Limits of CSP Problems and Efficient Parameter Testing (Revised Version)

Marek Karpinski* Roland Markó†

Abstract

We present a unified framework on the limits of *constraint satisfaction problems* (CSPs) and efficient parameter testing which depends only on array exchangeability and the method of cut decomposition without recourse to the weakly regular partitions. In particular, we formulate and prove a representation theorem for compact colored r-uniform directed hypergraph (r-graph) limits, and apply this to rCSP limits. We investigate the sample complexity of testable r-graph parameters, we discuss the generalized ground state energies and demonstrate that they are efficiently testable.

1 Introduction

We study the limits and efficient parameter testing properties of *Maximum Constraint Satisfaction Problems* of arity r (MAX-rCSP or rCSP for short), c.f. e.g., [4]. These two topics, limiting behavior and parameter estimation, are treated in the paper to a degree separately, as they require a different set of ideas and could be analyzed on their own right. The establishment of the underlying connection between convergence and testability is one of the main applications of the limit theory of dense discrete structures, see [10], [11].

In the first part of the paper we develop a general framework for the above 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 [10],[11],[24], and [26] for graphs including connections to statistical physics and complexity theory, and were subsequently extended to

Key words and phrases. Graph limits, constraint satisfaction problem, hypergraph parameter testing, exchangeability.

*Dept. of Computer Science and the Hausdroff Center for Mathematics, University of Bonn. Supported in part by DFG grants, the Hausdorff grant EXC59-1. Research partly supported by Microsoft Research New England. E-mail: marek@cs.uni-bonn.de

†Hausdorff Center for Mathematics, University of Bonn. Supported in part by a Hausdorff scholarship. E-mail: roland.marko@hcm.uni-bonn.de

hypergraphs by Elek and Szegedy [15] via the ultralimit method. The central concept of 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 node sampling. Our line of work particularly relies on ideas presented in [14] 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 an infinite random graph model that has the property of exchangeability. The precise definitions, references and results will be given in Section 2, 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 2.12 below. This extends the result of Diaconis and Janson [14], and of Lovász and Szegedy [25]. As an application, the description of the limit space of rCSPs is presented subsequent to the aforementioned theorem.

The second part of the paper, Sections 3 to 5, is dedicated to the introduction of a notion of efficient parameter testability of r-graphs and rCSP problems. We use the limit framework from the first part of the paper to formulate several results on it, which are proved with the aid of the cut decomposition method. We set our focus especially on parameters called *ground state energies* and study variants of them. These are in close relationship with MAX-rCSP problems, our results can be regarded as the continuous generalization of the former. We rely on the notion of parameter testing and sample complexity, that was introduced by Goldreich, Goldwasser, and Ron [17] and was employed in the graph limit theory in [11]. A graph parameter is testable in the sense of [11], when its value is estimable through a uniform sampling process, where the sample size only depends on the desired error gap, see Definition 3.3 below for the precise formulation. The characterization of the real functions on the graph space was carried out in [11], the original motivation of the current paper was to provide an analogous characterization for efficiently testable parameters. These latter are parameters, whose required sample size for the estimation is at most polynomial in the multiplicative inverse of the error.

The investigation of such parameters has been an active area of research for the finite setting in complexity theory. The method of exhaustive sampling in order to approximately solve NP-hard problems was proposed by Arora, Karger and Karpinski [5], their upper bound on the required sample size was still logarithmically increasing in the size of the problem. The approach in [5] enabled the employment of linear programming techniques. Subsequently, the testability of MAX-CUT was shown in [17], explicit upper bounds for the sample complexity in the general boolean MAX-rCSP were given by Alon, F. de la Vega, Kannan and Karpinski [4] using cut decomposition of r-arrays and sampling, that was inspired by the introduction of weak regularity by Frieze and Kannan [16]. In [4] and [16], the design of polynomial time approximation schemes (PTAS) in order to find not only an approximate value for MAX-rCSP, but also an assignment to the base variables that certify this value was an important subject, we did not pursue the generalization regarding this aspect in the current work. The achievements of these two aforementioned contributions turned out to be highly influential, and took also a key role in the first elementary treatment of graph limits and in the definition of the $\delta_\square$-metric in [10] that defines an equivalent topology on the limit space to the subgraph density convergence.

The best currently known upper bound on the sample complexity of MAX-rCSP is $\mathcal{O}(\varepsilon^{-4})$, and has been shown by Mathieu and Schudy [27], see also Alon, F. de la Vega, Kannan and Karpinski [4]. Unfortunately, the approach of [27] does not seem to have a natural counterpart in the continuous setting, although one can use their result on the *sample* to achieve an improved upper bound on the sample complexity. We mention that for the original problem we do not aim to produce an assignment for MAX-rCSP, or a partition for the ground state energy whose evaluation is nearly optimal as opposed to the above works, although we believe this could be done without serious difficulties.

Our contribution in the second part of the paper is the following. By employing a refined version of the proof of the main result of [4] adapted to the continuous setting we are able to prove the analogous efficient testability result for a general finite state space for ground state energies, see Theorem 4.4 in Section 4 for a precise formulation. Among the applications of this development we analyze the testability of the microcanonical version of ground state energies providing the first explicit upper bounds on efficiency. For the finite version a similar question was investigated by F. de la Vega, Kannan and Karpinski [13] by imposing additional global constraints (meaning a finite number of them with unbounded arity). Furthermore, the continuous version of the quadratic assignment problem is treated the first time in a sample complexity context, this subject is related to the recent contributions to the topic of approximate graph isomorphism and homomorphism, see [22] and [8].

1.1 Outline of the paper

The organization of this paper is as follows. In Section 2 we develop the limit theory for $\mathcal{K}$-decorated r-uniform directed hypergraphs with reference to previously known special (and in some way generic) cases, and use the representation of the limit to describe the limit space of rCSP problems. In Section 3 the basic notion of efficiency in context of parameter testing is given with some additional examples. The subsequent Section 4 contains the proof of Theorem 4.4 regarding ground state energies of r-graphons, and in the following Section 5 some variants are examined, in particular microcanonical energies and the quadratic assignment problem. We summarize possible directions of further research in Section 6.

2 Limit theory and related notions

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 (rCSP formula). *Let $V = \{x_1, x_2, \ldots, x_n\}$ be the set of variables, $x_e = (x_{e_1}, \ldots, x_{e_r}) \in V^r$ 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 $\omega(l_1, \ldots, l_n) = f(l_{e_1}, \ldots, l_{e_r})$ 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 $V(F) = \{x_1, x_2, \ldots, x_n\}$ 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 Theorem 2.1 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 2.2 (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 paper, 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 section 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 [24].

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 Theorem 3.3.

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.

2.1 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.

The basic content of the current subsection starts with the general setting given above, CSPs will be considered as a special case in this topic whose limit characterization will be derived at the end. Some of the basic cases are already settled regarding the representation of the limits, we refer to Lovász and Szegedy [24], [25], [23] for the $r = 2$, general $\mathcal{K}$, undirected case, to Elek and Szegedy [15] for the general r, $\mathcal{K} = \{0, 1\}$, undirected case; and Diaconis and Janson [14] 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 further generalized or applied by Zhao [30] to general r; respectively by Aroskar [7] to the directed case; respectively by Austin [9] general r and by Janson [20] to the directed case where the graph induces a partial order on the vertex set.

Definition of convergence Let $\mathcal{C}$ denote space $C(\mathcal{K})$ of continuous functionals on $\mathcal{K}$, and let $\mathcal{F} \subset \mathcal{C}$ 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 $\mathcal{C}$ in the L^∞-norm.

Denote by $\Pi(S) = \Pi^r(S)$ the set of all unlabeled S-decorated directed r-uniform hypergraphs for some arbitrary set S, where we will suppress r in the notation, when it is clear which r is meant (alternatively, $\Pi(S)$ denotes the isomorphism classes of the node labeled respective objects). The set $\Pi_k(S)$ denotes the elements of $\Pi(S)$ of vertex cardinality k. Let $\mathbb{G}(k, F)$ denote the random induced subformula of F on the set $S \subset V(F)$ that is chosen uniformly among the subsets of $V(F)$ of cardinality k. We define the homomorphism densities next.

Definition 2.3. *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*

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

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. Note that this way if $F, G \in \Pi^r$, 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 $\overline{t(F, .)}$ 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 $\tau_{\mathrm{inj}}(G)$ and $\tau^+_{\mathrm{inj}}(G)$, and the sets $\Pi_{\mathrm{inj}}(\mathcal{K}) = \tau_{\mathrm{inj}}(\Pi(\mathcal{K}))$ and $\Pi_{\mathrm{inj}}(\mathcal{K})^+$ are defined analogously. It was shown in [24] that

$$|t_{\mathrm{inj}}(F,G) - t(F,G)| \leq \frac{|V(F)|^2 \|F\|_\infty}{2|V(G)|} \tag{2.4}$$

for any pair $F \in \Pi(\mathcal{C})$ 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 [25].

Theorem 2.4. *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 $F \in \Pi(\mathcal{C})$, 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 $t_{\mathrm{inj}}(F,G_n)$, 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 $F' \in \Pi(\mathcal{C})$ 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

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

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 $F \in \Pi(\mathcal{C})$ 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 $M = \langle t(F, .) | F \in \Pi(\mathcal{C}) \rangle \subset C(\Pi_k(\mathcal{K}))$ is an algebra containing the constant function, and that it separates any two elements of $\Pi_k(\mathcal{K})$. It follows that $\langle t(F, .) | F \in \Pi(\mathcal{C}) \rangle$ 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 (2.4) 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 $F_1, F_2 \in \Pi(\mathcal{C})$ we have $t(F_1, G)t(F_2, G) = t(F_1F_2, G)$ for any $G \in \Pi_k(\mathcal{K})$, where the product F_1F_2 denotes the disjoint union of the two $\mathcal{C}$-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 $F/\mathcal{P} \in \Pi^{|\mathcal{P}|}(\mathcal{C})$ 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

$$\mathrm{inj}(F, G) = \sum_{\phi \in S_k} \prod_{i_1,\dots,i_r=1}^{k} F(i_1, \dots, i_r)(G(\phi(i_1), \dots, \phi(i_r))). \tag{2.5}$$

Now fix $G_1, G_2 \in \Pi_k(\mathcal{K})$ and let $F \in \Pi_k(\mathcal{C})$ 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 (2.5) 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 (2.4), 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 2.5. *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 2.4 hold, then we say that $(G_n)_{n=1}^\infty$ converges.*

We would like to add that, in the light of Theorem 2.4, 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 2.6. *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 2.4, 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. $\square$

Exchangeable arrays The correspondence analogous to the approach of Diaconis and Janson in [14] 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 2.7 (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 $\rho_1=\dots=\rho_r \in S_{\mathbb{N}}$.

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.

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 2.8. *[21] 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*

$$H(e)=f(U_\emptyset,U_{\{e_1\}},U_{\{e_2\}},\dots,U_{\overline{e}\setminus\{e_r\}},U_{\overline{e}}) \tag{2.6}$$

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 2.8 was first proved by de Finetti [12] (in the case $\mathcal{K} = \{0, 1\}$) and by Hewitt and Savage [18] (in the case of general $\mathcal{K}$) for $r = 1$, independently by Aldous [1] and Hoover [19] for $r = 2$, and by Kallenberg [21] for arbitrary $r \geq 3$. For equivalent formulations, proofs and further connections to related areas see the recent survey of Austin [9].

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 $\mathrm{Samp}(f)$ given by a function f as above.

With the aid of Theorem 2.8 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 [14, Section 2 to 5], combined with the compactification argument regarding the limit space from [25], see also [23, 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 [14].

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 Theorem 2.7 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 2.9. *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 $F \in \Pi_k(\mathcal{C})$, 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 2.4). 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 Theorem 2.6, 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 2.4 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 2.4 and Theorem 2.9 in order to formulate the following theorem, which is the crucial ingredient to the desired representation of limits.

Theorem 2.10. *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 $F \in \Pi_k(\mathcal{C})$ 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 2.9 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 2.4, Theorem 2.6, 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 2.9, 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 2.9.

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 $t(F, \Gamma) = \mathbb{E}t_{\mathrm{inj}}(F, H|_{[k]})$ between the equivalent objects.

□

The characterization of the aforementioned extreme points in Theorem 2.10 was given [14] 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 [14, Theorem 5.5.].

Theorem 2.11. *[14] 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 2.8 does not depend on the variable corresponding to the empty set.*

Graphons as limit objects 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 paper 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.

If $W \in \Xi^r$ and $F \in \Pi^r$, then the F-density of W is defined as

$$t(F, W) = \int\limits_{[0,1]^{\mathfrak{h}([k],r-1)}} \prod_{e \in E(F)} W(x_{\mathfrak{h}(e,r-1)}) \prod_{e \notin E(F)} (1 - W(x_{\mathfrak{h}(e,r-1)})) \mathrm{d}\lambda(x_{\mathfrak{h}([k],r-1)}). \tag{2.7}$$

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 [15].

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 (2.7) as

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

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 2.8. Furthermore, we define $\Gamma_W \in \overline{\Pi(\mathcal{K})^+} \setminus \Pi(\mathcal{K})^+$ to be the element associated to H_W through Theorem 2.9.

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 2.8). It is an immediate consequence of Theorem 2.9 and Theorem 2.10 above.

Theorem 2.12. *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 [25].

In previous works, for example in [14], 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 $0 \leq W^{\pi^*(u)}(x) = W^u(x_{\pi^*(\mathfrak{h}([r],r-1))})$ 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 2.12. 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.

In several applications 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, 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

$$|t_{\mathrm{inj}}(F,G) - t(F,W_G)| \le \frac{\binom{k}{2}}{n-\binom{k}{2}} \tag{2.9}$$

for each $F \in \Pi_k^r$, 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.

2.2 Representation of rCSP formulas as hypergraphs, and their convergence

In this subsection we elaborate on how homomorphism and sampling is meant in the CSP context, and formulate a representation the limit space in that context. Recall Theorem 2.1 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 Theorem 2.7 that the relation $\omega = (f; x_e) \in F[x_{i_1}, \dots, x_{i_k}]$ is equivalent to the relation

$$\phi(\omega) = (f_\phi, x_{\phi(e)}) \in F[x_{i_1}, \dots, x_{i_k}] \tag{2.10}$$

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 2.12 is the following.

Corollary 2.13. *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 $\mathfrak{eval}(F) \colon [n] \times \cdots \times [n] \to \{0, 1, \dots, d\}^{([q]^r)}$ that obeys the symmetry condition given after (2.2). This will be the way throughout the paper 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 $\{0,1,\dots,d\}^{([q]^r)}$ 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 Theorem 2.5. We describe here the special case for parallel multicolored graphs, see also [25].

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

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

forms an L^∞-dense subset in $C(L)$, therefore Theorem 2.4 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 2.12 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 (2.10) of the finite objects is inherited in the limit.

We state now the general evaluation rCSP version of Theorem 2.12.

Corollary 2.14. *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 Theorem 2.13 when elements of* L *are considered as maps from* $[q]^r$ *to non-negative integers.*

Exchangeable partition-indexed processes We conclude the subsection with a remark that is motivated by the array representation of rCSPs. 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 2.9 with rCSP limits is the following process, that preserves every piece of information contained in the evaluation representation.

Definition 2.15. *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 2.8 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.

3 Graph and graphon parameter testability

First we will invoke the method of sampling from $\mathcal{K}$-colored r-graphs and r-graphons, as well as inspect the metrics that will occur later.

Let $(U_S)_{S \in \mathfrak{h}([k],r)}$ be an independent uniform sample from $[0,1]$. Then for an r-graph G, respectively an r-graphon W, the random r-graphs $\mathbb{G}(k,G)$ and $\mathbb{G}(k,W)$ have vertex set $[k]$, and edge weights $W_G((U_{p_e(S)})_{S \in \mathfrak{h}(\hat{e},r)})$, respectively $W((U_{p_e(S)})_{S \in \mathfrak{h}(\hat{e},r)})$. Keep in mind, that $\mathbb{G}(k,G) \neq \mathbb{G}(k,W_G)$, the first term corresponds to sampling without, the second with replacement, but it is true that $\mathbb{P}(\mathbb{G}(k,G) \neq \mathbb{G}(k,W_G)) \leq \frac{r^2}{|V(G)|}$.

Norms and distances We also mention the definitions of the norms and distances that will play a important role in what follows. In the next definition each object is real-valued.

Definition 3.1. *The cut norm of an $n \times \cdots \times n$ 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 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 or r-arrays F and G on the same vertex set $[n]$ is

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

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

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

The continuous counterparts are described as follows.

Definition 3.2. *The cut norm of a naive r-graphon W is*

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

the cut distance of two naive r-graphons 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 norm even in the most simple case, when $\mathcal{K} = \{0,1\}$. Examples of subgraphs whose densities behave well with respect to the above norms are linear hypegraphs, that have the property that any two distinct edges intersect at most in one node.

Originally, in [10], testability of $(\mathcal{K}, r)$-graph parameters (which are real functions invariant under r-graph-isomorphisms) was defined as follows.

Definition 3.3. *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*

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

A $(\mathcal{K}, r)$-graphon parameter f is a functional on the space of r-graphons that is invariant under the action induced by measure preserving maps from $[0,1]$ to $[0,1]$, that is, $f(W) = f(W^\phi)$. Their testability is defined analogously to Definition 3.3.

Testing parameters A characterization of the testability of a graph parameter in terms of graph limits was developed in [10] 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 Theorem 3.3.

Theorem 3.4. *[10] 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 $k \geq k(\varepsilon)$ 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 $|V(G_1)|, |V(G_2)| \geq n_0$ 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 $G_n \to W$.

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.

Definition 3.5. *[26] $\mathcal{P}$ 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 $\mathbb{P}(\mathbb{G}(k, G) \in \mathcal{P}') \leq \frac{1}{3}$.

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 $\mathcal{P}'$, but not the characteristic of testability. The link below between the two notions is a simple consequence of the definitions. These concepts may be extended to the infinitary space of graphons, where a similar notion of sampling is available.

Lemma 3.6. *[26] $\mathcal{P}$ is a testable graph property if and only if $d_1(., \mathcal{P})$ is a testable graph parameter.*

We provide some remarks yielded by Theorem 3.4.

Remark 3.7. In the case $r = 2$, the testability of a graphon parameter is equivalent to continuity in the $\delta_\square$ distance.

Remark 3.8. The intuitive reason for the absence of an analogous, easily applicable characterization of testability for higher rank uniform hypergraphs as in Theorem 3.4 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.

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.

3.1 Examples of testable properties and parameters

We introduce now a notion of efficient parameter testability. Theorem 3.3 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 3.9. *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 Theorem 3.9 by saying that a testable parameter f is efficiently testable if its sample complexity is polynomial in $1/\varepsilon$.

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. [3] for a standard proof and a wide range of applications. We will also apply a more elaborate version of this concentration inequality below.

Lemma 3.10 (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 3.11. 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 [24], see also for hypergraphs Theorem 11 in [15].

Lemma 3.12. *[24, 15] 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

$$\mathbb{P}(|t(F,G) - t(F,\mathbb{G}(k,G))| > \varepsilon) < 2\exp\left(-\frac{\varepsilon^2 k}{18|V(F)|^2}\right). \tag{3.1}$$

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 $\mathcal{O}(\log(\frac{1}{\varepsilon})\varepsilon^{-2})$-testable. In the case of $(\mathcal{K}, r)$-graphs for arbitrary r the same as Theorem 3.12 holds, this can be shown by a straightforward application of the Azuma-Hoeffding inequality, Theorem 3.10, as in the original proofs.

Example 3.13. 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

$$\mathcal{E}_\phi(G,J,h) = \frac{1}{n^2}\sum_{1\leq i,j\leq q} J_{ij} e_G(\phi^{-1}(i), \phi^{-1}(j)) + \frac{1}{n}\sum_{1\leq i\leq q} h_i|\phi^{-1}(i)|, \tag{3.2}$$

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

$$\hat{\mathcal{E}}(G,J,h) = \max_{\phi\colon V(G)\to[q]} \mathcal{E}_\phi(G,J,h), \tag{3.3}$$

that is the *ground state energy* of the graph G (cf. [11]) 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 [29]. 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 $\mathcal{O}(\frac{1}{\varepsilon^4})$(see [4],[27]).

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 [11], (2.16), namely

$$\frac{1}{|V(G)|^2} \ln \hom(G, H) = \hat{\mathcal{E}}(G, J) + \mathcal{O}\left(\frac{1}{|V(G)|}\right), \tag{3.4}$$

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 section. Additionally we note, that an analogous statement to (3.4) 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 3.14. *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 $J^{i,j}(k,l) = \mathbb{I}_i(k)\mathbb{I}_j(l)$ that is defined below.

Example 3.15. 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 [10] will be presented in the next lemma.

Lemma 3.16. *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* $\max\{2^{\mathcal{O}\left(\frac{1}{\varepsilon^{2/\alpha}}\right)}, n_0(\varepsilon)\}$*-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

$$|f(G) - f(\mathbb{G}(k,G))| \leq C\left[\delta_\square(G, \mathbb{G}(k,G))\right]^\alpha < C\left(\frac{10}{\sqrt{\log_2 k}}\right)^\alpha, \tag{3.5}$$

with probability at least $1-\exp(-\frac{k^2}{2\log_2 k})$. The last probability bound in (3.5) is the statement of Theorem 2.9 of [10]. We may rewrite (3.5) by setting $\varepsilon = C\left(\frac{10}{\sqrt{\log_2 k}}\right)^\alpha$, the substitution implies that f is $2^{\mathcal{O}\left(\varepsilon^{-2/\alpha}\right)}$-testable, whenever $n_0(\varepsilon) \leq 2^{\mathcal{O}(\varepsilon^{-2/\alpha})}$. $\square$

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.

4 Testability of the ground state energy

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 $\phi = (\phi_1, \dots, \phi_q)$. 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 **??** of the energies of naive r-kernels, the version for true $(\mathcal{K}, r)$-graphons is

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

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 $\phi = (\mathbb{I}_{T_1}, \dots, \mathbb{I}_{T_q})$, 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 $J_{z_1,\dots,z_r}(x) = a_{z_1,\dots,z_r}\mathbb{I}_1(x)$ is a constant multiple of the indicator function of 1 we retrieve the original GSE notion in Theorem 3.13 and Theorem 3.14.

Remark 4.1. Ground state energies and subgraph densities are Lipschitz continuous graph parameters in the sense of Theorem 3.16 ([10],[11]), 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 4.2. *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 Theorem 2.2) 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 paper will be a generalization of the following theorem on sample complexity of rCSPs with respect to these factors.

The main result of [4] was the following.

Theorem 4.3. *[4] 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 [27] to $k \in \mathcal{O}(\varepsilon^{-4})$. 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 $(z_1,\ldots,z_r)$ 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 paper.

Theorem 4.4. *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*

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

A direct consequence of Theorem 4.4 is the corresponding result for layered ground state energies.

Corollary 4.5. *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 4.4 dealing with ground state energies of canonical form. $\square$

We start with the proof of Theorem 4.4 by providing the necessary background. We will proceed loosely along the lines of the proof of Theorem 4.3 from [4] 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 [4] carry through adapted to our continuous setting with some modifications, and we will also draw on tools from [10] and [11]. 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 4.6. *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*

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

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 $1 \leq t \leq \binom{k}{2}$, 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, $|Y_{t-1} - Y_t| \leq \frac{1}{k^2}\|J\|_\infty\|W\|_\infty$. By the Azuma-Hoeffding inequality, Theorem 3.10, it follows that

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

for any $\rho > 0$.

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

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

□

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

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

Proof. Trivially we have $\mathcal{E}(G,J)\geq\hat{\mathcal{E}}(G,J)$. 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

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

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 [4], and is closely related to the Weak Regularity Lemma, **??**, of [16], and its continuous version **??**. The result is a centerpiece of the cut decomposition method.

Lemma 4.8. *Let $\varepsilon>0$ arbitrary. For any bounded measurable function $W\colon[0,1]^r\to\mathbb{R}$ there exist an $s\leq\frac{1}{\varepsilon^2}$, 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 $B=\sum_{i=1}^s d_i\mathbb{I}_{S_i^1\times\cdots\times S_i^r}$ it holds that*

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

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

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

Proof. We construct stepwise the required rectangles and the respective coefficients implicitly. Let $W^0 = W$, and suppose that after the t'th 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 $W^t = W - \sum_{i=1}^t d_i \mathbb{I}_{S_i^1\times\dots\times S_i^r}$. We proceed to the $(t+1)$'st step, where two possible situations can occur. The first case is when

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

This implies by definition that there exist measurable subsets $S_{t+1}^1,\dots,S_{t+1}^r$ of $[0,1]$ such that $|\int_{S_{t+1}^1\times\dots\times S_{t+1}^r} W^t(x)\mathrm{d}\lambda(x)| \geq \varepsilon\|W\|_2$. We define d_{t+1} to be the average of W^t on the product set $S_{t+1}^1\times\dots\times S_{t+1}^r$, 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

$$\begin{aligned}\|W^t\|_2^2 - \|W^{t+1}\|_2^2 &= \int\limits_{S_{t+1}^1\times\dots\times S_{t+1}^r} (W^t)^2(x)\mathrm{d}\lambda(x) - \int\limits_{S_{t+1}^1\times\dots\times S_{t+1}^r} (W^t(x)-d_{t+1})^2\mathrm{d}\lambda(x)\\ &= d_{t+1}^2\lambda(S_{t+1}^1)\dots\lambda(S_{t+1}^r) \geq \varepsilon^2\|W\|_2^2. \end{aligned} \tag{4.6}$$

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

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

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

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

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

Next we state that the cut approximation provided by Theorem 4.8 is invariant under sampling. This is a crucial point of the whole argument, and is the r-dimensional generalization of Lemma 4.6 from [10].

Lemma 4.9. *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 [4].

Lemma 4.10. *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 Theorem 4.10 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

$$\begin{aligned} k^r\|G\|_\square \leq{}& E_{Q_1,\ldots,Q_r} \max_{Q_i'\subset Q_i} \max\{G(P_G(Q_1'),\ldots,P_G(Q_r')); (-G)(P_{-G}(Q_1'),\ldots,P_{-G}(Q_r'))\} \\ &+ \frac{rk^r}{\sqrt{p}}\|G\|_\infty, \end{aligned} \tag{4.7}$$

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

$$k^r\mathbb{E}_{U_{[k]}}\|G\|_\square \leq E_{Q_1,\ldots Q_r}\mathbb{E}_{U_Q}[\mathbb{E}_{U_{Q^c}} \max_{Q_i'\subset Q_i} \max\{G(P_G(Q_1')\cap Q^c,\ldots,P_G(Q_r')\cap Q^c);$$

$$(-G)(P_{-G}(Q_1') \cap Q^c, \dots, P_{-G}(Q_r') \cap Q^c)\}] + \frac{rk^r}{\sqrt{p}}\|G\|_\infty + pr^3k^{r-1}\|G\|_\infty, \tag{4.8}$$

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 $Q_1, \dots Q_r$ 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 $Y(Q_1', \dots, Q_r') = G(P_G(Q_1') \cap Q^c, \dots, P_G(Q_r') \cap Q^c)$, 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

$$\begin{aligned}\mathbb{E}_{U_{Q^c}} Y(Q_1', \dots, Q_r') &\leq \sum_{\substack{t_1,\dots,t_r \in Q^c \\ t_i \neq t_j}} \mathbb{E}_{U_{Q^c}} G(t_1, \dots, t_r) \mathbb{I}_{P_G(Q_1')}(t_1) \dots \mathbb{I}_{P_G(Q_r')}(t_r) + r^2k^{r-1}\|W\|_\infty \\ &\leq k^r \int_{T_1 \times \dots \times T_r} W(x) \mathrm{d}\lambda(x) + r^2k^{r-1}\|W\|_\infty \leq k^r\|W\|_\square + r^2k^{r-1}\|W\|_\infty.\end{aligned}$$

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

$$\mathbb{P}(Y(Q_1', \dots, Q_r') \geq \mathbb{E}_{U_{Q^c}} Y(Q_1', \dots, Q_r') + \rho k^r \|W\|_\infty) < \exp\left(-\frac{\rho^2 k}{8r^2}\right), \tag{4.9}$$

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 $1 - 2^{pr+1}\exp(-\frac{\rho^2 k}{8r^2})$, 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

$$\mathbb{E}_{U_{[k]}}\|G\|_\square \leq \|W\|_\square + \|W\|_\infty \left(\frac{r}{\sqrt{p}} + \frac{pr^3}{k} + \rho + \frac{r^2}{k} + 2^{pr+1}\exp\left(-\frac{\rho^2 k}{8r^2}\right)\right).$$

Let $p = \sqrt{k}$ and $\rho = \frac{4r^2}{\sqrt[4]{k}}$. Then

$$\mathbb{E}_{U_{[k]}}\|G\|_\square \leq \|W\|_\square + \|W\|_\infty \left(\frac{r}{\sqrt[4]{k}} + \frac{r^3}{\sqrt{k}} + \frac{4r^2}{\sqrt[4]{k}} + \frac{r^2}{k} + \exp\left(2\sqrt{k}r - 2r^2\sqrt{k}\right)\right)$$

$$\leq \|W\|_\square + \|W\|_\infty \left(\frac{\varepsilon}{16r} + \frac{\varepsilon^2}{2^8 r} + \frac{\varepsilon}{4} + \frac{\varepsilon^4}{2^{16} r^6} + \frac{\varepsilon^2}{2^8 r^6} \right) \leq \|W\|_\square + \varepsilon/2 \|W\|_\infty.$$

The direction concerning the lower bound, $\mathbb{E}\|G\|_\square \geq \|W\|_\square - \varepsilon/2$ 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

$$\begin{aligned} \mathbb{P}\left(|\|G\|_\square - \|W\|_\square| > \varepsilon \|W\|_\infty \right) &\leq \mathbb{P}\left(\left| \mathbb{E}\|G\|_\square - \frac{1}{k^r}\|G\|_\square \right| > \varepsilon/2 \|W\|_\infty \right) \\ &\leq 2 \exp\left(-\frac{\varepsilon^2 k}{32 r^2} \right). \end{aligned}$$

□

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, Theorem 3.10, and suits well the situation when the martingale jump sizes have inhomogeneous distribution. It can be found together with a proof in the survey [28] as Corollary 3.

Lemma 4.11 (Generalized Azuma-Hoeffding inequality)**.** *Let $k \geq 1$ and $(X_n)_{n=0}^k$ be a martingale sequence with respect to the natural filtration $(\mathcal{F}_n)_{n=1}^k$. If $|X_n - X_{n+1}| \leq d$ almost surely and $\mathbb{E}[(X_n - X_{n+1})^2 \mid \mathcal{F}_n] \leq \sigma^2$ for each $n \in [k]$, then for every $n \leq k$ and $\delta > 0$ it holds that*

$$\mathbb{P}(X_n - X_0 > \delta n) \leq \exp\left(-n \frac{\sigma^2}{d^2} \left((1 + \frac{\delta d}{\sigma^2}) \ln(1 + \frac{\delta d}{\sigma^2}) - \frac{\delta d}{\sigma^2} \right) \right). \tag{4.10}$$

Measurability for all of the following functions is assumed.

Lemma 4.12. *Let $c_m \colon [0,1] \to \mathbb{R}$, $U_{i,m} \colon [0,1] \to \mathbb{R}$ for $i = 1, \ldots, s$, $m = 1, \ldots, q$, $u \in \mathbb{R}^{s \times q}$, $\alpha \in \mathbb{R}$. Let d and σ be positive reals such that $\|c\|_\infty \leq d$ and $\|c\|_2 \leq \sigma$ and set $\gamma = \frac{\sigma^2}{d^2}$. If the optimum of the linear program*

$$\begin{aligned} &\text{maximize} && \int_0^1 \sum_{m=1}^q f_m(t) c_m(t) \mathrm{d}t && \\ &\text{subject to} && \int_0^1 f_m(t) U_{i,m}(t) \mathrm{d}t \leq u_{i,m} && \textit{for } i \in [s] \textit{ and } m \in [q] \\ & && 0 \leq f_m(t) \leq 1 && \textit{for } t \in [0,1] \textit{ and } m \in [q] \\ & && \sum_{m=1}^q f_m(t) = 1 && \textit{for } t \in [0,1] \end{aligned}$$

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

$$\begin{array}{lll}
\text{maximize} & \displaystyle\sum_{1\le n\le k}\sum_{m=1}^{q}\frac{1}{k}x_{n,m}c_m(X_n) & \\
\text{subject to} & \displaystyle\sum_{1\le n\le k}\frac{1}{k}x_{n,m}U_{i,m}(X_n) \le u_{i,m} - \delta\|U\|_\infty & \textit{for } i\in[s] \textit{ and } m\in[q] \\
& 0 \le x_{n,m} \le 1 & \textit{for } n\in[k] \textit{ and } m\in[q] \\
& \displaystyle\sum_{m=1}^{q} x_{n,m} = 1 & \textit{for } n\in[k]
\end{array}$$

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 1. 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 $B = \{\, x \mid x_{i,m} \le v_{i,m}\,\} \subset \mathbb{R}^{sq}$, this set is also a nonempty convex closed set. The absence of a solution to $Af \le v$ is equivalent to saying that $C \cap B$ is empty. It follows from the Separation Theorem for convex closed sets that there is a $0 \ne y' \in \mathbb{R}^{sq}$ 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 $y^T(Af) \le y^T v$ has no solution. □

From this lemma the finitary version follows without any difficulties.

Claim 2. 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 $A_{i,m}(t) = \sum_{n=1}^{k} \frac{B_{(i,m),n}}{k} \mathbb{I}_{[\frac{n-1}{k}, \frac{n}{k})}(t)$ 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 1 verifies the current claim. $\square$

The assumption of the lemma is by Claim 1 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 $h_1(t) = \sum_{m=1}^{q} \mathbb{I}_{T_m}(t) [\sum_{i=1}^{s} y_{i,m} U_{i,m}(t)]$ and $h_2(t) = \sum_{m=1}^{q} \mathbb{I}_{T_m}(t) \beta c_m(t)$. 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 $\|h_1\|_\infty = \| \sum_{i=1}^{s} \sum_{m=1}^{q} \mathbb{I}_{T_m} U_{i,m} y_{i,m} \|_\infty \leq \|U\|_\infty \sum_{i=1}^{s} \sum_{m=1}^{q} |y_{i,m}| \leq \|U\|_\infty$. Moreover, by Theorem 4.11 the event

$$\frac{1}{k} \sum_{n=1}^{k} h_2(X_n) < A_2 + \varepsilon \tag{4.11}$$

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 2. □

We start the principal part of the proof of the main theorem in this paper.

Proof of Theorem 4.4. It is enough to prove Theorem 4.4 for tuples of naive $([-d,d],r)$-digraphons. We first employ Theorem 4.6 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 Theorem 4.7 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

$$|\hat{\mathcal{E}}(\mathbb{G}(k,W))-\mathcal{E}(\mathbb{H}(k,W))|\le\varepsilon^2\|W\|_\infty$$

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

$$\mathcal{E}(\mathbb{H}(k,W))\ge\mathcal{E}(W)-\frac{r^2}{k^{1/4}}\|W\|_\infty \tag{4.12}$$

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

$$\begin{aligned}
\mathbb{E}\mathcal{E}(\mathbb{H}(k,W)) &\ge \mathbb{E}\mathcal{E}_y(\mathbb{H}(k,W))\\
&=\mathbb{E}\frac{1}{k^r}\sum_{z\in[q]^r}\sum_{n_1,\dots,n_r=1}^{k}W^z(U_{\mathfrak{h}(\{n_1,\dots,n_r\},r)})\prod_{j=1}^{r}y_{n_j,z_j}\\
&\ge\frac{k!}{k^r(k-r)!}\sum_{z\in[q]^r}\int_{[0,1]^{\mathfrak{h}([r])}}W^z(t_{\mathfrak{h}([r])})\prod_{j=1}^{r}\phi_{z_j}(t_j)\mathrm{d}\lambda(t)-\frac{r^2}{k}\|W\|_\infty\\
&\ge\mathcal{E}_\phi(W)-\frac{r^2}{k}\|W\|_\infty.
\end{aligned}$$

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 $1\le j\le k$. 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, Theorem 3.10, it follows that

$$\begin{aligned}
&\mathbb{P}\left(\mathcal{E}(\mathbb{H}(k,W)) < \mathcal{E}(W) - \frac{2r^2}{k^{1/4}}\|W\|_\infty\right) \\
&\quad\leq \mathbb{P}\left(\mathcal{E}(\mathbb{H}(k,W)) < \mathbb{E}\mathcal{E}(\mathbb{H}(k,W)) - \frac{r^2}{k^{1/4}}\|W\|_\infty\right) \\
&\quad= \mathbb{P}\left(Y_k < Y_0 - \frac{r^2}{k^{1/4}}\|W\|_\infty\right) \leq \exp\left(-\frac{r^2\sqrt{k}}{8}\right). \qquad (4.13)
\end{aligned}$$

So the lower bound (4.12) 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 Theorem 4.8, that is a variant the Cut Decomposition Lemma from [4] (closely related to the Weak Regularity Lemma by Frieze and Kannan [16]), and linear programming duality, in the form of Theorem 4.12. Recall the definition of the cut norm, for $W\colon [0,1]^r \to \mathbb{R}$, it is given as

$$\|W\|_\square = \max_{S^1,\dots,S^r \subset [0,1]} \left| \int\limits_{S^1\times\dots\times S^r} W(x)\mathrm{d}\lambda(x)\right|,$$

and for an r-array G by the expression

$$\|G\|_\square = \frac{1}{k^r}\max_{S^1,\dots,S^r \subset V(G)} \left|G(S^1,\dots,S^r)\right|.$$

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 Theorem 4.8. 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. Theorem 4.9 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 Theorem 4.12 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

$$\mathcal{E}_\phi(W) = \sum_{z\in[q]^r} \int\limits_{[0,1]^r} \prod_{j\in[r]} \phi_{z_j}(t_j) W^z(t) \mathrm{d}\lambda(t), \tag{4.14}$$

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

$$\mathcal{E}_{\mathrm{x}}(\mathbb{H}(k,W)) = \sum_{z\in[q]^r} \frac{1}{k^r} \sum_{n_1,\dots,n_r=1}^{k} \prod_{j\in[r]} x_{t_j,z_j} W^z(U_{n_1}, \dots, U_{n_r}). \tag{4.15}$$

We are going to establish a term-wise connection with respect to the parameter z in the previous formulas. Therefore we consider the function

$$\mathcal{E}^z_\phi(W^z) = \int\limits_{[0,1]^r} \prod_{j\in[r]} \phi_{z_j}(t_j) W^z(t) \mathrm{d}\lambda(t), \tag{4.16}$$

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 (4.14)-(4.16) make prefect sense even when the parameters ϕ and x are only vectors of bounded functions and reals respectively without forming partition.

Theorem 4.8 delivers for any $z \in [q]^r$ an integer $s_z \leq \frac{2^6 q^{2r}}{\varepsilon^2}$, 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

$$\|W^z - \sum_{i=1}^{s_z} d_{z,i} \mathbb{I}_{S_{z,i,1}\times\cdots\times S_{z,i,r}}\|_\square \leq \frac{\varepsilon}{8q^r} \|W^z\|_2,$$

and $\sum_{i=1}^{s_z} |d_{z,i}| \leq \frac{8q^r}{\varepsilon} \|W^z\|_2$. The cut function allows a sufficiently good approximation for $\mathcal{E}_\phi(W^z)$, for any ϕ. Let $D^z = \sum_{i=1}^{s_z} d_{z,i} \mathbb{I}_{S_{z,i,1}\times\cdots\times S_{z,i,r}}$. Then

$$|\mathcal{E}^z_\phi(W^z) - \mathcal{E}^z_\phi(D^z)| = \left| \int\limits_{[0,1]^r} \prod_{j\in[r]} \phi_{z_j}(t_j) \left[W^z(t) - D^z(t)\right] \mathrm{d}\lambda(t) \right|$$

$$\leq \|W^z - D^z\|_\square \leq \frac{\varepsilon}{8q^r}\|W^z\|_\infty.$$

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 $\hat{D}^z = \sum_{i=1}^{s_z} d_{z,i}\mathbb{I}_{\hat{S}_{z,i,1}\times\cdots\times\hat{S}_{z,i,r}}$. First we condition on the event from Theorem 4.9, call this event E_1, that is

$$E_1 = \bigcap_{z\in[q]^r} \left\{ \left| \|\mathbb{H}(k, W^z) - \hat{D}^z\|_\square - \|W^z - D^z\|_\square \right| < \frac{\varepsilon}{8q^r}\|W\|_\infty \right\}.$$

On E_1 it follows that for any x that is a fractional q-partition

$$\begin{aligned} |\mathcal{E}^z_{\mathrm{x}}(\mathbb{H}(k, W^z)) - \mathcal{E}^z_{\mathrm{x}}(\hat{D}^z)| &\leq \|\mathbb{H}(k, W^z) - \hat{D}^z\|_\square \\ &\leq \|W^z - D^z\|_\square + \frac{\varepsilon}{8q^r}\|W\|_\infty. \end{aligned}$$

This implies that

$$|\mathcal{E}_\phi(W) - \mathcal{E}_\phi(D)| \leq \frac{\varepsilon}{8}\|W\|_\infty \quad \text{and} \quad |\mathcal{E}_x(\mathbb{H}(k, W)) - \mathcal{E}_x(\mathbb{H}(k, D))| \leq \frac{\varepsilon}{4}\|W\|_\infty.$$

The probability that E_1 fails is at most $2q^r \exp\left(-\frac{\varepsilon^2 k}{2^{11}r^2q^{2r}}\right)$ whenever $k \geq \left(\frac{2^7q^rr^2}{\varepsilon}\right)^4$ due to Theorem 4.9, in the current theorem we have the condition $k \geq \left(\frac{2^{r+7}q^rr}{\varepsilon}\right)^4 \log(\frac{2^{r+7}q^rr}{\varepsilon})q^r$, which implies the aforementioned one. The failure probability of E_1 is then strictly less than $\frac{\varepsilon}{2^7}$.

Let $\mathcal{S} = \{ S_{z,i,j} \mid z \in [q]^r, 1 \leq i \leq s_z, 1 \leq j \leq r \}$ denote their set, and let $\mathcal{S}'$ stand for the corresponding set on the sample. Note that $s' = |\mathcal{S}| \leq 2^6rq^{3r}\frac{1}{\varepsilon^2}$ 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 $2^6rq^{2r}\frac{1}{\varepsilon^2}$ 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

$$I(b, \eta) = \left\{ \phi \mid \forall z \in [q]^r, 1 \leq i \leq s_z, 1 \leq j \leq r \colon \left| \int\limits_{S_{z,i,j}} \phi_{z_j}(t)\mathrm{d}t - b_{z,i,j} \right| \leq 2\eta \right\},$$

and

$$I'(b, \eta) = \left\{ \mathrm{x} \mid \forall z \in [q]^r, 1 \leq i \leq s_z, 1 \leq j \leq r \colon \left| \frac{1}{k} \sum_{U_n\in S_{z,i,j}} x_{n,z_j} - b_{z,i,j} \right| \leq \eta \right\}$$

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 4.13 (Local linearization)**.** *If $\eta \le \frac{\varepsilon}{16q^r 2^r}$, then for every $b \in \mathcal{A}$ there exist $l_0 \in \mathbb{R}$ and functions $l_1, l_2, \ldots, l_q \colon [0,1] \to \mathbb{R}$ 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 $\int_0^1 \sum_{m=1}^q l_m^2(t)\mathrm{d}t \le 2^{2r+9} r^2 q^{3r} \|W\|_\infty^2$.

Proof. Recall the decomposition of the energies as sums over $z \in [q]^r$ into terms

$$\begin{aligned}
\mathcal{E}_\phi^z(D^z) &= \sum_{i=1}^{s_z} d_{z,i} \int_{[0,1]^r} \prod_{j=1}^r \phi_{z_j}(t_j) \mathbb{I}_{S_{z,i,1} \times \cdots \times S_{z,i,r}}(t)\mathrm{d}t \\
&= \sum_{i=1}^{s_z} d_{z,i} \int_{[0,1]^r} \prod_{m=1}^q \prod_{\substack{j=1 \\ z_j = m}}^r \phi_m(t_j) \mathbb{I}_{S_{z,i,1} \times \cdots \times S_{z,i,r}}(t)\mathrm{d}t,
\end{aligned}$$

and

$$\mathcal{E}_{\mathrm{x}}^z(\hat{D}^z) = \sum_{i=1}^{s_z} d_{z,i} \frac{1}{k^r} \prod_{m=1}^q \prod_{\substack{j=1 \\ z_j = m}}^r \sum_{n \colon U_n \in S_{z,i,j}} x_{n,m}.$$

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 $1 \le i \le s_z$. Then

$$\prod_{j=1}^r \left[\int_0^1 \phi_{z_j}(t_j) \mathbb{I}_{S_{z,i,j}}(t_j)\mathrm{d}t_j \right] = B^i(z) + \sum_{j=1}^r \left[\int_0^1 \phi_{z_j}(t_j) \mathbb{I}_{S_{z,i,j}}(t_j)\mathrm{d}t_j - b_{z,i,j} \right] B^{i,j}(z) + \Delta$$

$$= (1-r)B^i(z) + \sum_{m=1}^{q} \int_0^1 \phi_m(t) \left[\sum_{j=1, z_j=m}^{r} \mathbb{I}_{S_{z,i,j}}(t) B^{i,j}(z) \right] \mathrm{d}t + \Delta,$$

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

$$\prod_{j=1}^{r} \left[\frac{1}{k} \sum_{n \colon U_n \in S_{z,i,j}} x_{n,z_j} - b_{z,i,j} + b_{z,i,j} \right]$$
$$= (1-r)B^i(z) + \sum_{m=1}^{q} \sum_{n=1}^{k} \frac{1}{k} x_{n,m} \left[\sum_{j=1, z_j=m}^{r} \mathbb{I}_{S_{z,i,j}}(U_n) B^{i,j}(z) \right] + \Delta',$$

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 $l_1, \ldots, l_q$. 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 $\mathcal{E}_{\mathrm{x}}(\mathbb{H}(k,D))$ 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

$$\sum_{j=1, z_j=m}^{r} \mathbb{I}_{S_{z,i,j}}(t) B^{i,j}(z)$$

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

$$\hat{\phi}_m(t) = \begin{cases} \phi_m(t) + \mathrm{sgn}(l_m(t)) & \text{if } t \in T \\ \phi_m(t) & \text{otherwise.} \end{cases}$$

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

$$\int_T \sum_{m=1}^{q} |l_m(t)| \mathrm{d}t = \int_0^1 \sum_{m=1}^{q} (\hat{\phi}_m(t) - \phi_m(t)) l_m(t) \mathrm{d}t$$

$$
\begin{aligned}
&\leq \left| \int_0^1 \sum_{m=1}^q \hat{\phi}_m(t) l_m(t)\mathrm{d}t - \mathcal{E}_{\hat{\phi}}(D) \right| + |\mathcal{E}_{\hat{\phi}}(D) - \mathcal{E}_{\phi}(D)| \\
&\qquad + \left| \int_0^1 \sum_{m=1}^q \phi_m(t) l_m(t)\mathrm{d}t - \mathcal{E}_{\phi}(D) \right| \\
&\leq \frac{5}{2^{r+3}} \varepsilon \|W\|_\infty + |\mathcal{E}_{\hat{\phi}}(D) - \mathcal{E}_{\phi}(D)|.
\end{aligned}
$$

We have to estimate the last term of the above expression.

$$
\begin{aligned}
|\mathcal{E}_{\hat{\phi}}(D) - \mathcal{E}_{\phi}(D)| &\leq \sum_{z\in[q]^r} \left| \int_{[0,1]^r} \left(\prod_{j=1}^r \phi_{z_j}(t_j) - \prod_{j=1}^r \hat{\phi}_{z_j}(t_j) \right) D^z(t)\mathrm{d}t \right| \\
&\leq 2\|W\|_\infty \sum_{z\in[q]^r} \int_{[0,1]^r} \sum_{j=1}^r \left| \prod_{i<j}^r \phi_{z_i}(t_i) \prod_{i>j}^r \hat{\phi}_{z_i}(t_i)(\hat{\phi}_{z_j}(t_j) - \phi_{z_j}(t_j)) \right| \mathrm{d}t \\
&\leq 2\|W\|_\infty 2^r q^{r-1} r \sum_{m=1}^q \|\phi_m - \hat{\phi}_m\|_1 \leq 2\|W\|_\infty 2^r q^r r\eta.
\end{aligned}
$$

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 $\sum_{m=1}^q |l_m(t)| \leq \left(\frac{5}{2^{r+3}} + \frac{r}{2^3} \right) \frac{\varepsilon}{\eta} \|W\|_\infty \leq (10 + 2^{r+1} r)\, q^r \|W\|_\infty$. These former bounds indicate

$$
\begin{aligned}
\int_0^1 \sum_{m=1}^q l_m^2(t)\mathrm{d}t &= \int_{[0,1]\setminus T} \sum_{m=1}^q l_m^2(t)\mathrm{d}t + \int_T \sum_{m=1}^q l_m^2(t)\mathrm{d}t \\
&\leq 2^{2r+8} r^2 q^{2r} \|W\|_\infty^2 + \|l\|_\infty \int_T \sum_{m=1}^q |l_m(t)|\mathrm{d}t \\
&\leq 2^{2r+8} r^2 q^{2r} \|W\|_\infty^2 + (2^{r+4} r q^r)(8 q^{2r} r) \|W\|_\infty^2 \\
&\leq 2^{2r+9} r^2 q^{3r} \|W\|_\infty^2.
\end{aligned}
$$

□

We return to the proof of the main theorem, and set $\eta = \frac{\varepsilon}{16 q^r 2^r}$. For each $b \in \mathcal{A}$ we apply Theorem 4.13, 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 $l_0, l_1, \dots,$ 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

$$\begin{array}{lll} \text{maximize} & l_0 + \sum_{n=1}^{k}\sum_{m=1}^{q} \frac{1}{k} x_{n,m} l_m(U_n) & \\ \text{subject to} & \mathrm{x} \in I'(b,\eta) & \\ & 0 \le x_{n,m} \le 1 & \text{for } n = 1,\dots,k \text{ and } m = 1,\dots,q \\ & \sum_{m=1}^{q} x_{n,m} = 1 & \text{for } m = 1,\dots,q \end{array}$$

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 Theorem 4.12 with $\delta = \eta$, $\sigma^2 = 2^{2r+9} r^2 q^{3r}\|W\|_\infty^2$, $d = \frac{8q^{2r}}{\varepsilon} r\|W\|_\infty$, and $\gamma = \frac{\sigma^2}{d^2}$, and attain that the probability that $E_2(b)$ fails is at most

$$\begin{aligned} &\exp\left(-\frac{k\eta^2}{2}\right) + \exp\left(-k\gamma\left((1+\frac{\varepsilon\|W\|_\infty}{\gamma d})\ln(1+\frac{\varepsilon\|W\|_\infty}{\gamma d}) - \frac{\varepsilon\|W\|_\infty}{\gamma d}\right)\right) \\ &\le \exp\left(-\frac{k\varepsilon^2}{2^8 q^{2r} 2^{2r}}\right) + \exp\left(-k\varepsilon^2 2^{2r+3} q^{-r}\left(\frac{1}{2^{4r+15} q^r r^2}\right)\right) \\ &= \exp\left(-\frac{k\varepsilon^2}{2^{2r+8} q^{2r}}\right) + \exp\left(-\frac{k\varepsilon^2}{2^{2r+12} q^{2r} r^2}\right) \le 2\exp\left(-\frac{k\varepsilon^2}{2^{2r+12} q^{2r} r^2}\right), \end{aligned}$$

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

$$\mathbb{P}(E_2) \ge 1 - 2\left(\frac{2^{r+3} q^r}{\varepsilon}\right)^{2^6 r q^{3r} \frac{1}{\varepsilon^2}} \exp\left(-\frac{k\varepsilon^2}{2^{2r+12} q^{2r} r^2}\right)$$

$$\geq 1 - 2\exp\left(\log\left(\frac{2^{r+3}q^r}{\varepsilon}\right)2^6 rq^{3r}\varepsilon^{-2} - \log(\frac{2^{r+7}q^r r}{\varepsilon})2^{2r+16}r^2q^{3r}\varepsilon^{-2}\right)$$
$$\geq 1 - 2\exp\left(-\log(\frac{2^{r+7}q^r r}{\varepsilon})2^{2r+15}r^2q^{3r}\varepsilon^{-2}\right)$$
$$\geq 1 - \varepsilon/4.$$

Therefore for $k \geq \left(\frac{2^{r+7}q^r r}{\varepsilon}\right)^4 \log(\frac{2^{r+7}q^r r}{\varepsilon})q^r$ we have that $\mathbb{P}(E_1 \cap E_2) \geq 1 - \varepsilon/2$. 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

$$\mathcal{E}_\phi(W) \geq \mathcal{E}_{\mathrm{x}}(\mathbb{H}(k, W)) - \varepsilon\|W\|_\infty.$$

This is what we wanted to show.

□

We can improve on the tail probability bound in Theorem 4.4 significantly by a constant factor strengthening of the lower threshold condition imposed on the sample size.

Corollary 4.14. *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*

$$\mathbb{P}(|\mathcal{E}(W) - \hat{\mathcal{E}}(\mathbb{G}(k, W))| > \varepsilon\|W\|_\infty) < 2\exp\left(-\frac{\varepsilon^2 k}{8r^2}\right). \tag{4.17}$$

Proof. For $k \geq \Theta^4 \log(\Theta)q^r$ we appeal to Theorem 4.4, 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 4.4 the Azuma-Hoeffding inequality can be applied, thus

$$\begin{aligned}\mathbb{P}(|\mathcal{E}(W) - \hat{\mathcal{E}}(\mathbb{G}(k, W))| > \varepsilon\|W\|_\infty) &\leq \mathbb{P}(|\mathbb{E}\hat{\mathcal{E}}(\mathbb{G}(k, W)) - \hat{\mathcal{E}}(\mathbb{G}(k, W))| > \varepsilon/2\|W\|_\infty) \\ &\leq 2\exp\left(-\frac{\varepsilon^2 k}{8r^2}\right).\end{aligned}$$

□

Remark 4.15. 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 $2^6 rq^{2r}\frac{1}{\varepsilon^2}$, gaining a factor of q^r. Therefore in this case the statement of Theorem 4.14 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 4.16. 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

$$f(G) = \max_{\mathcal{T}} g((t(F,(G,\mathcal{T})))_{F \in \mathcal{M}_{q,m_0}}), \tag{4.18}$$

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 (4.18) 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 4.4.

5 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 4.4 to some specific quadratic programming problems.

5.1 Microcanonical version

Next we will state and prove the microcanonical version of Theorem 4.4, that is the continuous generalization of the main result of [13] for an arbitrary number q of the states. To be able to do this, we require the microcanonical analog of Theorem 4.7, that will be a generalization of Theorem 5.5 from [11] 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 5.1. *Let for $\mathbf{a} = (a_1,\dots,a_q) \in \mathrm{Pd}_q$ (that is, $a_i \geq 0$ for each $i \in [q]$ and $\sum_i a_i = 1$) denote*

$$\Omega_{\mathbf{a}} = \left\{ \phi \textit{ fractional } q\textit{-partition of } [0,1] \mid \int_0^1 \phi_i(t)\mathrm{d}t = a_i \textit{ for } i \in [q] \right\},$$

$$\omega_{\mathbf{a}} = \left\{ \mathrm{x} \textit{ fractional } q\textit{-partition of } V(G) \mid \frac{1}{|V(G)|} \sum_{u \in V(G)} x_{u,i} = a_i \textit{ for } i \in [q] \right\},$$

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 **a***-partitions and fractional* **a***-partitions, respectively.*

We call the following expressions microcanonical ground state energies with respect to **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 **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 **a**-partition of $[0, 1]$, that we will receive an fractional **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 [11], we slightly generalize that approach.

Lemma 5.2. *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 **a**-partition ϕ a fractional **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 **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.

$$|\mathcal{E}_{\phi}(W) - \mathcal{E}_{\phi'}(W)| \leq \sum_{z \in [q]^r} \left| \int_{[0,1]^r} \phi_{z_1}(x_1) \dots \phi_{z_r}(x_r) - \phi'_{z_1}(x_1) \dots \phi'_{z_r}(x_r) \mathrm{d}\lambda(x) \right| \|W\|_{\infty}$$

$$\leq \sum_{z\in[q]^r}\sum_{m=1}^{r}\left|\int\limits_{[0,1]^r}(\phi_{z_m}(x_m)-\phi'_{z_m}(x_m))\prod_{j<m}\phi_{z_j}(x_j)\prod_{j>m}\phi'_{z_j}(x_j)\mathrm{d}\lambda(x)\right|\|W\|_\infty$$

$$=\sum_{z\in[q]^r}\sum_{m=1}^{r}\int\limits_{[0,1]}\left|\phi_{z_m}(x_m)-\phi'_{z_m}(x_m)\right|\mathrm{d}x_m\prod_{j<m}a_{z_j}\prod_{j>m}b_{z_j}\|W\|_\infty$$

$$=\sum_{m=1}^{r}\sum_{j=1}^{q}\int\limits_{[0,1]}\left|\phi_j(t)-\phi'_j(t)\right|\mathrm{d}t\left(\sum_{j=1}^{q}a_j\right)^{m-1}\left(\sum_{j=1}^{q}b_j\right)^{r-m-1}\|W\|_\infty$$

$$=r\|\mathbf{a}-\mathbf{b}\|_1\|W\|_\infty.$$

The same way we can find for any fractional $\mathbf{b}$-partition ϕ an fractional $\mathbf{a}$-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 [11], the proof proceeds similar to the one concerning the graph case that was dealt with in [11].

Lemma 5.3. *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 Theorem 5.2. 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 $\mathbf{a}$-partition x. A node i from $[n]$ is called bad in a fractional partition x, if at least two elements of $\{x_{i,1}, \dots, x_{i,q}\}$ 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 $-t_1 \le t \le t_2$ 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 **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.

$$\mathcal{E}_{\mathrm{x}^t}(G) = \mathcal{E}_{\mathrm{x}}(G) + c_1 t + \dots + c_r t^r,$$

where for $l \in [r]$ we have

$$c_l = \frac{1}{k^r} \sum_{z\in[q]^r} \sum_{\substack{u_1,\dots,u_l \in S \\ u_{l+1},\dots,u_r \in V\setminus S \\ \pi}} \beta_{u_1,z_{\pi(1)}} \dots \beta_{u_l,z_{\pi(l)}} x_{u_{l+1},z_{\pi(l+1)}} \dots x_{u_r,z_{\pi(r)}} G^z(u_{\pi(1)},\dots,u_{\pi(r)}),$$

where the second sum runs over permutations π of $[k]$ that preserves the ordering of the elements of $\{1,\dots,l\}$ and $\{l+1,\dots,r\}$ at the same time. We deform the entries corresponding to S through t in the direction so that $c_1 t \ge 0$ 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 $0 \le x_{i,j} + t\beta_{i,j} \le 1$, which implies that for $t\beta_{i,j} \le 0$ we have $|t\beta_{i,j}| \le x_{i,j}$. On the other hand, $\sum_j t\beta_{i,j} = 0$ for any t and i. Therefore $\sum_j |t\beta_{i,j}| = 2\sum_j |t\beta_{i,j}| \mathbb{I}_{\{t\beta_{i,j}\le 0\}} \le 2\sum_j x_{i,j} = 2$ for any $i \in [k]$. This simple fact enables us to upper bound the absolute value of the terms $c_l t^l$.

$$\begin{aligned}
|c_l t^l| &\le \frac{(k-q-1)^{r-l}}{k^r} \|G\|_\infty \sum_{z\in[q]^r} \sum_{\substack{u_1,\dots,u_l\in S \\ \pi}} |t\beta_{u_1,z_{\pi(1)}}| \dots |t\beta_{u_l,z_{\pi(l)}}| \\
&= \frac{(k-q-1)^{r-l}}{k^r} \|G\|_\infty \binom{r}{l} q^{r-l} \sum_{z\in[q]^l} \sum_{u_1,\dots,u_l\in S} |t\beta_{u_1,z_1}| \dots |t\beta_{u_l,z_l}| \\
&\le \frac{1}{k^l} \|G\|_\infty \binom{r}{l} q^{r-l} \left(\sum_{u\in S, j\in[q]} |t\beta_{u,j}| \right)^l \le \frac{1}{k^l} \|G\|_\infty \binom{r}{l} q^{r-l} (2q+2)^l.
\end{aligned}$$

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

$$\sum_{l=2}^{r} |c_l t^l| \le \frac{1}{k^2} \|G\|_\infty (3q+2)^r.$$

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 $m \le q$. For $u_i \in B$ we set $x''_{u_i,j} = \mathbb{I}_i(j)$, for the rest of the nodes we set $\mathrm{x}'' = \mathrm{x}'$, obtaining an integer **a**-partition of $[k]$. Finally, we estimate the cost of this operation. We get that

$$\mathcal{E}_{\mathrm{x}''}(G) \ge \mathcal{E}_{\mathrm{x}'}(G) - \frac{1}{k^r}\|G\|_\infty |B| k^{r-1} q^r.$$

The original fractional **a**-partition was arbitrary, therefore it follows that

$$\mathcal{E}_{\mathbf{a}}(G) - \hat{\mathcal{E}}_{\mathbf{a}}(G) \le \frac{1}{k}\|G\|_\infty 5^r q^{r+1}.$$

□

We are ready state the adaptation of Theorem 4.4 adapted to the microcanonical setting.

Theorem 5.4. *Let $r \ge 1$, $q \ge 1$, $\mathbf{a} \in \mathrm{Pd}_q$, and $\varepsilon > 0$. Then for any $[q]^r$-tuple of $([-d, d, r])$-graphons $W = (W^z)_{z \in [q]^r}$ and $k \ge \Theta^4 \log(\Theta) q^r$ 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 $k \ge \Theta^4 \log(\Theta) q^r$ 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 Theorem 5.3 and only introduce an initial error at most $\frac{1}{k}\|G\|_\infty 5^r q^{r+1} \le \frac{\varepsilon}{2}\|W\|_\infty$.

The lower bound on $\mathcal{E}_{\mathbf{a}}(\mathbb{G}(k, W))$ is the result of standard sampling argument combined with Theorem 5.2. 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

$$\mathbb{P}\left(\left| \frac{\sum_{n=1}^k y_{n,m}}{k} - a_m \right| \ge \varepsilon \right) \le 2\exp(-\varepsilon^2 k/2),$$

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 Theorem 5.2 to conclude

$$\mathbb{E}\mathcal{E}_{\mathbf{a}}(\mathbb{G}(k, W)) \ge \mathbb{E}\mathcal{E}_y(\mathbb{G}(k, W)) - (\varepsilon/2)\|W\|_\infty$$

$$
\begin{aligned}
&= \mathbb{E}\frac{1}{k^r}\sum_{z\in[q]^r}\sum_{n_1,\ldots,n_r=1}^{k} W(U_{n_1},\ldots,U_{n_r})\prod_{j=1}^{r} y_{n_j,z_j} - (\varepsilon/2)\|W\|_\infty \\
&\ge \frac{k!}{k^r(k-r)!}\sum_{z\in[q]^r}\int\limits_{[0,1]^r} W(t_1,\ldots,t_r)\prod_{j=1}^{r}\phi_{z_j}(t_j)\mathrm{d}t - \left(\frac{r^2}{k}+\varepsilon/2\right)\|W\|_\infty \\
&\ge \mathcal{E}_\phi(W) - \left(\frac{r^2}{k}+\varepsilon/2\right)\|W\|_\infty.
\end{aligned}
$$

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 4.4.

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 4.4, 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

$$
\begin{array}{lll}
\text{maximize} & l_0 + \displaystyle\sum_{n=1}^{k}\sum_{m=1}^{q}\frac{1}{k}x_{n,m}l_m(U_n) & \\
\text{subject to} & \mathrm{x}\in I'(b,\eta)\cap\omega_{\mathbf{a}} & \\
 & 0\le x_{n,m}\le 1 & \text{for } n=1,\ldots,k \text{ and } m=1,\ldots,q \\
 & \displaystyle\sum_{m=1}^{q} x_{n,m} = 1 & \text{for } n=1,\ldots,k
\end{array}
$$

has optimal value α, then the continuous linear program

$$
\begin{array}{lll}
\text{maximize} & l_0 + \displaystyle\int_0^1\sum_{m=1}^{q} l_m(t)\phi_m(t)\mathrm{d}t & \\
\text{subject to} & \phi\in I(b,\eta)\cap\left(\displaystyle\bigcup_{\mathbf{c}:\,|a_i-c_i|\le\eta}\Omega_{\mathbf{c}}\right) & \\
 & 0\le\phi_m(t)\le 1 & \text{for } t\in[0,1] \text{ and } m=1,\ldots,q \\
 & \displaystyle\sum_{m=1}^{q}\phi_m(t)=1 & \text{for } t\in[0,1]
\end{array}
$$

has optimal value at least $\alpha-\frac{\varepsilon}{2}\|W\|_\infty$.

Recall that $\eta=\frac{\varepsilon}{16q^r2^r}$. It follows by applying Theorem 4.12 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 4.4, and $E_3 = \cap_{b\in\mathcal{A}} E_3(b)$ we conclude that

$$\mathcal{E}_{\mathbf{a}}(\mathbb{G}(k,W)) \le \max_{\mathbf{c}\colon |a_i-c_i|\le\eta} \mathcal{E}_{\mathbf{c}}(W)+\varepsilon/2)\|W\|_\infty \le \mathcal{E}_{\mathbf{a}}(W)+(rq\eta+\varepsilon/2)\|W\|_\infty \le \mathcal{E}_{\mathbf{a}}(W)+\varepsilon\|W\|_\infty.$$

Also, like in Theorem 4.4, the probability of the required events to happen simultaneously is at least $1-\varepsilon/2$. This concludes the proof. $\square$

5.2 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 [6]. 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

$$\hat{\mathrm{Q}}_\rho(W,J) = \int\limits_{[0,1]^2} J(x,y)W(\rho(x),\rho(y))\mathrm{d}x\mathrm{d}y, \qquad \hat{\mathrm{Q}}(W,J) = \max_\rho \hat{\mathrm{Q}}_\rho(W,J),$$

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

$$\mathrm{Q}_\mu(W,J) = \int\limits_{[0,1]^2}\int\limits_{[0,1]^2} J(\alpha,\beta)W(x,y)\mathrm{d}\mu(\alpha,x)\mathrm{d}\mu(\beta,y)\mathrm{d}\alpha\mathrm{d}\beta,$$

and

$$\mathrm{Q}(W,J) = \max_\mu \mathrm{Q}_\mu(W,J),$$

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 $\mathrm{Q}_\rho(W,J) = \mathrm{Q}_\mu(W,J)$.

An r-dimensional generalization of the problem for J and $W\colon [0,1]^r \to \mathbb{R}$ is

$$\mathrm{Q}(W,J) = \max_\mu \int\limits_{[0,1]^r}\int\limits_{[0,1]^r} J(\alpha_1,\dots,\alpha_r)W(x_1,\dots,x_r)\mathrm{d}\mu(\alpha_1,x_1)\dots\mathrm{d}\mu(\alpha_r,x_r)\mathrm{d}\alpha_1\dots\mathrm{d}\alpha_r,$$

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

$$\mathrm{AC}(G) = \frac{1}{n^2}\max_\rho \sum_{i,j=1}^n G_{i,j}\mathbb{I}(\rho(i) < \rho(j)),$$

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

$$\hat{\mathrm{AC}}(W) = \sup_\phi \int\limits_{[0,1^2]} \mathbb{I}(\phi(x) > \phi(y))W(x,y)\mathrm{d}x\mathrm{d}y$$

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 paper. 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 5.5. *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 ([16]), **??**, 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 4.4 to the clustered QAP.

Lemma 5.6. *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

$$|\mathrm{Q}(W, J) - \mathrm{Q}(W, J')|$$

$$
\begin{aligned}
&\leq \max_{\mu} |\mathrm{Q}_{\mu}(W,J) - \mathrm{Q}_{\mu}(W,J')| \\
&= \max_{\mu} \left| \int\limits_{[0,1]^r} (J-J')(\alpha_1,\ldots,\alpha_r) \int\limits_{[0,1]^r} W(x)\mathrm{d}\mu(\alpha_1,x_1)\ldots\mathrm{d}\mu(\alpha_r,x_r)\mathrm{d}\alpha_1\ldots\mathrm{d}\alpha_r \right| \\
&\leq \max_{\mu} \int\limits_{[0,1]^r} |(J-J')(\alpha_1,\ldots,\alpha_r)| \|W\|_\infty \mathrm{d}\alpha_1\ldots\mathrm{d}\alpha_r \\
&= \|J-J'\|_1 \|W\|_\infty \leq \varepsilon \|W\|_\infty.
\end{aligned}
$$

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 $U_1,\ldots,U_k$, $\mathbb{G}'(k,J)$ and $\mathbb{G}'(k,J')$ by the distinct independent sample $Y_1,\ldots,Y_k$.

$$
\begin{aligned}
&|\mathrm{Q}(\mathbb{G}(k,W),\mathbb{G}(k,J)) - \mathrm{Q}(\mathbb{G}(k,W),\mathbb{G}(k,J'))| \\
&\quad \leq \max_{\rho} |\mathrm{Q}_{\rho}(\mathbb{G}(k,W),\mathbb{G}(k,J)) - \mathrm{Q}_{\rho}(\mathbb{G}(k,W),\mathbb{G}(k,J'))| \\
&\quad = \max_{\rho} \frac{1}{k^r} \left| \sum_{i_1,\ldots,i_r=1}^{k} (J-J')(Y_{i_1},\ldots,Y_{i_r}) \right| \|W\|_\infty. \qquad (5.1)
\end{aligned}
$$

We analyze the random sum on the right hand side of (5.1) by first upper bounding its expectation.

$$
\begin{aligned}
&\frac{1}{k^r}\mathbb{E}_Y \left| \sum_{i_1,\ldots,i_r=1}^{k} (J-J')(Y_{i_1},\ldots,Y_{i_r}) \right| \\
&\qquad \leq \frac{r^2}{k} [\|J\|_\infty + \|J'\|_\infty] + \mathbb{E}_Y |(J-J')(Y_1,\ldots,Y_r)| \\
&\qquad = \frac{2r^2}{k} + \varepsilon \leq 2\varepsilon.
\end{aligned}
$$

By the Azuma-Hoeffding inequality the sum is also sufficiently small in probability.

$$
\mathbb{P}\left(\frac{1}{k^r} \left| \sum_{i_1,\ldots,i_r=1}^{k} (J-J')(Y_{i_1},\ldots,Y_{i_r}) \right| \geq 4\varepsilon \right) \leq 2\exp(-\varepsilon^2 k/8) \leq \varepsilon.
$$

We obtain that

$$
|\mathrm{Q}(\mathbb{G}(k,W),\mathbb{G}'(k,J)) - \mathrm{Q}(\mathbb{G}(k,W),\mathbb{G}'(k,J'))| \leq 4\varepsilon \|W\|_\infty
$$

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 $b_i = \frac{1}{k}\sum_{j=1}^{k} \mathbb{I}_{S_i}(Y_j)$, where S_i is the ith step of J' with $\lambda(S_i) = a_i$. Then we have

$$
\mathrm{Q}(\mathbb{G}(k,W),\mathbb{G}(k,J')) = \hat{\mathcal{E}}_{\mathbf{b}}(\mathbb{G}(k,W),J'').
$$

It follows again from the Azuma-Hoeffding inequality that we have $\mathbb{P}(|a_i - b_i| > \varepsilon/q) \leq 2\exp(-\frac{\varepsilon^2 k}{2q^2})$ 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

$$\begin{aligned}|\mathrm{Q}(W, J) - \mathrm{Q}(\mathbb{G}(k, W), \mathbb{G}(k, J))| &\leq |\mathrm{Q}(W, J) - \mathrm{Q}(W, J')| \\ &+ |\mathcal{E}_{\mathbf{a}}(W, J'') - \hat{\mathcal{E}}_{\mathbf{b}}(\mathbb{G}(k, W), J'')| + |\mathrm{Q}(\mathbb{G}(k, W), \mathbb{G}(k, J)) - \mathrm{Q}(\mathbb{G}(k, W), \mathbb{G}(k, J'))| \\ &\leq (5 + 2r)\varepsilon\|W\|_\infty + |\mathcal{E}_{\mathbf{a}}(W, J'') - \hat{\mathcal{E}}_{\mathbf{a}}(\mathbb{G}(k, W), J'')|.\end{aligned}$$

By the application of Theorem 5.4 the claim of the lemma is verified.

□

Next we present the application of Theorem 5.6 for two special cases of QAP.

Corollary 5.7. *The optimal values of the d-dimensional geometric QAP and the maximum acyclic subgraph problem are efficiently testable. That is, let $d \geq 1$, 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 $k \geq k_0$ and any d-dimensional geometric QAP given by the pair (G, J) we have*

$$\mathbb{P}(|\mathrm{Q}(G, J) - \mathrm{Q}(\mathbb{G}(k, G), \mathbb{G}'(k, J))| \geq \varepsilon\|G\|_\infty\|J\|_\infty) \leq \varepsilon, \tag{5.2}$$

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 Theorem 5.6, since the size of J is not fixed and depends on G.

Proof. In the light of Theorem 5.6 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.

$$\|J - J'\|_1 = \int\limits_{[0,1]^2} |J(x) - J'(x)| \mathrm{d}x = \sum_{i,j=1}^{q} \int\limits_{P_i \times P_j} |J(x) - J'(x)| \mathrm{d}x \leq \sum_{i,j}^{q} \frac{1}{q^2}\varepsilon = \varepsilon.$$

By Theorem 5.6 and Theorem 5.4 it follows that the continuous d-dimensional metric QAP is $\mathcal{O}(\log(\frac{1}{\varepsilon})\frac{1}{\varepsilon^{4rd+4}})$-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.

$$\|J - J'\|_1 = \int\limits_{[0,1]^2} |J(x) - J'(x)|\mathrm{d}x = \sum_{i=1}^{q} \int\limits_{P_i \times P_i} |J(x) - J'(x)|\mathrm{d}x \leq \varepsilon.$$

Again, by Theorem 5.6 and Theorem 5.4 it follows that the AC is $\mathcal{O}(\log(\frac{1}{\varepsilon})\frac{1}{\varepsilon^{12}})$-testable.

□

6 Further Research

Our framework based on exchangeability principles allows us to extend the notion of a limit to the case of unbounded hypergraphs and efficient testability of ground state energies in this setting. The notion of exchangeability is crucial here. The notion of efficient testability in an unbounded case could be of independent interest, perhaps the results on ground state energy carry through for the setting when the r-graphons (induced by r-graphs) are in an L^p space for some $p \geq 1$.

Another 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 $1/\varepsilon$ for the efficiently testable problems is also a worthwhile question.

Acknowledgement

We thank Jennifer Chayes, Christian Borgs and Tim Austin for a number of interesting and stimulating discussions and the relevant new ideas in the early stages of this research.

References

[1] David J. Aldous. Representations for partially exchangeable arrays of random variables. *J. Multivariate Anal.*, 11(4):581–598, 1981.

[2] 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.

[3] 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. With an appendix on the life and work of Paul Erdős.

[4] Noga Alon, W. 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.

[5] 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.

[6] Sanjeev Arora, Alan 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.

[7] Ashwini Aroskar. *Limits, Regularity and Removal for Relational and Weighted Structures*. dissertation, CMU, 2012. URL `http://repository.cmu.edu/dissertations/144`.

[8] 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.

[9] Tim Austin. On exchangeable random variables and the statistics of large graphs and hypergraphs. *Probab. Surv.*, 5:80–145, 2008.

[10] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi. Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing. *Adv. Math.*, 219(6):1801–1851, 2008.

[11] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi. Convergent sequences of dense graphs II. Multiway cuts and statistical physics. *Ann. of Math. (2)*, 176(1):151–219, 2012.

[12] B. de Finetti. *Funzione Caratteristica Di un Fenomeno Aleatorio*, pages 251–299. 6. Memorie. Academia Nazionale del Linceo, 1931.

[13] Wenceslas Fernandez de la Vega, Ravi Kannan, and Marek Karpinski. Approximation of global max-csp problems. 2006. Technical Report TR06-124.

[14] Persi Diaconis and Svante Janson. Graph limits and exchangeable random graphs. *Rend. Mat. Appl. (7)*, 28(1):33–61, 2008.

[15] Gábor Elek and Balázs Szegedy. A measure-theoretic approach to the theory of dense hypergraphs. *Adv. Math.*, 231(3-4):1731–1772, 2012.

[16] Alan M. Frieze and Ravi Kannan. Quick approximation to matrices and applications. *Combinatorica*, 19(2):175–220, 1999.

[17] Oded Goldreich, Shafi Goldwasser, and Dana Ron. Property testing and its connection to learning and approximation. *J. ACM*, 45(4):653–750, 1998.

[18] Edwin Hewitt and Leonard J. Savage. Symmetric measures on Cartesian products. *Trans. Amer. Math. Soc.*, 80:470–501, 1955.

[19] D. N. Hoover. Relations on probability spaces and arrays of random variables (preprint), 1979.

[20] Svante Janson. Poset limits and exchangeable random posets. *Combinatorica*, 31(5): 529–563, 2011.

[21] Olav Kallenberg. Symmetries on random arrays and set-indexed processes. *J. Theoret. Probab.*, 5(4):727–765, 1992.

[22] 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.

[23] László Lovász. *Large networks and graph limits*, volume 60 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2012.

[24] László Lovász and Balázs Szegedy. Limits of dense graph sequences. *J. Combin. Theory Ser. B*, 96(6):933–957, 2006.

[25] László Lovász and Balázs Szegedy. Limits of compact decorated graphs, 2010. preprint, arXiv:1010.5155.

[26] László Lovász and Balázs Szegedy. Testing properties of graphs and functions. *Israel J. Math.*, 178:113–156, 2010.

[27] 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.

[28] Igal Sason. On refined versions of the azuma-hoeffding inequality with applications in information theory, 2011. preprint, arXiv:1111.1977.

[29] 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.

[30] Yufei Zhao. Hypergraph limits: a regularity approach, 2013. preprint, arXiv:1302.1634.