda(T_i)\lambda(T_j)}$$

$$+ 2\sum_{i=1}^{l} \lambda(N_i)$$

$$\leq 2\sum_{i,j=1}^{l} \lambda(N_i)\lambda(P_j) + \lambda(N_j)\lambda(P_i) + 2\sum_{i=1}^{l} \lambda(N_i)$$

$$= 6\sum_{i=1}^{l} \lambda(N_i).$$

Similarly,

$$\|W_{\mathcal{S}} - W_{\mathcal{T}_2}\|_1 \leq 6\sum_{i=1}^{l} \lambda(M_i),$$

and also

$$\|W_{\mathcal{T}_2} - W_{\mathcal{T}_1}\|_1 \leq 2\sum_{i=1}^{l} \lambda(M_i),$$

which implies the claim.

By Claim 1 it follows that

$$\left|\|\mathbf{W}_{\mathcal{R}'_{i-1}}\|_2^2 - \|\mathbf{W}_{\mathcal{R}''_{i-1}}\|_2^2\right| = \left|\sum_{\alpha,\beta=1}^{k} \int_{[0,1]^2} (W^{(\alpha,\beta)}_{\mathcal{R}'_{i-1}})^2(x,y) - (W^{(\alpha,\beta)}_{\mathcal{R}''_{i-1}})^2(x,y)\mathrm{d}x\mathrm{d}y\right|$$

$$\leq \sum_{\alpha,\beta=1}^{k} \left|\int_{[0,1]^2} (W^{(\alpha,\beta)}_{\mathcal{R}'_{i-1}}(x,y) - W^{(\alpha,\beta)}_{\mathcal{R}''_{i-1}}(x,y))(W^{(\alpha,\beta)}_{\mathcal{R}'_{i-1}}(x,y) + W^{(\alpha,\beta)}_{\mathcal{R}''_{i-1}}(x,y))\mathrm{d}x\mathrm{d}y\right|$$

$$\leq \sum_{\alpha,\beta=1}^{k} \|W^{(\alpha,\beta)}_{\mathcal{R}'_{i-1}} - W^{(\alpha,\beta)}_{\mathcal{R}''_{i-1}}\|_1$$

$$\leq 7k^2 \sum_{i=1}^{l} \lambda(R_i^1 \triangle R_i^2)$$

$$\leq \varepsilon^2/(2k^4).$$

We finish with the construction of $\mathcal{R}_i$ by refining $\mathcal{R}''_{i-1}$ into $\varepsilon^2/(14k^6 t_{\mathcal{R}'_{i-1}})$ sets in total of equal measure so that the resulting partition refines $\mathcal{R}''_{i-1}$. It follows that

$$\|\mathbf{W}_{\mathcal{R}_i}\|_2^2 - \|\mathbf{W}_{\mathcal{R}''_{i-1}}\|_2^2 \geq 0, \tag{3.25}$$

hence

$$\|\mathbf{W}_{\mathcal{R}_i}\|_2^2 - \|\mathbf{W}_{\mathcal{R}_{i-1}}\|_2^2 \geq \varepsilon^2/(2k^4). \tag{3.26}$$

The construction of the partitions terminates after at most $2k^4/\varepsilon^2$ steps. The partition $\mathcal{P}$ satisfies the norm conditions of the lemma, we are left to check whether it has the right number of classes. Similarly as above, let $s(0) = 1$ and $s(1) = 56m_0k^6/\varepsilon^2$, and further for $i \geq 1$ let $s(i+1) = 56k^6 s^2(i)/\varepsilon^2$. It is clear from the construction that $t_{\mathcal{R}_i} \leq s(i)$. Let $a = 56k^6/\varepsilon^2$, then it is not difficult to see that $s(i) = \frac{(a^2m_0)^{2^{i-1}}}{a}$. It follows that $t_{\mathcal{P}} \leq \frac{((2k)^{12}m_0/\varepsilon^4)^{2^{2k^4/\varepsilon^2}}}{(2k)^6/\varepsilon^2}$.

The graph case also follows analogously to the general partition case we dealt with above.

□

As seen in the proof, the upper bound on the number of classes in the statement of the lemma is not the sharpest we can prove, we stay with the simpler bound for the sake of readability. In the simple graph and graphon case the above reads as follows.

Corollary 3.3.16. *For every $\varepsilon > 0$ and $W \in \Xi^2$ there exists a partition $\mathcal{P} = (P_1, \ldots, P_m)$ of $[0,1]$ into $m \leq 16^{2^{1/\varepsilon^2}}/4$ parts, such that*

$$\|W - W_{\mathcal{P}}\|_{\square Q} \leq \varepsilon. \tag{3.27}$$

for each partition Q of $[0,1]$ into at most $t_{\mathcal{P}}$ classes.

With the additional condition that the partition classes should have the same measure the above is true with $m \leq \frac{(2^{12}/\varepsilon^4)^{2^{(2^4/\varepsilon^2)}}}{2^6/\varepsilon^2}$.

We illustrate the form of the original Regularity Lemma in the graphon context in order to provide a better understanding of the strength of the above statements.

Corollary 3.3.17. *For every $\varepsilon > 0$ and $W \in \Xi^2$ there exists a partition $\mathcal{P} = (P_1, \ldots, P_m)$ of $[0,1]$ into m parts that is at most a tower of twos of height $\mathrm{poly}(1/\varepsilon)$, such that*

$$\sup_{S_{i,j} \subset P_i, T_{i,j} \subset P_j} \sum_{i,j=1}^{m} \left| \int_{S_{i,j} \times T_{i,j}} W(x,y) - W_{\mathcal{P}}(x,y) \mathrm{d}x \mathrm{d}y \right| < \varepsilon, \tag{3.28}$$

where the supremum runs over all suitable measurable pairs $S_{i,j}$ and $T_{i,j}$ for $i, j \in [m]$.

If we impose on the the condition in the supremum that $S_{i,j}$ be the same for each $j \in [m]$ and disjoint if varying i, similarly but switching the indices for $T_{i,j}$, then the right hand side of (3.28) changes to the cut-$\mathcal{P}$-norm of $W - W_{\mathcal{P}}$. Another equivalent formulation of (3.28) is

$$\sup_{S_i,T_i \subset [0,1]} \sum_{i,j=1}^{m} \left| \int_{(S_j \cap P_i)\times(T_i \cap S_j)} W(x,y) - W_{\mathcal{P}}(x,y)\mathrm{d}x\mathrm{d}y \right| < \varepsilon. \tag{3.29}$$

3.3.2 Regular partitions of uniform hypergraphs

Again we start by presenting norms and distances, these generalize the graph in two directions. This is followed by the corresponding regularity lemmas that are verified by the virtue of Lemma 3.3.5 above. We start with the plain cut norm and distance that is a straight-forward generalization of the notions for graphs and was employed in [14]. We are going to utilize the norm as a substantial ingredient in Chapter 4.

Due to some shortcomings they can not be applied to certain problems in contrast to the genuine r-cut norms which are to be presented subsequently.

Definition 3.3.18. *The plain cut norm of an $n \times \cdots \times n$ real r-array A is*

$$\|A\|_{\square} = \frac{1}{n^r} \max_{S_1,\dots,S_r \subset [n]} |A(S_1,\dots,S_r)|,$$

and the 1-norm of A is

$$\|A\|_1 = \frac{1}{n^r} \sum_{i_1,\dots,i_r=1}^{n} |A(i_1,\dots,i_r)|.$$

The cut distance of two labeled r-graphs F and G on the same vertex set $[n]$ is

$$d_{\square}(F,G) = \|A_F - A_G\|_{\square},$$

where $F(S_1,\dots,S_r) = \sum_{i_j \in S_j} F(i_1,\dots,i_r)$. The edit distance of the same pair is

$$d_1(F,G) = \|A_F - A_G\|_1.$$

The continuous counterparts are described as follows.

Definition 3.3.19. *The cut norm of a naive r-kernel W is*

$$\|W\|_{\square} = \max_{S_1,\dots,S_r \subset [0,1]} \left| \int_{S_1\times\cdots\times S_r} W(x)\mathrm{d}\lambda(x) \right|,$$

the cut distance of two naive r-kernels W and U is

$$\delta_{\square}(W,U) = \inf_{\phi,\psi} \|W^{\phi} - U^{\psi}\|_{\square},$$

where the infimum runs over all measure-preserving permutations of $[0,1]$, *and the graphon* W^{ϕ} *is defined as* $W^{\phi}(x_1,\dots,x_r) = W(\phi(x_1),\dots,\phi(x_r))$. *The cut distance for arbitrary unlabeled r-graphs or r-arrays F and G is*

$$\delta_{\square}(F,G) = \delta_{\square}(W_F, W_G).$$

We remark that the above definition of the cut norm and distance is not satisfactory from one important aspect for $r \geq 3$: not all sub-r-graph densities are continuous functions in the topology induced by this distance even in the most simple case, when $\mathcal{K} = \{0,1\}$. Suppose that $r = 3$ and we have two random graph models, G_n and H_n on the vertex set $[n]$. We generate the first one by generating an auxiliary Erdős-Rényi 2-graph with density 1/2 (i.e., 2-edges are included independently with probability 1/2), and include as edges the triangles that appear, for the second we pick edges uniformly and independently at random with probability 1/8. For any triple of sets $S_1, S_2, S_3 \subset [n]$ both $e_{G_n}(S_1,S_2,S_3)$ and $e_{H_n}(S_1,S_2,S_3)$ is highly concentrated around the value $\frac{|S_1||S_2||S_3|}{8}$, therefore $d_{\square}(G_n,H_n)$ tends to 0 almost surely (a precise calculation could be executed using large deviation principles). However, the t^*-density of the graph F with $E(F) = \{\{1,2,3\},\{1,2,4\}\}$ and $V(F) = [4]$ tend to different limits almost surely, in particular $t^*(F,G_n) \to 1/32$, while $t^*(F,H_n) \to 1/64$.

Examples of subgraphs whose t^*-densities behave well with respect to the above norms are linear hypergraphs, that have the property that any two distinct edges intersect at most in one vertex.

Nevertheless, for a number of problems this concept is sufficient, and above all, computationally efficient, as we will see in Chapter 4 in the study of the testability of ground state energies. In particular, we will show a closely related result to the regularity lemma corresponding to this norm (not stated here explicitly), Lemma 4.2.4, where the approximation is achieved by a linear combination of indicator functions of rectangles.

The definitions of the genuine r-cut norms are given next that are stronger than the plain cut norms above for each r. To give a hint on their relation, we mention that the test sets here are sets of pairs, triplets, and so on, of the vertices according to the rank r, whereas for the plain cut norm they are simply subsets of the vertex set.

The norms below show a good behavior with regard to subgraph densities for any fixed r-graph, in contrast to the above case, we exploit this for our results in Chapter 6 dealing with uniform hypergraphs of higher rank. The r-cut norm independently from the current work also appeared in the investigation of hypergraph limits by Zhao [119], prior to this counting lemmas in the graph case indicated that such a norm might be useful in the treatment of uniform hypergraphs.

Moreover, we introduce the cut-$\mathcal{P}$-norm for r-graphs analogous to the graph case in Definition 3.3.12 and Definition 3.3.13, they are highly relevant for conducting the technical proofs in Chapter 6.

Definition 3.3.20. *Let* $r \geq 1$ *and A be a real r-array of size n. Then the genuine r-cut norm* *of A is*

$$\|A\|_{\square,r} = \frac{1}{n^r} \max_{\substack{S_i \subset [n]^{r-1} \setminus \mathrm{diag}([n]^{r-1}) \\ i \in [r]}} |A(r; S_1, \dots, S_r)|,$$

where $A(r; S_1, \dots, S_r) = \sum_{i_1,\dots,i_r=1}^{n} A(i_1, \dots, i_r) \prod_{j=1}^{r} \mathbb{1}_{S_j}(i_1, \dots, i_{j-1}, i_{j+1}, \dots, i_r)$*, and the maximum goes over sets* S_i *that are invariant under coordinate permutations as subsets of* $[n]^{r-1}$*.*

If $\mathcal{P} = (P_i)_{i=1}^t$ *is a partition of* $[n]^{r-1} \setminus \mathrm{diag}([n]^{r-1})$ *with symmetric classes, then the cut-*$(r, \mathcal{P})$*-norm of A is*

$$\|A\|_{\square,r,\mathcal{P}} = \frac{1}{n^r} \max_{\substack{S_i \subset [n]^{r-1} \setminus \mathrm{diag}([n]^{r-1}) \\ i \in [r]}} \sum_{j_1,\dots,j_r=1}^{t} |A(r; S_1 \cap P_{j_1}, \dots, S_r \cap P_{j_r})|,$$

where the maximum runs over symmetric sets.

The genuine r-cut norm of an r-kernel W is

$$\|W\|_{\square,r} = \sup_{\substack{S_i \subset [0,1]^{\mathfrak{h}([r-1])} \\ i \in [r]}} \left| \int_{\cap_{i\in[r]} p^{-1}_{\mathfrak{h}([r]\setminus\{i\})}(S_i)} W(x_{\mathfrak{h}([r],r-1)}) \mathrm{d}\lambda(x_{\mathfrak{h}([r],r-1)}) \right|,$$

where the supremum is taken over sets S_i *that are* $(r-1)$*-symmetric, and* $p_{\mathfrak{h}(e)}$ *is the natural projection from* $[0,1]^{\mathfrak{h}([r],r-1)}$ *onto* $[0,1]^{\mathfrak{h}(e)}$*. Furthermore, for a symmetric partition* $\mathcal{P} = (P_i)_{i=1}^t$ *of* $[0,1]^{\mathfrak{h}([r-1])}$ *the cut-*$(r, \mathcal{P})$*-norm of an r-kernel is defined by*

$$\|W\|_{\square,r,\mathcal{P}} = \sup_{\substack{S_i \subset [0,1]^{\mathfrak{h}([r-1])} \\ i \in [r]}} \sum_{j_1,\dots,j_r=1}^{t} \left| \int_{\cap_{i\in[r]} p^{-1}_{\mathfrak{h}([r]\setminus\{i\})}(S_i \cap P_{j_i})} W(x_{\mathfrak{h}([r],r-1)}) \mathrm{d}\lambda(x_{\mathfrak{h}([r],r-1)}) \right|,$$

where the supremum is taken over sets S_i *that are* $(r-1)$*-symmetric in* $[0,1]^{\mathfrak{h}([r-1])}$*.*

In what follows we sometimes omit the term genuine and the parameter r, so when we talk about cut norms of r-graphs, as default we mean genuine r-cut norms.

We remark that it is also true that

$$\|W\|_{\square,r} = \sup_{f_1,\dots,f_r \colon [0,1]^{\mathfrak{h}([r-1])} \to [0,1]} \left| \int_{[0,1]^{\mathfrak{h}([r],r-1)}} \prod_{i=1}^{r} f_i(x_{\mathfrak{h}([r]\setminus\{i\})}) W(x_{\mathfrak{h}([r],r-1)}) \mathrm{d}\lambda(x_{\mathfrak{h}([r],r-1)}) \right|, \tag{3.30}$$

where the supremum goes over functions f_i that are $(r-1)$-symmetric, and similarly for any symmetric partition $\mathcal{P} = (P_i)_{i=1}^t$ of $[0,1]^{\mathfrak{h}([r-1])}$ we have with the same conditions for the f_i's as above that

$$\|W\|_{\square,r,\mathcal{P}} = \sup_{f_1,\dots,f_r \colon [0,1]^{\mathfrak{h}([r-1])} \to [0,1]}$$

$$\sum_{j_1,\dots,j_r=1}^{t} \left| \int_{[0,1]^{\mathfrak{h}([r],r-1)}} \prod_{i=1}^{r} f_i(x_{\mathfrak{h}([r]\setminus\{i\})}) \mathbb{1}_{P_{j_i}}(x_{\mathfrak{h}([r]\setminus\{i\})}) W(x_{\mathfrak{h}([r],r-1)}) \mathrm{d}\lambda(x_{\mathfrak{h}([r],r-1)}) \right|.$$

It is immediate through the comparison with Definition 3.3.20 that $\|A\|_\square \leq \|A\|_{\square,r} \leq \|A\|_{\square,r,\mathcal{P}} \leq \|A\|_1$ for any r-array A, $\|W\|_\square \leq \|W\|_{\square,r} \leq \|W\|_{\square,r,\mathcal{P}} \leq \|W\|_1$ for any r-kernel W. The above norms give rise to a distance between r-graphons, and analogously for r-graphs. We present next the k-colored version.

Definition 3.3.21. *For two k-colored r-graphons* $\mathbf{U} = (U^\alpha)_{\alpha\in[k]}$ *and* $\mathbf{W} = (W^\alpha)_{\alpha\in[k]}$ *their r-cut distance is defined as*

$$d_{\square,r}(\mathbf{U},\mathbf{W}) = \sum_{\alpha=1}^{k} \|U^\alpha - W^\alpha\|_{\square,r},$$

and their cut-$\mathcal{P}$-distance as

$$d_{\square,r,\mathcal{P}}(\mathbf{U},\mathbf{W}) = \sum_{\alpha=1}^{k} \|U^\alpha - W^\alpha\|_{\square,r,\mathcal{P}}.$$

For two k-colored r-graphs $\mathbf{G} = (G^\alpha)_{\alpha\in[k]}$ *and* $\mathbf{H} = (H^\alpha)_{\alpha\in[k]}$ *their corresponding distances are defined as*

$$d_{\square,r}(\mathbf{G},\mathbf{H}) = d_{\square,r}(\mathbf{W_G},\mathbf{W_H}),$$

and

$$d_{\square,r,\mathcal{P}}(\mathbf{G},\mathbf{H}) = d_{\square,r,\mathcal{P}}(\mathbf{W_G},\mathbf{W_H}).$$

Distances between an r-graph and an r-graphon, as well as for r-kernels, is analogously defined.

Note that the norms introduced above are in general smaller or equal than the 1-norm of integrable functions, also $d_{\square,r}(\mathbf{U},\mathbf{W}) \leq d_{\square,r,\mathcal{P}}(\mathbf{U},\mathbf{W})$ hods for every pair. Their relevance will be clearer in the context of the next counting lemma, we include the standard proof only for completeness' sake.

Unfortunately, we cannot expect $d_{\square,r}$ to give rise to a metric analogous to $\delta_\square$ for graphs in terms of providing an equivalent characterization of r-graph convergence. One crucial property of the $\delta_\square$-distance is that a sampled graph is $\delta_\square$-close to the original, and the $\delta_\square$-deviation vanishes almost surely when the sample size tends to infinity. This happens to fail for the extensions of $d_{\square,r}$. We demonstrate this obstacle with two examples of randomly generated sequences, whose subgraph densities coincide in the limit, but are with positive probability far away in $d_{\square,r}$, and also no relabeling would improve the situation substantially.

Example 3.3.22. For any $n \geq 1$, let H_n and H'_n be independent random tripartite 3-graphs on $[3n]$ both generated by the same random process: we generate a bipartite Erdős-Rényi 2-graph G_n, G'_n respectively, between $[n]$ and $\{n+1,\dots,2n\}$ with density 1/2, and include the triplets $\{i,j,k\} \subset [3n]$ in H_n, H_n' respectively if $i \in [n]$, $j \in \{n+1,\dots,2n\}$, $k \in \{2n+1,\dots,3n\}$, and $ij \in E(G_n)$, respectively $ij \in E(G_n')$. Let S_1 denote the subgraph of the complete bipartite graph between $[n]$ and $\{n+1,\dots,2n\}$ where G_n and G_n' differ, further let S_2 and S_3 be the complete bipartite subgraphs between $[n]$ and $\{2n+1,\dots,3n\}$ and, $\{n+1,\dots,2n\}$ and $\{2n+1,\dots,3n\}$ respectively. Then

$$d_{\square,r}(H_n,H_n') \geq \frac{1}{27n^3}|A_{H_n}(r;S_1,S_2,S_3) - A_{H_n'}(r;S_1,S_2,S_3)| = \frac{1}{27n^2}|E(G_n)\triangle E(G_n')|.$$

We know that $|E(G_n)\triangle E(G_n')|$ is tightly concentrated around its expectation $n^2/2$, which means that $d_{\square,r}(H_n,H_n')$ is bounded away from 0 in the limit almost surely. However, it is not hard to show that the subgraph densities for each simple 3-graph F of H_n and H_n' converge to the same (deterministic) limit, see Lemma 3.3.24 below, so the interlaced sequence $H_1, H_1', H_2, H_2', \dots$ does converge in the sense of Definition 2.2.3, but is not Cauchy in $d_{\square,r}$.

The second example concerns non-partite graphs.

Example 3.3.23. For any $n \geq 1$, let H_n and H_n' be 3-graphs that are generated independently on $[n]$ again by independent Erdős-Rényi 2-graphs G_n, G_n' with density $1/2$, and include the triangles that appear as edges of the respective 3-graphs. Note that this measure is invariant under re-labeling the vertex set, therefore if we have a lower bound on $d_{\square,r}(H_n,H_n')$, then we also have one on $d_{\square,r}(H_n,H_n'')$ for any H_n'' obtained by relabeling H_n'. Let S_1 denote the subgraph of the complete graph on $[n]$ where G_n and G_n' differ, and S_2 and S_3 the subgraph of edges present in both models. Then

$$\begin{aligned} d_{\square,r}(H_n,H_n') &\geq \frac{1}{n^3}|A_{H_n}(r;S_1,S_2,S_3) - A_{H_n'}(r;S_1,S_2,S_3)| \\ &= \frac{1}{n^3}\sum_{ijk\in[n]^3} \mathbb{1}_{E(G_n)\triangle E(G_n')}(ij)\mathbb{1}_{E(G_n)\cap E(G_n')}(jk)\mathbb{1}_{E(G_n)\cap E(G_n')}(ki). \end{aligned}$$

For any pairs ij, the sum $\sum_{k\in[n]} \mathbb{1}_{E(G_n)\cap E(G_n')}(jk)\mathbb{1}_{E(G_n)\cap E(G_n')}(ki)$ is concentrated around $n/16$, so with high probability for at most $o(n)$ of them the sum fails to be below $n/32$.

Again, $|E(G_n)\triangle E(G_n')|$ is tightly concentrated around its expectation that is asymptotically equal to $n^2/4$, so almost surely we have that $d_{\square,r}(H_n,H_n')$ is asymptotically larger or equal than 2^{-7}.

The main advantage of the genuine r-cut norm $\|.\|_{\square,r}$ and distance $d_{\square,r}$ over the plain versions $\|.\|_\square$ and $d_\square$ is the counting lemma they entail.

Lemma 3.3.24. *Let* $\mathbf{U}$ *and* $\mathbf{W}$ *be two* (k,r)*-graphons. Then for every* $\mathbf{F} \in \Pi_q^{r,k}$ *it holds that*

$$|t(\mathbf{F},\mathbf{W}) - t(\mathbf{F},\mathbf{U})| \leq \binom{q}{r} d_{\square,r}(\mathbf{U},\mathbf{W}).$$

Proof. Fix q and $\mathbf{F} \in \Pi_q^{r,k}$. Then

$$|t(\mathbf{F},\mathbf{W}) - t(\mathbf{F},\mathbf{U})| = \left| \int_{[0,1]^{\mathfrak{h}([q],r-1)}} \prod_{e \in \binom{[q]}{r}} W^{\mathbf{F}(e)}(x_{\mathfrak{h}(e,r-1)}) - \prod_{e \in \binom{[q]}{r}} U^{\mathbf{F}(e)}(x_{\mathfrak{h}(e,r-1)}) \mathrm{d}\lambda(x) \right|$$

$$\leq \sum_{e \in \binom{[q]}{r}} \left| \int_{[0,1]^{\mathfrak{h}([q],r-1)}} [W^{\mathbf{F}(e)}(x_{\mathfrak{h}(e,r-1)}) - U^{\mathbf{F}(e)}(x_{\mathfrak{h}(e,r-1)})] \right.$$

$$\left. \prod_{f \in \binom{[q]}{r}, f \prec e} W^{\mathbf{F}(f)}(x_{\mathfrak{h}(f,r-1)}) \prod_{g \in \binom{[q]}{r}, e \prec g} U^{\mathbf{F}(g)}(x_{\mathfrak{h}(g,r-1)}) \mathrm{d}\lambda(x) \right|$$

$$\leq \sum_{e \in \binom{[q]}{r}} \|W^{\mathbf{F}(e)} - U^{\mathbf{F}(e)}\|_{\square,r} \leq \binom{q}{r} d_{\square,r}(\mathbf{U},\mathbf{W}),$$

where $\prec$ is an arbitrary total ordering of the elements of $\binom{q}{r}$, and the second inequality is the consequence of the formulation (3.30) of the r-cut norm. □

Let d_{tv} denote the total variation distance between probability measures on $\Pi_n^{r,[k]*}$, where $[k]* = [k] \cup \{\iota\}$ for $k \geq 1$ (without highlighting the specific parameters in the notion d_{tv}), that is $d_{\mathrm{tv}}(\mu,\nu) = \max_{\mathcal{F} \subset \Pi_n^{r,[k]*}} |\mu(\mathcal{F}) - \nu(\mathcal{F})|$, and let the measure $\mu(q,\mathbf{G})$, respectively $\mu(q,\mathbf{W})$, denote the probability distribution of the random r-graph $\mathbb{G}(q,\mathbf{G})$, respectively $\mathbb{G}(q,\mathbf{W})$, taking values in $\Pi_q^{r,[k]*}$. It is a standard observation then that

$$d_{\mathrm{tv}}(\mu(q,\mathbf{W}),\mu(q,\mathbf{U})) = \frac{1}{2} \sum_{\mathbf{F} \in \Pi_q^{r,[k]*}} |t(\mathbf{F},\mathbf{W}) - t(\mathbf{F},\mathbf{U})|, \tag{3.31}$$

and that $\mathbb{G}(q,\mathbf{W})$ and $\mathbb{G}(q,\mathbf{U})$ can be coupled in form of the random r-graphs $\mathbf{G}_1$ and $\mathbf{G}_2$, such that

$$d_{\mathrm{tv}}(\mu(q,\mathbf{W}),\mu(q,\mathbf{U})) = \frac{1}{2}\mathbb{P}(\mathbf{G}_1 \neq \mathbf{G}_2), \tag{3.32}$$

and further, for any coupling $\mathbf{G}_1'$ and $\mathbf{G}_2'$ it hods that $d_{\mathrm{tv}}(\mu(q,\mathbf{W}),\mu(q,\mathbf{U})) \leq \frac{1}{2}\mathbb{P}(\mathbf{G}_1' \neq \mathbf{G}_2')$.

For $\mathbf{G} \in \Pi_n^{r,k}$ note that

$$d_{\mathrm{tv}}(\mu(q,\mathbf{G}),\mu(q,\mathbf{W}_{\mathbf{G}})) \leq q^2/n, \tag{3.33}$$

where the right hand side is a simple upper bound on the probability that if we uniformly choose q elements of an n-element set, then we get at least two identical objects. The inequality (3.33) follows from the fact that conditioned on the event that the independent and uniform X_is for $i \in [q]$ fall in different intervals $[\frac{j-1}{n}, \frac{j}{n}]$ for $j \in [n]$ the distribution of $\mathbb{G}(q,\mathbf{W}_{\mathbf{G}})$ is the same as the distribution of $\mathbb{G}(q,\mathbf{G})$.

The next corollary is a direct consequence of Lemma 3.3.24.

Corollary 3.3.25. *If* $\mathbf{U}$ *and* $\mathbf{W}$ *are two k-colored r-graphons, then*

$$d_{\mathrm{tv}}(\mu(q,\mathbf{W}),\mu(q,\mathbf{U})) \le \frac{k^{q^r} q^r}{r!} d_{\square,r}(\mathbf{U},\mathbf{W}),$$

and there exists a coupling in form of $\mathbf{G}_1$ *and* $\mathbf{G}_2$ *of the random r-graphs* $\mathbb{G}(q,\mathbf{W})$ *and* $\mathbb{G}(q,\mathbf{U})$*, such that*

$$\mathbb{P}(\mathbf{G}_1 \neq \mathbf{G}_2) \le \frac{2k^{q^r} q^r}{r!} d_{\square,r}(\mathbf{U},\mathbf{W}).$$

A generalization of the notion of a step function in the case of graphs to the situation where we deal with r-graphs is given next.

Definition 3.3.26. *We call a k-colored r-graphon* $\mathbf{W}$ *with* $r \ge l$ *an* (r,l)*-step function if there exist positive integers* $t_l, t_{l+1}, \dots, t_r = k$*, l-symmetric partitions* $\mathcal{P}^S = (P^S_1, \dots, P^S_{t_l})$ *of* $[0,1]^{\mathfrak{h}([l])}$ *for each* $S \in \binom{[r]}{l}$*, and real arrays* $A^j_s \colon [t_{s-1}]^{\mathfrak{h}([s],s-1)} \to [0,1]$ *with* $j \in [t_s]$ *for* $l \le s \le r$ *such that* $\sum_{j\in[t_s]} A^j_s(i_{\mathfrak{h}([s],s-1)}) = 1$ *for any choice of* $i_{\mathfrak{h}([s],s-1)}$ *and for* $s \le r$ *so that* W^α *for* $\alpha \in [k]$ *is of the following form for each* $\alpha \in [k]$.

$$W^\alpha(x_{\mathfrak{h}([r])}) = \sum_{\substack{i_S=1\\ S\subset[r], l\le|S|}}^{t_{|S|}} A^\alpha_r(i_{\mathfrak{h}([r],r-1)})$$
$$\prod_{S\in\binom{[r]}{l}} \mathbb{1}_{P^S_{i_S}}(x_{\mathfrak{h}(S)}) \prod_{\substack{S\subset[r]\\ l+1\le|S|<r}} \mathbb{1}\Big(\sum_{j=1}^{i_S-1} A^j_{|S|}(i_{\mathfrak{h}(S,|S|-1)}) \le x_S \le \sum_{j=1}^{i_S} A^j_{|S|}(i_{\mathfrak{h}(S,|S|-1)})\Big).$$

We refer to the partitions $\mathcal{P}^S$ *as the steps of W. The step function is proper, if the partitions* $\mathcal{P}^S$ *can be chosen to be identical.*

The most simple example is the proper $(r,r-1)$-step function that can be written as

$$W^\alpha(x_{\mathfrak{h}([r])}) = \sum_{i_1,\dots,i_r=1}^{t_{r-1}} A^\alpha_r(i_1,\dots,i_r) \prod_{j=1}^{r} \mathbb{1}_{P_{i_j}}(x_{\mathfrak{h}([r]\setminus\{j\})}).$$

For any r-kernel W and any partition $\mathcal{P} = (P_1,\dots,P_t)$ of $[0,1]^{\mathfrak{h}([r],r-1)}$, the r-kernel $W_{\mathcal{P}}$ is the $(r,r-1)$-step function that with steps in $\mathcal{P}$ that takes the average value

$$1/\lambda(\cap_{j=1}^r p^{-1}_{\mathfrak{h}([r]\setminus\{j\})}(P_{i_j})) \int_{[0,1]^{\mathfrak{h}([r],r-1)}} W(x_{\mathfrak{h}([r])}) \prod_{j=1}^{r} \mathbb{1}_{P_{i_j}}(x_{\mathfrak{h}([r]\setminus\{j\})}) \mathrm{d}\lambda(x_{\mathfrak{h}([r])})$$

on the sets $\cap_{j=1}^r p^{-1}_{\mathfrak{h}([r]\setminus\{j\})}(P_{i_j})$ for each $i_1,\dots,i_r \in [t]$. The definition of averaging operation for (k,r)-graphons is analogous.

We describe in the following an intermediate version of the regularity lemma for edge k-colored r-graphons analogous to Lemma 3.3.15, the partition obtained here satisfies stronger conditions than those imposed by the Weak Regularity Lemma [59], and weaker than by Szemerédi's original. Note that in contrast to Lemma 3.3.15 the bound parameter t appears in the condition as a multiplying factor in the upper bound on the number classes of the test partitions $\mathcal{Q}$, this strengthening is necessary for the technical proofs in Chapter 6.

Lemma 3.3.27. *For every $r \geq 1$, $\varepsilon > 0$, $t \geq 1$, $k \geq 1$ and k-colored r-graphon $\mathbf{W}$ there exists an $(r-1)$-symmetric partition $\mathcal{P} = (P_1, \dots, P_m)$ of $[0,1]^{\mathfrak{h}([r-1])}$ into $m \leq (2^r t)^{2^{k^2/\varepsilon^2}} = t_{\mathrm{reg}}(r,k,\varepsilon,t)$ parts such that for any $(r-1)$-symmetric partition $\mathcal{Q}$ of $[0,1]^{\mathfrak{h}([r-1])}$ into at most mt classes we have*

$$d_{\square,r,\mathcal{Q}}(\mathbf{W},\mathbf{V}) \leq \varepsilon.$$

Proof. The proof proceeds analogously to the proof of Lemma 3.3.15. Fix an arbitrary $\varepsilon > 0$ and a $\mathbf{W} = (W^\alpha)_{\alpha \in [k]} \in \Xi^{r,k}$. We construct a sequence of partitions $\mathcal{R}_0, \mathcal{R}_1, \dots, \mathcal{R}_m$ of $[0,1]^{\mathfrak{h}([r-1])}$ such that $\mathcal{R}_0 = [0,1]^{\mathfrak{h}([r-1])}$ and each $\mathcal{R}_{i+1}$ refines the preceding $\mathcal{R}_i$. We stop the construction after the i_0th step when for every $(r-1)$-symmetric partition $\mathcal{Q}$ of $[0,1]^{\mathfrak{h}([r-1])}$ into at most $t_{\mathcal{R}_{i_0}} t$ classes we have

$$d_{\square,r,\mathcal{Q}}(\mathbf{W},\mathbf{W}_{\mathcal{R}_{i_0}}) \leq \varepsilon,$$

then we set $m = i_0$ and $\mathcal{P} = \mathcal{R}_{i_0}$.

As long as this event not occurs, we construct $\mathcal{R}_{i+1}$ as follows. There exists an $(r-1)$-symmetric $\mathcal{Q}$ partition of $[0,1]^{\mathfrak{h}([r-1])}$ into at most $t_{\mathcal{R}_i} t$ classes and an $\alpha \in [k]$ such that

$$d_{\square,r,\mathcal{Q}}(W^\alpha, W^\alpha_{\mathcal{R}_i}) > \varepsilon/k,$$

and in particular, there are $(r-1)$-symmetric sets $S_1, \dots, S_r \subset [0,1]^{\mathfrak{h}([r-1])}$ such that

$$\sum_{j_1,\dots,j_r=1}^{t_\mathcal{Q}} \left| \int_{\cap_{i\in[r]} p^{-1}_{\mathfrak{h}([r]\setminus\{i\})}(S_i \cap Q_{j_i})} W(x_{\mathfrak{h}([r],r-1)}) \mathrm{d}\lambda(x_{\mathfrak{h}([r],r-1)}) \right| > \varepsilon/k.$$

We define $\mathcal{R}_{i+1}$ as the coarsest common refinement of $\mathcal{R}_i$, $\mathcal{Q}$, and $\{S_1, \dots, S_r\}$. We argue as in the proof of Lemma 3.3.15 that $\|\mathbf{W}_{R_{i+1}}\|_2^2 - \|\mathbf{W}_{R_i}\|_2^2 > \varepsilon^2/k^2$, therefore the process terminates after at most k^2/ε^2 steps, that is $m \leq k^2/\varepsilon^2$. Let $s(0) = 1$, and for $i \geq 1$ let $s(i) = 2^r t s^2(i-1)$. Then it is not difficult to see that on one hand $t_{\mathcal{R}_i} \leq s(i)$ for every $i \geq 0$, and on the other that $s(i) = (2^r t)^{2^i - 1}$. We conclude that the partition $\mathcal{P}$ obtained as the output of the partition generating process satisfies the conditions of the lemma, in particular $t_\mathcal{P} \leq (2^r t)^{2^{k^2/\varepsilon^2}}$. □

We conclude the section with the cut-$\mathcal{P}$ version of the plain cut norm for r-graphs generalizing Definition 3.3.12 in weaker form than the cut-$(r, \mathcal{P})$-norm above in Definition 3.3.20. The corresponding counting lemma below makes it applicable in special cases of the setting of Chapter 6 leading to improved bounds there. Recall the notion of naive r-kernels introduced in Section 2.2.3.

Definition 3.3.28. *For any $r \geq 1$, real r-array A of size n, and partition $\mathcal{P} = (P_i)_{i=1}^t$ of $[n]$ the plain cut-$\mathcal{P}$-norm of A is*

$$\|A\|_{\square\mathcal{P}} = \frac{1}{n^r} \max_{\substack{S_i \subset [n] \\ i \in [r]}} \sum_{j_1,\dots,j_r=1}^{t} |A(S_1 \cap P_{j_1}, \dots, S_r \cap P_{j_r})|.$$

For any $r \geq 1$, and partition $\mathcal{P} = (P_i)_{i=1}^t$ of $[0,1]$ the plain cut-$\mathcal{P}$-norm of a naive r-kernel W is defined by

$$\|W\|_{\square\mathcal{P}} = \sup_{S_i \subset [0,1], i \in [r]} \sum_{j_1,\dots,j_r=1}^{t} \left| \int_{(S_1 \cap P_{j_1}) \times \dots \times (S_r \cap P_{j_r})} W(x_1, \dots, x_r) \mathrm{d}\lambda(x_1, \dots, x_r) \right|,$$

where the supremum is taken over measurable subsets $S_1, \dots, S_r$ of $[0,1]$. The plain cut-$\mathcal{P}$-distance $d_{\square\mathcal{P}}$ of graphs and naive graphons is defined analogously to Definition 3.3.12 exchanging the cut-$(r, \mathcal{P})$-norm for the plain cut-$\mathcal{P}$-norm.

The definition for the k-colored version is analogous. We will employ in Chapter 6 the following auxiliary lemmas that are analogous to Lemma 3.3.27, Corollary 3.3.25, respectively (with analogous proofs).

Lemma 3.3.29. *For every $r \geq 1$, $\varepsilon > 0$, $m_0 \colon [0,1] \to \mathbb{N}$, $k \geq 1$ and k-colored naive r-graphon $\mathbf{W}$ there exists a partition $\mathcal{P} = (P_1, \dots, P_t)$ of $[0,1]$ into $t \leq (2^{2r} m_0(\varepsilon))^{2^{k^2/\varepsilon^2}} / 2^r = t_{\mathrm{reg}}(r, k, \varepsilon, t)$ parts such that for any partition $\mathcal{Q}$ of $[0,1]$ into at most $\max\{m_0(\varepsilon), t\}$ classes we have*

$$d_{\square\mathcal{Q}}(\mathbf{W}, \mathbf{W}_{\mathcal{P}}) \leq \varepsilon.$$

We introduce the colored version of the t^* subgraph densities of r-graphs and r-graphons. Let F be a k-colored r-graph and $\hat{F}$ be a simple r-graph, both defined on $[q]$. Then for $G \in \Pi^{r,k}$ the $\hat{F}$-density of F in G is

$$t^*_{\hat{F}}(F, G) = \frac{1}{|V(G)|(|V(G)|-1)\dots(|V(G)|-q+1)} \sum_{\phi \colon [q] \to V(G)} \prod_{e \in \hat{F}} \mathbb{1}_{G^{F(e)}}(\phi(e)), \tag{3.34}$$

where the sum runs over injective ϕ maps, and for a $W \in \Xi^{r,k}$ it is

$$t^*_{\hat{F}}(F, W) = \int_{[0,1]^{\mathfrak{h}([q], r-1)}} \prod_{e \in \hat{F}} W^{F(e)}(x_{\mathfrak{h}(e, r-1)}) \mathrm{d}\lambda(x_{\mathfrak{h}([q], r-1)}). \tag{3.35}$$

The respective counting lemma is formulated next, we omit the proof since it is analogous to previous counting lemma proofs.

Lemma 3.3.30. *Let U and W be k-colored naive r-graphons. Then for every k-colored r-graph F and linear r-graph $\hat{F}$ we have*

$$|t^*_{\hat{F}}(F, W) - t^*_{\hat{F}}(F, U)| \le \binom{q}{r} d_\square(U, W).$$

3.4 Limit theory in testing

Testing parameters A characterization of the testability of a graph parameter in terms of graph limits was developed in [30] for $\mathcal{K} = \{0, 1\}$ in the undirected case, we will focus in the next paragraphs on this most simple setting and give an overview on previous work. Recall Definition 3.2.1.

Theorem 3.4.1. *[30] Let f be a simple graph parameter, then the following statements are equivalent.*

(i) The parameter f is testable.

(ii) For every $\varepsilon > 0$ there exists a $k(\varepsilon) \in \mathbb{N}$ such that for every $k \ge k(\varepsilon)$ and simple graph G on at least k vertices

$$|f(G) - \mathbb{E} f(\mathbb{G}(k, G))| < \varepsilon.$$

(iii) For every convergent sequence $(G_n)_{n=1}^\infty$ of simple graphs with $|V(G_n)| \to \infty$ the numerical sequence $(f(G_n))_{n=1}^\infty$ also converges.

(iv) For every $\varepsilon > 0$ there exist a $\varepsilon' > 0$ and a $n_0 \in \mathbb{N}$ such that for every pair G_1 and G_2 of simple graphs $|V(G_1)|, |V(G_2)| \ge n_0$ and $\delta_\square(G_1, G_2) < \varepsilon'$ together imply $|f(G_1) - f(G_2)| < \varepsilon$.

(v) There exists a $\delta_\square$-continuous functional f' on the space of graphons, so that $f(G_n) \to f'(W)$ whenever $G_n \to W$.

The above theorem also partially holds true for uniform hypergraph parameters of higher rank, and more generally, for $(\mathcal{K}, r)$-graphs. We will focus here on the basic case $\mathcal{K} = \{0, 1\}$.

Theorem 3.4.2. *A parameter f of simple r-uniform hypergraphs is testable if and only if for every convergent sequence $(G_n)_{n=1}^\infty$ of simple r-graphs with $|V(G_n)| \to \infty$ the numerical sequence $(f(G_n))_{n=1}^\infty$ also converges.*

Proof. For arbitrary $r \ge 1$, testability of an r-graph parameter f implies that for any convergent sequence $(H_n)_{n=1}^\infty$ of hypergraphs $(f(H_n)_{n=1}^\infty)$ also converges. This is due to the fact that convergence means distributional convergence of the random induced subgraphs. For every k the distribution of $\mathbb{G}(k, H_n)$ and $\mathbb{G}(k, H_m)$ are close for $m, n \geq n_0(k)$ for some $n_0(k)$, therefore they can be coupled in a way so that with high probability they coincide, and so the corresponding values $f(\mathbb{G}(k, H_n))$ and $f(\mathbb{G}(k, H_m))$ are equal with high probability. On the other hand, testability implies that we have closeness of $f(G)$ and $f(\mathbb{G}(k, G))$ for every G and $k \geq k(\varepsilon)$ with high probability.

For the converse direction we require a metric for (isomorphism classes of) r-graphs whose topology is identical to the subgraph convergence topology. Elek and Szegedy [49] proposed the rather abstract

$$\delta(G, H) = \inf_{\varepsilon > 0}\{\, \varepsilon \mid |t(F, G) - t(F, H)| < \varepsilon \text{ for every simple } F \text{ with } |V(F)| \leq 1/\varepsilon \,\}. \quad (3.36)$$

A convergent sequence of r-graphs is trivially Cauchy in this metric, and vice versa. Further, we claim that random subgraphs are typically close to the source graph.

Claim 2. There exists an absolute constant $c > 0$ such that if $k \geq c\varepsilon^{-5-r}$ and G is an r-graph on at least k vertices, then we have $\delta(G, \mathbb{G}(k, G)) \leq \varepsilon$ with probability at least $1 - \varepsilon$. In particular, $\delta(G, \mathbb{G}(k, G))$ tends to zero in probability when $k \to \infty$.

We only need to use Lemma 3.5.4: There are at most $(1/\varepsilon)2^{\binom{(1/\varepsilon)}{r}}$ simple r-graphs on at most $1/\varepsilon$ vertices, the probability that one fixed F fails to satisfy $|t(F, G) - t(F, \mathbb{G}(k, G))| < \varepsilon$ is upper bounded by $\exp(-\frac{\varepsilon^4 k}{18})$, see Lemma 3.5.4 below, so we can upper bound by the quantity ε the probability there will be one r-graph of size at most $1/\varepsilon$ that harms the bound through the right choice of $c > 0$, which proves the claim.

Now suppose that for every convergent $(G_n)_{n=1}^{\infty}$ with $|V(G_n)| \to \infty$, $(f(G_n))_{n=1}^{\infty}$ also converges. Then for every $\varepsilon > 0$ there are n_0 and $\gamma = \gamma(\varepsilon) > 0$ with $\gamma \leq \varepsilon$ such that if $\delta(G, H) < \gamma$ and $|V(G)|, |V(H)| \geq n_0$, then $|f(G) - f(H)| < \varepsilon$. To see the previous statement, suppose that the implication is not true, and consider two convergent sequences that demonstrate this with their pairwise δ-distance tends to 0, while the corresponding deviations of the f values are bounded from below by a positive constant. Then the interlaced sequence converges in the δ metric, this together with our assumption brings the contradiction.

Now suppose that f is not testable, and there is a $\varepsilon_0 > 0$ and a sequence of bad graphs $(G_n)_{n=1}^{\infty}$ such that $\mathbb{P}(|f(G_n) - f(\mathbb{G}(k_n, G_n))| > \varepsilon_0) > \varepsilon_0$ with $k_n \to \infty$. But from Claim 2 it follows that eventually there is an $m \geq n_o$ with $k_m \geq c(\varepsilon_0/2)^{-5-r}$ such that $\mathbb{P}(\delta(G_m, \mathbb{G}(k_m, G_m)) > \gamma(\varepsilon_0/2)) < \gamma(\varepsilon_0/2) \leq \varepsilon_0/2$, so $|f(G_m) - f(\mathbb{G}(k_m, G_m))| > \varepsilon_0/2$ with at most probability $\varepsilon_0/2$, which is a contradiction.

□

We turn to revisit some further results of [30] for the graphs case. From the practical viewpoint for the verification of the testability of a given parameter perhaps the next characterization is the most suitable. With the aid of it the task can be decomposed into three subproblems whose solution in general requires less effort.

Theorem 3.4.3. *[30] Let f be a simple graph parameter, then the following conditions together are equivalent to the testability of f.*

(i) For every $\varepsilon > 0$ there exists a $\varepsilon' > 0$ such that for every pair G_1 and G_2 of simple graphs on the same vertex set $d_\square(G_1, G_2) < \varepsilon'$ implies $|f(G_1) - f(G_2)| < \varepsilon$.

(ii) For every simple graph G the numerical sequence $(f(G[m]))_{m=1}^\infty$ converges for $m \to \infty$.

(iii) For every sequence $(G_n)_{n=1}^\infty$ of simple graphs with $|V(G_n)| \to \infty$ the sequence $(f(G_n) - f(G_n \cup K_1))_{n=1}^\infty$ converges.

It is an open problem whether there exists such an accessible characterization for r-graphs with $r \geq 3$. On a further note we mention that in the case $r = 2$, the testability of a graphon parameter is equivalent to continuity in the $\delta_\square$ distance.

Remark 3.4.4. The intuitive reason for the absence of an analogous, easily applicable characterization of testability for higher rank uniform hypergraphs as in Theorem 3.4.3 is that no natural notion of a suitable distance is available at the moment. The construction of such a metric would require to establish a standard method to compare a large hypergraph H_n to its random induced subgraph on a uniform sample similar to the behavior that the abstract δ-distance exhibits in the proof of Theorem 3.4.2.

The $\delta_\square$ metric for graphs is convenient because of its concise formulation and it induces a compact limit space, the main characteristic that is exploited that the total variation distance of probability measures of induced subgraphs of fixed size is continuous in this distance, any other δ_{var} with this property would fit into the above framework. We present the intuition for candidate distances for hypergraphons that share the above feature.

Let $\mathbf{U}$ and $\mathbf{W}$ be two arbitrary k-colored r-graphons given by $U^\alpha, W^\alpha \colon [0,1]^{\mathfrak{h}([r],r-1)} \to [0,1]$ for each $\alpha \in [k]$. Suppose that for any $l \leq r-1$ and $k' \geq 1$ we have already defined a distance $d_{l,k'}$ between k'-colored l-graphons. Then we define

$$d_{r,k}(\mathbf{U}, \mathbf{W}) = \inf_{\phi,\psi} \inf_{t \geq 1, V_1, V_2} \Big[d_{\square,r}(\mathbf{U}^\phi, \mathbf{V}_1) + d_{\square,r}(\mathbf{W}^\psi, \mathbf{V}_2) + d_{r-1,t}(\mathbf{W}(\mathcal{P}^1), \mathbf{W}(\mathcal{P}^2)) \Big], \quad (3.37)$$

where the first infimum runs over structure preserving maps from $[0,1]^{\mathfrak{h}([r],r-1)}$ to $[0,1]^{\mathfrak{h}([r],r-1)}$, see Definition 5.2.10 below, whereas the second infimum goes over $t \geq 1$ and a pair of $(r, r-1)$-step functions $\mathbf{V}_1 = (V_1^\alpha)_{\alpha\in[k]}$ and $\mathbf{V}_2 = (V_2^\alpha)_{\alpha\in[k]}$ both with t number of steps $\mathcal{P}^1$ and $\mathcal{P}^2$, whose steps can be paired such that P_i^1 corresponds to P_i^2, and on the t^r cells defined by the steps the corresponding values of V_1^α and V_2^α coincide for each $\alpha \in [k]$. Further, $\mathbf{W}(\mathcal{P}^1)$ and $\mathbf{W}(\mathcal{P}^2)$ are the t-colored $(r-1)$-graphons obtained from the partitions $\mathcal{P}^1$ and $\mathcal{P}^2$ by

$$(\mathbf{W}(\mathcal{P}^i))^\beta(x_{\mathfrak{h}([r-1],r-2)}) = \int_{[0,1]} \mathbb{1}_{P^i_\beta}(x_{\mathfrak{h}([r-1])}) \mathrm{d}\lambda(x_{[r-1]})$$

for each $\beta \in [t]$. We do not know for sure whether such a construction satisfies the triangle inequality, but it is not hard to show that the variational distance of the sample distributions is continuous under the above distance, we further conjecture that the metric space is compact (provided it is a genuine distance). Unfortunately, the above

$d_{r,k}$ distance does not seem to offer any immediate structural insight in general in contrast to the $\delta_\square$ distance in the graph case.

Testing properties The authors of [94] provided an insightful application of the graph limit machinery in order to construct a characterization for the testability of properties, their result follows through analytic steps exploiting the properties of the limit space. Most prior works on this topic relied explicitly on Szemerédi's Regularity Lemma, in [94] it can be noticed only in its disguised form as the compactness of the limit space. We briefly review the approach to highlight the methodological significance of going to the continuous limit space.

We start by describing some notion. By a graphon property we mean a measurable subset of the space Ξ^2, that is a family of symmetric two-variable $[0,1]$-valued functions that is invariant under graphon isomorphism. A graphon property is closed, if it is closed under the $\delta_\square$-metric, the closure $\overline{\mathcal{P}}$ of a graph property $\mathcal{P}$ constitutes of the graphons W such that $G_n \to W$ for some sequence $(G_n)_{n=1}^\infty \subset \mathcal{P}$. A graph property $\mathcal{P}$ is termed robust, if for every $\varepsilon > 0$ there exist $l \geq 1$ and $\delta > 0$ such that if G is a graph with at least l vertices such that $d_1(W_G, \overline{\mathcal{P}}) \leq \delta$, then $d_1(G, \mathcal{P}) \leq \varepsilon$.

The next result of [94] connects the testability of a graph property with that of its closure.

Theorem 3.4.5. *[94] The graph property $\mathcal{P}$ is testable if and only if it is robust and its closure $\overline{\mathcal{P}}$ is a testable graphon property.*

The second characterization in [94] is described in terms of $\delta_\square$-close induced subgraphs.

Theorem 3.4.6. *[94] For a graph property $\mathcal{P}$, its testability is equivalent to the following condition. For every $\varepsilon > 0$ there exists a $\delta > 0$ and a positive integer l such that for every $G \in \mathcal{P}$ and F that is an induced subgraph of G with at least l vertices and $\delta_\square(G, F) < \delta$, then $d_1(F, \mathcal{P}) < \varepsilon$.*

The next result was first proved by Alon and Shapira [10] in a stronger setting, namely where good tests are required to avoid false negatives, these test are called one-sided. The class of graph properties of concern here are hereditary in the sense, that they are closed under taking node-induced subgraphs, notable examples being perfect graphs, interval graphs, or triangle-free graphs. The first two classes cannot be characterized by a finite collection of forbidden induced subgraphs, whereas the testability of triangle-freeness is equivalent to the Triangle Removal Lemma. Prior to the result below it was known that properties closed under edge and node removal called monotone are testable, see [11].

Corollary 3.4.7. *[10][94][89] Every hereditary graph property is testable.*

Alon and Shapira [10] gave also a necessary condition for one-sided testability in form of semi-hereditary properties. The property $\mathcal{P}$ is semi-hereditary, if there exists a hereditary property $\mathcal{P}' \supset \mathcal{P}$, such that for any $\varepsilon > 0$ there exists an n_0 such that any graph of size at least n_0 that is ε-far from $\mathcal{P}$ has an induced subgraph not satisfying $\mathcal{P}'$. Partition based properties, such as having a large cut, are obviously not semi-hereditary, although they are testable, we have always have to settle with some small error probability, both false positives as well as negatives may occur. These are covered by Theorem 3.4.6, whose setting is more general.

3.5 Examples of testable properties and parameters

We introduce now a notion of efficient parameter testability. Definition 3.2.1 of testability does not ask for a specific upper bound on $k(\varepsilon)$ in terms of ε, but in applications the order of magnitude of this function may be an important issue once its existence has been verified. Therefore we introduce a more restrictive class of graph parameters, we refer to them as being efficiently testable.

Definition 3.5.1. *An r-graph parameter f is called β-testable for a family of measurable functions $\beta = \{\beta_i \mid \beta_i \colon \mathbb{R}^+ \to \mathbb{R}^+, i \in I\}$, if there exists an $i \in I$ such that for every $\varepsilon > 0$ and r-graph G we have*

$$\mathbb{P}(|f(G) - f(\mathbb{G}(\beta_i(\varepsilon), G))| > \varepsilon) < \varepsilon.$$

With slight abuse of notation we will also use the notion of β-testability for a family containing only a single function β. The term *efficient testability* will serve as shorthand for β-testability for some (family) of functions $\beta(\varepsilon)$ that are polynomial in $\frac{1}{\varepsilon}$. One could rephrase this in the light of Definition 3.5.1 by saying that a testable parameter f is efficiently testable if its sample complexity is polynomial in $1/\varepsilon$.

During the course of the thesis we will often deal with statistics that are required to be highly concentrated around their mean, this might be important for us even if their mean is not known to us in advance. A quite universal tool for this purpose is a Chernoff-type large deviation result, the Azuma-Hoeffding-inequality for martingales with bounded jumps. Mostly, we require the formulation given below, see e.g. [12] for a standard proof and a wide range of applications. We will also apply a more elaborate version of this concentration inequality in Chapter 4.

Lemma 3.5.2 (Azuma-Hoeffding-inequality)**.** *Let $(M_k)_{k=0}^n$ be a super-martingale with the natural filtration such that with probability 1 for every $k \in [n]$ we have $|M_k - M_{k-1}| \le c_k$. Then for every $\varepsilon > 0$ we have*

$$\mathbb{P}(|M_n - M_0| \ge \varepsilon) \le 2\exp\left(-\frac{\varepsilon^2}{2\sum_{k=1}^n c_k^2}\right).$$

We will list some examples of graph parameters, for which there is information available about their sample complexity implicitly or explicitly in the literature.

Example 3.5.3. One of the most basic testable simple graph parameters are subgraph densities $f_F(G) = t(F, G)$, where F is a simple graph. The next result was formulated as Theorem 2.5 in [91], see also for hypergraphs Theorem 11 in [49].

Lemma 3.5.4. *[91, 49] Let $\varepsilon > 0$ $q, r \geq 1$ be arbitrary. For any q-colored r-graphs F and G, and integer $k \geq |V(F)|$ we have*

$$\mathbb{P}(|t_{\mathrm{inj}}(F, G) - t_{\mathrm{inj}}(F, \mathbb{G}(k, G))| > \varepsilon) < 2\exp\left(-\frac{\varepsilon^2 k}{2|V(F)|^2}\right),$$

and

$$\mathbb{P}(|t(F, G) - t(F, \mathbb{G}(k, G))| > \varepsilon) < 2\exp\left(-\frac{\varepsilon^2 k}{18|V(F)|^2}\right). \tag{3.38}$$

For any q-colored r-graphon W we have

$$\mathbb{P}(|t(F, W) - t_{\mathrm{inj}}(F, \mathbb{G}(k, W))| > \varepsilon) < 2\exp\left(-\frac{\varepsilon^2 k}{2|V(F)|^2}\right),$$

and

$$\mathbb{P}(|t(F, W) - t(F, \mathbb{G}(k, W))| > \varepsilon) < 2\exp\left(-\frac{\varepsilon^2 k}{8|V(F)|^2}\right).$$

This implies that for any F that the parameter f_F is $O(\log(\frac{1}{\varepsilon})\varepsilon^{-2})$-testable. In the case of $(\mathcal{K}, r)$-graphs for arbitrary r the same as Lemma 3.5.4 holds, this can be shown by a straightforward application of the Azuma-Hoeffding inequality, Lemma 3.5.2, as in the original proofs.

Example 3.5.5. For $r = 2$, $q, n \in \mathbb{N}$, $J \in \mathbb{R}^{q\times q}$, $h \in \mathbb{R}^q$, and $G \in \Pi_n^2$ we consider the energy

$$\mathcal{E}_\phi(G, J, h) = \frac{1}{n^2} \sum_{1\leq i,j\leq q} J_{ij} e_G(\phi^{-1}(i), \phi^{-1}(j)) + \frac{1}{n} \sum_{1\leq i\leq q} h_i |\phi^{-1}(i)|, \tag{3.39}$$

of a partition $\phi\colon V(G) \to [q]$, and

$$\hat{\mathcal{E}}(G, J, h) = \max_{\phi\colon V(G)\to[q]} \mathcal{E}_\phi(G, J, h), \tag{3.40}$$

that is the *ground state energy* of the graph G (cf. [32]) with respect to J and h, where $e_G(S, T)$ denotes the number of edges going form S to T in G. These graph functions originate from statistical physics, for the rigorous mathematical treatment of the topic see e.g. Sinai's book [113]. The energy expression whose maximum is sought is also referred to as a Hamiltonian. In the literature this notion is also often to be found with negative sign or different normalization, more on this below.

This graph parameter can be expressed in the terminology applied for MAX-2CSP. Let the corresponding 2CSP formula to the pair (G,J) be F with domain $K = [q]$. The formula F is comprised of the constraints $(g_0; x_{(i,j)})$ for every edge (i, j) of G, where g_0 is the constraint type whose evaluation table is J, and additionally it contains n copies of $(g_1; x_i)$ for every vertex i of G, where g_1 is the constraint type in one variable with evaluation vector h. Then the optimal value of the objective function of the MAX-2CSP problem of the instance F is equal to $\hat{\mathcal{E}}(G, J, h)$. Note that this correspondence is consistent with the sampling procedure, that is, to the pair $(\mathbb{G}(k, G), J)$ corresponds the 2CSP formula $\mathbb{G}(k, F)$. Therefore $\hat{\mathcal{E}}(., J, h)$ has sample complexity $O(\frac{1}{\varepsilon^4})$(see [14],[98]).

These energies are directly connected to the number $\hom(G, H)$ of admissible vertex colorings of G by the colors $V(H)$ for a certain small weighted graph H. This was pointed out in [32], (2.16), namely

$$\frac{1}{|V(G)|^2} \ln \hom(G, H) = \hat{\mathcal{E}}(G, J) + O\left(\frac{1}{|V(G)|}\right), \tag{3.41}$$

where the edge weights of H are $\beta_{ij}(H) = \exp(J_{ij})$. The former line of thought of transforming ground state energies into MAX-2CSPs is also valid in the case of r-graphs and rCSPs for arbitrary r.

The results on the sample complexity of MAX-rCSP for $q = 2$ can be extended beyond the case of simple hypergraphs, higher dimensional Hamiltonians are also expressible as rCSP formulas. The generalization for arbitrary q and to r-graphons will follow in the next chapter. Additionally we note, that an analogous statement to (3.41) on testability of coloring numbers does not follow immediately for $r \geq 3$.

On the other hand, with the notion of the ground state energy available, we may rewrite the MAX-2CSP in a compact form as an energy problem. We will execute this task right away for limit objects. First, we introduce the ground state energy of a 2-kernel with respect to an interaction matrix J. The collection $\phi = (\phi_1, \dots, \phi_q)$ is a fractional q-partition of $[0, 1]$ with the components being measurable non-negative functions on $[0, 1]$, if for every $x \in [0, 1]$ it holds that $\sum_{i=1}^{q} \phi_i(x) = 1$.

Definition 3.5.6. *Let $q \geq 1$, $J \in \mathbb{R}^{q \times q}$. Then the ground state energy of the 2-kernel W with respect to J is*

$$\mathcal{E}(W, J) = \max_{\phi} \sum_{z \in [q]^2} J_z \int_{[0,1]^2} \phi_{z_1}(x) \phi_{z_2}(y) W(x, y) \mathrm{d}x \mathrm{d}y,$$

where ϕ runs over all fractional q-partitions of $[0, 1]$.

Let $K = [q]$, $L = \{0, 1, \dots, d\}^{[q]^2}$ and $(F_n)_{n=1}^{\infty}$ be a convergent sequence of 2CSP formulas. Consider the corresponding sequence of graphs $\mathfrak{eval}(F_n) = (\tilde{F}_n^z)_{z \in [q]^2}$ for each n, and let $W = (W^z)_{z \in [q]^2}$ be the respective limit. Let f be the $(L, 2)$-graph parameter so that $f(\mathfrak{eval}(F))$ is equal to the density of the MAX-2CSP value for the instance F. Then it is not hard to see that f can be extended to the limit space the following way

$$f(W) = \max_{\phi} \sum_{i,j=1}^{q} \int_{[0,1]^2} \phi_i(x) \phi_j(y) W^{(i,j)}(x, y) \mathrm{d}x \mathrm{d}y,$$

where ϕ runs over all fractional q-partitions of $[0, 1]$. The formula is a special case of the *layered ground state energy* with the interaction matrices defined by $J^{i,j}(k, l) = \mathbb{1}_i(k) \mathbb{1}_j(l)$ that is defined below in Chapter 4.

Example 3.5.7. The efficiency of testing a graph parameter can be investigated in terms of some additional continuity condition in the $\delta_\square$ metric. Direct consequence of results from [30] will be presented in the next lemma.

Lemma 3.5.8. *Let f be a simple graph parameter that is α-Hölder-continuous in the $\delta_\square$ metric in the following sense: There exists a $C > 0$ such that for every $\varepsilon > 0$ there exists $n_0(\varepsilon)$ so that if for the simple graphs G_1, G_2 it holds that $|V(G_1)|, |V(G_2)| \geq n_0(\varepsilon)$ and $\delta_\square(G_1, G_2) \leq \varepsilon$, then $|f(G_1) - f(G_2)| \leq C\delta_\square^\alpha(G_1, G_2)$. Then f is $\max\{2^{O\left(\frac{1}{\varepsilon^{2/\alpha}}\right)}, n_0(\varepsilon)\}$-testable.*

Proof. To see this, let us fix $\varepsilon > 0$. Then for an arbitrary simple graph G with $|V(G)| \geq n_0(\varepsilon)$ and $k \geq n_0(\varepsilon)$ we have

$$|f(G) - f(\mathbb{G}(k, G))| \leq C\,[\delta_\square(G, \mathbb{G}(k, G))]^\alpha < C\left(\frac{10}{\sqrt{\log_2 k}}\right)^\alpha, \tag{3.42}$$

with probability at least $1 - \exp(-\frac{k^2}{2\log_2 k})$. The last probability bound in (3.42) is the statement of Lemma 3.3.9, first proved in [30]. We may rewrite (3.42) by setting $\varepsilon = C\left(\frac{10}{\sqrt{\log_2 k}}\right)^\alpha$, the substitution implies that f is $2^{O\left(\varepsilon^{-2/\alpha}\right)}$-testable, whenever $n_0(\varepsilon) \leq 2^{O(\varepsilon^{-2/\alpha})}$. □

This latter approach is hard to generalize in a meaningful way to r-graphs for $r \geq 3$ because of the absence of a suitable metric, see the discussion above. The converse direction, namely formulating a qualitative statement about the continuity of f with respect to $\delta_\square$ obtained from the information about the sample complexity is also a worthwhile problem.

3.6 Further aspects of statistical physics

In this section we further motivate the term energies above for parameters determined by coloring configurations on graphs using the terminology of mean field theory in statistical physics.

We consider first simple graphs, this translates to pairwise interaction between sites in terms of physics. Let our model be given by $q \geq 1$, a symmetric real matrix J, and a q-dimensional real vector h, let further G be a simple graph. We define a probability distribution over the set of q-partitions ϕ of the node set $V(G)$ corresponding to the thermodynamical equilibrium called the Gibbs measure, each configuration ϕ occurs with probability proportional to $\exp(|V(G)|\mathcal{E}_\phi(G, J, h))$ (see (3.39). The normalizing factor $Z(G, J, h)$, i.e., the sum of the exponentials over the possible ϕ configurations, is of central importance, its limit behavior when $V(G)$ tends to infinity discloses several features of the Gibbs measure, and detects phase transitions. The measure is clearly a function of the entries of J and h, and roughly said a phase transition is a point where the correlation between the marginal distributions of ϕ at the nodes quickly changes, the configuration obtained following the Gibbs measure shifts from typically ordered to typically disordered. The parameters J and h of the measure are often driven by one variable T that stands for the temperature, this is the case in the Ising model.

In the common definition used by physicists this weight reads as

$$\exp\left(-\sum_{uv\in E(G)} J_{\phi(u)\phi(v)} + \sum_{u\in V(G)} h_{\phi(u)}\right),$$

which is suitable for cases where the average degree of the graphs is bounded from above by a constant, an example is the d-dimensional cubic lattice. A more universal weighting would be $\exp(\frac{|V(G)|^3}{|E(G)|}\mathcal{E}_\phi(G,J,h))$, that is just the previous formula with an average degree normalization in the exponent in order to ensure that it is $O(|V(G)|)$. We will use the first formulation without a negative sign and with zero magnetization h for its simplicity in the dense case.

For example, in the well-known Ising model, see below, each node can take one of two possible states (spins), where adjacent node pairs in the same state excel a repulsive interaction contributing -1 to the Hamiltonian, and in different states an attractive one that is reflected by a +1 contribution to the value of the Hamiltonian.

The overwhelming majority of the statistical physics literature deals with models on (possible infinite) lattice structures, whereas our study mainly concerns dense structures. The connection is given by a simplified version of the original physics model in which all pairs of nodes interact with each other in the same way with the further hypothesis that the fluctuation of the spin values is always negligible.

This method is called mean field approximation, it has no direct practical meaning for the description of an actual lattice system, however the behavior of the model often serves as a source of intuition regarding more realistic scenarios. A huge advantage of the model is that tracing the impact of the change of the interaction strength is less involved, also, it fits well into our framework as the case of complete graphs.

Free energies of r-graphs and kernels We will study the testability of the logarithm of the partition function per node, also called free energy of the system for r-uniform hypergraphs, this means a generalization of the pairwise interaction between adjacent nodes in certain states to multi-site interaction. For a fixed description of the characteristics of the states the free energy is a graph a parameter, and it is closely related to the ground state energies in a natural way. Such an aspect was first studied in [32].

Definition 3.6.1. *Let $q \geq 1$, J be a symmetric real r-array of size q. For an $(\mathbb{R}, r)$-graph G the partition function with respect to J is defined by*

$$Z(G,J) = \sum_{\phi\colon V(G)\to[q]} \exp(|V(G)|\mathcal{E}_\phi(G,J)), \tag{3.43}$$

where the sum goes over all q-partitions of $|V(G)|$, and

$$\mathcal{E}_\phi(G,J) = \frac{1}{|V(G)|^r} \sum_{i_1,\dots,i_r=1}^{q} J_{i_1,\dots,i_r} A_G(\phi^{-1}(i_1),\dots,\phi^{-1}(i_r)).$$

The ground state energy of G with respect to J is

$$\hat{\mathcal{E}}(G,J) = \max_\phi \mathcal{E}_\phi(G,J),$$

where the maximum runs over all q-partitions of $|V(G)|$.

The free energy of G with respect to J is

$$\hat{\mathcal{F}}(G,J) = -\frac{1}{|V(G)|} \ln Z(G,J). \tag{3.44}$$

We remark that if in the above definition the multiplicative factors $|V(G)|^{r-1}$ in the exponents in (3.43) would be replaced by $|V(G)|^r$ (thus, following the original terminology when studying lattices in statistical physics) and the normalization in the free energy expression in (3.44) would be changed to $\frac{1}{|V(G)|^r}$ at the same time, then we would obtain the situation described in (3.41). Therefore the modified definition would imply that the free energy is asymptotically equal to the ground state energy of the same system, meaning that their difference is $O(1/|V(G)|)$.

We also mention that allowing different levels of interactions on the set of sites up to the arity bound r would lead out of the space of uniform hypergraphs, however such hypergraphs can be encoded by uniform edge-colored ones.

The original mean field approximation for the Ising model, also called Curie-Weiss model, is a special instance of the above definition in the sense that G is assumed to be a complete graph K_n, the behavior of the free energy is studied when n is tending to infinity. This type of question regarding the limit behavior can be extended to any convergent sequence of dense r-graph sequences using the next counting lemma.

Lemma 3.6.2. *Let $q \geq 1$, J be a symmetric real r-array of size q. For two r-graphs G and H with the common vertex set $[n]$ we have that*

$$|\hat{\mathcal{E}}(G,J) - \hat{\mathcal{E}}(H,J)| \leq q^r d_\square(G,H)\|J\|_\infty,$$

and

$$|\hat{\mathcal{F}}(G,J) - \hat{\mathcal{F}}(H,J)| \leq q^r d_\square(G,H)\|J\|_\infty.$$

Proof. Let ϕ be an arbitrary q-partition of $[n]$. Then

$$|\mathcal{E}_\phi(G,J) - \mathcal{E}_\phi(H,J)| \leq \sum_{i_1,\dots,i_r=1}^{q} |J_{i_1,\dots,i_r}|\,|A_G(\phi^{-1}(i_1),\dots,\phi^{-1}(i_r)) - A_H(\phi^{-1}(i_1),\dots,\phi^{-1}(i_r))|$$

$$\leq q^r d_\square(G,H)\|J\|_\infty.$$

Let ϕ_0 be such that $\mathcal{E}_{\phi_0}(G,J) = \hat{\mathcal{E}}(G,J)$. Then

$$\begin{aligned}\hat{\mathcal{E}}(G,J) - \hat{\mathcal{E}}(H,J) &= \mathcal{E}_{\phi_0}(G,J) - \hat{\mathcal{E}}(H,J)\\ &\leq \mathcal{E}_{\phi_0}(G,J) - \mathcal{E}_{\phi_0}(G,J) \leq q^r d_\square(G,H)\,\|J\|_\infty.\end{aligned}$$

Bounding the negative of the above difference the same way gives the desired result. Further,

$$\begin{aligned}\frac{Z(G,J)}{Z(H,J)} &= \frac{\sum_\phi \exp(n\mathcal{E}_\phi(G,J))}{\sum_\phi \exp(n\mathcal{E}_\phi(H,J))}\\ &\leq \max_\phi \exp(n(\mathcal{E}_\phi(G,J) - \mathcal{E}_\phi(H,J))) \leq \exp(nq^r d_\square(G,H)\|J\|_\infty).\end{aligned}$$

By symmetry we have $\frac{Z(H,J)}{Z(G,J)} \leq \exp(nq^r d_\square(G,H)\|J\|_\infty)$, thus

$$|\hat{\mathcal{F}}(G,J) - \hat{\mathcal{F}}(H,J)| \leq 1/n|\ln Z(G,J) - \ln Z(H,J)| \leq q^r d_\square(G,H)\|J\|_\infty.$$

□

The version of the free energy for naive r-kernels and graphons seems at first less natural, than in the case of ground state energies.

Definition 3.6.3. *[32] Let $q \geq 1$, J be a symmetric real r-array of size q. Then the energy of a naive r-kernel W with respect to J and the fractional q-partition ϕ is*

$$\mathcal{E}_\phi(W,J) = \sum_{i_1,\dots,i_r=1}^{q} J_{i_1,\dots,i_r} \int_{[0,1]^r} W(x_1,\dots,x_r) \prod_{j=1}^{r} \phi_{i_j}(x_j)\mathrm{d}x_1\dots\mathrm{d}x_r,$$

the ground state energy of W with respect to J is

$$\mathcal{E}(W,J) = \sup_\phi \mathcal{E}_\phi(W,J).$$

The free energy of W with respect to J is

$$\mathcal{F}(W,J) = -\sup_\phi \left[\mathcal{E}_\phi(W,J) + \mathrm{ent}(\phi)\right],$$

where the supremum is taken over all fractional q-partitions of $[0,1]$, and

$$\mathrm{ent}(\phi) = -\int_0^1 \sum_{i=1}^{q} \phi_i(x)\log(\phi_i(x))\mathrm{d}x.$$

As a usual convention we set the function $x\log x$ to be 0 at 0.

The corresponding counting lemma for the graphon case is analogous to the graph case including its proof.

Lemma 3.6.4. *Let $q \geq 1$, J be a symmetric real r-array of size q. For two naive r-kernels U and W of common size n we have that*

$$|\mathcal{E}(U, J) - \mathcal{E}(W, J)| \leq q^r \delta_\square(U, W) \|J\|_\infty \|U\|_\infty \|W\|_\infty,$$

and

$$|\mathcal{F}(U, J) - \mathcal{F}(W, J)| \leq q^r \delta_\square(U, W) \|J\|_\infty \|U\|_\infty \|W\|_\infty.$$

The connection between the two versions of free energies $\hat{\mathcal{F}}(G, J)$ and $\mathcal{F}(W_G, J)$ for a fixed r-graph is not as straight-forward as in the case of GSEs, as to be shown in Chapter 4 below. We state a bound on their deviation without giving the detailed proof here, as it is only a slight generalization of Theorem 5.8. in [32] for the case $r = 2$, the alterations with respect to that proof are trivial.

Theorem 3.6.5. *[32] Let $q \geq 1$, J be a symmetric real r-array of size q. For an r-graph G we have*

$$|\hat{\mathcal{F}}(G, J) - \mathcal{F}(W_G, J)| \leq O\left(\frac{q^r}{\sqrt{\ln |V(G)|}}\right).$$

The proof method of [32] goes as follows. There is no direct way to relate the two quantities above, but it is possible to approximate $\mathcal{F}(W_G, J)$ by free energies of the equitable blow-ups of G, i.e., $\hat{\mathcal{F}}(G[k], J)$. To exploit this property one has to create a template graph for G with less number of nodes using the Weak Regularity Lemma to obtain a compressed representation H. Since by Lemma 3.6.2 above the free energies of blow-ups of H are close to the free energies of G, it only requires some technical surgery to estimate the impact of adding some small number of isolated vertices to finish the proof. In fact, sill an application of the continuous version of the counting lemma, Lemma 3.6.4, is needed as one approximates $\mathcal{F}(W_H, J)$ by $\hat{\mathcal{F}}(H[k], J)$ to conclude the proof with the aid of the first remark regarding the closeness of these two quantities.

As an immediate consequence of Theorem 3.6.5 we obtain the testability of the r-graph parameter $f(G) = \hat{\mathcal{F}}(G, J)$.

Corollary 3.6.6. *Let $q \geq 1$, J be a symmetric real r-array of size q, and let $\varepsilon > 0$ be an arbitrary real. There exists a $c > 0$, such that for any r-graph G and $k \geq \exp(c/\varepsilon^2)$ we have*

$$|\hat{\mathcal{F}}(G, J) - \hat{\mathcal{F}}(\mathbb{G}(k, G), J)| \leq \varepsilon$$

with probability at least $1 - \varepsilon$.

Proof. Fix $\varepsilon > 0$, and let G and k be as in the statement, further let F denote $\mathbb{G}(k, G)$. We have

$$|\hat{\mathcal{F}}(G,J) - \hat{\mathcal{F}}(\mathbb{G}(k,G),J)| \leq |\hat{\mathcal{F}}(G,J) - \mathcal{F}(W_G,J)|$$
$$+ |\mathcal{F}(W_G,J) - \mathcal{F}(W_F,J)| + |\mathcal{F}(W_F,J) - \hat{\mathcal{F}}(F,J)|.$$

We set c to be large enough so that the first and the third term are each bounded above by $\varepsilon/3$ due to Theorem 3.6.5. To estimate the middle term we apply Lemma 3.6.4 together with Lemma 3.3.9, perhaps by increasing c, we have

$$|\mathcal{F}(W_G,J) - \mathcal{F}(W_F,J)| \leq q^r \|J\|_\infty \delta_\square(G,F) \leq \varepsilon/3$$

with probability at least $1 - \varepsilon$.

□

Spin models in physics The problems we dealt with are defined on so-called spin models, see de la Harpe and Jones [42], and Wu [116, 117] for connections between the combinatorial and the physics aspects. Next we will list some well-known examples of these models from the physics literature, and describe how they fit into the above framework. Another important family of graph parameters appearing in statistical physics are vertex models, where a configuration in the graph case is given by an edge coloring, and the weight of a node in the energy formula is evaluated according to the states of the incident edges and a given weight function. These models also have a large literature, see [42], however their treatment would exceed the content of this thesis.

The spin models listed below have all a finite state space (except for the n-vector model, but this can be taken care of in some cases) with bounded interaction.

Ising model The most basic model for simple graphs is the Ising model, here $q = 2$ and $J_{i,j} = (-1)^{i+j}$, further $\mathcal{E}_\phi(G,J) = \frac{1}{|V(G)|^2} \sum_{uv \in E(G)} K(-1)^{\phi(u)+\phi(v)}$. Depending on the sign of K, sites in different states attract ($K < 0$) or repulse ($K > 0$) each other. For $K > 0$ it is also called the ferromagnetic Ising model, as for large K the system aligns into one state according to the Gibbs measure with high probability. The case $G = K_n$ is known as the Curie-Weiss model.

Potts model Let $q \geq 1$ be arbitrary, then the standard Potts model is given by $J_{i,j} = K\mathbb{1}_i(j)$, where $\mathbb{1}_x$ is the indicator of its index, so the weight function is $\mathcal{E}_\phi(G,J) = \frac{1}{|V(G)|^2} \sum_{u,v \in V(G)} K\mathbb{1}_{\phi(u)}(\phi(v))$. An alternative variant is known as the planar Potts model (its mean field specialization is also known as the Curie-Weiss clock model) defined by $J_{i,j} = K\cos(\frac{2\pi(i-j)}{q})$, for $q = 2, 3$ these two models are equivalent.

We mention that discarding the quadratic normalization in our formulation of the Potts model with considering q and K as variables, it is equivalent to the Tutte polynomial of graphs, that is also known as the dichromatic polynomial and contains information about a large variety of graph invariants such as the number of proper k-colorings, or spanning trees.

Biggs model Let q be arbitrary, and $T\colon \mathbb{Z}_q \to \mathbb{R}$ such that $T(i) = T(-i)$. The Biggs model generalizes the Potts model, it is given by $J_{i,j} = T((i-j) \mod q)$, in other words, $\mathcal{E}_\phi(G, J) = \frac{1}{|V(G)|^2} \sum_{uv \in E(G)} T((\phi(u) - \phi(v)) \mod q)$.

***n*-vector model** The most basic continuous state spin model family is called the n-vector model, where the spins come from the n-dimensional unit sphere, the interaction strength between two states is the inner product of the representing vectors, so $J_{uv} = \langle u, v\rangle$. The partition function in this setup is defined as an integral over all possible configurations with the usual spherical measure. If we specify the dimension n of the sphere of possible spin values we get other well-studied models. The case $n = 1$ corresponds to the Ising model above, the case $n = 2$ is also referred to as the XY-model, and $n = 3$ is known as the classical Heisenberg model. However, in terms of testability of the free energies the models with fixed dimension can be well approximated by finite state space models by discretizing the unit sphere.

We remark that in our framework it might be interesting to further generalize the above model and allow spins from the n-dimensional unit sphere for graphs with n vertices, this model seems to have close ties to semidefinite relaxations of combinatorial optimization problems on graphs. Unfortunately, the discretization method above would not lead to testability in this case, since the size of the ε-net for the auxiliary states grows with the vertex cardinality of the inspected graphs, and so does the bound in Corollary 3.6.6. So a corresponding testability result in this case still would require some new ideas.

Multisite Potts model This model is a generalization of the standard Potts model in the sense that influences of site groupings of higher rank also receive weights in the Hamiltonian, in combinatorics these groupings can be captured by hypergraphs defined on the set of sites. For $r = 3$ and arbitrary q, the 3-site Potts model corresponds to the interaction array $J_{i,j,k} = K\mathbb{1}_i(j)\mathbb{1}_j(k)\mathbb{1}_k(i)$ for $i, j, k \in [q]$, and the energy weight function takes the form

$$\mathcal{E}_\phi(G, J) = \frac{1}{|V(G)|^3} \sum_{uvw \in E(G)} K\mathbb{1}_{\phi(u)}(\phi(v))\mathbb{1}_{\phi(v)}(\phi(w))\mathbb{1}_{\phi(w)}(\phi(u)).$$

Zero-site interactions can be interpreted as the influence of some external magnetic field on the energy weight of the configurations. For example, a general instance of the Potts model with interaction graph $G = (G_1, G_2)$ on the site set $V(G)$ with G_1 being a 3-graph and G_2 being a 2-graph the energy weight of the spin configuration ϕ looks like

$$\mathcal{E}_\phi((G_1, G_2), J) = \frac{1}{|V(G)|^3} \sum_{uvw \in E(G_1)} K_1 \mathbb{1}_{\phi(u)}(\phi(v))\mathbb{1}_{\phi(v)}(\phi(w))\mathbb{1}_{\phi(w)}(\phi(u))$$

$$+ \frac{1}{|V(G)|^2} \sum_{uv \in E(G_2)} K_2 \mathbb{1}_{\phi(u)}(\phi(v)) + \frac{1}{|V(G_1)|} \sum_{u \in V(G)} K_3 \mathbb{1}_1(\phi(u)).$$

Remark 3.6.7. We would like to mention the consequences of Corollary 3.6.6 in terms of computational complexity for the value of the free energy. In general, it requires exponential time to determine the free energy, since the partition function is the sum of q^n energy terms, however Corollary 3.6.6 directly implies a PTAS for the value of the free energies that runs in constant time for fixed $\varepsilon > 0$, meaning that in this time we can compute an additive ε-approximation with high probability. Since the spin models listed above all have a finite state space (except for the n-vector model, but this can be handled if the dimension of the state space is bounded) and uniformly bounded interaction strength, Corollary 3.6.6 applies, with other words we can say that for these models the free energies considered as r-graph parameters are testable. As mentioned above, the situation in the n-vector model with increasing state dimension has not been resolved yet.

It is still an interesting question if it is possible to improve on upper bounds for the sample complexity of the parameter given by $f(G) = \mathcal{F}(W_G, J)$ using the framework of the forthcoming Chapter 4 concerning ground state energies.

CHAPTER 4

Testability of the ground state energy

4.1 Introduction

In this chapter we introduce a generalization of the notion of the ground state energy of graphs from [32], see Definition 3.5.6, to the space of r-graphons and r-kernels and reformulate the results of [14] regarding sample complexity of MAX-rCSP in that framework. The parameter derived from the maximal constraint satisfaction problem will also serve as an example for an efficiently testable parameter of the corresponding colored hypergraph. We will further generalize the main result of [14] in several directions.

Assume that $\mathcal{K}$ is a compact Polish space, and r is a positive integer. First we provide the basic definition of the energy of a $(\mathcal{K}, r)$-graphon $W\colon [0,1]^{\mathfrak{h}([r])} \to \mathcal{K}$ with respect to some $q \geq 1$, an r-array $J \in C(\mathcal{K})^{q\times\cdots\times q}$, and a fractional partition $\phi = (\phi_1, \ldots, \phi_q)$. With slight abuse of notation, the graphons in the upcoming parts of the section assume both the $\mathcal{K}$-valued and the probability measure valued form, it will be clear from the context which one of them is meant.

Recall Definition 3.6.3 of the energies of naive r-kernels, the version for true $(\mathcal{K}, r)$-graphons is

$$\mathcal{E}_\phi(W, J) = \sum_{z_1,\ldots,z_r=1}^{q} \int_{[0,1]^{\mathfrak{h}([r])}} J_{z_1,\ldots,z_r}(W(x_{\mathfrak{h}([r])})) \prod_{j=1}^{r} \phi_{z_j}(x_{\{j\}}) \mathrm{d}\lambda(x_{\mathfrak{h}([r])}). \tag{4.1}$$

The value of the above integral can be determined by first integrating over the coordinates corresponding to subsets of $[r]$ with at least two elements, and then over the remaining ones. The interior partial integral is then not dependent on ϕ, so it can be calculated in advance in the case when we want to optimize over all choices of fractional partitions. Therefore focusing attention on the naive kernel version does not lead to any loss of generality in terms of testing, see below.

When dealing with a so-called integer partition $\phi = (\mathbb{1}_{T_1}, \ldots, \mathbb{1}_{T_q})$, one is able to rewrite the former expression (4.1) as

$$\mathcal{E}_\phi(W,J) = \sum_{z_1,\dots,z_r=1}^{q} \int_{p^{-1}_{\mathfrak{h}([r],1)}(T_{z_1}\times\cdots\times T_{z_r})} J_{z_1,\dots,z_r}(W(x_{\mathfrak{h}([r])}))\mathrm{d}\lambda(x_{\mathfrak{h}([r])}),$$

where p_D stands for the projection of $[0,1]^{\mathfrak{h}([r])}$ to the coordinates contained in the set D.

The energy of a $(\mathcal{K},r)$-graph G on k vertices with respect to the $J \in C(\mathcal{K})^{q\times\cdots\times q}$ for the fractional q-partition $x_n = (x_{n,1},\dots,x_{n,q})$ for $n = 1,\dots,k$ (i. e., $x_{n,m} \in [0,1]$ and $\sum_m x_{n,m} = 1$) is defined as

$$\mathcal{E}_{\mathrm{x}}(G,J) = \frac{1}{k^r} \sum_{z_1,\dots,z_r=1}^{q} \sum_{n_1,\dots,n_r=1}^{k} J_{z_1,\dots,z_r}(G(n_1,\dots,n_r)) \prod_{j=1}^{r} x_{n_j,z_j}. \tag{4.2}$$

In the case when $\mathcal{K} = \{0,1\}$ and $J_{z_1,\dots,z_r}(x) = a_{z_1,\dots,z_r}\mathbb{1}_1(x)$ is a constant multiple of the indicator function of 1 we retrieve the original GSE notion in Example 3.5.5 and Definition 3.5.6.

Remark 4.1.1. Ground state energies and subgraph densities are Lipschitz continuous graph parameters in the sense of Lemma 3.5.8 ([30],[32]), but that result implies much weaker upper bounds on the sample complexity, than the best ones known to date. This is due to the fact, that $\delta_\square(G,\mathbb{G}(k,G))$ decreases with magnitude $1/\sqrt{\log k}$ in k, which is the result of the difficulty of finding a near optimal overlay between two graphons through a measure preserving permutation of $[0,1]$ in order to calculate their $\delta_\square$ distance. On the other hand, if the sample size $k(\varepsilon)$ is exponentially large in $1/\varepsilon$, then the distance $\delta_\square(G,\mathbb{G}(k,G))$ is small with high probability, therefore all Hölder-continuous graph parameters at G can be estimated simultaneously with high success probability by looking at the values at $\mathbb{G}(k,G)$).

Next we introduce the layered version of the ground state energy. This is a generalized optimization problem where we wish to obtain the optimal value corresponding to fractional partitions of the sums of energies over a finite layer set.

Definition 4.1.2. *Let $\mathfrak{E}$ be a finite layer set, $\mathcal{K}$ be a compact set, and $W = (W^{\mathfrak{e}})_{\mathfrak{e}\in\mathfrak{E}}$ be a tuple of $(\mathcal{K},r)$-graphons. Let q be a fixed positive integer and let $J = (J^{\mathfrak{e}})_{\mathfrak{e}\in\mathfrak{E}}$ with $J^{\mathfrak{e}} \in C(\mathcal{K})^{q\times\cdots\times q}$ for every $\mathfrak{e} \in \mathfrak{E}$. For a $\phi = (\phi_1,\dots,\phi_q)$ fractional q-partition of* $[0,1]$ *let*

$$\mathcal{E}_\phi(W,J) = \sum_{\mathfrak{e}\in\mathfrak{E}} \mathcal{E}_\phi(W^{\mathfrak{e}},J^{\mathfrak{e}})$$

and let

$$\mathcal{E}(W,J) = \max_\phi \mathcal{E}_\phi(W,J),$$

denote the layered ground state energy, where the maximum runs over all fractional q-partitions *of* $[0,1]$.

We define for $G = (G^{\mathfrak{e}})_{\mathfrak{e}\in\mathfrak{E}}$ *the energy* $\mathcal{E}_{\mathrm{x}}(G,J)$ *analogously as the energy sum over* $\mathfrak{E}$*, see (4.2) above, and* $\hat{\mathcal{E}}(G,J) = \max_{\mathrm{x}} \mathcal{E}_{\mathrm{x}}(G,J)$ *where the maximum runs over integer q-partitions* ($x_{n,m} \in \{0,1\}$ *), respectively* $\mathcal{E}(G,J) = \max_{\mathrm{x}} \mathcal{E}_{\mathrm{x}}(G,J)$*, where the maximum is taken over all fractional q-partitions* x.

Now we will rewrite the unweighted boolean limit MAX-rCSP (recall Definition 2.1.2) as a layered ground state energy problem. Let $\mathfrak{E} = \{0,1\}^r$, $\mathcal{K} = \{0,1,\dots,2^r\}$, $W = (W^z)_{z\in\{0,1\}^r}$ with W^z being $(\mathcal{K},r)$-graphons, and let

$$\alpha(W) = \max_{\phi} \sum_{z\in\{0,1\}^r} \int_{[0,1]^{\mathfrak{h}([r])}} \prod_{j=1}^{r} \phi(x_{\{j\}})^{z_j}(1-\phi(x_{\{j\}}))^{1-z_j} W^z(x)\mathrm{d}\lambda(x),$$

where the maximum is taken over all measurable functions $\phi\colon [0,1] \to [0,1]$. If $\mathfrak{eval}(F) = (F^z)_{z\in\{0,1\}^r}$ is a $(\mathcal{K}^{\mathfrak{E}},r)$-graph corresponding to a boolean rCSP formula F with k variables, then the finite integer version of α is given by

$$\hat{\alpha}(\mathfrak{eval}(F)) = \max_{\mathrm{x}} \frac{1}{k^r} \sum_{z\in\{0,1\}^r} \sum_{n_1,\dots,n_r=1}^{k} F^z(n_1,\dots,n_r) \prod_{j=1}^{r} x_{n_j,z_j},$$

where the maximum runs over integer 2-partitions of $[k]$. It is clear that $\hat{\alpha}(\mathfrak{eval}(F))$ is equal to the density of the optimum of the MAX-rCSP problem of F.

We return to the general setting and summarize the involved parameters in the layered ground state energy problem. These are the dimension r, the layer set $\mathfrak{E}$, the number of states q, the color set $\mathcal{K}$, the finite or limit case. Our main theorem on the chapter will be a generalization of the following theorem on sample complexity of rCSPs with respect to these factors.

The main result of [14] was the following.

Theorem 4.1.3. *[14] Let F be an unweighted boolean rCSP formula. Then for any* $\varepsilon > 0$ *and* $\delta > 0$ *we have that for* $k \in \mathcal{O}(\varepsilon^{-4}\log(\frac{1}{\varepsilon}))$ *it holds that*

$$\mathbb{P}\left(|\hat{\alpha}(\mathfrak{eval}(F)) - \hat{\alpha}(\mathbb{G}(k,\mathfrak{eval}(F)))| > \varepsilon\right) < \delta.$$

The upper bound on k in the above result was subsequently improved by Mathieu and Schudy [98] to $k \in O(\varepsilon^{-4})$. We will see in what follows that also the infinitary version of the above statement is true. It will be stated in terms of layered ground state energies of edge colored hypergraphs, and will settle the issue regarding the efficiency of testability of the mentioned parameters in the greatest generality with respect to the previously highlighted aspects. However, what the exact order of the magnitude of the sample complexity of the MAX-rCSP and the GSE problem is remains an open question.

In order to simplify the analysis we introduce the *canonical form* of the problem, that denote layered ground state energies of $[q]^r$-tuples of $([-d,d],r)$-graphons with the special interaction r-arrays $\hat{J}^z$ for each $z \in [q]^r$, that have the identity function $f(x) = x$ as the $(z_1, \dots, z_r)$ entry and the constant 0 function for the other entries. In most of what follows we will drop the dependence on J in the energy function when it is clear that we mean the aforementioned canonical $\hat{J}$, and will employ the notation $\mathcal{E}_x(G)$, $\mathcal{E}(G)$, $\hat{\mathcal{E}}(G)$, $\mathcal{E}_\phi(W)$, and $\mathcal{E}(W)$ (dependence on q is hidden in the notation), where G and W are $[q]^r$-tuples of $([-d,d],r)$-graphs and graphons, respectively. We are ready to state the main result of the chapter.

Theorem 4.1.4. *Let $r \geq 1$, $q \geq 1$, and $\varepsilon > 0$. Then for any $[q]^r$-tuple of $([-\|W\|_\infty, \|W\|_\infty], r)$-graphons $W = (W^z)_{z\in[q]^r}$ and $k \geq \Theta^4 \log(\Theta) q^r$ with $\Theta = \frac{2^{r+7} q^r r}{\varepsilon}$ we have*

$$\mathbb{P}(|\mathcal{E}(W) - \hat{\mathcal{E}}(\mathbb{G}(k,W))| > \varepsilon \|W\|_\infty) < \varepsilon. \tag{4.3}$$

We outline the organization of the chapter. We proceed with the proof of Theorem 4.1.4 in the next section, we employ the Cut Decomposition method that is a low-rank type approximation for arrays resembling the singular value decomposition for matrices closely related to the Weak Regularity Lemma, and linear programming duality.

This is followed by applications of the proof methodology to some variant energy problems. In particular, we analyze the testability of the generalization of ground state energies with an external magnetic field that is embodied in an additional bias towards certain states, and the microcanonical version of ground state energies, where we optimize over configurations such that their number of vertices that take certain states are fixed. We provide for these problems the first explicit upper bounds on testability in Theorem 4.3.2 and Theorem 4.3.6, respectively. For the finitary version of the microcanonical energy problem a similar question was investigated by Fernandez de la Vega, Kannan, and Karpinski [53] by imposing a finite number of additional global constraints with unbounded arity for common rCSPs.

Furthermore, the continuous version of the quadratic assignment problem is treated below in a sample complexity context, see Corollary 4.3.9, this subject is related to recent development in the topic of approximate graph isomorphism and homomorphism problems, see [88] and [22].

In the subsequent section we relax the condition on the colors to be from a compact space and treat graphs and graphons with real unbounded weights. We prove testability for graphons under an L^p condition, see Theorem 4.4.3.

We conclude the chapter with a topic of somewhat different flavor. In [32] it was shown that graph convergence is equivalent to convergent ground state energies, but is strictly stronger that convergence of unrestricted ground state energies. We introduce lower threshold ground state energies, where we optimize over configurations for that each state is represented by a minimal number vertices corresponding to the threshold. We prove a convergence hierarchy for these parameter families, see Theorem 4.5.11.

4.2 The main problem

This section contains the proof of Theorem 4.1.4 we will proceed loosely along the lines of the proof of Theorem 4.1.3 from [14] with most of the required lemmas being refinements of the respective ones in the proof of that theorem. We will formulate and verify these auxiliary lemmas one after another, afterwards we will compile them to prove the main statement. The arguments made in [14] carry through adapted to our continuous setting with some modifications, and we will also draw on tools from [30] and [32]. A direct consequence of Theorem 4.1.4 is the corresponding result for layered ground state energies.

Corollary 4.2.1. *Let $\mathfrak{E}$ be a finite layer set, $\mathcal{K}$ a compact Polish color set, $q \geq 1$, r-arrays $J = (J^{\mathfrak{e}})_{\mathfrak{e}\in\mathfrak{E}}$ with $J^{\mathfrak{e}} \in C(\mathcal{K})^{q\times\cdots\times q}$, and $\varepsilon > 0$. Then we have that for any $\mathfrak{E}$-tuple of $(\mathcal{K}, r)$-graphon $W = (W^{\mathfrak{e}})_{\mathfrak{e}\in\mathfrak{E}}$ and $k \geq \Theta^4 \log(\Theta) q^r$ with $\Theta = \frac{2^{r+7}q^r r}{\varepsilon}$ that*

$$\mathbb{P}(|\mathcal{E}(W,J) - \hat{\mathcal{E}}(\mathbb{G}(k,W),J)| > \varepsilon|\mathfrak{E}|\,\|J\|_\infty\,\|W\|_\infty) < \varepsilon.$$

Proof. We make no specific restrictions on the color set $\mathcal{K}$ and on the set $\mathfrak{E}$ of layers except for finiteness of the second, therefore it will be convenient to rewrite the layered energies $\mathcal{E}_\phi(W,J)$ into a more universal form as a sum of proper Hamiltonians in order to suppress the role of $\mathcal{K}$ and $\mathfrak{E}$. Let

$$\begin{aligned}\mathcal{E}_\phi(W,J) &= \sum_{\mathfrak{e}\in\mathfrak{E}} \sum_{z_1,\ldots,z_r\in[q]} \int_{[0,1]^{\mathfrak{h}([r])}} \prod_{j\in[q]} \phi_{z_j}(x_{\{j\}}) J^{\mathfrak{e}}_{z_1,\ldots,z_r}(W^{\mathfrak{e}}(x)) \mathrm{d}\lambda(x_{\mathfrak{h}([r])}) \\ &= \sum_{z_1,\ldots,z_r\in[q]} \int_{[0,1]^{\mathfrak{h}([r])}} \prod_{j\in[r]} \phi_{z_j}(x_{\{j\}}) \left[\sum_{\mathfrak{e}\in\mathfrak{E}} J^{\mathfrak{e}}_{z_1,\ldots,z_r}(W^{\mathfrak{e}}(x))\right] \mathrm{d}\lambda(x_{\mathfrak{h}([r])}).\end{aligned}$$

Motivated by this reformulation we introduce for every (W,J) pair a special auxiliary instance of the ground state problem that is defined for a $[q]^r$-tuple of $([-d,d],r)$-graphons, where $d = |\mathfrak{E}|\,\|J\|_\infty\,\|W\|_\infty$. For any $z \in [q]^r$, let $\hat{W}^z(x) = \sum_{\mathfrak{e}\in\mathfrak{E}} J^{\mathfrak{e}}_{z_1,\ldots,z_r}(W^{\mathfrak{e}}(x))$ for each $x \in [0,1]^{\mathfrak{h}([r])}$, and let the interaction matrices $\hat{J}^z$ be of the canonical form. We obtain for any fractional partition ϕ of $[0,1]$ into q parts that $\mathcal{E}_\phi(W,J) = \mathcal{E}_\phi(\hat{W},\hat{J})$, and also $\mathcal{E}_{\mathrm{x}}(\mathbb{G}(k,W),J) = \mathcal{E}_{\mathrm{x}}(\mathbb{G}(k,\hat{W}),\hat{J})$ for any fractional partition x, where the two random r-graphs are obtained via the same sample. Therefore, without loss of generality, we are able to reduce the statement of the corollary to the statement of Theorem 4.1.4 dealing with ground state energies of canonical form. □

We start with the proof of Theorem 4.1.4 by providing the necessary background. The first lemma tells us that in the real-valued case the energy of the sample and that of the averaged sample do not differ by a large amount.

Lemma 4.2.2. *Let W be a $([-d,d],r)$-graphon, $q \geq 1$, $J \in \mathbb{R}^{q\times\cdots\times q}$. Then for every $k \geq 1$ there* *is a coupling of* $\mathbb{G}(k, W)$ *and* $\mathbb{H}(k, W)$ *such that*

$$\mathbb{P}\left(|\hat{\mathcal{E}}(\mathbb{G}(k, W), J) - \hat{\mathcal{E}}(\mathbb{H}(k, W), J)| > \varepsilon\|J\|_\infty\|W\|_\infty\right) \le 2\exp\left(-k\left(\frac{\varepsilon^2 k}{2} - \log q\right)\right)$$

Proof. Let us fix a integer q-partition x of $[k]$, and furthermore let the two random r-graphs be generated by the same sample $(U_S)_{S\in\mathfrak{h}([k],r)}$. Then

$$\hat{\mathcal{E}}_{\mathrm{x}}(\mathbb{G}(k, W), J) = \frac{1}{k^r}\sum_{z_1,\dots,z_r=1}^{q}\sum_{n_1,\dots,n_r=1}^{k} J_{z_1,\dots,z_r} W((U_S)_{S\in\mathfrak{h}(\{n_1,\dots,n_r\},r)})\prod_{j=1}^{r} x_{n_j,z_j},$$

and

$$\begin{aligned}&\hat{\mathcal{E}}_{\mathrm{x}}(\mathbb{H}(k, W), J)\\ &= \frac{1}{k^r}\sum_{z_1,\dots,z_r=1}^{q}\sum_{n_1,\dots,n_r=1}^{k} J_{z_1,\dots,z_r}\mathbb{E}[W((U_S)_{S\in\mathfrak{h}(\{n_1,\dots,n_r\},r)}) \mid (U_S)_{S\in\mathfrak{h}(\{n_1,\dots,n_r\},1)}]\prod_{j=1}^{r} x_{n_j,z_j}.\end{aligned}$$

Let us enumerate the elements of $\binom{k}{2}$ as $e_1, e_2, \dots, e_{\binom{k}{2}}$, and define the martingale

$$Y_0 = \mathbb{E}[\hat{\mathcal{E}}_{\mathrm{x}}(\mathbb{G}(k, W), J) \mid \{\, U_j \mid j \in [k]\,\}],$$

and

$$Y_t = \mathbb{E}\left[\hat{\mathcal{E}}_{\mathrm{x}}(\mathbb{H}(k, W), J) \mid \{\, U_j \mid j \in [k]\,\} \cup \left(\cup_{j=1}^{t}\{\, U_S \mid e_j \subset S\,\}\right)\right]$$

for each $1 \le t \le \binom{k}{2}$, so that $Y_0 = \hat{\mathcal{E}}_{\mathrm{x}}(\mathbb{H}(k, W), J)$ and $Y_{\binom{k}{2}} = \hat{\mathcal{E}}_{\mathrm{x}}(\mathbb{G}(k, W), J)$. For each $t \in \binom{k}{2}$ we can upper bound the difference, $|Y_{t-1} - Y_t| \le \frac{1}{k^2}\|J\|_\infty\|W\|_\infty$. By the Azuma-Hoeffding inequality, Lemma 3.5.2, it follows that

$$\mathbb{P}(|Y_t - Y_0| \ge \rho) \le 2\exp\left(-\frac{\rho^2 k^4}{2\binom{k}{2}\|J\|_\infty^2\|W\|_\infty^2}\right) \le 2\exp\left(-\frac{\rho^2 k^2}{2\|J\|_\infty^2\|W\|_\infty^2}\right), \tag{4.4}$$

for any $\rho > 0$.

There are q^k distinct integer q-partitions of $[k]$, hence

$$\mathbb{P}\left(|\hat{\mathcal{E}}(\mathbb{G}(k, W), J) - \hat{\mathcal{E}}(\mathbb{H}(k, W), J)| > \varepsilon\|J\|_\infty\|W\|_\infty\right) \le 2\exp\left(-k\left(\frac{\varepsilon^2 k}{2} - \log q\right)\right). \tag{4.5}$$

□

In the following lemmas every r-graph or graphon is meant to be as bounded real-valued and directed.

We would like to point out in the beginning that in the finite case we are able to shift from the integer optimization problem to the relaxed one with having a reasonably good upper bound on the difference of the optimal values of the two.

Lemma 4.2.3. *Let G be a real-valued r-graph on $[k]$ and $J \in \mathbb{R}^{q\times\cdots\times q}$. Then*

$$|\mathcal{E}(G,J) - \hat{\mathcal{E}}(G,J)| \le r^2 \frac{1}{2k} \|G\|_\infty \|J\|_\infty.$$

Proof. Trivially we have $\mathcal{E}(G,J) \ge \hat{\mathcal{E}}(G,J)$. We define G' by setting all entries of G to 0 which have at least two coordinates which are the same (for $r = 2$ these are the diagonal entries). Thus, we get that

$$|\mathcal{E}(G,J) - \mathcal{E}(G',J)| \le \binom{r}{2} \frac{1}{k} \|G\|_\infty \|J\|_\infty.$$

Now assume that we are given a fractional partition $\overline{\mathrm{x}}$ so that $\mathcal{E}_{\overline{\mathrm{x}}}(G',J)$ attains the maximum $\mathcal{E}(G',J)$. We fix all the entries $\overline{x}_{n,1}, \dots \overline{x}_{n,q}$ of $\overline{\mathrm{x}}$ with $n = 2,\dots,k$ and regard $\mathcal{E}_{\overline{\mathrm{x}}}(G',J)$ as a function of $x_{1,1},\dots,x_{1,q}$. This function will be linear in the variables $x_{1,1},\dots,x_{1,q}$, and with the additional condition $\sum_{j=1}^r x_{1,j} = 1$ we obtain a linear program. By standard arguments this program possesses an integer valued optimal solution, so we are allowed to replace $\overline{x}_{1,1},\dots,\overline{x}_{1,q}$ by integers without letting $\mathcal{E}_{\overline{\mathrm{x}}}(G',J)$ decrease. We repeat this procedure for each $n \in [k]$, obtaining an integer optimum for $\mathcal{E}_{\overline{\mathrm{x}}}(G',J)$, which implies that $\mathcal{E}(G',J) = \hat{\mathcal{E}}(G',J)$. Hence, the claim follows. □

Next lemma is the continuous generalization of Theorem 4 from [14], and is closely related to the Weak Regularity Lemma, Lemma 3.3.4, of [59], and its continuous version Lemma 3.3.10. The result is a centerpiece of the cut decomposition method.

Lemma 4.2.4. *Let $\varepsilon > 0$ arbitrary. For any bounded measurable function $W\colon [0,1]^r \to \mathbb{R}$ there exist an $s \le \frac{1}{\varepsilon^2}$, measurable sets $S_i^j \subset [0,1]$ with $i = 1,\dots,s$, $j = 1,\dots,r$, and real numbers $d_1,\dots,d_s$ so that with $B = \sum_{i=1}^s d_i \mathbb{1}_{S_i^1\times\cdots\times S_i^r}$ it holds that*

(i) $\|W\|_2 \ge \|W - B\|_2$,

(ii) $\|W - B\|_\square < \varepsilon \|W\|_2$, *and*

(iii) $\sum_{i=1}^s |d_i| \le \frac{1}{\varepsilon}\|W\|_2$.

Proof. We construct stepwise the required rectangles and the respective coefficients implicitly. Let $W^0 = W$, and suppose that after the tth step of the construction we have already obtained every set $S_i^j \subset [0,1]$ with $i = 1,\dots,t$, $j = 1,\dots,r$, and the real numbers $d_1,\dots,d_t$. Set $W^t = W - \sum_{i=1}^t d_i \mathbb{1}_{S_i^1\times\cdots\times S_i^r}$. We proceed to the $(t+1)$st step, where two possible situations can occur. The first case is when

$$\|W^t\|_\square \ge \varepsilon \|W\|_2.$$

This implies by definition that there exist measurable subsets $S^1_{t+1}, \dots, S^r_{t+1}$ of $[0,1]$ such that $|\int_{S^1_{t+1}\times\cdots\times S^r_{t+1}} W^t(x)\mathrm{d}\lambda(x)| \geq \varepsilon\|W\|_2$. We define d_{t+1} to be the average of W^t on the product set $S^1_{t+1} \times \cdots \times S^r_{t+1}$, and proceed to the $(t+2)$nd step. In the case of

$$\|W^t\|_\square < \varepsilon\|W\|_2$$

we are ready with the construction and set $s = t$.

We analyze the first case to obtain an upper bound on the total number of steps required by the construction. So suppose that the first case above occurs. Then

$$\begin{aligned}\|W^t\|_2^2 - \|W^{t+1}\|_2^2 &= \int_{S^1_{t+1}\times\cdots\times S^r_{t+1}} (W^t)^2(x)\mathrm{d}\lambda(x) - \int_{S^1_{t+1}\times\cdots\times S^r_{t+1}} (W^t(x) - d_{t+1})^2\mathrm{d}\lambda(x) \\ &= d^2_{t+1}\lambda(S^1_{t+1})\dots\lambda(S^r_{t+1}) \geq \varepsilon^2\|W\|_2^2. \end{aligned} \tag{4.6}$$

This means that the square of the 2-norm of W^t decreases in t in every step when the first case occurs in the construction by at least $\varepsilon^2\|W\|_2^2$, therefore it can happen only at most $\frac{1}{\varepsilon^2}$ times, with other words $s \leq \frac{1}{\varepsilon^2}$. It is also clear that the 2-norm decreases in each step, so we are left to verify the upper bound on the sum of the absolute values of the coefficients d_i. From (4.6) we get, that

$$\|W\|_2^2 = \sum_{t=1}^{s} \|W^{t-1}\|_2^2 - \|W^t\|_2^2 \geq \sum_{t=1}^{s} d_t^2\lambda(S^1_t)\dots\lambda(S^r_t).$$

We also know for every $t \leq s$ that $|d_t|\lambda(S^1_t)\dots\lambda(S^r_t) \geq \varepsilon\|W\|_2$. Hence,

$$\sum_{t=1}^{s} |d_t|\varepsilon\|W\|_2 \leq \sum_{t=1}^{s} d_t^2\lambda(S^1_t)\dots\lambda(S^r_t) \leq \|W\|_2^2,$$

and therefore $\sum_{t=1}^{s} |d_t| \leq \frac{1}{\varepsilon}\|W\|_2$. □

Next we state that the cut approximation provided by Lemma 4.2.4 is invariant under sampling. This is a crucial point of the whole argument, and is the r-dimensional generalization of Lemma 4.6 from [30].

Lemma 4.2.5. *For any $\varepsilon > 0$ and bounded measurable function $W\colon [0,1]^r \to \mathbb{R}$ we have that*

$$\mathbb{P}\left(|\|\mathbb{H}(k,W)\|_\square - \|W\|_\square| > \varepsilon\|W\|_\infty\right) < 2\exp\left(-\frac{\varepsilon^2 k}{32r^2}\right)$$

for every $k \geq \left(\frac{16r^2}{\varepsilon}\right)^4$.

Proof. Fix an arbitrary $0 < \varepsilon < 1$, $r \geq 2$, and further let W be a real-valued naive r-kernel. Set the sample size to $k \geq \left(\frac{16r^2}{\varepsilon}\right)^4$. Let us consider the array representation of

$\mathbb{H}(k, W)$ and denote the r-array $A_{\mathbb{H}(k,W)}$ by G that has zeros on the diagonal. We will need the following lemma from [14].

Lemma 4.2.6. *G is a real r-array on some finite product set $V_1 \times \cdots \times V_r$, where V_i are copies of V of cardinality k. Let $S_1 \subset V_1, \ldots, S_r \subset V_r$ be fixed subsets and Q_1 a uniform random subset of $V_2 \times \cdots \times V_r$ of cardinality p. Then*

$$G(S_1, \ldots, S_r) \leq E_{Q_1} G(P(Q_1 \cap S_2 \times \cdots \times S_r), S_2, \ldots, S_r) + \frac{k^r}{\sqrt{p}} \|G\|_2,$$

where $P(Q_1) = P_G(Q_1) = \{ x_1 \in V_1 \mid \sum_{(y_2,\ldots y_r) \in Q_1} G(x_1, y_2, \ldots, y_r) > 0 \}$ and the 2-norm denotes $\|G\|_2 = \left(\frac{\sum_{x_i \in V_i} G^2(x_1,\ldots,x_r)}{|V_1|\ldots|V_r|} \right)^{1/2}$.

If we apply Lemma 4.2.6 repeatedly r times to the r-arrays G and $-G$, then we arrive at an upper bound on $G(S_1, \ldots, S_r)$ ($(-G)(S_1, \ldots, S_r)$ respectively) for any collection of the $S_1, \ldots, S_r$ which does not depend on the particular choice of these sets any more, so we get that

$$k^r \|G\|_\square \leq E_{Q_1,\ldots,Q_r} \max_{Q'_i \subset Q_i} \max\{G(P_G(Q'_1), \ldots, P_G(Q'_r)); (-G)(P_{-G}(Q'_1), \ldots, P_{-G}(Q'_r))\} + \frac{rk^r}{\sqrt{p}} \|G\|_\infty, \quad (4.7)$$

since $\|G\|_2 \leq \|G\|_\infty$.

Let us recall that G stands for the random $\mathbb{H}(k, W)$. We are interested in the expectation $\mathbb{E}$ of the left hand side of (4.7) over the sample that defines G. Now we proceed via the method of conditional expectation. We establish an upper bound on the expectation of right hand side of (4.7) over the sample $U_1, \ldots, U_k$ for each choice of the tuple of sets $Q_1, \ldots, Q_r$. This bound does not depend on the actual choice of the Q_i's, so if we take the average (over the Q_i's), that upper bound still remains valid.

In order to do this, let us fix $Q_1, \ldots, Q_r$, set Q to be the set of elements of $V(G)$ which are contained in at least one of the Q_i's, and fix also the sample points of $U_Q = \{ U_i \mid i \in Q \}$. Take the expectation $\mathbb{E}_{U_{Q^c}}$ only over the remaining U_i sample points.

To this end, by Fubini we have the estimate

$$k^r \mathbb{E}_{U_{[k]}} \|G\|_\square \leq E_{Q_1,\ldots Q_r} \mathbb{E}_{U_Q} [\mathbb{E}_{U_{Q^c}} \max_{Q'_i \subset Q_i} \max\{G(P_G(Q'_1) \cap Q^c, \ldots, P_G(Q'_r) \cap Q^c); (-G)(P_{-G}(Q'_1) \cap Q^c, \ldots, P_{-G}(Q'_r) \cap Q^c)\}] + \frac{rk^r}{\sqrt{p}} \|G\|_\infty + pr^3 k^{r-1} \|G\|_\infty, \quad (4.8)$$

where $U_S = \{ U_i \mid i \in S \}$.

Our goal is to uniformly upper bound the expression in the brackets in (4.8) so that in the dependence on the particular $Q_1, \ldots Q_r$ and the sample points from U_Q vanishes. To achieve this, we consider additionally a tuple of subsets $Q'_i \subset Q_i$, and introduce the random variable $Y(Q'_1, \ldots, Q'_r) = G(P_G(Q'_1) \cap Q^c, \ldots, P_G(Q'_r) \cap Q^c)$, where the randomness comes from U_{Q^c} exclusively. Let

$$T_i = \{ x_i \in [0,1] \mid \sum_{(y_1,\dots,y_{i-1},y_{i+1},\dots y_r)\in Q_i'} W(U_{y_1},\dots,U_{y_{i-1}},x_i,U_{y_{i+1}},\dots U_{y_r}) > 0 \}$$

for $i \in [r]$. Note that $t_i \in P_G(Q_i')$ is equivalent to $U_{t_i} \in T_i$. Then

$$\begin{aligned}\mathbb{E}_{U_{Q^c}} Y(Q_1',\dots,Q_r') &\le \sum_{\substack{t_1,\dots,t_r\in Q^c\\ t_i\neq t_j}} \mathbb{E}_{U_{Q^c}} G(t_1,\dots,t_r)\mathbb{1}_{P_G(Q_1')}(t_1)\dots\mathbb{1}_{P_G(Q_r')}(t_r) + r^2k^{r-1}\|W\|_\infty \\ &\le k^r \int_{T_1\times\dots\times T_r} W(x)\mathrm{d}\lambda(x) + r^2k^{r-1}\|W\|_\infty \le k^r\|W\|_\square + r^2k^{r-1}\|W\|_\infty.\end{aligned}$$

By the Azuma-Hoeffding inequality we also have high concentration of the random variable $Y(Q_1',\dots,Q_r')$ around its mean, that is

$$\mathbb{P}(Y(Q_1',\dots,Q_r') \ge \mathbb{E}_{U_{Q^c}} Y(Q_1',\dots,Q_r') + \rho k^r\|W\|_\infty) < \exp\left(-\frac{\rho^2 k}{8r^2}\right), \tag{4.9}$$

since modification of one sampled element changes the value of $Y(Q_1',\dots,Q_r')$ by at most $2rk^{r-1}\|W\|_\infty$. Analogous upper bounds on the expectation and the tail probability hold for each of the expressions $(-G)(P_{-G}(Q_1'),\dots,P_{G}(Q_r'))$.

With regard to the maximum expression in (4.8) over the Q_i' sets we have to this end either that the concentration event from (4.9) holds for each possible choice of the Q_i' subsets for both expressions in the brackets in (4.8), this has probability at least $1-2^{pr+1}\exp(-\frac{\rho^2k}{8r^2})$, or it fails for some choice. In the first case we can employ the upper bound $k^r\|W\|_\square + (r^2k^{r-1}+\rho k^r)\|W\|_\infty$, and in the event of failure we still have the trivial upper bound of $k^r\|W\|_\infty$. Eventually we presented an upper bound on the expectation that does not depend on the choice of $Q_1,\dots,Q_r$, and the sample points from U_Q. Hence by taking expectation and assembling the terms, we have

$$\mathbb{E}_{U_{[k]}}\|G\|_\square \le \|W\|_\square + \|W\|_\infty\left(\frac{r}{\sqrt{p}} + \frac{pr^3}{k} + \rho + \frac{r^2}{k} + 2^{pr+1}\exp\left(-\frac{\rho^2k}{8r^2}\right)\right).$$

Let $p = \sqrt{k}$ and $\rho = \frac{4r^2}{\sqrt[4]{k}}$. Then

$$\begin{aligned}\mathbb{E}_{U_{[k]}}\|G\|_\square &\le \|W\|_\square + \|W\|_\infty\left(\frac{r}{\sqrt[4]{k}} + \frac{r^3}{\sqrt{k}} + \frac{4r^2}{\sqrt[4]{k}} + \frac{r^2}{k} + \exp\left(2\sqrt{k}r - 2r^2\sqrt{k}\right)\right) \\ &\le \|W\|_\square + \|W\|_\infty\left(\frac{\varepsilon}{16r} + \frac{\varepsilon^2}{2^8r} + \frac{\varepsilon}{4} + \frac{\varepsilon^4}{2^{16}r^6} + \frac{\varepsilon^2}{2^8r^6}\right) \le \|W\|_\square + \varepsilon/2\|W\|_\infty.\end{aligned}$$

The direction concerning the lower bound, $\mathbb{E}\|G\|_\square \ge \|W\|_\square - \varepsilon/2$ follows from a standard sampling argument, the idea is that we can project each set $S \subset [0,1]$ to a set $\hat{S} \subset [k]$ through the sample, which will fulfill the desired conditions, we leave the details to the reader. Concentration follows by the Azuma-Hoeffding inequality. We conclude that

$$\begin{aligned}\mathbb{P}\left(|\|G\|_\square - \|W\|_\square| > \varepsilon\|W\|_\infty\right) &\le \mathbb{P}\left(\left|\mathbb{E}\|G\|_\square - \frac{1}{k^r}\|G\|_\square\right| > \varepsilon/2\|W\|_\infty\right) \\ &\le 2\exp\left(-\frac{\varepsilon^2 k}{32r^2}\right).\end{aligned}$$

□

The two previous results Lemma 4.2.4 and Lemma 4.2.5 together imply a generalization of Lemma 3.3.9 for r-graphs and r-graphons for arbitrary r formulating that typical induced subgraphs are close in the $\delta_\square$ to the source object. We formulate the precise statement below, however we are not going to use it in the proof of Theorem 4.1.4. Its proof is analogous to the case of Lemma 3.3.9 in [30]: for a given instance, the application of Lemma 4.2.4 ensures the existence of $\delta_\square$-close graphon of bounded complexity (a step function) in the quality of the approximation. Lemma 4.2.5 tells us that the projected approximation on the sample is also of sufficient quality with respect to the sampled problem with high probability, the proof concludes with the description of an overlay of the steps of the two approximating graphons to infer that they can be rearranged to be identical on a large subset of the unit cube. We refer for details to [30].

Lemma 4.2.7. *Let $\varepsilon > 0$ and let U be an r-graphon with $0 \le U \le 1$. Then for $q \ge 2^{100r/\varepsilon^2}$ we have*

$$\mathbb{P}\left(\delta_\square(U, \mathbb{G}(q, U)) \ge \varepsilon\right) \le \exp\left(-4^{100/\varepsilon^2}\frac{\varepsilon^2}{50}\right). \tag{4.10}$$

Next we state a result on the relationship of a continuous linear program (LP) and its randomly sampled finite subprogram. We will rely on the next concentration result that is a generalization of the Azuma-Hoeffding inequality, Lemma 3.5.2, and suits well the situation when the martingale jump sizes have inhomogeneous distribution. It can be found together with a proof in the survey [112] as Corollary 3.

Lemma 4.2.8 (Generalized Azuma-Hoeffding inequality). *Let $k \ge 1$ and $(X_n)_{n=0}^k$ be a martingale sequence with respect to the natural filtration $(\mathcal{F}_n)_{n=1}^k$. If $|X_n - X_{n+1}| \le d$ almost surely and $\mathbb{E}[(X_n - X_{n+1})^2 \mid \mathcal{F}_n] \le \sigma^2$ for each $n \in [k]$, then for every $n \le k$ and $\delta > 0$ it holds that*

$$\mathbb{P}(X_n - X_0 > \delta n) \le \exp\left(-n\frac{\sigma^2}{d^2}\left((1 + \frac{\delta d}{\sigma^2})\ln(1 + \frac{\delta d}{\sigma^2}) - \frac{\delta d}{\sigma^2}\right)\right). \tag{4.11}$$

Measurability for all of the following functions is assumed.

Lemma 4.2.9. *Let $c_m\colon [0,1] \to \mathbb{R}$, $U_{i,m}\colon [0,1] \to \mathbb{R}$ for $i = 1,\dots,s$, $m = 1,\dots,q$, $u \in \mathbb{R}^{s\times q}$, $\alpha \in \mathbb{R}$. Let d and σ be positive reals such that $\|c\|_\infty \le d$ and $\|c\|_2 \le \sigma$ and set $\gamma = \frac{\sigma^2}{d^2}$. If the optimum of the linear program*

$$\begin{aligned}
&\text{maximize} && \int_0^1 \sum_{m=1}^{q} f_m(t)c_m(t)\mathrm{d}t && \\
&\text{subject to} && \int_0^1 f_m(t)U_{i,m}(t)\mathrm{d}t \le u_{i,m} && \text{for } i \in [s] \text{ and } m \in [q] \\
& && 0 \le f_m(t) \le 1 && \text{for } t \in [0,1] \text{ and } m \in [q] \\
& && \sum_{m=1}^{q} f_m(t) = 1 && \text{for } t \in [0,1]
\end{aligned}$$

is less than α, then for any $\varepsilon, \delta > 0$ and $k \in \mathbb{N}$ and a uniform random sample $\{X_1,\dots,X_k\}$ of $[0,1]^k$ the optimum of the sampled linear program

$$\begin{aligned}
&\text{maximize} && \sum_{1\le n\le k} \sum_{m=1}^{q} \frac{1}{k} x_{n,m} c_m(X_n) && \\
&\text{subject to} && \sum_{1\le n\le k} \frac{1}{k} x_{n,m} U_{i,m}(X_n) \le u_{i,m} - \delta\|U\|_\infty && \text{for } i \in [s] \text{ and } m \in [q] \\
& && 0 \le x_{n,m} \le 1 && \text{for } n \in [k] \text{ and } m \in [q] \\
& && \sum_{m=1}^{q} x_{n,m} = 1 && \text{for } n \in [k]
\end{aligned}$$

is less than $\alpha + \varepsilon$ with probability at least

$$1 - \left[\exp\left(-\frac{\delta^2 k}{2}\right) + \exp\left(-k\gamma\left((1+\frac{\varepsilon}{\gamma d})\ln(1+\frac{\varepsilon}{\gamma d}) - \frac{\varepsilon}{\gamma d}\right)\right)\right].$$

Proof. We require a continuous version of Farkas' Lemma.

Claim 3. Let $(Af)_{i,m} = \int_0^1 A_{i,m}(t) f_m(t)\mathrm{d}t$ for the bounded measurable functions $A_{i,m}$ on $[0,1]$ for $i \in [s]$ and $m \in [q]$, and let $v \in \mathbb{R}^{sq}$. There is no fractional q-partition solution $f = (f_1,\dots,f_q)$ to $Af \le v$ if and only if, there exists a non-zero $0 \le y \in \mathbb{R}^{sq}$ with $\|y\|_1 = 1$ such that there is no fractional q-partition solution f to $y^T(Af) \le y^T v$.

For clarity we remark that in the current claim and the following one Af and v are indexed by a pair of parameters, but are regarded as 1-dimensional vectors in the multiplication operation.

Proof. One direction is trivial: if there is a solution f to $Af \le v$, then it is also a solution to $y^T(Af) \le y^T v$ for any $y \ge 0$.

We turn to show the opposite direction. Let

$$C = \{Af \mid f \text{ is a fractional } q\text{-partition of } [0,1]\}.$$

The set C is a nonempty convex closed subset of $\mathbb{R}^{sq}$ containing 0. Let $B = \{x \mid x_{i,m} \leq v_{i,m}\} \subset \mathbb{R}^{sq}$, this set is also a nonempty convex closed set. The absence of a solution to $Af \leq v$ is equivalent to saying that $C \cap B$ is empty. It follows from the Separation Theorem for convex closed sets that there is a $0 \neq y' \in \mathbb{R}^{sq}$ such that $y'^T c < y'^T b$ for every $c \in C$ and $b \in B$. Additionally every coordinate $y'_{i,m}$ has to be non-positive. To see this suppose that $y'_{i_0,m_0} > 0$, we pick a $c \in C$ and $b \in B$, and send b_{i_0,m_0} to minus infinity leaving every other coordinate of the two points fixed (b will still be an element of B), for b_{i_0} small enough the inequality $y'^T c < y'^T b$ will be harmed eventually. We conclude that for any f we have $y'^T(Af) < y'^T v$, hence for $y = \frac{-y'}{\|y'\|_1}$ the inequality $y^T(Af) \leq y^T v$ has no solution. □

From this lemma the finitary version follows without any difficulties.

Claim 4. Let B be a real $sq \times k$ matrix, and let $v \in \mathbb{R}^{sq}$. There is no fractional q-partition $x \in \mathbb{R}^{kq}$ so that $Bx \leq v$ if and only if, there is a non-zero $0 \leq y \in \mathbb{R}^{sq}$ with $\|y\|_1 = 1$ such that there is no fractional q-partition $x \in \mathbb{R}^{kq}$ so that $y^T Bx \leq y^T v$.

Proof. Let $A_{i,m}(t) = \sum_{n=1}^{k} \frac{B_{(i,m),n}}{k} \mathbb{1}_{[\frac{n-1}{k},\frac{n}{k})}(t)$ for $i = 1, \ldots, s$. The nonexistence of a fractional q-partition $x \in \mathbb{R}^{kq}$ so that $Bx \leq v$ is equivalent to nonexistence of a fractional q-partition f so that $Af \leq v$. For any nonzero $0 \leq y$, the nonexistence of a fractional q-partition $x \in \mathbb{R}^{kq}$ so that $y^T Bx \leq y^T v$ is equivalent to the nonexistence of a fractional q-partition f so that $y^T(Af) \leq y^T v$. Applying Claim 3 verifies the current claim. □

The assumption of the lemma is by Claim 3 equivalent to the statement that there exists a nonzero $0 \leq y \in \mathbb{R}^{sq}$ and $0 \leq \beta$ with $\sum_{i=1}^{s} \sum_{m=1}^{q} y_{i,m} + \beta = 1$ such that

$$\int_0^1 \sum_{i=1}^{s} \sum_{m=1}^{q} y_{i,m} U_{i,m}(t) f_m(t) \mathrm{d}t - \int_0^1 \beta \sum_{m=1}^{q} c_m(t) f_m(t) \leq \sum_{i=1}^{s} \sum_{m=1}^{q} y_{i,m} u_{i,m} - \beta\alpha$$

has no solution f among fractional q-partitions. This is equivalent to the condition

$$\int_0^1 h(t)\mathrm{d}t > A,$$

where $h(t) = \min_m [\sum_{i=1}^{s} y_{i,m} U_{i,m}(t) - \beta c_m(t)]$, and $A = \sum_{i=1}^{s} \sum_{m=1}^{q} y_{i,m} u_{i,m} - \beta\alpha$. Let $T_m = \{t \mid h(t) = \sum_{i=1}^{s} y_{i,m} U_{i,m}(t) - \beta c_m(t)\}$ for $m \in [q]$ and define the functions $h_1(t) = \sum_{m=1}^{q} \mathbb{1}_{T_m}(t) [\sum_{i=1}^{s} y_{i,m} U_{i,m}(t)]$ and $h_2(t) = \sum_{m=1}^{q} \mathbb{1}_{T_m}(t) \beta c_m(t)$. Clearly, $h(t) = h_1(t) - h_2(t)$. Set also $A_1 = \sum_{i=1}^{s} \sum_{m=1}^{q} y_{i,m} u_{i,m}$ and $A_2 = \beta\alpha$. Fix an arbitrary $\delta > 0$ and $k \geq 1$. By the Azuma-Hoeffding inequality it follows that with probability at least $1 - \exp(-\frac{k\delta^2}{2})$ we have that

$$\frac{1}{k}\sum_{n=1}^{k} h_1(X_n) > A_1 - \delta\|h_1\|_\infty.$$

Note that $\|h_1\|_\infty = \|\sum_{i=1}^{s}\sum_{m=1}^{q} \mathbb{1}_{T_m} U_{i,m} y_{i,m}\|_\infty \leq \|U\|_\infty \sum_{i=1}^{s}\sum_{m=1}^{q} |y_{i,m}| \leq \|U\|_\infty$. Moreover, by Lemma 4.2.8 the event

$$\frac{1}{k}\sum_{n=1}^{k} h_2(X_n) < A_2 + \varepsilon \tag{4.12}$$

has probability at least $1 - \exp\left(-k\gamma\left((1+\frac{\varepsilon}{\gamma d})\ln(1+\frac{\varepsilon}{\gamma d}) - \frac{\varepsilon}{\gamma d}\right)\right)$. Thus,

$$\frac{1}{k}\sum_{n=1}^{k} h(X_n) > \sum_{i=1}^{s}\sum_{m=1}^{q} y_{i,m}(u_{i,m} - \delta\|U\|_\infty) - \beta(\alpha + \varepsilon)$$

with probability at least

$$1 - \left[\exp\left(-\frac{\delta^2 k}{2}\right) + \exp\left(-k\gamma\left((1+\frac{\varepsilon}{\gamma d})\ln(1+\frac{\varepsilon}{\gamma d}) - \frac{\varepsilon}{\gamma d}\right)\right)\right].$$

We conclude the proof by noting that the last event is equivalent to the event in the statement of our lemma by Claim 4. □

We start the principal part of the proof of the main theorem in this chapter.

Proof of Theorem 4.1.4. It is enough to prove Theorem 4.1.4 for tuples of naive $([-d,d], r)$-digraphons. We first employ Lemma 4.2.2 to replace the energy $\hat{\mathcal{E}}(\mathbb{G}(k,W))$ by the energy of the averaged sample $\hat{\mathcal{E}}(\mathbb{H}(k,W))$ without altering the ground state energy of the sample substantially with high probability. Subsequently, we apply Lemma 4.2.3 to change from the integer version of the energy $\hat{\mathcal{E}}(\mathbb{H}(k,W))$ to the relaxed one $\mathcal{E}(\mathbb{H}(k,W))$. That is

$$|\hat{\mathcal{E}}(\mathbb{G}(k,W)) - \mathcal{E}(\mathbb{H}(k,W))| \leq \varepsilon^2\|W\|_\infty$$

with probability at least $1-\varepsilon^2$.

We begin with the main argument by showing that the ground state energy of the sample can not be substantially smaller than that of the original, formally

$$\mathcal{E}(\mathbb{H}(k,W)) \geq \mathcal{E}(W) - \frac{r^2}{k^{1/4}}\|W\|_\infty \tag{4.13}$$

with high probability. In what follows $\mathbb{E}$ denotes the expectation with respect to the uniform independent random sample $(U_S)_{S\in\mathfrak{h}([k],r)}$ from $[0,1]$. To see the correctness of the inequality, we consider a fixed fractional partition ϕ of $[0,1]$, and define the random fractional partition of $[k]$ as $y_{n,m} = \phi_m(U_n)$ for every $n \in [k]$ and $m \in [q]$. Then we have that

$$\begin{aligned}
\mathbb{E}\mathcal{E}(\mathbb{H}(k,W)) &\geq \mathbb{E}\mathcal{E}_y(\mathbb{H}(k,W)) \\
&= \mathbb{E}\frac{1}{k^r}\sum_{z\in[q]^r}\sum_{n_1,\ldots,n_r=1}^{k} W^z(U_{\mathfrak{h}(\{n_1,\ldots,n_r\},r)})\prod_{j=1}^{r} y_{n_j,z_j} \\
&\geq \frac{k!}{k^r(k-r)!}\sum_{z\in[q]^r}\int_{[0,1]^{\mathfrak{h}([r])}} W^z(t_{\mathfrak{h}([r])})\prod_{j=1}^{r}\phi_{z_j}(t_j)\mathrm{d}\lambda(t) - \frac{r^2}{k}\|W\|_\infty \\
&\geq \mathcal{E}_\phi(W) - \frac{r^2}{k}\|W\|_\infty.
\end{aligned}$$

This argument proves the claim in expectation, concentration will be provided by standard martingale arguments. For convenience, we define a martingale by $Y_0 = \mathbb{E}\mathcal{E}(\mathbb{H}(k,W))$ and $Y_j = E\left[\mathcal{E}(\mathbb{H}(k,W)) \mid \{\, U_S \mid S \in \mathfrak{h}([j], r-1)\,\}\right]$ for $1 \leq j \leq k$. The difference $|Y_j - Y_{j+1}| \leq \frac{2r}{k}\|W\|_\infty$ is bounded from above for any j, thus by the inequality of Azuma and Hoeffding, Lemma 3.5.2, it follows that

$$\begin{aligned}
&\mathbb{P}\left(\mathcal{E}(\mathbb{H}(k,W)) < \mathcal{E}(W) - \frac{2r^2}{k^{1/4}}\|W\|_\infty\right) \\
&\quad\leq \mathbb{P}\left(\mathcal{E}(\mathbb{H}(k,W)) < \mathbb{E}\mathcal{E}(\mathbb{H}(k,W)) - \frac{r^2}{k^{1/4}}\|W\|_\infty\right) \\
&\quad= \mathbb{P}\left(Y_k < Y_0 - \frac{r^2}{k^{1/4}}\|W\|_\infty\right) \leq \exp\left(-\frac{r^2\sqrt{k}}{8}\right). \qquad (4.14)
\end{aligned}$$

So the lower bound (4.13) on $\mathcal{E}(\mathbb{H}(k,W))$ is established. Note that by the condition regarding k we can establish (rather crudely) the upper bound $\exp(-\frac{r^2\sqrt{k}}{8}) \leq \varepsilon 2^{-7}$.

Now we turn to prove that $\mathcal{E}(\mathbb{H}(k,W)) < \mathcal{E}(W) + \varepsilon$ holds also with high probability for $k \geq \left(\frac{2^{r+7}q^r r}{\varepsilon}\right)^4 \log(\frac{2^{r+7}q^r r}{\varepsilon})q^r$. Our two main tools will be Lemma 4.2.4, that is a variant the Cut Decomposition Lemma from [14] (closely related to the Weak Regularity Lemma by Frieze and Kannan [59]), and linear programming duality, in the form of Lemma 4.2.9. Recall the definition of the cut norm, for $W\colon [0,1]^r \to \mathbb{R}$, it is given as

$$\|W\|_\square = \max_{S^1,\ldots,S^r\subset[0,1]}\left|\int_{S^1\times\cdots\times S^r} W(x)\mathrm{d}\lambda(x)\right|,$$

and for an r-array G by the expression

$$\|G\|_\square = \frac{1}{k^r}\max_{S^1,\ldots,S^r\subset V(G)}\left|G(S^1,\ldots,S^r)\right|.$$

Before starting the second part of the technical proof, we present an informal outline. Our task is to certify that there is no assignment of the variables on the sampled energy problem, which produces an overly large value relative to the ground state energy of the continuous problem. For this reason we build up a cover of subsets over the set of fractional partitions of the variables of the finite problem, also build a cover of subsets over the fractional partitions of the original continuous energy problem, and establish an association scheme between the elements of the two in such a way, that with high probability we can state that the optimum on one particular set of the cover of the sampled energy problem does not exceed the optimal value of the original problem on the associated set of the other cover. To be able to do this, first we have to define these two covers, this is done with the aid of the cut decomposition, see Lemma 4.2.4. We will replace the original continuous problem by an auxiliary one, where the number of variables will be bounded uniformly in terms of our error margin ε. Lemma 4.2.5 makes it possible for us to replace the sampled energy problem by an auxiliary problem with the same complexity as for the continuous problem. This second replacement will have a straightforward relationship to the approximation of the original problem. We will produce the cover sets of the two problems by localizing the auxiliary problems, association happens through the aforementioned straightforward connection. Finally, we will linearize the local problems, and use the linear programming duality principle from Lemma 4.2.9 to verify that the local optimal value on the sample does not exceed the local optimal value on the original problem by an infeasible amount, with high probability.

Recall that for a $\phi = (\phi_1, \dots, \phi_q)$ a fractional q-partition of $[0,1]$ the energy is given by the formula

$$\mathcal{E}_{\phi}(W) = \sum_{z \in [q]^r} \int_{[0,1]^r} \prod_{j \in [r]} \phi_{z_j}(t_j) W^z(t) \mathrm{d}\lambda(t), \tag{4.15}$$

and for an $\mathrm{x} = (x_{1,1}, x_{1,2}, \dots, x_{1,q}, x_{2,1}, \dots, x_{k,q})$ a fractional q-partition of $[k]$ by

$$\mathcal{E}_{\mathrm{x}}(\mathbb{H}(k, W)) = \sum_{z \in [q]^r} \frac{1}{k^r} \sum_{n_1, \dots, n_r = 1}^{k} \prod_{j \in [r]} x_{t_j, z_j} W^z(U_{n_1}, \dots, U_{n_r}). \tag{4.16}$$

We are going to establish a term-wise connection with respect to the parameter z in the previous formulas. Therefore we consider the function

$$\mathcal{E}^z_{\phi}(W^z) = \int_{[0,1]^r} \prod_{j \in [r]} \phi_{z_j}(t_j) W^z(t) \mathrm{d}\lambda(t), \tag{4.17}$$

it follows that $\mathcal{E}_\phi(W) = \sum_{z\in[q]^r} \mathcal{E}^z_\phi(W^z)$. Analogously we consider

$$\mathcal{E}^z_{\mathrm{x}}(\mathbb{H}(k, W^z)) = \frac{1}{k^r} \sum_{n_1,\dots,n_r=1}^{k} \prod_{j\in[r]} x_{t_j,z_j} W^z(U_{n_1},\dots,U_{n_r}),$$

so $\mathcal{E}_{\mathrm{x}}(\mathbb{H}(k, W)) = \sum_{z\in[q]^r} \mathcal{E}^z_{\mathrm{x}}(\mathbb{H}(k, W^z))$ with the sampled graphs on the right generated by the same sample points. Note that the formulas (4.15)-(4.17) make prefect sense even when the parameters ϕ and x are only vectors of bounded functions and reals respectively without forming partition.

Lemma 4.2.4 delivers for any $z \in [q]^r$ an integer $s_z \le \frac{2^6 q^{2r}}{\varepsilon^2}$, measurable sets $S_{z,i,j} \subset [0, 1]$ with $i = 1,\dots, s_z$, $j = 1,\dots, r$, and the real numbers $d_{z,1},\dots, d_{z,s_z}$ such that the conditions of the lemma are satisfied, namely

$$\|W^z - \sum_{i=1}^{s_z} d_{z,i} \mathbb{1}_{S_{z,i,1}\times\cdots\times S_{z,i,r}}\|_\square \le \frac{\varepsilon}{8q^r}\|W^z\|_2,$$

and $\sum_{i=1}^{s_z} |d_{z,i}| \le \frac{8q^r}{\varepsilon}\|W^z\|_2$. The cut function allows a sufficiently good approximation for $\mathcal{E}_\phi(W^z)$, for any ϕ. Let $D^z = \sum_{i=1}^{s_z} d_{z,i} \mathbb{1}_{S_{z,i,1}\times\cdots\times S_{z,i,r}}$. Then

$$\begin{aligned}|\mathcal{E}^z_\phi(W^z) - \mathcal{E}^z_\phi(D^z)| &= \left|\int_{[0,1]^r} \prod_{j\in[r]} \phi_{z_j}(t_j)\,[W^z(t) - D^z(t)]\,\mathrm{d}\lambda(t)\right| \\ &\le \|W^z - D^z\|_\square \le \frac{\varepsilon}{8q^r}\|W^z\|_\infty.\end{aligned}$$

We apply the cut approximation to W^z for every $z \in [q]^r$ to obtain the $[q]^r$-tuple of naive r-kernels $D = (D^z)_{z\in[q]^r}$. We define the push-forward of this approximation for the sample $\mathbb{H}(k, W)$. To do this we need to define the subsets $[k] \supset \hat{S}_{z,i,j} = \{\, m \mid U_m \in S_{z,i,j} \,\}$. Let $\hat{D}^z = \sum_{i=1}^{s_z} d_{z,i} \mathbb{1}_{\hat{S}_{z,i,1}\times\cdots\times\hat{S}_{z,i,r}}$. First we condition on the event from Lemma 4.2.5, call this event E_1, that is

$$E_1 = \bigcap_{z\in[q]^r} \left\{ \Big| \|\mathbb{H}(k, W^z) - \hat{D}^z\|_\square - \|W^z - D^z\|_\square \Big| < \frac{\varepsilon}{8q^r}\|W\|_\infty \right\}.$$

On E_1 it follows that for any x that is a fractional q-partition

$$\begin{aligned}|\mathcal{E}^z_{\mathrm{x}}(\mathbb{H}(k, W^z)) - \mathcal{E}^z_{\mathrm{x}}(\hat{D}^z)| &\le \|\mathbb{H}(k, W^z) - \hat{D}^z\|_\square \\ &\le \|W^z - D^z\|_\square + \frac{\varepsilon}{8q^r}\|W\|_\infty.\end{aligned}$$

This implies that

$$|\mathcal{E}_\phi(W) - \mathcal{E}_\phi(D)| \le \frac{\varepsilon}{8}\|W\|_\infty \quad \text{and} \quad |\mathcal{E}_x(\mathbb{H}(k,W)) - \mathcal{E}_x(\mathbb{H}(k,D))| \le \frac{\varepsilon}{4}\|W\|_\infty.$$

The probability that E_1 fails is at most $2q^r \exp\left(-\frac{\varepsilon^2 k}{2^{11} r^2 q^{2r}}\right)$ whenever $k \ge \left(\frac{2^7 q^r r^2}{\varepsilon}\right)^4$ due to Lemma 4.2.5, in the current theorem we have the condition $k \ge \left(\frac{2^{r+7} q^r r}{\varepsilon}\right)^4 \log(\frac{2^{r+7} q^r r}{\varepsilon}) q^r$, which implies the aforementioned one. The failure probability of E_1 is then strictly less than $\frac{\varepsilon}{2^7}$.

Let $\mathcal{S} = \{ S_{z,i,j} \mid z \in [q]^r, 1 \le i \le s_z, 1 \le j \le r \}$ denote their set, and let $\mathcal{S}'$ stand for the corresponding set on the sample. Note that $s' = |\mathcal{S}| \le 2^6 r q^{3r} \frac{1}{\varepsilon^2}$ in general, but in some cases the W^z functions are constant multiples of each other, so the cut approximation can be chosen in a way that $S_{z,i,j}$ does not depend on $z \in [q]^r$, and in this case we have the slightly refined upper bound $2^6 r q^{2r} \frac{1}{\varepsilon^2}$ for s', consequences of this in the special case are discussed in the remark after the proof. Let $\eta > 0$ be arbitrary, and define the sets

$$I(b,\eta) = \left\{ \phi \mid \forall z \in [q]^r, 1 \le i \le s_z, 1 \le j \le r \colon \left| \int_{S_{z,i,j}} \phi_{z_j}(t)\mathrm{d}t - b_{z,i,j} \right| \le 2\eta \right\},$$

and

$$I'(b,\eta) = \left\{ \mathrm{x} \mid \forall z \in [q]^r, 1 \le i \le s_z, 1 \le j \le r \colon \left| \frac{1}{k} \sum_{U_n \in S_{z,i,j}} x_{n,z_j} - b_{z,i,j} \right| \le \eta \right\}$$

For a collection of non-negative reals $\{b_{z,i,j}\}$. At this point in the definitions of the above sets we do not require ϕ and x to be fractional q-partitions, but to be vectors of bounded functions and vectors respectively. We will use the grid points $\mathcal{A} = \{ (b_{z,i,j})_{z,i,j} \mid \forall z,i,j \colon b_{z,i,j} \in [0,1] \cap \eta\mathbb{Z} \}$.

On every nonempty set $I(b,\eta)$ we can produce a linear approximation of $\mathcal{E}_\phi(D)$ (linearity is meant in the functions ϕ_m) which carries through to a linear approximation of $\mathcal{E}_\mathrm{x}(\mathbb{H}(k,D))$ via sampling. The precise description of this is given in the next auxiliary result.

Lemma 4.2.10 (Local linearization). *If $\eta \le \frac{\varepsilon}{16 q^r 2^r}$, then for every $b \in \mathcal{A}$ there exist $l_0 \in \mathbb{R}$ and functions $l_1, l_2, \dots, l_q \colon [0,1] \to \mathbb{R}$ such that for every $\phi \in I(b,\eta)$ it holds that*

$$\left| \mathcal{E}_\phi(D) - l_0 - \int_0^1 \sum_{m=1}^q l_m(t)\phi_m(t)\mathrm{d}t \right| < \frac{\varepsilon}{2^{r+3}}\|W\|_\infty,$$

and for every $\mathrm{x} \in I'(b,\eta)$ *we have*

$$\left|\mathcal{E}_{\mathrm{x}}(\mathbb{H}(k,D)) - l_0 - \sum_{n=1}^{k}\sum_{m=1}^{q}\frac{1}{k}x_{n,m}l_m(U_i)\right| < \frac{\varepsilon}{2^{r+5}}\|W\|_\infty.$$

Additionally we have that $l_1, l_2, \ldots, l_q$ *are bounded from above by* $\frac{8q^{2r}}{\varepsilon}\|W\|_\infty$ *and* $\int_0^1 \sum_{m=1}^{q} l_m^2(t)\mathrm{d}t \leq 2^{2r+9}r^2q^{3r}\|W\|_\infty^2$.

Proof. Recall the decomposition of the energies as sums over $z \in [q]^r$ into terms

$$\begin{aligned}\mathcal{E}_\phi^z(D^z) &= \sum_{i=1}^{s_z} d_{z,i}\int_{[0,1]^r}\prod_{j=1}^{r}\phi_{z_j}(t_j)\mathbb{1}_{S_{z,i,1}\times\cdots\times S_{z,i,r}}(t)\mathrm{d}t\\ &= \sum_{i=1}^{s_z} d_{z,i}\int_{[0,1]^r}\prod_{m=1}^{q}\prod_{\substack{j=1\\ z_j=m}}^{r}\phi_m(t_j)\mathbb{1}_{S_{z,i,1}\times\cdots\times S_{z,i,r}}(t)\mathrm{d}t,\end{aligned}$$

and

$$\mathcal{E}_{\mathrm{x}}^z(\hat{D}^z) = \sum_{i=1}^{s_z} d_{z,i}\frac{1}{k^r}\prod_{m=1}^{q}\prod_{\substack{j=1\\ z_j=m}}^{r}\sum_{n:\,U_n\in S_{z,i,j}} x_{n,m}.$$

We linearize and compare the functions $\mathcal{E}_\phi^z(D^z)$ and $\mathcal{E}_{\mathrm{x}}^z(\hat{D}^z)$ term-wise. In the end we will sum up the errors and deviations occurred at each term. Let $b \in \mathcal{A}$ and $\eta > 0$ as in the statement of the lemma with $I(b,\eta)$ being nonempty. Let us fix an arbitrary $\phi \in I(b,\eta)$, $z \in [q]^r$, and $1 \leq i \leq s_z$. Then

$$\begin{aligned}\prod_{j=1}^{r}\left[\int_0^1 \phi_{z_j}(t_j)\mathbb{1}_{S_{z,i,j}}(t_j)\mathrm{d}t_j\right] &= B^i(z) + \sum_{j=1}^{r}\left[\int_0^1 \phi_{z_j}(t_j)\mathbb{1}_{S_{z,i,j}}(t_j)\mathrm{d}t_j - b_{z,i,j}\right]B^{i,j}(z) + \Delta\\ &= (1-r)B^i(z) + \sum_{m=1}^{q}\int_0^1 \phi_m(t)\left[\sum_{j=1,z_j=m}^{r}\mathbb{1}_{S_{z,i,j}}(t)B^{i,j}(z)\right]\mathrm{d}t + \Delta,\end{aligned}$$

where $B^i(z)$ stands for $\prod_{j=1}^r b_{z,i,j}$, $B^{i,j}(z) = \prod_{l\neq j} b_{z,i,l}$, and $|\Delta| \leq 4\eta^2 2^r$. Analogously for an arbitrary fixed element $\mathrm{x} \in I'(b,\eta)$ and a term of $\mathcal{E}_{\mathrm{x}}^z(\hat{D}^z)$ we have

$$\begin{aligned}&\prod_{j=1}^{r}\left[\frac{1}{k}\sum_{n:\,U_n\in S_{z,i,j}} x_{n,z_j} - b_{z,i,j} + b_{z,i,j}\right]\\ &\quad= (1-r)B^i(z) + \sum_{m=1}^{q}\sum_{n=1}^{k}\frac{1}{k}x_{n,m}\left[\sum_{j=1,z_j=m}^{r}\mathbb{1}_{S_{z,i,j}}(U_n)B^{i,j}(z)\right] + \Delta',\end{aligned}$$

where $|\Delta'| \leq \eta^2 2^r$.

If we multiply these former expressions by the respective coefficient $d_{z,i}$ and sum up over i and z, then we obtain the final linear approximation consisting of the constant l_0 and the functions $l_1, \dots, l_q$. We would like to add that these objects do not depend on η if $I(b, \eta)$ is nonempty, only the accuracy of the approximation does. As overall error in approximating the energies we get in the first case of $\mathcal{E}_\phi(D)$ at most $32\eta^2 2^r \frac{q^{2r}}{\varepsilon}\|W\|_\infty \leq \frac{\varepsilon}{2^{r+3}}\|W\|_\infty$, and in the second case of $\mathcal{E}_x(\mathbb{H}(k,D))$ at most $\frac{\varepsilon}{2^{r+5}}\|W\|_\infty$.

Now we turn to prove the upper bound on $|l_m(t)|$. Looking at the above formulas we could write out $l_m(t)$ explicitly, for our upper bound it is enough to note that $\left[\sum_{j=1, z_j=m}^r \mathbb{1}_{S_{z,i,j}}(t) B^{i,j}(z)\right]$ is at most r. So it follows that for any $t \in [0,1]$ it holds that

$$|l_m(t)| \leq \frac{8q^{2r}}{\varepsilon} r\|W\|_\infty.$$

It remains to verify the assertion regarding $\int_0^1 \sum_{m=1}^q l_m^2(t)\mathrm{d}t$. Note that $I(b,\eta) \subset I(b, 2\eta)$, so we can apply the same linear approximation to elements ψ of $I(b, 2\eta)$ as above with a deviation of at most $\frac{\varepsilon}{2^{r+1}}\|W\|_\infty$ from $\mathcal{E}_\psi(D)$. Let ϕ be an arbitrary element of $I(b,\eta)$, and let $T \subset [0,1]$ denote the set of measure η corresponding to the largest $\sum_{m=1}^q |l_m(t)|$ values. Define

$$\hat{\phi}_m(t) = \begin{cases} \phi_m(t) + \operatorname{sgn}(l_m(t)) & \text{if } t \in T \\ \phi_m(t) & \text{otherwise.} \end{cases}$$

Then $\hat{\phi} \in I(b, 2\eta)$, since $\|\phi_m - \hat{\phi}_m\|_1 \leq \eta$ for each $m \in [q]$, but $\hat{\phi}$ is not necessarily a fractional partition. Therefore we have

$$\begin{aligned} \int_T \sum_{m=1}^q |l_m(t)|\mathrm{d}t &= \int_0^1 \sum_{m=1}^q (\hat{\phi}_m(t) - \phi_m(t)) l_m(t)\mathrm{d}t \\ &\leq \left| \int_0^1 \sum_{m=1}^q \hat{\phi}_m(t) l_m(t)\mathrm{d}t - \mathcal{E}_{\hat{\phi}}(D) \right| + |\mathcal{E}_{\hat{\phi}}(D) - \mathcal{E}_\phi(D)| \\ &\quad + \left| \int_0^1 \sum_{m=1}^q \phi_m(t) l_m(t)\mathrm{d}t - \mathcal{E}_\phi(D) \right| \\ &\leq \frac{5}{2^{r+3}} \varepsilon\|W\|_\infty + |\mathcal{E}_{\hat{\phi}}(D) - \mathcal{E}_\phi(D)|. \end{aligned}$$

We have to estimate the last term of the above expression.

$$\begin{aligned} |\mathcal{E}_{\hat{\phi}}(D) - \mathcal{E}_\phi(D)| &\leq \sum_{z \in [q]^r} \left| \int_{[0,1]^r} \left(\prod_{j=1}^r \phi_{z_j}(t_j) - \prod_{j=1}^r \hat{\phi}_{z_j}(t_j) \right) D^z(t)\mathrm{d}t \right| \\ &\leq 2\|W\|_\infty \sum_{z \in [q]^r} \int_{[0,1]^r} \sum_{j=1}^r \left| \prod_{i<j} \phi_{z_i}(t_i) \prod_{i>j} \hat{\phi}_{z_i}(t_i)(\hat{\phi}_{z_j}(t_j) - \phi_{z_j}(t_j)) \right| \mathrm{d}t \end{aligned}$$

$$\leq 2\|W\|_\infty 2^r q^{r-1} r \sum_{m=1}^{q} \|\phi_m - \hat{\phi}_m\|_1 \leq 2\|W\|_\infty 2^r q^r r\eta.$$

We conclude that

$$\int_T \sum_{m=1}^{q} |l_m(t)| \mathrm{d}t \leq \left(\frac{5}{2^{r+3}} + \frac{r}{2^3}\right) \varepsilon \|W\|_\infty.$$

This further implies that for each $t \notin T$ we have $\sum_{m=1}^{q} |l_m(t)| \leq \left(\frac{5}{2^{r+3}} + \frac{r}{2^3}\right) \frac{\varepsilon}{\eta} \|W\|_\infty \leq \left(10 + 2^{r+1} r\right) q^r \|W\|_\infty$. These former bounds indicate

$$\begin{aligned}
\int_0^1 \sum_{m=1}^{q} l_m^2(t)\mathrm{d}t &= \int_{[0,1]\setminus T} \sum_{m=1}^{q} l_m^2(t)\mathrm{d}t + \int_T \sum_{m=1}^{q} l_m^2(t)\mathrm{d}t \\
&\leq 2^{2r+8} r^2 q^{2r} \|W\|_\infty^2 + \|l\|_\infty \int_T \sum_{m=1}^{q} |l_m(t)|\mathrm{d}t \\
&\leq 2^{2r+8} r^2 q^{2r} \|W\|_\infty^2 + (2^{r+4} r q^r)(8q^{2r} r)\|W\|_\infty^2 \\
&\leq 2^{2r+9} r^2 q^{3r} \|W\|_\infty^2.
\end{aligned}$$

□

We return to the proof of the main theorem, and set $\eta = \frac{\varepsilon}{16q^r 2^r}$. For each $b \in \mathcal{A}$ we apply Lemma 4.2.10, so that we have for any $\phi \in I(b,\eta)$ and $\mathrm{x} \in I'(b,\eta)$ that

$$\left| \mathcal{E}_\phi(W) - l_0 - \sum_{m=1}^{q} \int_0^1 \phi_m(t) l_m(t)\mathrm{d}t \right| = \frac{\varepsilon}{2^{r+3}} \|W\|_\infty,$$

$$\left| \mathcal{E}_{\mathrm{x}}(\mathbb{H}(k,W)) - l_0 - \sum_{n=1}^{k} \frac{1}{k} x_{n,m} l_m(U_n) \right| = \frac{\varepsilon}{2^{r+5}} \|W\|_\infty,$$

since η is small enough. Note that $l_0, l_1, \ldots,$ and l_q inherently depend on b. We introduce the event $E_2(b)$, which stands for the occurrence of the following implication:

If the linear program

maximize	$l_0 + \sum_{n=1}^{k} \sum_{m=1}^{q} \frac{1}{k} x_{n,m} l_m(U_n)$	
subject to	$\mathrm{x} \in I'(b,\eta)$	
	$0 \leq x_{n,m} \leq 1$	for $n = 1, \ldots, k$ and $m = 1, \ldots, q$
	$\sum_{m=1}^{q} x_{n,m} = 1$	for $m = 1, \ldots, q$

has optimal value α, then the continuous linear program

$$\begin{array}{lll} \text{maximize} & l_0 + \int_0^1 \sum_{m=1}^{q} l_m(t)\phi_m(t)\mathrm{d}t & \\ \text{subject to} & \phi \in I(b,\eta) & \\ & 0 \le \phi_m(t) \le 1 & \text{for } t \in [0,1] \text{ and } m = 1,\dots,q \\ & \sum_{m=1}^{q} \phi_m(t) = 1 & \text{for } t \in [0,1] \end{array}$$

has optimal value at least $\alpha - (\varepsilon/2)\|W\|_\infty$.

We apply Lemma 4.2.9 with $\delta = \eta$, $\sigma^2 = 2^{2r+9}r^2q^{3r}\|W\|_\infty^2$, $d = \frac{8q^{2r}}{\varepsilon}r\|W\|_\infty$, and $\gamma = \frac{\sigma^2}{d^2}$, and attain that the probability that $E_2(b)$ fails is at most

$$\begin{aligned} &\exp\left(-\frac{k\eta^2}{2}\right) + \exp\left(-k\gamma\left((1+\frac{\varepsilon\|W\|_\infty}{\gamma d})\ln(1+\frac{\varepsilon\|W\|_\infty}{\gamma d}) - \frac{\varepsilon\|W\|_\infty}{\gamma d}\right)\right) \\ &\le \exp\left(-\frac{k\varepsilon^2}{2^8q^{2r}2^{2r}}\right) + \exp\left(-k\varepsilon^2 2^{2r+3}q^{-r}\left(\frac{1}{2^{4r+15}q^r r^2}\right)\right) \\ &= \exp\left(-\frac{k\varepsilon^2}{2^{2r+8}q^{2r}}\right) + \exp\left(-\frac{k\varepsilon^2}{2^{2r+12}q^{2r}r^2}\right) \le 2\exp\left(-\frac{k\varepsilon^2}{2^{2r+12}q^{2r}r^2}\right), \end{aligned}$$

where we used that $(1+x)\ln(1+x) - x \ge (1+x)(x - x^2/2) - x = x^2/2 - x^3/2 \ge x^2/4$ for $0 \le x \le \frac{1}{4}$. Denote by E_2 the event that for each $b \in \mathcal{A}$ the event $E_2(b)$ occurs. Then we have

$$\begin{aligned} \mathbb{P}(E_2) &\ge 1 - 2\left(\frac{2^{r+3}q^r}{\varepsilon}\right)^{2^6rq^{3r}\frac{1}{\varepsilon^2}} \exp\left(-\frac{k\varepsilon^2}{2^{2r+12}q^{2r}r^2}\right) \\ &\ge 1 - 2\exp\left(\log\left(\frac{2^{r+3}q^r}{\varepsilon}\right)2^6rq^{3r}\varepsilon^{-2} - \log(\frac{2^{r+7}q^r r}{\varepsilon})2^{2r+16}r^2q^{3r}\varepsilon^{-2}\right) \\ &\ge 1 - 2\exp\left(-\log(\frac{2^{r+7}q^r r}{\varepsilon})2^{2r+15}r^2q^{3r}\varepsilon^{-2}\right) \\ &\ge 1 - \varepsilon/4. \end{aligned}$$

Therefore for $k \ge \left(\frac{2^{r+7}q^r r}{\varepsilon}\right)^4 \log(\frac{2^{r+7}q^r r}{\varepsilon})q^r$ we have that $\mathbb{P}(E_1 \cap E_2) \ge 1 - \varepsilon/2$. We only need to check that conditioned on E_1 and E_2 our requirements are fulfilled. For this, consider an arbitrary fractional q-partition of $[k]$ denoted by x. For some $b \in \mathcal{A}$ we have that $\mathrm{x} \in I'(b,\eta)$. If we sum up the error gaps that were allowed for the Cut Decomposition and at the local linearization stage, then the argument we presented above yields that there exists a $\phi \in I(b, \eta)$ such that conditioned on the event $E_1 \cap E_2$ it holds

$$\mathcal{E}_\phi(W) \geq \mathcal{E}_x(\mathbb{H}(k, W)) - \varepsilon \|W\|_\infty.$$

This is what we wanted to show.

□

We can improve on the tail probability bound in Theorem 4.1.4 significantly by a constant factor strengthening of the lower threshold condition imposed on the sample size.

Corollary 4.2.11. *Let $r \geq 1$, $q \geq 1$, and $\varepsilon > 0$. Then for any $[q]^r$-tuple of $([-d, d], r)$-graphons $W = (W^z)_{z \in [q]^r}$ and $k \geq \Theta^4 \log(\Theta) q^r$ with $\Theta = \frac{2^{r+10} q^r r}{\varepsilon}$ we have that*

$$\mathbb{P}(|\mathcal{E}(W) - \hat{\mathcal{E}}(\mathbb{G}(k, W))| > \varepsilon \|W\|_\infty) < 2 \exp\left(-\frac{\varepsilon^2 k}{8 r^2}\right). \tag{4.18}$$

Proof. For $k \geq \Theta^4 \log(\Theta) q^r$ we appeal to Theorem 4.1.4, hence

$$\begin{aligned}|\mathcal{E}(W) - \mathbb{E}\hat{\mathcal{E}}(\mathbb{G}(k, W))| &\leq \mathbb{P}(|\mathcal{E}(W) - \hat{\mathcal{E}}(\mathbb{G}(k, W))| > \varepsilon/8 \|W\|_\infty) 2 \|W\|_\infty + \varepsilon/8 \|W\|_\infty \\ &< \varepsilon/2 \|W\|_\infty.\end{aligned}$$

Using a similar martingale construction to the one in the first part of the proof of Theorem 4.1.4 the Azuma-Hoeffding inequality can be applied, thus

$$\begin{aligned}\mathbb{P}(|\mathcal{E}(W) - \hat{\mathcal{E}}(\mathbb{G}(k, W))| > \varepsilon \|W\|_\infty) &\leq \mathbb{P}(|\mathbb{E}\hat{\mathcal{E}}(\mathbb{G}(k, W)) - \hat{\mathcal{E}}(\mathbb{G}(k, W))| > \varepsilon/2 \|W\|_\infty) \\ &\leq 2 \exp\left(-\frac{\varepsilon^2 k}{8 r^2}\right).\end{aligned}$$

□

Remark 4.2.12. A simple investigation of the above proof also exposes that in the case when the W^z's are constant multiples of each other then we can employ the same cut decomposition to all of them with the right scaling, which implies that the upper bound on $|\mathcal{S}|$ can be strengthened to $2^6 r q^{2r} \frac{1}{\varepsilon^2}$, gaining a factor of q^r. Therefore in this case the statement of Corollary 4.2.11 is valid with the improved lower bound condition $\left(\frac{2^{r+10} q^r r}{\varepsilon}\right)^4 \log(\frac{2^{r+10} q^r r}{\varepsilon})$ on k.

Remark 4.2.13. Suppose that f is the following simple graph parameter. Let $q \geq 1$, $m_0 \geq 1$, and g be a polynomial of l variables and degree d with values between 0 and 1 on the unit cube, where l is the number of unlabeled node-q-colored graphs on m_0 vertices, whose set we denote by $\mathcal{M}_{q,m_0}$. Note that $l \leq 2^{m_0^2/2} q^{m_0}$. Let then

$$f(G) = \max_{\mathcal{T}} g((t(F, (G, \mathcal{T})))_{F \in \mathcal{M}_{q,m_0}}), \tag{4.19}$$

where the maximum goes over all node-q-colorings of G, and $(G, \mathcal{T})$ denotes the node-q-colored graph by imposing $\mathcal{T}$ on the node set of G. Using the identity $t(F_1, G)t(F_2, G) = t(F_1 \cup F_2, G)$, where $F_1 \cup F_2$ is the disjoint union of the (perhaps colored) graphs F_1 and F_2, we can replace in (4.19) g by g' that is linear, and its variables are indexed by $\mathcal{M}_{q,dm_0}$. Then it becomes clear that f can be regarded as a ground state energy of dm_0-dimensional arrays by associating to every G an tuple $(A^z)_{z\in[q]^r}$ with $r = dm_0$, where the entries $A^z(i_1, \dots, i_r)$ are the coefficients of g' corresponding to the element of $\mathcal{M}_{q,dm_0}$ given by the pair z and $G|_{(i_1,\dots,i_r)}$. We conclude that f is efficiently testable by Theorem 4.1.4.

4.3 Testability of variants of the ground state energy

In the current section we derive further testability results using the techniques employed in the proofs of the previous section, and apply Theorem 4.1.4 to some specific quadratic programming problems.

4.3.1 External magnetic field

Using the physics language formulation, one might consider the generalization of the ground state energy optimization problem of Definition 4.1.2 with some external magnetic field that introduces a bias towards certain states. The field is represented by a real vector $h \in \mathbb{R}^q$, the components of h correspond to the states in $[q]$, the energies and the GSE problem in this slightly extended setup are given next.

Definition 4.3.1. *Let $\mathfrak{E}$ be a finite layer set, $\mathcal{K}$ be a compact set, and $W = (W^{\mathfrak{e}})_{\mathfrak{e}\in\mathfrak{E}}$ be a tuple of $(\mathcal{K}, r)$-graphons. Let q be a fixed positive integer and $J = (J^{\mathfrak{e}})_{\mathfrak{e}\in\mathfrak{E}}$ with $J^{\mathfrak{e}} \in C(\mathcal{K})^{q\times\dots\times q}$ for every $\mathfrak{e} \in \mathfrak{E}$, and $h = (h_1, \dots, h_q) \in \mathbb{R}^q$. For a $\phi = (\phi_1, \dots, \phi_q)$ fractional q-partition of* $[0, 1]$ *let*

$$\mathcal{E}(W, J, h) = \max_{\phi} \mathcal{E}_{\phi}(W, J) + \sum_{i=1}^{q} h_i \int_0^1 \phi_i(t)\mathrm{d}t$$

denote the layered ground state energy of W with respect to J and h, where the maximum runs over all fractional q-partitions of $[0, 1]$.

We define for $G = (G^{\mathfrak{e}})_{\mathfrak{e}\in\mathfrak{E}}$ the discrete versions as

$$\hat{\mathcal{E}}(G, J, h) = \max_{\mathrm{x}} \mathcal{E}_{\mathrm{x}}(G, J) + \frac{1}{|V(G)|} \sum_{i=1}^{q} \sum_{u\in V(G)} h_i x_{u,i},$$

where the maximum runs over integer q-partitions ($x_{n,m} \in \{0,1\}$ *), and*

$$\mathcal{E}(G,J,h) = \max_{\mathrm{x}} \mathcal{E}_{\mathrm{x}}(G,J) + \frac{1}{|V(G)|}\sum_{i=1}^{q}\sum_{u\in V(G)} h_i x_{u,i},$$

where the maximum is taken over all fractional q-partitions x.

The canonical from introduced above can also be employed here, thus J can always be assumed to a have a special form, let the corresponding ground state energies be denoted by $\mathcal{E}(W,h)$ and $\mathcal{E}(G,h)$. The generalization of Theorem 4.1.4 with the magnetic field bias is the following result. The proof will only be sketched and relies strongly on the proof of Theorem 4.1.4.

Theorem 4.3.2. *Let* $r \geq 1$, $q \geq 1$, $h \in \mathbb{R}^q$, *and* $\varepsilon > 0$. *Then for any* $[q]^r$*-tuple of* $([-d,d],r)$*-graphons* $W = (W^z)_{z\in[q]^r}$ *and* $k \geq \Theta^4 \log(\Theta) q^r$ *with* $\Theta = \frac{2^{r+7}q^r r}{\varepsilon}$ *we have that*

$$\mathbb{P}(|\mathcal{E}(W,h) - \hat{\mathcal{E}}(\mathbb{G}(k,W),h)| > \varepsilon(\|W\|_\infty + \|h\|_\infty)) < \varepsilon. \tag{4.20}$$

Proof. The proof is virtually identical to the case without an external field. The only difference is in the coefficients in the local linearization stage, Lemma 4.2.10. At that point, if we denote the output functions of the original lemma by l_m^0, we have to define $l_m(t) = l_m^0(t) + h_m$ for the adapted lemma (which we don not state here explicitly), where $\mathcal{E}_\phi(D)$ is replaced by $\mathcal{E}_\phi(D) + \sum_{i=1}^q h_i \int_0^1 \phi_i(t)\mathrm{d}t$, and so is the discrete energy $\mathcal{E}_{\mathrm{x}}(\hat{D})$ in an analogous way by $\mathcal{E}_{\mathrm{x}}(\hat{D}) + \frac{1}{|V(G)|}\sum_{i=1}^q \sum_{u\in V(G)} h_i x_{u,i}$. Therefore for the current problem the upper bound for $l_1, l_2, \ldots, l_q$ is modified to $\frac{8q^{2r}}{\varepsilon}\|W\|_\infty + \|h\|_\infty$, and further $\int_0^1 \sum_{m=1}^q l_m^2(t)\mathrm{d}t \leq 2^{2r+9}r^2q^{3r}\|W\|_\infty^2 + \|h\|_\infty^2$. The final steps of the proof are again identical to the proof of Theorem 4.1.4, the new bound in Theorem 4.3.2 is the consequence of the reasoning above.

□

4.3.2 Microcanonical version

Next we will state and prove the microcanonical version of Theorem 4.1.4, that is the continuous generalization of the main result of [53] for an arbitrary number q of the states. To be able to do this, we require the microcanonical analog of Lemma 4.2.3, that will be a generalization of Theorem 5.5 from [32] for arbitrary r-graphs (except for the fact that we are not dealing with node weights), and its proof will also follow the lines of the aforementioned theorem. Before stating the lemma, we outline some notation and state yet another auxiliary lemma.

Definition 4.3.3. *Let for* $\mathbf{a} = (a_1, \ldots, a_q) \in \mathrm{Pd}_q$ *(that is,* $a_i \geq 0$ *for each* $i \in [q]$ *and* $\sum_i a_i = 1$*) denote*

$$\Omega_{\mathbf{a}} = \left\{ \phi \text{ fractional q-partition of } [0,1] \mid \int_0^1 \phi_i(t)\mathrm{d}t = a_i \text{ for } i \in [q] \right\},$$

$$\omega_{\mathbf{a}} = \left\{ \mathrm{x} \textit{ fractional } q\textit{-partition of } V(G) \mid \frac{1}{|V(G)|} \sum_{u \in V(G)} x_{u,i} = a_i \textit{ for } i \in [q] \right\},$$

and

$$\hat{\omega}_{\mathbf{a}} = \left\{ \mathrm{x} \textit{ integer } q\textit{-partition of } V(G) \mid \left| \frac{\sum_{u \in V(G)} x_{u,i}}{|V(G)|} - a_i \right| \leq \frac{1}{|V(G)|} \textit{ for } i \in [q] \right\}.$$

The elements of the above sets are referred to as integer $\mathbf{a}$*-partitions and fractional* $\mathbf{a}$*-partitions, respectively.*

We call the following expressions microcanonical ground state energies with respect to $\mathbf{a}$ *for* $(\mathcal{K}, r)$*-graphs and graphons and* $C(\mathcal{K})$*-valued r-arrays J, in the finite case we add the term fractional and integer respectively to the name. Denote*

$$\mathcal{E}_{\mathbf{a}}(W, J) = \max_{\phi \in \Omega_{\mathbf{a}}} \mathcal{E}_{\phi}(W, J), \quad \mathcal{E}_{\mathbf{a}}(G, J) = \max_{\mathrm{x} \in \omega_{\mathbf{a}}} \mathcal{E}_{\mathrm{x}}(G, J), \quad \hat{\mathcal{E}}_{\mathbf{a}}(G, J) = \max_{\mathrm{x} \in \hat{\omega}_{\mathbf{a}}} \mathcal{E}_{\mathrm{x}}(G, J).$$

The layered versions for a finite layer set $\mathfrak{E}$*, and the canonical versions* $\mathcal{E}_{\mathbf{a}}(W)$*,* $\mathcal{E}_{\mathbf{a}}(G)$*, and* $\hat{\mathcal{E}}_{\mathbf{a}}(G)$ *are defined analogously.*

The requirements for an x to be an integer fractional $\mathbf{a}$-partition (that is $\phi \in \Omega_{\mathbf{a}}$) are rather strict and we are not able to guarantee with high probability that if we sample from an fractional $\mathbf{a}$-partition of $[0,1]$, that we will receive an fractional $\mathbf{a}$-partition on the sample, in fact this will not happen with probability 1. To tackle this problem we need to establish an upper bound on the difference of two microcanonical ground state energies with the same parameters. This was done in the two dimensional case in [32], we slightly generalize that approach.

Lemma 4.3.4. *Let* $r \geq 1$*, and* $q \geq 1$*. Then for any* $[q]^r$*-tuple of naive r-kernels* $W = (W^z)_{z \in [q]^r}$*, and probability distributions* $\mathbf{a}, \mathbf{b} \in \mathrm{Pd}_q$ *we have*

$$|\mathcal{E}_{\mathbf{a}}(W) - \mathcal{E}_{\mathbf{b}}(W)| \leq r \|W\|_\infty \|\mathbf{a} - \mathbf{b}\|_1.$$

The analogous statement is true for a $[q]^r$*-tuple of* $([-d,d], r)$*-digraphs* $G = (G^z)_{z \in [q]^r}$*,*

$$|\mathcal{E}_{\mathbf{a}}(G) - \mathcal{E}_{\mathbf{b}}(G)| \leq r \|G\|_\infty \|\mathbf{a} - \mathbf{b}\|_1.$$

Proof. We will find for each fractional $\mathbf{a}$-partition ϕ a fractional $\mathbf{b}$-partition ϕ' and vice versa, so that the corresponding energies are as close to each other as in the statement. So let $\phi = (\phi_1, \dots, \phi_q)$ be an arbitrary fractional $\mathbf{a}$-partition, we define ϕ'_i so that the following holds: if $a_i \geq b_i$ then $\phi'_i(t) \leq \phi_i(t)$ for every $t \in [0,1]$, otherwise $\phi'_i(t) \geq \phi_i(t)$ for every $t \in [0,1]$. It is easy to see that such a $\phi' = (\phi'_1, \dots, \phi'_q)$ exists. Next we estimate the energy deviation.

$$
\begin{aligned}
|\mathcal{E}_{\phi}(W) - \mathcal{E}_{\phi'}(W)| &\leq \sum_{z \in [q]^r} \left| \int_{[0,1]^r} \phi_{z_1}(x_1) \ldots \phi_{z_r}(x_r) - \phi'_{z_1}(x_1) \ldots \phi'_{z_r}(x_r) \mathrm{d}\lambda(x) \right| \|W\|_\infty \\
&\leq \sum_{z \in [q]^r} \sum_{m=1}^{r} \left| \int_{[0,1]^r} (\phi_{z_m}(x_m) - \phi'_{z_m}(x_m)) \prod_{j<m} \phi_{z_j}(x_j) \prod_{j>m} \phi'_{z_j}(x_j) \mathrm{d}\lambda(x) \right| \|W\|_\infty \\
&= \sum_{z \in [q]^r} \sum_{m=1}^{r} \int_{[0,1]} \left| \phi_{z_m}(x_m) - \phi'_{z_m}(x_m) \right| \mathrm{d}x_m \prod_{j<m} a_{z_j} \prod_{j>m} b_{z_j} \|W\|_\infty \\
&= \sum_{m=1}^{r} \sum_{j=1}^{q} \int_{[0,1]} \left| \phi_j(t) - \phi'_j(t) \right| \mathrm{d}t \left(\sum_{j=1}^{q} a_j \right)^{m-1} \left(\sum_{j=1}^{q} b_j \right)^{r-m-1} \|W\|_\infty \\
&= r \|\mathbf{a} - \mathbf{b}\|_1 \|W\|_\infty.
\end{aligned}
$$

The same way we can find for any fractional **b**-partition ϕ an fractional **a**-partition ϕ' so that their respective energies differ at most by $r\|\mathbf{a} - \mathbf{b}\|_1 \|W\|_\infty$. This implies the first statement of the lemma. The finite case is proven in a completely analogous fashion. □

We are ready to show that the difference of the fractional and the integer ground state energies is $o(|V(G)|)$ whenever all parameters are fixed, this result is a generalization with respect to the dimension in the non-weighted case of Theorem 5.5 of [32], the proof proceeds similar to the one concerning the graph case that was dealt with in [32].

Lemma 4.3.5. *Let $q, r, k \geq 1, \mathbf{a} \in \mathrm{Pd}_q$, and $G = (G^z)_{z \in [q]^r}$ be a tuple of $([-d, d], r)$-graphs on $[k]$. Then*

$$
|\mathcal{E}_{\mathbf{a}}(G) - \hat{\mathcal{E}}_{\mathbf{a}}(G)| \leq \frac{1}{k} \|G\|_\infty 5^r q^{r+1}.
$$

Proof. The inequality $\mathcal{E}_{\mathbf{a}}(G) \leq \hat{\mathcal{E}}_{\mathbf{a}}(G) + \frac{1}{k} \|G\|_\infty 5^r q^{r+1}$ follows from Lemma 4.3.4. Indeed, for this bound a somewhat stronger statement it possible,

$$
\hat{\mathcal{E}}_{\mathbf{a}}(G) \leq \max_{\mathbf{b}\colon |b_i - a_i| \leq 1/k} \mathcal{E}_{\mathbf{b}}(G) \leq \mathcal{E}_{\mathbf{a}}(G) + r \frac{q}{k} \|G\|_\infty.
$$

Now we will show that $\hat{\mathcal{E}}_{\mathbf{a}}(G) \geq \mathcal{E}_{\mathbf{a}}(G) - \frac{1}{k} \|G\|_\infty 5^r q^{r+1}$. We consider an arbitrary fractional **a**-partition x. A node i from $[n]$ is called bad in a fractional partition x, if at least two elements of $\{x_{i,1}, \ldots, x_{i,q}\}$ are positive. We will reduce the number of fractional entries of the bad nodes of x step by step until we have at most q of them, and keep track of the cost of each conversion, at the end we round the corresponding fractional entries of the remaining bad nodes in some certain way.

We will describe a step of the reduction of fractional entries. For now assume that we have at least $q + 1$ bad nodes and select an arbitrary set S of cardinality $q + 1$ of them. To each element of S corresponds a q-tuple of entries and each of these q-tuples has at least two non-$\{0,1\}$ elements.

We reduce the number of fractional entries corresponding to S while not disrupting any entries corresponding to nodes that lie outside of S. To do this we fix for each $i \in [q]$ the sums $\sum_{v \in S} x_{v,i}$ and for each $v \in S$ the sums $\sum_{i=1}^{q} x_{v,i}$ (these latter are naturally fixed to be 1), in total $2q+1$ linear equalities. We have at least $2q+2$ fractional entries corresponding to S, therefore there exists a subspace of solutions of dimension at least 1 for the $2q+1$ linear equalities. That is, there is a family of fractional partitions parametrized by $-t_1 \leq t \leq t_2$ for some $t_1, t_2 > 0$ that obey our $2q+1$ fixed equalities and have the following form. Let $x^t_{i,j} = x_{i,j} + t\beta_{i,j}$, where $\beta_{i,j} = 0$ if $i \notin S$ or $x_{i,j} \in \{0,1\}$, and $\beta_{i,j} \neq 0$ else, together these entries define x^t. The boundaries $-t_1$ and t_2 are non-zero and finite, because eventually an entry corresponding to S would exceed 1 or would be less than 0 with t going to plus, respectively minus infinity. Therefore at these boundary points we still have an fractional $\mathbf{a}$-partition that satisfies our selected equalities, but the number of fractional entries decreases by at least one. We will formalize how the energy behaves when applying this procedure.

$$\mathcal{E}_{\mathrm{x}^t}(G) = \mathcal{E}_{\mathrm{x}}(G) + c_1 t + \cdots + c_r t^r,$$

where for $l \in [r]$ we have

$$c_l = \frac{1}{k^r} \sum_{z \in [q]^r} \sum_{\substack{u_1,\ldots,u_l \in S \\ u_{l+1},\ldots,u_r \in V \setminus S \\ \pi}} \beta_{u_1, z_{\pi(1)}} \ldots \beta_{u_l, z_{\pi(l)}} x_{u_{l+1}, z_{\pi(l+1)}} \ldots x_{u_r, z_{\pi(r)}} G^z(u_{\pi(1)}, \ldots, u_{\pi(r)}),$$

where the second sum runs over permutations π of $[k]$ that preserves the ordering of the elements of $\{1,\ldots,l\}$ and $\{l+1,\ldots,r\}$ at the same time. We deform the entries corresponding to S through t in the direction so that $c_1 t \geq 0$ until we have eliminated at least one fractional entry, that is we set $t = -t_1$, if $c < 0$, and $t = t_2$ otherwise. Note, that as x^t is a fractional partition, therefore $0 \leq x_{i,j} + t\beta_{i,j} \leq 1$, which implies that for $t\beta_{i,j} \leq 0$ we have $|t\beta_{i,j}| \leq x_{i,j}$. On the other hand, $\sum_j t\beta_{i,j} = 0$ for any t and i. Therefore $\sum_j |t\beta_{i,j}| = 2\sum_j |t\beta_{i,j}| \mathbb{1}_{\{t\beta_{i,j} \leq 0\}} \leq 2\sum_j x_{i,j} = 2$ for any $i \in [k]$. This simple fact enables us to upper bound the absolute value of the terms $c_l t^l$.

$$\begin{aligned}
|c_l t^l| &\leq \frac{(k-q-1)^{r-l}}{k^r} \|G\|_\infty \sum_{z \in [q]^r} \sum_{\substack{u_1,\ldots,u_l \in S \\ \pi}} |t\beta_{u_1, z_{\pi(1)}}| \ldots |t\beta_{u_l, z_{\pi(l)}}| \\
&= \frac{(k-q-1)^{r-l}}{k^r} \|G\|_\infty \binom{r}{l} q^{r-l} \sum_{z \in [q]^l} \sum_{u_1,\ldots,u_l \in S} |t\beta_{u_1, z_1}| \ldots |t\beta_{u_l, z_l}| \\
&\leq \frac{1}{k^l} \|G\|_\infty \binom{r}{l} q^{r-l} \left(\sum_{u \in S, j \in [q]} |t\beta_{u,j}| \right)^l \leq \frac{1}{k^l} \|G\|_\infty \binom{r}{l} q^{r-l} (2q+2)^l.
\end{aligned}$$

It follows that in each step of elimination of a fractional entry of x we have to admit a decrease of the energy value of at most

$$\sum_{l=2}^{r} |c_l t^l| \leq \frac{1}{k^2}\|G\|_\infty (3q+2)^r.$$

There are in total kq entries in x, therefore, since in each step the number of fractional entries is reduced by at least 1, we can upper bound the number of required steps for reducing the cardinality of bad nodes to at most q by $k(q-1)$, and conclude that we admit an overall energy decrease of at most $\frac{1}{k}\|G\|_\infty(q-1)(3q+2)^r$ to construct from x a fractional partition x′ with at most q nodes with fractional entries In the second stage we proceed as follows. Let $B = \{u_1, \dots, u_m\}$ be the set of the remaining bad nodes of x′, with $m \leq q$. For $u_i \in B$ we set $x''_{u_i,j} = \mathbb{1}_i(j)$, for the rest of the nodes we set x″ = x′, obtaining an integer **a**-partition of $[k]$. Finally, we estimate the cost of this operation. We get that

$$\mathcal{E}_{x''}(G) \geq \mathcal{E}_{x'}(G) - \frac{1}{k^r}\|G\|_\infty |B| k^{r-1} q^r.$$

The original fractional **a**-partition was arbitrary, therefore it follows that

$$\mathcal{E}_{\mathbf{a}}(G) - \hat{\mathcal{E}}_{\mathbf{a}}(G) \leq \frac{1}{k}\|G\|_\infty 5^r q^{r+1}.$$

□

We are ready state the adaptation of Theorem 4.1.4 adapted to the microcanonical setting.

Theorem 4.3.6. *Let $r \geq 1$, $q \geq 1$, $\mathbf{a} \in \mathrm{Pd}_q$, and $\varepsilon > 0$. Then for any $[q]^r$-tuple of $([-d,d,r])$-graphons $W = (W^z)_{z\in[q]^r}$ and $k \geq \Theta^4 \log(\Theta) q^r$ with $\Theta = \frac{2^{r+7}q^r r}{\varepsilon}$ we have*

$$\mathbb{P}\left(|\mathcal{E}_{\mathbf{a}}(W) - \hat{\mathcal{E}}_{\mathbf{a}}(\mathbb{G}(k,W))| > \varepsilon\|W\|_\infty\right) < \varepsilon.$$

Proof. Let W be as in the statement and $k \geq \Theta^4 \log(\Theta) q^r$ with $\Theta = \frac{2^{r+7}q^r r}{\varepsilon}$. We start with pointing out that we are allowed to replace the quantity $\hat{\mathcal{E}}_{\mathbf{a}}(\mathbb{G}(k,W))$ by $\mathcal{E}_{\mathbf{a}}(\mathbb{G}(k,W))$ in the statement of the theorem by Lemma 4.3.5 and only introduce an initial error at most $\frac{1}{k}\|G\|_\infty 5^r q^{r+1} \leq \frac{\varepsilon}{2}\|W\|_\infty$.

The lower bound on $\mathcal{E}_{\mathbf{a}}(\mathbb{G}(k,W))$ is the result of standard sampling argument combined with Lemma 4.3.4. Let us consider a fixed **a**-partition ϕ of $[0,1]$, and define the random fractional partition of $[k]$ as $y_{n,m} = \phi_m(U_n)$ for every $n \in [k]$ and $m \in [q]$. The partition y is not necessarily an fractional **a**-partition, but it can not be very far from being one. For $m \in [q]$ it holds that

$$\mathbb{P}\left(\left|\frac{\sum_{n=1}^{k} y_{n,m}}{k} - a_m\right| \geq \varepsilon\right) \leq 2\exp(-\varepsilon^2 k/2),$$

therefore for our choice of k the sizes of the partition classes obey $|\frac{1}{k}\sum_{n=1}^{k} y_{n,m} - a_m| < \frac{\varepsilon}{2(q+1)}$ for every $m \in [q]$ with probability at least $1 - \varepsilon/2$.

We appeal to Lemma 4.3.4 to conclude

$$\begin{aligned}
\mathbb{E}\mathcal{E}_{\mathbf{a}}(\mathbb{G}(k,W)) &\geq \mathbb{E}\mathcal{E}_{y}(\mathbb{G}(k,W)) - (\varepsilon/2)\|W\|_\infty \\
&= \mathbb{E}\frac{1}{k^r}\sum_{z\in[q]^r}\sum_{n_1,\ldots,n_r=1}^{k} W(U_{n_1},\ldots,U_{n_r})\prod_{j=1}^{r} y_{n_j,z_j} - (\varepsilon/2)\|W\|_\infty \\
&\geq \frac{k!}{k^r(k-r)!}\sum_{z\in[q]^r}\int_{[0,1]^r} W(t_1,\ldots,t_r)\prod_{j=1}^{r}\phi_{z_j}(t_j)\mathrm{d}t - \left(\frac{r^2}{k} + \varepsilon/2\right)\|W\|_\infty \\
&\geq \mathcal{E}_{\phi}(W) - \left(\frac{r^2}{k} + \varepsilon/2\right)\|W\|_\infty.
\end{aligned}$$

The concentration of the random variable $\mathcal{E}_{\mathbf{a}}(\mathbb{G}(k,W))$ can be obtained through martingale arguments identical to the technique used in the proof of the lower bound in Theorem 4.1.4.

For the upper bound on $\mathcal{E}_{\mathbf{a}}(\mathbb{G}(k,W))$ we are going to use the cut decomposition and local linearization, the approach to approximate the energy of $\mathcal{E}_{\phi}(W)$ and $\mathcal{E}_{\mathrm{x}}(\mathbb{G}(k,W))$ for certain partitions ϕ, respectively x is completely identical to the proof of Theorem 4.1.4, therefore we borrow all the notation from there, and we do not refer to again in what follows.

Now we consider a $b \in \mathcal{A}$ and define the event $E_3(b)$ that is occurrence the following implication.

If the linear program

$$\begin{aligned}
&\text{maximize} && l_0 + \sum_{n=1}^{k}\sum_{m=1}^{q}\frac{1}{k}x_{n,m}l_m(U_n) \\
&\text{subject to} && \mathrm{x} \in I'(b,\eta) \cap \omega_{\mathbf{a}} \\
& && 0 \leq x_{n,m} \leq 1 \qquad \text{for } n = 1,\ldots,k \text{ and } m = 1,\ldots,q \\
& && \sum_{m=1}^{q} x_{n,m} = 1 \qquad \text{for } n = 1,\ldots,k
\end{aligned}$$

has optimal value α, then the continuous linear program

$$\begin{aligned}
&\text{maximize} && l_0 + \int_0^1 \sum_{m=1}^{q} l_m(t)\phi_m(t)\mathrm{d}t \\
&\text{subject to} && \phi \in I(b,\eta) \cap \left(\bigcup_{\mathbf{c}:\, |a_i - c_i| \le \eta} \Omega_{\mathbf{c}} \right) \\
& && 0 \le \phi_m(t) \le 1 && \text{for } t \in [0,1] \text{ and } m = 1,\dots,q \\
& && \sum_{m=1}^{q} \phi_m(t) = 1 && \text{for } t \in [0,1]
\end{aligned}$$

has optimal value at least $\alpha - \frac{\varepsilon}{2}\|W\|_\infty$.

Recall that $\eta = \frac{\varepsilon}{16q^r 2^r}$. It follows by applying Lemma 4.2.9 that $E_3(b)$ has probability at least $1 - 2\exp\left(-\frac{k\varepsilon^2}{2^{2r+12}q^{2r}r^2}\right)$. When conditioning on E_1, the event from the proof of Theorem 4.1.4, and $E_3 = \cap_{b\in\mathcal{A}} E_3(b)$ we conclude that

$$\mathcal{E}_{\mathbf{a}}(\mathbb{G}(k,W)) \le \max_{\mathbf{c}:\, |a_i - c_i| \le \eta} \mathcal{E}_{\mathbf{c}}(W) + \varepsilon/2)\|W\|_\infty \le \mathcal{E}_{\mathbf{a}}(W) + (rq\eta + \varepsilon/2)\|W\|_\infty \le \mathcal{E}_{\mathbf{a}}(W) + \varepsilon\|W\|_\infty.$$

Also, like in Theorem 4.1.4, the probability of the required events to happen simultaneously is at least $1 - \varepsilon/2$. This concludes the proof. □

4.3.3 Quadratic assignment and maximum acyclic subgraph problem

The two optimization problems that are the subject of this subsection, the quadratic assignment problem (QAP) and maximum acyclic subgraph problem (AC), are known to be NP-hard, similarly to MAX-rCSP that was investigated above. The first polynomial time approximation schemes were designed for the QAP by Arora, Frieze and Kaplan [20]. Dealing with a QAP means informally that one aims to minimize the transportation cost of his enterprise that has n production locations and n types of production facilities. This is to be achieved by an optimal assignment of the facilities to the locations with respect to the distances (dependent on the location) and traffic (dependent on the type of the production). In formal, terms this means that we are given two real quadratic matrices of the same size, G and $J \in \mathbb{R}^{n\times n}$, and the objective is to calculate

$$\mathrm{Q}(G,J) = \frac{1}{n^2} \max_{\rho} \sum_{i,j=1}^{n} J_{i,j} G_{\rho(i),\rho(j)},$$

where ρ runs over all permutations of $[n]$. We speak of metric QAP, if the entries of J are all non-negative with zeros on the diagonal, and obey the triangle inequality, and

d-dimensional geometric QAP if the rows and columns of J can be embedded into a d-dimensional L^p metric space so that distances of the images are equal to the entries of J.

The continuous analog of the problem is the following. Given the measurable functions $W, J\colon [0,1]^2 \to \mathbb{R}$, we are interested in obtaining

$$\hat{Q}_\rho(W,J) = \int_{[0,1]^2} J(x,y)W(\rho(x),\rho(y))\mathrm{d}x\mathrm{d}y, \qquad \hat{Q}(W,J) = \max_\rho \hat{Q}_\rho(W,J),$$

where ρ in the previous formula runs over all measure preserving permutations of $[0,1]$. In even greater generality we introduce the QAP with respect to fractional permutations of $[0,1]$. A fractional permutation μ is a probability kernel, that is $\mu\colon [0,1]\times\mathcal{L}([0,1]) \to [0,1]$ so that

(i) for any $A \in \mathcal{L}([0,1])$ the function $\mu(.,A)$ is measurable,

(ii) for any $x \in [0,1]$ the function $\mu(x,.)$ is a probability measure on $\mathcal{L}([0,1])$, and

(iii) for any $A \in \mathcal{L}([0,1])$ $\int_0^1 \mathrm{d}\mu(x,A) = \lambda(A)$.

Here $\mathcal{L}([0,1])$ is the σ-algebra of the Borel sets of $[0,1]$.

Then we define

$$Q_\mu(W,J) = \int_{[0,1]^2}\int_{[0,1]^2} J(\alpha,\beta)W(x,y)\mathrm{d}\mu(\alpha,x)\mathrm{d}\mu(\beta,y)\mathrm{d}\alpha\mathrm{d}\beta,$$

and

$$Q(W,J) = \max_\mu Q_\mu(W,J),$$

where the maximum runs over all fractional permutations. For each measure preserving permutation ρ one can consider the fractional permutation μ with the probability measure $\mu(\alpha,.)$ is defined as the atomic measure $\delta_{\rho(\alpha)}$ concentrated on $\rho(\alpha)$, for this choice of μ we have $Q_\rho(W,J) = Q_\mu(W,J)$.

An r-dimensional generalization of the problem for J and $W\colon [0,1]^r \to \mathbb{R}$ is

$$Q(W,J) = \max_\mu \int_{[0,1]^r}\int_{[0,1]^r} J(\alpha_1,\dots,\alpha_r)W(x_1,\dots,x_r)\mathrm{d}\mu(\alpha_1,x_1)\dots\mathrm{d}\mu(\alpha_r,x_r)\mathrm{d}\alpha_1\dots\mathrm{d}\alpha_r,$$

where the maximum runs over all fractional permutations μ of $[0,1]$. The definition of the finitary case in r dimensions is analogous.

A special QAP is the maximum acyclic subgraph problem (AC). Here we are given a weighted directed graph G with vertex set of cardinality n, and our aim is to determine the maximum of the total value of edge weights of a subgraph of G that contains no directed cycle. We can formalize this as follows. Let $G \in \mathbb{R}^{n\times n}$ be the input data, then the maximum acyclic subgraph density is

$$\mathrm{AC}(G) = \frac{1}{n^2} \max_{\rho} \sum_{i,j=1}^{n} G_{i,j} \mathbb{1}(\rho(i) < \rho(j)),$$

where ρ runs over all permutations of $[n]$.

This can be thought of as a QAP with the restriction that J is the upper triangular $n \times n$ matrix with zeros on the diagonal and all nonzero entries being equal to 1. However in general AC cannot be reformulated as metric QAP. The continuous version of the problem

$$\hat{\mathrm{AC}}(W) = \sup_{\phi} \int_{[0,1^2]} \mathbb{1}(\phi(x) > \phi(y)) W(x,y) \mathrm{d}x \mathrm{d}y$$

for a function $W\colon [0,1]^2 \to \mathbb{R}$ is defined analogous to the QAP, where the supremum runs over measure preserving permutations $\phi\colon [0,1] \to [0,1]$, as well as the relaxation $\mathrm{AC}(W)$, where the supremum runs over probability kernels.

Both the QAP and the AC problems resemble the ground state energy problems that were investigated in previous parts of this chapter. In fact, if the number of clusters of the distance matrix J in the QAP is bounded from above by an integer that is independent from n, then this special QAP is a ground state energy with the number of states q equal to the number of clusters of J. By the number of clusters we mean here the smallest number m such that there exists an $m \times m$ matrix J' so that J is a blow-up of J', that is not necessarily equitable. To establish an approximation to the solution of the QAP we will only need the cluster condition approximately, and this will be shown in what follows.

Definition 4.3.7. *We call a measurable function $J\colon [0,1]^r \to \mathbb{R}$ ν-clustered for a non-increasing function $\nu\colon \mathbb{R}^+ \to \mathbb{R}^+$, if for any $\varepsilon > 0$ there exists another measurable function $J'\colon [0,1]^r \to \mathbb{R}$ that is a step function with $\nu(\varepsilon)$ steps and $\|J - J'\|_1 < \varepsilon \|J\|_\infty$.*

Note, that by the Weak Regularity Lemma ([59]), Lemma 3.3.4, any J can be well approximated by a step function with $\nu(\varepsilon) = 2^{\frac{1}{\varepsilon^2}}$ steps in the cut norm. To see why it is likely that this approximation will not be sufficient for our purposes, consider an arbitrary $J\colon [0,1]^r \to \mathbb{R}$. Suppose that we have an approximation in the cut norm of J at hand denoted by J'. Define the probability kernel $\mu_0(\alpha, .) = \delta_\alpha$ and the naive r-kernel $W_0 = J - J'$. In this case $|\mathrm{Q}_{\mu_0}(W, J) - \mathrm{Q}_{\mu_0}(W, J')| = \|J - J'\|_2^2$. This 2-norm is not granted to be small compared to ε by any means.

In some special cases, for example if J is a d-dimensional geometric array or the array corresponding to the AC, we are able to require bounds on the number of steps required for the 1-norm approximation of J that are sub-exponential in $\frac{1}{\varepsilon}$. By the aid of this fact we can achieve good approximation of the optimal value of the QAP via sampling. Next we state an application of Theorem 4.1.4 to the clustered QAP.

Lemma 4.3.8. *Let $\nu : \mathbb{R}^+ \to \mathbb{R}^+$ be nondecreasing, and let $J \colon [0,1]^r \to \mathbb{R}$ be a ν-clustered measurable function. Then there exists an absolute constant $c > 0$ so that for every $\varepsilon > 0$, every naive r-kernel W, and $k \geq c \log(\frac{\nu(\varepsilon)^r}{\varepsilon})(\frac{\nu(\varepsilon)^r}{\varepsilon})^4$ we have*

$$\mathbb{P}(|\mathrm{Q}(W,J) - \mathrm{Q}(\mathbb{G}(k,W), \mathbb{G}'(k,J))| \geq \varepsilon \|W\|_\infty \|J\|_\infty) \leq \varepsilon,$$

where $\mathbb{G}(k,W)$ and $\mathbb{G}'(k,J)$ are generated by distinct independent samples.

Proof. Without loss of generality we may assume that $\|J\|_\infty \leq 1$. First we show that under the cluster condition we can introduce a microcanonical ground state energy problem whose optimum is close to $\mathrm{Q}(W,J)$, and the same holds for the sampled problem. Let $\varepsilon > 0$ be arbitrary and J' be an approximating step function with $q = \nu(\varepsilon)$ steps. We may assume that $\|J'\|_\infty \leq 1$. We set $\mathbf{a} = (a_1, \dots, a_q)$ to be the vector of the sizes of the steps of J', and construct from J' a real r-array of size q in the natural way by associating to each class of the steps of J' an element of $[q]$ (indexes should respect $\mathbf{a}$), and set the entries of the r-array corresponding to the value of the respective step of J'. We will call the resulting r-array J''. From the definitions it follows that

$$\mathrm{Q}(W,J') = \mathcal{E}_{\mathbf{a}}(W,J'')$$

for every r-kernel W. On the other hand we have

$$\begin{aligned}
|\mathrm{Q}(W,J)-&\mathrm{Q}(W,J')| \\
&\leq \max_\mu \left|\mathrm{Q}_\mu(W,J) - \mathrm{Q}_\mu(W,J')\right| \\
&= \max_\mu \left| \int_{[0,1]^r} (J-J')(\alpha_1,\dots,\alpha_r) \int_{[0,1]^r} W(x) \mathrm{d}\mu(\alpha_1,x_1)\dots\mathrm{d}\mu(\alpha_r,x_r) \mathrm{d}\alpha_1\dots\mathrm{d}\alpha_r \right| \\
&\leq \max_\mu \int_{[0,1]^r} |(J-J')(\alpha_1,\dots,\alpha_r)| \|W\|_\infty \mathrm{d}\alpha_1\dots\mathrm{d}\alpha_r \\
&= \|J-J'\|_1 \|W\|_\infty \leq \varepsilon \|W\|_\infty.
\end{aligned}$$

Now we proceed to the sampled version of the optimization problem. First we gain control over the difference between the QAPs corresponding to J and J'. $\mathbb{G}(k,W)$ is induced by the sample $U_1,\dots,U_k$, $\mathbb{G}'(k,J)$ and $\mathbb{G}'(k,J')$ by the distinct independent sample $Y_1,\dots,Y_k$.

$$\begin{aligned}
|\mathrm{Q}(\mathbb{G}(k,W),&\mathbb{G}(k,J)) - \mathrm{Q}(\mathbb{G}(k,W),\mathbb{G}(k,J'))| \\
&\leq \max_\rho |\mathrm{Q}_\rho(\mathbb{G}(k,W),\mathbb{G}(k,J)) - \mathrm{Q}_\rho(\mathbb{G}(k,W),\mathbb{G}(k,J'))| \\
&= \max_\rho \frac{1}{k^r} \left| \sum_{i_1,\dots,i_r=1}^{k} (J-J')(Y_{i_1},\dots,Y_{i_r}) \right| \|W\|_\infty. \qquad (4.21)
\end{aligned}$$

We analyze the random sum on the right hand side of (4.21) by first upper bounding its expectation.

$$\begin{aligned}\frac{1}{k^r}\mathbb{E}_Y\left|\sum_{i_1,\ldots,i_r=1}^{k}(J-J')(Y_{i_1},\ldots,Y_{i_r})\right| \\ &\le \frac{r^2}{k}\left[\|J\|_\infty+\|J'\|_\infty\right]+\mathbb{E}_Y|(J-J')(Y_1,\ldots,Y_r)| \\ &= \frac{2r^2}{k}+\varepsilon\le 2\varepsilon.\end{aligned}$$

By the Azuma-Hoeffding inequality the sum is also sufficiently small in probability.

$$\mathbb{P}\left(\frac{1}{k^r}\left|\sum_{i_1,\ldots,i_r=1}^{k}(J-J')(Y_{i_1},\ldots,Y_{i_r})\right|\ge 4\varepsilon\right)\le 2\exp(-\varepsilon^2k/8)\le\varepsilon.$$

We obtain that

$$|Q(\mathbb{G}(k,W),\mathbb{G}'(k,J))-Q(\mathbb{G}(k,W),\mathbb{G}'(k,J'))|\le 4\varepsilon\|W\|_\infty$$

with probability at least $1-\varepsilon$, if k is such as in the statement of the lemma. Set $\mathbf{b}=(b_1,\ldots,b_q)$ to be the probability distribution for that $b_i=\frac{1}{k}\sum_{j=1}^{k}\mathbb{1}_{S_i}(Y_j)$, where S_i is the ith step of J' with $\lambda(S_i)=a_i$. Then we have

$$Q(\mathbb{G}(k,W),\mathbb{G}(k,J'))=\hat{\mathcal{E}}_{\mathbf{b}}(\mathbb{G}(k,W),J'').$$

It follows again from the Azuma-Hoeffding inequality that we have $\mathbb{P}(|a_i-b_i|>\varepsilon/q)\le 2\exp(-\frac{\varepsilon^2k}{2q^2})$ for each $i\in[q]$, thus we have $\|\mathbf{a}-\mathbf{b}\|_1<\varepsilon$ with probability at least $1-\varepsilon$. We can conclude that with probability at least $1-2\varepsilon$ we have

$$\begin{aligned}|Q(W,J)-Q(\mathbb{G}(k,W),\mathbb{G}(k,J))| &\le |Q(W,J)-Q(W,J')| \\ &+|\mathcal{E}_{\mathbf{a}}(W,J'')-\hat{\mathcal{E}}_{\mathbf{b}}(\mathbb{G}(k,W),J'')|+|Q(\mathbb{G}(k,W),\mathbb{G}(k,J))-Q(\mathbb{G}(k,W),\mathbb{G}(k,J'))| \\ &\le (5+2r)\varepsilon\|W\|_\infty+|\mathcal{E}_{\mathbf{a}}(W,J'')-\hat{\mathcal{E}}_{\mathbf{a}}(\mathbb{G}(k,W),J'')|.\end{aligned}$$

By the application of Theorem 4.3.6 the claim of the lemma is verified.

□

Next we present the application of Lemma 4.3.8 for two special cases of QAP.

Corollary 4.3.9. *The optimal values of the d-dimensional geometric QAP and the maximum acyclic subgraph problem are efficiently testable. That is, let $d\ge 1$, for every $\varepsilon>0$ there exists an integer $k_0=k_0(\varepsilon)$ such that k_0 is a polynomial in $1/\varepsilon$, and for every $k\ge k_0$ and any d-dimensional geometric QAP given by the pair (G,J) we have*

$$\mathbb{P}(|Q(G,J)-Q(\mathbb{G}(k,G),\mathbb{G}'(k,J))|\ge\varepsilon\|G\|_\infty\|J\|_\infty)\le\varepsilon, \qquad (4.22)$$

where $\mathbb{G}(k, W)$ and $\mathbb{G}'(k, J)$ are generated by distinct independent samples. The formulation regarding the testability of the maximum acyclic subgraph problem is analogous.

Note that testability here is meant in the sense of the statement of Lemma 4.3.8, since the size of J is not fixed and depends on G.

Proof. In the light of Lemma 4.3.8 it suffices to show that for both cases any feasible J is ν-clustered, where $\nu(\varepsilon)$ is polynomial in $1/\varepsilon$. For both settings we have $r = 2$.

We start with the continuous version of the d-dimensional geometric QAP given by the measurable function $J\colon [0,1]^2 \to \mathbb{R}^+$, and an instance is given by the pair (W, J), where W is a 2-kernel. Note, that d refers to the dimension corresponding to the embedding of the indices of J into an L^p metric space, not the actual dimension of J. We are free to assume that $0 \leq J \leq 1$, simply by rescaling. By definition, there exists a measurable embedding $\rho\colon [0,1] \to [0,1]^d$, so that $J(i,j) = \|\rho(i) - \rho(j)\|_p$ for every $(i,j) \in [0,1]^2$. Fix $\varepsilon > 0$ and consider the partition $\mathcal{P}' = (T_1, \dots, T_\beta) = ([0, \frac{1}{\beta}), [\frac{1}{\beta}, \frac{2}{\beta}), \dots, [\frac{\beta-1}{\beta}, 1])$ of the unit interval into $\beta = \lceil \frac{2\sqrt[p]{d}}{\varepsilon} \rceil$ classes. Define the partition $\mathcal{P} = (P_1, \dots, P_q)$ of $[0,1]$ consisting of the classes $\rho^{-1}(T_{i_1} \times \dots \times T_{i_d})$ for each $(i_1, \dots, i_d) \in [\beta]^r$, where $|\mathcal{P}| = q = \beta^d = \frac{2^d d^{d/p}}{\varepsilon^d}$. We construct the approximating step function J' of J by averaging J on the steps determined by the partition classes of $\mathcal{P}$. It remains to show that this indeed is a sufficient approximation in the L^1-norm.

$$\|J - J'\|_1 = \int_{[0,1]^2} |J(x) - J'(x)| \mathrm{d}x = \sum_{i,j=1}^{q} \int_{P_i \times P_j} |J(x) - J'(x)| \mathrm{d}x \leq \sum_{i,j}^{q} \frac{1}{q^2} \varepsilon = \varepsilon.$$

By Lemma 4.3.8 and Theorem 4.3.6 it follows that the continuous d-dimensional metric QAP is $O(\log(\frac{1}{\varepsilon}) \frac{1}{\varepsilon^{4rd+4}})$-testable, and so is the discrete version of it.

Next we show that the AC is also efficiently testable given by the upper triangular matrix J whose entries above the diagonal are 1. Note that here we have $r = 2$. Fix $\varepsilon > 0$ and consider the partition $\mathcal{P} = (P_1, \dots, P_q)$ with $q = \frac{1}{\varepsilon}$, and set J' to 0 on every step $P_i \times P_j$ whenever $i \geq j$, and to 1 otherwise. This function is indeed approximating J in the L^1-norm.

$$\|J - J'\|_1 = \int_{[0,1]^2} |J(x) - J'(x)| \mathrm{d}x = \sum_{i=1}^{q} \int_{P_i \times P_i} |J(x) - J'(x)| \mathrm{d}x \leq \varepsilon.$$

Again, by Lemma 4.3.8 and Theorem 4.3.6 it follows that the AC is $O(\log(\frac{1}{\varepsilon}) \frac{1}{\varepsilon^{12}})$-testable. □

4.4 Energies of unbounded graphs and graphons

In the present section we will investigate upper bounds on the sample size that is required approximate to ground state energies from Definition 4.1.2 of unbounded families of kernels and weighted graphs. A subfamily of this class in question was recently also considered by Lovász [89]. Namely, in [89, Section 17.2] it was shown that any sequence of convergent unbounded symmetric 2-kernels whose elements are all members of $\cap_{p\geq 1} L^p$ has a continuous limit object that has the same strict finiteness property for each of its moments. For this particular kind of result the membership in $\cap_{p\geq 1} L^p$ is essential, since otherwise for a kernel not included in L^p for a particular p the density of a path of length $2p$ is infinite, and for the original convergence notion the finiteness of all subgraph densities is required.

For our approach we are allowed to deal with a broader family of graphons, for reasonable results only the much weaker property of uniform integrability is necessary, or equivalently membership in a single L^p space. The statement will be formulated in terms of the p-norm of the graphon and the graph. Motivation for our setup originates from approximation theory of optimization problems, it was shown by Fernandez de la Vega and Karpinski [50] that MAX-CUT has PTAS (polynomial time approximation scheme) for families of graphs whose the edge weights are unbounded, but still obey some condition analogous to uniform integrability.

On a further note we mention that such families were considered by Borgs, Chayes, Cohn, and Zhao [33, 34] in terms of their convergence in energies we deal with here and in a normalized version of the cut distance $\delta_\square$.

Definition 4.4.1. *Let $\kappa\colon \mathbb{R}^+ \to \mathbb{R}^+$ be a non-increasing measurable function. We call a real-valued measurable function f on $\mathbb{R}^d$ uniformly κ-integrable if for each $\varepsilon > 0$ we have*

$$\|f \mathbb{1}_{\{f\geq\kappa(\varepsilon)\}}\|_1 < \varepsilon.$$

Let $r \geq 1$, and $\kappa_i\colon \mathbb{R}^+ \to \mathbb{R}^+$ be non-increasing measurable functions for each $i \in [r]$. We call an r-kernel W κ-integrable for $\kappa = (\kappa_1, \ldots, \kappa_r)$, if for every $e \subset [r]$ with $|e| = l$ the marginal functions W_e of W given by $W_e(x_{e_1}, \ldots, x_{e_l}) = \int_{[0,1]^{[r]\setminus e}} W(x_1, \ldots, x_r) \mathrm{d}\lambda(x_{\mathfrak{h}([r]\setminus e,1)})$ are κ_l-integrable functions for each $1 \leq l \leq r$. We call a symmetric r-kernel an L^p-graphon if it is in L^p as a function.

For random variables instead of graphons and functions the definitions are analogous.

Remark 4.4.2. The previous definition attempts to distinguish heterogeneous graphs with heterogeneous and homogeneous degree distribution. We will see below why this distinction is advantageous, and indeed the degrees of a graph can be bounded by a linear function of the vertex cardinality even if the edge weights of the graph are super-cubic. Imagine an infinite sequence $(G_i)_{i=1}^{\infty}$ of Erdős-Rényi graphs with density 2^{-2i} for G_i on the common node set $[n]$, and add them up weighted by 2^i to obtain a weighted graph H with unbounded edge weights. Then with high probability the degrees of H are bounded above by 2.

If we restrict our attention to families of L^p-graphons instead of a uniformly bounded family of graphons, the statement about efficient testability of energies (definitions are identical to the bounded case) still remains valid outlined in the first section of the chapter, and we obtain a generalization of Theorem 4.1.4.

Theorem 4.4.3. *Let $q \geq 1$, $J \in \mathbb{R}^{q\times\cdots\times q}$ an r-array, and let $p > 1$. For every $\varepsilon > 0$, every naive r-graphon W with $\mathbb{E}|W|^p < \infty$ and $k \geq \Theta^4 \log(\Theta)$ with $\Theta = 2^{r+7}\left(\frac{4q^r r}{\varepsilon}\right)^{1+1/(p-1)}$ we have*

$$\mathbb{P}(|\mathcal{E}(W,J) - \hat{\mathcal{E}}(\mathbb{G}(k,W),J)| \geq \varepsilon\|J\|_\infty\|W\|_p) \leq \varepsilon.$$

We can translate the requirement of the existence of certain moments easily to a type of uniform integrability, we will need this connection in the proof of Theorem 4.4.3, that is presented at the end of the section.

Note that there is a strong connection between the property of κ-integrability and the existence of the moments of a random variable. The asymptotic magnitude of κ as a function of $1/\varepsilon$ is in our analysis of major importance as ε approaches 0, therefore we will state the relationship in a quantitative form, and will also provide the proof of this folklore result.

Lemma 4.4.4. *Let X be an integrable random variable. Then for any $p > 1$ we have the following two statements.*

(i) If $\mathbb{E}|X|^p < \infty$, then X is κ-integrable with $\kappa(\varepsilon) = (\frac{\|X\|_p^p}{\varepsilon})^{\frac{1}{p-1}}$, where $\|X\|_p = (\mathbb{E}|X|^p)^{1/p}$.

(ii) If X is κ-integrable with $\kappa(\varepsilon) = K\varepsilon^{-\frac{1}{p-1}}$, then $\mathbb{E}|X|^q < \infty$ for any $1 < q < p$.

Proof. We fix $p > 1$, and start with *(i)*. Suppose that $\mathbb{E}|X|^p < \infty$ and set $\kappa(\varepsilon) = (\frac{\|X\|_p^p}{\varepsilon})^{\frac{1}{p-1}}$. Then we have

$$\mathbb{E}|X|\mathbb{1}_{\{|X|\geq\kappa(\varepsilon)\}} \leq \frac{1}{\kappa(\varepsilon)^{p-1}}\mathbb{E}|X|^p < \varepsilon.$$

For *(ii)* let $1 < q < p$ and $\kappa(\varepsilon) = K\varepsilon^{-\frac{1}{p-1}}$ for $\varepsilon < 1$. Note that $\mathbb{E}|X| < C$ for some $C > 0$. Then

$$\begin{aligned}\mathbb{E}|X|^q &= \mathbb{E}[|X|\int_0^\infty (q-1)\mathbb{1}_{\{|X|\geq t\}}t^{q-2}\mathrm{d}t] \\ &= (q-1)\int_0^\infty t^{q-2}\mathbb{E}[|X|\mathbb{1}_{\{|X|\geq t\}}]\mathrm{d}t \\ &\leq C(q-1) + K^{p-1}(q-1)\int_1^\infty t^{-1-(p-q)}\mathrm{d}t < \infty,\end{aligned}$$

where the finiteness of the integral is implied by $1 < q < p$.

□

We will need the quantitative version of the Weak Law of Large Numbers for naive r-kernels, originally proved for $r = 1$ by Khinchin (see [78] for the usual statement and proof). We include the proof here for the sake of completeness.

Lemma 4.4.5. *Let $r \geq 1$ and $W\colon [0,1]^r \to \mathbb{R}$ be an r-symmetric r-kernel such that $W \in L^p([0,1]^r)$ for some $p > 1$ real. Then for any $\varepsilon > 0$ and $n \geq \varepsilon^{-3-4/(p-1)} 2^{\frac{(r+5)p+1}{p-1}} r^{\frac{1}{p-1}}$, and the uniform independent random sample $X_1, \ldots, X_n$ from $[0,1]$ we have*

$$\mathbb{P}\left(\left| \frac{\sum_{i_1,\ldots,i_r=1}^n W(X_{i_1}, \ldots, X_{i_r})}{n^r} - \mathbb{E}W(X_1, \ldots, X_r) \right| > \varepsilon \|W\|_p \right) < \varepsilon.$$

Proof. Let $\varepsilon > 0$, $p > 1$, and let $n \in \mathbb{N}$ be fixed and as large as in the statement of the lemma. Let W be an arbitrary naive r-kernel that is contained in $L^p([0,1]^r)$, and set $\delta = 2^{-r-5}\varepsilon^3 \|W\|_p$. An easy consequence of Jensen's inequality is that for any nonempty subset S of $[r]$ it is true that $\mathbb{E}|W_S|^p = \mathbb{E}|\mathbb{E}[W(X_1, \ldots, X_r) \mid \{ X_i \mid i \in S \}]|^p \leq \mathbb{E}|W|^p$, since the pth power of the absolute value is a convex function. Further, with $\Delta(\gamma) = \left(\frac{1}{\gamma}\right)^{1/(p-1)} \|W\|_p$ we have by Lemma 4.4.4 for each $\gamma > 0$ that

$$\mathbb{E}|W| \mathbb{1}_{\{|W_S| \geq \Delta(\gamma)\}} \leq \frac{1}{\Delta(\gamma)^{p-1}} \mathbb{E}|W_S|^p \leq \frac{1}{\Delta(\gamma)^{p-1}} \mathbb{E}|W|^p = \gamma \|W\|_p. \tag{4.23}$$

We will truncate W level-wise, first we eliminate vertices with high degree, then edges containing pairs whose co-degree is large, an proceed this way to finally exclude all edges with large weights over our chosen threshold. More precisely, we define the graphons

$$W'(x_1, \ldots, x_r) = W(x_1, \ldots, x_r) \prod_{\emptyset \neq S \subset [r]} \mathbb{1}_{\{|W_S(x_1, \ldots, x_{|S|})| \leq \delta n^{|S|}\}}$$

that has bounded edge weights, and the remainder

$$W''(x_1, \ldots, x_r) = W(x_1, \ldots, x_r) - W'(x_1, \ldots, x_r)$$

for each $x_1, \ldots, x_r \in [0,1]$. Note that

$$|\mathbb{E}W - \mathbb{E}W'| \leq \sum_{\emptyset \neq S \subset [r]} \mathbb{E}|W| \mathbb{1}_{\{|W_S| > \delta n^{|S|}\}} \leq \varepsilon/4 \|W\|_p, \tag{4.24}$$

since $\delta n \geq \Delta(2^{-r} \frac{\varepsilon}{4})$ by the condition on n in the statement of the lemma. Then

$$\begin{aligned} &\mathbb{P}\left(\left| \frac{\sum_{i_1,\ldots,i_r=1}^n W(X_{i_1}, \ldots, X_{i_r})}{n^r} - \mathbb{E}W \right| > \varepsilon \|W\|_p \right) \\ &\leq \mathbb{P}\left(\left| \frac{\sum_{i_1,\ldots,i_r=1}^n W'(X_{i_1}, \ldots, X_{i_r})}{n^r} - \mathbb{E}W' \right| > \varepsilon/4 \|W\|_p \right) \\ &\quad + \mathbb{P}\left(\left| \frac{\sum_{i_1,\ldots,i_r=1}^n W''(X_{i_1}, \ldots, X_{i_r})}{n^r} \right| > \varepsilon/2 \|W\|_p \right). \end{aligned} \tag{4.25}$$

We estimate the first term of (4.25) by Chebysev's inequality.

$$\mathbb{P}\left(\left|\frac{\sum_{i_1,\dots,i_r=1}^n W'(X_{i_1},\dots,X_{i_r})}{n^r} - \mathbb{E}W'\right| > \varepsilon/4\|W\|_p\right) \le \frac{16\mathbb{E}\left(\sum_{i_1,\dots,i_r=1}^n [W'(X_{i_1},\dots,X_{i_r}) - \mathbb{E}W']\right)^2}{n^{2r}\varepsilon^2\|W\|_p^2}. \tag{4.26}$$

We expand the square in the expectation on the left of (4.26), and group the terms with respect to the set of common variables in the parameters of the two random variables of the product. Note that if there are no common variables in the factors, then the factors are independent, so the corresponding product vanishes in expectation.

$$\begin{aligned}
&\mathbb{E}(\sum_{i_1,\dots,i_r=1}^n [W'(X_{i_1},\dots,X_{i_r}) - \mathbb{E}W'])^2 \\
&= \sum_{\emptyset\neq S\subset[r]} \sum_{\substack{i_1,\dots,i_r=1\\ j_1,\dots,j_r=1\\ i_k=j_k\Leftrightarrow k\in S}}^n [\mathbb{E}W'(X_{i_1},\dots,X_{i_r})W'(X_{j_1},\dots,X_{j_r}) - (\mathbb{E}W')^2] \\
&\le \sum_{\emptyset\neq S\subset[r]} \sum_{\substack{i_1,\dots,i_r=1\\ j_1,\dots,j_r=1\\ i_k=j_k\Leftrightarrow k\in S}}^n \mathbb{E}\left[\mathbb{E}[W'(X_{i_1},\dots,X_{i_r}) \mid \{X_{i_k} \mid k\in S\}]\mathbb{E}[W'(X_{j_1},\dots,X_{j_r}) \mid \{X_{j_k} \mid k\in S\}]\right] \\
&\le \sum_{\emptyset\neq S\subset[r]} \sum_{\substack{i_1,\dots,i_r=1\\ j_1,\dots,j_r=1\\ i_k=j_k\Leftrightarrow k\in S}}^n \mathbb{E}|W|\delta n^{|S|} = \sum_{s=1}^r \binom{r}{s} n^{2r-s}\mathbb{E}|W|\delta n^s \le \|W\|_p\delta n^{2r}2^r.
\end{aligned} \tag{4.27}$$

Plugging in the previous estimate into (4.26) we arrive at

$$\mathbb{P}\left(\left|\frac{\sum_{i_1,\dots,i_r=1}^n W'(X_{i_1},\dots,X_{i_r})}{n^r} - \mathbb{E}W'\right| > \varepsilon/4\|W\|_p\right) \le \frac{\|W\|_p\delta 2^{r+4}}{\varepsilon^2\|W\|_p^2}. \tag{4.28}$$

Recall that $\delta = 2^{-r-5}\varepsilon^3\|W\|_p$, thus we have that the right hand side of (4.28) is at most $\frac{\varepsilon}{2}$.

We turn to bound the second term in (4.25).

$$\begin{aligned}
&\mathbb{P}\left(\left|\frac{\sum_{i_1,\dots,i_r=1}^n W''(X_{i_1},\dots,X_{i_r})}{n^r}\right| > \varepsilon/2\|W\|_p\right) \\
&\quad\le \sum_{i_1=1}^n \mathbb{P}\left(\left|\sum_{i_2,\dots,i_r=1}^n W''(X_{i_1},\dots,X_{i_r})\right| > 0\right) \\
&\quad\le n[\mathbb{P}(|W_1(X_1)| > \delta n) + \mathbb{P}(\{|W_1(X_1)| \le \delta n\} \cap (\cup_{i_2=1}^n\{|\sum_{i_3,\dots,i_r=1}^n W''(X_{i_1},\dots,X_{i_r})| > 0\})]
\end{aligned}$$

$$\le \sum_{j=1}^{r} n^j \mathbb{P}(|W_{\{1,\dots,j\}}(X_1,\dots,X_j)| > \delta n^j)$$
$$\le \sum_{j=1}^{r} n^j \frac{1}{\delta n^j} \mathbb{E}|W_{[j]}| \mathbb{1}_{\{|W_{[j]}|>\delta n^j\}}$$
$$\le \sum_{j=1}^{r} \frac{1}{\delta} \frac{\mathbb{E}|W|^p}{\delta^{p-1}(n^{p-1})^j}$$
$$\le \frac{r\mathbb{E}|W|^p}{\delta^p n^{p-1}}. \tag{4.29}$$

Because of the choice of the lower bound on n we have that (4.29) is at most $\frac{\varepsilon}{2}$, which implies the statement of the lemma. □

We turn to the proof of the main result of this section.

Proof of Theorem 4.4.3.
Fix $\varepsilon > 0$, let $W\colon [0,1]^r \to \mathbb{R}$ and k be as in the statement, in particular $W \in L^p([0,1]^r)$. We define U as the truncated version of W, that is $U = W\mathbb{1}_{\{|W|\le \|W\|_p (4q^r/\varepsilon)^{1/(p-1)}\}}$. For any fractional q-partition $\phi = (\phi_1,\dots,\phi_q)$ we have

$$\begin{aligned}
|\mathcal{E}_\phi(W,J) - \mathcal{E}_\phi(U,J)| &= \left|\sum_{z\in[q]^r} J(z) \int_{[0,1]^r} \prod_{j=1}^{r} \phi_{z_j}(x_j)(W-U)(x)\mathrm{d}\lambda(x)\right| \\
&\le \|J\|_\infty \sum_{z\in[q]^r} \int_{[0,1]^r} |W(x) - U(x)|\mathrm{d}\lambda(x) \\
&= \|J\|_\infty q^r \|W-U\|_1 \\
&= \|J\|_\infty q^r \mathbb{E}|W|\mathbb{1}_{\{|W|>\|W\|_p(4q^r/\varepsilon)^{1/(p-1)}\}} \\
&\le \|J\|_\infty q^r \frac{\varepsilon\|W\|_p}{4q^r} = \frac{\varepsilon}{4}\|J\|_\infty \|W\|_p.
\end{aligned}$$

It follows that $|\mathcal{E}(W,J) - \mathcal{E}(U,J)| \le \frac{\varepsilon}{4}\|J\|_\infty\|W\|_p$.

Since $W \in L^p([0,1]^r)$ and U is bounded it follows that with setting $V = W - U$ we have $V \in L^p([0,1]^r)$, further, $\|V\|_p \le \|W\|_p$. Now we apply Lemma 4.4.5 to V with k to obtain that for uniform independent random sample $X_1,\dots,X_k$ from $[0,1]$ we have

$$\mathbb{P}\left(\left|\frac{\sum_{i_1,\dots,i_r=1}^{k} V(X_{i_1},\dots,X_{i_r})}{k^r} - \mathbb{E}V(X_1,\dots,X_r)\right| > \frac{\varepsilon}{4q^r}\|V\|_p\right) < \frac{\varepsilon}{4q^r},$$

since $k \ge \varepsilon^{-3-4/(p-1)} q^{3+4/(p-1)} 2^{\frac{(r+11)p+3}{p-1}} r^{\frac{1}{p-1}}$. Let us condition on the event

$$E_1 = \left\{\left|\frac{\sum_{i_1,\dots,i_r=1}^{k} V(X_{i_1},\dots,X_{i_r})}{k^r} - \mathbb{E}V(X_1,\dots,X_r)\right| \le \frac{\varepsilon}{4q^r}\|V\|_p\right\},$$

this implies trivially the event

$$\left\{\left|\frac{\sum_{i_1,\dots,i_r=1}^k V(X_{i_1},\dots,X_{i_r})}{k^r} - \mathbb{E}V(X_1,\dots,X_r)\right| \le \frac{\varepsilon}{4q^r}\|W\|_p\right\}. \tag{4.30}$$

Let x be an arbitrary integer q-partition of $[k]$, and let the random graphs below be generated by the same sample, then

$$\begin{aligned}
|\mathcal{E}_{\mathrm{x}}(\mathbb{G}(k,W),J)&-\mathcal{E}_{\mathrm{x}}(\mathbb{G}(k,U),J)| \\
&= \frac{1}{k^r}\left|\sum_{z\in[q]^r} J(z)\sum_{i_1,\dots,i_r=1}^{k}(W(X_{i_1},\dots,X_{i_r}) - U(X_{i_1},\dots,X_{i_r}))\prod_{j=1}^{r} x_{i_j,z_j}\right| \\
&\le \|J\|_\infty q^r \frac{\sum_{i_1,\dots,i_r=1}^k V(X_{i_1},\dots,X_{i_r})}{k^r} \\
&\le q^r\|J\|_\infty\left[\frac{\varepsilon}{4q^r}\|W\|_p + \|W-U\|_1\right] \\
&\le \frac{\varepsilon}{2}\|J\|_\infty\|W\|_p,
\end{aligned}$$

where the second inequality holds in the event of (4.30). Consequently,

$$|\hat{\mathcal{E}}(\mathbb{G}(k,W),J) - \hat{\mathcal{E}}(\mathbb{G}(k,U),J)| \le \frac{\varepsilon}{2}\|J\|_\infty\|W\|_p.$$

We know that $\|U\|_\infty \le \|W\|_p(4q^r/\varepsilon)^{1/(p-1)}$, therefore by Corollary 4.2.1 and Remark 4.2.12 we have

$$\begin{aligned}
\mathbb{P}\Big(|\mathcal{E}(U,J) - \hat{\mathcal{E}}(\mathbb{G}(k,U),J)| > \varepsilon^{1+1/(p-1)}q^{-r/(p-1)}2^{-2/(p-1)-2}\|U\|_\infty\|J\|_\infty\Big)& \\
\le \varepsilon^{1+1/(p-1)}q^{-r/(p-1)}2^{-2/(p-1)-2}&,
\end{aligned} \tag{4.31}$$

since $k \ge \Theta^4\log(\Theta)$ with $\Theta = \frac{2^{r+9+2/(p-1)}q^{r(1+1/(p-1))}r}{\varepsilon^{1+1/(p-1)}}$. We condition on the event

$$E_2 = \left\{|\mathcal{E}(U,J) - \hat{\mathcal{E}}(\mathbb{G}(k,U),J)| \le \frac{\varepsilon}{4}\|J\|_\infty\|W\|_p\right\},$$

which is the same as the complement of the event in the argument in (4.31).

The failure probability of both E_1 and E_2 is at most $\varepsilon/2$, so they hold simultaneously with probability at least $1-\varepsilon$. In this case

$$\begin{aligned}
|\mathcal{E}(W,J) - \hat{\mathcal{E}}(\mathbb{G}(k,W),J)| \le |\mathcal{E}(W,J) - \mathcal{E}(U,J)| &+ |\mathcal{E}(U,J) - \hat{\mathcal{E}}(\mathbb{G}(k,U),J)| \\
&+ |\hat{\mathcal{E}}(\mathbb{G}(k,W),J) - \hat{\mathcal{E}}(\mathbb{G}(k,U),J)| \\
\le \varepsilon\|J\|_\infty\|W\|_p&,
\end{aligned}$$

which concludes the proof.

4.5 Lower threshold ground state energies

Recall the notion of the microcanonical ground state energies in Definition 4.3.3. The main goal of this section is to introduce and to reveal the properties of an intermediate object between the microcanonical ground state energy (MGSE) and ground state energy (GSE) of weighted graphs defined in [32]. The main contribution here is that we give a convergence hierarchy with respect to the aforementioned intermediate objects that are Hamiltonians subject to certain conditions. This can be regarded as a refined version of Theorem 2.9. (ii) from [32] combined with the equivalence assertion of Theorem 2.8. (v) from the same paper. In short, these state that MGSE convergence implies GSE convergence. Counterexamples are provided indicating that the implication is strict. We also reprove with the aid of the established hierarchy one of the main results of [28]. Our motivation comes from cluster analysis, where the minimal cut problem is a central subject of research. The graph limit theory sheds new light on this, especially their statistical physics correspondence suits for application in the cluster analysis setting.

Next we introduce the central object of investigation in the current section. The set Sym_q denotes the set of real symmetric $q \times q$ matrices.

Definition 4.5.1. *Let G be a weighted graph, $q \geq 1$, $J \in \mathrm{Sym}_q$, and $0 \leq c \leq 1/q$. We define the set $A_c = \{\, \mathbf{a} \in \mathrm{Pd}_q \mid a_i \geq c, i = 1, \dots, q \,\}$, and with its help the lower threshold ground state energy (LTGSE):*

$$\hat{\mathcal{E}}^c(G, J) = \inf_{\mathbf{a} \in A_c} \hat{\mathcal{E}}_{\mathbf{a}}(G, J). \tag{4.32}$$

In a similar manner we introduce the LTGSEs for a graphon W for $q \geq 1$, $J \in \mathrm{Sym}_q$, and $0 \leq c \leq 1/q$ lower threshold:

$$\mathcal{E}^c(W, J) = \inf_{\mathbf{a} \in A_c} \mathcal{E}_{\mathbf{a}}(W, J). \tag{4.33}$$

This section is organized as follows. In Section 4.5.1 we prove yet another equivalent condition to the convergence of a simple graph sequence relying on a subclass of MGSE, the reasoning will be instrumental for the proof of our main result in the subsequent subsection. In Section 4.5.2 we study the convergence of LTGSEs, see (4.33) for their definition. We will consider $c\colon \mathbb{N} \to [0, 1]$ threshold functions with the property that $c(q)q$ is constant as a function of q. For this case we will prove that if $0 \leq c_1(q) < c_2(q) \leq 1/q$ (for all q), then the convergence of $(\mathcal{E}^{c_2(q)}(W_n, J))_{n \geq 1}$ for all $q \geq 1$, $J \in \mathrm{Sym}_q$ implies the convergence of $(\mathcal{E}^{c_1(q)}(W_n, J))_{n \geq 1}$ for all $q \geq 1$, $J \in \mathrm{Sym}_q$.

In Section 4.5.3 we provide some examples of graphs and graphons which support the fact, that the implication of convergence in Section 4.5.2 is strict in the sense that convergence of LTGSE sequences with smaller threshold do not imply convergence of LTGSE sequences with larger threshold in general. We also present a one-parameter family of block-diagonal graphons whose elements can be distinguished by LTGSEs for any threshold $c > 0$.

4.5.1 Microcanonical convergence

We start by showing that for each discrete probability distribution with rational probabilities there exists a uniform probability distribution, such that the microcanonical ground state energies (MGSE) of it can be expressed as MGSEs corresponding to the uniform distribution.

Lemma 4.5.2. *Let $q \geq 1$, and $\mathbf{a} \in \mathrm{Pd}_q$ be such that $\mathbf{a} = (\frac{k_1}{q'}, \frac{k_2-k_1}{q'}, \dots, \frac{k_q-k_{q-1}}{q'})$, where q' is a positive integer, $k_1 \leq k_2 \leq \cdots \leq k_q = q'$ are non-negative integers. Then for all $J \in \mathrm{Sym}_q$ there exists a $J' \in \mathrm{Sym}_{q'}$, such that for $\mathbf{b} = (1/q', \dots, 1/q') \in \mathrm{Pd}_{q'}$ and every graphon W it holds that*

$$\mathcal{E}_{\mathbf{a}}(W, J) = \mathcal{E}_{\mathbf{b}}(W, J').$$

Proof. Set $k_0 = 0$. If the ith component of $\mathbf{a}$ is 0, then erase this component from $\mathbf{a}$, and also erase the ith row and column of J. This transformation clearly will have no effect on the value of the GSE. Let us define the $q' \times q'$ matrix J' by blowing up rows and columns of J in the following way. For each $u, v \in [q']$ let $J'_{uv} = J_{ij}$, where $k_{i-1} < u \leq k_i$ and $k_{j-1} < v \leq k_j$. The matrix J' defined this way is clearly symmetric.

Now we will show that for every fractional q-partition with distribution $\mathbf{a}$ there exists a fractional q'-partition ρ' with distribution $\mathbf{b}$, and vice versa, such that $\mathcal{E}_\rho(W, J) = \mathcal{E}_{\rho'}(W, J')$. On one hand, for $1 \leq u \leq q'$ let $\rho'_u = \frac{\rho_i}{k_i - k_{i-1}}$, where $k_{i-1} < u \leq k_i$. Then

$$\begin{aligned}
\mathcal{E}_{\rho'}(W, J') &= -\sum_{u,v=1}^{q'} J'_{u,v} \int_{[0,1]^2} \rho'_u(x)\rho'_v(y) W(x,y) \mathrm{d}x\mathrm{d}y \\
&= -\sum_{i,j=1}^{q} J_{i,j} \sum_{l=k_{i-1}+1}^{k_i} \sum_{h=k_{j-1}+1}^{k_j} \int_{[0,1]^2} \rho'_l(x)\rho'_h(y) W(x,y) \mathrm{d}x\mathrm{d}y \\
&= \mathcal{E}_\rho(W, J).
\end{aligned}$$

On the other hand, for $1 \leq i \leq q$ let $\rho_i = \sum_{l=k_{i-1}+1}^{ki} \rho'_l$. Then

$$\begin{aligned}
\mathcal{E}_\rho(W, J) &= -\sum_{i,j=1}^{q} J_{i,j} \int_{[0,1]^2} \rho_i(x)\rho_j(y) W(x,y) \mathrm{d}x\mathrm{d}y \\
&= -\sum_{i,j=1}^{q} J_{i,j} \sum_{l=k_{i-1}+1}^{k_i} \sum_{h=k_{j-1}+1}^{k_j} \int_{[0,1]^2} \rho'_l(x)\rho'_h(y) W(x,y) \mathrm{d}x\mathrm{d}y \\
&= -\sum_{u,v=1}^{q'} J'_{u,v} \int_{[0,1]^2} \rho'_u(x)\rho'_v(y) W(x,y) \mathrm{d}x\mathrm{d}y = \mathcal{E}_{\rho'}(W, J').
\end{aligned}$$

So we conclude that

$$\mathcal{E}_{\mathbf{a}}(W,J) = \inf_{\rho\in\omega_{\mathbf{a}}} \mathcal{E}_{\rho}(W,J) = \inf_{\rho'\in\omega_{\mathbf{b}}} \mathcal{E}_{\rho'}(W,J') = \mathcal{E}_{\mathbf{b}}(W,J').$$

□

Recall Lemma 4.3.4 that states that the MGSE with fixed parameters W, J are close, whenever their corresponding probability distribution parameters are close to each other. We need here only the following special case.

Lemma 4.5.3. *Let $q \geq 1$, $J \in \mathrm{Sym}_q$, and W be an arbitrary graphon. Then for $\mathbf{a}, \mathbf{b} \in \mathrm{Pd}_q$ we have that*

$$|\mathcal{E}_{\mathbf{a}}(W,J) - \mathcal{E}_{\mathbf{b}}(W,J)| < 2\|\mathbf{a}-\mathbf{b}\|_1\|W\|_\infty\|J\|_\infty.$$

With the aid of the two previous lemmas we are able to prove the main assertion of the section. In the statement of the following theorem the LTGSE expression $\mathcal{E}^{1/q}(W,J)$ (which is equal to $\mathcal{E}_{\mathbf{b}}(W,J)$, with $\mathbf{b} = (1/q,\dots,1/q)$) appears, the notion will further be generalized in what follows later on.

Theorem 4.5.4. *Let I be a bounded interval, and $(W_n)_{n\geq 1}$ a sequence of graphons from Ξ_I^2. If for all $q \geq 1$ and $J \in \mathrm{Sym}_q$ the sequences $(\mathcal{E}^{1/q}(W_n,J))_{n\geq 1}$ converge, then for all $q \geq 1$, $\mathbf{a} \in \mathrm{Pd}_q$ and $J \in \mathrm{Sym}_q$ the sequences $(\mathcal{E}_{\mathbf{a}}(W_n,J))_{n\geq 1}$ converge.*

Proof. Let $q \geq 1$, $\mathbf{a} \in \mathrm{Pd}_q$ and $J \in \mathrm{Sym}_q$ be arbitrary and fixed. We will prove that whenever the conditions of the theorem are satisfied, then $(\mathcal{E}_{\mathbf{a}}(W_n,J))_{n\geq 1}$ is Cauchy convergent. Fix an arbitrary $\varepsilon > 0$. Let q' be such that $4\frac{q}{q'}\|I\|_\infty\|J\|_\infty < \frac{\varepsilon}{3}$, and let $\mathbf{b} \in \mathrm{Pd}_q$ be such that $b_i = [a_i/q']$ $(i = 1,\dots,q-1)$, $b_q = 1 - \sum_{i=1}^{q-1} b_i$ (where $[x]$ is the lower integer part x). Then

$$\|\mathbf{a}-\mathbf{b}\|_1 = \sum_{i=1}^{q} |a_i - b_i| \leq 2\frac{q-1}{q'} < 2\frac{q}{q'}.$$

$\mathbf{b}$ is a q'-rational distribution, so by Lemma 4.5.2 there exists $J' \in \mathrm{Sym}_{q'}$, such that for all $n \geq 1$

$$\mathcal{E}_{\mathbf{b}}(W_n,J) = \mathcal{E}^{1/q'}(W_n,J').$$

It follows from the conditions of the theorem that there exists $n_0 \in \mathbb{N}$ such that for all $m,n \geq n_0$ it is true that $\left|\mathcal{E}^{1/q'}(W_n,J') - \mathcal{E}^{1/q'}(W_m,J')\right| < \frac{\varepsilon}{3}$. Applying Lemma 4.5.3 to all $m,n \geq n_0$ we get that

$$\begin{aligned}
|\mathcal{E}_{\mathbf{a}}(W_n,J) - \mathcal{E}_{\mathbf{a}}(W_m,J)| &\leq |\mathcal{E}_{\mathbf{a}}(W_n,J) - \mathcal{E}_{\mathbf{b}}(W_n,J)| \\
&\quad + |\mathcal{E}_{\mathbf{b}}(W_n,J) - \mathcal{E}_{\mathbf{b}}(W_m,J)| + |\mathcal{E}_{\mathbf{b}}(W_m,J) - \mathcal{E}_{\mathbf{a}}(W_m,J)| \\
&\leq 2\|\mathbf{a}-\mathbf{b}\|_1\|I\|_\infty\|J\|_\infty + \left|\mathcal{E}^{1/q'}(W_n,J') - \mathcal{E}^{1/q'}(W_m,J')\right| \\
&\quad + 2\|\mathbf{a}-\mathbf{b}\|_1\|I\|_\infty\|J\|_\infty \\
&\leq \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon.
\end{aligned}$$

□

We remark that Theorem 4.5.4 also appears in [32] as Corollary 7.4, but its proof follows a different line of thought in the present work.

4.5.2 Convergence hierarchies

In various cases of testing, for certain cuts of graphs neither the notion of ground state energies, nor the notion of microcanonical ground state energies are satisfactory. For example when investigating clusteredness of a graph in a certain sense these notions become useless, because the partition for which energies attain the minimal value are trivial partitions. On the other hand, in several applications one only asks for a lower bound on the size of these classes to keep a grade of freedom of the ground state case and at the same time achieve a certain balance with respect to the sizes of classes. This setting can be regarded as an intermediate energy notion that manages to get rid of values corresponding to trivial partitions. Recall Definition 4.5.1 of the lower threshold ground state energies.

The next theorem will deliver an upper bound on the difference of the MGSEs of G and W_G for fixed $\mathbf{a}$ an J, W_G is the graphon constructed form the adjacency matrix of G in the natural way. A straightforward consequence of this will be the analogous statement for the LTGSEs.

Theorem 4.5.5. *[32] Let G be a weighted graph, $q \geq 1$, $\mathbf{a} \in \mathrm{Pd}_q$ and $J \in \mathrm{Sym}_q$. Then*

$$\left|\hat{\mathcal{E}}_{\mathbf{a}}(G, J) - \mathcal{E}_{\mathbf{a}}(W_G, J)\right| \leq 6q^3 \frac{\alpha_{\max}(G)}{\alpha_G} \beta_{\max}(G) \|J\|_\infty.$$

Since the upper bound in the theorem for a given q is not dependent on $\mathbf{a}$, it is easily possible to apply it to the LTGSEs.

Corollary 4.5.6. *Let G be a weighted graph, $q \geq 1$, $0 \leq c \leq 1/q$ and $J \in \mathrm{Sym}_q$. Then*

$$\left|\hat{\mathcal{E}}^c(G, J) - \mathcal{E}^c(W_G, J)\right| \leq 6q^3 \frac{\alpha_{\max}(G)}{\alpha_G} \beta_{\max}(G) \|J\|_\infty.$$

Based on the preceding facts we are able to perform analysis on the LTGSEs the same way as the authors of [32] did in the case of MGSE.

Corollary 4.5.7. *Let G_n be a sequence of weighted graphs with uniformly bounded edge weights. Then if $\frac{\alpha_{\max}(G_n)}{\alpha_{G_n}} \to 0$ $(n \to \infty)$, then for all $q \geq 1$, $0 \leq c \leq 1/q$ and $J \in \mathrm{Sym}_q$ the sequences $(\hat{\mathcal{E}}^c(G_n, J))_{n \geq 1}$ converge if, and only if $(\mathcal{E}^c(W_{G_n}, J))_{n \geq 1}$ converge, and then*

$$\lim_{n \to \infty} \hat{\mathcal{E}}^c(G_n, J) = \lim_{n \to \infty} \mathcal{E}^c(W_{G_n}, J).$$

Recall the definition of testability, Definition 3.2.1. It was shown in [30], among presenting other characterizations, that the testability of a graph parameter f is equivalent to the existence of a $\delta_\square$-continuous extension $\hat{f}$ of f to the space Ξ_I^2, where extension here means that $f(G_n) - \hat{f}(W_{G_n}) \to 0$ whenever $|V(G_n)| \to \infty$ (see Theorem 3.4.1, originally from [30]). Using this we are able to present yet another consequence of Theorem 4.5.5, that was verified earlier using a different approach in [28] (see also [27], Chapter 4).

Corollary 4.5.8. *For all $q \geq 1$, $0 \leq c \leq 1/q$ and $J \in \mathrm{Sym}_q$ the simple graph parameter $f(G) = \hat{\mathcal{E}}^c(G, J)$ is testable. Choosing J appropriately, $f(G)$ can be regarded as a type of balanced multiway minimal cut in [28].*

Proof. Let $q \geq 1$, $0 \leq c \leq 1/q$ and $J \in \mathrm{Sym}_q$ be fixed, and we define $\hat{f}(W) = \mathcal{E}^c(W, J)$. It follows from Corollary 4.5.6 that $f(G_n) - \hat{f}(W_{G_n}) \to 0$ whenever $|V(G_n)| \to \infty$. It remains to show that $\hat{f}$ is $\delta_\square$-continuous. To elaborate on this issue, let $U, W \in \Xi_I^2$ and ϕ be a measure-preserving permutation of $[0,1]$ such that $\delta_\square(U, W) = \|U - W^\phi\|_\square$, and let $\rho = (\rho_1, \dots, \rho_q)$ be an arbitrary fractional partition. Then

$$\begin{aligned} \left|\mathcal{E}_\rho(U, J) - \mathcal{E}_\rho(W^\phi, J)\right| &\leq \sum_{i,j=1}^{q} \left|J_{ij}\right| \left|\int_{[0,1]^2} (U - W^\phi)(x, y)\rho_i(x)\rho_j(y)\mathrm{d}x\mathrm{d}y\right| \\ &\leq q^2\|J\|_\infty\|U - W^\phi\|_\square = q^2\|J\|_\infty\delta_\square(U, W). \end{aligned} \tag{4.34}$$

This implies our claim, as $\mathcal{E}_{\mathbf{a}}(W, J) = \mathcal{E}_{\mathbf{a}}(W^\phi, J)$ for any $\mathbf{a} \in \mathrm{Pd}_q$ and ϕ measure preserving permutation, and the fact that the right-hand side of (4.34) does not depend on $\mathbf{a}$, and that by definition $\mathcal{E}^c(W, J) = \inf_{\mathbf{a} \in A_c} \mathcal{E}_{\mathbf{a}}(W, J)$.

□

In order to analyze the convergence relationship of LTGSEs with different thresholds for a given graph sequence it is sensible to consider c as a function of q. We restrict our attention to lower threshold functions c with $c(q)q$ being constant, which means that in the case of graphons the total size of the thresholds stays the same relative to the size of the interval $[0,1]$ (in the case of graphs relative to the cardinality of the vertex set). The main statement of the current section informally asserts that the convergence of LTGSEs with larger lower threshold imply convergence of all LTGSEs with smaller ones. By the results of the previous section we know that in the case of $c(q) = 1/q$ the convergence of these LTGSEs is equivalent convergence of the MGSEs for all probability distributions, and by this, according to [32], to left convergence of graphs. Moreover, in the case of $c(q) = 0$ it is equivalent to the convergence of the unrestricted GSEs, that property is known to be strictly weaker than left convergence. For technical purposes we introduce *general* LTGSEs and will refer to the previously presented notion in all that follows as *homogeneous* LTGSEs.

Definition 4.5.9. *Let $q \geq 1$, $\mathbf{x} = (x_1, \dots, x_q)$, $x_1, \dots x_q \geq 0$ and $\sum_{i=1}^{q} x_i \leq 1$, and let $A_{\mathbf{x}} = \{\, \mathbf{a} \in \mathrm{Pd}_q \mid \; a_i \geq x_i, i = 1, \dots, q \,\}$. For a graphon W and $J \in \mathrm{Sym}_q$ we call the following expression* *the lower threshold ground state energy corresponding to* $\mathbf{x}$*:*

$$\mathcal{E}^{\mathbf{x}}(W, J) = \inf_{\mathbf{a} \in A_{\mathbf{x}}} \mathcal{E}_{\mathbf{a}}(W, J).$$

The definition of $\hat{\mathcal{E}}^{\mathbf{x}}(G, J)$ *for graphs is analogous.*

Similarly to Lemma 4.5.2, the convergence of homogeneous LTGSEs is equivalent to the convergence of certain general LTGSEs.

Lemma 4.5.10. *Let* I *be a bounded interval,* $(W_n)_{n\geq 1}$ *a sequence of graphons in* Ξ_I^2*. Let* c *be a lower threshold function, so that* $c(q)q = h$ *for all* $q \geq 1$ *and for some* $0 \leq h \leq 1$*. If for all* $q \geq 1$ *and* $J \in \mathrm{Sym}_q$ *the sequences* $(\mathcal{E}^{c(q)}(W_n, J))_{n\geq 1}$ *converge, then for all* $q \geq 1$*, all*

$$\mathbf{x} = (x_1, \dots, x_q) \quad x_1, \dots, x_q \geq 0 \qquad \sum_{i=1}^{q} x_i = h, \tag{4.35}$$

and $J \in \mathrm{Sym}_q$*, the sequences* $(\mathcal{E}^{\mathbf{x}}(W_n, J))_{n\geq 1}$ *also converge.*

Proof. Fix an arbitrary graphon W from Ξ_I^2, $q \geq 1$ and $J \in \mathrm{Sym}_q$, and an arbitrary vector $\mathbf{x}$ that satisfies condition (4.35). Select for each of these vectors $\mathbf{x}$ a positive vector $\mathbf{x}'$ that obeys the condition (4.35), and that has components which are integer multiples of $c(q')$ (q' will be chosen later), so that

$$\|\mathbf{x} - \mathbf{x}'\|_1 \leq 2qc(q') = 2h\frac{q}{q'}.$$

The sets $A_{\mathbf{x}}$ and $A_{\mathbf{x}'}$ have Hausdorff distance in the L^1-norm at most $\|\mathbf{x} - \mathbf{x}'\|_1$, in particular for every $\mathbf{a} \in A_{\mathbf{x}}$ there exists a $\mathbf{b} \in A_{\mathbf{x}'}$, such that $\|\mathbf{a} - \mathbf{b}\|_1 \leq \|\mathbf{x} - \mathbf{x}'\|_1$, and vice versa. Let $\varepsilon > 0$ be arbitrary, and $\mathbf{a} \in A_{\mathbf{x}}$ be such that $\mathcal{E}^{\mathbf{x}}(W, J) + \varepsilon > \mathcal{E}_{\mathbf{a}}(W, J)$ holds. Then by applying Lemma 4.5.3 we have that

$$\begin{aligned}
\mathcal{E}^{\mathbf{x}'}(W, J) - \mathcal{E}^{\mathbf{x}}(W, J) &< \mathcal{E}^{\mathbf{x}'}(W, J) - \mathcal{E}_{\mathbf{a}}(W, J) + \varepsilon \\
&\leq \mathcal{E}_{\mathbf{b}}(W, J) - \mathcal{E}_{\mathbf{a}}(W, J) + \varepsilon \\
&\leq 2\|\mathbf{a} - \mathbf{b}\|_1 \|W\|_\infty \|J\|_\infty + \varepsilon \\
&\leq 2\|\mathbf{x} - \mathbf{x}'\|_1 \|W\|_\infty \|J\|_\infty + \varepsilon.
\end{aligned}$$

The lower bound of the difference can be handled similarly, and therefore by the arbitrary choice of ε it holds that

$$\left|\mathcal{E}^{\mathbf{x}'}(W, J) - \mathcal{E}^{\mathbf{x}}(W, J)\right| \leq 2\|\mathbf{x} - \mathbf{x}'\|_1 \|W\|_\infty \|J\|_\infty \leq 4h\frac{q}{q'}\|I\|_\infty \|J\|_\infty.$$

With completely analogous line of thought to the proof of Lemma 4.5.2, one can show that there exists a $J' \in \mathrm{Sym}_{q'}$ such that $\mathcal{E}^{\mathbf{x}'}(W, J) = \mathcal{E}^{c(q')}(W, J')$. Finally, choose q' small enough in order to satisfy $4h\frac{q}{q'}\|I\|_\infty \|J\|_\infty < \frac{\varepsilon}{3}$, and $n_0 > 0$ large enough, so that for all

$m, n \geq n_0$ the relation

$$\left|\mathcal{E}^{c(q')}(W_n, J') - \mathcal{E}^{c(q')}(W_m, J')\right| < \frac{\varepsilon}{3}$$

holds.

Then for all $m, n \geq n_0$:

$$\begin{aligned}|\mathcal{E}^{\mathbf{x}}(W_n, J) - \mathcal{E}^{\mathbf{x}}(W_m, J)| &< \left|\mathcal{E}^{\mathbf{x}}(W_n, J) - \mathcal{E}^{\mathbf{x}'}(W_n, J)\right| \\ &+ \left|\mathcal{E}^{c(q')}(W_n, J') - \mathcal{E}^{c(q')}(W_m, J')\right| + \left|\mathcal{E}^{\mathbf{x}'}(W_m, J) - \mathcal{E}^{\mathbf{x}}(W_m, J)\right| \\ &< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon.\end{aligned}$$

We did not only prove the statement of the lemma, but we also showed that the convergence is uniform in the sense that n_0 does not depend on $\mathbf{x}$ for fixed q and J. □

With the aid of the former lemma we can now prove that if all homogeneous LTGSEs with large thresholds converge, then all homogeneous LTGSEs with smaller ones also converge.

Theorem 4.5.11. *Let I be a bounded interval, $(W_n)_{n\geq 1}$ a sequence of graphons in Ξ_I^2. Let c_1, c_2 be two lower threshold functions, so that $c_1(q)q = h_1 < h_2 = c_2(q)q$ for all $q \geq 1$ for some $0 \leq h_1, h_2 \leq 1$. If for every $q \geq 1$ and $J \in \mathrm{Sym}_q$ the sequences $(\mathcal{E}^{c_2(q)}(W_n, J))_{n\geq 1}$ converge, then for every $q \geq 1$ and $J \in \mathrm{Sym}_q$ the sequences $(\mathcal{E}^{c_1(q)}(W_n, J))_{n\geq 1}$ also converge.*

Proof. From Lemma 4.5.10 it follows that if the conditions of the theorem are satisfied then for every $q \geq 1$, every

$$\mathbf{x} = (x_1, \ldots, x_q) \quad x_1, \ldots x_q \geq 0 \qquad \sum_{i=1}^{q} x_i = h_2, \tag{4.36}$$

and $J \in \mathrm{Sym}_q$ the sequences $(\mathcal{E}^{\mathbf{x}}(W_n, J))$ converge, for fixed q and J uniformly in $\mathbf{x}$.

Fix q. Our aim is to find for all $\mathbf{a} \in A_{c_1(q)}$ an $\mathbf{x}$, so that the condition (4.36) is satisfied, $\mathbf{a} \in A_{\mathbf{x}}$ and $A_{\mathbf{x}} \subseteq A_{c_1(q)}$, where $c_1(q) \leq x_i \leq a_i$ for $i = 1, \ldots, q$. As $h_1 < h_2 \leq 1$, there exists such an $\mathbf{x}$ for all $\mathbf{a} \in A_{c_1(q)}$, let us denote it by $\mathbf{x_a}$, for convenience set $(\mathbf{x_a})_i = \frac{h_1}{q} + \frac{a_i - h_1}{1-h_1}(h_2 - h_1)$. According to this correspondence we have $A_{c_1(q)} = \bigcup_{\mathbf{a}\in A_{c_1(q)}} A_{\mathbf{x_a}}$. So for an arbitrary graphon W and $J \in \mathrm{Sym}_q$ we have

$$\mathcal{E}^{c_1(q)}(W, J) = \inf_{\mathbf{a}\in A_{c_1(q)}} \mathcal{E}^{\mathbf{x_a}}(W, J).$$

We fix $\varepsilon > 0$, $J \in \mathrm{Sym}_q$, and apply Lemma 4.5.10 for the case that the conditions of the theorem are satisfied. Then there exists a $n_0 \in \mathbb{N}$, so that for all $n, m > n_0$, for all $\mathbf{x}$ which satisfies (4.36), and implies

$$|\mathcal{E}^{\mathbf{x}}(W_n, J) - \mathcal{E}^{\mathbf{x}}(W_m, J)| < \varepsilon.$$

Let $\varepsilon' > 0$ be arbitrary and $\mathbf{b} \in A_{c_1(q)}$ such that $\mathcal{E}^{c_1(q)}(W_m, J) + \varepsilon' > \mathcal{E}^{\mathbf{x}_\mathbf{b}}(W_m, J)$. Then

$$\begin{aligned}\mathcal{E}^{c_1(q)}(W_n, J) - \mathcal{E}^{c_1(q)}(W_m, J) &< \mathcal{E}^{c_1(q)}(W_n, J) - \mathcal{E}^{\mathbf{x}_\mathbf{b}}(W_m, J) + \varepsilon' \\ &\leq \mathcal{E}^{\mathbf{x}_\mathbf{b}}(W_n, J) - \mathcal{E}^{\mathbf{x}_\mathbf{b}}(W_m, J) + \varepsilon' < \varepsilon + \varepsilon'.\end{aligned}$$

The lower bound of $\mathcal{E}^{c_1(q)}(W_n, J) - \mathcal{E}^{c_1(q)}(W_m, J)$ can be established completely similarly and as ε' was arbitrary, it follows that

$$\left|\mathcal{E}^{c_1(q)}(W_n, J) - \mathcal{E}^{c_1(q)}(W_m, J)\right| < \varepsilon,$$

which verifies the statement of the theorem. □

A direct consequence is the version of Theorem 4.5.11 for weighted graphs.

Corollary 4.5.12. *Let G_n be a sequence of weighted graphs with uniformly bounded edge weights, and $\frac{\alpha_{\max}(G_n)}{\alpha_{G_n}} \to 0$ $(n \to \infty)$. Let c_1 and c_2 be two lower threshold functions, so that $c_1(q)q = h_1 < h_2 = c_2(q)q$ for all $q \geq 1$ for some $0 \leq h_1, h_2 \leq 1$. If for every $q \geq 1$ and $J \in \mathrm{Sym}_q$ the sequences $(\hat{\mathcal{E}}^{c_2(q)}(G_n, J))_{n\geq 1}$ converge, then for every $q \geq 1$ and $J \in \mathrm{Sym}_q$ the sequences $(\hat{\mathcal{E}}^{c_1(q)}(G_n, J))_{n\geq 1}$ also converge.*

The proof of Corollary 4.5.12 can be easily given through the combination of the results of Theorem 4.5.5 and Theorem 4.5.11.

Concluding this subsection we would like to mention a natural variant of the LTGSEs, the upper threshold ground state energies (UTGSE). Here we will only give an informal description of the definition and the results and leave the details to the reader, everything carries through analogously to the above. The homogeneous UTGSE, denoted by $\hat{\mathcal{E}}^{c\uparrow}(G, J)$, is determined by a formula similar to (4.32) with the set A_c replaced by A^c, that is the set of probability distributions whose components are at most c, the general variant of the UTGSE is defined in the same manner. The equivalence corresponding to the one stated in Lemma 4.5.10 between the general and the homogeneous version's convergence follows by the same blow-up trick as there, here for $c(q)q = h \geq 1$. The counterpart of Theorem 4.5.11 also holds true in the following form for $1 \leq c_2(q)q \leq c_1(q)q \leq q$: If for every $q \geq 1$ and $J \in \mathrm{Sym}_q$ the sequences $(\mathcal{E}^{c_2(q)\uparrow}(W_n, J))_{n\geq 1}$ converge, then for every $q \geq 1$ and $J \in \mathrm{Sym}_q$ the sequences $(\mathcal{E}^{c_1(q)\uparrow}(W_n, J))_{n\geq 1}$ also converge. This conclusion comes not unexpected, it says, as in the LTGSE case, that less restriction on the set A^c weakens the convergence property of a graph sequence.

4.5.3 Counterexamples

In this subsection we provide an example of a graphon family whose elements can be distinguished for a larger $c_2(q)$ lower threshold function for some pair of $q_0 \geq 1$ and $J_0 \in \mathrm{Sym}_q$ by looking at $\mathcal{E}^{c_2(q_0)}(W, J_0)$, but whose LTGSEs are identical for some smaller $c_1(q)$ lower threshold function for all $q \geq 1$ and $J \in \mathrm{Sym}_q$. Based on this it is possible to construct a sequence of graphs, whose $c_1(q)$-LTGESs converge for every $q \geq 1$ and $J \in \mathrm{Sym}_q$, but not the $c_2(q)$-LTGSEs through the same randomized method presented in [32] to show a non-convergent graph sequence with convergent ground state energies.

In the second part of the subsection we demonstrate that there exist a family of graphons, where elements can be distinguished from each other by looking only at their LTGSEs for an arbitrary small, but positive $c(q)$ lower threshold function, but whose corresponding GSEs without any threshold are identical.

Example 4.5.13. An example which can be treated relatively easily are block-diagonal graphons which are defined for the parameters $0 \leq \alpha \leq 1$, $0 \leq \beta_1, \beta_2$ as

$$W(x,y) = \begin{cases} \beta_1 & \text{, if } 0 \leq x, y \leq \alpha \\ \beta_2 & \text{, if } \alpha < x, y \leq 1 \\ 0 & \text{, else.} \end{cases}$$

In the case of $c(q)q = h$, $1 - \alpha \geq h$, $\beta_2 = 0$, for arbitrary $q \geq 1$ and $J \in \mathrm{Sym}_q$ we have $\mathcal{E}(W, J) = \mathcal{E}^{c(q)}(W, J)$. Choosing $\beta_1 = \frac{1}{\alpha^2}$, we get a one parameter family of graphons which have identical $c(q)$-LTGSEs parametrized by α with $0 < \alpha \leq 1 - h$. This means that $\mathcal{E}^{c(q)}(W(\alpha), J) = \mathcal{E}(I, J)$, where I stands for the constant 1 graphon.

For every $\alpha_0 > 1 - h$ there are $q \geq 1$ and $J \in \mathrm{Sym}_q$, so that the former equality does not hold anymore. Let $J_q \in \mathrm{Sym}_q$ be the $q \times q$ matrix, whose diagonal entries are 0, all other entries being -1 (this is the q-partition minimal cut problem). Then $\mathcal{E}(I, J_q) = 0$ for all $q \geq 1$, but for q_0 large enough $\mathcal{E}^{c(q_0)}(W(\alpha_0), J_{q_0}) > 0$, we leave the details to the reader.

With the aid of the previous example it is possible to construct a sequence of graphs which verify that in Theorem 4.5.11 the implication of the convergence property of the sequence is strictly one-way. This example is degenerate in the sense that the graphs consist of a quasi-random part and a sub-dense part with the bipartite graph spanned between the two parts also being sub-dense.

Example 4.5.14. Let us consider block-diagonal graphons with $0 < \alpha < 1$, $\beta_1, \beta_2 > 0$. It was shown in [32] that if we restrict our attention to a subfamily of block-diagonal graphons, where $\alpha^2\beta_1 + (1-\alpha)^2\beta_2$ is constant, then in these subfamilies the corresponding GSEs are identical. Let $c(q)$ be an arbitrarily small positive threshold function. Next we will show that the $c(q)$-LTGSEs determine the parameters of the block-diagonal graphon at least for a one-parameter family (up to graphon equivalence, since $(\alpha, \beta_1, \beta_2)$ belongs to the same equivalence class as $(1-\alpha, \beta_2, \beta_1)$). The constant δ_{ij} is 1, when $i = j$, and 0 otherwise.

The value of the expression $\alpha^2\beta_1 + (1-\alpha)^2\beta_2$ is determined by the MAX-CUT problem by $\mathcal{E}(W, J)$ with $q = 2$ and $J_{ij} = 1 - \delta_{ij}$.

In the second step let q_0 be as large so that $c(q_0) < \min(\alpha, 1 - \alpha)$ holds, and let J be the $q_0 \times q_0$ matrix with entries $J_{ij} = -\delta_{i1}\delta_{j1}$. In this case $\mathcal{E}(W, J) = 0$, but simple calculus gives $-\mathcal{E}^{c(q_0)}(W, J) = -\frac{\beta_1\beta_2}{\beta_1+\beta_2}c(q_0)^2$. Hence $\frac{1}{\beta_1} + \frac{1}{\beta_2}$ is determined by the LTGSEs.

The extraction of a third dependency of the parameters from $c(q)$-LTGSEs requires little more effort, we will only sketch details here. First consider α's with $\min(\alpha, 1-\alpha) \geq$

$c(2)$. Let for $q = 2$ and $k \geq 1$ be

$$J_k = \begin{pmatrix} 1 & -k \\ -k & 2 \end{pmatrix}.$$

For every α with $\min(\alpha, 1-\alpha) \geq c(2)$ we have

$$\lim_{k\to\infty} -\mathcal{E}^{c(2)}(W, J_k) = \alpha^2\beta_1 + (1-\alpha)^2\beta_2 + \max(\alpha^2\beta_1, (1-\alpha)^2\beta_2).$$

Now apply the notion of the general lower threshold: for $q = 2$ let $c_1(n) = 2c(2)/n$ and $c_2(n) = 2c(2)(n-1)/n$ two threshold functions , let us first consider the threshold $\mathbf{x}_n = (c_1(n), c_2(n))$. If $\alpha \geq c_1(n)$ or $1 - \alpha \geq c_1(n)$, then analogously to the case of the homogeneous lower thresholds

$$\lim_{k\to\infty} -\mathcal{E}^{\mathbf{x}_n}(W, J_k) = 2\alpha^2\beta_1 + (1-\alpha)^2\beta_2 \quad \text{or} \quad \alpha^2\beta_1 + 2(1-\alpha)^2\beta_2.$$

If for example $\alpha < c_1(n)$, then it is easy to see that the LTGSE is going to infinity, because for some k_0, for all $k > k_0$:

$$-\mathcal{E}^{\mathbf{x}_n}(W, J_k) < -k(c_1(n) - \alpha)c_2(n)\min\{\beta_1, \beta_2\} + 2(\alpha^2\beta_1 + (1-\alpha)^2\beta_2).$$

So for fixed n then

$$\lim_{k\to\infty} -\mathcal{E}^{\mathbf{x}_n}(W, J_k) = -\infty.$$

To actually be able to extract the expression $\alpha^2\beta_1 + (1-\alpha)^2\beta_2 + \max(\alpha^2\beta_1, (1-\alpha)^2\beta_2)$, we only have to consider the lower threshold obtained by swapping the bounds, $\mathbf{x}'_n = (c_2(n), c_1(n))$.

Then, if $\alpha \geq c_1(n)$ or $1 - \alpha \geq c_1(n)$, we have

$$\begin{aligned}\max\{&\lim_{k\to\infty} -\mathcal{E}^{\mathbf{x}_n}(W, J_k), \lim_{k\to\infty} -\mathcal{E}^{\mathbf{x}'_n}(W, J_k)\} \\ &= \alpha^2\beta_1 + (1-\alpha)^2\beta_2 + \max\{\alpha^2\beta_1, (1-\alpha)^2\beta_2\},\end{aligned}$$

otherwise

$$\max\{\lim_{k\to\infty} -\mathcal{E}^{\mathbf{x}_n}(W, J_k), \lim_{k\to\infty} -\mathcal{E}^{\mathbf{x}'_n}(W, J_k)\} = -\infty.$$

For every α there is a minimal n_0 so that one of the conditions $\alpha \geq c_1(n)$ and $1-\alpha \geq c_1(n)$ is satisfied, and for $n < n_0$ the LTGSEs corresponding to $\mathbf{x}_n$ and $\mathbf{x}'_n$ tend to infinity when k goes to infinity. Therefore the expression $\alpha^2\beta_1 + (1-\alpha)^2\beta_2 + \max(\alpha^2\beta_1, (1-\alpha)^2\beta_2)$ is determined by $c(q)$-LTGSEs.

Consider the one-parameter block-diagonal graphon family analyzed in [32], that is $W(\alpha) = W(\alpha, \frac{1}{\alpha}, \frac{1}{1-\alpha})$, where $0 < \alpha < 1$. In this case the values of our first two expressions are constant, for every $0 < \alpha < 1$ we have

$$\frac{1}{\beta_1} + \frac{1}{\beta_2} = 1,$$

$$\alpha^2\beta_1 + (1-\alpha)^2\beta_2 = 1.$$

But by applying the third expression for $c(q)$-LTGSEs, we extract $\max\{\alpha^2\beta_1, (1-\alpha)^2\beta_2\} = \max\{\alpha, 1-\alpha\}$, which determines the graphon uniquely in this family up to equivalence.

CHAPTER 5

Limits of weighted hypergraph sequences via the ultralimit method and applications

5.1 Introduction

We have seen in the previous chapter an effective proof for the testability characteristic of ground state energies which are special graph parameters. Consider the problem of finding the density of the maximal cut in a simple graph as a particular GSE problem. This has several natural counterparts for 3-uniform hypergraphs, for example one can ask to find a 3-coloring of the vertex set of a 3-graph so that the number of trichromatic 3-edges is maximal, this optimization problem is again a GSE, and therefore testable as a 3-graph parameter by the results of Chapter 4.

Another natural generalization is for 3-graphs a setting, when we focusing on colorings of vertex pairs instead of singletons. A concrete example for such a problem is to consider each pair of vertices of a 3-graph and color them with two colors. Here we are counting the number of the "good" 3-edges of the original simple 3-graph that consist of a triples whose underlying 2-colored 2-graph (obtained from the above coloring) is not monochromatic. Again, we can formulate the corresponding optimization problem where the objective is to find the coloring that maximizes the number of "good" edges. The 3-graph parameter obtained this way is not a GSE in the sense of Definition 4.1.2.

We will show that these hypergraph parameters which will be precisely defined in Definition 5.3.1 are testable by means of the ultralimit method and the machinery developed by Elek and Szegedy [49]. To our knowledge these parameters were not previously studied in a testability context. Our original motivation was the special case to establish that the cut norm $\|.\|_{\square,r}$ in Definition 3.3.20 is approximately preserved under going to a randomly sampled subgraph. Such a result would be contributing to the progress towards presenting an analogous metric for r-graphs to $\delta_\square$ in Definition 3.3.6 for graphs that generates the same topology as the r-graph convergence of Definition 2.2.3.

From the notational perspective and the theoretical background this section slightly stands out from the rest of the thesis. First we give a brief summary of the notions that were used in [49] in order to produce a representation for the limit space of simple r-graphs. This representation entails a structural connection between ultraproduct spaces and Borel spaces, and has led to a new analytical proof method for several results for simple r-graphs such as the Regularity Lemma, the Removal Lemma, or the testability assertion about hereditary r-graph properties. Subsequently, technical results proved in [49] that are relevant here are mentioned, for more details and complete proofs we refer to the paper [49].

Recall that a sequence of simple r-graphs $(G_n)_{n\geq 1}$ is convergent if for every simple F the numerical sequences $t(F, G_n)$ converge when n tends to infinity, see Definition 2.2.3.

We proceed in the framework of non-standard analysis. As default, we will treat any sequence of graphs, not only convergent ones, the main advantage of non-standard techniques is that we are not required to pass to subsequences in order to speak of limits. This paradigm is best observable in the case of sequences of reals in the unit interval, where each infinite sequence $\{x_i\}_{i=1}^{\infty}$ has a well-defined limit object in $[0, 1]$ denoted by $\lim_{\omega} x_i$ contrary to the standard setting.

The content of the chapter connects to other parts of the thesis in multiple ways. The motivating example of generalized GSE is another step forward from Chapter 4 in understanding testable hypergraph parameters. The result regarding this question, Theorem 5.3.4, will receive later in the thesis a substantially different effective proof in Chapter 6, as Corollary 6.8.1. We also reprove the main contributions of Chapter 2, where a representation of limits of $(\mathcal{K}, r)$-graph sequences was provided by probabilistic arguments. The existence, Theorem 5.4.11, and the uniqueness, Theorem 5.4.12, assertions of the limit objects in the current chapter are consequences of a deep structural correspondence between the spaces of graph sequences, ultraproducts, and Borel-measurable functions on the unit square. This also allows for extending the Regularity Lemma, which was dealt with in Chapter 3, to the setting where no algebraic structure on the set of edge colors is required, see Theorem 5.5.2 and Corollary 5.5.5.

The outline for the main part of the current chapter is as follows. Section 5.2 introduces the basic concepts of ultralimit analysis, and at the same time develops the analogous theory to [49] for r-graphs whose edges are colored from a finite color set. These concepts are already sufficient to prove the testability of the generalized GSE in Section 5.3. We proceed to the even more general case of compact colored r-graphs in Section 5.4, here the color set is a compact Polish space, and we require more advanced techniques, than in the previous sections, however still substantially rely on the framework of [49] in combination of results in [93]. We conclude the chapter in Section 5.5 with an application of the correspondence established in Section 5.4 to prove a version of the Regularity Lemma for compact colored hypergraphs.

5.2 The ultralimit method and limits of finitely colored graphs

We start by introducing the basic notations for ultraproduct measure spaces.

Definition 5.2.1. *The set $\omega \subset 2^{\mathbb{N}}$ is a non-principal ultrafilter if*

(i) $\emptyset \notin \omega$*,*

(ii) if $A \in \omega$ *and* $A \subset B$*, then* $B \in \omega$*,*

(iii) if $A, B \in \omega$*, then* $A \cap B \in \omega$*,*

(iv) for any $A \subset \mathbb{N}$ *either* $A \in \omega$ *or* $\mathbb{N} \setminus A \in \omega$*, and*

(v) there is no $a \in \mathbb{N}$*, such that* $\omega = \{ X \mid a \in X \}$*.*

One can show using the axiom of choice that a non-principal ultrafilter exists, although it is not possible to explicitly construct it. For all what follows let us fix a non-principal ultrafilter ω, it is not important which one we choose, since the measure algebras obtained below are homeomorphic for different ultrafilters.

Let us fix a non-principal ultrafilter ω on $\mathbb{N}$, and let $X_1, X_2, \dots$ be a sequence of finite sets of increasing size. We define the infinite product set $\hat{X} = \prod_{i=1}^{\infty} X_i$ and the equivalence relation $\sim$ between elements of $\hat{X}$, so that $p \sim q$ if and only if $\{ i \mid p_i = q_i \} \in \omega$. Set $\mathbf{X} = \hat{X}/\sim$, this set is called the ultraproduct of the X_i sets, and it serves as the base set of the ultraproduct measure space. Further, let $\mathcal{P}$ denote the algebra of subsets of $\mathbf{X}$ of the form $A = [\{A_i\}_{i=1}^{\infty}]$, where $A_i \subset X_i$ for each i, and $[.]$ denotes the equivalence class under $\sim$ (for convenience, $p = [\{p_i\}_{i=1}^{\infty}] \in [\{A_i\}_{i=1}^{\infty}]$ exactly in the case when $\{ i \mid p_i \in A_i \} \in \omega$).

We define a measure on the sets belonging to $\mathcal{P}$ through the ultralimit of the counting measure on the sets X_i, that is, $\mu(A) = \lim_\omega \frac{|A_i|}{|X_i|}$, where the ultralimit of a bounded real numerical sequence $\{x_i\}_{i=1}^{\infty}$ is denoted by $x = \lim_\omega x_i$, and is defined by the property that for every $\varepsilon > 0$ we have $\{ i \mid |x - x_i| < \varepsilon \} \in \omega$. One can see that the limit exists for every bounded sequence and is unique, therefore well-defined, this is a consequence of basic properties of a non-principal ultrafilter. The set $\mathcal{N} \subset 2^{\mathbf{X}}$ of μ-null sets is the family of sets N for which there exists an infinite sequence of supersets $\{A^i\}_{i=1}^{\infty} \subset \mathcal{P}$ such that $\mu(A^i) \leq 1/i$. Finally, we define the σ-algebra $\mathcal{B} = \mathcal{B}_{\mathbf{X}}$ on $\mathbf{X}$ by the σ-algebra generated by $\mathcal{P}$ and $\mathcal{N}$, and set the measure $\mu(B) = \mu(A)$ for each $B \in \mathcal{B}$, where $A \triangle B \in \mathcal{N}$ and $A \in \mathcal{P}$. Again, everything is well-defined, see [49], so we obtain the ultraproduct measure space $(\mathbf{X}, \mathcal{B}, \mu)$.

Definition 5.2.2. *Let* $(X, \mathcal{A}, \mu)$ *be a measure space, and let* $A \sim B$ *for* $A, B \in \mathcal{A}$ *whenever* $\mu(A \triangle B) = 0$*. We define the distance* d_μ *on the measure algebra* $\mathcal{B} = \mathcal{A}/\sim$ *by* $d_\mu([A], [B]) = \mu(A \triangle B)$ *and say that the measure space* $(X, \mathcal{A}, \mu)$ *is separable if the metric space* $(\mathcal{B}, d_\mu)$ *is separable.*

Note that the measure space $(\mathbf{X}, \mathcal{B}, \mu)$ is in general non-separable, this can be seen by means of contradiction. Suppose that there exists a countable d_μ-dense set in $\mathcal{P} \subset \mathcal{B}$, then the ultraproduct $\mathbf{G}$ of random subsets G_i of X_i generated by independent uniform coin tosses for any element of X_i with success probability 1/2 has with probability 1 that $\mu(\mathbf{G} \triangle \mathbf{H}) = 1/2$ for any member $\mathbf{H}$ of the above dense set of the σ-algebra, so by our assumption on countability this holds jointly with probability 1. Since we can always choose a d_μ-dense set in $\mathcal{B}$ to be in $\mathcal{P}$ we have a contradiction.

Let $X_1, X_2, \dots$ and $Y_1, Y_2, \dots$ be two increasing sequences of finite sets with ultraproducts $\mathbf{X}$ and $\mathbf{Y}$ respectively, then it is true that the ultraproduct of the product sequence $X_1 \times Y_1, X_2 \times Y_2, \dots$ is the product $\mathbf{X} \times \mathbf{Y}$ (in the sense that the two infinite products $\prod_{i=1}^\infty X_i \times Y_i$ and $\prod_{i=1}^\infty X_i \times \prod_{i=1}^\infty Y_i$ coincide, and one can check easily that the two ways of defining the equivalence $\sim$ lead to the same space), but the σ-algebra $\mathcal{B}_{\mathbf{X}\times\mathbf{Y}}$ of the measure space can be strictly larger than the σ-algebra generated by $\mathcal{B}_\mathbf{X} \times \mathcal{B}_\mathbf{Y}$, and this is a crucial point when the aim is to construct a separable representation of the ultraproduct measure space of product sets.

Let r be some positive integer, and again $X_1, X_2, \dots$ a sequence of finite sets as above. For any $e \subset [r]$ we define the ultraproduct measure spaces $(\mathbf{X}^e, \mathcal{B}_{\mathbf{X}^e}, \mu_{\mathbf{X}^e})$, also let P_e denote the natural projection from $\mathbf{X}^{[r]}$ to $\mathbf{X}^e$. Furthermore let $\sigma(e)$ denote the sub-σ-algebra of $\mathcal{B}_{\mathbf{X}^{[r]}}$ given by $P_e^{-1}(\mathcal{B}_{\mathbf{X}^e})$, and $\sigma(e)^*$ be the sub-σ-algebra $\langle \sigma(f) \mid f \subset e, |f| < |e| \rangle$. Note that in general $\sigma(e)$ is strictly larger than $\sigma(e)^*$. This can be seen by presenting elements of $\sigma(e)$ independent of $\sigma(e)^*$ of measure strictly between 0 and 1, a generic example for this is the ultraproduct of random subsets G_i^e of X_i^e where each element is included with probability 1/2 independent from each other. The ultraproduct $\mathbf{G}^e$ has with probability 1 measure 1/2 and is independent from any element of $\sigma(e)^*$.

This property will be exploited during the characterization of the limit object as functions on a separable space, where this strict inclusion will become apparent. We denote the measure $\mu_{\mathbf{X}^e}$ simply by μ_e and the σ-algebra $\mathcal{B}_{\mathbf{X}^e}$ by $\mathcal{B}_e$.

Definition 5.2.3. *Let r be a positive integer. We call a measure preserving map $\phi\colon \mathbf{X}^{[r]} \to [0,1]^{\mathfrak{h}([r])}$ a separable realization if*

(i) for any permutation $\pi \in S_{[r]}$ of the coordinates we have for all $\mathbf{x} \in \mathbf{X}^{[r]}$ that $\pi^*(\phi(\mathbf{x})) = \phi(\pi(\mathbf{x}))$, where π^* is the permutation of the power set of $[r]$ induced by π, and

(ii) for any $e \in \mathfrak{h}([r])$ and any measurable $A \subset [0,1]$ we have that $\phi_e^{-1}(A) \in \sigma(e)$ and $\phi_e^{-1}(A)$ is independent of $\sigma(e)^*$.

Let $K_r(X)$ denote the set of r-tuples of X that have no repetition in the components. We are interested in the limiting behavior of sequences of symmetric k-partitions (or edge-k-colored r-graphs on the vertex sets $X_1, X_2, \dots$) of the sequence $K_r(X_1), K_r(X_2), \dots$, this leads us to the first, relatively easy, generalization of the main result of [49]. The convergence definition, Definition 2.2.3, in this case can be reformulated in the following general way.

Let $G_i = (G_i^1, \dots, G_i^k)$ be a symmetric partition of $K_r(X_i)$ for each $i \in \mathbb{N}$, then $(G_i)_{i=1}^\infty$ converges if for every k-colored r-graph F the numerical sequences $t(F, G_i)$ converge, as in Chapter 2. The ultralimit method enables us to handle the cases where the convergence does not hold without passing to subsequences, we describe the approach next. Let us denote the size of F by m and let $F(e)$ be the color of $e \in \binom{[m]}{r}$, then $t(F, G_i)$ can be written as the measure of a subset of X_i^m. We show this by explicitly presenting the set denoted by $T(F, G_i)$, so let

$$T(F, G_i) = \bigcap_{e \in \binom{[m]}{r}} P_e^{-1}(P_{s_e}(G_i^{F(e)})), \tag{5.1}$$

where P_e is the natural projection from $X_i^{[m]}$ to X_i^e, and P_{s_e} is a bijection going from $X_i^{[r]}$ to X_i^e induced by an arbitrary but fixed bijection s_e between e and $[r]$. We define the induced subgraph density of the ultraproduct of k-colored r-graphs formally following (5.1), if $\mathbf{G} = (\mathbf{G}^1, \dots, \mathbf{G}^k)$ is a $\mathcal{B}_{[r]}$-measurable k-partition of $\mathbf{X}^{[r]}$ and F is as above then let

$$T(F, \mathbf{G}) = \bigcap_{e \in \binom{[m]}{r}} P_e^{-1}(P_{s_e}(\mathbf{G}^{F(e)})), \tag{5.2}$$

that is a measurable subset of $\mathbf{X}^{[m]}$. Further, let $t(F, \mathbf{G}) = \mu_{[m]}(T(F, \mathbf{G}))$.

It is not difficult to see that $\frac{|T(F,G_i)|}{|V(G_i)|^m} = t(F, G_i)$. Forming the ultraproduct of a series of sets commutes with finite intersection, therefore $[\{T(F, G_i)\}_{i=1}^\infty] = T(F, [\{G_i\}_{i=1}^\infty])$ and $\lim_\omega t(F, G_i) = t(F, [\{G_i\}_{i=1}^\infty])$. Observe that all of the above notions make perfect sense and the identities hold true for directed colored r-graphs, that is, when the adjacency arrays of the G^α's are not necessarily symmetric.

We call a measurable subset of $[0,1]^{\mathfrak{h}([r])}$ an r-set graphon if it is r-symmetric, we can turn it into a proper $(\{0,1\}, r)$-graphon in the sense of Chapter 2 by generating the marginal with respect to the coordinate corresponding to $[r]$. Analogously a k-colored r-set graphon is a measurable partition of $[0,1]^{\mathfrak{h}([r])}$ into k classes invariant under coordinate permutations induced by permuting $[r]$. These objects can serve as representations of the ultraproducts of r-graph sequences in the sense that the numerical sequences of subgraph densities converge to densities defined for r-set graphons in accordance with the notation in Chapter 2, we will provide the definition next.

Definition 5.2.4. *Let F be a k-colored r-graph on m vertices, and $W = (W^1, \dots, W^k)$ be a k-colored r-set graphon. Then $T(F, W) \subset [0,1]^{\mathfrak{h}([m],r)}$ denotes the set of the symmetric maps $g\colon \mathfrak{h}([m], r) \to [0,1]$ that satisfy that for each $e \in \binom{[m]}{r}$ it holds that $(g(f))_{f \in \mathfrak{h}(e)} \in W^{F(e)}$. For the Lebesgue measure of $T(F, W)$ we write $t(F, W)$, this expression is referred to as the density of F in W.*

The reader may easily verify that the above definition of density agrees with the notions in Chapter 2, especially the formula (2.8).

We will rely on a basic, but not trivial statement from measure theory due to Maharam, and that gives a sufficient condition for the existence of the independent complement of a sub-σ-algebra that is nested in some larger σ-algebra.

Lemma 5.2.5. *Let $\mathcal{B} \subset \mathcal{A}$ two separable σ-algebras on X with probability measure μ. If for any $k \geq 1$ there exists an equiv-partition $\mathcal{S}_k \subset \mathcal{A}$ of X into k parts so that $\mathcal{S}_k$ is independent of $\mathcal{B}$, then there is a σ-algebra $C \subset \mathcal{A}$ that is independent of $\mathcal{B}$ and $\langle \mathcal{B}, C \rangle$ is dense in $\mathcal{A}$ (i.e., C is an independent complement of $\mathcal{B}$ in $\mathcal{A}$).*

For the proof we refer to [49]. One of the main technical results of [49] is the following.

Theorem 5.2.6. *[49] Let r be an arbitrary positive integer and let $\mathcal{A}$ be a separable sub-σ-algebra of $\mathcal{B}_{[r]}$. Then there exists a separable realization $\phi\colon \mathbf{X}^{[r]} \to [0,1]^{\mathfrak{h}([r])}$ such that for every $A \in \mathcal{A}$ there exists a measurable $B \subset [0,1]^{\mathfrak{h}([r])}$ such that $\mu_{[r]}(A \triangle \phi^{-1}(B)) = 0$.*

This last result means that any sequence of subsets G_i of X_i^r is representable by a subset W of $[0,1]^{\mathfrak{h}([r])}$. This representation also has the property that with the usual homomorphism density definitions of hypergraphs the ultralimit $\lim_\omega t(F, G_i)$ is equal to $t(F, W)$ for any r-uniform hypergraph F, this is yet another consequence of ϕ being measure preserving. The only thing that is needed to show this, is that the densities are measures of certain sets of $\mathbf{X}^{[m]}$, respectively $[0,1]^{\mathfrak{h}([m],r)}$, and the so-called lifting of ϕ establishes a measure preserving relationship (see [49] for details).

A lifting of a separable realization $\phi\colon \mathbf{X}^{[r]} \to [0,1]^{\mathfrak{h}([r])}$ of degree m for $m \geq r$ is a measure preserving map $\psi\colon \mathbf{X}^{[m]} \to [0,1]^{\mathfrak{h}([m],r)}$ that satisfies $p_{\mathfrak{h}([r])} \circ \psi = \phi \circ P_{[r]}$, and it is equivariant under coordinate permutations in S_m, where $p_{\mathfrak{h}([r])}$ and $P_{[r]}$ are the natural projections from $[0,1]^{\mathfrak{h}([m],r)}$ to $[0,1]^{\mathfrak{h}([r])}$, and from $\mathbf{X}^{[m]}$ to $\mathbf{X}^{[r]}$ respectively. The next lemma is central to relate the sub-r-graph densities of ultraproducts to the corresponding densities in r-set graphons.

Lemma 5.2.7. *[49] For every separable realization ϕ and integer $m \geq r$ there exists a degree m lifting ψ.*

The next statement is the colored version of the homomorphism correspondence in [49] (Lemma 3.3. in that paper).

Lemma 5.2.8. *Let ϕ be a separable realization and $W = (W^1, \dots, W^k)$ be a k-colored r-set graphon, and let $\mathbf{H} = (\mathbf{H}^1, \dots, \mathbf{H}^k)$ be a k-colored ultraproduct with $\mu_{[r]}(\mathbf{H}^\alpha \triangle \phi^{-1}(W^\alpha)) = 0$ for each $\alpha \in [k]$. Let ψ be a degree m lifting of ϕ and F be a k-colored r-graph on m vertices. Then $\mu_{[m]}(\psi^{-1}(T(F,W)) \triangle T(F,\mathbf{H})) = 0$, and consequently $t(F,W) = t(F,\mathbf{H})$ for each F.*

Proof. By definition we have that

$$T(F,\mathbf{H}) = \bigcap_{e \in \binom{[m]}{r}} P_e^{-1}(P_{s_e}(\mathbf{H}^{F(e)}))$$

and

$$T(F,W) = \bigcap_{e \in \binom{[m]}{r}} p_{\mathfrak{h}([r])}^{-1}(p_{s_e}(W^{F(e)})).$$

Due to the fact that ψ commutes with coordinate permutations from S_n and the conditions we imposed on the symmetric difference of $\mathbf{H}^\alpha$ and $\phi^{-1}(W^\alpha)$ the statement follows. □

This previous result directly implies a representation of the limit objects of convergent sequences of k-colored r-graph sequences, that is a special case of Theorem 2.2.10.

Corollary 5.2.9. *Let $\{G_i\}_{i=1}^\infty$ be a convergent sequence of k-colored r-graphs. Then there exists a k-colored r-set graphon U such that for any k-colored r-graph F it holds that* $\lim_{i\to\infty} t(F, G_i) = t(F, U)$.

We turn to describe the relationship of two r-set graphons whose F-densities coincide for each F. For this purpose we have to introduce two types of transformations and clarify their connection. Let us define the σ-algebras $\mathcal{A}_S$, $\mathcal{A}_S^*$, and $\mathcal{B}_S \subset \mathcal{L}_{[0,1]^{\mathfrak{h}([r])}}$ for each $S \subset [r]$, the σ-algebra $\mathcal{B}_S = p_S^{-1}(\mathcal{L}_{[0,1]})$, $\mathcal{A}_S$ is the generated σ-algebra $\langle \mathcal{B}_T \mid T \subset S\rangle$, and $\mathcal{A}_S^*$ is $\langle \mathcal{B}_T \mid T \subset S, T \neq S\rangle$, where $\mathcal{L}_{[0,1]^t}$ denotes the Lebesgue measurable subsets of the unit cube with the dimension given by the index.

Definition 5.2.10. *[49] We say that the measurable map $\phi\colon [0,1]^{\mathfrak{h}([r])} \to [0,1]^{\mathfrak{h}([r])}$ is structure preserving if it is measure preserving, for any $S \subset [r]$ we have $\phi^{-1}(\mathcal{A}_S) \subset \mathcal{A}_S$, for any measurable $I \subset [0,1]$ we have $\phi^{-1}(p_S^{-1}(I))$ is independent of $\mathcal{A}_S^*$, and for any $\pi \in S_r$ we have $\pi^* \circ \phi = \phi \circ \pi^*$, where π^* is the coordinate permuting action induced by π.*

Let $\mathcal{L}^{\mathfrak{h}([r])}$ denote the measure algebra of $([0,1]^{\mathfrak{h}([r])}, \mathcal{L}_{[0,1]^{\mathfrak{h}([r])}}, \lambda)$.

Definition 5.2.11. *[49] We call an injective homomorphism $\Phi\colon \mathcal{L}^{\mathfrak{h}([r])} \to \mathcal{L}^{\mathfrak{h}([r])}$ a structure preserving embedding if it is measure preserving, for any $S \subset [r]$ we have $\Phi(\mathcal{B}_S) \subset \mathcal{A}_S$, $\Phi(\mathcal{B}_S)$ is independent from $\mathcal{A}_S^*$, and for any $\pi \in S_r$ we have $\pi^* \circ \Phi = \Phi \circ \pi^*$.*

Another result from [49] sheds light on the build-up of structure preserving embeddings.

Lemma 5.2.12. *[49] Suppose that $\Phi\colon \mathcal{L}^{\mathfrak{h}([r])} \to \mathcal{L}^{\mathfrak{h}([r])}$ is a structure preserving embedding of a measure algebra into itself. Then there exists a structure preserving map $\phi\colon [0,1]^{\mathfrak{h}([r])} \to [0,1]^{\mathfrak{h}([r])}$ that represents Φ in the sense that for each $[U] \in \mathcal{L}^{\mathfrak{h}([r])}$ it holds that $\Phi([U]) = [\phi^{-1}(U)]$, where U is a representative of $[U]$.*

A random coordinate system τ is the ultraproduct function on $\mathbf{X}^{[r]}$ of the random symmetric functions $\tau_n\colon [n]^r \to [0,1]^{\mathfrak{h}([r])}$ that are for each n given by a uniform random point Z_n in $[0,1]^{\mathfrak{h}([n],r)}$ so that $(\tau_n(i_1,\dots,i_r))_e = (Z_n)_{p_e(i_1,\dots,i_r)}$. An important property of the random mapping τ_n is that for any r-set graphon and positive integer n it holds that $(\tau_n)^{-1}(U) = \mathbb{G}(n,U)$, when the random sample Z_n that governs the two objects is the same.

Lemma 5.2.13. *[49] Let U be an r-set graphon, and let $\mathbf{H} = [\{\mathbb{G}(n,U)\}_{n=1}^\infty]$. Then the random coordinate system $\tau = [\{\tau_n\}_{n=1}^\infty]$ is a separable realization such that with probability one we have $\mu_{[r]}(\mathbf{H}\triangle\tau^{-1}(U)) = 0$.*

A direct consequence is the statement for k-colored r-set graphons.

Corollary 5.2.14. *Let $U = (U^1, \dots, U^k)$ be a k-colored r-set graphon, and let $\mathbf{H} = (\mathbf{H}^1, \dots, \mathbf{H}^k)$ be a k-colored ultraproduct in $\mathbf{X}^{[r]}$, where $\mathbf{H}^\alpha = [\{\mathbb{G}(n, U^\alpha)\}_{n=1}^\infty]$ for each $\alpha \in [k]$. Then the random coordinate system τ is with probability one a separable realization such that we have $\mu_{[r]}(\mathbf{H}^\alpha \triangle \tau^{-1}(U^\alpha)) = 0$ for each $\alpha \in [k]$.*

The following result is a generalization of the uniqueness assertion of [49], and states that subgraph densities determine an r-set graphon up to structure preserving transformations, the proof is also similar to the one given in [49] in the uncolored case. We say if the conditions of the theorem below apply for two graphons that they are equivalent.

Theorem 5.2.15. *Let $U = (U^1, \dots, U^k)$ and $V = (V^1, \dots, V^k)$ be two k-colored r-set graphons such that for each k-colored r-graph F it holds that $t(F, U) = t(F, V)$. Then there exist two structure preserving maps ν_1 and ν_2 from $[0,1]^{\mathfrak{h}([r])}$ to $[0,1]^{\mathfrak{h}([r])}$ such that $\mu_{[r]}(\nu_1^{-1}(U^\alpha) \triangle \nu_2^{-1}(V^\alpha)) = 0$ for each $\alpha \in [k]$.*

Proof. The equality $t(F, U) = t(F, V)$ for each F implies that $\mathbb{G}(n, U)$ and $\mathbb{G}(n, V)$ have the same distribution Y_n for each n. Let $\mathbf{H} = [\{Y_n\}_{n=1}^\infty]$, then Corollary 5.2.14 implies that there exist separable realizations ϕ_1 and ϕ_2 such that $\mu_{[r]}(\mathbf{H}^\alpha \triangle \phi_1^{-1}(U^\alpha)) = 0$ and $\mu_{[r]}(\mathbf{H}^\alpha \triangle \phi_2^{-1}(V^\alpha)) = 0$ for each $\alpha \in [k]$, therefore also $\mu_{[r]}(\phi_1^{-1}(U^\alpha) \triangle \phi_2^{-1}(V^\alpha)) = 0$. Set $\mathcal{A} = \sigma(\phi_1^{-1}(\mathcal{L}_{[0,1]^{\mathfrak{h}([r])}}), \phi_2^{-1}(\mathcal{L}_{[0,1]^{\mathfrak{h}([r])}}))$ that is a separable σ-algebra on $\mathbf{X}^{[r]}$ so by Theorem 5.2.6 there exists a separable realization ϕ_3 such that for each measurable $A \subset [0,1]^{\mathfrak{h}([r])}$ the element $\phi_i^{-1}(A)$ of $\mathcal{A}$ can be represented by a subset of $[0,1]^{\mathfrak{h}([r])}$ denoted by $\psi_i(A)$. It is easy to check that the maps ψ_1 and ψ_2 defined this way are structure preserving embeddings from $\mathcal{L}^{\mathfrak{h}([r])} \to \mathcal{L}^{\mathfrak{h}([r])}$ satisfying $\lambda(\psi_1(U^\alpha) \triangle \psi_2(V^\alpha)) = 0$ for each $\alpha \in [k]$. We conclude that by Lemma 5.2.12 there are structure preserving ν_1 and ν_2 such that $\lambda(\nu_1^{-1}(U^\alpha) \triangle \nu_2^{-1}(V^\alpha)) = 0$ for each $\alpha \in [k]$. □

5.3 Testability of energies by non-effective methods

We define a parameter of r-uniform hypergraphs that is a generalization of the ground state energies of [32] in the case of graphs. This notion encompasses several important quantities, therefore its testability is central to many applications.

Definition 5.3.1. *For a set $H \subset \binom{[n]}{r}$, a real r-array J of size q, and a symmetric partition $\mathcal{P} = (P^1, \dots, P^q)$ of $\binom{[n]}{r-1}$ we define the r-energy*

$$\mathcal{E}_{\mathcal{P},r}(H, J) = \frac{1}{n^r} \sum_{i_1,\dots,i_r=1}^{q} J(i_1, \dots, i_r) e_H(r; P_{i_1}, \dots, P_{i_r}),$$

where $e_H(r; S_1, \dots, S_r) = |\{ (u_1, \dots, u_r) \in [n]^r \mid A_{S_j}(u_1, \dots, u_{j-1}, u_{j+1}, \dots, u_r) = 1$ for all $j = 1, \dots, r$ and $A_H(u_1, \dots, u_r) = 1 \}|$.

Let $H = (H^\alpha)_{\alpha\in[k]}$ be a k-colored r-uniform hypergraph on the vertex set $[n]$ and J^α a be real $q\times\cdots\times q$ r-array with $\|J\|_\infty \le 1$ for each $\alpha \in [k]$. Then the energy for a partition $\mathcal{P}$ as above is

$$\mathcal{E}_{\mathcal{P},r}(\mathbf{H}, J) = \sum_{\alpha\in[k]} \mathcal{E}_{\mathcal{P},r-1}(H^\alpha, J^\alpha).$$

The maximum of the energy over all partitions $\mathcal{P}$ of $\binom{[n]}{r-1}$ is called the r-ground state energy (rGSE) of H with respect to J, and is denoted by

$$\mathcal{E}_r(\mathbf{H}, J) = \max_{\mathcal{P}} \mathcal{E}_{\mathcal{P},r}(\mathbf{H}, J).$$

The rGSE can also be defined for $(\{0,1\}, r)$-graphons, and more generally, for kernels.

Definition 5.3.2. *For an r-kernel W, a real r-array J of size q, and a symmetric partition $\mathcal{P} = (P^1, \dots, P^q)$ of $[0,1]^{\mathfrak{h}([r-1])}$ we define the energy*

$$\mathcal{E}_{\mathcal{P},r}(W, J) = \sum_{i_1,\dots,i_r\in[q]} J(i_1,\dots,i_r) \int_{\cap_{j\in[r]} p^{-1}_{[r]\setminus\{j\}}(P^{i_j})} W(x_{\mathfrak{h}([r],r-1)})\mathrm{d}\lambda(x_{\mathfrak{h}([r],r-1)}).$$

Let $W = (W^\alpha)_{\alpha\in[k]}$ be a k-colored r-graphon and J^α a be real $q\times\cdots\times q$ r-array with $\|J\|_\infty \le 1$ for each $\alpha \in [k]$. Then the energy for a partition $\mathcal{P}$ as above is

$$\mathcal{E}_{\mathcal{P},r}(W, J) = \sum_{\alpha\in[k]} \mathcal{E}_{\mathcal{P},r-1}(W^\alpha, J^\alpha).$$

and the rGSE of W with respect to J, and is denoted by

$$\mathcal{E}_r(W, J) = \sup_{\mathcal{P}} \mathcal{E}_{\mathcal{P},r-1}(W, J),$$

where the supremum runs over all symmetric partitions $\mathcal{P} = (P^1, \dots, P^q)$ of $[0,1]^{\mathfrak{h}([r-1])}$.

The generalization of free energies (recall Definition 3.6.1) corresponding to this formulation of the rGSE is straight-forward, their analysis is expected to require novel methods. One of the obstacles for applying the previously presented framework in Chapter 3 and Chapter 4 is that state configurations applied to the blow-up of a graph cannot be traced back to configurations on the original easily. For 3-graphs in the q state model of an rGSE one particular pair of nodes can be in q states, whereas in the k-fold blow up the induced complete bipartite graph between the nodes corresponding to the two originals has $\Omega(\exp(kq))$ isomorphism classes of configurations, determining the corresponding weight as a function of weights of the original graph is an incomparably harder task than in the case discussed in Chapter 3.

Definitions of the above energies are analogous in the directed, and the weighted case, and also for r-kernels. The next lemma tells us about the distribution of the rGSE when taking a random sample $\mathbb{G}(n, H)$ of an $H \in \Pi^{r,k}$.

Lemma 5.3.3. *The expression $\mathcal{E}_r(\mathbb{G}(n,H),J)$ is highly concentrated around its mean, that is for every $\varepsilon > 0$ it holds that*

$$\mathbb{P}(|\mathcal{E}_r(\mathbb{G}(n,H),J) - \mathbb{E}\mathcal{E}_r(\mathbb{G}(n,H),J)| \geq \varepsilon\|J\|_\infty) \leq 2\exp\left(-\frac{\varepsilon^2 n}{8r^2}\right).$$

Proof. We can assume that $\|J\|_\infty \leq 1$. The random r-graph $\mathbb{G}(n,H)$ is generated by picking random nodes from $V(H)$ without repetition, let X_i denote the ith random element of $V(H)$ that has been selected. Define the martingale $Y_i = \mathbb{E}[\mathcal{E}_r(\mathbb{G}(n,H),J) \mid X_1,\dots,X_i]$ for $0 \leq i \leq n$. It has the property that $Y_0 = \mathbb{E}[\mathcal{E}_r(\mathbb{G}(n,H),J)]$ and $Y_n = \mathcal{E}_r(\mathbb{G}(n,H),J)$, whereas the jumps $|Y_i - Y_{i-1}|$ are bounded above by $\frac{2r}{n}$ for each $i \in [n]$. The last observation is the consequence of the fact that for any partition $\mathcal{P}$ of $\binom{[n]}{r-1}$ only at most rn^{r-1} terms in the sum constituting $\mathcal{E}_{\mathcal{P},r-1}(H,J)$ are affected by changing the placing of X_i in the classes of $\mathcal{P}$. Applying the Azuma-Hoeffding inequality to the martingale verifies the statement of the lemma. □

The same concentration result as above applies also to $\mathcal{E}_r(\mathbb{G}(n,W),J)$. The next result is the main contribution in the current section.

Theorem 5.3.4. *For any $J = (J^1,\dots,J^k)$ with J^α being a real r-array of size q for each $\alpha \in [k]$ the parameter of k-colored r-graphs $\mathcal{E}_r(.,J)$ is testable.*

Proof. We may assume that $\|J^\alpha\|_\infty \leq 1$ for every α without losing generality. We proceed by contradiction. Suppose there exist an $\varepsilon > 0$ and a sequence of k-colored r-uniform hypergraphs $\{H_n\}_{n=1}^\infty$ with $V(H_n) = [m_n]$, where $(m_n)_{n=1}^\infty$ is tending to infinity, whose elements are such that for each n with probability at least ε we have that $\mathcal{E}_r(H_n,J) + \varepsilon \leq \mathcal{E}_r(\mathbb{G}(n,H_n),J)$. Let $G_n = (G_n^1,\dots,G_n^k)$ denote the random k-colored hypergraph $\mathbb{G}(n,H_n)$, that is for each n with $G_n^\alpha = \mathbb{G}(n,H_n^\alpha)$. The previous event can be reformulated as stating that for each n with probability at least ε there is a partition $\mathcal{P}_n = (P_n^1,\dots,P_n^q)$ of $\binom{[n]}{r-1}$ such that the expression

$$\frac{1}{n^r}\sum_{\alpha=1}^{k}\sum_{i_1,\dots,i_r=1}^{q} J^\alpha(i_1,\dots,i_r)e_{G_n^\alpha}(r;P_n^{i_1},\dots,P_n^{i_r})$$

is larger than

$$\frac{1}{m_n^r}\sum_{\alpha=1}^{k}\sum_{i_1,\dots,i_r=1}^{q} J^\alpha(i_1,\dots,i_r)e_{H_n^\alpha}(r;R_n^{i_1},\dots,R_n^{i_r}) + \varepsilon$$

for any partition $\mathcal{R}_n = (R_n^1,\dots,R_n^q)$ of $\binom{[m_n]}{r-1}$.

Let $\mathbf{H} = (\mathbf{H}^1,\dots,\mathbf{H}^k)$ denote the ultraproduct of the hypergraph sequence $\{H_n\}_{n=1}^\infty$ that is a k-partition in the measure space $(\mathbf{X}_1^{[r]},\mathcal{B}_1,\mu_1)$, and let $\sigma_1(S)$ and $\sigma_1(S)^*$ denote the sub-σ-algebras of $\mathcal{B}_1$ corresponding to subsets S of $[r]$. Due to Theorem 5.2.6 there exists a separable realization $\phi_1 \colon \mathbf{X}_1^{[r]} \to [0,1]^{\mathfrak{h}([r])}$ such that there is a k-colored r-set graphon $\mathbf{W} = (W^1, \dots, W^k)$ satisfying $\mu_1(\phi_1^{-1}(W^\alpha) \triangle \mathbf{H}^\alpha) = 0$ for each $\alpha \in [k]$.

Let $\mathbf{G}(s)$ stand for the point-wise ultraproduct realization of the $\{G_n(s)\}_{n=1}^\infty \subset \mathbf{X}_2^{[r]}$ for all $s \in \mathbb{S}$, where $(\mathbb{S}, \mathcal{S}, \nu)$ denotes the underlying joint probability space for the random hypergraphs, and $(\mathbf{X}_2^{[r]}, \mathcal{B}_2, \mu_2)$ is the ultraproduct measure space in the case of the sample sequence of the sets $[1], [2], \dots$, $\sigma_2(S)$ and $\sigma_2(S)^*$ are the corresponding sub-σ-algebras. Note that the ultraproducts $\mathbf{G}(s)$ are not k-partitions of the same ultraproduct space as $\mathbf{H}$, moreover, it is possible that the σ-algebra generated by $\{\, \mathbf{G}(s) \mid s \in \mathbb{S} \,\}$ together with μ_2 form a non-separable measure algebra that prevents us from using Theorem 5.2.6 directly.

Suppose that for some n we have that $\mathbb{E}\mathcal{E}_r(G_n, J) < \mathcal{E}_r(H_n, J) + 3/4\varepsilon$. This assumption implies by Lemma 5.3.3 that $\mathbb{P}(\mathcal{E}_r(G_n, J) \geq \mathcal{E}_r(H_n, J) + \varepsilon) \leq \mathbb{P}(\mathcal{E}_r(G_n, J) \geq \mathbb{E}\mathcal{E}_r(G_n, J) + \varepsilon/4) \leq 2\exp(-\frac{\varepsilon^2 n}{64 r^2})$. The last bound is strictly smaller than ε when n is chosen sufficiently large, therefore it contradicts the main assumption for large n. Therefore we can argue that $\mathbb{E}\mathcal{E}_r(G_n, J) \geq \mathcal{E}_r(H_n, J) + 3/4\varepsilon$ for large n, throwing away a starting piece of the sequence $\{H_n\}_{n=1}^\infty$ we may assume that the inequality holds for all n. Note that this omission does not affect the ultraproduct spaces.

A second application of Lemma 5.3.3 leads to a lower bound on the probability that $\mathcal{E}_r(G_n, J)$ is close to $\mathcal{E}_r(H_n, J)$, namely $\mathbb{P}(\mathcal{E}_r(G_n, J) \leq \mathcal{E}_r(H_n, J) + \varepsilon/2) \leq 2\exp(-\frac{\varepsilon^2 n}{64 r^2})$. Hence, by invoking the Borel-Cantelli Lemma, we infer that with probability one the event $\mathcal{E}_r(G_n, J) \leq \mathcal{E}_r(H_n, J) + \varepsilon/2$ can occur only for finitely many n, let the M_1 denote the (random) threshold for which is true that $\mathcal{E}_r(G_n, J) > \mathcal{E}_r(H_n, J) + \varepsilon/2$ for every $n \geq M_1$. It follows that $\lim_\omega \mathcal{E}_r(G_n, J) > \lim_\omega \mathcal{E}_r(H_n, J) + \varepsilon/2$ with probability 1.

Next we will show that with probability one $\mathbf{G}$ is equivalent to $\mathbf{H}$ in the sense that for each k-colored r-graph F it holds that $t(F, \mathbf{G}) = t(F, \mathbf{H})$. Then, since there are countably many test graphs F, we can conclude that the equality holds simultaneously for all F with probability 1.

We have seen above in the paragraph after (5.2) that for every fixed k-colored r-uniform hypergraph $t(F, \mathbf{H}) = \lim_\omega t(F, H_n)$. On the other hand the subgraph densities in random induced subgraphs are highly concentrated around their mean, and the mean is equal to the corresponding density in the source graph, that is

$$\mathbb{P}(|t(F, G_n) - t(F, H_n)| \geq \delta) \leq 2\exp\left(-\frac{\delta^2 n}{2|V(F)|^2}\right)$$

for any $\delta > 0$, this follows with basic martingale techniques, see Lemma 3.5.4 originally proved in [49] for the almost identical statement together.

The Borel-Cantelli Lemma implies then for every fixed F that with probability one for each $\delta > 0$ there exists a (random) $n_0(\delta)$ such that for each $n \geq n_0(\delta)$ it is true that $|t(F, G_n) - t(F, H_n)| < \delta/2$. Let us fix $\delta > 0$ and $F \in \Pi^{r,k}$. Since the set $\{\, n \mid |t(F, H_n) - t(F, \mathbf{H})| < \delta/2 \,\}$ belongs to ω by the definition of the ultraproduct function, it holds that $\{ n \mid |t(F, G_n) - t(F, \mathbf{H})| < \delta \} \in \omega$ as a consequence of

$$\begin{aligned}\{ n \,\| t(F, G_n) - t(F, \mathbf{H})| < \delta \} \\ \supset (\{ n \mid |t(F, G_n) - t(F, H_n)| < \delta/2 \} \cap \{ n \mid |t(F, H_n) - t(F, \mathbf{H})| < \delta/2 \}) \in \omega.\end{aligned}$$

Consequently, $\lim_\omega t(F, G_n) = t(F, \mathbf{H})$ with probability one for each F, and the limit equation holds simultaneously for each F also with probability one, since their number is countable.

Let us pick a realization $\{G_n(s)\}_{n=1}^\infty$ of $\{G_n\}_{n=1}^\infty$ such that it satisfies $\lim_\omega \mathcal{E}_r(G_n(s), J) - \lim_\omega \mathcal{E}_r(H_n, J) \geq \varepsilon/2$ and $\lim_\omega t(F, G_n(s)) = t(F, \mathbf{H})$ for each F, the preceding discussion implies that such a realization exists, in fact almost all of them are like this. Furthermore, let us consider the sequence of partitions $\mathcal{P}_n = (P_n^1, \dots, P_n^q)$ of $\binom{[n]}{r-1}$ that realize $\mathcal{E}_r(G_n(s), J)$, and define $T_n^{i,j} \subset [n]^r \setminus \mathrm{diag}([n]^r)$ through the inverse images of the projections $A_{T_n^{i,j}} = (p_j^n)^{-1}(A_{P_n^i})$ for $i \in [q]$, $j \in [r]$, and $n \in \mathbb{N}$, where p_j^n is the projection that maps an r-array of size n onto an $(r-1)$-array by erasing the jth coordinate. Note that the $T_n^{i,j}$ sets are not completely r-symmetric, but are invariant under coordinate permutations from $S_{[r]\setminus\{j\}}$ for the corresponding $j \in [r]$. A further property is that and T_n^{i,j_1} can be obtained from T_n^{i,j_2} swapping the coordinates corresponding to j_1 and j_2.

We additionally define the ultraproducts of these sets by $\mathbf{P}^i = [\{P_n^i\}_{n=1}^\infty] \subset \mathbf{X}_2^{[r-1]}$ and $\mathbf{T}^{i,j} = [\{T_n^{i,j}\}_{n=1}^\infty] \subset \mathbf{X}_2^{[r]}$, it is clear that $\mathbf{T}^{i,j} \in \sigma_2([r] \setminus \{j\})$ for each pair of i and j, so $\cap_{(i,j)\in I} \mathbf{T}^{i,j} \in \sigma_2([r])^*$ for any $I \subset [q] \times [r]$, and that $\mathbf{X}_2^{[r-1]} = \cup_i \mathbf{P}^i$. The same symmetry assumptions apply for the $\mathbf{T}^{i,j}$ sets as for the $T_n^{i,j}$ sets described above.

We also require the fact that these ultraproduct sets defined above establish a correspondence between the rGSE of $\mathbf{G}(s)$ and the ultralimit of the sequence of energies $\{\mathcal{E}_r(G_n(s), J)\}_{n=1}^\infty$.

This can be seen as follows: Recall that

$$\mathcal{E}_r(G_n(s), J) = \frac{1}{n^r} \sum_{\alpha=1}^{k} \sum_{i_1,\dots,i_r=1}^{q} J^\alpha(i_1, \dots, i_r) |G_n^\alpha \cap (\cap_{j=1}^r T_n^{i_j,j})|$$

This formula together with the identities $[\{G_n^\alpha(s) \cap (\cap_{j=1}^r T_n^{i_j,j})\}_{n=1}^\infty] = \mathbf{G}^\alpha(s) \cap (\cap_{j=1}^r \mathbf{T}^{i_j,j})$, and that the ultralimit of subgraph densities equals the corresponding subgraph densities of the ultraproduct imply that

$$\lim_\omega \mathcal{E}_r(G_n(s), J) = \sum_{\alpha=1}^{k} \sum_{i_1,\dots,i_r=1}^{q} J^\alpha(i_1, \dots, i_r) \mu_2(\mathbf{G}^\alpha(s) \cap (\cap_{j=1}^r \mathbf{T}^{i_j,j})).$$

Now consider the separable sub-σ-algebra $\mathcal{A}$ of $\mathcal{B}_2$ generated by the collection of the sets $\mathbf{G}^1(s), \dots, \mathbf{G}^k(s), \mathbf{T}^{1,1}, \dots, \mathbf{T}^{q,r}$. Then by Theorem 5.2.6 there exists a separable realization $\phi_2\colon \mathbf{X}_2^{[r]} \to [0,1]^{\mathfrak{h}([r])}$ and measurable sets $U^1, \dots, U^k, V^{1,1}, \dots, V^{q,r}$ such that $\mu_2(\phi_2^{-1}(U^\alpha) \triangle \mathbf{G}^\alpha(s)) = 0$ for each $\alpha \in [k]$ and $\mu_2(\phi_2^{-1}(V^{i,j}) \triangle \mathbf{T}^{i,j}) = 0$ for every $i \in [q]$, $j \in [r]$.

Additionally, since ϕ is a separable realization, we can assume that each of the $V^{i,j}$ sets only depends on the coordinates corresponding to the sets in $\mathfrak{h}([r] \setminus \{j\})$, is invariant under coordinate permutations induced by elements of S_r that fix j, and V^{i,j_1} can be obtained from V^{i,j_2} by relabeling the coordinates according to the S_r permutation swapping j_1 and j_2. Also, we can assume that $(U^1, \ldots, U^k)$ form a k-colored r-set graphon $\mathbf{U}$. Most importantly, the separable realization ϕ_2 is measure preserving, so we have that

$$\lim_{\omega} \mathcal{E}_r(G_n(s), J) = \sum_{\alpha=1}^{k} \sum_{i_1,\ldots,i_r=1}^{q} J^{\alpha}(i_1, \ldots, i_r)\lambda(U^{\alpha} \cap (\cap_{j=1}^{r} V^{i_j,j})). \tag{5.3}$$

On the other hand we established that $t(F, \mathbf{G}(s)) = t(F, \mathbf{H})$ for each F,which implies $t(F, \mathbf{U}) = t(F, \mathbf{W})$, therefore the uniqueness statement of Theorem 5.2.15 ensures the existence of two structure preserving measurable maps $\nu_1, \nu_2 \colon [0,1]^{\mathfrak{h}([r])} \to [0,1]^{\mathfrak{h}([r])}$ such that $\lambda(\nu_1^{-1}(W^{\alpha}) \triangle \nu_2^{-1}(U^{\alpha})) = 0$ for each $\alpha \in [k]$.

Now let us define the sets $\mathbf{S}^{i,j} = \phi_1^{-1}(\nu_2(\nu_1^{-1}(V^{i,j})))$, these satisfy exactly the same symmetry and measurability properties as the $\mathbf{T}^{i,j}$ sets above, by the measure preserving nature of the maps involved we have

$$\lim_{\omega} \mathcal{E}_r(G_n(s), J) = \sum_{\alpha=1}^{k} \sum_{i_1,\ldots,i_r=1}^{q} J^{\alpha}(i_1, \ldots, i_r)\mu_1(\mathbf{H}^{\alpha} \cap (\cap_{j=1}^{r} \mathbf{S}^{i_j,j})). \tag{5.4}$$

The properties of structure preserving maps imply that $\mathbf{S}^{i,j} \in \sigma_1([r] \setminus \{j\})$ for each i, j, so $\cap_{(i,j)\in I} \mathbf{S}^{i,j} \in \sigma_1([r])^*$ for any $I \subset [q] \times [r]$. Also, the ultraproduct construction makes it possible to assert the existence of a sequence of partitions $\mathcal{R}_n = (R_n^1, \ldots, R_n^q)$ of $\binom{[m_n]}{r-1}$ for ω-almost every n such that $\mu_{[r]\setminus\{j\}}(\mathbf{S}^{i,j} \triangle [\{(p_j^{m_n})^{-1}(R_n^i)\}_{n=1}^{\infty}]) = 0$ for each i, j. But again by the correspondence principle between ultralimits of sequences and ultraproducts in Lemma 5.2.8 applied to (5.3) and (5.4) we have

$$\lim_{\omega} \mathcal{E}_{\mathcal{R}_n,r}(H_n, J) = \lim_{\omega} \mathcal{E}_r(G_n(s), J),$$

which contradicts $\lim_{\omega} \mathcal{E}_r(G_n(s), J) - \lim_{\omega} \mathcal{E}_r(H_n, J) \geq \varepsilon/2$.

□

5.4 Ultralimits of compact colored graphs

The separable correspondence enables us to switch to more tangible r-set graphons in the analysis of simple hypergraph limits instead of the abstract subsets of the ultraproduct space. This also allows for the utilization of the toolbox of classical real analysis.

Our goal next is to produce an analogous correspondence to Corollary 5.2.9 for limits of colored hypergraph limits, where the edge colors are coming from some compact

Polish space possibly without algebraic structure. For this case we had already given a proof in Chapter 2 with significantly different methods drawing on previous results in the area of exchangeable arrays. The main purpose of the new approach in the current section lies in the additional structural connection that is obtained, this serves as the main tool for establishing a regularity lemma for compact colored r-graphs in the subsequent section.

We fix $r \geq 1$, the sets $X_1, X_2, \ldots$ be as above. Let $\mathcal{K}$ be a compact Polish space (recall that this means that the space is separable and completely metrizable), and consider a sequence of colored r-graphs (or functions) $G_i \colon X_i^r \to \mathcal{K}$ that are invariant under coordinate permutations and take the special value ι on the diagonal, which is introduced for the sake of completeness. It will be more convenient for our purposes to consider these as functions that take values from the space of probability measures on $\mathcal{K}$, denoted by $\mathcal{P}(\mathcal{K})$, endowed with the weak topology. That is, let $\hat{G}_i(x_i) = \delta_{G_i(x_i)}$ for $x_i \in X_i^r$. We will sometimes omit the hat sign and identify the two functions G and $\hat{G}$. Now we are going to define the ultraproduct of $\mathcal{P}(\mathcal{K})$-valued functions, and thereby also of $\mathcal{K}$-valued functions.

We denote by $\langle \mu, f \rangle$ the expectation of a measurable functional f on $\mathcal{K}$ with respect to the probability measure $\mu \in \mathcal{P}(\mathcal{K})$, that is, $\int_{\mathcal{K}} f(t) \mathrm{d}\mu(t)$.

The ultraproduct of $\{G_i\}_{i=1}^{\infty}$ ($\mathcal{K}$-valued or $\mathcal{P}(\mathcal{K})$-valued) is the function $\mathbf{G} \colon \mathbf{X}^r \to \mathcal{P}(\mathcal{K})$ such that for $\mathbf{p} = [\{p_i\}_{i=1}^{\infty}]$ with $p_i \in X_i^r$ and any continuous non-negative functional f on $\mathcal{K}$ we have that $\langle \mathbf{G}(\mathbf{p}), f \rangle = \lim_{\omega} \langle G_i(p_i), f \rangle$ for general $\mathcal{P}(\mathcal{K})$-valued G_i functions, and $\langle \mathbf{G}(\mathbf{p}), f \rangle = \lim_{\omega} f(G_i(p_i))$ for a special $\mathcal{K}$-valued sequence.

Proposition 5.4.1. *The ultraproduct function $\mathbf{G}$ is well-defined and measurable with respect to the relevant σ-algebras.*

Proof. Let for a fixed $\mathbf{p} \in \mathbf{X}^r$ be $\lim_{\omega} \langle G_i(p_i), f \rangle = f_{\mathbf{p}}$, then $f \mapsto f_{\mathbf{p}}$ is a positive bounded linear functional on $C(\mathcal{K})$, so by the Riesz Representation Theorem on a compact space there exists a unique probability measure $\mathbf{G}(\mathbf{p})$ that satisfies $\langle \mathbf{G}(\mathbf{p}), f \rangle = \lim_{\omega} \langle G_i(p_i), f \rangle$. Also, $[\{p_i\}_{i=1}^{\infty}] = [\{q_i\}_{i=1}^{\infty}]$ then $\{ i \mid p_i = q_i \} \in \omega$, therefore for any f we have $\{ i \mid \langle G_i(p_i), f \rangle = \langle G_i(q_i), f \rangle \} \in \omega$, hence $\lim_{\omega} \langle G_i(p_i), f \rangle = \lim_{\omega} \langle G_i(q_i), f \rangle$.

We are left to check whether $\mathbf{G}$ is measurable with respect to the σ-algebras $\sigma([r])$ and the Borel sets of the weak topology on $\mathcal{P}(\mathcal{K})$. The measurable sets on $\mathcal{P}(\mathcal{K})$ are generated by the sets $\{ \mu \mid \langle \mu, g \rangle \geq 0 \}$ for some $g \in C(\mathcal{K})$ therefore it suffices to show that $\mathbf{A}_g = \{ \mathbf{p} \mid \langle \mathbf{G}(\mathbf{p}), g \rangle \geq 0 \} \in \sigma([r])$ for arbitrary $g \in C(\mathcal{K})$. Let $A_g^{i,\varepsilon} = \{ p_i \mid \langle G_i(p_i), g \rangle \geq -\varepsilon \} \subset X_i^r$ for $\varepsilon > 0$ and $i \in \mathbb{N}$, and also let $\mathbf{A}_g^{\varepsilon} = [\{A_g^{i,\varepsilon}\}_{i=1}^{\infty}] \in \sigma([r])$. We will show that $\cap_{n \geq 1} \mathbf{A}_g^{1/n} = \mathbf{A}_g$, and that implies $\mathbf{A}_g \in \sigma([r])$. It is clear that $\cap_{n \geq 1} \mathbf{A}_g^{1/n} \supset \mathbf{A}_g$. We fix a $\mathbf{p} = [\{p_i\}_{i=1}^{\infty}] \in \cap_{n \geq 1} \mathbf{A}_g^{1/n}$. Then we have $\lim_{\omega} \langle G_i(p_i), g \rangle \geq -\frac{1}{n}$ for any n, that is, for any $\varepsilon > 0$ the ultrafilter ω contains $\{ i \mid \langle G_i(p_i), g \rangle > -\frac{1}{n} - \varepsilon \}$, which implies that $\mathbf{p} \in \mathbf{A}_g$. □

Let $\mathcal{F} \subset C(\mathcal{K})$ be a countable family such that the linear subspace generated by $\mathcal{F}$ is $\|.\|_{\infty}$-dense in $C(\mathcal{K})$. Recall Definition 2.2.1 that says that for $F \in \Pi(\mathcal{F})$ with $V(F) = [m]$

and $G_i\colon X_i^r \to \mathcal{K}$, the frequency of F in G is

$$t(F,G_i) = \frac{1}{|X_i|^m} \sum_{\phi\colon [m]\to X_i} \prod_{e\in\binom{[m]}{r}} F(e)(G_i(\phi(e))). \tag{5.5}$$

If G is $\mathcal{P}(\mathcal{K})$-valued, then we define

$$t(F,G_i) = \frac{1}{|X_i|^m} \sum_{\phi\colon [m]\to X_i} \prod_{e\in\binom{[m]}{r}} \langle G_i(\phi(e)), F(e)\rangle, \tag{5.6}$$

which agrees with (5.5) if the deployed colors are point measures.

Further, recall Definition 2.2.3, that tells us that one of the several equivalent formulations of the convergence of a sequence $(G_i)_{i=1}^\infty$ is the simultaneous convergence of the numerical sequences $t(F,G_i)$ for each $F \in \mathcal{F}$. We define now the subgraph densities of the ultraproduct functions.

Definition 5.4.2. *For an $F \in \Pi_m(C(\mathcal{K}))$ and an ultraproduct function $\mathbf{G}\colon \mathbf{X}^r \to \mathcal{P}(\mathcal{K})$ the density of F in $\mathbf{G}$ is defined as*

$$t(F,\mathbf{G}) = \int_{\mathbf{X}^m} \prod_{e\in\binom{[m]}{r}} \langle \mathbf{G}(P_e(\mathbf{x})), F(e)\rangle \mathrm{d}\mu_{[m]}(\mathbf{x}).$$

Let $f_i\colon X_i \to [-d,d]$ be uniformly bounded real functions, then their ultraproduct $\mathbf{f} = [\{f_i\}_{i=1}^\infty]$ is defined by $\mathbf{f}(\mathbf{x}) = \lim_\omega f_i(p_i)$, where $\mathbf{x} = [\{x_i\}_{i=1}^\infty]$. We formulate an integration formula for ultraproduct functions that was proven in [49].

Lemma 5.4.3. *[49] For an ultraproduct function $\mathbf{f} = [\{f_i\}_{i=1}^\infty]$ on $\mathbf{X}$ is measurable and we have*

$$\int_{\mathbf{X}} \mathbf{f}\mathrm{d}\mu = \lim_\omega \frac{\sum_{x_i\in X_i} f_i(x_i)}{|X_i|}.$$

The integration formula above directly implies that ultraproduct functions can be considered as the limit objects of $\mathcal{P}(\mathcal{K})$-valued functions.

Lemma 5.4.4. *Let $\{G_i\}_{i=1}^\infty$ be a sequence of $\mathcal{P}(\mathcal{K})$-colored r-graphs such that $V(G_i) = X_i$. Then for every $F \in \Pi(C(\mathcal{K}))$ we have $\lim_\omega t(F,G_i) = t(F,\mathbf{G})$, where $\mathbf{G}\colon \mathbf{X}^r \to \mathcal{P}(\mathcal{K})$ is the ultraproduct of the G_i functions.*

Proof. Let $m \ge 1$ and $F \in \Pi_m(C(\mathcal{K}))$ be arbitrary, let the functions G_i and $\mathbf{G}$ be as in the statement. Let us rewrite (5.6) as

$$t(F,G_i) = \frac{1}{|X_i|^m} \sum_{x_i=(x_i^1,\ldots,x_i^m)\in X_i^m} \prod_{e\in\binom{[m]}{r}} \langle G_i(x_i^{e_1},\ldots,x_i^{e_r}), F(e)\rangle = \frac{1}{|X_i|^m} \sum_{x_i\in X_i^m} \prod_{e\in\binom{[m]}{r}} \langle G_i(P_e(x_i)), F(e)\rangle. \tag{5.7}$$

We claim that $[\{\prod_{e\in\binom{[m]}{r}}\langle G_i(P_e(.)),F(e)\rangle\}_{i=1}^{\infty}] = \prod_{e\in\binom{[m]}{r}}\langle \mathbf{G}(P_e(.)),F(e)\rangle$. This is a consequence of the definition of the ultraproduct function $\mathbf{G}$, of the interchangeability of the projection P_e operation with taking ultraproducts, i.e, $P_e([\{x_i\}_{i=1}^{\infty}]) = [\{P_e(x_i)\}_{i=1}^{\infty}]$, and of the product of ultralimits $\lim_\omega x_i \lim_\omega y_i$ being equal to the ultralimit of the products $\lim_\omega x_i y_i$. Now Lemma 5.4.3 finishes the proof.

□

We are now ready with the first step of establishing useful limit objects for sequences of r-uniform weighted hypergraphs. Next we turn to the representation of the abstract, non-separable space $\mathbf{X}^r$ by the well-understood Lebesgue-space of $[0,1]^{\mathfrak{h}([r])}$. This will be followed by connecting the probability measure valued functions between the spaces so that r-graphons can truly serve as limit objects of r-graph sequences. We start with the introduction of some additional notation related to subgraph densities.

Again, as in the case of k-colored r-graphs, we want to express the homomorphism densities $t(F,G_i)$ as measures of certain subsets of X_i^m. For this purpose we rewrite the more general (5.6) as

$$t(F,G_i) = \frac{1}{|X_i|^m}\sum_{\phi\colon [m]\to X_i}\prod_{e\in\binom{[m]}{r}}\int_0^1 \mathbb{1}_{\{\langle G_i(.),F(e)\rangle\geq a_e\}}(\phi(e))\mathrm{d}a_e \tag{5.8}$$

$$= \int_{[0,1]^{\binom{[m]}{r}}}\frac{1}{|X_i|^m}\sum_{\phi\colon [m]\to X_i}\prod_{e\in\binom{[m]}{r}}\mathbb{1}_{\{\langle G_i(.),F(e)\rangle\geq a_e\}}(\phi(e))\mathrm{d}\lambda(a), \tag{5.9}$$

where $a = (a_e)_{e\in\binom{[m]}{r}}$. Let $G_i^{f,a} = \{\, p_i \in X_i^r \mid \langle G_i(p_i), f\rangle \geq a \,\}$ for every $a\in[0,1]$ and $f\in\mathcal{F}$. We define

$$T(F,a,G_i) = \bigcap_{e\in\binom{[m]}{r}} P_e^{-1}(P_{s_e}(G_i^{F(e),a_e})),$$

that is the set of maps $\phi\colon [m]\to X_i$ given by elements of X_i^m that satisfy for each $e\in\binom{[m]}{r}$ that $\langle G_i(\phi(e)),F(e)\rangle \geq a_e$. Now we have

$$t(F,G_i) = \int_{[0,1]^{\binom{[m]}{r}}}\frac{T(F,a,G_i)}{|X_i|^m}\,\mathrm{d}\lambda(a). \tag{5.10}$$

Let for an ultraproduct function $\mathbf{G}\colon \mathbf{X}^r\to\mathcal{P}(\mathcal{K})$ the measurable set $\mathbf{G}^{f,a} = \{\,\mathbf{p}\in\mathbf{X}^r \mid \langle \mathbf{G}(\mathbf{p}), f\rangle \geq a\,\}$ for any $a\in[0,1]$ and $f\in\mathcal{F}$, and further, let

$$T(F,a,\mathbf{G}) = \bigcap_{e\in\binom{[m]}{r}} P_e^{-1}(P_{s_e}(\mathbf{G}^{F(e),a_e}))$$

for $F\in\Pi_m(\mathcal{F})$ and $a\in[0,1]^{\binom{[m]}{r}}$.

We reformulate $t(F, \mathbf{G})$ from Definition 5.4.2 as

$$t(F, \mathbf{G}) = \int_{[0,1]^{\binom{[m]}{r}}} \mu_{[m]}(T(F, a, \mathbf{G}))\mathrm{d}\lambda(a). \tag{5.11}$$

One of our main results is the following, we will use Theorem 5.2.6, that was proved in [49].

Theorem 5.4.5. *Let $\mathcal{K}$ be a compact Polish space, and $\mathbf{G}\colon \mathbf{X}^r \to \mathcal{P}(\mathcal{K})$ be a $\sigma([r])$-measurable function. Then there exists a separable realization $\phi\colon \mathbf{X}^r \to [0,1]^{\mathfrak{h}([r])}$ so that there is a Borel measurable map $W\colon [0,1]^{\mathfrak{h}([r])} \to \mathcal{P}(\mathcal{K})$ so that $\mu_{[r]}$-almost everywhere $\mathbf{G} = W \circ \phi$.*

Proof. Let $\mathcal{F}$ be a countable subset of $C(\mathcal{K})$ so that the linear subspace generated by $\mathcal{F}$ is L^∞-dense in $C(\mathcal{K})$, and each of its elements take values between 0 and 1. Then let for $a \in \mathbb{R}$ and $f \in \mathcal{F}$ the $\mathcal{B}_{[r]}$-measurable set $\mathbf{G}^{f,a} \subset \mathbf{X}^r$ be defined as $\mathbf{G}^{f,a} = \{\, \mathbf{p} \mid \langle \mathbf{G}(\mathbf{p}), f\rangle \geq a \,\}$. Consider the σ-algebra $\mathcal{A}$ generated by the countable collection $\{\, \mathbf{G}^{f,a} \mid f \in \mathcal{F}, a \in \mathbb{Q} \,\}$ of measurable subsets of $\mathbf{X}^r$. Note that $\mathcal{A}$ is separable, so that we can apply Theorem 5.2.6. We obtain that there exists a separable realization $\phi\colon \mathbf{X}^r \to [0,1]^{\mathfrak{h}([r])}$ such that for each $f \in \mathcal{F}$, and $a \in \mathbb{Q}$ there exists a measurable subset $W^{f,a}$ of $[0,1]^{\mathfrak{h}([r])}$ such that $\mu_{[r]}(\mathbf{G}^{f,a} \triangle \phi^{-1}(W^{f,a})) = 0$. Further, we can always choose these sets in a way so that for each f the family $\{W^{f,a}\}_{a\in\mathbb{Q}}$ is monotone increasing in a. To see this, consider for any $f \in \mathcal{F}$ and $a \in \mathbb{Q}$ the set $U^{f,a} = \cup_{b\geq a, b\in\mathbb{Q}} W^{f,b}$, then

$$\begin{aligned} \mu_{[r]}(\mathbf{G}^{f,a} \triangle \phi^{-1}(U^{f,a})) &= \mu_{[r]}((\cup_{b\geq a, b\in\mathbb{Q}} \mathbf{G}^{f,b}) \triangle \phi^{-1}(\cup_{b\geq a, b\in\mathbb{Q}} W^{f,b})) \\ &\leq \sum_{b\geq a, b\in\mathbb{Q}} \mu_{[r]}(\mathbf{G}^{f,b} \triangle \phi^{-1}(W^{f,b})) = 0. \end{aligned}$$

We define for each $f \in \mathcal{F}$ the measurable function $W^f\colon [0,1]^{\mathfrak{h}([r])} \to \mathbb{R}$ by $W^f = \inf_{a\in\mathbb{Q}\cap[0,1]} a\mathbb{1}_{W^{f,a}}$. We have that $\lambda(W^{f,a} \triangle (W^f)^{-1}([a,\infty))) = 0$ for each $f \in \mathcal{F}$, and $a \in \mathbb{Q}$. It also holds true that $\mu_{[r]}$-almost everywhere $W^f \circ \phi = \langle \mathbf{G}, f\rangle$ for every $f \in \mathcal{F}$, where $(\langle \mathbf{G}, f\rangle)(\mathbf{p}) = \langle \mathbf{G}(\mathbf{p}), f\rangle$. Since the cardinality of $\mathcal{F}$ is countable it follows that the exceptional set $\mathbf{N} \subset \mathbf{X}^r$ where the above equality does not hold for some f has also measure 0. Suppose that for $\mathbf{p}_1, \mathbf{p}_2 \in \mathbf{X}^r \setminus \mathbf{N}$ it holds that $\phi(\mathbf{p}_1) = \phi(\mathbf{p}_2)$, then for every f we have $\langle \mathbf{G}(\mathbf{p}_1), f\rangle = \langle \mathbf{G}(\mathbf{p}_2), f\rangle$, and further the probability measures $\mathbf{G}(\mathbf{p}_1)$ and $\mathbf{G}(\mathbf{p}_2)$ on $\mathcal{K}$ coincide. We can now define W on the image of $\mathbf{X}^r \setminus \mathbf{N}$ to be the probability measure defined by the measure taken by $\mathbf{G}$ on the ϕ-ancestors, the equality $\langle W, f\rangle = W^f$ is true by the uniqueness assertion of the Riesz Representation Theorem, measurability follows from the measurability of ϕ and $\mathbf{G}$. This concludes the proof of the theorem.

□

We have shown that for any $\mathbf{G}$ there is a $\phi\colon \mathbf{X}^r \to [0,1]^{\mathfrak{h}([r])}$ separable realization and a function W such that $\mathbf{G}$ is almost everywhere the pull-back of W by ϕ. It remains to show if $\mathbf{G}$ is the ultraproduct of a convergent sequence $\{G_i\}_{i=1}^\infty$, then the densities of elements of $\Pi(\mathcal{F})$ in the corresponding W also converge.

Let us recall the definition of these densities in $(\mathcal{K}, r)$-graphons in Equation (2.8). A slight variant of these are graphons that take $P(\mathcal{K})$ values instead of $\mathcal{K}$. If $F \in \Pi_m(\mathcal{F})$, and $W\colon [0,1]^{\mathfrak{h}([r])} \to \mathcal{P}(\mathcal{K})$, then

$$t(F, W) = \int_{[0,1]^{\mathfrak{h}([m],r)}} \prod_{e \in \binom{[m]}{r}} \langle W(x_{\mathfrak{h}(e)}), F(e) \rangle \mathrm{d}\lambda(x_{\mathfrak{h}([m],r)}). \tag{5.12}$$

We may rewrite the above expression using the level sets as in the case of the discrete case above. For $f \in \mathcal{F}$ and $a \in [0,1]$, let $W^{f,a} = \{ x \in [0,1]^{\mathfrak{h}([r])} \mid \langle W(x), f \rangle \geq a \}$. Define further the set

$$T(F, a, W) = \bigcap_{e \in \binom{[m]}{r}} p^{-1}_{\mathfrak{h}(e)}(L_{s_e}(W^{F(e),a_e})) \subset [0,1]^{\mathfrak{h}([m],r)}$$

for $F \in \Pi_m(\mathcal{F})$ and $a \in [0,1]^{\binom{[m]}{r}}$, where $p_{\mathfrak{h}(e)}\colon [0,1]^{\mathfrak{h}([m],r)} \to [0,1]^{\mathfrak{h}(e)}$ is the natural projection, and $L_{s_e}\colon [0,1]^{\mathfrak{h}([r])} \to [0,1]^{\mathfrak{h}(e)}$ is the isomorphism aligning coordinates corresponding to a previously fixed bijection $s_e\colon [r] \to e$.

Then we can rewrite $t(F, W)$ as

$$t(F, W) = \int_{[0,1]^{\binom{[m]}{r}}} \lambda(T(F, a, W)) \mathrm{d}\lambda(a). \tag{5.13}$$

Lemma 5.4.6. *If $\phi\colon \mathbf{X}^r \to [0,1]^{\mathfrak{h}([r])}$ is a separable realization, and $\mathbf{G}\colon \mathbf{X}^r \to \mathcal{P}(\mathcal{K})$ and $W\colon [0,1]^{\mathfrak{h}([r])} \to \mathcal{P}(\mathcal{K})$ are such that $\mathbf{G} = W \circ \phi$ almost everywhere, then for every $F \in \Pi(\mathcal{F})$ we have $t(F, \mathbf{G}) = t(F, W)$.*

Proof. Let $F \in \Pi_m(\mathcal{F})$ be arbitrary. We want to show that for every $a \in [0,1]^{\binom{[m]}{r}}$ it holds that $\mu_{[m]}(T(F, a, \mathbf{G})) = \lambda(T(F, a, W))$, by (5.11) and (5.13) this implies the statement of the lemma. Lemma 5.2.7 implies that there exists a degree m lifting ψ of the separable realization ϕ. Also, since $\mathbf{G}$ and $W \circ \phi$ are equal almost everywhere we have that $\mu_{[r]}(\mathbf{G}^{f,a} \triangle \phi^{-1}(W^{f,a})) = 0$ for every $f \in \mathcal{F}$ and $a \in [0,1]$. As ψ is a lifting of ϕ it follows that for every $e \in \binom{[m]}{r}$ the analog of the previous also holds for cylinders, that is $\mu_{[m]}(P_e^{-1}(P_{s_e}(\mathbf{G}^{f,a})) \triangle \psi^{-1}(p^{-1}_{\mathfrak{h}(e)}(L_{s_e}(W^{f,a})))) = 0$. This immediately implies that $\mu_{[m]}(T(F, a, \mathbf{G}) \triangle \psi^{-1}(T(F, a, W))) = 0$, and since ψ is measure preserving this concludes the proof.

□

An immediate consequence is one of the main results in this section that concerns the existence of a limit object.

Theorem 5.4.7. *Let $\{G\}_{i=1}^{\infty}$ be a sequence of $\mathcal{K}$-colored (or $\mathcal{P}(\mathcal{K})$ colored) r-uniform hypergraphs so that for every n the sequence of probability measures on $\mathcal{K}$-colored random hypergraphs on $[n]$, $\{\mathbb{G}(n, G_i)\}_{i=1}^{\infty}$, converges weakly (or equivalently in the Prokhorov metric) when i tends to infinity. Then there exists a $\mathcal{P}(\mathcal{K})$-valued r-graphon W on $[0,1]^{\mathfrak{h}([r])}$ such that the measures*

$\{\mathbb{G}(n, G_i)\}_{i=1}^\infty$ converge to $\mathbb{G}(n, W)$ weakly, or equivalently the sequences $\{t(F, G_i)\}_{i=1}^\infty$ converge to $t(F, W)$ for every $F \in \Pi(\mathcal{F})$.

For our purposes it will be useful to consider not the ultraproduct function $\mathbf{G}$, but its conditional expectation $\hat{\mathbf{G}}(\mathbf{p}) = \mathbb{E}[\mathbf{G}(\mathbf{p}) \mid \sigma([r])^*]$ with respect to the σ-algebra of cylinder sets. Such a measure valued function can be defined in a sensible way, and it always exists, that is the consequence of a generalization of the Radon-Nikodym Theorem given next, a proof can be found for example in [48].

Theorem 5.4.8 (Radon-Nikodym-Dunford-Pettis). *Let $(X, \mu, \mathcal{A})$ be a probability measure space and f be an $\mathcal{A}$-measurable L^*-valued function, where L is a Banach space, and let $\mathcal{B} \subset \mathcal{A}$ be a sub-σ-algebra. Then there is an essentially unique weak -*-$\mathcal{B}$-measurable function $\mathbb{E}[f \mid \mathcal{B}]$ such that for any $v \in L$ and $B \in \mathcal{B}$ we have*

$$\int_B \langle f, v \rangle \mathrm{d}\mu = \int_B \langle \mathbb{E}[f \mid \mathcal{B}], v \rangle \mathrm{d}\mu.$$

For convenience, in our case $L = C(\mathcal{K})$, $L^* = \mathcal{P}(\mathcal{K})$, and the weak-*-measurability means here simply measurability with respect to the weak topology.

The advantage of this operation will become clear when we want to construct a random $\mathcal{K}$-valued hypergraph through the separable realization of $\mathbf{G}$, mainly it helps to avoid redundancy when we talk about equivalent r-graphons.

We remark, that if the target space $\mathcal{K}$ has a metric structure, then one can define the ultraproduct function $\mathbf{G}$ more directly as a $\mathcal{K}$-valued function instead of the general case where it is $\mathcal{P}(\mathcal{K})$-valued, because we have the notion of the metric ultraproduct at hand on the space $\mathcal{K}$ (analogous to the numerical ultraproduct). In the end we would arrive at the same point as in the general case: we represent $\mathbf{G}$ as a point-measure valued function, then project it to the space of $\sigma([r])^*$-measurable probability measure valued functions via the Radon-Nikodym-Dunford-Pettis Theorem.

Using these concept combined with the above correspondence we show now that $\mathcal{P}(\mathcal{K})$-valued r-graphons defined as functions on $[0,1]^{\mathfrak{h}([r],r-1)}$ already serve as a suitable limit space in contrast to ones defined on $[0,1]^{\mathfrak{h}([r])}$. Such objects are equivalent to the $(\mathcal{K}, r)$-graphons (recall their space, $\Xi^r(\mathcal{K})$) introduced in Section 2.2.3, and reduce the dimension of the domain of the function W obtained in Theorem 5.4.7 from $\mathfrak{h}([r])$ to $\mathfrak{h}([r], r-1)$.

Lemma 5.4.9. *For every measurable $\mathbf{G}\colon \mathbf{X}^r \to \mathcal{P}(\mathcal{K})$ and $F \in \Pi(\mathcal{F})$ we have $t(F, \mathbf{G}) = t(F, \mathbb{E}[\mathbf{G} \mid \sigma([r])^*])$. Also, for every $W\colon [0,1]^{\mathfrak{h}([r],r-1)} \to \mathcal{P}(\mathcal{K})$ we have $t(F, W) = t(F, \mathbb{E}[W \mid \mathcal{A}^*_{[r]}])$.*

Proof. Let $F \in \Pi_m(\mathcal{F})$. For the first statement we recall $t(F, \mathbf{G})$ that was formally defined by the expression

$$t(F, \mathbf{G}) = \int_{\mathbf{X}^m} \prod_{e \in \binom{[m]}{r}} \langle \mathbf{G}(P_e^{-1}(\mathbf{x})), F(e) \rangle \mathrm{d}\mu_{[m]}(\mathbf{x}). \tag{5.14}$$

Each of the product terms $\langle \mathbf{G}(P_e^{-1}(.)), F(e)\rangle$ inside the above integral considered as real-valued functions on $\mathbf{X}^m$ are measurable with respect to $\sigma(e) \subset \sigma([m])$ for the respective $e \in \binom{[m]}{r}$. We can employ the integration rule in [49, Proposition 5.3] that tells us that if $g_j \colon \mathbf{X}^m \to [-d, d]$ are bounded $\sigma(A_i)$-measurable functions with $A_i \subset [m]$ for $i \in [n]$, and $\mathcal{B} = \langle \sigma(A_1 \cap A_j) \mid j \geq 2\rangle$, then

$$\int_{\mathbf{X}^m} \prod_{j=1}^{n} g_j(\mathbf{x}) \mathrm{d}\mu_{[m]}(\mathbf{x}) = \int_{\mathbf{X}^m} \mathbb{E}[g_1 \mid \mathcal{B}](\mathbf{x}) \prod_{j=2}^{n} g_j(\mathbf{x}) \mathrm{d}\mu_{[m]}(\mathbf{x}).$$

Hence in our case,

$$\int_{\mathbf{X}^m} \prod_{e \in \binom{[m]}{r}} \langle \mathbf{G}(P_e^{-1}(\mathbf{x})), F(e)\rangle \mathrm{d}\mu_{[m]}(\mathbf{x}) = \int_{\mathbf{X}^m} \prod_{e \in \binom{[m]}{r}} \mathbb{E}[\langle \mathbf{G}(P_e^{-1}(.)), F(e)\rangle \mid \sigma(e)^*](\mathbf{x}) \mathrm{d}\mu_{[m]}(\mathbf{x}). \tag{5.15}$$

Since $\mathbb{E}[\langle \mathbf{G}(P_e^{-1}(.)), f\rangle \mid \sigma(e)^*] = \mathbb{E}[\langle \mathbf{G}(.), f\rangle \mid \sigma([r])^*]$ for every $e \in \binom{[m]}{r}$ and $f \in \mathcal{F}$, and by definition $\mathbb{E}[\langle \mathbf{G}(.), f\rangle \mid \sigma([r])^*] = \langle \mathbb{E}[\mathbf{G} \mid \sigma([r])^*], f\rangle$ we conclude that $t(F, \mathbf{G}) = t(F, \mathbb{E}[\mathbf{G} \mid \sigma([r])^*])$.

Regarding the second statement we write

$$t(F, W) = \int_{[0,1]^{\mathfrak{h}([m],r)}} \prod_{e \in \binom{[m]}{r}} \langle W(x_{\mathfrak{h}(e)}), F(e)\rangle \mathrm{d}\lambda(x_{\mathfrak{h}([m],r)}), \tag{5.16}$$

and for each e we pull in the integral over x_e into the respective product term, integrating first inside the factors is then no different from taking the conditional expectation with respect to the respective σ-algebra $\mathcal{A}_e^*$. Therefore it follows that $t(F, W) = t(F, \mathbb{E}[W \mid \mathcal{A}_{[r]}^*])$. □

Next we formulate a version of Lemma 5.4.6.

Lemma 5.4.10. *For every measurable* $\mathbf{G} \colon \mathbf{X}^r \to \mathcal{P}(\mathcal{K})$ *there exists a separable realization* $\phi \colon \mathbf{X}^r \to [0,1]^{\mathfrak{h}([r])}$ *and a* $(\mathcal{K}, r)$*-graphon* $U \colon [0,1]^{\mathfrak{h}([r],r-1)} \to \mathcal{P}(\mathcal{K})$ *such that* $\mathbb{E}[\mathbf{G} \mid \sigma([r])^*] = U \circ p_{\mathfrak{h}([r],r-1)} \circ \phi$ *almost everywhere, then for every* $F \in \Pi(\mathcal{F})$ *we have* $t(F, \mathbf{G}) = t(F, U)$.

Proof. Let $\hat{\mathbf{G}} = \mathbb{E}[\mathbf{G} \mid \sigma([r])^*]$, and let $\mathbf{G}^{f,a} = \{\mathbf{p} \mid \langle \mathbf{G}(\mathbf{p}), f\rangle \geq a\}$ and $\hat{\mathbf{G}}^{f,a} = \{\mathbf{p} \mid \langle \hat{\mathbf{G}}(\mathbf{p}), f\rangle \geq a\}$ for each $f \in \mathcal{F}$ and $a \in [0,1]$. Let $\mathcal{D}$ denote the separable generated by the collection of the sets $\mathbf{G}^{f,a}$ and $\hat{\mathbf{G}}^{f,a}$, and $\mathcal{D}' \subset \sigma([r])^*$ be the one generated only by the $\hat{\mathbf{G}}^{f,a}$ sets. We know from Theorem 5.2.6 and Theorem 5.4.5 that there exists a separable realization ϕ such that for each $A \in \mathcal{D}$ there is a measurable $V_A \subset [0,1]^{\mathfrak{h}([r])}$ such that $\mu_{[r]}(A \triangle \phi^{-1}(V_A)) = 0$, and there is a measurable $W \colon [0,1]^{\mathfrak{h}([r])} \to \mathcal{P}(\mathcal{K})$ such that $\mathbf{G} = W \circ \phi$ almost everywhere.

We claim that U defined by $U(p_{\mathfrak{h}([r],r-1)}(x)) = \mathbb{E}[W \mid \mathcal{A}_{[r]}^*](x)$ satisfies the requirements of the statement, in particular that $\mathbb{E}[W \mid \mathcal{A}_{[r]}^*] \circ \phi = \hat{\mathbf{G}}$ almost everywhere. In order to verify $\mathbb{E}[W \mid \mathcal{A}_{[r]}^*] \circ \phi = \mathbb{E}[\mathbf{G} \mid \sigma([r])^*]$ almost everywhere we need to check the two conditions that defines a conditional expectation. First, we note the fact that $U \circ p_{\mathfrak{h}([r],r-1)} \circ \phi$ is $\sigma([r])^*$-measurable, this is trivial since ϕ is a separable realization. For the second requirement, it suffices to show $\langle \mathbb{E}[W \mid \mathcal{A}^*_{[r]}] \circ \phi, f\rangle = \langle \hat{\mathbf{G}}, f\rangle$ almost everywhere for arbitrary $f \in \mathcal{F}$, since $\mathcal{F}$ is countable. To do this, let us fix $f \in \mathcal{F}$. Since $\langle \hat{\mathbf{G}}, f\rangle$ is $\mathcal{D}'$-measurable it is enough to show that

$$\int_B \langle \mathbb{E}[W \mid \mathcal{A}^*_{[r]}] \circ \phi(\mathbf{x}), f\rangle \mathrm{d}\mu_{[r]}(\mathbf{x}) = \int_B \langle \hat{\mathbf{G}}(\mathbf{x}), f\rangle \mathrm{d}\mu_{[r]}(\mathbf{x}) \tag{5.17}$$

for arbitrary $B \in \mathcal{D}'$, instead for all elements of $\sigma([r])^*$. The left-hand side of (5.17) can be written as

$$\int_B \int_0^1 \mathbb{1}_{\mathbf{G}^{f,a}}(\mathbf{x}) \mathrm{d}a \mathrm{d}\mu_{[r]}(\mathbf{x}) = \int_0^1 \mu_{[r]}(\mathbf{G}^{f,a} \cap B) \mathrm{d}a. \tag{5.18}$$

Since $B \in \mathcal{D}$, there exists a $V_B \subset [0,1]^{\mathfrak{h}([r])}$ such that $\mu_{[r]}(B \triangle \phi^{-1}(V_B)) = 0$. As $B \in \sigma([r])^*$ and ϕ is a separable realization, $V_B \in \mathcal{A}^*_{[r]}$. Let $W^{f,a} = \{ x \in [0,1]^{\mathfrak{h}([r])} \mid \langle W(x), f\rangle \geq a \}$ for $f \in \mathcal{F}$ and $a \in [0,1]$. Then the right-hand side of (5.17) is equal to

$$\begin{aligned}
\int_B \langle \mathbb{E}[W \mid \mathcal{A}^*_{[r]}] \circ \phi(\mathbf{x}), f\rangle \mathrm{d}\mu_{[r]}(\mathbf{x}) &= \int_{\phi^{-1}(V_B)} \langle \mathbb{E}[W \mid \mathcal{A}^*_{[r]}] \circ \phi(\mathbf{x}), f\rangle \mathrm{d}\mu_{[r]}(\mathbf{x}) \\
&= \int_{V_B} \langle \mathbb{E}[W \mid \mathcal{A}^*_{[r]}](x), f\rangle \mathrm{d}\lambda(x_{\mathfrak{h}([r])}) \\
&= \int_{V_B} \langle W(x), f\rangle \mathrm{d}\lambda(x_{\mathfrak{h}([r])}) \\
&= \int_{V_B} \int_0^1 \mathbb{1}_{W^{f,a}}(x) \mathrm{d}a \mathrm{d}\lambda(x_{\mathfrak{h}([r])}) \\
&= \int_0^1 \lambda(W^{f,a} \cap V_B) \mathrm{d}a,
\end{aligned} \tag{5.19}$$

the second equality is the consequence of ϕ being measure-preserving, whereas the third is true by the definition of the conditional expectation and $V_B \in \mathcal{A}^*_{[r]}$.

For any $f \in \mathcal{F}$ and $a \in [0,1]$ we have $\mu_{[r]}(\mathbf{G}^{f,a} \triangle \phi^{-1}(W^{f,a})) = 0$ since $\mathbf{G} = W \circ \phi$ almost everywhere, consequently $\mu_{[r]}(\mathbf{G}^{f,a} \cap B) = \lambda(W^{f,a} \cap V_B)$, as ϕ is measure-preserving. Comparing (5.18) and (5.19) verifies the equality (5.17).

It remains to show that for every $F \in \Pi(\mathcal{F})$ we have $t(F, \mathbf{G}) = t(F, U)$, this follows from Lemma 5.4.6 and Lemma 5.4.9.

□

From the preceding Lemma 5.4.10 directly follows an improved version of Theorem 5.4.7, already shown to hold true in Chapter 2 through different methods.

Theorem 5.4.11. *Let $\{G\}_{i=1}^{\infty}$ be a sequence of $\mathcal{K}$-colored (or $\mathcal{P}(\mathcal{K})$ colored) r-uniform hypergraphs so that for every n the sequence of probability measures on $\mathcal{K}$-colored random* *hypergraphs on* $[n]$, $\{\mathbb{G}(n,G_i)\}_{i=1}^\infty$, *converges weakly when i tends to infinity. Then there exists a* $(\mathcal{K},r)$*-graphon* $W\colon [0,1]^{\mathfrak{h}([r],r-1)}$ *such that the measures of* $\{\mathbb{G}(n,G_i)\}_{i=1}^\infty$ *converge to* $\mathbb{G}(n,W)$ *weakly.*

We are going to characterize the equivalence classes of $\mathcal{P}(\mathcal{K})$-valued hypergraphons determined by the equality of the distributions of sampled $(\mathcal{K},r)$-graphs, that is $U\sim W$ exactly in the case, when $t(F,U)=t(F,W)$ for any $F\in\mathcal{F}$. The main statement is going to be a straight-forward analog of the result in the simple hypergraph case, [49], where $\mathcal{K}=\{0,1\}$, and the k-colored r-graph case above. The proof is also a slight variant of the proof given for the latter case. We note that for any U it is true that U can be chosen as the limit of $\{\mathbb{G}(n,U)\}_{n=1}^\infty$ in the sense of Theorem 5.4.11.

Theorem 5.4.12. *Let* $\mathcal{K}$ *be a compact Polish space, and let U and W be two* $(\mathcal{K},r)$*-graphons on* $[0,1]^{\mathfrak{h}([r],r-1)}$ *such that* $t(F,U)=t(F,W)$ *for any* $F\in\mathcal{F}$. *Then there exist structure-preserving maps* ν_1 *and* ν_2 *from* $[0,1]^{\mathfrak{h}([r],r-1)}$ *to* $[0,1]^{\mathfrak{h}([r],r-1)}$ *such that* $U\circ\nu_1=W\circ\nu_2$ *almost everywhere.*

Proof. Let U and W be as in the statement of the theorem. We define U' to be a $\mathcal{K}$-valued map on $[0,1]^{\mathfrak{h}([r])}$ such that for every $x\in[0,1]^{\mathfrak{h}([r],r-1)}$ the distribution of $U'(x,Y)$ is equal to $U(x)$ for a uniformly chosen random Y from $[0,1]$, and analogously define W'.

Let τ be the ultraproduct of the random coordinate systems τ_n that is with probability one a separable realization from $\mathbf{X}^r$ to $[0,1]^{\mathfrak{h}([r])}$, where $\mathbf{X}^r$ is the ultraproduct of $\{[n]^r\}_{n=1}^\infty$, see [49, Lemma 4.9]. Now we have that if τ_n and $\mathbb{G}(n,U)$ are driven by the same random $Z_n\in[0,1]^{\mathfrak{h}([n],r)}$, then $\tau_n^{-1}(U')=\mathbb{G}(n,U)$. Further, for every $f\in\mathcal{F}$ and $a\in[0,1]$ let $U^{f,a}=\{x\in[0,1]^{\mathfrak{h}([r])}\mid f(U'(x))\ge a\}$, then also $\tau_n^{-1}(U^{f,a})=\mathbb{G}(n,U^{f,a})$. By Lemma 5.2.13 we have that with probability 1 for every $f\in\mathcal{F}$ and $a\in[0,1]\cap\mathbb{Q}$ it is true that

$$\mu_{[r]}([\{\mathbb{G}(n,U^{f,a})\}_{n=1}^\infty]\triangle\tau^{-1}(U^{f,a}))=0. \tag{5.20}$$

The same assertion holds for W, where $W^{f,a}$ is defined analogously.

The equality $t(F,U)=t(F,W)$ for each $F\in\mathcal{F}$ implies by Theorem 2.2.2 that $\mathbb{G}(n,U)$ and $\mathbb{G}(n,W)$ have the same distribution, so they can be coupled as the common random objects Y_n for each n. For any $f\in\mathcal{F}$ and $a\in[0,1]$ the random graph $\mathbb{G}(n,U^{f,a})$ is trivially $\mathbb{G}(n,U)$-measurable (similarly for W), so the coupling also satisfies that there are random objects $Y_n^{f,a}$ that are equal to the common realization of $\mathbb{G}(n,U^{f,a})$ and $\mathbb{G}(n,U^{f,a})$.

Let $\mathbf{H}^{f,a}=[\{Y_n^{f,a}\}_{n=1}^\infty]$, then the above discussion implies that there exist separable realizations ϕ_1 and ϕ_2 that are realizations of ultraproducts of two random coordinate sytems such that $\mu_{[r]}(\mathbf{H}^{f,a}\triangle\phi_1^{-1}(U^{f,a}))=0$ and $\mu_{[r]}(\mathbf{H}^{f,a}\triangle\phi_2^{-1}(W^{f,a}))=0$ for each $f\in\mathcal{F}$ and $a\in[0,1]\cap\mathbb{Q}$, therefore also $\mu_{[r]}(\phi_1^{-1}(U^{f,a})\triangle\phi_2^{-1}(W^{f,a}))=0$. Set $\mathcal{D}=\sigma(\phi_1^{-1}(\mathcal{A}_{[r]}),\phi_2^{-1}(\mathcal{A}_{[r]}))$ that is a separable σ-algebra on $\mathbf{X}^{[r]}$, so by Theorem 5.2.6 there exists a separable realization ϕ_3 such that for each measurable $D\subset[0,1]^{\mathfrak{h}([r])}$ the element $\phi_i^{-1}(D)$ of $\mathcal{D}$ can be represented by a subset of $[0,1]^{\mathfrak{h}([r])}$ denoted by $\psi_i(D)$. As in the proof of Theorem 5.2.15, ψ_1 and ψ_2 defined this way are structure preserving embeddings from $\mathcal{L}^{\mathfrak{h}([r])}\to\mathcal{L}^{\mathfrak{h}([r])}$ satisfying $\lambda(\psi_1(U^{f,a})\triangle\psi_2(W^{f,a}))=0$ for each $f\in\mathcal{F}$ and $a\in[0,1]\cap\mathbb{Q}$.

If follows that by Lemma 5.2.12 there are structure preserving $\hat{\nu}_1$ and $\hat{\nu}_2$ such that $\lambda(\hat{\nu}_1^{-1}(U^{f,a})\triangle\hat{\nu}_2^{-1}(W^{f,a})) = 0$ for each $f \in \mathcal{F}$ and $a \in [0,1] \cap \mathbb{Q}$. This implies that $U' \circ \hat{\nu}_1 = W' \circ \hat{\nu}_2$ almost everywhere on $[0,1]^{\mathfrak{h}([r])}$. We conclude that ν_1 and ν_2 defined by $\nu_i(x_{\mathfrak{h}([r],r-1)}) = P_{\mathfrak{h}([r],r-1)}(\hat{\nu}_i(x_{\mathfrak{h}([r],r-1)}, 0))$ for $i = 1, 2$ are structure-preserving satisfying the conditions of the theorem since $\hat{\nu}_1$ and $\hat{\nu}_2$ were structure-preserving, that is $U \circ \nu_1 = W \circ \nu_2$ almost everywhere. □

5.5 Regularity Lemma for $(\mathcal{K}, r)$-graphs

One of the most important tools in graph theory and in the broader field of combinatorics in recent years is the Regularity Lemma. It states that for every graph there exists a quasi-random approximating graph of bounded size, where this bound is only a function of the error margin of the approximation, but not of the graph in consideration. We dealt with this topic already in-depth in Section 3.3, in particular we stated and proved several versions of varying strength. In this section we prove a general version of applicable to colored uniform hypergraphs, the result here directly implies Lemma 3.3.27 except for the concrete upper bound on the number of partition classes required for the regular approximation. This shortcoming is a consequence of the non-standard methods employed in the current chapter. In particular we will rely on Lemma 5.4.10 that ensures the existence of a certain separable realization.

The proof given here follows the framework of the proof for the simple r-graph case given by Elek and Szegedy [49]. That version was dealt with in [49] was first proved for $r = 3$ in a slightly different setting by Frankl and Rödl [58] and Gowers [68], and by Rödl and Skokan [106] for general r. For the related development, see Nagle, Rödl, and Schacht [99], Rödl and Schacht [105], and Gowers [69].

First we provide the definitions of the necessary notions. We fix $r \geq 1$. For a finite set X, $\delta \geq 0$, and $l \geq 1$ we call

$$\mathcal{H} = \{H_k^j\}_{\substack{k=1,\dots,r \\ j=1,\dots,l}}$$

a δ-equitable l-hyperpartition of $K_r(X)$, if for every $1 \leq k \leq r$ the sets $\{H_k^j\}_{j=1,\dots,l}$ form a partition of $K_k(X)$ with symmetric classes such that $\big||H_k^i| - |H_k^j|\big| \leq \delta |K_k(X)|$ for $i, j \in [l]$. For an l-hyperpartition $\mathcal{H}$ we can define $\mathcal{H}$-cells that are classes of a certain partition of $K_r(X)$. For each $\kappa \colon \mathfrak{h}([r]) \to [l]$ we define the set $C_\kappa = \cap_{e \in \mathfrak{h}([r])} P_e^{-1}(H_{|e|}^{\kappa(e)})$ as the intersection of cylinder sets induced by certain classes of the hyperpartition $\mathcal{H}$. We call the union over the S_r-orbits $B_\kappa = \cup_{\pi \in S_r} C_{\kappa \circ \pi^*}$ $\mathcal{H}$-cells.

Further, for $\varepsilon \geq 0$ we call a k-uniform hypergraph G represented by a symmetric subset of $K_k(X)$ ε-regular, if for any k-collection of $(k-1)$-uniform hypergraphs $F_1, \dots, F_k$ considered as subsets of $K_{k-1}(X)$ with the k-uniform cylinder intersection

$L = \cap_{i\in[k]} P^{-1}_{[k]\setminus\{i\}}(F_i)$ we have

$$\left|\frac{|G|}{|K_k(X)|} - \frac{|G \cap L|}{|L|}\right| \le \varepsilon, \tag{5.21}$$

whenever $|L| \ge \varepsilon |K_k(X)|$.

In this section we consider the target space $\mathcal{K}$ to be a compact metric space with the distance d, and diameter at most 1 instead of the more general compact Polish spaces we studied in Section 5.4. We next give the definition of the metric we will employ on $\mathcal{P}(\mathcal{K})$ in order to measure the error of a regular approximation.

Definition 5.5.1. *Let $(\mathcal{K}, d)$ be a compact metric space of diameter at most 1. Then the map $d_W \colon \mathcal{P}(\mathcal{K}) \times \mathcal{P}(\mathcal{K}) \to \mathbb{R}^+ \cup \{0\}$ with*

$$d_W(\mu, \nu) = \sup_{f\colon \mathcal{K}\to[0,1]\in \mathrm{Lip}_1(\mathcal{K})} |\langle \mu, f\rangle - \langle \nu, f\rangle|$$

is a metric on $\mathcal{P}(\mathcal{K})$ and is called the 1-Wasserstein metric, where $\mathrm{Lip}_1(\mathcal{K})$ denotes the set of Lipschitz-continuous functionals on $\mathcal{K}$ in the metric d with Lipschitz constant at most 1.

We call a set $T_\varepsilon \subset \mathrm{Lip}_1(\mathcal{K})$ of $[0,1]$-valued 1-Lipschitz functions on $\mathcal{K}$ an ε-net if for every $[0,1]$-valued 1-Lipschitz function g there is an $f \in T_\varepsilon$ so that $\|f - g\|_\infty \le \varepsilon$. For any ε-net T_ε we have for every pair $\mu, \nu \in \mathcal{P}(\mathcal{K})$ that

$$\left| d_W(\mu, \nu) - \sup_{f \in T_\varepsilon} |\langle \mu, f\rangle - \langle \nu, f\rangle| \right| \le \varepsilon. \tag{5.22}$$

Moreover, the Arzelà–Ascoli theorem for compact Polish spaces implies that there exists for every $\varepsilon > 0$ a finite ε-net T_ε.

We have introduced the necessary terminology to be able to state our version of the Regularity Lemma next.

Theorem 5.5.2 (Colored hypergraph regularity lemma)**.** *Let $r \ge 1$ and $(\mathcal{K}, d)$ be a compact metric space. Then for any $\varepsilon > 0$ and $F \colon \mathbb{N} \to (0,1)$ there exist $c = c(\varepsilon, F)$ and $N_0 = N_0(\varepsilon, F)$ so that for every $(\mathcal{K}, r)$-graph $G \colon K_r(X) \to \mathcal{K}$ that is symmetric with $|X| \ge N_0$ there is an $F(l)$-equitable l-hyperpartition $\mathcal{H} = \{H^j_k\}_{\substack{k=1,\dots,r\\ j=1,\dots,l}}$ of $K_r(X)$ for some $l \le c$ such that each H^j_k is $F(l)$-regular, and there exists a symmetric map $Q \colon K_r(X) \to \mathcal{P}(\mathcal{K})$ that is constant on $\mathcal{H}$-cells such that*

$$\left| \left\{ x \in K_r(X) \mid d_W(\delta_{G(x)}, Q(x)) \ge \varepsilon \right\} \right| \le \varepsilon |K_r(X)|. \tag{5.23}$$

We outline the corresponding notions for ultraproduct measure spaces. Let now $r \ge 1$ and $(\mathbf{X}^r, \mathcal{B}_{[r]}, \mu_{[r]})$ be an arbitrary ultraproduct space, let $K_r(\mathbf{X})$ denote the set of elements of $\mathbf{X}^r$ without repetitions. Trivially $K_r(\mathbf{X}) \in \mathcal{B}_{[r]}$ and $\mu_{[r]}(K_r(\mathbf{X})) = 1$. Then a

δ-equitable l-hyperpartition of $K_r(\mathbf{X})$ is similar to the finite case a collection of partitions

$$\hat{\mathcal{H}} = \{\mathbf{H}_k^j\}_{\substack{k=1,\dots,r \\ j=1,\dots,l}}$$

with symmetric classes such that $|\mu_{[k]}(\mathbf{H}_k^i) - \mu_{[k]}(\mathbf{H}_k^j)| \leq \delta$ for $i, j \in [l]$, and $k \in [r]$. The definition of $\hat{\mathcal{H}}$-cells is identical to the finite case.

Note that any hyperpartition of $K_k(\mathbf{X})$ can be extended to one of $\mathbf{X}^k$ by adding a zero measure set to the collection of partition classes.

Next we state and prove the version of the hypergraph regularity lemma for $\mathcal{P}(\mathcal{K})$-valued ultraproduct functions, this will be followed by the reduction of the finitary lemma to this one. Both proofs will follow the approach in [49] and in [93], but require some new ideas.

Theorem 5.5.3 (Infinitary weighted regularity lemma)**.** *Let $r \geq 1$, $(\mathbf{X}^r, \mathcal{B}_{[r]}, \mu_{[r]})$ be an ultraproduct measure space, and $(\mathcal{K}, d)$ a compact metric space. For any $\varepsilon > 0$ and any $\sigma([r])$-measurable $\mathbf{G}\colon \mathbf{X}^r \to \mathcal{P}(\mathcal{K})$ there exist a positive integer l, a 0-equitable l-hyperpartition $\hat{\mathcal{H}}$ of $K_r(\mathbf{X})$ such that the classes $\mathbf{H}_k^j$, are independent from $\sigma([k])^*$ for any $1 \leq k \leq r$ and $1 \leq j \leq l$, and a measurable function $\mathbf{T}\colon \mathbf{X}^r \to \mathcal{P}(\mathcal{K})$ that is constant on $\hat{\mathcal{H}}$-cells satisfying*

$$\mu_{[r]}(\{\, \mathbf{x} \in K_r(\mathbf{X}) \mid d_W(\mathbf{G}(\mathbf{x}), \mathbf{T}(\mathbf{x})) \geq \varepsilon \,\}) \leq \varepsilon. \tag{5.24}$$

Proof. Fix an arbitrary $\varepsilon > 0$ and $\mathbf{G}$. By Lemma 5.4.10 there exists a separable realization $\phi\colon \mathbf{X}^r \to [0,1]^{\mathfrak{h}([r])}$ and measurable function $W\colon [0,1]^{\mathfrak{h}([r])} \to \mathcal{P}(\mathcal{K})$ such that $\mathbf{G} = W \circ \phi$ almost everywhere. Let $T_{\varepsilon/2} = \{f_1, \dots, f_t\}$ an $\varepsilon/2$-net in the set of $[0,1]$-valued functions in $\mathrm{Lip}_1(\mathcal{K})$ in the $\|.\|_\infty$-norm. Since $\mathrm{Lip}_1(\mathcal{K}) \subset C(\mathcal{K})$ we have for each $i \in [t]$ that $\langle \mathbf{G}, f_i \rangle = \langle W \circ \phi, f_i \rangle$ almost everywhere. Let

$$C = \{\, y \in \mathbb{R}^t \mid \text{ there exist a } \mu \in \mathcal{P}(\mathcal{K}) \text{ such that } y_i = \langle \mu, f_i \rangle \text{ for all } i \in [t] \,\}. \tag{5.25}$$

It is trivial that C is a compact convex subset of $[0,1]^t$, therefore there is a partition of C into a finite number of non-empty parts $S_1, \dots, S_m$ such that the diameter in d_W of each S_j is at most $\varepsilon/2$. Let $s_j \in S_j$ be an arbitrary element for each $j \in [m]$ and let μ_j be a corresponding probability measure, that is $(s_j)_i = \langle \mu_j, f_i \rangle$ for each $i \in [t]$.

Let $\hat{W}\colon [0,1]^{\mathfrak{h}([r])} \to C$ such that $(\hat{W}(x))_i = \langle W, f_i \rangle$ for each $i \in [t]$. Let $L_j = \hat{W}^{-1}(S_j)$ for each $j \in [m]$, then $(L_1, \dots, L_m)$ is a partition of $[0,1]^{\mathfrak{h}([r])}$ into r-symmetric measurable parts. It follows by measurability that there exists an integer l such that there are symmetric measurable sets $(M_1, \dots, M_m)$ that constitute a partition of $[0,1]^{\mathfrak{h}([r])}$ satisfying the following conditions. Each M_j is the union of l-boxes of the form $\bigtimes_{e \in \mathfrak{h}([r])} \left[\frac{j_e - 1}{l}, \frac{j_e}{l}\right)$ for some $j_e \in [l]$ for all $e \in \mathfrak{h}([r])$ and $\sum_{j=1}^m \lambda(L_j \triangle M_j) \leq \varepsilon$.

We define $U\colon [0,1]^{\mathfrak{h}([r])} \to \mathcal{P}(\mathcal{K})$ to be equal to μ_j on M_j. Except for the set $N = \cup_{j=1}^m (L_j \triangle M_j)$ we have for each $x \in [0,1]^{\mathfrak{h}([r])}$ and $i \in [t]$ that $|\langle W(x), f_i \rangle - \langle U(x), f_i \rangle| \leq \varepsilon/2$, so by (5.22) we have that $d_W(W(x), U(x)) \leq \varepsilon$ for each x not in N.

Now we define the hyperpartition $\hat{\mathcal{H}}$ on $K_r(\mathbf{X})$. Let

$$\mathbf{H}_k^j = \phi_{[k]}^{-1}\left(\left[\frac{j-1}{l}, \frac{j}{l}\right)\right) \setminus K_k(\mathbf{X})$$

for $k \in [r]$ and $j \in [l]$. Note that the $\mathbf{H}_k^j$ sets are independent from $\sigma([k])^*$, as ϕ is a separable realization. We are ready to construct the approximating function $\mathbf{T}$ that is constant on the $\hat{\mathcal{H}}$-cells: simply set $\mathbf{T} = U \circ \phi$.

It remains to show that $\mathbf{T}$ is a suitable approximation of $\mathbf{G}$. If $\mathbf{x} \in K_r(\mathbf{X}) \setminus \phi^{-1}(N)$, then by the definition of $\mathbf{T}$ and the properties of W and U we have $d_W(\mathbf{G}(\mathbf{x}), \mathbf{T}(\mathbf{x})) \leq \varepsilon$. On the other hand, as ϕ is measure-preserving we have $\mu_{[r]}(\phi^{-1}(N)) \leq \varepsilon$, which concludes the proof.

□

Now turn to the proof of the main theorem of the section.

Proof of Theorem 5.5.2. Suppose that the statement of the theorem is false. Then there exists an $\varepsilon > 0$ and an $F\colon \mathbb{N} \to (0,1)$ so that there is a sequence of r-uniform hypergraphs G_i on X_i with $\{|X_i|\}_{i=1}^{\infty}$ strictly monotone increasing such that G_i has no $F(s)$-equitable s-hyperpartition for any $s \leq i$ satisfying the desired conditions. Let $\mathbf{X}^r$ be the ultraproduct space of the sets $\{X_i\}_{i=1}^{\infty}$, and $\mathbf{G}\colon \mathbf{X}^r \to \mathcal{P}(\mathcal{K})$ denote the ultraproduct of the G_i functions and let us apply the infinitary regularity lemma, Theorem 5.5.3. Then there exists a 0-equitable l-hyperpartition $\hat{\mathcal{H}}$ of $K_r(\mathbf{X})$ such that the classes $\mathbf{H}_k^j$ are independent from $\sigma([k])^*$ for any $1 \leq k \leq r$ and $1 \leq j \leq l$ and there exists a measurable function $\mathbf{T}\colon \mathbf{X}^r \to \mathcal{P}(\mathcal{K})$ that is constant on $\hat{\mathcal{H}}$-cells, and for

$$\mathbf{N} = \{\, \mathbf{x} \in K_r(\mathbf{X}) \mid d_W(\mathbf{G}(\mathbf{x}), \mathbf{T}(\mathbf{x})) \geq \varepsilon/2 \,\} \tag{5.26}$$

we have

$$\mu_{[r]}(\mathbf{N}) \leq \varepsilon/2. \tag{5.27}$$

For each $k \in [r]$, it follows from the definition of the measure space $(\mathbf{X}^k, \mathcal{B}_{[k]}, \mu_{[k]})$, that there exist sets $H_{k,i}^j \subset K_k(X_i)$ such that $\{H_{k,i}^j\}_{j=1,\dots,l}$ form for ω-almost all i an $F(l)$-equitable l-partition of $K_k(X_i)$, and $\mu_{[k]}(\mathbf{H}_k^j \triangle [\{H_{k,i}^j\}_{i=1}^{\infty}]) = 0$, for these indices let $\mathcal{H}_i$ denote the hyperpartition $\{H_{k,i}^j\}_{\substack{k=1,\dots,r\\ j=1,\dots,l}}$.

We will only deal with the indices above where we defined hyperpartitions in detail, for all the remaining ones let $Q_i\colon K_r(X_i) \to \mathcal{P}(\mathcal{K})$ be arbitrary measurable symmetric functions. Now for the important indices, whose set is in ω, let for i the function value $Q_i(x_i)$ be defined as the value of $\mathbf{T}$ on the $\hat{\mathcal{H}}$-cell

$$\mathbf{B}_\kappa = \cup_{\pi \in S_r} \cap_{e \in \mathfrak{h}([r])} P_e^{-1}(\mathbf{H}_{|e|}^{\kappa(\pi^*(e))}) \tag{5.28}$$

for some $\kappa\colon \mathfrak{h}([r]) \to [l]$, whenever

$$x_i \in \cup_{\pi \in S_r} \cap_{e \in \mathfrak{h}([r])} P_e^{-1}(H_{|e|,i}^{\kappa(\pi^*(e))}). \tag{5.29}$$

Note that $\mathbf{T} = \lim_\omega Q_i$ $\mu_{[r]}$-almost everywhere.

It is trivial now that for almost all i we have that Q_i is a measurable function constant on $\mathcal{H}_i$-cells. Further, let $T_{\varepsilon/4} = \{f_1, \dots, f_t\}$ be an $\varepsilon/4$-net in the set of $[0,1]$-valued functions in $\mathrm{Lip}_1(\mathcal{K})$ in the $\|.\|_\infty$-norm. Define

$$N_i = \{\, x_i \in K_r(X_i) \mid d_W(\delta_{G_i(x_i)}, Q_i(x_i)) \geq \varepsilon \,\},$$

$$N_i' = \{\, x_i \in K_r(X_i) \mid \max_{j \in [t]} |f_j(G_i(x_i)) - \langle Q_i(x_i), f_j \rangle| \geq \frac{\varepsilon}{2} \,\},$$

and

$$\mathbf{N}' = \{\, \mathbf{x} \in K_r(\mathbf{X}) \mid \max_{j \in [t]} |\langle \mathbf{G}(\mathbf{x}), f_j \rangle - \langle \mathbf{T}(\mathbf{x}), f_j \rangle| \geq \frac{\varepsilon}{2} \,\}.$$

Since $\mathbf{G} - \mathbf{T} = \lim_\omega \delta_{G_i} - Q_i$ almost everywhere in the sense that for every $f \in C(\mathcal{K})$ we have

$$\langle \mathbf{G}(\mathbf{x}), f \rangle - \langle \mathbf{T}(\mathbf{x}), f \rangle = \lim_\omega \langle \delta_{G_i(x_i)}, f \rangle - \langle Q_i(x_i), f \rangle,$$

and ω is closed under finite intersection, it holds by (5.22) that

$$\mathbf{N} \supset \mathbf{N}' \supset [\{N_i'\}_{n=1}^\infty] \supset [\{N_i\}_{n=1}^\infty].$$

Therefore $\lim_\omega \frac{|N_i|}{|X_i|^r} \leq \mu_{[r]}(\mathbf{N}) \leq \varepsilon/2$, and consequently $|N_i| \leq \varepsilon |K_r(X_i)|$ for ω-almost all i.

It follows from our assumption that for ω-almost all i there are $k \in [r]$ and $j \in [l]$ such that $H_{k,i}^j$ is not ε-regular, hence there are $k_0 \in [r]$ and $j_0 \in [l]$ such that $H_{k_0,i}^{j_0}$ is not ε-regular for ω-almost all i. This means that for almost all i there are k_0-uniform cylinder intersections $L_i \subset K_{k_0}(X_i)$ such that

$$\left| \frac{|H_{k_0,i}^{j_0}|}{|K_{k_0}(X_i)|} - \frac{|H_{k_0,i}^{j_0} \cap L_i|}{|L_i|} \right| \geq \varepsilon, \tag{5.30}$$

and $|L_i| \geq \varepsilon |K_{k_0}(X_i)|$.

Let $\mathbf{L} = [\{L_i\}_{i=1}^\infty]$. We know on one hand that $\mathbf{L}$ is $\sigma([k_0])^*$-measurable and $\mu_{[k_0]}(\mathbf{L}) = \lim_\omega \frac{|L_i|}{|K_{k_0}(X_i)|} \geq \varepsilon$, and on the other that $\mathbf{H}_{k_0}^{j_0} = [\{H_{k_0,i}^{j_0}\}_{i=1}^\infty]$ and $\mathbf{H}_{k_0}^{j_0}$ is independent of $\sigma([k_0])^*$, so

$$\mu_{[k_0]}(\mathbf{H}_{k_0}^{j_0}) \mu_{[k_0]}(\mathbf{L}) = \mu_{[k_0]}(\mathbf{H}_{k_0}^{j_0} \cap \mathbf{L}). \tag{5.31}$$

However, (5.30) implies that $|\mu_{[k_0]}(\mathbf{H}_{k_0}^{j_0})\mu_{[k_0]}(\mathbf{L}) - \mu_{[k_0]}(\mathbf{H}_{k_0}^{j_0} \cap \mathbf{L})| \geq \varepsilon$, which together with $\mu_{[k_0]}(\mathbf{L}) > 0$ leads to a contradiction, which concludes the proof.

□

We illustrate the relation of Theorem 5.5.2 to versions of the Regularity Lemma in the uncolored case by giving a corollary that is analogous to Lemma 3.3.2 and handles the case $r = 2$.

Definition 5.5.4. *Let G be an undirected $(\mathcal{K}, 2)$-graph, and A and B subsets of its vertex set, and let $\varepsilon \geq 0$. We call the pair (A, B) $(\mathcal{K}, \varepsilon)$-regular, if they are disjoint and for any $A' \subset A$ and $B' \subset B$ with $|A'| \geq \varepsilon|A|$ and $|B'| \geq \varepsilon|B|$ it holds that*

$$d_W\left(\frac{\sum_{\substack{a\in A\\ b\in B}} \delta_{G(a,b)}}{|A||B|}, \frac{\sum_{\substack{a\in A'\\ b\in B'}} \delta_{G(a,b)}}{|A'||B'|}\right) \leq \varepsilon. \tag{5.32}$$

With other words, a pair (A, B) is regular, if the empirical distribution of elements of $\mathcal{K}$ given by their presence on the edges between large subsets $A' \subset A$, $B' \subset B$ is as expected, namely close to the empirical distribution taken on all edges running between A and B. Now a special case of Theorem 5.5.2 reads almost identically as Lemma 3.3.2.

Corollary 5.5.5. *For any $\varepsilon > 0$ there exists a positive integer $\frac{1}{\varepsilon} \leq M(\varepsilon)$ such that the following holds. For any undirected $(\mathcal{K}, 2)$-graph G there exists a partition of $V(G)$ into at most $m \leq M(\varepsilon)$ parts $(V_1, \ldots, V_m)$ such that*

(i) for any $i, j \in [m]$ we have $\big||V_i| - |V_j|\big| \leq 1$, and

(ii) at least $(1 - \varepsilon)m^2$ of the pairs $(i, j) \in [m]^2$ satisfy that (V_i, V_j) is $(\mathcal{K}, \varepsilon)$-regular.

CHAPTER 6

Complexity of Nondeterministic Testing

6.1 Introduction and results

Graph and hypergraph parameters are real-valued functions defined on the space of uniform hypergraphs of some given rank invariant under relabeling the vertex set. Testing a parameter value associated to an instance in the dense model means to produce an estimation by only having access to a small portion of the data that describes it. The test data is provided as a uniform random subset of the vertex set with the induced substructure of the hypergraph on this subset exposed. A certain parameter is said to be testable if for every given tolerated error the estimation is within the error range of the parameter value with high probability, and the size of the selected random subset does only depend on the size of this permitted error and not on the size of the instance, for precise definitions see Definition 3.2.1 below. Similar notions apply to testing graph properties, in that situation one also uses uniform sampling in order to separate the cases where an instance has the property or is far from having it, where the distance is measured by the number of edge modifications required, see Definition 3.2.2. The development in this direction resulted in a number of randomized sub-linear time algorithms for the corresponding decision problems, see [13], [14], [66], for the background in approximation theory of NP-hard problems for dense structures, see [18].

Several attempts were made to characterize the parameters in terms of the sample size that is needed for carrying out the above task of testing. The notion nondeterministic testability was introduced by Lovász and Vesztergombi [97] in the framework of graph property testing, and encompasses an a priori weaker characteristic than the original testability. The authors of [97] defined a certain property to be nondeterministically testable if there exists another property of colored (edge or node) graphs that is testable in the traditional sense and serves as a certificate for the original.

Definition 6.1.1. *[97] The simple graph property $\mathcal{P}$ is non-deterministically testable if there exist integers $k \geq m$ and a testable property $\mathcal{Q}$ of k-colored directed graphs called witness such that a simple graph G has exactly in that case $\mathcal{P}$, when there exists a $\mathbf{G} \in \mathcal{Q}$ that is a (k, m)-coloring of G. The edge-k-colored directed graph $\mathbf{G}$ is a (k, m)-coloring of G, if after erasing all edges of $\mathbf{G}$ colored with an element of $[m+1, \ldots, k]$ and discarding the orientation, coloring, and multiplicity of the remaining edges we end up with G. We say in this case that G is the shadow of $\mathbf{G}$.*

Definition 6.1.2. *The graph parameter f is non-deterministically testable if there exist integers $k \geq m$ and a testable k-colored directed graph parameter g called witness such that for any simple graph G the value $f(G) = \max_{\mathbf{G}} g(\mathbf{G})$ where the maximum goes over the set of (k, m)-colorings of G.*

The corresponding definition for r-uniform hypergraphs (in short, r-graphs) is analogous. The choice of maximizing over the g-values in Definition 6.1.2 is somewhat arbitrary, in a more general sense we could have $f(G) = g(\text{argmax}_{\mathbf{G}} L(g(\mathbf{G})))$ for an α-Hölder continuous function L from $\mathbb{R}$ to $\mathbb{R}$. Also, stronger formulations of being a witness can be employed, such as permitting only undirected instances or imposing $k = 2m$.

The problem regarding the relationship of the class of parameters that are testable and those who are non-deterministically testable was first studied in the framework of dense graph limits and property testing by Lovász and Vesztergombi [97] in the spirit of the general "P vs. NP" question, that is a central problem in theoretical computer science. Using the particular notion of nondeterminism above they were able to prove that any non-deterministically testable graph property is also testable, which implies the analogous statement for parameters.

Theorem 6.1.3. *[97] Every non-deterministically testable graph property $\mathcal{P}$ is testable. The same equivalence holds for parameter testing.*

However, no explicit relationship was provided between on one hand, the sample size required for estimating the f value, and on the other, the two factors, the number of colors k and m, and the sample complexity of the witness g. The reason for the non-efficient characteristic of the result is that the authors exploited various consequences of the next remarkable fact.

Recall that graphons are bounded symmetric measurable functions on the unit square, their cut norm $\|.\|_\square$ given in Definition 3.3.6. At this point we wish to stress that it is weaker than the L^1-norm, and the $\delta_\square$-distance induced by it has a compact unit ball, this is not the case for the δ_1-distance generated by the $\|.\|_1$-norm.

Fact. If $(W_n)_{n\geq 1}$ is a sequence of graphons and $\|W_n\|_\square \to 0$ when n tends to infinity, then for any measurable function $Z\colon [0,1]^2 \to [-1,1]$ it is true that $\|W_n Z\|_\square \to 0$, where the product is taken point-wise.

Although the above statement is true for all Z, the convergence is not uniform and its rate depends heavily on the structure of Z.

The relationship of the magnitude of the sample complexity of a testable property $\mathcal{P}$ and its non-deterministic certificate $\mathcal{Q}$ was analyzed by Gishboliner and Shapira

[61] relying on Szemerédi's Regularity Lemma and its connections to graph property testing unveiled by Alon, Fischer, Newman, and Shapira [16]. In the upper bound given in [61] the height of the exponential tower was not bounded and growing as a function of the inverse of the accuracy $1/\varepsilon$.

Theorem 6.1.4. *[61] Every non-deterministically testable graph property $\mathcal{P}$ is testable. If the sample complexity of the witness property $\mathcal{Q}$ for each $\varepsilon > 0$ is $q_{\mathcal{Q}}(\varepsilon)$, then the sample complexity of $\mathcal{P}$ for each $\varepsilon > 0$ is at most $\mathrm{tf}(cq_{\mathcal{Q}}(\varepsilon/2))$ for some universal constant $c > 0$, where $\mathrm{tf}(t)$ is the exponential tower of twos of height t.*

In the current chapter we improve on the result of [61] in terms of parameters first and subsequently in terms of properties by using a weaker type of regularity approach which eliminates the tower-type dependence on the sample complexity of the witness parameter. The function $\exp^{(t)}$ stands for the t-fold iteration of the exponential function ($\exp^{(0)}(x) = x$).

Theorem 6.1.5. *Let f be a nondeterministically testable simple graph parameter with witness parameter g of k-colored digraphs, and let the corresponding sample complexity be q_g. Then f is testable with sample complexity q_f, and there exists a constant $c > 0$ only depending on k but not on f or g such that for any $\varepsilon > 0$ the inequality $q_f(\varepsilon) \leq \exp^{(3)}(cq_g^2(\varepsilon/2))$ holds.*

Section 6.2 is devoted to the proof of Theorem 6.1.5.

The previous works mentioned above dealt with graphs, it was asked in [97] if the concept can be employed for hypergraphs. The notion of an r-uniform hypergraph (in short, r-graph) parameter and its testability can be defined completely analogously to the graph case, the same applies for nondeterministic testability. Naturally, first the question arises whether the deterministic and the nondeterministic testability are equivalent for higher rank hypergraphs, and secondly, if the answer to the first question is positive, then what can be said about the relationship of the sample complexity of the parameter and that of its witness parameter. The statements that are analogous to the main results of [61], and [97], as well as to Theorem 6.1.5 do not follow immediately for uniform hypergraphs of higher rank from the proof for graphs, like-wise to the generalizations of the hypergraph version of the Regularity Lemma new tools and notions are required to handle these cases. In the current chapter we prove the equivalence of the two testability notions for uniform hypergraphs of higher rank and make progress for both questions posed above.

Theorem 6.1.6. *Every non-deterministically testable r-graph parameter f is testable. If g is the parameter of k-edge-colored r-graphs that certifies the testability of f, then $q_f(\varepsilon) \leq \exp^{(4(r-1)+1)}(c_{r,k}q_g(\varepsilon)/\varepsilon)$ for some constant $c_{r,k} > 0$ depending only on r and k, but not on f or g. Here $\exp^{(t)}$ denotes the t-fold iteration of the exponential function for $t \geq 1$, and $\exp^{(0)}$ is the identity function.*

We describe the proof of Theorem 6.1.6 in Section 6.5. We will need a result concerning the testability of the r-cut norm that is related to the Gowers norm of functions, see Definition 3.3.20, we deal with this in Section 6.4.

We also show that testing nondeterministically testable properties is as hard as parameter testing with our method in the sense that the same complexity bounds apply.

Theorem 6.1.7. *Every nondeterministically testable r-graph property is testable, the sample complexity dependence is the same as in the parameter testing case.*

Section 6.6 deals with the proof of Theorem 6.1.7.

A $2k$-coloring of a graph is a $(2k, m)$-coloring with $m = k$. This technical restriction facilitates our proofs for r-graphs. Also, for higher rank hypergraphs we only deal with witnesses that are undirected for simplicity, but we believe that the case of employing general directed witnesses as in the graph case can be dealt with analogously.

Further, we can derive significantly better bounds, if we use more restrictive notions of nondeterministic testing, such as weakly non-deterministic testing, see Section 6.3, and the case of witnesses that only depend on densities of linear subgraphs, see Section 6.7.

We conclude the chapter by presenting some applications of the main theorems above in Section 6.8, among other results we derive an effective proof for Theorem 5.3.4 from Chapter 5.

6.2 Graph parameter testing

We will exploit the continuity of a testable graph parameter with respect to the cut norm and distance, and the connection of this characteristic to the sample complexity of the parameter. We require two results, the first one quantifies the above continuity. We generally assume that the sample complexity satisfies $q_g(\varepsilon) \geq 1/\varepsilon$, also $\varepsilon \leq 1$ and $k \geq 2$.

Lemma 6.2.1. *Let g be a testable k-colored digraph parameter with sample complexity at most q_g. Then for any $\varepsilon > 0$ and two graphs, $\mathbf{G}$ and $\mathbf{H}$, with $|V(\mathbf{G})|, |V(\mathbf{H})| \geq \left(\frac{2q_g^2(\varepsilon/4)}{\varepsilon}\right)^{1/(q_g(\varepsilon/4)-1)}$ satisfying $\delta_\square(\mathbf{G}, \mathbf{H}) \leq k^{-2q_g^2(\varepsilon/4)}$ we have*

$$|g(\mathbf{G}) - g(\mathbf{H})| \leq \varepsilon.$$

Proof. Let $\varepsilon > 0$, $\mathbf{G}$ and $\mathbf{H}$ be as in the statement, and set $q = q_g(\varepsilon/4)$. Then we have

$$\begin{aligned} |g(\mathbf{G}) - g(\mathbf{H})| \leq & |g(\mathbf{G}) - g(\mathbb{G}(q, \mathbf{G}))| + |g(\mathbb{G}(q, \mathbf{W_G})) - g(\mathbb{G}(q, \mathbf{G}))| \\ & + |g(\mathbb{G}(q, \mathbf{W_G})) - g(\mathbb{G}(q, \mathbf{W_H}))| + |g(\mathbb{G}(q, \mathbf{H})) - g(\mathbb{G}(q, \mathbf{W_H}))| \\ & + |g(\mathbf{H}) - g(\mathbb{G}(q, \mathbf{H}))|. \end{aligned} \tag{6.1}$$

The first and the last term on the right of (6.1) can be each upper bounded by $\varepsilon/4$ with cumulative failure probability $\varepsilon/2$ due to the assumptions of the lemma. To deal with the second term we require the fact that $\mathbb{G}(q, \mathbf{G})$ and $\mathbb{G}(q, \mathbf{W_G})$ have the same distribution conditioned on the event that the X_i variables that define $\mathbb{G}(q, \mathbf{W_G})$ lie in different classes of the canonical equiv-partition of $[0,1]$ into $|V(\mathbf{G})|$ classes. The failure probability of the latter event can be upper bounded by $q^2/2|V(\mathbf{G})|^{q-1}$, which is at most $\varepsilon/4$, analogously for the fourth term. Until this point we have not dealt with the relationship of the two random objects $\mathbb{G}(q, \mathbf{W_G})$ and $\mathbb{G}(q, \mathbf{W_H})$, therefore the above discussion is valid for every coupling of them.

In order to handle the third term we upper bound the probability that the two random graphs are different by means of an appropriate coupling, since clearly in the event of identity the third term of (6.1) vanishes. More precisely, we will show that $\mathbb{G}(q, \mathbf{W_G})$ and $\mathbb{G}(q, \mathbf{W_H})$ can be coupled in such a way that $\mathbb{P}(\mathbb{G}(q, \mathbf{W_G}) \neq \mathbb{G}(q, \mathbf{W_H})) < 1 - \varepsilon$. We utilize that for a fixed k-colored digraph $\mathbf{F}$ on q vertices we can upper bound the deviation of the subgraph densities of $\mathbf{F}$ in $\mathbf{G}$ and $\mathbf{H}$ through the cut distance of these graphs, see Lemma 3.3.7. In particular,

$$|\mathbb{P}(\mathbb{G}(q, \mathbf{W_G}) = \mathbf{F}) - \mathbb{P}(\mathbb{G}(q, \mathbf{W_H}) = \mathbf{F})| \leq \binom{q}{2} \delta_\square(\mathbf{W_G}, \mathbf{W_H}).$$

Therefore in our case

$$\sum_{\mathbf{F}} |\mathbb{P}(\mathbb{G}(q, \mathbf{W_G}) = \mathbf{F}) - \mathbb{P}(\mathbb{G}(q, \mathbf{W_H}) = \mathbf{F})| \leq k^{2\binom{q}{2}} \binom{q}{2} k^{-2q^2} \leq \varepsilon,$$

where the sum goes over all labeled k-colored digraphs $\mathbf{F}$ on q vertices.

Since there are only finitely many possible target graphs for the random objects, we can couple $\mathbb{G}(q, \mathbf{W_G})$ and $\mathbb{G}(q, \mathbf{W_H})$ so that in the end we have $\mathbb{P}(\mathbb{G}(q, \mathbf{W_G}) \neq \mathbb{G}(q, \mathbf{W_H})) \leq \varepsilon$, see Corollary 3.3.25 for details. This implies that with positive probability (in fact, with at least $1 - 2\varepsilon$) the sum of the five terms on the right hand side of (6.1) does not exceed ε, so the statement of the lemma follows.

□

We will also require the following statement which can be regarded as the quantitative counterpart of Lemma 3.2 from [97]. It clarifies why the cut-$\mathcal{P}$-norm, (Definition 3.3.12, Definition 3.3.13) and the need for the accompanying regularity lemma, Lemma 3.3.15, are essential for our intent.

Lemma 6.2.2. *Let $k \geq 2$, $\varepsilon > 0$, U be a step function with steps $\mathcal{P} = (P_1, \ldots, P_t)$ and V be a graphon with $\|U - V\|_{\square\mathcal{P}} \leq \varepsilon$. For any k-colored digraphon $\mathbf{U} = (U^{(1,1)}, \ldots, U^{(k,k)})$ that is a step function with steps from $\mathcal{P}$ and a (k,m)-coloring of U there exists a (k,m)-coloring $\mathbf{V} = (V^{(1,1)}, \ldots, V^{(k,k)})$ of V so that $\|\mathbf{U} - \mathbf{V}\|_\square = \sum_{\alpha,\beta=1}^{k} \|U^{(\alpha,\beta)} - V^{(\alpha,\beta)}\|_\square \leq k^2\varepsilon$.*

If $V = W_G$ for a simple graph G on $n \geq 16/\varepsilon^2$ nodes and $\mathcal{P}$ is an $\mathcal{I}_n$-partition of $[0,1]$ *then there is a (k,m)-coloring $\mathbf{G}$ of G that satisfies the above conditions and $\|\mathbf{U} - \mathbf{W_G}\|_\square \leq 2k^2\varepsilon$.*

Proof. Fix $\varepsilon > 0$, and let U, V, and $\mathbf{U}$ be as in the statement of the lemma. Then $\sum_{\alpha,\beta=1}^{k} U^{(\alpha,\beta)} = 1$, let M be the subset of $[k]^2$ such that its elements have at least one component that is at most m so we have $\sum_{(\alpha,\beta)\in M} U^{(\alpha,\beta)} = U$ by definition. For $(\alpha,\beta) \in M$ set $V^{(\alpha,\beta)} = \frac{VU^{(\alpha,\beta)}}{U}$ on the set where $U > 0$ and $V^{(\alpha,\beta)} = \frac{V}{k^2-(k-m)^2}$ where $U = 0$, furthermore for $(\alpha,\beta) \notin M$ set $V^{(\alpha,\beta)} = \frac{(1-V)U^{(\alpha,\beta)}}{1-U}$ on the set where $U < 1$ and $V^{(\alpha,\beta)} = \frac{1-V}{(k-m)^2}$ where $U = 1$. We will show that the k-colored digraphon $\mathbf{V}$ defined this way satisfies the conditions, in particular for each $(\alpha,\beta) \in [k]^2$ we have $\|U^{(\alpha,\beta)} - V^{(\alpha,\beta)}\|_\square \le \varepsilon$. We will explicitly perform the calculation only for $(\alpha,\beta) \in M$, the other case is analogous. Fix $S, T \subset [0,1]$, then

$$\begin{aligned}\left|\int_{S\times T} U^{(\alpha,\beta)} - V^{(\alpha,\beta)}\right| &= \left|\int_{S\times T, U>0} U^{(\alpha,\beta)} - V^{(\alpha,\beta)} + \int_{S\times T, U=0} U^{(\alpha,\beta)} - V^{(\alpha,\beta)}\right| \\ &\le \sum_{i,j=1}^{t}\left|\int_{(S\cap P_i)\times(T\cap P_j), U>0} \frac{U^{(\alpha,\beta)}}{U}(U-V) + \int_{(S\cap P_i)\times(T\cap P_j), U=0} \frac{1}{k^2-(k-m)^2}(U-V)\right| \\ &= \sum_{i,j=1}^{t}\left|\int_{(S\cap P_i)\times(T\cap P_j)} (U-V)\left[\mathbb{1}_{U>0}\frac{U^{(\alpha,\beta)}}{U} + \mathbb{1}_{U=0}\frac{1}{k^2-(k-m)^2}\right]\right| \\ &\le \sum_{i,j=1}^{t}\left|\int_{(S\cap P_i)\times(T\cap P_j)} (U-V)\right| \\ &= \|U-V\|_{\square\mathcal{P}} \le \varepsilon.\end{aligned}$$

The second inequality is a consequence of $\left[\mathbb{1}_{U>0}\frac{U^{(\alpha,\beta)}}{U} + \mathbb{1}_{U=0}\frac{1}{k^2-(k-m)^2}\right]$ being a constant between 0 and 1 on each of the rectangles $P_i \times P_j$.

We prove now the second statement of the lemma concerning graphs with $V = W_G$ and a partition $\mathcal{P}$ that is an $\mathcal{I}_n$-partition. The general discussion above delivers the existence of $\mathbf{V}$ that is a (k,m)-coloring of W_G, which can be regarded as a fractional coloring of G, as $\mathbf{V}$ is constant on the sets associated with nodes of G. For $|V(G)| = n$ we get for each $ij \in \binom{[n]}{2}$ a probability distribution on $[k]^2$ with $\mathbb{P}\left(Z_{ij} = (\alpha,\beta)\right) = n^2 \int_{[\frac{i-1}{n},\frac{i}{n}]\times[\frac{j-1}{n},\frac{j}{n}]} V^{(\alpha,\beta)}(x,y)\mathrm{d}x\mathrm{d}y$. For each pair ij we make an independent random choice according to this measure, and color (i,j) by the first, and (j,i) by the second component of Z_{ij} to get a proper (k,m)-coloring $\mathbf{G}$ of G. It remains to conduct the analysis of the deviation in the statement of the lemma, we will show that this is small with high probability with respect to the randomization, which in turn implies the existence. We have

$$\begin{aligned}\|\mathbf{U} - \mathbf{W_G}\|_\square &\le \|\mathbf{U} - \mathbf{V}\|_\square + \|\mathbf{V} - \mathbf{W_G}\|_\square \\ &\le k^2\varepsilon + \sum_{\alpha,\beta=1}^{k} \|V^{(\alpha,\beta)} - W_{\mathbf{G}}^{(\alpha,\beta)}\|_\square\end{aligned}$$

For each $(\alpha,\beta) \in [k]^2$ we have that $\mathbb{P}\left(\|V^{(\alpha,\beta)} - W_{\mathbf{G}}^{(\alpha,\beta)}\|_\square \ge 4/\sqrt{n}\right) \le 2^{-n}$, this result is exactly Lemma 4.3 in [30], see also Lemma 4.2.2 for a related result. This implies for $n \geq 16/\varepsilon^2$ the existence of a suitable coloring, which in turn finishes the proof of the lemma.

□

Remark 6.2.3. Actually we can perform the same proof to verify the existence of a k-coloring $\mathbf{V}$ such that $d_{\square\mathcal{P}}(\mathbf{U},\mathbf{V}) \leq k^2\varepsilon$. On the other hand, we can not weaken the condition on the closeness of U and V, a small cut-norm of $U - V$ does not imply the existence of a suitable coloring $\mathbf{V}$, for example in the case when the number of steps of $\mathbf{U}$ is exponential in $1/\|U - V\|_\square$.

We proceed towards the proof of the main statement of the subsection. Before we can outline that we require yet another specific lemma.

Let $\mathcal{M}_{\Delta,n}$ denote the set of $\mathcal{I}_n$-step functions U that have steps $\mathcal{P}_U$ with $|\mathcal{P}_U| \leq t_k(\Delta,1)$ classes, and values between 0 and 1, where t_k is the function from Lemma 3.3.15. In order to verify Theorem 6.1.5 we will condition on the event that is formulated in the following lemma. Recall Definition 3.3.12 for the deviation $d_{W,\mathcal{P}}(V)$.

Lemma 6.2.4. *Let G be a simple graph on n vertices and $\Delta > 0$. Then for $q \geq 2^{2^{(2k^4/\Delta^2)+4}}$ we have*

$$|d_{U,\mathcal{P}_U}(G) - d_{U,\mathcal{P}_U}(\mathbb{G}(q,G))| \leq \Delta, \tag{6.2}$$

for each $U \in \mathcal{M} = \mathcal{M}_{\Delta,n}$ simultaneously with probability at least $1 - \exp(-\frac{\Delta^2 q}{2^7})$, whenever $n \geq 4q/\Delta$.

Proof. Let G and $\Delta > 0$ be arbitrary, and q be such that it satisfies the conditions of the lemma. For technical convenience we assume that n is an integer multiple of q, let us introduce the quantity $t_1 = t_k(\Delta,1) = 2^{2^{\frac{k^4}{\Delta^2}+2}-2}$, and denote $\mathbb{G}(q,G)$ by F. For the case when q is not a divisor of n then we just add at most q isolated vertices to G to achieve the above condition, by this operation $d_{U,\mathcal{P}}(G)$ is changed by at most q/n. Also, we can couple in a way such that $d_{U,\mathcal{P}}(\mathbb{G}(q,G))$ remains unchanged with probability at least $1 - q/n$.

We will show that there exists an $\mathcal{I}_n$-permutation ϕ of $[0,1]$ such that $\|W_G - W_F^\phi\|_{\square\mathcal{Q}} < \Delta$ for any $\mathcal{I}_n$-partition $\mathcal{Q}$ of $[0,1]$ into at most t_1 classes with high probability. Applying Lemma 3.3.15 with the error parameter $\Delta/4$ and $m_0(\Delta) = t_k(\Delta,1)$ for approximating W_G by a step function we can assert that there exists an $\mathcal{I}_n$-partition $\mathcal{P}$ of $[0,1]$ into $t_\mathcal{P}$ classes with $t_\mathcal{P} \leq t_2$ with $t_2 = t_k(\Delta/4, t_k(\Delta,1)) = 2^{2^{(2k^4/\Delta^2)+2}+2^{(k^4/\Delta^2)+1}} \leq 2^{2^{(2k^4/\Delta^2)+3}}$ such that for every $\mathcal{I}_n$-partition $\mathcal{Q}$ into $t_\mathcal{Q}$ classes $t_\mathcal{Q} \leq \max\{t_\mathcal{P}, t_1\}$ it holds that

$$\|W_G - (W_G)_\mathcal{P}\|_{\square\mathcal{Q}} \leq \Delta/4.$$

We only need here

$$\sup_{\mathcal{Q}:t_\mathcal{Q}\leq t_1} \|W_G - (W_G)_\mathcal{P}\|_{\square\mathcal{Q}} \leq \Delta/4. \tag{6.3}$$

This property is by Remark 3.3.14 equivalent to stating that

$$\max_{Q} \max_{A \in \mathbb{A}} \max_{S,T \subset [0,1]} \sum_{i,j=1}^{t_1} A_{i,j} \int_{S \times T} (W_G - (W_G)_{\mathcal{P}})(x,y) \mathbb{1}_{Q_i}(x) \mathbb{1}_{Q_j}(y) \mathrm{d}x \mathrm{d}y \leq \Delta/4, \tag{6.4}$$

where $\mathbb{A}$ is the set of all $t_1 \times t_1$ matrices with -1 or $+1$ entries.

We can reformulate the above expression (6.4) by putting

$$J = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

and defining the tensor product $B_A = A \otimes J$, so that $B_{i,j}^{\alpha,\beta} = A_{ij} J_{\alpha,\beta}$ for each $A \in \mathbb{A}$. The first matrix J corresponds to the $\mathcal{I}_n$-partition $(S \cap T, S \setminus T, T \setminus S, [0,1] \setminus (S \cup T)) = (T_1, T_2, T_3, T_4)$ generated by a pair (S,T) of $\mathcal{I}_n$-sets of $[0,1]$ so that for any function $U \colon [0,1]^2 \to \mathbb{R}$ it holds that

$$\sum_{i,j=1}^{4} J_{ij} \int_{[0,1]^2} U(x,y) \mathbb{1}_{T_i}(x) \mathbb{1}_{T_j} \mathrm{d}x \mathrm{d}y = \int_{S \times T} U(x,y) \mathrm{d}x \mathrm{d}y.$$

It follows that the inequality (6.4) is equivalent to saying

$$\max_{A \in \mathbb{A}} \max_{\hat{Q}} \sum_{i,j=1}^{t_1} \sum_{\alpha,\beta=1}^{4} (B_A)_{i,j}^{\alpha,\beta} \int_{[0,1]^2} (W_G - (W_G)_{\mathcal{P}})(x,y) \mathbb{1}_{Q_i^\alpha}(x) \mathbb{1}_{Q_j^\beta}(y) \mathrm{d}x \mathrm{d}y \leq \Delta/4, \tag{6.5}$$

where the second maximum goes over all $\mathcal{I}_n$-partitions $\hat{Q} = (Q_i^\alpha)_{\substack{i \in [t_1] \\ \alpha \in [4]}}$ into $4t_1$ classes. Let us substitute an arbitrary graphon U for $W_G - (W_G)_{\mathcal{P}}$ in (6.5) and define

$$\hat{h}_{A,\hat{Q}}(U) = \sum_{\substack{1 \leq i,j \leq t_1 \\ 1 \leq \alpha,\beta \leq 4}} (B_A)_{i,j}^{\alpha,\beta} \int_{[0,1]^2} U(x,y) \mathbb{1}_{Q_i^\alpha}(x) \mathbb{1}_{Q_j^\beta}(y) \mathrm{d}x \mathrm{d}y$$

and

$$\hat{h}_A(U) = \max_{\hat{Q}} h_{A,\hat{Q}}(U)$$

as the expression whose optima is sought for a fixed $A \in \mathbb{A}$.

For notational convenience only lower indices will be used when referring to the entries of B_A. We introduce a relaxed version h_A of the above function $\hat{h}_A$ by replacing the requirement on $\hat{Q}$ being an $\mathcal{I}_n$-partition, instead we define

$$h_{A,f}(U) = \sum_{1\le i,j\le 4t_1} (B_A)_{i,j} \int_{[0,1]^2} U(x,y) f_i(x) f_j(y)\mathrm{d}x\mathrm{d}y$$

with $f = (f_i)_{i\in[4t_1]}$ being a fractional $\mathcal{I}_n$-partition into $4t_1$ classes, that is, each component of f is a non-negative $\mathcal{I}_n$-function, and their sum is the constant 1 function. Further, we define $h_A(U) = \max_f h_{A,f}(U)$, where f runs over all fractional $\mathcal{I}_n$-partitions into $4t_1$ parts. It is easy to see that

$$|\hat{h}_A(U) - h_A(U)| \le 1/n,$$

since the two functions coincide when U is 0 on the diagonal blocks. Denote $U' = W_{\mathbb{H}(q,U)}$, where the graphon is given by the increasing order of the sample points $\{\, X_i \mid i \in [q] \,\}$. We wish to upper bound the probability that the deviation $|h_A(U)-h_A(U')|$ exceeds $\Delta/4$, for some $A \in \mathbb{A}$. Similarly as above, $|\hat{h}_A(U') - h_A(U')| \le 1/q$.

We remark, that a simple approach would be using a slight variant of the counting lemma Lemma 3.3.7 that $|h_A(U) - h_A(U')| \le 16t_1^2\delta_\square(U,U')$ together with a version of Lemma 3.3.9 for kernels with perhaps negative values, this way we would have to impose a lower bound on q that is exponential in t_1 in order to satisfy the statement of the lemma. We can do slightly better using more involved methods also employed in Chapter 4.

We require the notion of ground state energies from [32], see Chapter 4.

$$\hat{\mathcal{E}}(G,J) = \max_{Q} \sum_{i,j=1}^{s} J_{i,j} \int_{[0,1]^2} \mathbb{1}_{Q_i}(x)\mathbb{1}_{Q_j}(y) W_G(x,y)\mathrm{d}x\mathrm{d}y,$$

where the maximum runs over all $\mathcal{I}_n$-partitions Q into s parts when $|V(G)| = n$.

Further,

$$\mathcal{E}(U,J) = \sup_{f} \sum_{i,j=1}^{s} J_{i,j} \int_{[0,1]^2} f_i(x) f_j(y) U(x,y)\mathrm{d}x\mathrm{d}y,$$

where the supremum runs over all fractional partitions f into s parts.

Recall Corollary 4.2.11 from Chapter 4. It states in the current setting that there is an absolute constant $c > 0$ such that for any $s \ge 1$, $\rho > 0$, 2-kernel U, real matrix J, and $q \ge c\Theta^4 \log(\Theta)$ with $\Theta = \frac{s^2}{\rho}$ we have

$$\mathbb{P}(|\mathcal{E}(U,J) - \hat{\mathcal{E}}(\mathbb{G}(q,U),J)| > \rho\|U\|_\infty) < 2\exp\left(-\frac{\rho^2 q}{32}\right). \tag{6.6}$$

We have seen above that $\hat{h}_A(U) = \hat{\mathcal{E}}(U,B_A)$ and $h_A(U) = \mathcal{E}(U,B_A)$. Since $q \ge t_2^2 \ge 2^{85}t_1^{10}/\Delta^5$ we can apply Corollary 4.2.11 for each $A \in \mathbb{A}$ with $s = 4t_1$, and $\rho = \Delta/4$.

This shows eventually that with probability at least $1 - 2^{16t_1^2+1}\exp\left(-\frac{\Delta^2 q}{2^9}\right) \leq 1 - \exp\left(-\frac{\Delta^2 q}{2^8}\right)$ we have that

$$\max_{A\in\mathbb{A}} \max_{\hat{Q}} \sum_{i,j=1}^{t_1} \sum_{\alpha,\beta=1}^{4} (B_A)_{i,j}^{\alpha,\beta} \int_{[0,1]^2} (W_{\mathbb{G}(q,U)})(x,y) \mathbb{1}_{Q_i^\alpha}(x) \mathbb{1}_{Q_j^\beta}(y) \mathrm{d}x\mathrm{d}y \leq \Delta/2, \tag{6.7}$$

where the second maximum runs over all $\mathcal{I}_q$-partitions $\hat{Q}$ of $[0,1]$ into $4t_1$ parts. Denote this event by E_1.

This however is equivalent to saying that for every Q partition into t_Q classes $t_Q \leq t_1$ it is true that

$$||W_{\mathbb{G}(q,U)}||_{\square Q} \leq \Delta/2. \tag{6.8}$$

The second estimate we require concerns the closeness of the step function $(W_G)_{\mathcal{P}}$ and its sample $W_{\mathbb{H}(q,(W_G)_{\mathcal{P}})}$. Our aim is to overlay these two functions via measure preserving permutations of $[0,1]$, such that the measure of the subset of $[0,1]^2$ where they differ is as small as possible.

Let $V = W_{\mathbb{H}(q,(W_G)_{\mathcal{P}})}$, this $\mathcal{I}_n$-function is well-defined this way and is a step function with steps forming the $\mathcal{I}_n$-partition $\mathcal{P}'$. This latter $\mathcal{I}_n$-partition of $[0,1]$ is the image of $\mathcal{P}$ induced by the sample $\{X_1,\dots,X_q\}$ and the map $i \mapsto [\frac{i-1}{q}, \frac{i}{q})$. Let ψ be a measure preserving $\mathcal{I}_n$-permutation of $[0,1]$ that satisfies that for each $i \in [t_{\mathcal{P}}]$ the volumes $\lambda(P_i \triangle \psi(P_i')) = |\lambda(P_i) - \lambda(P_i')|$. Let $\mathcal{P}''$ denote the partition with classes $P_i'' = \psi(P_i')$ and $V' = (V)^\psi$ (note that V' and V are equivalent as graphons), furthermore let N be the (random) subset of $[0,1]^2$ where the two functions $(W_G)_{\mathcal{P}}$ and V' differ. Then

$$\mathbb{E}[\lambda(N)] \leq 2\mathbb{E}[\sum_{i=1}^{t_{\mathcal{P}}} |\lambda(P_i) - \lambda(P_i')|]. \tag{6.9}$$

The random variables $\lambda(P_i')$ for each i can be interpreted as the proportion of positive outcomes out of q independent Bernoulli trials with success probability $\lambda(P_i)$. By Cauchy-Schwarz it follows that

$$\mathbb{E}[\sum_{i=1}^{t_{\mathcal{P}}} |\lambda(P_i) - \lambda(P_i')|] \leq \sqrt{t_{\mathcal{P}}\mathbb{E}[\sum_{i=1}^{t_{\mathcal{P}}} (\lambda(P_i) - \lambda(P_i'))^2]} \leq \sqrt{\frac{t_2}{q}}. \tag{6.10}$$

This calculation yields that $\mathbb{E}[\lambda(N)] \leq \sqrt{\frac{4t_2}{q}} \leq \Delta/8$ by the choice of q, since $q \geq t_2^2$. Standard concentration result gives us that $\lambda(N)$ is also small in probability if q is chosen large enough. For convenience, define the martingale $M_l = \mathbb{E}[\lambda(N)|X_1,\dots,X_l]$ for $1 \leq l \leq q$, and notice that the martingale differences are uniformly bounded, $|M_l - M_{l-1}| \leq \frac{4}{q}$. The Azuma-Hoeffding inequality then yields

$$\mathbb{P}(\lambda(N) \geq \Delta/4) \leq \mathbb{P}(\lambda(N) \geq \mathbb{E}[\lambda(N)] + \Delta/8) \leq \exp(-\Delta^2 q/2^{11}). \tag{6.11}$$

Define the event E_2 that holds whenever $\lambda(N) \leq \Delta/4$, and condition on E_1 and E_2, the failure probability of each one is at most $\exp(-\frac{\Delta^2 q}{2^{11}})$.

It follows that $\|V^\psi - (W_G)_\mathcal{P}\|_1 \leq \lambda(N) \leq \Delta/4$. Now employing the triangle inequality and the bound (6.8) we get for all $\mathcal{I}_n$-partitions $\mathcal{Q}$ into t parts that

$$\|W_G - (W_F)^\psi\|_{\square\mathcal{Q}} \leq \|W_G - (W_G)_\mathcal{P}\|_{\square\mathcal{Q}} + \|(W_G)_\mathcal{P} - V^\psi\|_1 + \|V^\psi - (W_F)^\psi\|_{\square\psi(\mathcal{Q})} \leq \Delta.$$

Now let $U \in \mathcal{M}_{\Delta,n}$ be arbitrary, and let $\mathcal{P}_U$ denote the partition consisting of the steps of U. Let ϕ be the $\mathcal{I}_n$-permutation of $[0,1]$ that is optimal in the sense that $d_{U,\mathcal{P}_U}(G) = \|U - (W_G)^\phi\|_{\square\mathcal{P}_U}$. Then

$$\begin{aligned} d_{U,\mathcal{P}_U}(G) - d_{U,\mathcal{P}_U}(F) &\leq \|U - (W_G)^\phi\|_{\square\mathcal{P}_U} - \|U - (W_F)^{(\psi\circ\phi)}\|_{\square\mathcal{P}_U} \\ &\leq \|W_G - (W_F)^\psi\|_{\square\phi^{-1}(\mathcal{P}_U)} \leq \Delta. \end{aligned}$$

The lower bound on the above difference can be handled in a similar way, therefore we have that $|d_{U,\mathcal{P}_U}(G) - d_{U,\mathcal{P}_U}(F)| \leq \Delta$ for every $U \in \mathcal{M}_{\Delta,n}$.

We conclude the proof with mentioning that the failure probability of the two events E_1 and E_2 taking place simultaneously is at most $\exp(-\frac{\Delta^2 q}{2^7})$.

□

We are now ready to conduct the proof of the main result of the chapter concerning graph parameters.

Proof of Theorem 6.1.5.

Let us fix $\varepsilon > 0$ and the simple graph G with n vertices. We introduce the error parameter $\Delta = \frac{k^{-2q_g^2(\varepsilon/2)}}{4k^2+1}$ and set $q \geq 2^{2^{(2k^4/\Delta^2)+4}}$. To establish the lower bound on $f(\mathbb{G}(q,G))$ not much effort is required: we pick a (k,m)-coloring $\mathbf{G}$ of G that certifies the value $f(G)$, that is, $g(\mathbf{G}) = f(G)$. Then the (k,m)-coloring of $\mathbf{F} = \mathbb{G}(q,\mathbf{G})$ of $\mathbb{G}(q,G)$ induced by $\mathbf{G}$ satisfies $g(\mathbf{F}) \geq g(\mathbf{G}) - \varepsilon/2$ with probability at least $1-\varepsilon/2$ since $q \geq q_g(\varepsilon/2)$, due to the testability property of g, which in turn implies $f(\mathbb{G}(q,G)) \geq f(G) - \varepsilon/2$ with probability at least $1 - \varepsilon/2$.

The problem concerning the upper bound in terms of q on $f(\mathbb{G}(q,G))$ is the difficult part of the proof, the rest of it deals with this case. Recall that $\mathcal{M}_{\Delta,n}$ denotes the set of the $[0,1]$-valued proper $\mathcal{I}_n$-step functions that have at most $t_k(\Delta,1)$ steps. Let us condition on the event in the statement of Lemma 6.2.4, that is for all $U \in \mathcal{M}_{\Delta,n}$ it holds that $|d_{U,\mathcal{P}_U}(G) - d_{U,\mathcal{P}_U}(\mathbb{G}(q,G))| \leq \Delta$. Let $\mathcal{N}$ be the set of all k-colored digraphs $\mathbf{W}$ that are $\mathcal{I}_n$-step functions with at most $t_k(\Delta,1)$ steps $\mathcal{P}$, and that satisfy $d_{U,\mathcal{P}}(G) \leq 2\Delta$ for $U = \sum_{(\alpha,\beta)\in M} W^{(\alpha,\beta)}$.

Our main step in the proof is that, conditioned on the aforementioned event, we construct for each (k,m)-coloring of F a corresponding coloring of G so that the g values of the two colored instances are sufficiently close. We elaborate on this argument in the following.

Let us fix an arbitrary (k,m)-coloring of F denoted by $\mathbf{F}$. According to Lemma 3.3.15 there exists a $\mathbf{W}$ that is a proper $\mathcal{I}_n$-step function with at most $t_k(\Delta,1)$ steps $\mathcal{P}_\mathbf{W}$ such that there exists an $\mathcal{I}_n$-permutation ϕ of $[0,1]$ such that $d_{\square\mathcal{P}}((\mathbf{W_F})^\phi, \mathbf{W}) \leq \Delta$. Therefore, setting $U = \sum_{(\alpha,\beta)\in M} W^{(\alpha,\beta)}$ we have $d_{U,\mathcal{P}}(F) \leq \Delta$ and $U \in \mathcal{M}_{\Delta,n}$. This in turn implies that $d_{U,\mathcal{P}}(G) \leq 2\Delta$, and consequently $\mathbf{W} \in \mathcal{N}$. It follows from Lemma 6.2.2 that there exists a (k,m)-coloring of G denoted by $\mathbf{G}$ such that $d_\square(\mathbf{W}, (\mathbf{W_G})^\psi) \leq 4k^2\Delta$ for some ψ that is an $\mathcal{I}_n$-permutation of $[0,1]$.

Therefore we get that $\delta_\square(\mathbf{G}, \mathbf{F}) \leq (4k^2+1)\Delta$. By virtue of Lemma 6.2.1 we can assert that $|g(\mathbf{G}) - g(\mathbf{F})| \leq \varepsilon/2$. This finishes our argument, as $\mathbf{F}$ was arbitrary, and the failure probability of the conditioned event in the analysis of the upper bound is at most $\varepsilon/2$. □

6.3 Weak nondeterminism

We introduce an even more restrictive notion of nondeterminism corresponding to node colorings (Definition 6.1.2 used throughout the chapter is a special case of the nondeterminism notion used commonly in complexity theory). Relying on this new concept we are able to improve on the upper bound of the sample complexity using a simplified version of our approach applied in the proof of Theorem 6.1.5 without significant alterations.

We formulate the definition of a stronger property than the previously defined nondeterministic testability. The notion itself may seem at first more involved, but in fact it only corresponds to the case, where the witness parameter g of f for a graph G is evaluated only on the set of node-colorings of G instead of edge-colorings in order to define the maximum expression. This modification will enable us to rely only on the cut-norm and the corresponding regularity lemmas instead of the cut-$\mathcal{P}$-norm that was employed in the general case, thus leads us to improved upper bounds on the sample complexity of f with respect to that of g. This time we only treat the case of undirected graph colorings in detail, the directed case is analogous.

We will introduce the set of colorings of G called node-(k,m)-colorings. Let $\mathcal{T} = (T_1,\dots,T_k)$ be a partition of $V(G)$ and $\mathcal{D} = ((D_1,\dots,D_m),(D'_1,\dots,D'_m))$ be two partitions of $[k]^2$, together they induce two partitions, $\mathcal{C} = ((C_1,\dots,C_m),(C'_1,\dots,C'_m))$, of $V(G)^2$ such that each class is of the form $C_\alpha = \cup_{(i,j)\in D_\alpha} T_i \times T_j$ and $C'_\alpha = \cup_{(i,j)\in D'_\alpha} T_i \times T_j$ respectively. A node-(k,m)-coloring of G is defined by some $\mathcal{C}$ of the previous form and is the $2m$-tuple of simple graphs $\mathbf{G} = (G_1,\dots,G_m,\tilde{G}_1,\dots,\tilde{G}_m)$ with $G_\alpha = G[C_\alpha]$ and $\tilde{G}_\alpha = G^c[C'_\alpha]$. Here G^c stands for the complement of G (the union of G and its complement is the directed complete graph with all loops present), and $G[C_\alpha]$ is the union of induced labeled subgraphs of G between T_i and T_j for each $(i,j) \in D_\alpha$ for $i \neq j$, in the case of $i = j$ the term in the union is the induced labeled subgraph of G on the node set T_i.

These special edge-$2m$-colored graphs that can serve as node-(k,m)-colorings are given by a triple $(G,\mathcal{T},\mathcal{D})$, where G is a simple graph, $\mathcal{T}$ is a partition of $V(G)$ into k parts, and $\mathcal{D}$ is a pair of partitions of $[k]^2$ into m parts. In the case of r-uniform hypergraphs for arbitrary $r \geq 2$ a node-(k,m)-colorings is also a triple $(G,\mathcal{T},\mathcal{D})$, the only difference in comparison to the graph case is that $\mathcal{D}$ is a pair of partitions of $[k]^r$ into m parts, the rest of the description is analogous.

Definition 6.3.1. *The r-uniform hypergraph parameter f is weakly non-deterministically testable if there exist integers m and k with $m \leq k^r$ and a testable edge-$2m$-colored directed r-graph parameter g such that for any simple r-graph G we have $f(G) = \max_{\mathbf{G}} g(\mathbf{G})$, where the maximum goes over the set of node-(k, m)-colorings of G.*

We present two approaches to handle this variant of the nondeterministic testability. The first method follows the proof framework introduced in the previous section in the general case for graphs, its adaptation to the current setting results an improvement on the upper bound on the sample complexity to a 2-fold exponential of the sample complexity of the witness a parameter and is also applicable to the corresponding property testing setting. The second idea entails the graph case as well as the r-uniform hypergraph setting for arbitrary rank r of the weak setting. We manage to reduce the upper bound on the sample complexity further to only exponential dependence. This approach does seems to be more problem specific, than the previous one, and it does not directly yield an analogous statement in property testing.

First approach

The following lemma is the analogous result to Lemma 6.2.2 that can be employed in the proof of the variant of Theorem 6.1.5 for the special case of weakly nondeterministically testable graph parameters.

Lemma 6.3.2. *Let $\varepsilon > 0$, let U and V be arbitrary graphons with $\|U - V\|_\square \leq \varepsilon$, and also let $k \geq 2$. For any $\mathbf{U} = (U^{(1)}, \dots, U^{(m)}, \tilde{U}^{(1)}, \dots, \tilde{U}^{(m)})$ node-(k, m)-coloring of U there exists a node-(k, m)-coloring of V denoted by $\mathbf{V} = (V^{(1)}, \dots, V^{(k)}, \tilde{V}^{(1)}, \dots, \tilde{V}^{(m)})$ so that $d_\square(\mathbf{U}, \mathbf{V}) = \sum_{i=1}^m \|U^{(i)} - V^{(i)}\|_\square + \sum_{i=1}^m \|\tilde{U}^{(i)} - \tilde{V}^{(i)}\|_\square \leq 2k^2\varepsilon$. If $V = W_G$ for some simple graph G on n nodes and each $U^{(i)}$ is an $\mathcal{I}_n$-step function then there is a coloring $\mathbf{G}$ of G such that $d_\square(\mathbf{U}, \mathbf{W_G}) \leq 2k^2\varepsilon$.*

Proof. Our approach is quite elementary: consider the partition $\mathcal{T}$ of $[0, 1]$ and C that is a pair of partitions of $[0, 1]^2$ corresponding to a pair of partitions $\mathcal{D}$ of $[k]^2$ as above that together with U describe $\mathbf{U}$ as above, and define $V^{(i)} = V\mathbb{1}_{C_i}$ and $\tilde{V}^{(i)} = (1 - V)\mathbb{1}_{C'_i}$ for each $i \in [m]$. Then

$$\|U^{(i)} - V^{(i)}\|_\square \leq \sum_{(\alpha,\beta)\in D_i} \|(U - V)\mathbb{1}_{T_\alpha \times T_\beta}\|_\square \leq \varepsilon |D_i| \tag{6.12}$$

for each $i \in [m]$, and the same upper bound applies to $\|\tilde{U}^{(i)} - \tilde{V}^{(i)}\|_\square$. Summing up over i gives the result stated in the lemma.

The argument showing the part regarding simple graphs is identical.

□

Note that in Lemma 6.2.2 we required U and V to be close in the cut-$\mathcal{P}$-norm for some partition $\mathcal{P}$, and U to be a $\mathcal{P}$ step function to guarantee for each $\mathbf{U}$ the existence of $\mathbf{V}$ that is close to it in the cut distance of k-colored digraphons. Using the fact that in the weakly non-deterministic framework cut-closeness of instances implies the cut-closeness of the sets of their node-(k,m)-colorings we can formulate the next corollary of Theorem 6.1.5 that is one of the main results of this subsection.

Corollary 6.3.3. *Let f be a weakly non-deterministically testable graph parameter with witness parameter g of node-(k,m)-colored graphs with the corresponding sample complexity q_g. Then f is testable with sample complexity q_f, and we have that $q_f(\varepsilon) \leq \exp^{(2)}(cq_g^2(\varepsilon/2))$ for some $c > 0$ large enough that does depend only on k and not on f for any $\varepsilon > 0$.*

Proof. We will give only a sketch of the proof, as it is almost identical to that of Theorem 6.1.5, and we automatically refer to that, including the notation used in the current proof. Let G be a simple graph on n nodes, and let $\varepsilon > 0$ be fixed, $q \geq \exp^{(2)}(cq_g^2(\varepsilon/2))$ for some constant $c > 0$ that will be specified later. The part concerning the lower bound of $f(\mathbb{G}(q,G))$ is completely identical to the general case.

For the upper bound set $\Delta = \exp(-cq_q^2(\varepsilon))$. We condition on the event $\delta_\square(G, \mathbb{G}(q,G)) \leq \Delta$, whose failure probability is sufficiently small due to Lemma 3.3.9, i.e. for $q \geq 2^{100/\Delta^2}$ it is at most $\exp\left(-4^{100/\Delta^2}\frac{\Delta^2}{50}\right)$. We define c to be large enough so that the above lower bound on q holds true. Now we select an arbitrary node-(k,m)-coloring $\mathbf{F}$ of $\mathbb{G}(q,G)$ and apply the Weak Regularity Lemma, Lemma 3.3.4, for $2m$-colored graphons, Lemma 3.3.11, in the $\mathcal{I}_n$-step function case with error parameter $\Delta/(2k^2+1)$ (keeping in mind that $m \leq k^2$) to get a tuple of $\mathcal{I}_n$-step functions forming $\mathbf{U}$ with at most $t'_{2k^2}(\Delta/(2k^2+1))$ steps. We define the $\mathcal{I}_n$-step function graphon $U = \sum_{i=1}^m U_i$ and note that our condition implies that $\delta_\square(G,U) \leq 2\Delta$, since $\delta_\square(G,U) \leq \delta_\square(G,\mathbb{G}(q,G)) + \delta_\square(\mathbb{G}(q,G),U)$. To finish the proof we apply Lemma 6.3.2, it implies the existence of a coloring $\mathbf{G}$ of G so that $\delta_\square(\mathbf{G},\mathbf{F}) \leq (2k^2+1)\Delta$. Applying Lemma 6.2.1 delivers the desired result by establishing that $|g(\mathbf{F}) - g(\mathbf{G})| \leq \varepsilon$.

□

Second approach

Recall the notion of layered ground state energies of r-arrays of Chapter 4 for arbitrary $r \geq 1$.

Let $r,k \geq 1$, and $G = (G^z)_{z\in[k]^r}$ be $[k]^r$-tuple of real r-arrays of size n, and $\mathcal{T} = (T_1,\dots,T_k)$ a partition of $[n]$ into k parts. Then

$$\mathcal{E}_{\mathcal{T}}(G) = \sum_{z\in[k]^r} \frac{1}{n^r} \sum_{i_1,\dots,i_r=1}^{n} G^z(i_1,\dots,i_r) \prod_{j=1}^{r} \mathbb{1}_{T_{z_j}}(i_j), \tag{6.13}$$

and

$$\hat{\mathcal{E}}(G) = \max_{\mathcal{T}} \mathcal{E}_{\mathcal{T}}(G), \tag{6.14}$$

where the maximum runs over all integer partitions $\mathcal{T}$ of $[n]$ into k parts.

We will make use of Theorem 4.1.4 that deals with the testability of layered GSE, in particular the dependence of the upper bound on the sample complexity on the dimension r.

We are ready to state and prove the main theorem of the section that includes a further improvement fo the upper bound on the sample complexity compared to our first approach in the weak nondeterministic testing setting.

Theorem 6.3.4. *Let $r \geq 1$ and f be a weakly non-deterministically testable r-graph parameter with witness parameter g of node-(k,m)-colored graphs, and let the corresponding sample complexity functions be q_f and q_g. Then f is testable and there exist a $c_{r,k} > 0$ that does depend only on r and k, but not on f such that for any $\varepsilon > 0$ we have $q_f(\varepsilon) \leq \exp(c_{r,k} q_g(\varepsilon/8))$.*

Proof. Let $r \geq 1$ be arbitrary, and f be a weakly nondeterministically testable r-graph parameter with a certificate specified by the constants k and $m \leq k^r$, and the testable $2m$-colored r-graph parameter g. Then

$$f(G) = \max_{\mathcal{T},\mathcal{D}} g(\mathbf{G}(G,\mathcal{T},\mathcal{D})),$$

where the maximum goes over every pair $(\mathcal{T},\mathcal{D})$, where $\mathcal{T}$ is a partition of $V(G)$ into k parts, and $\mathcal{D}$ is a pair of partitions of $[k]^r$ into m parts, and $\mathbf{G}(G,\mathcal{T},\mathcal{D})$ is the edge $2m$-colored graph defined by its parameters as seen above. Define the for each fixed $\mathcal{D}$ the node-k-colored r-graph (i.e., a simple r-graph together with a k-coloring of its nodes) parameter $g^{\mathcal{D}}(G,\mathcal{T}) = g(\mathbf{G}(G,\mathcal{T},\mathcal{D}))$ and the simple r-graph parameter $f^{\mathcal{D}}(G) = \max_{\mathcal{T}} g^{\mathcal{D}}(G,\mathcal{T})$.

Let $\varepsilon > 0$ be arbitrary, define

$$g^{\varepsilon}(\mathbf{G}(G,\mathcal{P},\mathcal{D})) = \sum_{F,\mathcal{T}} t(\mathbf{G}(F,\mathcal{T},\mathcal{D}),\mathbf{G}(G,\mathcal{P},\mathcal{D}))g(\mathbf{G}(F,\mathcal{T},\mathcal{D})),$$

where the sum goes over all simple r-graphs F on $q_0 = q_g(\varepsilon/8)$ vertices and partitions $\mathcal{T}$ of $[q_0]$ into k parts. By the testability of g we have

$$|g^{\varepsilon}(\mathbf{G}(G,\mathcal{P},\mathcal{D})) - g(\mathbf{G}(G,\mathcal{P},\mathcal{D}))| \leq \varepsilon/4, \tag{6.15}$$

for each permitted tuple $(G,\mathcal{P},\mathcal{D})$. Analogously we define

$$g^{\varepsilon,\mathcal{D}}(G,\mathcal{P}) = \sum_{F,\mathcal{T}} t(\mathbf{G}(F,\mathcal{T},\mathcal{D}),\mathbf{G}(G,\mathcal{P},\mathcal{D}))g(\mathbf{G}(F,\mathcal{T},\mathcal{D})),$$

and

$$f^{\varepsilon,\mathcal{D}}(G) = \max_{\mathcal{T}} g^{\varepsilon,\mathcal{D}}(G,\mathcal{T}).$$

It follows from (6.15) that for any G simple r-graph

$$|f^{\varepsilon,\mathcal{D}}(G) - f^{\mathcal{D}}(G)| \leq \varepsilon/4,$$

and for any $\varepsilon > 0$ and $q \geq 1$ we have

$$|f(G) - f(\mathbb{G}(q,G))| \leq \max_{\mathcal{D}} |f^{\varepsilon,\mathcal{D}}(G) - f^{\varepsilon,\mathcal{D}}(\mathbb{G}(q,G))| + \varepsilon/2. \tag{6.16}$$

For any $\varepsilon > 0$ and $\mathcal{D}$ that is a pair of partitions of $[k]^r$ into m parts the parameter $f^{\varepsilon,\mathcal{D}}$ can be re-written as an energy of q_0-arrays: For G of size $[n]$ let $H = (H^z)_{z\in[k]^{q_0}}$ so that for each $z \in [k]^{q_0}$ the real q_0-array H^z is defined by

$$H^z(i_1,\dots,i_{q_0}) = g^{\mathcal{D}}(G[(i_1,\dots,i_{q_0})], \mathcal{P}_z(i_1,\dots,i_{q_0}))$$

for each $(i_1,\dots,i_{q_0}) \in [n]^{q_0}$, where $\mathcal{P}_z(i_1,\dots,i_{q_0}) = (P_1,\dots,P_k)$ is a partition of $(i_1,\dots,i_{q_0})$ given by $P_l = \{\, i_j \mid z_j = l\,\}$ for $l \in [k]$. Then for each $\mathcal{T}$ that is a partition of $[n]$ into k parts we can assert that

$$\begin{aligned} g^{\varepsilon,\mathcal{D}}(G,\mathcal{T}) &= \sum_{z\in[k]^{q_0}} \frac{1}{n^{q_0}} \sum_{i_1,\dots,i_r=1}^{n} g^{\mathcal{D}}(G[(i_1,\dots,i_{q_0})], \mathcal{P}_z(i_1,\dots,i_{q_0})) \prod_{j=1}^{r} \mathbb{1}_{T_{z_j}}(i_j) \\ &= \sum_{z\in[k]^{q_0}} \frac{1}{n^{q_0}} \sum_{i_1,\dots,i_r=1}^{n} H^z(i_1,\dots,i_{q_0}) \prod_{j=1}^{r} \mathbb{1}_{T_{z_j}}(i_j) \\ &= \mathcal{E}_{\mathcal{T}}(H), \end{aligned}$$

and further

$$f^{\varepsilon,\mathcal{D}}(G) = \max_{\mathcal{T}} g^{\varepsilon,\mathcal{D}}(G,\mathcal{T}) = \max_{\mathcal{T}} \mathcal{E}_{\mathcal{T}}(H) = \hat{\mathcal{E}}(H).$$

Analogously it holds for any $q \geq q_0$ that $f^{\varepsilon,\mathcal{D}}(\mathbb{G}(q,G)) = \hat{\mathcal{E}}(\mathbb{G}(q,H))$.

This implies by Corollary 4.2.11 that for $q \geq \Theta^4 \log(\Theta)$ with $\Theta = \frac{2^{q_0+11}k^{q_0}q_0}{\varepsilon}$ and each fixed $\mathcal{D}$ that

$$\mathbb{P}(|f^{\varepsilon,\mathcal{D}}(G) - f^{\varepsilon,\mathcal{D}}(\mathbb{G}(q,G))| > \varepsilon/2) < 2\exp\left(-\frac{\varepsilon^2 q}{32q_0^2}\right).$$

The probability that the event in the previous formula occurs for some $\mathcal{D}$ is at most $k^{2rk^r} 2\exp\left(-\frac{\varepsilon^2 q}{32q_0^2}\right)$, therefore by recalling (6.16) we can conclude that there exists a constant $c_{r,k} > 0$ not depending on other specifics of f such that for each simple graph G and $q \geq \exp(c_{r,k} q_g(\varepsilon/8))$ it holds that

$$\mathbb{P}(|f(G) - f(\mathbb{G}(q,G))| > \varepsilon) < \varepsilon.$$

□

6.4 Effective upper bound for the r-cut norm of a sampled r-graph

We are going to establish upper and lower bounds for the genuine r-cut norm (see Definition 3.3.20) of an r-kernel using certain subgraph densities mentioned in (2.4) and (2.10). Let W be an r-kernel, the r-cut norm of W reads as

$$\|W\|_{\square,r} = \sup_{\substack{S_i \subset [0,1]^{\mathfrak{h}([r-1])} \\ i\in[r]}} \left| \int_{\cap_{i\in[r]} p^{-1}_{\mathfrak{h}([r]\setminus\{i\})}(S_i)} W(x_{\mathfrak{h}([r],r-1)}) \mathrm{d}\lambda(x_{\mathfrak{h}([r],r-1)}) \right|,$$

where the supremum is taken over sets S_i that are $(r-1)$-symmetric.

Let further $H \subset \binom{[q]}{r}$ be a simple r-graph on q vertices, recall

$$t^*(H,W) = \int_{[0,1]^{\mathfrak{h}([q],r-1)}} \prod_{e\in H} W(x_{\mathfrak{h}(e,r-1)}) \mathrm{d}\lambda(x),$$

this expression is a variant of the subgraph densities discussed in Chapter 2 that in the case of $W = W_G$ counts the graph homomorphisms from H to W that preserve adjacency but not necessarily non-adjacency. Using the previously introduced terminology we can write

$$t^*(H,W) = \sum_{H\subset F\subset \binom{q}{r}} t(F,W).$$

Let K_r^2 denote the simple r-graph that is the 2-fold blow-up of the r-graph consisting of r vertices and one edge. That is, $V(K_r^2) = \{v_1^0,\dots,v_r^0,v_1^1,\dots,v_r^1\}$ and $E(K_r^2) = \{\,\{v_1^{i_1},\dots,v_r^{i_r}\} \mid i_1,\dots,i_r \in \{0,1\}\,\}$, alternatively we may regard K_r^2 as a subset of $\binom{[2r]}{r}$.

It was shown by Borgs, Chayes, Lovász, Sós, and Vesztergombi [30] for $r = 2$ with tools from functional analysis that for any symmetric 2-kernel W with $\|W\|_\infty \le 1$ we have

$$\frac{1}{4} t^*(K_2^2, W) \le \|W\|_{\square,2} \le [t^*(K_2^2,W)]^{1/4}, \tag{6.17}$$

where $[t^*(K_2^2,.)]^{1/4}$ is also called the trace norm or the Schatten norm of the integral operator T_W corresponding to W. We remark that in the above case K_2^2 stands for the 4-cycle. We establish here a similar relation between $t^*(K_r^2,W)$ and $\|W\|_{\square,r}$ for general r.

It is not hard to show that for any $r \ge 1$ and r-kernel W it holds that $t^*(K_r^2,W) \ge 0$.

$$t^*(K_r^2,W) = \int_{[0,1]^{\mathfrak{h}(V(K_r^2),r-1)}} \prod_{i_1,\dots,i_r\in\{0,1\}} W(x_{\mathfrak{h}(\{v_1^{i_1},\dots,v_r^{i_r}\},r-1)}) \mathrm{d}\lambda(x)$$

$$= \int_{[0,1]^{T_1}} \int_{[0,1]^{T_2}} \prod_{i_1,\dots,i_r \in \{0,1\}} W(x_{\mathfrak{h}(\{v_1^{i_1},\dots,v_r^{i_r}\},r-1)}) \mathrm{d}\lambda(x_{T_2}) \mathrm{d}\lambda(x_{T_1})$$

$$= \int_{[0,1]^{T_1 \cup T_2^2}} \left[\int_{[0,1]^{T_2^0}} \prod_{i_1,\dots,i_{r-1} \in \{0,1\}} W(x_{\mathfrak{h}(\{v_1^{i_1},\dots,v_{r-1}^{i_{r-1}},v_r^0\},r-1)}) \mathrm{d}\lambda(x_{T_2^0}) \right]$$

$$\left[\int_{[0,1]^{T_2^1}} \prod_{i_1,\dots,i_{r-1} \in \{0,1\}} W(x_{\mathfrak{h}(\{v_1^{i_1},\dots,v_{r-1}^{i_{r-1}},v_r^1\},r-1)}) \mathrm{d}\lambda(x_{T_2^1}) \right] \mathrm{d}\lambda(x_{T_1 \cup T_2^2})$$

$$= \int_{[0,1]^{T_1}} \left[\int_{[0,1]^{T_3 \setminus T_1}} \prod_{i_1,\dots,i_{r-1} \in \{0,1\}} W(x_{\mathfrak{h}(\{v_1^{i_1},\dots,v_{r-1}^{i_{r-1}},u\},r-1)}) \mathrm{d}\lambda(x_{T_3 \setminus T_1}) \right]^2 \mathrm{d}\lambda(x_{T_1}),$$

where $T_1 = \mathfrak{h}(V(K_r^2) \setminus \{v_r^0, v_r^1\}, r-1)$, $T_2 = \mathfrak{h}(V(K_r^2), r-1) \setminus T_1$, and T_2^i is the subset of T_2 whose elements contain v_r^i, but not v_r^{i+1}, for $i = 0, 1$. Further, $T_2^2 = T_2 \setminus (T_2^0 \cup T_2^1)$ and $T_3 = \mathfrak{h}(V(K_r^2) \setminus \{v_r^0, v_r^1\} \cup \{u\}, r-1)$, note that the T_2^2 coordinates do not actually appear in the integrand. We used Fubini's theorem, that enabled to integrate first over variables with coordinates from T_2 while the variables with T_1 coordinates were fixed, which we could then use to identify v_r^0 and v_r^1.

In the proof of (6.17) the authors drew on tools from functional analysis and the fact that a 2-kernel describes an integral operator, those concepts do not have a natural counterpart for r-kernels. However, we can provide an analogous result by the repeated application of Fubini's theorem and the Cauchy-Schwarz inequality in the L^2-space.

Lemma 6.4.1. *For any $r \geq 1$ and r-kernel W with $\|W\|_\infty \leq 1$ we have*

$$2^{-r} t^*(K_r^2, W) \leq \|W\|_{\square,r} \leq [t^*(K_r^2, W)]^{1/2^r}. \tag{6.18}$$

Proof. The lower bound on $\|W\|_{\square,r}$ is straightforward, and K_r^2 could even be replaced by any other simple r-graph with 2^r edges, we only need to use a simplified version of Lemma 3.3.24 for kernels, with setting one of the kernels to 0.

For the other direction, let us fix a collection of arbitrary symmetric measurable functions $f_1, \dots, f_r \colon [0,1]^{\mathfrak{h}([r-1])} \to [0,1]$. Set $V = \{v_1, \dots, v_r\}$ and for any $l \geq 1$ and $i_1, \dots, i_l \in \{0,1\}$ let

$$V^{i_1,\dots,i_l} = \{v_1^{i_1}, \dots, v_l^{i_l}, v_{l+1}, \dots, v_r\}.$$

Further, let $V_j = V \setminus \{v_j\}$ and for $l+1 \leq j \leq r$ let $V_j^{i_1,\dots,i_l} = V^{i_1,\dots,i_l} \setminus \{v_j\}$. Let us introduce the index sets $T_1 = \mathfrak{h}(V_1)$, $S_1 = \mathfrak{h}(V, r-1) \setminus T_1$, $S_1^0 = \mathfrak{h}(V^0, r-1) \setminus T_1$, and $S_1^1 = \mathfrak{h}(V^1, r-1) \setminus T_1$. For $1 \leq l \leq r-1$ we define the sets

$$T_{l+1} = \cup_{i_1,\dots,i_l \in \{0,1\}} \mathfrak{h}(V_{l+1}^{i_1,\dots,i_l}),$$

$$S_{l+1} = (T_l \cup S_l^0 \cup S_l^1) \setminus T_{l+1},$$

and

$$S_{l+1}^0 = \{ (e \setminus \{v_{l+1}\}) \cup \{v_{l+1}^0\} \mid e \in S_{l+1} \}, \qquad S_{l+1}^1 = \{ (e \setminus \{v_{l+1}\}) \cup \{v_{l+1}^0\} \mid e \in S_{l+1} \}.$$

Then we have

$$\begin{aligned}
&\left| \int_{[0,1]^{\mathfrak{h}(V,r-1)}} \prod_{j=1}^{r} f_j(x_{\mathfrak{h}(V_j)}) W(x_{\mathfrak{h}(V,r-1)}) \mathrm{d}\lambda(x_{\mathfrak{h}(V,r-1)}) \right| \\
&\quad = \left| \int_{[0,1]^{T_1}} f_1(x_{\mathfrak{h}(V_1)}) \int_{[0,1]^{S_1}} \prod_{j=2}^{r} f_j(x_{\mathfrak{h}(V_j)}) W(x_{\mathfrak{h}(V,r-1)}) \mathrm{d}\lambda(x_{S_1}) \mathrm{d}\lambda(x_{T_1}) \right| \\
&\quad \le \left[\int_{[0,1]^{T_1}} f_1^2(x_{\mathfrak{h}(V_1)}) \lambda(x_{T_1}) \right]^{1/2} \left[\int_{[0,1]^{T_1}} \left(\int_{[0,1]^{S_1}} \prod_{j=2}^{r} f_j(x_{\mathfrak{h}(V_j)}) W(x_{\mathfrak{h}(V,r-1)}) \mathrm{d}\lambda(x_{S_1}) \right)^2 \mathrm{d}\lambda(x_{T_1}) \right]^{1/2} \\
&\quad \le \Bigg[\int_{[0,1]^{T_1}} \left(\int_{[0,1]^{S_1^0}} \prod_{j=2}^{r} f_j(x_{\mathfrak{h}(V_j^0)}) W(x_{\mathfrak{h}(V^0,r-1)}) \mathrm{d}\lambda(x_{S_1^0}) \right) \\
&\qquad\qquad \left(\int_{[0,1]^{S_1^1}} \prod_{j=2}^{r} f_j(x_{\mathfrak{h}(V_j^1)}) W(x_{\mathfrak{h}(V^1,r-1)}) \mathrm{d}\lambda(x_{S_1^1}) \right) \mathrm{d}\lambda(x_{T_1}) \Bigg]^{1/2},
\end{aligned}$$

where we used $\|f_1\|_\infty \le 1$ and the identity $\int (\int f(x,y)\mathrm{d}y)^2 \mathrm{d}x = \int f(x,y) f(x,z) \mathrm{d}y\mathrm{d}z\mathrm{d}x$ in the previous inequality. We proceed by upper bounding the last expression through repeated application of this reformulation combined with Cauchy-Schwarz.

$$\begin{aligned}
&\Bigg[\int_{[0,1]^{T_l}} \left(\prod_{i_1,\dots,i_{l-1}\in\{0,1\}} f_l(x_{\mathfrak{h}(V_l^{i_1,\dots,i_{l-1}})}) \right) \\
&\qquad \left(\int_{[0,1]^{S_l}} \prod_{i_1,\dots,i_{l-1}\in\{0,1\}} \prod_{j=l+1}^{r} f_j(x_{\mathfrak{h}(V_j^{i_1,\dots,i_{l-1}})}) W(x_{\mathfrak{h}(V^{i_1,\dots,i_{l-1}})}) \mathrm{d}\lambda(x_{S_l}) \right) \mathrm{d}\lambda(x_{T_l}) \Bigg]^{\frac{1}{2^{l-1}}} \\
&\quad \le \left[\int_{[0,1]^{T_l}} \left(\prod_{i_1,\dots,i_{l-1}\in\{0,1\}} f_l(x_{\mathfrak{h}(V_l^{i_1,\dots,i_{l-1}})}) \right)^2 \mathrm{d}\lambda(x_{T_l}) \right]^{\frac{1}{2l}} \\
&\qquad \left[\int_{[0,1]^{T_l}} \left(\int_{[0,1]^{S_l}} \prod_{i_1,\dots,i_{l-1}\in\{0,1\}} \prod_{j=l+1}^{r} f_j(x_{\mathfrak{h}(V_j^{i_1,\dots,i_{l-1}})}) W(x_{\mathfrak{h}(V^{i_1,\dots,i_{l-1}})}) \mathrm{d}\lambda(x_{S_l}) \right)^2 \mathrm{d}\lambda(x_{T_l}) \right]^{\frac{1}{2^l}}
\end{aligned}$$

$$
\begin{aligned}
&\leq \Bigg[\int_{[0,1]^{T_l}} \Bigg(\int_{[0,1]^{S_l^0}} \prod_{i_1,\dots,i_{l-1}\in\{0,1\}} \prod_{j=l+1}^{r} f_j(x_{\mathfrak{h}(V_j^{i_1,\dots,i_{l-1},0})}) W(x_{\mathfrak{h}(V^{i_1,\dots,i_{l-1},0})}) \mathrm{d}\lambda(x_{S_l^0})\Bigg) \\
&\qquad \Bigg(\int_{[0,1]^{S_l^1}} \prod_{i_1,\dots,i_{l-1}\in\{0,1\}} \prod_{j=l+1}^{r} f_j(x_{\mathfrak{h}(V_j^{i_1,\dots,i_{l-1},1})}) W(x_{\mathfrak{h}(V^{i_1,\dots,i_{l-1},1})}) \mathrm{d}\lambda(x_{S_l^1})\Bigg) \mathrm{d}\lambda(x_{T_l})\Bigg]^{\frac{1}{2^l}} \\
&= \Bigg[\int_{[0,1]^{T_l\cup S_l^0\cup S_l^1}} \prod_{i_1,\dots,i_l\in\{0,1\}} \prod_{j=l+1}^{r} f_j(x_{\mathfrak{h}(V_j^{i_1,\dots,i_l})}) W(x_{\mathfrak{h}(V^{i_1,\dots,i_l})}) \mathrm{d}\lambda(x_{S_l^0}) \mathrm{d}\lambda(x_{S_l^1}) \mathrm{d}\lambda(x_{T_l})\Bigg]^{\frac{1}{2^l}} \\
&= \Bigg[\int_{[0,1]^{T_{l+1}}} \Bigg(\prod_{i_1,\dots,i_l\in\{0,1\}} f_{l+1}(x_{\mathfrak{h}(V_{l+1}^{i_1,\dots,i_l})})\Bigg) \\
&\qquad \Bigg(\int_{[0,1]^{S_{l+1}}} \prod_{i_1,\dots,i_l\in\{0,1\}} \prod_{j=l+2}^{r} f_j(x_{\mathfrak{h}(V_j^{i_1,\dots,i_l})}) W(x_{\mathfrak{h}(V^{i_1,\dots,i_l})}) \mathrm{d}\lambda(x_{S_{l+1}})\Bigg) \mathrm{d}\lambda(x_{T_{l+1}})\Bigg]^{\frac{1}{2^l}} \\
&\quad \vdots \\
&\leq \Bigg[\int_{[0,1]^{T_r\cup S_r^0\cup S_r^1}} \prod_{i_1,\dots,i_r\in\{0,1\}} W(x_{\mathfrak{h}(V^{i_1,\dots,i_r})}) \mathrm{d}\lambda(x_{T_r\cup S_r^0\cup S_r^1})\Bigg]^{\frac{1}{2^r}} = t^*(K_r^2, W)^{1/2^r},
\end{aligned}
$$

where in subsequent inequalities we first used the Cauchy-Schwarz inequality, and afterwards that $\|f_j\|_\infty \leq 1$ for any $j \in [r]$. As the test functions $f_1, \dots, f_r$ were arbitrary the statement of the lemma follows. □

Utilizing the previous result we can obtain a quantitative upper bound on the r-cut norm of a graph sampled from a kernel for arbitrary r.

Lemma 6.4.2. *Let $r, k \geq 1$. For any $\varepsilon > 0$ and $t \geq 1$ there exists an integer $q_{\mathrm{cut}}(r, k, \varepsilon, t) \leq cr^2(t^r k^2(1/\varepsilon))^{2^{r+1}}$ for some universal constant $c > 0$ such that for any k-tuple of r-kernels $U_1, \dots, U_k$ that take values in $[-1, 1]$, and any integer $q \geq q_{\mathrm{cut}}(r, k, \varepsilon, t)$ it holds with probability at least $1 - \varepsilon$ that if*

$$\sum_{l=1}^{k} \|U_l\|_{\square,r} \leq \left(\frac{\varepsilon}{kt^r}\right)^{2^r} 2^{-r-1},$$

then

$$\sup_{Q, t_Q \leq t} \sum_{l=1}^{k} \|W_{\mathbb{H}(q,U_l)}\|_{\square,r,Q} \leq \varepsilon.$$

where the supremum goes over $(r-1)$-symmetric partitions Q of $[0,1]^{\mathfrak{h}([r-1])}$ into at most t classes.

Proof. Let $r, k, t \geq 1$ and $\varepsilon > 0$ be fixed, and let $U_1, \ldots, U_k$ and q be arbitrary. We have by a straight-forward variant of Lemma 3.5.4 for any r-kernel U, positive integer q, and $F \in \Pi^r$ we have that

$$\mathbb{P}(|t^*(F, U) - t^*(F, \mathbb{H}(q, U))| \geq \delta) \leq 2\exp(-\frac{\delta^2 q}{8|V(F)|^2})$$

for any $\delta > 0$, in particular for $F = K_r^2$ we have

$$\mathbb{P}(|t^*(K_r^2, U) - t^*(K_r^2, \mathbb{H}(q, U))| \geq \delta) \leq 2\exp(-\frac{\delta^2 q}{32r^2}). \tag{6.19}$$

Then we can estimate $\sup_{Q, t_Q \leq t} \sum_{l=1}^{k} ||W_{\mathbb{H}(q,U_l)}||_{\square,r,Q}$ using Lemma 6.4.1. Set $\delta = \frac{1}{2}\left(\frac{\varepsilon}{kt^r}\right)^{2^r}$, and let q be large enough so that $2k\exp\left(-\frac{\delta^2 q}{32r^2}\right) < \varepsilon$. Let $\mathbb{A}$ denote the set of all r-arrays of size t with $\{-1, 1\}$ entries. Then we have

$$\begin{aligned}
&\sup_{Q, t_Q \leq t} \sum_{l=1}^{k} ||W_{\mathbb{H}(q,U_l)}||_{\square,r,Q} \\
&= \sup_{Q, t_Q \leq t} \max_{A \in \mathbb{A}} \sup_{\substack{T_j^l \subset [0,1]^{\mathfrak{h}([r-1])} \\ j \in [r], l \in [k]}} \\
&\quad \sum_{l=1}^{k} \sum_{i_1, \ldots, i_r = 1}^{t} A(i_1, \ldots, i_r) \int_{[0,1]^{\mathfrak{h}([r], r-1)}} W_{\mathbb{H}(q,U_l)}(x_{\mathfrak{h}([r], r-1)}) \prod_{j=1}^{r} \mathbb{1}_{T_j^l \cap Q_{i_j}}(x_{\mathfrak{h}([r] \setminus \{j\})}) d\lambda(x_{\mathfrak{h}([r], r-1)}) \\
&\leq t^r \sum_{l=1}^{k} ||W_{\mathbb{H}(q,U_l)}||_{\square,r} \\
&\leq t^r \sum_{l=1}^{k} t^*(K_r^2, W_{\mathbb{H}(q,U_l)})^{1/2^r} \\
&= t^r \sum_{l=1}^{k} t^*(K_r^2, \mathbb{H}(q, U_l))^{1/2^r} \\
&\leq t^r \sum_{l=1}^{k} (t^*(K_r^2, U_l) + \delta)^{1/2^r} \\
&\leq t^r \sum_{l=1}^{k} (2^r ||U_l||_{\square,r} + \delta)^{1/2^r} \leq \varepsilon,
\end{aligned}$$

and the assumptions of the calculation, in particular the fourth inequality, hold true with probability at least $1 - \varepsilon$. For convenience, the first inequality is true by definition, the third holds under the event in (6.19), whereas the second and the fourth are the consequence of Lemma 6.4.1.

□

6.5 Hypergraph parameter testing

In this section we prove Theorem 6.1.6 that generalizes Theorem 6.1.5 from the case of graphs to r-graphs with arbitrary $r \geq 2$. The proof relies crucially on Lemma 6.4.2 from the previous section. We start with presenting the required notation.

In the current section coloring is always meant as edge coloring. A *k-coloring* of a t-colored r-graph $\mathbf{G} = (G^\alpha)_{\alpha \in [t]}$ is a tk-colored r-graph $\hat{\mathbf{G}} = (G^{\alpha,\beta})_{\alpha \in [t], \beta \in [k]}$ with colors from the set $[t] \times [k]$, where each of the original color classes indexed by $\alpha \in [t]$ is retrieved by taking the union of the new classes corresponding to (α, β) over all $\beta \in [k]$, that is $G^\alpha = \cup_{\beta \in [k]} G^{\alpha,\beta}$. This last operation is called *k-discoloring* of a $[t] \times [k]$-colored graph, we denote it by $[\hat{\mathbf{G}}, k] = \mathbf{G}$. We will sometimes write tk-colored for $[t] \times [k]$-colored graphs when it is clear from the context what we mean.

Similar to the finitary case, a *k-coloring* of a t-colored $\mathbf{W} = (W^\alpha)_{\alpha \in [t]} \in \Xi^{r,t}$ is a tk-colored r-graphon $\hat{\mathbf{W}} = (W^{\alpha,\beta})_{\alpha \in [t], \beta \in [k]}$ with colors from the set $[t] \times [k]$ so that $\sum_{\beta \in [k]} W^{(\alpha,\beta)}(x) = W^\alpha(x)$ for each $x \in [0,1]^{\mathfrak{h}([r], r-1)}$ and $\alpha \in [t]$. The k-discoloring $[\hat{\mathbf{W}}, k]$ of $\hat{\mathbf{W}}$ is defined analogously to the discrete case, and simple r-graphons are treated as 2-colored.

The next lemma is analogous to Lemma 6.2.2. It describes under what metric conditions a k-coloring of a t-colored graphon can be transfered to another object so that the two tk-colored graphons are close in a certain sense. For the sake if completeness we sketch the proof.

Lemma 6.5.1. *Let $\varepsilon > 0$, $t \geq 2$, $\mathbf{U}$ be a t-colored r-graphon that is an $(r, r-1)$-step function with steps $\mathcal{P} = (P_1, \ldots, P_m)$ and $\mathbf{V}$ be a t-colored r-graphon with $d_{\square,r,\mathcal{P}}(\mathbf{U}, \mathbf{V}) \leq \varepsilon$. For any $k \geq 1$ and $[t] \times [k]$-colored r-graphon $\hat{\mathbf{U}}$ that is an $(r, r-1)$-step function with steps from $\mathcal{P}$ such that $[\hat{\mathbf{U}}, k] = \mathbf{U}$ there exists a k-coloring of $\mathbf{V}$ denoted by $\hat{\mathbf{V}}$ so that*

$$d_{\square,r,\mathcal{P}}(\hat{\mathbf{U}}, \hat{\mathbf{V}}) \leq k\varepsilon.$$

Proof. Fix $\varepsilon > 0$, $t \geq 2$, and let $\mathbf{U} = (U^\alpha)_{\alpha \in [t]}$, $\mathbf{V} = (V^\alpha)_{\alpha \in [t]}$ and $\hat{\mathbf{U}} = (U^{\alpha,\beta})_{\alpha \in [t], \beta \in [k]}$ as in the statement of the lemma. Then $\sum_{\alpha=1}^t U^\alpha = 1$ and $\sum_{\beta=1}^k U^{\alpha,\beta} = U^\alpha$ for each $\alpha \in [t]$. Let us define $\hat{\mathbf{V}} = (V^{\alpha,\beta})_{\alpha \in [t], \beta \in [k]}$ that is a k-coloring of $\mathbf{V}$. Set $V^{\alpha,\beta} = V^\alpha [\mathbb{1}_{U^\alpha = 0} \frac{1}{k} + \mathbb{1}_{U^\alpha > 0} \frac{U^{\alpha,\beta}}{U^\alpha}]$, it is easy to see that the factor in the brackets is an $(r, r-1)$-step function with steps $\mathcal{P} = (P_1, \ldots, P_m)$. We estimate the deviation of each pair $U^{\alpha,\beta}$ and $V^{\alpha,\beta}$ from above in the r-cut norm, for this we fix the $(r-1)$-symmetric $S_1, \ldots, S_r \subset [0,1]^{\mathfrak{h}([r-1])}$. Then we have

$$\left| \int_{\cap_{l \in [r]} p_l^{-1}(S_l)} U^{\alpha,\beta} - V^{\alpha,\beta} \right| \leq \sum_{j_1, \ldots, j_r = 1}^{t} \left| \int_{\cap_{l \in [r]} p_l^{-1}(S_l \cap P_{j_l})} U^{\alpha,\beta} - V^{\alpha,\beta} \right|$$

$$= \sum_{j_1,\dots,j_r=1}^{t} \left| \int_{\cap_{l\in[r]} p_l^{-1}(S_l \cap P_{j_l})} (U^\alpha - V^\alpha)[\mathbb{1}_{U^\alpha=0}\frac{1}{k} + \mathbb{1}_{U^\alpha>0}\frac{U^{\alpha,\beta}}{U^\alpha}] \right|$$
$$\leq \|U^\alpha - V^\alpha\|_{\square,r,\mathcal{P}}.$$

Taking the maximum over all $(r-1)$-symmetric measurable r-tuples $S_1,\dots,S_r$ and summing up over all choices of α and β delivers the upper bound we were after. □

The central tool in the main proof is the following lemma which can also be of independent interest. Informally it states that every coloring of a sampled r-graph can be projected onto the graphon from which the graph was sampled from, such that another sampling procedure with a much smaller sample size cannot distinguish the two colored objects.

Lemma 6.5.2. *For every $r \geq 1$, proximity parameter $\delta > 0$, palette sizes $t, k \geq 1$, sampling depth $q_0 \geq 1$ there exists an integer $q_{\text{tv}} = q_{\text{tv}}(r, \delta, q_0, t, k) \geq 1$ such that for every $q \geq q_{\text{tv}}$ the following holds. Let $\mathbf{U} = (U^\alpha)_{\alpha\in[t]}$ be a t-colored r-graphon and let V^α denote $W_{\mathbb{G}(q,U^\alpha)}$ for each $\alpha \in [t]$, also let $\mathbf{V} = (V^\alpha)_{\alpha\in[t]}$, so $\mathbf{W}_{\mathbb{G}(q,\mathbf{U})} = \mathbf{V}$. Then with probability at least $1-\delta$ there exists for every k-coloring $\hat{\mathbf{V}} = (V^{\alpha,\beta})_{\alpha\in[t],\beta\in[k]}$ of $\mathbf{V}$ a k-coloring $\hat{\mathbf{U}} = (U^{\alpha,\beta})_{\alpha\in[t],\beta\in[k]}$ of $\mathbf{U} = (U^\alpha)_{\alpha\in[t]}$ such that we have*

$$d_{\text{tv}}(\mu(q_0, \hat{\mathbf{W}}), \mu(q_0, \hat{\mathbf{U}})) \leq \delta.$$

Proof. We proceed by induction with respect to r. The statement is not difficult to verify for $r = 1$. In this case the 1-graphons U^α and V^α can be regarded as indicator functions of measurable subsets B^α and A^α of $[0,1]$ (so for each $\alpha \in [k]$ we have $U^\alpha = \mathbb{1}_{B^\alpha}$ and $V^\alpha = \mathbb{1}_{A^\alpha}$) that form two partitions associated to $\mathbf{U}$ and $\mathbf{V}$ respectively. Note that $(A^\alpha)_{\alpha\in[k]}$ is obtained from $(B^\alpha)_{\alpha\in[k]}$ by the sampling process. A k-coloring corresponds to a refinement of these partitions with each original class being divided into k measurable parts, that is $A^\alpha = \cup^*_{\beta\in[k]} A^{\alpha,\beta}$ and $V^{\alpha,\beta} = \mathbb{1}_{A^{\alpha,\beta}}$. Moreover, $|t(\mathbf{F}, \hat{\mathbf{U}}) - t(\mathbf{F}, \hat{\mathbf{V}})| = |\prod_{l=1}^{q_0} \lambda(B^{\mathbf{F}(l)}) - \prod_{l=1}^{q_0} \lambda(A^{\mathbf{F}(l)})|$ for any of k-coloring $\hat{\mathbf{U}}$ of $\mathbf{U}$ and for any $[t]\times[k]$-colored $\mathbf{F}$ on q_0 vertices. We may define a suitable coloring of $\mathbf{U}$ by partitioning each of the sets B^α into parts $(B^{\alpha,\beta})_{\beta\in[k]}$ so that the classes satisfy $\lambda(B^{\alpha,\beta}) = \lambda(B^\alpha)\frac{\lambda(A^{\alpha,\beta})}{\lambda(A^\alpha)}$ when $\lambda(A^\alpha) > 0$, and $\lambda(B^{\alpha,\beta}) = \lambda(B^\alpha)\frac{1}{k}$ otherwise for each $\beta \in [k]$. Then by setting $U^{\alpha,\beta} = \mathbb{1}_{B^{\alpha,\beta}}$ and $\hat{\mathbf{U}} = (U^{\alpha,\beta})_{\alpha\in[t],\beta\in[k]}$ we have

$$d_{\text{tv}}(\mu(q_0, \hat{\mathbf{W}}), \mu(q_0, \hat{\mathbf{U}})) = \frac{1}{2} \sum_{\mathbf{F}:|V(\mathbf{F})|=q_0} |t(\mathbf{F}, \hat{\mathbf{U}}) - t(\mathbf{F}, \hat{\mathbf{V}})| \leq \frac{q_0 t^{q_0}}{2} \max_{\alpha\in[t]} |\lambda(A^\alpha) - \lambda(B^\alpha)|,$$

where the sum runs over all $[t]\times[k]$-colored 1-graphs $\mathbf{F}$ on q_0 vertices.

The probability that for a fixed $\alpha \in [t]$ the deviation $|\lambda(A^\alpha) - \lambda(B^\alpha)|$ exceeds $\frac{2\delta}{q_0 t^{q_0}}$ is at most $2\exp(-\frac{2\delta^2 q}{q_0^2 t^{2q_0}})$, the union bound gives the upper bound $2t\exp(-\frac{2\delta^2 q}{q_0^2 t^{2q_0}})$ for the probability that

$$d_{\mathrm{tv}}(\mu(q_0, \hat{\mathbf{W}}), \mu(q_0, \hat{\mathbf{U}})) \le \delta$$

fails for our particular choice for the coloring $\hat{\mathbf{U}}$ of $\mathbf{U}$. We note that the failure probability can be made arbitrary small with the right choice of q, so in particular smaller than δ, therefore $q_{\mathrm{tv}}(1, \delta, q_0, t, k) = \ln(2t/\delta)\frac{q_0^2 t^{2q_0}}{2\delta^2}$ that satisfies the conditions of the lemma.

Now assume that we have already verified the statement of the lemma for $r-1$ and any other choice of the other parameters of q_{tv}. Let us proceed to the proof of the case for r-graphons, therefore let $\delta > 0$, $t, k, q_0 \ge 1$ be arbitrary and fixed, q to be determined below, and $\mathbf{U}$, $\mathbf{V}$, and $\hat{\mathbf{V}}$ as in the condition of the lemma. We start by explicitly constructing a k-coloring $\hat{\mathbf{U}}$ for $\mathbf{U}$, in the second part of the proof we verify that the construction is suitable.

In a nutshell, we proceed as follows. We approximate $\hat{\mathbf{V}}$ by the step function $\hat{\mathbf{Z}}$, and set $\mathbf{Z} = [\hat{\mathbf{Z}}, k]$, and also approximate $\mathbf{U}$ by $\mathbf{W}_1$. Let $\mathbf{W}_2$ be the sampled version of $\mathbf{W}_1$ generated by the same process as $\mathbf{V}$. This way $\mathbf{W}_2$ and $\mathbf{Z}$ are close, hence we can color $\mathbf{W}_2$ using the coloring $\hat{\mathbf{Z}}$ of $\mathbf{Z}$ to obtain $\hat{\mathbf{W}}_2$. The coloring $\hat{\mathbf{W}}_2$ is then transferred onto $\mathbf{W}_1$ using the induction hypothesis applied to the marginals of the step sets of $\mathbf{W}_1$ and $\mathbf{W}_2$ to obtain $\hat{\mathbf{W}}_1$ with $[\hat{\mathbf{W}}_1, k] = \mathbf{W}_1$. Finally, we color $\mathbf{U}$ exploiting the proximity of $\mathbf{U}$ and $\mathbf{W}_1$ using the coloring $\hat{\mathbf{W}}_1$ of $\mathbf{W}_1$.

Our construction may fail to meet the criteria of the lemma, this can be caused at two points in the above outline. For one, it may happen, that $\mathbf{W}_2$ does not approximate $\mathbf{V}$ well enough, and the second time, when we transfer $\hat{\mathbf{W}}_2$ onto $\mathbf{W}_1$ using the induction hypothesis with $r-1$, as the current lemma leaves space for probabilistic error. These two events are independent from the particular choice of $\hat{\mathbf{V}}$ and their probability can be made sufficiently small, we aim for to show this. We proceed now to the technical details.

Let $\Delta = \Psi(r, \delta, q_0, t, k) = \frac{\delta}{4k(kt)^{q_0^r} q_0^r}$. Set $t_2 = t_{\mathrm{reg}}(r, tk, \Delta, 1)$ and $t_1 = t_{\mathrm{reg}}(r, t, (\Delta/t_2^r t)^{2^r} 2^{-r-1}, t_2)$, where t_{reg} is the function from Lemma 3.3.27, and define

$$q_{\mathrm{tv}}(r, \delta, q_0, t, k) = \max\{q_{\mathrm{tv}}(r-1, \delta/4, q_0, t_1, t_2), q_{\mathrm{cut}}(r, t, \Delta, t_2)\},$$

where q_{cut} is the function from Lemma 6.4.2.

Note that $t_2 \le \exp^{(2)}(c(1/\Delta)^3)$ and $t_1 \le \exp^{(4)}(c(1/\Delta)^3)$ for a large enough constant $c > 0$. If we assume that $q_{\mathrm{tv}}(r-1, \delta, q_0, t, k) \le \exp^{(d)}(c_{r-1}\left(\frac{1}{\Delta'}\right)^3)$ for some positive integer d and real $c_{r-1} > 0$, where $\Delta' = \Psi(r-1, \delta, q_0, t, k) = \frac{\delta}{4k(kt)^{q_0^{r-1}} q_0^{r-1}}$, then it follows

$$q_{\mathrm{tv}}(r-1, \delta/4, q_0, t_1, t_2) \le \exp^{(d+4)}(c_r(1/\Delta)^3) \tag{6.20}$$

for some $c_r > 0$. Since we can adjust the constant factor c_{r-1} in a way that $q_{\mathrm{tv}}(r-1, \delta/4, q_0, t_1, t_2) \ge q_{\mathrm{cut}}(r, t, \Delta, t_2)$ holds for any possible choice of the parameters we conclude that $q_{\mathrm{tv}}(r, \delta, q_0, t, k)$ is upper bounded by $\exp^{(4(r-1))}(c_r(1/\Delta)^3)$.

Let $q \ge q_{\mathrm{tv}}(r, \delta, q_0, t, k)$ be arbitrary. We describe now the steps of constructing $\hat{\mathbf{U}}$ that satisfies the conditions of the lemma.

We approximate $\hat{\mathbf{V}}$ by some function $\hat{\mathbf{Z}}$ that is only given implicitly by means of Lemma 3.3.27 and is of the form $\hat{\mathbf{Z}} = \hat{\mathbf{V}}_{\mathcal{R}}$. We have

$$\sup_{Q, t_Q \le t_{\mathcal{R}}} d_{\square,r,Q}(\hat{\mathbf{V}}, \hat{\mathbf{Z}}) \le \Delta,$$

and $t_{\mathcal{R}} \le t_2$ holds. Also, by Corollary 3.3.25 we have

$$d_{\mathrm{tv}}(\mu(q_0, \hat{\mathbf{V}}), \mu(q_0, \hat{\mathbf{Z}})) \le \delta/(8k). \tag{6.21}$$

We set $\mathbf{Z} = [\hat{\mathbf{Z}}, k]$, consequently

$$\sup_{Q, t_Q \le t_{\mathcal{R}}} d_{\square,r,Q}(\mathbf{V}, \mathbf{Z}) \le \Delta, \tag{6.22}$$

and $\mathbf{Z} = \mathbf{V}_{\mathcal{R}}$. Note that $\mathbf{Z}$ and $\hat{\mathbf{Z}}$ depend on $\hat{\mathbf{V}}$.

Define the r-arrays $B_1, \dots, B_t$ such that for each $\alpha \in [t]$ it holds that

$$Z^{\alpha}(x_{\mathfrak{h}([r],r-1)}) = \sum_{i_1,\dots,i_r=1}^{t_{\mathcal{R}}} B_{\alpha}(i_1, \dots, i_r) \prod_{l=1}^{r} \mathbb{1}_{R_{i_l}}(x_{\mathfrak{h}([r]\setminus\{l\})}),$$

further define also the r-arrays $(B_{\alpha}^{\beta})_{\alpha\in[t],\beta\in[k]}$ such that

$$Z^{\alpha,\beta}(x_{\mathfrak{h}([r],r-1)}) = \sum_{i_1,\dots,i_r=1}^{t_{\mathcal{R}}} B_{\alpha}^{\beta}(i_1, \dots, i_r) \prod_{l=1}^{r} \mathbb{1}_{R_{i_l}}(x_{\mathfrak{h}([r]\setminus\{l\})})$$

for each $\alpha \in [t], \beta \in [k]$. Clearly, $B_{\alpha}(i_1, \dots, i_r) = \sum_{\beta=1}^{k} B_{\alpha}^{\beta}(i_1, \dots, i_r)$ for each $i_1, \dots, i_r \in [t_{\mathcal{R}}]$.

We apply again Lemma 3.3.27 with the proximity parameter $(\Delta/t_2^r t)^{2^r} 2^{-r-1}$ and the multiplier parameter t_2 in the condition regarding the partitions Q to the r-graphon $\mathbf{U}$ to approximate it by $\mathbf{U}_{\mathcal{P}} = \mathbf{W}_1 = (W_1^1, \dots, W_1^t)$ with steps in $\mathcal{P}$ that satisfies

$$\sup_{Q, t_Q \le t_{\mathcal{P}} t_2} d_{\square,r,Q}(\mathbf{W}_1, \mathbf{U}) \le (\Delta/t_2^r t)^{2^r} 2^{-r-1}, \tag{6.23}$$

where the supremum runs over all $(r-1)$-symmetric partitions Q of $[0,1]^{\mathfrak{h}([r-1])}$ with at most $t_{\mathcal{P}} t_2$ classes, and $t_{\mathcal{P}} \le t_1$.

Applying structure preserving transformations to $[0,1]^{\mathfrak{h}([r-1])}$ the classes of $\mathcal{P}$ can be considered as piled up, meaning that for each $y \in [0,1]^{\mathfrak{h}([r-1],r-2)}$ the fibers $\{y\} \times [0,1]$ are partitioned by the intersections with the classes of $\mathcal{P}$ into the intervals $[0, a_1), [a_1, a_2), \dots, [a_{t_1-1}, a_{t_1}]$ with $\{y\} \times [a_{j-1}, a_j) = (\{y\} \times [0,1]) \cap P_j$. We introduce the r-dimensional real arrays $A_1, \dots, A_t$ in order to describe the explicit form of the W_1^{α}

graphons. So,

$$W_1^\alpha(x_{\mathfrak{h}([r],r-1)}) = \sum_{i_1,\dots,i_r=1}^{t_{\mathcal{P}}} A_\alpha(i_1,\dots,i_r) \prod_{l=1}^{r} \mathbb{1}_{P_{i_l}}(x_{\mathfrak{h}([r]\setminus\{l\})}).$$

Define $\mathbf{W}_2 = (W_2^\alpha)_{\alpha\in[t]}$ to be the r-graphon representing $\mathbb{G}(q,\mathbf{W}_1)$, so W_2^α represents $\mathbb{G}(q,W_1^\alpha)$ for each $\alpha \in [t]$. The steps of $\mathbf{W}_2$ are denoted by $\mathcal{P}'$. Then it follows from Lemma 6.4.2 that

$$\sup_{Q,t_Q\le t_2} d_{\square,r,Q}(\mathbf{W}_2,\mathbf{V}) \le \Delta,$$

with probability at least $1-\Delta$, so consequently

$$d_{\square,r,\mathcal{R}}(\mathbf{W}_2,\mathbf{V}) \le \Delta,$$

with the same upper bound on the failure probability as above. Furthermore, with (6.22) we have

$$d_{\square,r,\mathcal{R}}(\mathbf{W}_2,\mathbf{Z}) \le 2\Delta. \tag{6.24}$$

Also,

$$W_2^\alpha(x_{\mathfrak{h}([r],r-1)}) = \sum_{i_1,\dots,i_r=1}^{t_{\mathcal{P}}} A_\alpha(i_1,\dots,i_r) \prod_{l=1}^{r} \mathbb{1}_{P'_{i_l}}(x_{\mathfrak{h}([r]\setminus\{l\})}),$$

for each $\alpha \in [t]$ and

$$P'_j = \cup_{(p_1,\dots,p_{r-1})\in I_j} [\frac{p_1-1}{q},\frac{p_1}{q}] \times\cdots\times [\frac{p_r-1}{q},\frac{p_r}{q}] \times [0,1]\times\cdots\times[0,1]$$

with $I_j = \{\, (p_1,\dots,p_{r-1}) \mid X_{r[\{p_1,\dots,p_{r-1}\}]} \in P_j \,\}$ for every $j \in [t_{\mathcal{P}}]$. Note that $\mathcal{P}' = (P'_j)_{j\in[t_{\mathcal{P}}]}$ is a symmetric partition.

We define a k-coloring $\hat{\mathbf{W}}_2$ of $\mathbf{W}_2$ that satisfies

$$d_{\square,r,\mathcal{R}}(\hat{\mathbf{Z}},\hat{\mathbf{W}}_2) \le 2k\Delta.$$

Such a k-coloring exists by Lemma 6.5.1 and (6.24). It follows by Corollary 3.3.25 that

$$d_{\mathrm{tv}}(\mu(q_0,\hat{\mathbf{Z}}),\mu(q_0,\mathbf{W}_2)) \le \delta/4. \tag{6.25}$$

The graphon $\hat{\mathbf{W}}_2$ is a symmetric step function with steps that form the coarsest partition that refines both $\mathcal{P}'$ and $\mathcal{R}$, we denote this $(r-1)$-symmetric partition of $[0,1]^{\mathfrak{h}([r-1])}$ by $\mathcal{S}$, the number of its classes satisfies $t_{\mathcal{S}} = t_{\mathcal{P}'}t_{\mathcal{R}} \le t_1t_2$.

Let us define the $t_{\mathcal{P}}$-colored $(r-1)$-graphon $\mathbf{w} = (w^1,\dots,w^{t_{\mathcal{P}}})$ that is obtained from the classes of the partition $\mathcal{P}$ by integrating over the coordinate corresponding to the set $[r-1]$, that is $w^i(x_{\mathfrak{h}([r-1],r-2)}) = \int_0^1 \mathbb{1}_{P_i}(x_{\mathfrak{h}([r-1])})\mathrm{d}x_{[r-1]}$. In the same way we define the $t_{\mathcal{P}}$-colored $(r-1)$-graphon $\mathbf{u} = (u^1,\dots,u^{t_{\mathcal{P}}})$ corresponding to the partition $\mathcal{P}'$, as well as the $[t_{\mathcal{P}}]\times[t_{\mathcal{R}}]$-colored $\hat{\mathbf{u}} = (u^{i,j})_{i\in[t_{\mathcal{P}}],j\in[t_{\mathcal{R}}]}$, where it holds that $\mathbf{u} = [\hat{\mathbf{u}},t_{\mathcal{R}}]$ and $\hat{\mathbf{u}}$ is the $t_{\mathcal{R}}$-coloring of $\mathbf{u}$ corresponding to the partition $\mathcal{S}$. Note that $\mathbf{w}$, $\mathbf{u}$, and $\hat{\mathbf{u}}$ are

$(r-1)$-symmetric, since their origin partitions were symmetric. As the partition $\mathcal{P}'$ was constructed via the same sampling procedure as $\mathbf{V}$ and $\mathbf{W}_2$, it holds that $\mathbf{u} = \mathbb{G}(q, \mathbf{w})$ and $u^i = \mathbb{G}(q, w^i)$ for each $i \in [t_{\mathcal{P}}]$.

We can assert that due to the induction hypothesis there exists a $t_{\mathcal{R}}$-coloring $\hat{\mathbf{w}} = (w^{i,j})_{i\in[t_{\mathcal{P}}],j\in[t_{\mathcal{R}}]}$ of $\mathbf{w}$ that satisfies

$$d_{\mathrm{tv}}(\mu(q_0, \hat{\mathbf{w}}), \mu(q_0, \hat{\mathbf{u}})) \le \delta/4$$

with probability at least $1 - \delta/4$ for $q \ge q_{\mathrm{tv}}(r-1, \delta/4, q_0, t_1, t_2)$.

We construct a k-coloring for $\mathbf{W}_1$ next. Recall the proof of Lemma 6.5.1, therefore we have that

$$\begin{aligned} &W_2^{\alpha,\beta}(x_{\mathfrak{h}([r],r-1)}) \\ &\quad = \sum_{i_1,\dots,i_r=1}^{t_{\mathcal{P}}} \sum_{j_1,\dots,j_r=1}^{t_{\mathcal{R}}} A_\alpha(i_1,\dots,i_r)\left[\frac{B_\alpha^\beta(j_1,\dots,j_r)}{B_\alpha(j_1,\dots,j_r)}\mathbb{1}_{B_\alpha>0} + \frac{1}{k}\mathbb{1}_{B_\alpha=0}\right]\prod_{m=1}^{r}\mathbb{1}_{P'_{i_m}\cap R_{j_m}}(x_{\mathfrak{h}([r]\setminus\{m\})}), \end{aligned} \tag{6.26}$$

and set $A_\alpha^\beta((i_1,j_1),\dots,(i_r,j_r)) = A_\alpha(i_1,\dots,i_r)\left[\frac{B_\alpha^\beta(j_1,\dots,j_r)}{B_\alpha(j_1,\dots,j_r)}\mathbb{1}_{B_\alpha>0} + \frac{1}{k}\mathbb{1}_{B_\alpha=0}\right]$ for all $\alpha \in [t]$, $\beta \in [k]$ and $((i_1,j_1),\dots,(i_r,j_r)) \in ([t_{\mathcal{P}}]\times[t_{\mathcal{R}}])^r$.

We utilize the $t_{\mathcal{R}}$-coloring $\hat{\mathbf{w}}$ of the $(r-1)$-graphon $\mathbf{w}$ to construct a refined partition of $\mathcal{P}$ that resembles $\mathcal{S}$ in order to enable the construction of a k-coloring of $\mathbf{W}_1$ along the same lines as in (6.26). Let

$$\begin{aligned} P_{i,j} = \{\, & x \in [0,1]^{\mathfrak{h}([r-1])} \mid \\ & \sum_{l=1}^{i-1} w^l(x_{\mathfrak{h}([r-1],r-2)}) + \sum_{l=1}^{j-1} w^{i,l}(x_{\mathfrak{h}([r-1],r-2)}) \le x_{[r-1]} < \sum_{l=1}^{i-1} w^l(x_{\mathfrak{h}([r-1],r-2)}) + \sum_{l=1}^{j} w^{i,l}(x_{\mathfrak{h}([r-1],r-2)})\,\} \end{aligned}$$

for each $i \in [t_{\mathcal{P}}]$ and $j \in [t_{\mathcal{R}}]$. Let $\mathcal{P}'' = (P_{i,j})_{i\in[t_{\mathcal{P}}],j\in[t_{\mathcal{R}}]}$.

Clearly, $(P_{i,j})_{j\in[t_{\mathcal{R}}]}$ is an $(r-1)$-symmetric $t_{\mathcal{R}}$-partition of the set P_i, and $w^{i,j}(x_{\mathfrak{h}([r-1],r-2)}) = \int_0^1 \mathbb{1}_{P_{i,j}}(x_{\mathfrak{h}([r-1])})\mathrm{d}x_{[r-1]}$. We are able now to describe the k-coloring of the $\mathbf{W}_1$, define

$$W_1^{\alpha,\beta}(x_{\mathfrak{h}([r],r-1)}) = \sum_{i_1,\dots,i_r=1}^{t_{\mathcal{P}}} \sum_{j_1,\dots,j_r=1}^{t_{\mathcal{R}}} A_\alpha^\beta((i_1,j_1),\dots,(i_r,j_r)) \prod_{m=1}^{r}\mathbb{1}_{P_{i_m,j_m}}(x_{\mathfrak{h}([r]\setminus\{m\})}). \tag{6.27}$$

Note that $\hat{\mathbf{W}}_1 = (W_1^{\alpha,\beta})_{\alpha\in[t],\beta\in[k]}$ is a step function whose steps form the partition $\mathcal{P}''$ that refines $\mathcal{P}$, but the regularity property (6.23) of $\mathbf{W}_1$ allows for

$$d_{\square,r,\mathcal{P}''}(\mathbf{U}, \mathbf{W}_1) \le (\Delta/t_2^r t)^{2^r} 2^{-r-1} \le \Delta/2. \tag{6.28}$$

We will next elaborate on the correctness of the inductive step of the construction. Let us consider the tk-colored random r-graph $\mathbb{G}(q_0, \hat{\mathbf{W}}_1)$, it is generated by the independent uniformly distributed $[0,1]$-valued random variables $\{ Y_S \mid S \in \mathfrak{h}([q_0], r) \}$. The color of each edge $e = \{e_1, \dots, e_2\} \in \binom{[q_0]}{r}$ is decided by determining first the unique tuple (up to coordinate permutations) $((i_1, j_1), \dots, (i_r, j_r)) \in ([t_{\mathcal{P}}] \times [t_{\mathcal{R}}])^r$ such that $(Y_S)_{S \in \mathfrak{h}(e \setminus \{e_l\})} \in P_{i_l, j_l}$, and then check for which pair $\alpha \in [t]$, $\beta \in [k]$ it holds that

$$\sum_{l=1}^{\alpha-1} A_l((i_1, j_1), \dots, (i_r, j_r)) + \sum_{l=1}^{\beta-1} A_\alpha^l((i_1, j_1), \dots, (i_r, j_r)) < Y_e$$
$$\leq \sum_{l=1}^{\alpha-1} A_l((i_1, j_1), \dots, (i_r, j_r)) + \sum_{l=1}^{\beta} A_\alpha^l((i_1, j_1), \dots, (i_r, j_r)),$$

then add the color (α, β) to e with the corresponding index. It is convenient to view this process as first randomly $t_{\mathcal{P}''}$-coloring an $(r-1)$-uniform template hypergraph $\mathbf{G}_1$, whose edges are the simplices of the original edges, here we add a color (i, j) to an $(r-1)$-edge e' whenever $(Y_S)_{S \in \mathfrak{h}(e')} \in P_{i,j}$, and conditioned on $\mathbf{G}_1$ we subsequently make independent choices for each edge to determine their color based on the arrays A_α^β by means of the random variables $\{ Y_S \mid S \in \binom{[q_0]}{r} \}$ at the top level.

Let us turn to the tk-colored $\mathbb{G}(q_0, \hat{\mathbf{W}}_2)$, the above description of the random process generating this object remains conceptually valid also for this random graph, the r-arrays A_α^β are identical to the case above, only the partition $\mathcal{P}''$ has to be altered to $\mathcal{S}$. Similarly as above, we introduce the random $(r-1)$-uniform $t_{\mathcal{P}''}$-colored hypergraph $\mathbf{G}_2$ that is generated as above adapted to $\mathbb{G}(q_0, \hat{\mathbf{W}}_2)$. That means that the $(r-1)$-edges are colored by indices of the classes that form the partition $\mathcal{S}$ through the process that generates $\mathbb{G}(q_0, \hat{\mathbf{W}}_2)$, see above. The key observation here is that conditioned on $\mathbf{G}_1 = \mathbf{G}_2$, one can couple $\mathbb{G}(q_0, \hat{\mathbf{W}}_1)$ and $\mathbb{G}(q_0, \hat{\mathbf{W}}_2)$ so that the two random graphs coincide with conditional probability 1. Recall that a coupling is only another name for a joint probability space for the two random objects with the marginal distributions following $\mu(q_0, \mathbf{W}_1)$ and $\mu(q_0, \mathbf{W}_2)$, respectively. As the conditional distributions for the choices of colors for the r-edges are identical, provided that $\mathbf{G}_1 = \mathbf{G}_2$, the coupling is trivial. In order to construct a good unconditional coupling we require another coupling, now of $\mathbf{G}_1$ and $\mathbf{G}_2$, so that $\mathbb{P}(\mathbf{G}_1 \neq \mathbf{G}_2)$ is negligibly small for our purposes, and whose existence is exactly what the induction hypothesis in the case of $(r-1)$-uniform hypergraphs ensures, when q is large enough.

As $q \geq q_{\mathrm{tv}}(r-1, \delta/4, q_0, t_1, t_2)$, the induction hypothesis enables us to use that there exists for any $\hat{\mathbf{u}}$ a $\hat{\mathbf{w}}$ so that $d_{\mathrm{tv}}(\mu(q_0, \hat{\mathbf{u}}), \mu(q_0, \hat{\mathbf{w}})) \leq \delta/4$ holds with probability at least $1 - \delta/4$ for each $\hat{\mathbf{u}}$ simultaneously, which in turn implies that there is a coupling of the $t_1 t_2$-colored random $(r-1)$-graphs $\mathbf{G}_1$ and $\mathbf{G}_2$ so that $\mathbb{P}(\mathbf{G}_1 \neq \mathbf{G}_2) \leq \delta/2$.

It follows that there exists a coupling of $\mathbb{G}(q_0, \hat{\mathbf{W}}_1)$ and $\mathbb{G}(q_0, \hat{\mathbf{W}}_2)$ such that $\mathbb{P}(\mathbb{G}(q_0, \hat{\mathbf{W}}_1) \neq \mathbb{G}(q_0, \hat{\mathbf{W}}_2)) \leq \delta/2$ due to the discussion above, which in turn implies

$$d_{\mathrm{tv}}(\mu(q_0, \hat{\mathbf{W}}_1), \mu(q_0, \hat{\mathbf{W}}_2)) \leq \delta/4. \tag{6.29}$$

Since $\hat{\mathbf{W}}_1$ has at most $t_{\mathcal{P}}t_2$ steps, Lemma 6.5.1 together with (6.28) provides the existence of $\hat{\mathbf{U}}$ with $[\hat{\mathbf{U}}, k] = \mathbf{U}$ and the bound $d_{\square,r}(\hat{\mathbf{U}}, \hat{\mathbf{W}}_1) \leq \frac{k\Delta}{2}$. The application of Corollary 3.3.25 implies

$$d_{\mathrm{tv}}(\mu(q_0, \hat{\mathbf{W}}_1), \mu(q_0, \hat{\mathbf{U}})) \leq \delta/16. \tag{6.30}$$

Evoking the triangle inequality and summing up the upper bounds on the respective deviations (6.21), (6.25), (6.29), and (6.30) we conclude that

$$\begin{aligned} d_{\mathrm{tv}}(\mu(q_0, \hat{\mathbf{V}}), \mu(q_0, \hat{\mathbf{U}})) &\leq d_{\mathrm{tv}}(\mu(q_0, \hat{\mathbf{V}}), \mu(q_0, \hat{\mathbf{Z}})) + d_{\mathrm{tv}}(\mu(q_0, \hat{\mathbf{Z}}), \mu(q_0, \hat{\mathbf{W}}_2)) \\ &+ d_{\mathrm{tv}}(\mu(q_0, \hat{\mathbf{W}}_2), \mu(q_0, \hat{\mathbf{W}}_1)) + d_{\mathrm{tv}}(\mu(q_0, \hat{\mathbf{W}}_1), \mu(q_0, \hat{\mathbf{U}})) \leq \left(\frac{1}{8k} + \frac{1}{4} + \frac{1}{4} + \frac{1}{16}\right)\delta < \delta, \end{aligned}$$

the overall error probability is at most $\delta/2 + \Delta/2$, which is at most δ.

□

With Lemma 6.5.2 at hand we can overcome the difficulties we discussed in Chapter 3 caused by the properties of the r-cut norm for $r \geq 3$ in contrast to the case $r = 2$, we turn to prove the main result of the paper.

Proof of Theorem 6.1.6. We regard simple hypergraphs as 2-colored r-graphs, in the following the term simple should be understood this way at each appearance. Let the $2k$-colored witness parameter of the nondeterministically testable r-graph parameter f be denoted by g, whose sample complexity is at most $q_g(\varepsilon)$ for each proximity parameter $\varepsilon > 0$. Set $d(r, \varepsilon, q_0, k, t) = \frac{[q_g(\varepsilon)^r \ln(tk) - \ln(\varepsilon)][2(tk)^{q_0^r} q_0^2]}{\varepsilon^2}$. Let $\varepsilon > 0$ be fixed and define $q_f(\varepsilon) = \max\{q_{\mathrm{tv}}(r, \varepsilon/4, q_g(\varepsilon/4), k, 3); \frac{4}{\varepsilon} q_g^2(\varepsilon/2); d(r, \varepsilon/4, q_g(\varepsilon/4), k, 2)\}$. We will show that for every $q \geq q_f(\varepsilon)$ the condition

$$\mathbb{P}(|f(G) - f(\mathbb{G}(q_f(\varepsilon), G)| > \varepsilon) < \varepsilon.$$

is satisfied for each G. Let $q \geq q_f(\varepsilon)$ arbitrary but fixed and G be a fixed simple r-graph on n vertices.

First we show that $f(\mathbb{G}(q, G)) \geq f(G) - \varepsilon/4$ with probability at least $1 - \varepsilon/4$. For this let us select a k-coloring $\mathbf{G}$ of G such that $f(G) = g(\mathbf{G})$, then the random k-colored graph $\mathbf{F} = \mathbb{G}(q, \mathbf{G})$ is a k-coloring of $\mathbb{G}(q, G)$, therefore $f(\mathbb{G}(q, G)) \geq g(\mathbf{F})$, but since $q \geq q_q(\varepsilon/4)$ we know from the testability of g that $g(\mathbf{F}) \geq g(\mathbf{G}) - \varepsilon/4$ with probability at least $1 - \varepsilon/4$, which verifies our claim.

The more difficult part is to show that $f(\mathbb{G}(q, G)) \leq f(G) + \varepsilon$ with failure probability at most $\varepsilon/2$. Let us denote the random r-graph $\mathbb{G}(q, G)$ by F. We claim that with probability at least $1 - \varepsilon/2$ there exists for any k-coloring $\mathbf{F}$ of F a k-coloring $\mathbf{G}$ of G such that $|g(\mathbf{F}) - g(\mathbf{G})| \leq \varepsilon$, this suffices to verify the statement of the theorem.

Our proof exploits that the difference of the g values between two colored r-graphs $\mathbf{F}$ and $\mathbf{G}$ can be upper bounded by

$$|g(\mathbf{F}) - g(\mathbf{G})| \leq |g(\mathbf{F}) - g(\mathbb{G}(q_g(\varepsilon/4), \mathbf{F}))| + |g(\mathbf{G}) - g(\mathbb{G}(q_g(\varepsilon/4), \mathbf{G}))| \leq \varepsilon/2,$$

whenever there exists a coupling of the two random $2k$-colored r-graphs $\mathbb{G}(q_g(\varepsilon/4), \mathbf{G})$ and $\mathbb{G}(q_g(\varepsilon/4), \mathbf{F})$ appearing in the above formula such that they are equal with probability larger than $\varepsilon/2$. Set $q_0 = q_g(\varepsilon/4)$. We will show that with high probability for every $\mathbf{F}$ there exists a $\mathbf{G}$ that satisfies the previous conditions.

Recall that coupling is a probability space together with the random r-graphs $\mathbf{G}_1$ and $\mathbf{G}_2$ defined on it such that $\mathbf{G}_1$ has the same marginal distribution as $\mathbb{G}(q_0, \mathbf{G})$ and $\mathbf{G}_2$ has the same as $\mathbb{G}(q_0, \mathbf{F})$, their joint distribution is constructed in a way that serves our current purposes by maximizing the probability that they coincide. When the target spaces are finite as in our case then a coupling that satisfies the above condition can be easily constructed whenever $d_{\mathrm{tv}}(\mu(q_0, \mathbf{G}), \mu(q_0, \mathbf{F})) \leq 1 - \varepsilon/2$, see (3.32).

By Lemma 6.5.2 for 3-colored r-graphs (there are 3 types of entries in the graphon representation of simple r-graphs, edges, non-edges, and diagonal elements) it follows that with probability at least $1 - \varepsilon/4$ for each $\mathbf{F}$ there exists a $3k$-colored $\mathbf{U}$ with $[\mathbf{U}, k] = W_G$ such that $d_{\mathrm{tv}}(\mu(q_0, \mathbf{U}), \mu(q_0, \mathbf{W_F})) \leq \varepsilon/4$. Let us condition on this event and let $\mathbf{F}$ be fixed. From (3.33) we know that $d_{\mathrm{tv}}(\mu(q_0, \mathbf{G}), \mu(q_0, \mathbf{W_G})) \leq q_0^2/n \leq \varepsilon/4$ and $d_{\mathrm{tv}}(\mu(q_0, \mathbf{F}), \mu(q_0, \mathbf{W_F})) \leq q_0^2/q \leq \varepsilon/4$. It remains to produce a $3k$-coloring of W_G from any fixed $3k$-colored $\mathbf{U}$ (k out of the $3k$ colors of $\mathbf{U}$ correspond exclusively to diagonal cubes, so can be neglected). We do this by randomization, let $(X_{\{i\}})_{i \in [n]}$ be independent uniform random variables distributed on $[\frac{i-1}{n}, \frac{i}{n}]$, and let $(X_S)_{S \in \mathfrak{h}([n],r) \setminus \mathfrak{h}([n],1)}$ be independent uniform random variables on $[0,1]$. We can define $\mathbf{W_G}$ to take the color $\mathbf{U}(X_{\mathfrak{h}(e)})$ on the set $[\frac{e_1-1}{n}, \frac{e_1}{n}] \times \cdots \times [\frac{e_r-1}{n}, \frac{e_r}{n}] \times [0,1] \times \cdots \times [0,1]$ for $e = \{e_1, \ldots, e_r\} \in \binom{[n]}{r}$. For any fixed $\mathbf{H} \in \Pi_{q_0}^{r,2k}$ basic martingale methods deliver

$$\mathbb{P}(|t(\mathbf{H}, \mathbf{W_G}) - t(\mathbf{H}, \mathbf{U})| \geq \delta) \leq 2 \exp(-\frac{\delta^2 n}{2 q_0^2})$$

for any $\delta > 0$, therefore when setting $\delta = \frac{\varepsilon}{4(2k)^{q_0^r}}$ we get that

$$d_{\mathrm{tv}}(\mu(q_0, \mathbf{U}), \mu(q_0, \mathbf{W_G})) = \frac{1}{2} \sum_{\mathbf{H} \in \Pi_{q_0}^{r,2k}} |t(\mathbf{H}, \mathbf{W_G}) - t(\mathbf{H}, \mathbf{U})| \leq \varepsilon/4$$

with probability at least $1 - \varepsilon/4$, since $n \geq q \geq d(r, \varepsilon/4, q_0, k, 2) = \frac{[q_0^r \ln(2k) - \ln(\varepsilon/4)][32(2k)^{q_0^r} q_0^2]}{\varepsilon^2}$. Summing up terms gives

$$\begin{aligned} d_{\mathrm{tv}}(\mu(q_0,\mathbf{F}), \mu(q_0, \mathbf{G})) &\leq d_{\mathrm{tv}}(\mu(q_0, \mathbf{F}), \mu(q_0, \mathbf{W_F})) + d_{\mathrm{tv}}(\mu(q_0, \mathbf{W_F}), \mu(q_0, \mathbf{U})) \\ &+ d_{\mathrm{tv}}(\mu(q_0, \mathbf{U}), \mu(q_0, \mathbf{W_G})) + d_{\mathrm{tv}}(\mu(q_0, \mathbf{W_G}), \mu(q_0, \mathbf{G})) \leq \varepsilon, \end{aligned}$$

with failure probability at most $\varepsilon/2$, this concludes our proof.

□

6.6 Nondeterministically testable properties

The concept of nondeterministic testing was originally introduced for testing properties by Lovász and Vesztergombi [97], and remarkable progress has been made in that context, see [61] and [97], the estimation of parameters, which was our main issue in the preceding sections, is in close relationship to that notion. For related developments in combinatorial property testing using regularity methods we refer to the respective part in Chapter 1.

We present a slight variant of the definition of testability of properties in the nondeterministic sense and construct a tester from the tester of the witness property with the aid of Lemma 6.5.2 that achieves the same sample complexity as in the parameter testing case. This result connects our contribution to previous efforts more directly compared to the content of the preceding sections, and answers the question posed in [97] asking whether the equivalence of the two testability notions persists for uniform hypergraphs of higher rank similar to the case of graphs.

We recall Definition 3.2.2. An r-graph property $\mathcal{P}$ is testable, if there exists another r-graph property $\hat{\mathcal{P}}$ called the sample property, such that

(a) $\mathbb{P}(\mathbb{G}(q, G) \in \hat{\mathcal{P}}) \geq \frac{2}{3}$ for every $G \in \mathcal{P}$ and $q \geq 1$, and

(b) for every $\varepsilon > 0$ there is an integer $q_{\mathcal{P}}(\varepsilon) \geq 1$ such that for every $q \geq q_{\mathcal{P}}(\varepsilon)$ and every G that is ε-far from $\mathcal{P}$ on at least q vertices we have that $\mathbb{P}(\mathbb{G}(q, G) \in \hat{\mathcal{P}}) \leq \frac{1}{3}$.

Testability for colored r-graphs is defined analogously.

We remark that ε-far here means that one has to modify at least $\varepsilon|V(G)|^2$ edges in order to obtain an element of $\mathcal{P}$. Note that $\frac{1}{3}$ and $\frac{2}{3}$ in the definition can be replaced by arbitrary constants $0 < a < b < 1$, this change may alter the corresponding certificate $\hat{\mathcal{P}}$ and the function $q_{\mathcal{P}}$, but not the characteristic of $\mathcal{P}$ being testable or not. Let $\mathcal{P}_n$ denote the elements of $\mathcal{P}$ of size n.

Next we formulate the definition of nondeterministic testability in terms of k-colorings.

Definition 6.6.1. *An r-graph property $\mathcal{P}$ is nondeterministically testable, if there exists an integer $k \geq 1$ and a $2k$-colored r-graph property Q called the witness property that is testable in the sense of Definition 3.2.2 satisfying $[Q, k] = \{ [\mathbf{G}, k] \mid \mathbf{G} \in Q \} = \mathcal{P}$.*

We conduct the proof of the main theorem in this section next.

Proof of Theorem 6.1.7. Let $\mathcal{P}$ be a nondeterministically testable property with witness property Q of $2k$-colored r-graphs. We employ the combinatorial language with counting subgraph densities when referring to Q and its testability, and the probabilistic language of picking random subgraphs in a uniform way when handling $\mathcal{P}$ in order to facilitate readability.

Let $\hat{Q}$ be the corresponding sample property that certifies the testability of Q and q_Q be the sample complexity corresponding to the thresholds 1/5 and 4/5, that is

(i) if $\mathbf{G} \in \mathcal{Q}$, then for every and $q \geq 1$ we have $t_{\text{inj}}(\hat{\mathcal{Q}}_q, \mathbf{G}) \geq 4/5$, and

(ii) for every $\varepsilon > 0$, if $\mathbf{G}$ is ε-far from $\mathcal{Q}$, then for every $q \geq q_{\mathcal{Q}}(\varepsilon)$ we have that $t_{\text{inj}}(\hat{\mathcal{Q}}_q, \mathbf{G}) \leq 1/5$.

Our task is to construct a property $\hat{\mathcal{P}}$ together with a function $q_{\mathcal{P}}$ such that they fulfill the conditions of Definition 3.2.2. We are free to specify the error thresholds by the remark after Definition 3.2.2, we set them to 2/5 and 3/5.

Let n be a positive integer, and let $\varepsilon_n > 0$ be the infimum of all positive reals δ that satisfy $n \geq \max\{q_{\text{tv}}(r, 1/10, q_{\mathcal{Q}}(\delta), 3, k); 100 q^2_{\mathcal{Q}}(\delta); d(r, 1/10, q_{\mathcal{Q}}(\delta), k, 2)\}$, where q_{tv} and d are as in Lemma 6.5.2. Define for each n the set

$$\hat{\mathcal{P}}_n = \{ H \in \Pi^r_n \mid \text{there exists a } k\text{-coloring } \mathbf{H} \text{ of } H \text{ such that } t_{\text{inj}}(\hat{\mathcal{Q}}_{q_{\mathcal{Q}}(\varepsilon_n)}, \mathbf{H}) \geq 3/5 \},$$

and let $\hat{\mathcal{P}} = \cup_{n=1}^{\infty} \hat{\mathcal{P}}_n$.

We set $q_{\mathcal{P}}(\varepsilon) = \max\{q_{\text{tv}}(r, 1/10, q_{\mathcal{Q}}(\varepsilon), 3, k); 100 q^2_{\mathcal{Q}}(\varepsilon); d(r, 1/10, q_{\mathcal{Q}}(\delta), k, 2)\}$. We are left to check if the two conditions for testability of $\mathcal{P}$ hold with $\hat{\mathcal{P}}$ and $q_{\mathcal{P}}$ described as above. Assume for the rest of the proof that $n \geq q_{\mathcal{P}}(\varepsilon_n)$ for each n for simplicity, the general case follows along the same lines with some technical difficulties.

First let $G \in \mathcal{P}$, we have to show that for every $q \geq 1$ integer we have that $\mathbb{G}(q, G) \in \hat{\mathcal{P}}_q$ with probability at least 3/5.

The condition $G \in \mathcal{P}$ implies that there exists a a k-coloring $\mathbf{G}$ of G such that $\mathbf{G} \in \mathcal{Q}$. From the testability of $\mathcal{Q}$ it follows that $t_{\text{inj}}(\hat{\mathcal{Q}}_l, \mathbb{G}(l, \mathbf{G})) \geq 4/5$ for any $l \geq 1$. Let $q \geq 1$ be arbitrary, and let F denote $\mathbb{G}(q, G)$, furthermore let $\mathbf{F} = \mathbb{G}(q, \mathbf{G})$ generated by the same random process as F, so $\mathbf{F}$ is a k-coloring of F. We know by a standard sampling argument that

$$\mathbb{P}(|t_{\text{inj}}(\hat{\mathcal{Q}}_{q_{\mathcal{Q}}(\varepsilon_q)}, \mathbf{G}) - t_{\text{inj}}(\hat{\mathcal{Q}}_{q_{\mathcal{Q}}(\varepsilon_q)}, \mathbf{F})| \geq 1/5) \leq 2 \exp\left(-\frac{q}{50 q^2_{\mathcal{Q}}(\varepsilon_q)} \right), \tag{6.31}$$

and the right hand side of (6.31) is less than 2/5 by the choice of ε_q, since by definition $q \geq 100 q^2_{\mathcal{Q}}(\varepsilon_q)$. It follows that $t_{\text{inj}}(\hat{\mathcal{Q}}_{q_{\mathcal{Q}}(\varepsilon_q)}, \mathbf{F}) \geq 3/5$ with probability at least 3/5, so by the definition of $\hat{\mathcal{P}}$ we have that $F \in \hat{\mathcal{P}}_q$ with probability at least 3/5, which is what we wanted to show.

To verify the second condition we proceed by contradiction. Suppose that G is ε-far from $\mathcal{P}$, but at the same time there exists an $l \geq q_{\mathcal{P}}(\varepsilon)$ such that $F \in \hat{\mathcal{P}}_l$ with probability larger than 2/5, where $F = \mathbb{G}(l, G)$.

In this case, the latter condition implies that with probability larger than 2/5 there exists a k-coloring $\mathbf{F}$ of F such that $t_{\text{inj}}(\hat{\mathcal{Q}}_{q_{\mathcal{Q}}(\varepsilon_l)}, \mathbf{F}) \geq 3/5$. By Lemma 6.5.2 and the proof of Theorem 6.1.6 there exists a k-coloring $\mathbf{G}$ of G such that $d_{\text{tv}}(\mu(q_{\mathcal{Q}}(\varepsilon_l), \mathbf{F}), \mu(q_{\mathcal{Q}}(\varepsilon_l), \mathbf{G})) \leq 22/100$ with probability at least 4/5, in particular

$$|t_{\text{inj}}(\hat{\mathcal{Q}}_{q_{\mathcal{Q}}(\varepsilon_l)}, \mathbf{F}) - t_{\text{inj}}(\hat{\mathcal{Q}}_{q_{\mathcal{Q}}(\varepsilon_l)}, \mathbf{G})| \leq \frac{22}{100},$$

which implies that with probability at least 1/5 there exist a $\mathbf{G}$ such that $t_{\mathrm{inj}}(\hat{Q}_{q_Q(\varepsilon_l)}, \mathbf{G}) > \frac{3}{10}$. We can drop the probabilistic assertion and can say that there exists a k-coloring $\mathbf{G}$ of G such that $t_{\mathrm{inj}}(\hat{Q}_{q_Q(\varepsilon_l)}, \mathbf{G}) > \frac{3}{10}$, because G and the density expression are deterministic.

On the other hand, the fact that G is ε-far from $\mathcal{P}$ implies that any k-coloring $\mathbf{G}$ of G is ε-far from $\mathcal{Q}$, which means that $t_{\mathrm{inj}}(\hat{Q}_q, \mathbf{G}) \le 1/5$ for any k-coloring $\mathbf{G}$ of G and $q \ge q_Q(\varepsilon)$. But we know that $q_Q(\varepsilon_l) \ge q_Q(\varepsilon)$, since $\varepsilon_l \le \varepsilon$ which delivers the contradiction. The last inequality is the consequence of our definitions, ε_l is the infimum of the $\delta > 0$ that satisfy $l \ge q_{\mathcal{P}}(\delta)$, and on the other hand, $l \ge q_{\mathcal{P}}(\varepsilon)$.

□

6.7 Parameters depending on densities of linear hypergraphs

We present a special case of the above general notion of ND-testability for higher rank uniform hypergraphs that preserves several useful properties of the graph case, $r = 2$. Restricting our attention to this sub-class we are able to essentially remove the dependence on r in the bound given by Theorem 6.1.6 on the sample complexity.

A simple linear r-graph is an r-graph that satisfies that any distinct pair of its edges intersect at most in one vertex.

Definition 6.7.1. *We call an r-graph parameter f linearly non-deterministically testable if it is non-deterministically testable and its witness parameter g does only depend on the $\hat{F}$-densities $t^*_{\hat{F}}(F, .)$ of F for simple linear hypergraphs $\hat{F}$ and arbitrary colored r-graphs F with the same vertex set as $\hat{F}$.*

This density notion was formally introduced in (3.34) and (3.35) for $G \in \Pi^{r,k}$, $\hat{F} \in \Pi^r_q$, and $F \in \Pi^{r,k}_q$ as

$$t^*_{\hat{F}}(F, G) = \frac{1}{|V(G)|(|V(G)| - 1) \dots (|V(G)| - q + 1)} \sum_{\phi\colon [q] \to V(G)} \prod_{e \in \hat{F}} \mathbb{1}_{G^{F(e)}}(\phi(e)), \tag{6.32}$$

where the sum runs over injective ϕ maps, and for a naive k-colored r-graphon W as

$$t^*_{\hat{F}}(F, W) = \int_{[0,1]^{[\mathfrak{h}([q],1)]}} \prod_{e \in \hat{F}} W^{F(e)}(x_{\mathfrak{h}(e,1)}) \mathrm{d}\lambda(x_{\mathfrak{h}([q],1)}). \tag{6.33}$$

It generalizes the non-induced t^*-densities of simple r-graphs to the space of colored r-graphs.

In this section we depart from the general r-graphon notion and use instead objects called *naive r-graphons* and *naive r-kernels*, see Section 2.2.3. These differ from true graphons and kernels in their domain that is the r-dimensional unit cube and whose coordinates correspond to nodes of r-edges instead of proper subsets of the node set of an r-edge. They can be transformed into genuine graphons by adding dimension to the domain in a way that the values taken do not depend on the entries corresponding to the new dimensions. This way we can think of naive graphons as a special subclass of graphons, sampling is defined analogously to the general case. Note that for $r = 2$ the naive notion does not introduce any restriction as all proper subsets of a 2-element set are singletons. We require the notion of ground state energies of r-graphs, naive r-graphons, and kernels form Chapter 4, see also [32] and [14].

Let $s \geq 1$, J be an r-array of size s, and G be an arbitrary r-graph. Recall Definition 3.5.6 of the ground state energy (GSE) of the r-graph G with respect to the r-array J that reads as

$$\hat{\mathcal{E}}(G, J) = \max_{\phi} \frac{1}{|V(G)|^r} \sum_{i_1,\dots,i_r=1}^{q} J_{i_1,\dots,i_r} A_G(\phi^{-1}(i_1), \dots, \phi^{-1}(i_r)), \tag{6.34}$$

where the maximum runs over all q-partitions of $|V(G)|$. Analogously, the GSE of a naive r-kernel U with respect to J is

$$\mathcal{E}(U, J) = \max_{f} \sum_{i_1,\dots,i_r=1}^{s} J(i_1, \dots, i_r) \int_{[0,1]^r} \prod_{j=1}^{r} f_{i_j}(x_j) U(x_1, \dots, x_r) \mathrm{d}\lambda(x),$$

where the maximum runs over all fractional partitions f of $[0, 1]$ into s parts.

Recall the definition of the plain cut norm for r-graphs and kernels, Definition 3.3.18 and Definition 3.3.19, for instance, the plain cut norm of a naive r-kernel W is

$$\|W\|_\square = \sup_{S_i \subset [0,1], i \in [r]} \left| \int_{S_1 \times \cdots \times S_r} W(x) \mathrm{d}\lambda(x) \right|,$$

where the supremum is taken over measurable sets $S_i \subset [0, 1]$ for each $i \in [r]$. Recall further Definition 3.3.28 of the plain cut-$\mathcal{P}$-norm for a partition $\mathcal{P}$ of $[0, 1]$, the corresponding regularity lemma, Lemma 3.3.29, and the counting lemma, Lemma 3.3.30. We further require a version of the coloring lemma Lemma 6.2.2, we formulate this tool next, but refrain from giving the detailed proof as it is identical to the proof of Lemma 6.2.2.

Lemma 6.7.2. *Let $k \geq 1$, $\varepsilon > 0$, U be a (naive) step function with steps $\mathcal{P} = (P_1, \dots, P_t)$ and V be a (naive) r-graphon with $d_{\square\mathcal{P}}(U, V) \leq \varepsilon$. For any (naive) k-colored r-graphon $\mathbf{U} = (U^1, \dots, U^k)$ that is a (naive) step function with steps from $\mathcal{P}$ and a k-coloring of U there exists a k-coloring $\mathbf{V} = (V^1, \dots, V^k)$ of V so that $d_\square(\mathbf{U}, \mathbf{V}) = \sum_{\alpha=1}^{k} \|U^{(\alpha)} - V^{(\alpha)}\|_\square \leq k\varepsilon$.*

Next we state and prove the main contribution of this section.

Theorem 6.7.3. *Let $r \geq 2$, and let f be a linearly non-deterministically testable r-graph parameter with witness parameter g of k-colored r-graphs, and let the corresponding sample complexity be q_g. Then f is testable with sample complexity q_f, and there exists a constant $c > 0$ only depending on k and r but not on f or g such that for any $\varepsilon > 0$ we have*

$$q_f(\varepsilon) \leq \exp^{(3)}(c q_g^2(\varepsilon/2)).$$

Proof. The proof is almost identical to the case of 2-graphs in Section 6.2, we will sketch it in the framework of Lemma 6.5.2, from there the statement follows in a similar way as the proof of Theorem 6.1.6. The main distinction between the general hypergraph setting and the current linear setting is that we do not require for each coloring $\mathbf{F}$ of $F = \mathbb{G}(q, G)$ to have a corresponding coloring $\mathbf{G}$ of G such that their q_0-sampled distributions $\mu(q_0, \mathbf{F})$ and $\mu(q_0, \mathbf{G})$ are close in the total variation distance, here it is enough to impose that they are close in $d_\square$ distance. This relaxed condition implies that the densities $t^*_{\hat{H}}(\mathbf{H}, \mathbf{F})$ and $t^*_{\hat{H}}(\mathbf{H}, \mathbf{G})$ for linear hypergraphs $\hat{H}$ are close, and as a consequence the deviation of $g(\mathbf{F})$ and $g(\mathbf{G})$ is sufficiently small. The different norm employed in the measurements of the proximity allows us to remove the inductive part that is contained in the general proof in Lemma 6.5.2.

Let f and g be as in the statement of the theorem, and let G be an arbitrary r-graph and W_G a 3-colored naive r-graphon that represents it (the colors correspond to edges, non-edges, and diagonal entries respectively). Let $q \geq \exp^{(3)}(cq_g^2(\varepsilon/2))$ for some $c > 0$ that is chosen to be large enough, and let F denote the random r-graph $\mathbb{G}(q, G)$, and let W_F be its 3-colored representative naive graphon. It is easy to see as in the general case that $f(F) \geq f(G) - \varepsilon/4$ with probability at least $1 - \varepsilon/4$, in fact this is even true with much smaller q, see the proof of Theorem 6.1.5.

We establish an upper bound on $f(F)$ next. We will show first that with probability at least $1 - \varepsilon/4$ there exist for every k-coloring $\mathbf{V} = (V^{\alpha,\beta})_{\alpha\in[3],\beta\in[k]}$ of W_F a k-coloring $\mathbf{U} = (U^{\alpha,\beta})_{\alpha\in[3],\beta\in[k]}$ of W_G such that $d_{\square\mathcal{Q}}(\mathbf{U}, \mathbf{V}) \leq \Delta$, where $\Delta = \exp(-c' q_g^2(\varepsilon/2))$ for a suitably chosen $c' > 0$. Let $W_1 = (W_G)_{\mathcal{P}}$ be a naive r-graphon that satisfies

$$\sup_{t_\mathcal{Q} \leq t_\mathcal{P} t_2} d_{\square\mathcal{Q}}(W_G, W_1) \leq \Delta/8k, \tag{6.35}$$

by Lemma 3.3.29 there exists such a naive $(r, 1)$-step function with at most $t_1 = t_{\text{reg}}(r, 2, \Delta/8k, t_2)$ steps, where $t_2 = t_{\text{reg}}(r, 3k, \Delta/8k, 1)$. Further, let W_2' be the naive $(r, 1)$-step function associated with $\mathbb{G}(q, W_1)$ with its steps forming the partition $\mathcal{P}''$. There exists a measure-preserving permutation ϕ of $[0, 1]$ such that W_2 given by $W_2(x_1, \ldots, x_r) = W_2'(\phi(x_1), \ldots, \phi(x_r))$ is another valid representation of $\mathbb{G}(q, W_1)$ with steps $\mathcal{P}'$, and having the additional property that the measure of the set where W_1 and W_2 differ is at most $r \sum_i |\lambda(P_i) - \lambda(P_i'')|$. In particular by the choice of q it is true that $\|W_1 - W_2\|_1 \leq r\sqrt{t_1/q} \leq \Delta/8k$ with probability at least $1 - \varepsilon/8$ for a large enough $c' > 0$.

Further, the bound in (6.35) can be rewritten as a GSE problem in the sense of Definition 4.1.2, applying Corollary 4.2.11 leads to the assertion that

$$\sup_{t_\mathcal{Q} \leq t_\mathcal{P} t_2} d_{\square\mathcal{Q}}(W_F, W_2) \leq \Delta/4k, \tag{6.36}$$

with probability at least $1 - \Delta/8k$, which is larger than $1 - \varepsilon/8$.

We condition on the aforementioned two events, they occur jointly with probability at least $1 - \varepsilon/4$. Now let $\mathbf{V}$ be an arbitrary k-coloring of W_F, it follows by Lemma 3.3.29 that there exists a $3k$-colored naive $(r, 1)$-step function $\mathbf{Z} = \mathbf{V}_\mathcal{R} = (Z^{\alpha,\beta})_{\alpha\in[3],\beta\in[k]}$ with steps forming $\mathcal{R}$ such that

$$\sup_{t_Q \le t_{\mathcal{R}}} d_{\square Q}(\mathbf{V}, \mathbf{Z}) \le \Delta/8k, \tag{6.37}$$

and $t_{\mathcal{R}} \le t_2$. Let the naive r-graphon Z denote the k-discoloring of $\mathbf{Z}$. Then we have $Z = (W_F)_{\mathcal{R}}$ and

$$\sup_{t_Q \le t_{\mathcal{R}}} d_{\square Q}(W_F, Z) \le \Delta/8k. \tag{6.38}$$

Together with (6.36) it follows that

$$\sup_{t_Q \le t_{\mathcal{R}}} d_{\square Q}(W_2, Z) \le \Delta/4k. \tag{6.39}$$

An application of Lemma 6.7.2 together with the bound in (6.39) ensures the existence of a k-coloring $\mathbf{W}_2$ of W_2 that is a naive $(r,1)$-step function with the steps comprising $\mathcal{S}$ that is the coarsest common refinement of $\mathcal{P}'$ and $\mathcal{R}$, and that satisfies

$$d_{\square}(\mathbf{W}_2, \mathbf{Z}) \le \Delta. \tag{6.40}$$

Now we construct a k-coloring of W_1 by simply copying $\mathbf{W}_2$ on the set on $[\cup_i (P_i \cap P_i')]^r$, and defining it in arbitrary way on the rest of $[0,1]^r$, paying attention to keep it a k-coloring of W_1 and not increase the number of steps above $t_{\mathcal{R}}$. For the $\mathbf{W}_1$ obtained this way we have

$$d_1(\mathbf{W}_1, \mathbf{W}_2) = \sum_{\alpha,\beta} \|W_1^{\alpha,\beta} - W_2^{\alpha,\beta}\|_1 \le 2k\|W_1 - W_2\|_1 \le \Delta/4. \tag{6.41}$$

Employing again Lemma 6.7.2 with (6.35) we obtain a k-coloring $\mathbf{U}$ of W_G that satisfies

$$d_{\square}(\mathbf{U}, \mathbf{W}_1) \le \Delta,$$

hence

$$d_{\square}(\mathbf{U}, \mathbf{V}) \le 4\Delta.$$

With a further randomization we can form a proper k-coloring $\mathbf{G}$ of G that satisfies

$$d_{\square}(\mathbf{W}_{\mathbf{G}}, \mathbf{V}) \le 5\Delta.$$

Finally, we use

$$|g(\mathbf{F}) - g(\mathbf{G})| \le |g(\mathbf{F}) - g(\mathbb{G}(q_g(\varepsilon/4), \mathbf{F}))| + |g(\mathbf{G}) - g(\mathbb{G}(q_g(\varepsilon/4), \mathbf{G}))| \le \varepsilon/2,$$

whenever there exists a coupling of the random $2k$-colored r-graphs $\mathbb{G}(q_g(\varepsilon/4), \mathbf{G})$ and $\mathbb{G}(q_g(\varepsilon/4), \mathbf{F})$ appearing in the above formula such that their densities $t^*_{\hat{H}}(\mathbf{H}, \mathbf{G})$ and

$t^*_{\hat{H}}(\mathbf{H}, \mathbf{F})$ are pairwise equal for each $\mathbf{H} \in \Pi^{r,2k}_{q_0}$ and each linear $\hat{H} \in \Pi^r_{q_0}$ with probability larger than $\varepsilon/2$, where $q_0 = q_g(\varepsilon/4)$. Such a coupling exists by Lemma 3.3.30 and standard probabilistic assumptions, thus we have $f(G) \geq f(F) - \varepsilon/2$ with probability at least $1 - \varepsilon/4$, that concludes the proof.

□

6.8 Applications

The characterization of testable properties of r-uniform hypergraphs for $r \geq 3$ is a well-studied area, for instance it was established by Rödl and Schacht [104] that hereditary properties (properties that are preserved under the removal of vertices) are testable generalizing the situation formerly known in the graph case. Nevertheless, several analogous question to the graph case have remained open. We present some of these in this section together with the proofs for positive results as an application of Theorem 6.1.6 and Theorem 6.1.7.

6.8.1 Energies and partition problems

Recall the family of r-ground state energies (rGSE) of r-uniform hypergraphs introduced in Chapter 5 in Definition 5.3.1. The notion is a generalization of the ground state energies (GSE) of Borgs, Chayes, Lovász, Sós, and Vesztergombi [32] introduced in the case of graphs (see also Chapter 4 for a different generalization), for connections to statistical physics, in particular to the Ising and the Curie-Weiss model, see [32] and Chapter 3 in the current work. The original GSE notion encompasses several important graph optimization problems, such as the maximal cut density and multiway cut densities for graphs, therefore its testability is central to several applications.

For a simple r-graph $H \subset \binom{[n]}{r}$, a real r-array J of size q, the r-ground state energy of H with respect to J is

$$\mathcal{E}_r(H, J) = \max_{\mathcal{P}} \frac{1}{n^r} \sum_{i_1,\dots,i_r=1}^{q} J(i_1, \dots, i_r) e_H(r; P_{i_1}, \dots, P_{i_r}),$$

where $e_H(r; S_1, \dots, S_r) = |\{ (u_1, \dots, u_r) \in [n]^r \mid A_{S_j}(u_1, \dots, u_{j-1}, u_{j+1}, \dots, u_r) = 1$ for all $j = 1, \dots, r$ and $A_H(u_1, \dots, u_r) = 1 \}|$, and the maximum is taken over all partitions $\mathcal{P}$ of $\binom{[n]}{r-1}$.

We have already proved the rGSE to be testable, see Theorem 5.3.4. As an application of Theorem 6.1.6, we derive a substantially different new proof.

Corollary 6.8.1. *For any $r, q \geq 1$ and real r-array J of size t the generalized ground state energy $\mathcal{E}_r(., J)$ is a testable r-graph parameter.*

The proof in Chapter 5 for the above result used ultralimits and was therefore non-effective. Here, a rather straightforward application of Theorem 6.1.6 gives us Corollary 6.8.1 that does not rely on non-effective tools, we could provide an explicit upper bound on the sample complexity, and in this sense the result is new.

The above problem of testing of the rGSE is a special case of the question regarding testability of general partition problems. These properties were first dealt with systematically in the graph property testing setting in [66], where the authors showed their testability. They form also the most prominent family of non-trivial properties from the testing perspective in the dense model that are testable with polynomial sample complexity known to date, for further background we direct the reader to Chapter 1, in particular to the part on property testing.

A test for example for the maximal cut density can be obtained from a collection of partition problems with two classes only constraining the edge density between the two distinct parts for each integer multiple of ε in $[0, 1]$.

Researched focused on partition problems for hypergraphs was initiated by Fischer, Matsliah, and Shapira [57] defining a framework that slightly extended the notions of [66]. In the setup of [57] the problem is formulated as in [66] as a question of existence of a vertex partition of a hypergraph with prescribed class sizes given by a real vector that satisfies that the r-partite sub-hypergraphs spanned by each r-tuple of classes contain a certain number of edges given by a real array. The additional feature of the approach is that it can also handle tuples of uniform hypergraphs (perhaps of different rank) sharing a common vertex set that is the subject of the partitioning, the partition problem defined again by density tensors comprises constraints on edge densities between classes for each of the component hypergraphs. In [57] it was shown that such properties are testable with polynomial sample complexity.

A further generalization has been investigated by Rozenberg [107] dealing the first time with constraints imposed on partitions of pairs, triplets, and so on of the vertices on one hand, and the edge densities filtered by these partitions on the other. However, the cells that the edge density constraints are applied to in [107] are not partitioning the edge set as in the previous approaches, rather layers of partitions corresponding to partitions of $[r]$ for r-graphs are considered. Let us illustrate the framework for 3-graphs with the partitioning understood as coloring. In [57], only vertices are colored, and the number of edges whose vertices have certain colors are constrained, in [107] also pairs, triplets and further tuples of vertices receive colors, and based on this the number of edges can be constrained that fulfill the condition that a pair of vertices (as a tuple) has a certain color and the third vertex (as a singleton) has also some other given color. However, in [107] only colorings disjoint subsets of the r-edges are allowed to yield a constraint, for instance it is not possible to have a condition on the number of pair-monochromatic edges, that is, 3-edges whose three underlaying pairs have the same color. The positive result obtained in [107] is also somewhat weaker than testability, the term pseudo-testability is introduced in order to formalize the conclusion.

Our approach allows for more general constraints on edge densities, the definition of the general partition problem follows next.

Definition 6.8.2. *Let $r \geq 1$, and Φ denote the set of all maps $\phi\colon \mathfrak{h}([r], r-1) \to [k]$ that are assigning to each element of the set of proper subsets of $[r]$, $\mathfrak{h}([r], r-1)$, a color $[k]$. We define a density tensor by $\psi = \langle \langle \rho_i^s \rangle_{s \in [r-1], i \in [k]}, \langle \mu_\phi \rangle_{\phi \in \Phi} \rangle$, where each component is in $[0, 1]$.*

Let H be an r-graph with vertex set $V = V(H)$ of cardinality n and for each $1 \le s \le r-1$ let $\mathcal{P}(s)$ be a partition of $\binom{V}{s}$ into k parts, and let $\mathcal{P} = (\mathcal{P}(s))_{s=1}^{r-1}$. Then the density tensor corresponding to the pair $(H, \mathcal{P})$ is given by

$$\rho_i^s(H, \mathcal{P}) = \frac{|P_i(s)|}{n^s} \quad \textit{for all} \quad s \in [r-1] \textit{ and } i \in [k],$$

and

$$\mu_\phi(H, \mathcal{P}) = \frac{|\{\, e \in [n]^r \mid \overline{e} \in H \textit{ and } \overline{p_A(e)} \in P_{\phi(A)}(|A|) \textit{ for all } A \in \mathfrak{h}([r], r-1) \,\}|}{|\{\, e \in [n]^r \mid \overline{p_A(e)} \in P_{\phi(A)}(|A|) \textit{ for all } A \in \mathfrak{h}([r], r-1) \,\}|}$$

for all $\phi \in \Phi$, where $\overline{v}$ is the set that consists of the components of the vector v.

We say that H satisfies the partition problem given by a density tensor ψ if there exists a collection of partitions $\mathcal{P}$ of its vertex tuples as above so that the tensor yielded by the pair $(H, \mathcal{P})$ is equal to ψ.

We remark that the above partition problem property is non-hereditary. An application of Theorem 6.1.7 yields the following corollary.

Corollary 6.8.3. *For any $r, k \ge 1$ and any density tensor $\psi = \langle \langle \rho_i^s \rangle_{s \in [r-1], i \in [k]}, \langle \mu_\phi \rangle_{\phi \in \Phi} \rangle$, the partition property given by the tensor is testable.*

6.8.2 Logical formulas

The characterization of testability in terms of logical formulas was initiated by Alon, Fischer, Krivelevich, and Szegedy [13] who showed that properties expressible by certain first order formulas are testable, while there exist some first order formulas that generate non-testable properties. The result can be formulated as follows.

Theorem 6.8.4. *[13] Let $l, k \ge 1$ and ϕ be a quantifier-free first order formula of arity $l + k$ containing only adjacency and equality. The graph property given by the truth assignments of the formula $\exists u_1, \dots, u_l \forall v_1, \dots, v_k \phi(u_1, \dots, u_l, v_1, \dots, v_k)$ with the variables being vertices is testable.*

Without going into further details at the moment we mention that any $\exists\forall$ property of graphs is indistinguishable by a tester from the existence of a node-coloring that is proper in the sense that the colored graph does not contain subgraphs of a certain set of forbidden node-colored graphs, see [13].

Our focus is directed at the positive results of [13], those were generalized into two directions. First, by Jordan and Zeugmann [77] to the framework of relational structures in the sense that ϕ is allowed to contain several r-ary relations with even $r \ge 3$ whereas the $\exists\forall$ prefix remains the same concerning vertices. Secondly, by Lovász and Vesztergombi [97] staying in the graph property testing setting to a restricted class of second order formulas, where existential quantifiers for 2-ary relationships are added ahead of the above formula in Theorem 6.8.4 so that they can be included in

ϕ, see Corollary 4.1 in [97]. This subclass of second order logic is also referred to as monadic second order logic (MSOL), see [89]. Our framework allows for extending these results even further.

Corollary 6.8.5. *Let $r_1, \dots, r_m, l, k \geq 1$ be arbitrary, and let $r = \max r_i$. Any r-graph property that is expressible by the truth assignments of the second order formula*

$$\exists L_1, \dots, L_m \exists u_1, \dots, u_l \forall v_1, \dots, v_k \phi(L_1, \dots, L_m, u_1, \dots, u_l, v_1, \dots, v_k) \tag{6.42}$$

is testable, where L_i are symmetric r_i-ary predicate symbols and $u_1, \dots, u_l, v_1, \dots, v_k$ are nodes, and ϕ is a quantifier-free first order expression containing adjacency, equality, and the symmetric r_i-ary predicates L_i for each $i \in [m]$.

Proof (Sketch). We first note that any collection of the relations $L_1, \dots, L_m$ can be encoded into one edge-colored r-uniform hypergraph with at most 2^{rm} colors with an additional compatibility requirement. An edge color for $e \in \binom{[n]}{r}$ consists of a $2^r - 1$-tuple corresponding to non-empty subsets of $[r]$, where the entry corresponding to $S \subset [r]$ in the tuple is determined by the evaluation of $p_S(e)$ in the relations L_i that have arity $|S|$. We can reconstruct the predicates from a coloring whenever the color of any pair of edges e and e' is such that their entries in the tuple corresponding to the power set of $e \cap e'$ coincide, for $r = 2$ this means some combinations of colors (determined by a partition of the colors) for incident edges are forbidden. This compatibility criteria for 2^{rm}-colored r-graphs is known to be a testable property, from here on this will be seen as a default condition.

For a fixed tuple $L_1, \dots, L_m$ of relations of arity at most r the property corresponding to the first order expression $\forall v_1, \dots, v_k \phi(L_1, \dots, L_m, v_1, \dots, v_k)$ is equivalent to the property of 2^{rm}-colored r-graphs that is defined by forbidding certain subgraphs of size at most k. This is testable by the following theorem of Austin and Tao [24] that generalizes the result of Rödl and Schacht [104].

Theorem 6.8.6. *[24] For any $r, k \geq 1$, every hereditary property of k-colored r-graphs is testable.*

We sketch now the proof that the properties corresponding to the more general formula (6.42) in the statement of the corollary are indistinguishable from the existence of a further node-coloring on top of the edge-colored graphs such that no subgraph appears from a certain set of forbidden subgraphs. We follow the argument of [13] (see also [77], and [97]).

Two properties are said to be indistinguishable in this sense whenever for every $\varepsilon > 0$ there exists an $n_0 = n_0(\varepsilon)$ such that any graph on $n \geq n_0$ vertices that has one property can be modified by at most εn^r edge additions or removals to obtain a graph that has the other property, and vice versa. The testability behavior of the two properties is identical.

Consider $L_1, \dots, L_m$ as fixed, then the property of 2^{rm}-colored r-graphs corresponding to $\exists u_1, \dots, u_l \forall v_1, \dots, v_k \phi(L_1, \dots, L_m, u_1, \dots, u_l, v_1, \dots, v_k)$ is indistinguishable to from the existence of the following proper coloring. Every node gets either color $(0, 0)$ or

(a, b), where a represents a 2^{rm}-colored r-graph on l nodes, and b represents an l-tuple of 2^{rm}-colored edges. A coloring is proper if there are at most l nodes colored by $(0, 0)$, further for any other color appearing, the first component a is identical. Now a colored subgraph of size k is forbidden if considering the edge-colored graph on $V = \{v_1, \dots, v_k\}$ (without node colors) supplemented by a graph on $\{u_1, \dots, u_l\}$ together with their connection to V given by the node colors on V the evaluation of the formula $\phi(L_1, \dots, L_m, u_1, \dots, u_l, v_1, \dots, v_k)$ is false.

It is not hard to see that for this coloring property Theorem 6.8.6 applies since it is hereditary, therefore it is testable. Now if we let $L_1, \dots, L_m$ to be arbitrary and apply Theorem 6.1.6, then we obtain the testability of the property given by (6.42) in the statement of the corollary. □

6.8.3 Estimation of the distance to properties and tolerant testers

We can also express the property of being close to a given property in the nondeterministic framework, and can show the testability here. This problem was introduced first for graphs by Fischer and Newman [56], in this paper the authors show the equivalence of testability and estimability of the distance of a property, in [97] one direction of this was reproved for graphs. To our knowledge the generalization for r-graphs with $r \geq 3$ has not been considered yet. Recall that d_1 is the normalized edit distance.

Corollary 6.8.7. *For any $r \geq 2$, testable r-graph property $\mathcal{P}$, and real $c \geq 0$, the property determined by $d_1(., \mathcal{P}) < c$ is testable.*

Proof. The proof is identical to the one given in [97]: for any $r \geq 2$, testable r-graph property $\mathcal{P}$ and real $c > 0$ there exists a testable property of 4-colored r-graphs that witnesses the property of $d_1(., \mathcal{P}) < c$ in the nondeterministic testing sense. Let G be an arbitrary r-graph, then we consider the 2-colorings of G where $(1, 1)$ and $(1, 2)$ color the edges of G, and $(2, 1)$ and $(2, 2)$ the non-edges. The 4-colored witness property $\mathcal{Q}$ is then that the edges with the colors $(1, 1)$ and $(2, 1)$ together form a member of $\mathcal{P}$, and additionally there are at most cn^r edges colored by $(1, 2)$ or $(2, 1)$. The property $\mathcal{Q}$ is trivially testable, therefore Theorem 6.1.7 implies the statement. □

The above corollary is a generalization of the concept of tolerant testers. In this case, we expect a good test with parameters $0 \leq \delta < \varepsilon$ to distinguish between the cases that a graph is δ-close (another way of saying not δ-far) to the property or that it is ε-far. Here the sample size should only depend on the two thresholds. We see that taking $\delta = 0$ is the same as the usual testing. On the other hand for a fixed δ_0, the testing of the property $d_1(., \mathcal{P}) \leq \delta_0$ with an $\varepsilon - \delta_0$ margin of error is exactly the task described above. We can conclude that every testable hypergraph property is also tolerantly testable.

On a further note we mention that with the same method we can also show that the r-graph parameter $f(G) = d_1(G, \mathcal{P})$ is also testable for any testable $\mathcal{P}$.

CHAPTER 7

Conclusion and further research

In this thesis we proved the equivalence of nondeterministic testing to usual testing for r-uniform hypergraphs for arbitrary rank r for both properties and parameters, we further gave upper bounds on the sample complexity that improved previously known bounds related to this approach of characterizing testability, see [97], [61].

A major open question is whether it is possible to prove that the testing of the witness and the testing of the original property are computationally equivalent, in the sense that their sample complexities are in a polynomial relationship. In this sense, improvement for upper bounds in magnitude would be also a reasonable contribution. Also, lower bounds in this regard would be welcome, currently we are only aware of the trivial one. More specifically, one might try to improve upper bounds for weakly nondeterministically testable parameters and properties, an implicit appearance of this setting were partition problems in [66] and multiway cuts testing in [14]. In both cases the witness is the easiest non-trivial property that imposes conditions on the number of the edge counts of different colors, and the desired polynomial dependence was established via problem-specific arguments. As a further particular question, on may ask whether it is possible to gain something from considering only witnesses that are testable with a sample size that is polynomial in the multiplicative inverse of the error parameter.

The general upper bound given in Theorem 6.1.6 is dependent on the rank r, it would be interesting to see if it is possible to remove this dependence in a similar way as it was shown in the special case of linearly nondeterministically testable parameters. The gap between the two bounds here is still wide open. Currently no non-trivial lower bound on the sample complexity for general r in our framework is known, in the original dense property testing setting there are some properties that admit no tester that only makes a polynomial number of queries, such as triangle-freeness and other properties defined by forbidden families of subgraphs or induced subgraphs.

The partition problems introduced in [66] have lead to further applications, this development was presented in [57]. As mentioned, the framework of [57] also dealt with tuples of hypergraphs extending the result of [66], this enabled the analysis of the number of 4-cycles appearing in the bipartite graphs induced by the pairs of the partition classes instead of only observing the edge density by means of adding an auxiliary 4-graph to the simple graph. An alternative characterization of the notion of a regular bipartite graph says that a pair of classes is regular if and only if the number of 4-cycles spanned by them is minimal, with other words their density is approximately the fourth power of the edge density. Using this together with the result regarding the testability of partition problems the authors of [57] were able to show that satisfying a certain regularity instance is also testable. This achievement in turn implies an algorithmic version of the Regularity Lemma. In this manner, Corollary 6.8.3 might be of further use for testing regular partitions of r-uniform hypergraphs by utilizing concepts that emerged during the course of research towards an algorithmic version of the Hypergraph Regularity Lemma (see for example Haxell, Nagle, and Rödl [71]) in a similar way to the approach in [57].

On a further thought, one may depart from the setting of general dense r-graphs in favor of other classes of combinatorial objects in order to define and study their ND-testability. Such classes are for example *semi-algebraic hypergraphs* that admit a regularity lemma that produces a polynomial number of classes as a function of the multiplicative inverse of the proximity parameter, thus they are good candidates for an improvement in the upper bounds of the sample complexity. Other choices could be graph families with bounded Minkowski-dimension with regard to a similarity metric defined between the vertices of the graph in question, see [89].

Additionally we mention a possible direction for further study towards the characterization of *locally repairable* properties, see [24], that appears to be promising. This characteristic is stronger than testability in that respect that in this setup there should exist a local edge modifying algorithm applied to graphs G that are close to a given property that observes only some piece $A \subset V(G)$ of bounded size of the graph and its connection to individual vertex pairs uv and decides upon the adjacency of u and v depending only on this information, i.e., the induced subgraph on $A \cup \{u, v\}$ (we only gave here an impression of the weak version, for details see [24]). The output of this algorithm should be a graph that is close to the input and actually satisfies the property. We may define nondeterministically locally repairable properties in a straight-forward way analogous to ND-testing by requiring a certain locally repairable property of edge-colored graphs that reduces to the original property after the discoloring procedure. It has been established in [24] that hereditary graph properties are locally repairable, but there are examples of hereditary properties of directed graphs and 3-graphs that are testable, but not locally repairable. It would be compelling to investigate analogous problems concerning nondeterministically locally repairable properties.

As a final remark on nondeterministic property testing we mention a generalization proposed by Fischer [54]. Similar to our setup, he dealt with colorings, but from one aspect in a less general sense by allowing only witness properties defined by a finite collection of forbidden colored subgraphs. These problems are called $\mathcal{F}$-colorability and $\mathcal{F}$-pair-colorability for node and edge colorings, respectively, in both cases positive testability results were obtained. The author of [54] introduced $\mathcal{F}$-colorability with restrictions, the so-called $(\alpha, \mathcal{F})$-colorability, here we are looking for a coloring that avoids certain subgraphs, but additionally the sizes of the color classes have to obey upper and lower bounds given by α. Also for this setup, testability was shown, the situation in the pair coloring version is an open problem. It is straight-forward to generalize the notion of nondeterministic testing to *nondeterministic testing with restrictions* on the sizes of the color classes. This is truly a generalization, as it is not ensured that the intersection of a witness property and a set of restrictions results in a testable property, and certainly not in a hereditary one, that could take the place of the witness of the new property. It would be interesting to see whether our methods can handle questions in this setting.

As a further main result of the thesis, we proved a characterization of the limit space of sequences of r-uniform hypergraphs, whose edges are colored with elements taken from compact space, especially from the interval $[0,1]$. It turned out that the limit objects are measurable functions on $[0,1]^{2^r-2}$ (with symmetries dependent on the properties of the discrete objects) taking values from the probability distributions on the compact color set. This outcome is in accordance with the case of the simple ($\{0,1\}$-colored) graphs [91, 43], and r-uniform hypergraphs [49], where the range is a subset of $[0,1]$, and the values correspond to Bernoulli-measures. Previous results in this direction include a similar characterization for 2-graphs with compact edge colors [93].

One major open problem is in this area whether there exists a a metric carrying a similar structural statement for r-uniform hypergraphs as the $\delta_\square$-metric for simple graphs, and at the same time explaining the compactness of the limit space. On a different note, it would be interesting to see if the characterization of testability via limits outlined in [94] can be generalized for the hypergraph setting.

We showed the testability of layered ground state energies in the limit space with a sample complexity upper bound $\Theta^4 \log(\Theta) q^r$, where $\Theta = \frac{2^{r+7} q^r r}{\varepsilon}$. This notion of GSE encompasses multiway cuts of graphs as well as any family of MAX-rCSPs, as a corollary we get the discrete version rather easily. The result itself was implicitly known for $r = 2$ as a combination of developments in [32] and [98].

This kind of proof method is novel in the sense that it directly relates the optimum of the continuous optimization problem which is a graphon parameter to the optimum of the corresponding problem on the sample without inserting an intermediate step and then using the discrete analog of our result, this concept was proposed by Borgs, Chayes, Lovász, Sós, and Vesztergombi [32]. We roughly follow the approach of [14], our reasoning is self-contained, and perhaps more transparent (several error terms vanish in the limit) than [14], we get a better bound on the failure probability that decreases with the desired additive error instead of being constant, and extend the method to q-state GSE that corresponds to non-Boolean MAX-rCSP. Our hope is that a refinement of our approach will be suitable to determine the exact sample complexity of MAX-rCSP, the gap between the currently best lower, $\Omega(\varepsilon^{-2})$ and upper bound $O(\varepsilon^{-4})$ is still of considerable size.

Our method could only handle the case of GSE, where the number of states was finite. It is an intriguing question what happens for a continuous state space. It seems that if the state space has finite Minkowski-dimension (as in the case of the d-dimensional unit sphere for fixed d), then our proof follows without a large amount of adjustment from the finite state case. It is not clear what happens in the case if we deal with a GSE that corresponds to the semidefinite relaxation of the MAX-CUT density of some simple graph. Are such parameters even testable, and if yes, then how can we relate their sample complexity to that of the unrelaxed versions? The question is even more compelling as we know that semidefinite programs can be solved in polynomial time, in contrast, MAX-CUT is NP-hard.

Analogously to the generalized ground state energies, which were derived from partitions of vertex tuples, one may similarly study free energies, since a Gibbs-like measure can be also defined in this setup. We are not aware of any prior work on these structures in statistical physics, the closest connection is perhaps to monomer-dimer problems. Testability of these r-graph parameters is an open problem, even at the first step towards understanding the problem difficulties not present in the basic case emerge: The free energy of a blow-up is much harder to analyze, as the bipartite graphs between the classes corresponding to vertices come into play, therefore the enumeration problem is not nearly as trivial as in the basic node coloring case.

From a more general perspective, the meta-problem is to characterize more precisely the class of problems which are efficiently parameter testable as opposed to the hard ones. Improving the bounds in the multiplicative inverse of the permitted error for the efficiently testable problems is also a worthwhile question.

Bibliography

[1] Nir Ailon and Noga Alon. Hardness of fully dense problems. *Inform. and Comput.*, 205(8):1117–1129, 2007.

[2] David J. Aldous. Representations for partially exchangeable arrays of random variables. *J. Multivariate Anal.*, 11(4):581–598, 1981.

[3] Noga Alon. Testing subgraphs in large graphs. *Random Structures Algorithms*, 21 (3-4):359–370, 2002. Random structures and algorithms (Poznan, 2001).

[4] Noga Alon and Jacob Fox. Easily testable graph properties. *Combin. Probab. Comput.*, 24(4):646–657, 2015.

[5] Noga Alon and Michael Krivelevich. Testing k-colorability. *SIAM J. Discrete Math.*, 15(2):211–227 (electronic), 2002.

[6] Noga Alon and Assaf Naor. Approximating the cut-norm via Grothendieck's inequality. *SIAM J. Comput.*, 35(4):787–803 (electronic), 2006.

[7] Noga Alon and Asaf Shapira. Testing satisfiability. *J. Algorithms*, 47(2):87–103, 2003.

[8] Noga Alon and Asaf Shapira. Testing subgraphs in directed graphs. *J. Comput. System Sci.*, 69(3):353–382, 2004.

[9] Noga Alon and Asaf Shapira. A characterization of easily testable induced subgraphs. *Combin. Probab. Comput.*, 15(6):791–805, 2006.

[10] Noga Alon and Asaf Shapira. A characterization of the (natural) graph properties testable with one-sided error. *SIAM J. Comput.*, 37(6):1703–1727, 2008.

[11] Noga Alon and Asaf Shapira. Every monotone graph property is testable. *SIAM J. Comput.*, 38(2):505–522, 2008.

[12] Noga Alon and Joel H. Spencer. *The probabilistic method*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc., Hoboken, NJ, third edition, 2008.

[13] Noga Alon, Eldar Fischer, Michael Krivelevich, and Mario Szegedy. Efficient testing of large graphs. *Combinatorica*, 20(4):451–476, 2000.

[14] Noga Alon, Wenceslas Fernandez de la Vega, Ravi Kannan, and Marek Karpinski. Random sampling and approximation of MAX-CSP problems. In *Proceedings of the Thirty-Fourth Annual ACM Symposium on Theory of Computing*, pages 232–239, 2002. Also appeared in J. Comput. System Sci., 67(2):212–243,2003.

[15] Noga Alon, Konstantin Makarychev, Yury Makarychev, and Assaf Naor. Quadratic forms on graphs. *Invent. Math.*, 163(3):499–522, 2006.

[16] Noga Alon, Eldar Fischer, Ilan Newman, and Asaf Shapira. A combinatorial characterization of the testable graph properties: it's all about regularity. *SIAM J. Comput.*, 39(1):143–167, 2009.

[17] Gunnar Andersson and Lars Engebretsen. Sampling methods applied to dense instances of non-Boolean optimization problems. In *Randomization and approximation techniques in computer science (Barcelona, 1998)*, volume 1518 of *Lecture Notes in Comput. Sci.*, pages 357–368. Springer, Berlin, 1998.

[18] Sanjeev Arora, David R. Karger, and Marek Karpinski. Polynomial time approximation schemes for dense instances of *NP*-hard problems. In *Proceedings of the Twenty-Seventh Annual ACM Symposium on Theory of Computing*, pages 284–293, 1995. Also appeared in J. Comput. System Sci., 58(1):193–210, 1999.

[19] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *J. ACM*, 45(3):501–555, 1998.

[20] Sanjeev Arora, Alan M. Frieze, and Haim Kaplan. A new rounding procedure for the assignment problem with applications to dense graph arrangement problems. *Math. Program.*, 92(1, Ser. A):1–36, 2002.

[21] Ashwini Aroskar. *Limits, Regularity and Removal for Relational and Weighted Structures*. dissertation, CMU, 2012.

[22] Vikraman Arvind, Johannes Köbler, Sebastian Kuhnert, and Yadu Vasudev. Approximate graph isomorphism. In *Mathematical foundations of computer science 2012*, volume 7464 of *Lecture Notes in Comput. Sci.*, pages 100–111. Springer, Heidelberg, 2012.

[23] Tim Austin. On exchangeable random variables and the statistics of large graphs and hypergraphs. *Probab. Surv.*, 5:80–145, 2008.

[24] Tim Austin and Terence Tao. Testability and repair of hereditary hypergraph properties. *Random Structures Algorithms*, 36(4):373–463, 2010.

[25] Boaz Barak, Moritz Hardt, Thomas Holenstein, and David Steurer. Subsampling mathematical relaxations and average-case complexity. In *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 512–531. SIAM, Philadelphia, PA, 2011.

[26] Itai Benjamini and Oded Schramm. Recurrence of distributional limits of finite planar graphs. *Electron. J. Probab.*, 6:no. 23, 13 pp. (electronic), 2001.

[27] Marianna Bolla. *Spectral clustering and biclustering*. John Wiley & Sons, Ltd., Chichester, 2013. Learning large graphs and contingency tables.

[28] Marianna Bolla, Tamás Kói, and András Krámli. Testability of minimum balanced multiway cut densities. *Discrete Appl. Math.*, 160(7-8):1019–1027, 2012.

[29] Béla Bollobás and Oliver Riordan. Metrics for sparse graphs. In *Surveys in combinatorics 2009*, volume 365 of *London Math. Soc. Lecture Note Ser.*, pages 211–287. Cambridge Univ. Press, Cambridge, 2009.

[30] Christian Borgs, Jennifer Chayes, László Lovász, Vera T. Sós, and Katalin Vesztergombi. Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing. *Adv. Math.*, 219(6):1801–1851, 2008.

[31] Christian Borgs, Jennifer Chayes, and László Lovász. Moments of two-variable functions and the uniqueness of graph limits. *Geom. Funct. Anal.*, 19(6):1597–1619, 2010.

[32] Christian Borgs, Jennifer Chayes, László Lovász, Vera T. Sós, and Katalin Vesztergombi. Convergent sequences of dense graphs II. Multiway cuts and statistical physics. *Ann. of Math. (2)*, 176(1):151–219, 2012.

[33] Christian Borgs, Jennifer T. Chayes, Henry Cohn, and Yufei Zhao. An l^p theory of sparse graph convergence i: limits, sparse random graph models, and power law distributions, 2014. preprint, arXiv:1401.2906.

[34] Christian Borgs, Jennifer T. Chayes, Henry Cohn, and Yufei Zhao. An l^p theory of sparse graph convergence ii: Ld convergence, quotients, and right convergence, 2014. preprint,arXiv:1408.0744.

[35] Moses Charikar and Anthony Wirth. Maximizing quadratic programs: Extending grothendieck's inequality. In *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '04, pages 54–60, Washington, DC, USA, 2004. IEEE Computer Society.

[36] Sourav Chatterjee and Persi Diaconis. Estimating and understanding exponential random graph models. *Ann. Statist.*, 41(5):2428–2461, 2013.

[37] Sourav Chatterjee and S. R. S. Varadhan. The large deviation principle for the erdős-rényi random graph. *Eur. J. Comb.*, 32(7):1000–1017, 2011.

[38] David Conlon and Jacob Fox. Bounds for graph regularity and removal lemmas. *Geom. Funct. Anal.*, 22(5):1191–1256, 2012.

[39] David Conlon and Jacob Fox. Graph removal lemmas. In *Surveys in combinatorics 2013*, volume 409 of *London Math. Soc. Lecture Note Ser.*, pages 1–49. Cambridge Univ. Press, Cambridge, 2013.

[40] Artur Czumaj and Christian Sohler. Testing hypergraph colorability. *Theoret. Comput. Sci.*, 331(1):37–52, 2005.

[41] Bruno de Finetti. *Funzione Caratteristica Di un Fenomeno Aleatorio*, pages 251–299. 6. Memorie. Academia Nazionale del Linceo, 1931.

[42] Pierre de la Harpe and Vaughan F.R. Jones. Graph invariants related to statistical mechanical models: examples and problems. *J. Combin. Theory Ser. B*, 57(2):207–227, 1993.

[43] Persi Diaconis and Svante Janson. Graph limits and exchangeable random graphs. *Rend. Mat. Appl. (7)*, 28(1):33–61, 2008.

[44] Persi Diaconis, Susan Holmes, and Svante Janson. Threshold graph limits and random threshold graphs. *Internet Math.*, 5(3):267–320 (2009), 2008.

[45] Persi Diaconis, Susan Holmes, and Svante Janson. Interval graph limits. *Ann. Comb.*, 17(1):27–52, 2013.

[46] Petros Drineas, Ravi Kannan, and Michael W. Mahoney. Sampling subproblems of heterogeneous Max-Cut problems and approximation algorithms. *Random Structures Algorithms*, 32(3):307–333, 2008.

[47] Gábor Elek. On limits of finite graphs. *Combinatorica*, 27(4):503–507, 2007.

[48] Gábor Elek. Samplings and observables. invariants of metric measure spaces, 2012. preprint,arXiv:1205.6936.

[49] Gábor Elek and Balázs Szegedy. A measure-theoretic approach to the theory of dense hypergraphs. *Adv. Math.*, 231(3-4):1731–1772, 2012.

[50] Wenceslas Fernandez de la Vega and Marek Karpinski. Polynomial time approximation of dense weighted instances of MAX-CUT. *Random Structures Algorithms*, 16(4):314–332, 2000.

[51] Wenceslas Fernandez de la Vega and Marek Karpinski. A polynomial time approximation scheme for subdense MAX-CUT. *Electronic Colloquium on Computational Complexity (ECCC)*, (044), 2002.

[52] Wenceslas Fernandez de la Vega, Ravi Kannan, Marek Karpinski, and Santosh Vempala. Tensor decomposition and approximation schemes for constraint satisfaction problems. In *STOC'05: Proceedings of the 37th Annual ACM Symposium on Theory of Computing*, pages 747–754. ACM, New York, 2005.

[53] Wenceslas Fernandez de la Vega, Ravi Kannan, and Marek Karpinski. Approximation of global max-csp problems. *Electronic Colloquium on Computational Complexity (ECCC)*, 2006. Technical Report TR06-124.

[54] Eldar Fischer. Testing graphs for colorability properties. *Random Structures Algorithms*, 26(3):289–309, 2005.

[55] Eldar Fischer and Arie Matsliah. Testing graph isomorphism. *SIAM J. Comput.*, 38(1):207–225, 2008.

[56] Eldar Fischer and Ilan Newman. Testing versus estimation of graph properties. *SIAM J. Comput.*, 37(2):482–501 (electronic), 2007.

[57] Eldar Fischer, Arie Matsliah, and Asaf Shapira. Approximate hypergraph partitioning and applications. *SIAM J. Comput.*, 39(7):3155–3185, 2010.

[58] Peter Frankl and Vojtěch Rödl. Extremal problems on set systems. *Random Structures Algorithms*, 20(2):131–164, 2002.

[59] Alan M. Frieze and Ravi Kannan. Quick approximation to matrices and applications. *Combinatorica*, 19(2):175–220, 1999.

[60] Stefanie Gerke and Angelika Steger. The sparse regularity lemma and its applications. In *Surveys in combinatorics 2005*, volume 327 of *London Math. Soc. Lecture Note Ser.*, pages 227–258. Cambridge Univ. Press, Cambridge, 2005.

[61] Lior Gishboliner and Asaf Shapira. Deterministic vs non-deterministic graph property testing. *Israel J. Math.*, 204(1):397–416, 2014.

[62] Michel X. Goemans and David P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. Assoc. Comput. Mach.*, 42(6):1115–1145, 1995.

[63] Oded Goldreich. Introduction to testing graph properties. In *Studies in complexity and cryptography*, volume 6650 of *Lecture Notes in Comput. Sci.*, pages 470–506. Springer, Heidelberg, 2011.

[64] Oded Goldreich and Dana Ron. Property testing in bounded degree graphs. *Algorithmica*, 32(2):302–343, 2002.

[65] Oded Goldreich and Luca Trevisan. Three theorems regarding testing graph properties. *Random Structures Algorithms*, 23(1):23–57, 2003.

[66] Oded Goldreich, Shafi Goldwasser, and Dana Ron. Property testing and its connection to learning and approximation. *J. ACM*, 45(4):653–750, 1998.

[67] W. Timothy Gowers. Lower bounds of tower type for Szemerédi's uniformity lemma. *Geom. Funct. Anal.*, 7(2):322–337, 1997.

[68] W. Timothy Gowers. Quasirandomness, counting and regularity for 3-uniform hypergraphs. *Combin. Probab. Comput.*, 15(1-2):143–184, 2006.

[69] W. Timothy Gowers. Hypergraph regularity and the multidimensional Szemerédi theorem. *Ann. of Math. (2)*, 166(3):897–946, 2007.

[70] Martin Grötschel, László Lovász, and Alexander Schrijver. *Geometric algorithms and combinatorial optimization*, volume 2 of *Algorithms and Combinatorics*. Springer-Verlag, Berlin, second edition, 1993.

[71] Penny E. Haxell, Brendan Nagle, and Vojtěch Rödl. An algorithmic version of the hypergraph regularity method. *SIAM J. Comput.*, 37(6):1728–1776, 2008.

[72] Edwin Hewitt and Leonard J. Savage. Symmetric measures on Cartesian products. *Trans. Amer. Math. Soc.*, 80:470–501, 1955.

[73] Jan Hladky, Andras Mathe, Viresh Patel, and Oleg Pikhurko. Poset limits can be totally ordered, 2012. preprint,arXiv:1211.2473.

[74] D. N. Hoover. Relations on probability spaces and arrays of random variables (preprint), 1979.

[75] Carlos Hoppen, Yoshiharu Kohayakawa, Carlos Gustavo Moreira, Balázs Ráth, and Rudini Menezes Sampaio. Limits of permutation sequences. *J. Combin. Theory Ser. B*, 103(1):93–113, 2013.

[76] Svante Janson. Poset limits and exchangeable random posets. *Combinatorica*, 31 (5):529–563, 2011.

[77] Charles Jordan and Thomas Zeugmann. Testable and untestable classes of first-order formulae. *J. Comput. System Sci.*, 78(5):1557–1578, 2012.

[78] Olav Kallenberg. Symmetries on random arrays and set-indexed processes. *J. Theoret. Probab.*, 5(4):727–765, 1992.

[79] Marek Karpinski and Roland Markó. Limits of CSP problems and efficient parameter testing, 2014. preprint, arXiv:1406.3514.

[80] Marek Karpinski and Roland Markó. Complexity of nondeterministic graph parameter testing, 2014. preprint, arXiv:1408.3590.

[81] Marek Karpinski and Roland Markó. On the complexity of nondeterministically testable hypergraph parameters, 2015. preprint, arXiv:1503.07093.

[82] Marek Karpinski and Roland Markó. Explicit bounds for nondeterministically testable hypergraph parameters, 2015. preprint, arXiv:1509.03046.

[83] Marek Karpinski and Warren Schudy. Sublinear time construction schemes for dense MAX-CSP problems and their extensions (unpublished), 2011.

[84] Subhash Khot and Assaf Naor. Grothendieck-type inequalities in combinatorial optimization. *Comm. Pure Appl. Math.*, 65(7):992–1035, 2012.

[85] János Komlós, Ali Shokoufandeh, Miklós Simonovits, and Endre Szemerédi. The regularity lemma and its applications in graph theory. In *Theoretical aspects of computer science (Tehran, 2000)*, volume 2292 of *Lecture Notes in Comput. Sci.*, pages 84–112. Springer, Berlin, 2002.

[86] András Krámli and Roland Markó. Lower threshold ground state energy and testability of minimal balanced cut density. *Ann. Univ. Sci. Budapest. Sect. Comput.*, 42:231–247, 2014.

[87] Dávid Kunszenti-Kovács, László Lovász, and Balázs Szegedy. Multigraph limits, unbounded kernels,and Banach space decorated graphs, 2014. preprint, arXiv:1406.7846.

[88] Michael Langberg, Yuval Rabani, and Chaitanya Swamy. Approximation algorithms for graph homomorphism problems. In *Approximation, randomization and combinatorial optimization*, volume 4110 of *Lecture Notes in Comput. Sci.*, pages 176–187. Springer, Berlin, 2006.

[89] László Lovász. *Large networks and graph limits*, volume 60 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2012.

[90] László Lovász and Vera T. Sós. Generalized quasirandom graphs. *J. Combin. Theory Ser. B*, 98(1):146–163, 2008.

[91] László Lovász and Balázs Szegedy. Limits of dense graph sequences. *J. Combin. Theory Ser. B*, 96(6):933–957, 2006.

[92] László Lovász and Balázs Szegedy. Szemerédi's lemma for the analyst. *Geom. Funct. Anal.*, 17(1):252–270, 2007.

[93] László Lovász and Balázs Szegedy. Limits of compact decorated graphs, 2010. preprint, arXiv:1010.5155.

[94] László Lovász and Balázs Szegedy. Testing properties of graphs and functions. *Israel J. Math.*, 178:113–156, 2010.

[95] László Lovász and Balázs Szegedy. Regularity partitions and the topology of graphons. In *An irregular mind*, volume 21 of *Bolyai Soc. Math. Stud.*, pages 415–446. János Bolyai Math. Soc., Budapest, 2010.

[96] László Lovász and Balázs Szegedy. Finitely forcible graphons. *J. Combin. Theory Ser. B*, 101(5):269–301, 2011.

[97] László Lovász and Katalin Vesztergombi. Non-deterministic graph property testing. *Combin. Probab. Comput.*, 22(5):749–762, 2013.

[98] Claire Mathieu and Warren Schudy. Yet another algorithm for dense max cut: go greedy. In *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 176–182. ACM, New York, 2008.

[99] Brendan Nagle, Vojtěch Rödl, and Mathias Schacht. The counting lemma for regular k-uniform hypergraphs. *Random Structures Algorithms*, 28(2):113–179, 2006.

[100] Sofia C. Olhede and Patrick J. Wolfe. Network histograms and universality of blockmodel approximation. *CoRR*, pages –1–1, 2013.

[101] Peter Orbanz and Daniel Roy. Bayesian models of graphs, arrays and other exchangeable random structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 99(PrePrints):1, 2014.

[102] Prasad Raghavendra and David Steurer. How to round any CSP. In *2009 50th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2009)*, pages 586–594. IEEE Computer Soc., Los Alamitos, CA, 2009.

[103] Vojtěch Rödl and Richard A. Duke. On graphs with small subgraphs of large chromatic number. *Graphs Combin.*, 1(1):91–96, 1985.

[104] Vojtěch Rödl and Mathias Schacht. Property testing in hypergraphs and the removal lemma [extended abstract]. In *STOC'07—Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, pages 488–495. ACM, New York, 2007.

[105] Vojtěch Rödl and Mathias Schacht. Regular partitions of hypergraphs: regularity lemmas. *Combin. Probab. Comput.*, 16(6):833–885, 2007.

[106] Vojtěch Rödl and Jozef Skokan. Regularity lemma for k-uniform hypergraphs. *Random Structures Algorithms*, 25(1):1–42, 2004.

[107] Eyal Rozenberg. *Lower Bounds and Structural Results in Property Testing of Dense Combinatorial Structures*. dissertation, Technion, 2012.

[108] Ronitt Rubinfeld and Asaf Shapira. Sublinear time algorithms. *SIAM J. Discrete Math.*, 25(4):1562–1588, 2011.

[109] Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM J. Comput.*, 25(2):252–271, 1996.

[110] Mark Rudelson and Roman Vershynin. Sampling from large matrices: an approach through geometric functional analysis. *J. ACM*, 54(4):Art. 21, 19 pp. (electronic), 2007.

[111] Imre Z. Ruzsa and Endre Szemerédi. Triple systems with no six points carrying three triangles. In *Combinatorics (Proc. Fifth Hungarian Colloq., Keszthely, 1976), Vol. II*, volume 18 of *Colloq. Math. Soc. János Bolyai*, pages 939–945. North-Holland, Amsterdam-New York, 1978.

[112] Igal Sason. On refined versions of the Azuma-Hoeffding inequality with applications in information theory, 2011. preprint, arXiv:1111.1977.

[113] Ya. G. Sinaĭ. *Theory of phase transitions: rigorous results*, volume 108 of *International Series in Natural Philosophy*. Pergamon Press, Oxford-Elmsford, N.Y., 1982. Translated from the Russian by J. Fritz, A. Krámli, P. Major and D. Szász.

[114] Christian Sohler. Almost optimal canonical property testers for satisfiability. In *53rd Annual IEEE Symposium on Foundations of Computer Science, FOCS 2012, New Brunswick, NJ, USA, October 20-23, 2012*, pages 541–550, 2012.

[115] Endre Szemerédi. Regular partitions of graphs. In *Problèmes combinatoires et théorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976)*, volume 260 of *Colloq. Internat. CNRS*, pages 399–401. CNRS, Paris, 1978.

[116] Fa-Yueh Wu. The Potts model. *Rev. Modern Phys.*, 54(1):235–268, 1982.

[117] Fa-Yueh Wu. Potts model and graph theory. *J. Statist. Phys.*, 52(1-2):99–112, 1988.

[118] Grigory Yaroslavtsev. Going for speed: Sublinear algorithms for dense r-CSPs, 2014. preprint,arXiv:1407.7887.

[119] Yufei Zhao. Hypergraph limits: a regularity approach. *Random Structures Algorithms*, 47(2):205–226, 2015.

Notation Index

Subject Index