AN L^p THEORY OF SPARSE GRAPH CONVERGENCE II: LD CONVERGENCE, QUOTIENTS, AND RIGHT CONVERGENCE

CHRISTIAN BORGS, JENNIFER T. CHAYES, HENRY COHN, AND YUFEI ZHAO

Abstract. We extend the L^p theory of sparse graph limits, which was introduced in a companion paper, by analyzing different notions of convergence. Under suitable restrictions on node weights, we prove the equivalence of metric convergence, quotient convergence, microcanonical ground state energy convergence, microcanonical free energy convergence, and large deviation convergence. Our theorems extend the broad applicability of dense graph convergence to all sparse graphs with unbounded average degree, while the proofs require new techniques based on uniform upper regularity. Examples to which our theory applies include stochastic block models, power law graphs, and sparse versions of W-random graphs.

Contents

1. Introduction

In the companion paper [3], we developed a theory of graph convergence for sequences of sparse graphs whose average degrees tend to infinity. These results fill a major gap in the theory of convergent graph sequences, which dealt primarily with either bounded degree graphs or dense graphs. While progress in this direction was made by Bollobás and Riordan in [2], their approach required a "bounded density" condition that excludes many graphs of interest. For example, it cannot handle graphs with heavy-tailed degree distributions such as power laws. To accommodate these and other graphs excluded by the bounded density condition, we generalized the Bollobás-Riordan approach in [3] to graphs obeying a condition we called L^p upper regularity. We then showed that when $p > 1$, every sequence of L^p upper

Zhao was supported by a Microsoft Research PhD Fellowship and internships at Microsoft Research New England.

regular graphs contains a subsequence converging to a symmetric, measurable function $W\colon [0,1]^2 \to \mathbb{R}$ that is in $L^p([0,1]^2)$. Such a function is an L^p *graphon*. Conversely, only L^p upper regular sequences can converge to L^p graphons, and so our results characterize these limits. The work of Bollobás and Riordan in [2] and the prior work on dense graph sequences amount to the special case $p = \infty$, while L^p graphons with $p < \infty$ describe limiting behaviors that occur only in the sparse setting. Thus, the L^p theory of graphons completes the previous L^∞ theory to provide a rich setting for limits of sparse graph sequences with unbounded average degree.

One attractive feature of dense graph limits is that many definitions of convergence coincide, and it is natural to ask whether the same is true for sparse graphs. After all, there are many ways to formulate the idea that two graphs are similar. For example, one could base convergence on subgraph counts or quotients. Furthermore, statistical physics provides many numerical measures for similarity, such as ground state energies or free energies.

Let us first address the question of subgraph counts. For dense graphs, the sequence $(G_n)_{n\geq 0}$ converges under the cut metric if and only if the F-density in G_n converges for all graphs F, where the F-density is the probability that a random map from F to G_n is a homomorphism [7]. One might guess that suitably normalized F-densities would characterize sparse graph convergence as well, but this fails dramatically: for sparse graphs, cut metric convergence does not determine subgraph densities (see Section 2.9 of [3]). This is not merely a technicality, but rather a fundamental fact about sparse graphs. We must therefore give up on convergence of subgraph counts as a criterion for sparse graph convergence.

By contrast, we show in this paper that several other widely studied forms of convergence are indeed equivalent to cut metric convergence in the sparse setting. Thus, with the exception of subgraph counts, the scope and consequences of sparse graph convergence are comparable with those of dense graph convergence.

We will consider several notions of convergence motivated by statistical physics and the theory of graphical models from machine learning, such as convergence of ground state energies and free energies, as well as convergence of quotients,[1] which encode "global" graph properties of interest to computer scientists, such as max-cut and min-bisection. We will also analyze the notion of large deviation (LD) convergence, which was recently introduced for graph sequences with bounded degrees [4] and can easily be adapted to our more general context. For bounded degree graphs, LD convergence was strictly stronger than convergence of quotients or other notions introduced before, but we will see that in our setting it is equivalent to these other forms of convergence.

All these question can be studied for L^p upper regular sequences of sparse graphs, but they can also be studied directly for L^p graphons. While the former might be more interesting from the point of view of applications, the latter turns out to be more elegant from an abstract point of view. We therefore first develop the theory for sequences of graphons, and then prove our results for sparse graph sequences.

We begin in Section 2 with motivation, definitions, and precise statements of our results, with some ancillary results stated in Section 3. We begin the proofs in Section 4 by completing the cases that do not require the notion of upper regularity. We then make use of upper regularity to deal with graphons in Section 5 and graphs

[1] Quotient convergence is also called partition convergence in some of the literature.

in Section 6. Finally, in Section 7, we show that any sequence whose quotients, microcanonical free energies, or ground state energies converge to those of a graphon must be upper regular, which completes the proofs.

Before turning to these details, though, we will explain the motivations behind the different types of convergence analyzed in this paper.

1.1. **Motivation.** When formulating a notion of convergence for growing sequences of graphs, one is immediately faced with the problem of deciding when to consider two large graphs on different numbers of vertices to be similar.

One natural approach is to compare summary statistics, such as weighted counts of homomorphisms to or from small graphs. Convergence based on these statistics is called *left convergence* if it uses homomorphisms *from* small graphs and *right convergence* if it uses homomorphisms *to* small graphs. Left convergence amounts to using subgraph counts, and as discussed in the previous section it is not a useful tool for characterizing sparse graph convergence. By contrast, right convergence is far more useful in the sparse setting. It amounts to using statistical physics models, and it encompasses quantities such as max-cut, min-bisection, etc. that are important in combinatorial optimization.

The advantage of using summary statistics is that they can easily be normalized to compare graphs on different numbers of nodes. For a more direct approach, one must find other ways to compare such graphs.

One way to deal with this is to blow up both graphs to obtain two new graphs on a common, much larger set of vertices. Conceptually, the most elegant way to do this is probably an infinite blow-up, replacing the vertex sets of both graphs by the interval $[0,1]$ and the adjacency matrices by appropriate step functions on $[0,1]^2$. Comparing the two graphs then reduces to comparing two functions on $[0,1]^2$, leading to the notion of convergence in the cut norm. A priori, this has the problem that relabeling the nodes of a graph would change its representation as a function on $[0,1]^2$, but this can be cured by defining the distance as the cut distance of "aligned" step functions, where alignments are formalized as measure preserving transformations from $[0,1] \to [0,1]$, chosen in such a way that the resulting two functions are as close to each other as possible. The resulting definition is known as *cut metric convergence*, and it was analyzed for sparse graphs in [3].

Another way to deal with the different vertex sets is to "squint your eyes" and look at whether the results are similar. More formally, one divides the vertex sets of both graphs into q blocks, and then averages the adjacency matrices over the respective blocks, leading to two $q \times q$ matrices representing the edge densities between various blocks (we call these matrices q-quotients). One might want to call two graphs similar if their q-quotients are close, but we are again faced with an alignment problem, now of a slightly different kind: different ways of dividing the vertex set of a graph into blocks produce different quotients. While some of the quotients of a graph contain useful information about the graph (for example those corresponding to Szemerédi partitions), others might not. Unfortunately, it is not a priori clear which of the q-quotients of a graph represent its properties well and which do not. We solve this problem by defining two graphs to be similar if the *sets* of their q-quotients are close, measured in the Hausdorff distance between subsets of the metric space of weighted graphs on q nodes.

The four notions of convergence describe informally above, namely left convergence, right convergence, convergence in metric, and convergence of quotients, were

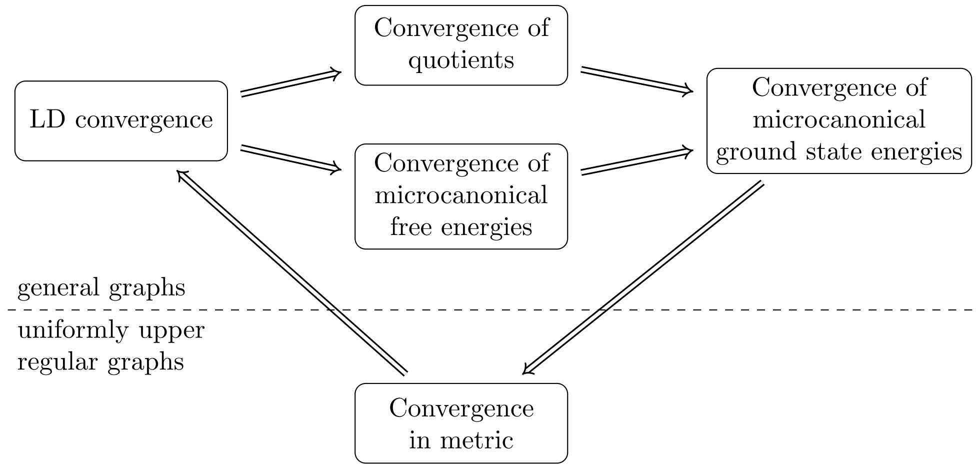

FIGURE 1. Implications between different notions of sparse graph convergence.

already introduced in [7, 8] in the context of sequences of dense graphs. But we felt it to be useful to review the motivation behind these notions, before addressing the extra complications stemming from the fact that we want to analyze sparse graphs.

In this paper we also discuss a fifth notion of convergence: large deviation convergence (LD convergence), which was recently introduced [4] to discuss convergence of bounded degree graphs. Roughly speaking, LD convergence keeps track of not just the possible quotients of a graph but also how often they occur.

Figure 1 illustrates the implications among these concepts. In the upper half of the figure, we see that LD convergence is the strongest notion and ground state energy convergence is the weakest. To complete the cycle and prove that they are all equivalent to metric convergence, we require one hypothesis, namely uniform upper regularity. This notion first arose in [3], and we review its definition below; intuitively, it ensures that subsequential limits are graphons rather than more subtle objects. Indeed, it is possible to state our results using just the fact that limits can be expressed in terms of graphons, without explicitly referring to upper regularity. We will chose this approach when stating our results in Theorem 2.10.

2. Definitions and main results

2.1. **Notation.** We begin with some notation. As usual, a weighted graph $G = (V, \alpha, \beta)$ consists of a set $V = V(G)$ of vertices, vertex weights $\alpha_x \geq 0$ for $x \in V$, and edge weights $\beta_{xy} = \beta_{yx} \in \mathbb{R}$ for $x, y \in V$. We use $E(G)$ to denote the set of edges of G, i.e., the set of pairs $\{x, y\}$ such that $\beta_{xy} \neq 0$. If we consider several graphs at the same time, then we make the dependence on G explicit, denoting the edge weights by $\beta_{xy}(G)$ and the vertex weights by $\alpha_x(G)$. The maximal node weight of G will be denoted by

$$\alpha_{\max}(G) = \max_{x \in V(G)} \alpha_x(G),$$

and the total node weight of G will be denoted by

$$\alpha_G = \sum_{x \in V(G)} \alpha_x(G).$$

We will always assume that α_G is strictly positive. If U is a subset of $V(G)$, we will use $\alpha_U(G)$ to denote the total weight of U, i.e., $\alpha_U(G) = \sum_{x\in U}\alpha_x(G)$. We say a sequence $(G_n)_{n\geq 0}$ of graphs has *no dominant nodes* if $\alpha_{\max}(G_n)/\alpha_{G_n} \to 0$ as $n\to\infty$. Finally, if $c \in \mathbb{R}$, we often use cG to denote the weighted graph with vertex weights identical to those of G and edge weights $\beta_{xy}(cG) = c\beta_{xy}(G)$. If G is a simple graph with edge set $E(G)$, we often identify it with the weighted graph with vertex weights 1 and edge weights $\beta_{xy} = \mathbf{1}_{xy\in E(G)}$. In this case, α_G is just the number of vertices in G, and $(\beta_{xy}(G))_{x,y\in V(G)}$ is the adjacency matrix. As usual, we use $[n]$ to denote the set $[n] = \{1,\dots,n\}$ and $\mathbb{N}$ to denote the set of positive integers. Finally, we define the density of a weighted graph G to be

$$\|G\|_1 = \sum_{x,y\in V(G)} \frac{\alpha_x(G)\alpha_y(G)}{\alpha_G^2}|\beta_{xy}(G)|.$$

Note that for an unweighted graph without self-loops, $\|G\|_1$ is just the edge density $2|E(G)|/|V(G)|^2$.

2.2. **Convergence in metric.** One of the main topics studied in [3] is the conditions under which a sequence of sparse graphs contains a subsequence that converges in metric. This question led us to the notion of L^p upper regularity, and more generally uniform upper regularity. Upper regularity plays an important role in the proofs in the present paper, but it is not essential for stating our main results. We therefore defer the discussion of uniform upper regularity to Section 2.7 and restrict ourselves here to just defining convergence in metric. For examples, see Section 3.3.

As already discussed, it is convenient to define this distance by embedding the space of graphs into the set of functions from $[0,1]^2$ into the reals.

Definition 2.1. An L^p *graphon* is a measurable, symmetric function $W\colon [0,1]^2 \to \mathbb{R}$ such that

$$\|W\|_p := \left(\int |W(x,y)|^p\,dx\,dy\right)^{1/p} < \infty.$$

Here symmetry means $W(x,y) = W(y,x)$ for all $(x,y)\in[0,1]^2$. If we do not specify p, we assume that W is in L^1 and call it simply a graphon, rather than an L^1 graphon.

On the set of graphons, one defines the cut norm $\|\cdot\|_\square$ by

$$\|W\|_\square = \sup_{S,T\subseteq[0,1]}\left|\int_{S\times T} W(x,y)\,dx\,dy\right|, \tag{2.1}$$

where the supremum is over measurable sets $S,T\subseteq[0,1]$; this notion goes back to the classic paper of Frieze and Kannan [10] on the "weak regularity" lemma. One then defines the *cut distance* between two graphons U and W by

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

where the infimum is over all invertible maps $\phi\colon[0,1]\to[0,1]$ such that both ϕ and its inverse are measure preserving, and W^ϕ is defined by $W^\phi(x,y) = W(\phi(x),\phi(y))$ (see [6, 12, 7]); such a map ϕ is called a *measure-preserving bijection.* After identifying graphons with cut distance zero, the space of graphons equipped with the metric $\delta_\square$ becomes a metric space.

To define the cut distance between two weighted graphs, we assign a graphon W^G to a weighted graph G as follows: let $n = |V(G)|$, identify $V(G)$ with $[n]$, and let $I_1, \dots, I_n$ be consecutive intervals in $[0,1]$ of lengths $\alpha_1(G)/\alpha_G, \dots, \alpha_n(G)/\alpha_G$, respectively. We then define W^G to be the step function that is constant on sets of the form $I_u \times I_v$ with

$$W^G(x,y) = \beta_{uv}(G) \qquad \text{if} \qquad (x,y) \in I_u \times I_v. \tag{2.2}$$

Informally, we consider the adjacency matrix of G and replace each entry (u,v) by a square of size $\alpha_u(G)\alpha_v(G)/\alpha_G^2$ with the constant function β_{uv} on this square.

With this definition, one easily checks that the density of a weighted graph G can be expressed as $\|G\|_1 = \|W^G\|_1$. For dense graphs, one can define a distance $\delta_\square(G,G')$ between two graphs by just considering the cut distance between W^G and $W^{G'}$. But for sparse graphs, the inequality

$$\delta_\square(W^G, W^{G'}) \le \|W^G - W^{G'}\|_\square \le \|W^G\|_1 + \|W^{G'}\|_1$$

means the cut distance is not very informative, since under this metric all sparse graph sequences are Cauchy sequences.

To overcome this problem, we identify weighted graphs whose edge weights only differ by a multiplicative factor.[2] Explicitly, we introduce the distance

$$\delta_{\square,\mathrm{norm}}(G,G') = \delta_\square\left(\frac{1}{\|G\|_1}W^G, \frac{1}{\|G'\|_1}W^{G'}\right), \tag{2.3}$$

where in the degenerate case of a graph G with $\|G\|_1 = 0$ we define $\frac{1}{\|G\|_1}W^G$ to be zero. As an example, with this definition, two random graphs $G_{n,p}$ for different p can be shown to be close in the metric $\delta_{\square,\mathrm{norm}}$, as are two random graphs with different numbers of nodes, at least as long as $pn \to \infty$ as $n \to \infty$ (see Section 3.3).

Definition 2.2 ([3])**.** Let $(G_n)_{n\ge 0}$ be a sequence of weighted graphs, and let W be a graphon. We say that $(G_n)_{n\ge 0}$ is *convergent in metric* if $(G_n)_{n\ge 0}$ is a Cauchy sequence in the metric $\delta_{\square,\mathrm{norm}}(G,G')$ defined in (2.3), and we say that G_n *converges to W in metric* if $\delta_\square\left(\frac{1}{\|G_n\|_1}W^{G_n}, W\right) \to 0$. (Again, we set $\frac{1}{\|G_n\|_1}W^{G_n} = 0$ if $\|G_n\|_1 = 0$.)

2.3. **Convergence of quotients.** The next object we define is convergence of quotients. To formalize this, consider a weighted graph G and a partition $\mathcal{P} = (V_1, \dots, V_q)$ of $V(G)$ into q parts, some of which could be empty. Equivalently, consider a map $\phi\colon V(G) \to [q]$ (related to $\mathcal{P}$ by setting $\phi(x) = i$ iff $x \in V_i$). We will define a quotient $G/\phi = G/\mathcal{P}$ as a pair $(\alpha,\beta) = (\alpha(G/\phi), \beta(G/\phi))$, where $\alpha \in \mathbb{R}^q$ is a vector encoding the total vertex weights of the classes in $\mathcal{P}$ and $\beta \in \mathbb{R}^{q\times q}$ is a matrix encoding the number of edges (weighted by their edge weights) between different classes. Explicitly,

$$\alpha_i(G/\phi) = \frac{\alpha_{V_i}(G)}{\alpha_G} \tag{2.4}$$

and

$$\beta_{ij}(G/\phi) = \frac{1}{\|G\|_1} \sum_{(u,v)\in V_i\times V_j} \frac{\alpha_u(G)}{\alpha_G}\frac{\alpha_v(G)}{\alpha_G}\beta_{uv}(G). \tag{2.5}$$

(In the degenerate case where G has no edges and $\|G\|_1 = 0$, we set $\beta(G/\phi) = 0$.) We call G/ϕ a *q-quotient* of G, and we denote the set of all q-quotients of G

[2] Of course, this slightly decreases our ability to distinguish between dense graphs.

by $\mathcal{S}_q(G)$. Note that without the normalization factor $\frac{1}{\|G\|_1}$ in (2.5), the weights $\beta_{ij}(G/\phi)$ would scale with the density of G, which means that all quotients of a sparse sequence would tend to zero. We have chosen this factor in such a way that

$$\|\beta(G/\phi)\|_1 = \sum_{i,j} |\beta_{ij}(G/\phi)| \le 1,$$

with equality if and only if G has non-negative weights and density $\|G\|_1 > 0$.

We will consider $\mathcal{S}_q(G)$ as a subset of

$$\mathcal{S}_q = \Big\{(\alpha, \beta) \in [0,1]^q \times [-1,1]^{q\times q} : \sum_{i\in[q]} \alpha_i = 1 \text{ and } \sum_{i,j\in[q]} |\beta_{ij}| \le 1\Big\}, \tag{2.6}$$

equipped with the usual ℓ_1 distance on $\mathbb{R}^{q+q^2}$,

$$d_1((\alpha,\beta),(\alpha',\beta')) = \sum_{i\in[q]} |\alpha_i - \alpha_i'| + \sum_{i,j\in[q]} |\beta_{ij} - \beta_{ij}'|, \tag{2.7}$$

which turns $\mathcal{S}_q$ into a compact metric space $(\mathcal{S}_q, d_1)$, a fact we will use repeatedly in this paper. For $\mathbf{a} \in \triangle_q$, we define the subspace

$$\mathcal{S}_{\mathbf{a}} = \{(\alpha,\beta) \in \mathcal{S}_q : \alpha = \mathbf{a}\}, \tag{2.8}$$

which is closed and hence also compact. Note that our normalizations are a little different from those in [8], in order to ensure compactness.[3]

The quotients of a graph G allow one to express many properties of interest to combinatorialists and computer scientists in a compact form. For example, the size of a maximal cut in a simple graph G,

$$\mathrm{MaxCut}(G) = \max_{W\subseteq V(G)} \sum_{(x,y)\in W\times(V(G)\setminus W)} \beta_{xy}(G),$$

can be expressed as

$$\mathrm{MaxCut}(G) = |E(G)| \max_{(\alpha,\beta)\in\mathcal{S}_2(G)} \big(\beta_{12} + \beta_{21}\big).$$

Restricting oneself to the subset of quotients $(\alpha,\beta) \in \mathcal{S}_2(G)$ such that $\alpha_1 = \alpha_2 = 1/2$, one can express quantities like min- or max-bisection, and considering $\mathcal{S}_q(G)$ for $q > 2$, one obtains weighted versions of max-cut for partitions into more than two sets.

To define convergence of quotients, we need the Hausdorff metric on subsets of a metric space (X, d). As usual, it is a metric d^{Hf} on the set of nonempty compact subsets of X, defined by

$$d^{\mathrm{Hf}}(S, S') = \max\left\{\sup_{x\in S} d(x, S'), \sup_{y\in S'} d(y, S)\right\},$$

where

$$d(x, S) = \inf_{y\in S} d(x, y).$$

If d is a complete metric, then so is d^{Hf} (see [13]), and the same holds for total boundedness. Thus, starting from the metric space $(\mathcal{S}_q, d_1)$, this gives a metric d_1^{Hf}

[3] Specifically, the analogue of (2.4) and (2.5) in [8] would be to use $\beta_{ij}(G/\phi)/\big(\alpha_i(G/\phi)\alpha_j(G/\phi)\big)$ instead of $\beta_{ij}(G/\phi)$, while modifying the definition of d_1 accordingly. This would encode essentially the same information, but the analogue of $\mathcal{S}_q$ would not be compact.

on the space of nonempty compact subsets of $\mathcal{S}_q$, and this space inherits compactness from the compactness of $(\mathcal{S}_q, d_1)$.

Definition 2.3. Let $(G_n)_{n\geq 0}$ be a sequence of weighted graphs. We say that the sequence $(G_n)_{n\geq 0}$ has *convergent quotients* if for each q, there exists a closed set $\mathcal{S}_q^\infty \subseteq \mathcal{S}_q$ such that $\mathcal{S}_q(G_n)$ converges to $\mathcal{S}_q^\infty$ in the Hausdorff metric.

Remark 2.4. Note that the closedness of the set $\mathcal{S}_q^\infty \subseteq \mathcal{S}_q$ can be assumed without loss of generality (every set has Hausdorff distance zero from its closure, which is why the Hausdorff metric is restricted to closed sets). Furthermore, because of the compactness of the Hausdorff metric, convergence of quotients is equivalent to the statement that the quotients $\mathcal{S}_q(G_n)$ form a Cauchy sequence. It is then easy to verify that the limiting set $\mathcal{S}_q^\infty$ can be expressed as[4]

$$\mathcal{S}_q^\infty = \big\{(\alpha,\beta) \in \mathcal{S}_q : d_1\big((\alpha,\beta), \mathcal{S}_q(G_n)\big) \to 0\big\}.$$

2.4. **Statistical physics and multiway cuts.** Next we define some notions motivated by concepts from statistical physics (or, for a different audience, by the concept of graphical models in machine learning).

Consider a weighted graph G. We will randomly color the vertices of G with q colors; i.e., we will consider random maps $\phi\colon V(G) \to [q]$. We allow for all possible maps, not just proper colorings, and call such a map a *spin configuration.* To make the model nontrivial, different spin configurations get different weights, based on a symmetric $q \times q$ matrix J with entries $J_{ij} \in \mathbb{R}$ called the *coupling matrix.* Given G and J, a map $\phi\colon V(G) \to [q]$ then gets an *energy*

$$E_\phi(G,J) = -\frac{1}{\|G\|_1} \sum_{u,v\in V(G)} \frac{\alpha_u(G)\alpha_v(G)}{\alpha_G^2} \beta_{uv}(G) J_{\phi(v)\phi(u)}. \tag{2.9}$$

(If G has no edges, we set this term equal to zero.) Given a vector $\mathbf{a} = (a_1, \dots, a_q)$ of nonnegative real numbers adding up to 1 (we denote the set of these vectors by $\triangle_q$), we consider configurations ϕ such that the (weighted) fraction of vertices mapped onto a particular color $i \in [q]$ is near to a_i. More precisely, we consider configurations ϕ in

$$\Omega_{\mathbf{a},\varepsilon}(G) = \left\{\phi\colon [q] \to V(G) : \left|\frac{\alpha_{\phi^{-1}(\{i\})}(G)}{\alpha_G} - a_i\right| \leq \varepsilon \text{ for all } i \in [q]\right\}.$$

On $\Omega_{\mathbf{a},\varepsilon}(G)$ we then define a probability distribution

$$\mu_{G,J}^{(\mathbf{a},\varepsilon)}(\phi) = \frac{1}{Z_{G,J}^{(\mathbf{a},\varepsilon)}} e^{-|V(G)|E_\phi(G,J)},$$

where $Z_{G,J}^{(\mathbf{a},\varepsilon)}$ is the normalization factor

$$Z_{G,J}^{(\mathbf{a},\varepsilon)} = \sum_{\phi\in\Omega_{\mathbf{a},\varepsilon}(G)} e^{-|V(G)|E_\phi(G,J)}. \tag{2.10}$$

[4]To see why, note that if $(\alpha,\beta) \in \mathcal{S}_q^\infty$, then

$$d_1((\alpha,\beta), \mathcal{S}_q(G_n)) \leq d_1^{\mathrm{Hf}}(\mathcal{S}_q^\infty, S_q(G_n)) \to 0,$$

while if $d_1((\alpha,\beta), \mathcal{S}_q(G_n)) \to 0$, then combining this limit with $d_1^{\mathrm{Hf}}(\mathcal{S}_q^\infty, S_q(G_n)) \to 0$ and the fact that $\mathcal{S}_q^\infty$ is closed shows that $(\alpha,\beta) \in \mathcal{S}_q^\infty$.

The distribution $\mu_{G,J}^{(\mathbf{a},\varepsilon)}$ is usually called the *microcanonical Gibbs distribution of the model J on G*, and $Z_{G,J}^{(\mathbf{a},\varepsilon)}$ is called the *microcanonical partition function*.

In this paper, we will not analyze the particular properties of the distribution $\mu_{G,J}^{(\mathbf{a},\varepsilon)}$, but we will be interested in the normalization factor, or more precisely its normalized logarithm

$$F_{\mathbf{a},\varepsilon}(G,J) = -\frac{1}{|V(G)|}\log Z_{G,J}^{(\mathbf{a},\varepsilon)}, \tag{2.11}$$

which is called the *microcanonical free energy*. We will also be interested in the dominant term contributing to $Z_{G,J}^{(\mathbf{a},\varepsilon)}$, or more precisely its normalized logarithm, the *microcanonical ground state energy*

$$E_{\mathbf{a},\varepsilon}(G,J) = \min_{\phi\in\Omega_{\mathbf{a},\varepsilon}(G)} E_\phi(G,J). \tag{2.12}$$

Note that the energy $E_\phi(G,J)$ has been normalized in such a way that $|E_\phi(G,J)| \le \|J\|_\infty$ (where $\|J\|_\infty = \max_{i,j\in[q]}|J_{ij}|$), and $\Omega_{\mathbf{a},\varepsilon}(G)\ne\emptyset$ as long as $\varepsilon \ge \alpha_{\max}(G)/\alpha_G$, implying that under this condition, $Z_{G,J}^{(\mathbf{a},\varepsilon)} \ge e^{-|V(G)|E_{\mathbf{a},\varepsilon}(G,J)} \ge e^{-|V(G)|\|J\|_\infty}$. Thus, for fixed J the microcanonical energies and free energies are of order one.

Definition 2.5. Let $(G_n)_{n\ge0}$ be a sequence of weighted graphs. We say that

(i) $(G_n)_{n\ge0}$ has *convergent microcanonical ground state energies* if the limit

$$E_{\mathbf{a}}(J) = \lim_{\varepsilon\to0}\limsup_{n\to\infty} E_{\mathbf{a},\varepsilon}(G,J) = \lim_{\varepsilon\to0}\liminf_{n\to\infty} E_{\mathbf{a},\varepsilon}(G,J) \tag{2.13}$$

exists for all $q\in\mathbb{N}$, $\mathbf{a}\in\triangle_q$, and symmetric $J\in\mathbb{R}^{q\times q}$, and

(ii) $(G_n)_{n\ge0}$ has *convergent microcanonical free energies* if the limit

$$F_{\mathbf{a}}(J) = \lim_{\varepsilon\to0}\limsup_{n\to\infty} F_{\mathbf{a},\varepsilon}(G,J) = \lim_{\varepsilon\to0}\liminf_{n\to\infty} F_{\mathbf{a},\varepsilon}(G,J) \tag{2.14}$$

exists for all $q\in\mathbb{N}$, $\mathbf{a}\in\triangle_q$, and symmetric $J\in\mathbb{R}^{q\times q}$.

Recall that the microcanonical ground state energy describes the largest term contributing to the microcanonical partition function $Z_{G,J}^{(\mathbf{a},\varepsilon)}$. Using the fact that this partition function contains at least one and at most $q^{|V(G)|}$ terms, we will see that a scaling argument shows that convergence of the microcanonical free energies implies convergence of the microcanonical ground state energies. On the other hand, the energy of a configuration ϕ can be expressed in terms of the quotient G/ϕ as

$$E_\phi(G,J) = -\langle\beta(G/\phi),J\rangle, \tag{2.15}$$

where

$$\langle\beta,J\rangle = \sum_{i,j}\beta_{ij}J_{ij}.$$

Using this identity, we will express the microcanonical ground state energy as a minimum over quotients, which in turn can be used to show that convergence of quotients implies convergence of the microcanonical ground state energies.

The following theorem gives a precise statement of these facts. We will restate the theorem as part of Lemma 3.2 and Theorem 3.3 and prove it in Section 4.

Theorem 2.6. *Let $q\in\mathbb{N}$ and let $(G_n)_{n\ge0}$ be a sequence of weighted graphs.*

(i) *If $\mathcal{S}_q(G_n)$ converges to a closed set $\mathcal{S}_q^\infty$ in the Hausdorff metric, then the limit* (2.13) *exists for all $\mathbf{a} \in \triangle_q$ and all symmetric $J \in \mathbb{R}^{q\times q}$ and can be expressed as*

$$E_{\mathbf{a}}(J) = - \max_{(\alpha,\beta)\in\mathcal{S}_q^\infty\cap\mathcal{S}_{\mathbf{a}}} \langle \beta, J\rangle.$$

(ii) *Let $\mathbf{a} \in \triangle_q$. If $|V(G_n)| \to \infty$ and the limit* (2.14) *exists for all symmetric $J \in \mathbb{R}^{q\times q}$, then the limit* (2.13) *exists for all such J and*

$$E_{\mathbf{a}}(J) = \lim_{\lambda\to\infty} \frac{1}{\lambda} F_{\mathbf{a}}(\lambda J).$$

Remark 2.7. Definition 2.5 differs from that given in [8] for dense graphs in that we are taking the double limit of first sending $n \to \infty$ and then sending $\varepsilon \to 0$, rather than a single limit with an n-dependent $\varepsilon = \varepsilon_n$. (In [8], ε_n was chosen to be $\alpha_{\max}(G_n)/\alpha_{G_n}$, even though all theorems involving the microcanonical free energies required the additional assumption that G_n has vertex weights one, corresponding to $\varepsilon_n = 1/|V(G_n)|$). While there is some merit to the simplicity of a single limit, here we decided to follow the spirit of the definitions from mathematical statistical physics, where the formulation of a double limit is standard; it is also more consistent with the double limits usually taken in the theory of large deviations, where an n-dependent ε usually makes no sense.

However, the two definitions are equivalent if G_n is dense with bounded edge weights and vertex weights one (this follows from Theorem 2.15 below, because such graphs are L^∞ upper regular). Thus, as far as the results of [8] are concerned, there is no difference between the two definitions.

2.5. **Large deviation convergence.** As we have seen in the last section, the quotients of a graph G provide enough information to calculate the microcanonical ground state energies (2.12), since the quotients tell us which energies $E_\phi(G, J)$ can be realized. However, to calculate the microcanonical free energies (2.11) we need to know a little more, namely how often a term with given energy appears in the sum (2.10).

This leads to the notion of large deviation convergence (LD convergence), which was first introduced in the context of bounded degree graphs [4], where it turned out to be strictly stronger than convergence of quotients. Roughly speaking, this notion codifies how often a given quotient $(\alpha, \beta) \in \mathcal{S}_q(G)$ appears in a sum of the form (2.10). Or, put differently, it specifies the probability that for a uniformly random map $\phi\colon V(G) \to [q]$, the quotient G/ϕ is approximately equal to (α, β). The precise definition is as follows:

Definition 2.8. Let $q \in \mathbb{N}$, let $(G_n)_{n\geq 0}$ be a sequence of weighted graphs, and let $\mathcal{P}_{q,G_n}$ be the probability distribution of G_n/ϕ when $\phi\colon V(G_n) \to [q]$ is chosen uniformly at random. We say that $(G_n)_{n\geq 0}$ is *q-LD convergent* if $|V(G_n)| \to \infty$ and

$$\begin{aligned}\lim_{\varepsilon\to 0} \liminf_{n\to\infty} & \frac{\log \mathcal{P}_{q,G_n}\big[d_1((\alpha,\beta), G_n/\phi) \leq \varepsilon\big]}{|V(G_n)|} \\ &= \lim_{\varepsilon\to 0} \limsup_{n\to\infty} \frac{\log \mathcal{P}_{q,G_n}\big[d_1((\alpha,\beta), G_n/\phi) \leq \varepsilon\big]}{|V(G_n)|}\end{aligned} \tag{2.16}$$

and say it is *q-LD convergent* with rate function $I_q\colon \mathcal{S}_q \to [0,\infty]$ if the above limit is equal to $-I_q((\alpha,\beta))$. We say that $(G_n)_{n\geq 0}$ is *LD convergent* if it is q-LD convergent for all $q\in\mathbb{N}$.

The following theorem states that LD convergence is at least as strong as convergence of quotients and convergence of the microcanonical free energies. We prove it in Sections 4.3 and 4.4.

Theorem 2.9. *Let $q\in\mathbb{N}$ and let $(G_n)_{n\geq 0}$ be a sequence of weighted graphs. If $(G_n)_{n\geq 0}$ is q-LD convergent with rate function I_q, then the following hold:*

(i) *The sets of quotients $\mathcal{S}_q(G_n)$ converge to the closed set*
$$\mathcal{S}_q(I_q) = \{(\alpha,\beta)\in\mathcal{S}_q : I_q((\alpha,\beta)) < \infty\}$$
in the Hausdorff metric.

(ii) *For all $\mathbf{a}\in\triangle_q$ and all symmetric $J\in\mathbb{R}^{q\times q}$, the microcanonical free energies converge to*
$$F_{\mathbf{a}}(I_q,J) = \inf_{(\alpha,\beta)\in\mathcal{S}_{\mathbf{a}}}\Big(-\langle\beta,J\rangle + I_q((\alpha,\beta))\Big) - \log q.$$

2.6. **Limiting expressions for convergent sequences of graphs.** The results stated so far, namely Theorems 2.6 and 2.9, raise the question of whether the four notions of convergence considered in these theorems are equivalent. They also raise the question of whether the limits of the quotients, microcanonical ground state energies, and free energies as well as the rate functions I_q can be expressed in terms of a limiting graphon. It turns out that the answers to these two questions are related, and that we have equivalence if we postulate convergence to a graphon $W\in L^1$.

We need some definitions. All of them rely on the notion of a *fractional partition* of $[0,1]$ into q classes (briefly, a *fractional q-partition*), which we define as a q-tuple of measurable functions $\rho_1,\dots,\rho_q\colon[0,1]\to[0,1]$ such that $\rho_1(x)+\cdots+\rho_q(x)=1$ for all $x\in[0,1]$. We denote the set of fractional q-partitions by FP_q. To each fractional partition $\rho\in\mathsf{FP}_q$, we assign a weight vector $\alpha(\rho)=(\alpha_1(\rho),\dots,\alpha_q(\rho))\in\triangle_q$ and an entropy $\mathsf{Ent}(\rho)\in[0,\log q]$ by setting
$$\alpha_i(\rho) = \int_0^1 \rho_i(x)\,dx$$
and
$$\mathsf{Ent}(\rho) = \int_0^1 \mathsf{Ent}_x(\rho)\,dx \quad\text{with}\quad \mathsf{Ent}_x(\rho) = -\sum_{i=1}^q \rho_i(x)\log\rho_i(x)$$
(with $0\log 0 = 0$). Let
$$\widehat{\mathcal{S}}_q = \Big\{(\alpha,\beta)\in[0,1]^q\times\mathbb{R}^{q\times q} : \sum_{i\in[q]}\alpha_i = 1\Big\}$$
(in comparison with the definition (2.6) of $\mathcal{S}_q$, we do not restrict β). Given a graphon W and a fractional q-partition $\rho\in\mathsf{FP}_q$, we then define the quotient W/ρ to be the pair $(\alpha,\beta)\in\widehat{\mathcal{S}}_q$ where
$$\alpha_i(W/\rho) = \alpha_i(\rho)$$
and
$$\beta_{ij}(W/\rho) = \int_{[0,1]^2}\rho_i(x)\rho_j(y)W(x,y)\,dx\,dy$$

for $i, j \in [q]$. We call W/ρ a *fractional q-quotient* of W. Let $\widehat{\mathcal{S}}_q(W)$ denote the set of all fractional q-quotients of W, and for $\mathbf{a} \in \triangle_q$, let $\widehat{\mathcal{S}}_{\mathbf{a}}(W)$ denote the set of pairs in $\widehat{\mathcal{S}}_q(W)$ whose first coordinate equals $\mathbf{a}$. It will be shown in Proposition 5.5 that $\widehat{\mathcal{S}}_q(W)$ is compact.

Next we define the microcanonical ground state energies and free energies of a graphon W. Given an integer $q \geq 1$ and a symmetric matrix $J \in \mathbb{R}^{q\times q}$, we define the *energy of a fractional partition* $\rho \in \mathsf{FP}_q$ to be

$$\mathcal{E}_\rho(W, J) = -\sum_{i,j} J_{ij} \int_{[0,1]^2} \rho_i(x)\rho_j(y)W(x, y)\, dx\, dy.$$

For $\mathbf{a} \in \triangle_{\mathbf{q}}$, the *microcanonical ground state energy* is defined as

$$\mathcal{E}_{\mathbf{a}}(W, J) = \inf_{\rho:\alpha(\rho)=\mathbf{a}} \mathcal{E}_\rho(W, J), \tag{2.17}$$

while the *microcanonical free energy* is defined as

$$\mathcal{F}_{\mathbf{a}}(W, J) = \inf_{\rho:\alpha(\rho)=\mathbf{a}} \Big(\mathcal{E}_\rho(W, J) - \mathsf{Ent}(\rho)\Big). \tag{2.18}$$

The infima in these equations are over all fractional q-partitions of $[0, 1]$ such that $\alpha(\rho) = \mathbf{a}$. Note that all these quantities are well defined because $0 \leq \mathsf{Ent}_x(\rho) \leq \log q$ and

$$|\mathcal{E}_\rho(W, J)| \leq \|J\|_\infty \|W\|_1.$$

Finally, the LD rate function $I_q(F, W)$ is defined as

$$I_q((\alpha, \beta), W) = \inf_{\rho\in\mathsf{FP}_q:W/\rho=(\alpha,\beta)} (\log q - \mathsf{Ent}(\rho)), \tag{2.19}$$

Note that $I_q((\alpha, \beta), W) \in [0, \log q]$ if $(\alpha, \beta) \in \widehat{\mathcal{S}}_q(W)$ and $I_q((\alpha, \beta), W) = \infty$ if $(\alpha, \beta) \notin \widehat{\mathcal{S}}_q(W)$.

We are now ready to state the main theorem of this paper. Recall that graphons are assumed to be L^1 (and not necessarily L^∞, as in some papers in the literature).

Theorem 2.10. *Let W be a graphon, and let $(G_n)_{n\geq 0}$ be a sequence of weighted graphs with no dominant nodes, in the sense that $\alpha_{\max}(G_n)/\alpha_{G_n} \to 0$. Then the following statements are equivalent:*

(i) *$(G_n)_{n\geq 0}$ converges to W in metric.*
(ii) *For all $q \in \mathbb{N}$, $\mathcal{S}_q(G_n) \to \widehat{\mathcal{S}}_q(W)$ in the Hausdorff metric d_1^{Hf}.*
(iii) *The microcanonical ground state energies of $(G_n)_{n\geq 0}$ converge to those of W.*

If all the vertices of G_n have weight one, then the following two statements are also equivalent to (i):

(iv) *$(G_n)_{n\geq 0}$ is LD convergent with rate function $I_q = I_q(\cdot, W)$.*
(v) *The microcanonical free energies of $(G_n)_{n\geq 0}$ converge to those of W.*

We prove this theorem in Section 6.

2.7. **Uniform upper regularity.** It is natural to ask whether one can state Theorem 2.10 without reference to the limiting graphon W. It turns out that the answer is yes, and in fact this reformulation (Theorem 2.15) will play a key role in the proof. To state this theorem, we need the notion of upper regularity, which first arose in our study of subsequential metric convergence in [3] and plays a key role both in that paper and in this one.

To define this concept, we define the L^p norm of a weighted graph G to be

$$\|G\|_p = \left(\sum_{x,y \in V(G)} \frac{\alpha_x(G)\alpha_y(G)}{\alpha_G^2} |\beta_{xy}(G)|^p \right)^{1/p},$$

and for $p = \infty$ we set

$$\|G\|_\infty = \max_{\substack{x,y \in V(G) \\ \alpha_x(G),\alpha_y(G)>0}} |\beta_{xy}(G)|.$$

As we already have seen in Section 2.2, when studying graph convergence for sparse graphs, it is natural to reweight the edge weights by $\frac{1}{\|G\|_1}$ to obtain a weighted graph which does not go to zero for trivial reasons. In order to control the now possibly large entries of the adjacency matrix of the weighted graph $\frac{1}{\|G\|_1}G$, one might want to require the L^p norm of $\frac{1}{\|G\|_1}G$ to be bounded, but this turns out to be too restrictive. Instead, we will use a weaker condition, which requires the L^p norm of $\frac{1}{\|G\|_1}G$ to be bounded "on average," at least when the averages are taken over sufficiently large blocks. To make this precise, we need some additional notation.

Given a weighted graph G and a partition $\mathcal{P} = \{V_1, \dots, V_q\}$ of $V(G)$ into disjoint sets $V_1, \dots, V_q$, we define $G_{\mathcal{P}}$ to be the weighted graph with the same vertex weights as G and edge weights which are defined by averaging over the blocks $V_i \times V_j$, suitably weighted by the vertex weights:

$$\beta_{xy}(G_{\mathcal{P}}) = \frac{1}{\alpha_{V_i}(G)\alpha_{V_j}(G)} \sum_{(u,v)\in V_i \times V_j} \alpha_u(G)\alpha_v(G)\beta_{uv}(G) \tag{2.20}$$

if $(x, y) \in V_i \times V_j$ and $\alpha_{V_i}(G)\alpha_{V_j}(G) > 0$, while we set $\beta_{xy}(G_{\mathcal{P}}) = 0$ if either x or y lie in a block $V_k(G)$ with total node weight $\alpha_{V_k}(G) = 0$.

Definition 2.11. Let G be a weighted graph, let $C, \eta > 0$, and let $p > 1$. We say that G is (C, η)-*upper L^p regular* if $\alpha_{\max}(G) \le \eta\alpha_G$ and

$$\|G_{\mathcal{P}}\|_p \le C\|G\|_1$$

for all partitions $\mathcal{P} = \{V_1, \dots, V_q\}$ for which $\min_i \alpha_{V_i} \ge \eta\alpha_G$. We say that a sequence of graphs $(G_n)_{n\ge0}$ is C-*upper L^p regular* if there exists a sequence $\eta_n \to 0$ such that G_n is (C, η_n)-upper regular, and we say that $(G_n)_{n\ge0}$ is L^p *upper regular* if there exists a $C < \infty$ such that $(G_n)_{n\ge0}$ is C-upper L^p regular.

The definition of L^1 upper regularity always holds vacuously, but the following definition of uniform upper regularity turns out to be the correct L^1 analogue, as described in Appendix C of [3]. It is closely related to the notion of uniform integrability of a set of graphons (see Section 5.2), and it is the notion we will need in this paper.

Definition 2.12. Let $\eta > 0$ and let $K\colon (0,\infty) \to (0,\infty)$ be any function. We say that a weighted graph G is (K,η)-*upper regular* if $\alpha_{\max}(G) \le \eta\alpha_G$ and

$$\sum_{x,y\in V(G)} \frac{\alpha_x(G)\alpha_y(G)}{\alpha_G^2}\frac{|\beta_{xy}(G_{\mathcal{P}})|}{\|G\|_1}\mathbf{1}_{|\beta_{xy}(G_{\mathcal{P}})|\ge K(\varepsilon)\|G\|_1} \le \varepsilon \tag{2.21}$$

for all $\varepsilon > 0$ and all partitions $\mathcal{P} = \{V_1,\dots,V_q\}$ for which $\min_i \alpha_{V_i} \ge \eta\alpha_G$. We say that a sequence of graphs $(G_n)_{n\ge0}$ is K-*upper regular* if there exists a sequence $\eta_n \to 0$ such that G_n is (K,η_n)-upper regular, and we say that $(G_n)_{n\ge0}$ is *uniformly upper regular* if there exists a function $K\colon(0,\infty)\to(0,\infty)$ such that $(G_n)_{n\ge0}$ is K-upper regular.

Note that the properties of L^p upper regularity and uniform upper regularity require $(G_n)_{n\ge0}$ to have *no dominant nodes*, a property we already encountered in Theorem 2.10. One of the main results of [3] is the following theorem.

Theorem 2.13 (Theorem C.7 in [3])**.** *Let $(G_n)_{n\ge0}$ be a uniformly upper regular sequence of weighted graphs. Then $(G_n)_{n\ge0}$ contains a subsequence that is convergent in metric. Furthermore, if $(G_n)_{n\ge0}$ is convergent in metric, then there exists a graphon W such that G_n converges to W in metric.*

Conversely, it was shown in [3] that every sequence of weighted graphs which converges in metric to a graphon and has no dominant nodes must be upper regular. The precise statement is given by the following theorem, which follows immediately from Corollary 2.11 and Proposition C.5 in [3].

Theorem 2.14 ([3])**.** *Let $(G_n)_{n\ge0}$ be a sequence of weighted graphs without dominant nodes, and assume that G_n converges to some graphon W in metric. Then $(G_n)_{n\ge0}$ is uniformly upper regular. If W is in L^p, then $(G_n)_{n\ge0}$ is L^p-upper regular.*

A uniformly upper regular sequence of simple graphs must have unbounded average degree, by Proposition C.15 in [3]. This corresponds to the fact that graphons are not the appropriate limiting objects for graphs with bounded average degree (although they apply to all other sparse graphs).

Returning to the subject of this paper, the question of whether the five versions of convergence defined in Sections 2.2 through 2.5 are equivalent, we are now ready to state our results without reference to a limiting graphon.

Theorem 2.15. *Let $(G_n)_{n\ge0}$ be a uniformly upper regular sequence of weighted graphs. Then the following three statements are equivalent:*

(i) *$(G_n)_{n\ge0}$ is convergent in metric.*
(ii) *$(G_n)_{n\ge0}$ has convergent quotients.*
(iii) *$(G_n)_{n\ge0}$ has convergent microcanonical ground state energies.*

If all the vertices of G_n have weight one, then the following two statements are also equivalent to (i):

(iv) *$(G_n)_{n\ge0}$ is LD convergent.*
(v) *$(G_n)_{n\ge0}$ has convergent microcanonical free energies.*

Note that by Theorems 2.6 and 2.9, we already know that (iv) implies both (v) and (ii), and that both (v) and (ii) imply (iii); in fact, we need neither node weights one, nor the assumption of upper regularity. So the important part of this theorem is that under the assumption of uniform upper regularity, convergence in metric implies convergence of quotients (and LD convergence, if we assume node weights one), and convergence of the microcanonical ground state energies implies convergence in metric. We prove Theorem 2.15 in Section 6.

One may want to know whether the assumption of upper regularity is actually necessary for these conclusions to hold. The answer is yes, by the following example.

Example 2.16. Let $c_n \in \mathbb{N}$ be such that $c_n \to \infty$ and $c_n/n \to 0$ as $n \to \infty$, and let G_n be the disjoint union of a complete graph on c_n nodes with $n - c_n$ isolated nodes. Then $(G_n)_{n\geq 0}$ is LD convergent (and hence has convergent quotients, microcanonical free energies and microcanonical ground state energies); see Section 3.3.6 below. However, $(G_n)_{n\geq 0}$ is not a Cauchy sequence in the normalized cut metric $\delta_{\square,\mathrm{norm}}$ from (2.3) and hence does not converge to any graphon in metric (see the proof of Proposition 2.12(a) in [3]).

The following theorem states that convergence of the quotients, microcanonical ground state energies, or microcanonical free energies to those of a graphon W all imply upper regularity, as does LD convergence with a rate function $I_q(\cdot, W)$ given in terms of a graphon W. It is the analogue of Theorem 2.14 for these notions of convergence.

Theorem 2.17. *Let $(G_n)_{n\geq 0}$ be a sequence of weighted graphs with no dominant nodes, and let W be a graphon. Then any of the following conditions implies that $(G_n)_{n\geq 0}$ is uniformly upper regular.*

(i) *The microcanonical ground state energies of $(G_n)_{n\geq 0}$ converge to those of W.*
(ii) *For all $q \in \mathbb{N}$, $\mathcal{S}_q(G_n) \to \widehat{\mathcal{S}}_q(W)$ in the Hausdorff metric d_1^{Hf}.*
(iii) *The microcanonical free energies of $(G_n)_{n\geq 0}$ converge to those of W.*
(iv) *$(G_n)_{n\geq 0}$ is LD convergent with rate function $I_q = I_q(\cdot, W)$.*

Note that the first two assertions in this theorem already follow by combining Theorem 2.14 with Theorem 2.10(i)–(iii). However, this is not how our proofs of Theorems 2.10 and 2.17 proceed. Instead of proving Theorem 2.10 directly, we use uniform upper regularity to prove Theorem 2.15 in Section 6. Then Theorem 2.17 is exactly what we need to deduce Theorem 2.10 from Theorem 2.15, and we prove Theorem 2.17 in Section 7.

3. Further definitions, remarks, and examples

3.1. **Convergence of free energies and ground state energies.** In addition to the microcanonical quantities introduced in Section 2.4, statistical physicists often analyze the unrestricted probability measure

$$\mu_{G,J,h}(\phi) = \frac{1}{Z_{G,J,h}} e^{-|V(G)|E_\phi(G,J)+|V(G)|\langle h,\alpha(G/\phi)\rangle},$$

where h is a vector in $\mathbb{R}^q$ called the *magnetic field*,

$$\langle h, \alpha \rangle = \sum_{i\in[q]} h_i \alpha_i,$$

and $Z_{G,J,h}$ is the normalization factor

$$Z_{G,J,h} = \sum_{\phi\colon V(G)\to[q]} e^{-|V(G)|E_\phi(G,J)+|V(G)|\langle h,\alpha(G/\phi)\rangle}, \tag{3.1}$$

usually called the *partition function.* The normalized logarithm of the partition function is the *free energy*

$$F(G,J,h) = -\frac{1}{|V(G)|}\log Z_{G,J,h}$$

of the model (J,h) on G, and the maximizer in the sum (3.1), or more precisely its normalized logarithm, is the *ground state energy*

$$E(G,J,h) = \min_{\phi\colon [q]\to V(G)} \Big(E_\phi(G,J,h) - \langle h, \alpha(G/\phi)\rangle\Big).$$

Definition 3.1.

(i) $(G_n)_{n\geq 0}$ has *convergent ground state energies* if the limit

$$E(J,h) = \lim_{n\to\infty} E(G_n,J,h) \tag{3.2}$$

exists for all $q\in\mathbb{N}$, symmetric $J\in\mathbb{R}^{q\times q}$, and $h\in\mathbb{R}^q$.

(ii) $(G_n)_{n\geq 0}$ has *convergent free energies* if the limit

$$F(J,h) = \lim_{n\to\infty} F(G_n,J,h) \tag{3.3}$$

exists for all $q\in\mathbb{N}$, symmetric $J\in\mathbb{R}^{q\times q}$, and $h\in\mathbb{R}^q$.

These notions are implied by the microcanonical versions, and convergence of free energies implies convergence of ground state energies. This is the content of the following lemma, which we will prove in Section 4.1. Note that part (iii) is a restatement of Theorem 2.6(ii).

Lemma 3.2. *Let $(G_n)_{n\geq 0}$ be a sequence of weighted graphs with $|V(G_n)|\to\infty$, and let $q\in\mathbb{N}$. Then the following hold:*

(i) *Let J be a symmetric matrix in $\mathbb{R}^{q\times q}$, and assume that the limit* (2.14) *exists for all $\mathbf{a}\in\triangle_q$. Then the limit* (3.3) *exists for all $h\in\mathbb{R}^q$, and*

$$F(J,h) = \inf_{\mathbf{a}\in\triangle_q}\big(F_{\mathbf{a}}(J) - \langle \mathbf{a}, h\rangle\big).$$

(ii) *Let J be a symmetric matrix in $\mathbb{R}^{q\times q}$, and assume that the limit* (2.13) *exists for all $\mathbf{a}\in\triangle_q$. Then the limit* (3.2) *exists for all $h\in\mathbb{R}^q$, and*

$$E(J,h) = \inf_{\mathbf{a}\in\triangle_q}\big(E_{\mathbf{a}}(J) - \langle \mathbf{a}, h\rangle\big).$$

(iii) *Let $\mathbf{a}\in\triangle_q$, and assume that the limit* (2.14) *exists for all symmetric $J\in\mathbb{R}^{q\times q}$. Then the limit* (2.13) *exists for all such J, and*

$$E_{\mathbf{a}}(J) = \lim_{\lambda\to\infty}\frac{1}{\lambda}F_{\mathbf{a}}(\lambda J).$$

(iv) *Assume that the limit* (3.3) *exists for all $h\in\mathbb{R}^d$ and all symmetric $J\in\mathbb{R}^{q\times q}$. Then the limit* (3.2) *exists for all $h\in\mathbb{R}^d$ and all symmetric $J\in\mathbb{R}^{q\times q}$, and*

$$E(J,h) = \lim_{\lambda\to\infty}\frac{1}{\lambda}F(\lambda J,\lambda h).$$

Convergence of the ground state and free energies is strictly weaker than that of the microcanonical versions. See Section 3.3.5 for an example.

On the other hand, we can use (2.15) to express both the microcanonical ground state energies $E_{\mathbf{a},\varepsilon}(G,J)$ and the unrestricted ground state energies $E(G,J,h)$ as minima over quotients. Using this fact, it is not hard to show that convergence of quotients implies convergence of the ground state energies as well as the microcanonical ground state energies. This is the content of the following theorem, which again holds for an arbitrary sequence, with no assumption about upper regularity. We prove the theorem (which encompasses Theorem 2.6(i)) in Section 4.2.

Theorem 3.3. *Let $q \in \mathbb{N}$ and let $(G_n)_{n\geq 0}$ be a sequence of weighted graphs such that $\mathcal{S}_q(G_n)$ converges to a closed set $\mathcal{S}_q^\infty$ in the Hausdorff metric. Then the limit* (2.13) *exists for all $\mathbf{a} \in \triangle_q$ and all symmetric $J \in \mathbb{R}^{q\times q}$ and can be expressed as*

$$E_{\mathbf{a}}(J) = - \max_{(\alpha,\beta)\in\mathcal{S}_q^\infty\cap\mathcal{S}_{\mathbf{a}}} \langle \beta, J\rangle.$$

and the limit (3.2) *exists for all symmetric $J \in \mathbb{R}^{q\times q}$ and all $h \in \mathbb{R}^q$ and can be expressed as*

$$E(J,h) = \min_{(\alpha,\beta)\in\mathcal{S}_q^\infty} \big(-\langle \beta, J\rangle - \langle \alpha, h\rangle\big),$$

Much as in Section 2.6, we can write down limiting expressions for a graphon W. The *ground state energy of the model (J,h) on W* is

$$\mathcal{E}(W,J,h) = \inf_{\rho\in\mathsf{FP}_q} \left(\mathcal{E}_\rho(W,J) - \sum_i h_i \int_{[0,1]} \rho_i(x)\,dx\right)$$

and its *free energy* is defined as

$$\mathcal{F}(W,J,h) = \inf_{\rho\in\mathsf{FP}_q} \left(\mathcal{E}_\rho(W,J) - \sum_i h_i \int_{[0,1]} \rho_i(x)\,dx - \mathsf{Ent}(\rho)\right).$$

It follows from Lemma 3.2 and Theorem 2.10 that if $(G_n)_{n\geq 0}$ has no dominant nodes and converges to W in metric, then its ground state energies converge to those of W, and if all the vertices of G_n have weight one, then the free energies also converge to those of W.

3.2. **LD convergence.**

Remark 3.4. It is not hard to see that $(G_n)_{n\geq 0}$ is q-LD convergent if and only if $\mathcal{P}_{q,G_n}$ obeys a large deviation principle with speed $|V(G_n)|$, i.e., if there exists a lower semicontinuous function $I_q \colon \mathcal{S}_q \to [0,\infty]$ such that

$$\begin{aligned}
-\inf_{(\alpha,\beta)\in\mathring{S}} I_q((\alpha,\beta)) &\leq \liminf_{n\to\infty} \frac{\log \mathcal{P}_{q,G_n}\big[G_n/\phi \in \mathring{S}\big]}{|V(G_n)|} \\
&\leq \limsup_{n\to\infty} \frac{\log \mathcal{P}_{q,G_n}\big[G_n/\phi \in \bar{S}\big]}{|V(G_n)|} \leq -\inf_{(\alpha,\beta)\in\bar{S}} I_q((\alpha,\beta))
\end{aligned} \tag{3.4}$$

for all sets $S \subseteq \mathcal{S}_q$. Here $\bar{S}$ denotes the closure of S and $\mathring{S}$ its interior.

Indeed, assume that (3.4) holds for some lower semicontinuous function $I_q \colon \mathcal{S}_q \to [0,\infty]$. By the lower semicontinuity of I_q,

$$I_q((\alpha,\beta)) = \lim_{\varepsilon\to 0} \inf\{I_q((\alpha',\beta')) : d_1((\alpha,\beta),(\alpha',\beta')) < \varepsilon\},$$

which implies (2.16) when inserted into (3.4). It turns out that (2.16) is also sufficient for (3.4) to hold. Indeed, under the assumption that the underlying metric space is compact (which is the case here), the equality of the two limits in (2.16) implies that $\mathcal{P}_{q,G_n}$ obeys a large deviation principle with rate function given by I_q; see, for example, Theorem 4.1.11 in [9] for the proof.

3.3. **Examples.** In this section we give some examples of convergent graph sequences, as well as a few counterexamples in which the equivalences in Theorem 2.15 fail (of course because uniform upper regularity does not hold).

3.3.1. *Erdős-Rényi random graphs.* The simplest example of a uniformly upper regular sequence—in fact an L^∞-upper regular sequence—is the standard Erdős-Rényi random graphs $G_{n,p}$ obtained by connecting each pair of distinct vertices in $[n]$ independently with probability p. Here p can depend on n, as long as $pn \to \infty$ as $n \to \infty$. Under this condition, $G_{n,p}$ converges with probability one to the constant graphon $W = 1$. This can proved in several ways, for example by showing that in expectation all the quotients in $\mathcal{S}_q(G_{n,p})$ converge to the corresponding quotients in $\widehat{\mathcal{S}}_q(W)$ and proving concentration with the help of Azuma's inequality.

3.3.2. *Stochastic block models.* Next we consider the block models obtained as follows. Fix $k \in \mathbb{N}$, a symmetric matrix $B = (b_{ij})_{i,j\in[k]}$ with entries $b_{ij} \ge 0$ satisfying $k^{-2}\sum_{i,j} b_{ij} = 1$, and a target density $\rho_n \le 1/\max b_{ij}$. Divide $[n]$ into k blocks $V_1, \dots, V_k$ of equal size (or, in the case where n is not divisible by k of sizes differing by at most 1) and define $p_{uv} = \rho_n b_{ij}$ if $(u,v) \in V_i \times V_j$. Then we connect vertices u and v with probability p_{uv}. If $n\rho_n \to \infty$ as $n \to \infty$, then the resulting graph converges with probability one to the step function W that is equal to b_{ij} on the block $(\frac{i-1}{n}, \frac{i}{n}] \times (\frac{j-1}{n}, \frac{j}{n}]$. The proof can again be obtained by proving that the quotients converge in expectation, followed by a concentration argument.

3.3.3. *Power law graphs.* Starting again with the vertex set $[n]$, connect $i \neq j$ with probability $\min(1, n^\beta (ij)^{-\alpha})$, where $0 < \alpha < 1$ and $0 \le \beta < 2\alpha$. In other words, the expected degree distribution follows an inverse power law with exponent α, while the n^β scaling factor ensures that the probabilities do not become too small. If $\beta > 2\alpha - 1$, then the expected number of edges is superlinear, and a similar argument to the one used in the above two examples shows convergence, this time to a graphon that is not in L^∞, namely $W(x,y) = (1-\alpha)^2 (xy)^{-\alpha}$.

3.3.4. *W-random graphs.* Our fourth example provides a construction of a sequence $(G_n)_{n\ge 0}$ of simple graphs that converge to a given graphon W with non-negative entries $W(x,y) \ge 0$. Normalizing W so that $\int_{[0,1]^2} W = 1$ and fixing a target density ρ_n, we proceed by first choosing n i.i.d. variables $x_1, \dots x_n$ uniformly in $[0,1]$, and then defining a random graph $G_n(W, \rho_n)$ on $\{1, \dots, n\}$ by connecting each pair $\{i,j\} \in \binom{[n]}{2}$ independently with probability $\min\{1, \rho_n W(x_i, y_i)\}$. Assuming that $\rho_n \to 0$ and $n\rho_n \to \infty$, the graphs G_n converge to W under the normalized cut metric with probability one, by Theorem 2.14 in [3]. If W is a step function, this is more or less equivalent to the convergence of stochastic block models, while for general graphons W, one can proceed by first approximating W by a step function.

3.3.5. *Convergence of free energies without convergence of microcanonical free energies.* Our next example is a generalization of Example 6.3 from [8] to the sparse setting, and is based on the observation that for an arbitrary sequence of graphs G_n, the free energies of G_n and a disjoint union of G_n with itself are identical (this follows from the fact that for two disjoint graphs G and G', the partition function on $G \cup G'$ factors into that of G times that of G'). If we take G_n to be equal to $G_{n,p}$ if n is odd, and equal to a disjoint union of two copies of $G_{n,p}$ if n is even, then we get convergence of the free energies. By contrast, in the notions of convergence from Theorem 2.10, the odd subsequence converges to $W = 1$, while the even one converges to the block graphon W' that is equal to 2 on $[0,1/2]^2 \cup [1/2,1]^2$ and 0 elsewhere. In particular, the min-bisection of the even subsequence converges to zero, while the min-bisection of the odd sequence converges to 1/2. This shows that the microcanonical ground state energies are not convergent, which implies that the microcanonical free energies don't converge either.

3.3.6. *LD convergence without metric convergence.* This is Example 2.16 from Section 2.7, consisting of a graph G_n that is the disjoint union of a complete graph on c_n nodes with $n - c_n$ isolated nodes. A random q-quotient is then determined by how many elements of the clique there are in each part and how many elements of the non-clique. Calling these numbers $b_1, \dots, b_q$ and $a_1, \dots, a_q$, we have $b_1 + \cdots + b_q = c_n$ and $a_1 + \cdots + a_q = n - c_n$, and this occurs with probability

$$q^{-n}\binom{c_n}{b_1, \dots, b_q}\binom{n-c_n}{a_1, \dots, a_q}.$$

Everything else is determined from this data: $\alpha_i = (a_i + b_i)/n$, $\beta_{ij} = b_i b_j/(c_n(c_n-1))$ if $i \neq j$, and $\beta_{ii} = b_i(b_i-1)/(c_n(c_n-1))$. If $c_n \in \mathbb{N}$ is such that $c_n \to \infty$ and $c_n/n \to 0$ as $n \to \infty$, then in the rate function, the choice of $b_1, \dots, b_q$ gets wiped out by the choice of $a_1, \dots, a_q$, leading to LD convergence with rate function

$$I_q((\alpha, \beta)) = \log q + \sum_{i=1}^{q} \alpha_i \log \alpha_i$$

as long as $\beta \in \mathbb{R}^{q \times q}$ satisfies $\beta_{ii} \geq 0$, $\beta_{ij} = \sqrt{\beta_{ii}\beta_{jj}}$, and $\sum_i \sqrt{\beta_{ii}} = 1$ (while $I_q((\alpha, \beta)) = \infty$ otherwise). On the other hand, $(G_n)_{n \geq 0}$ is not a Cauchy sequence in the normalized cut metric $\delta_{\square,\mathrm{norm}}$ from (2.3) and hence does not converge to any graphon in metric (see the proof of Proposition 2.12(a) in [3]).

3.3.7. *Convergence of quotients without convergence of the microcanonical free energies.* We close our example section with an example from [4] (Example 5 from that paper) which shows that without the assumption of upper regularity, convergence of quotients does not imply convergence of the microcanonical free energies, and hence does not imply LD-convergence either. Before stating this example, we note that whenever H_n is a sequence of regular bipartite graphs, $c_n \to \infty$, and G_n is the union of c_n disjoint copies of H_n, then the quotients of G_n converge to the convex hull of the quotients of a graph consisting of a single edge. To see why, consider a map from the vertex set of G_n into $[q]$. Since G_n is regular, the corresponding quotient does not change if we replace G_n by a disjoint union of $|E(G_n)|$ edges (and map each of the split vertices to the same element of $[q]$ as its original vertex in G_n). Thus, the quotient is in the convex hull of the quotients of a single edge. On the other hand, each quotient of a single edge can be realized in the bipartite graph H_n, showing that each quotient in the convex hull can be arbitrarily well approximated in G_n if $c_n \to \infty$.

To get a sequence G_n without convergent microcanonical free energies we specialize to the case where H_n consists of a 4-cycle when n is even and a 6-cycle when n is odd. The free energies of G_n are then equal to the free energies of the 4-cycle when n is even and those of the 6-cycle when n is odd. But it is easy to check that the 4-cycle has different free energies from a 6-cycle, implying that G_n does not have convergent free energies, and hence does not have convergent microcanonical free energies either (an alternative proof was given in [4], where it was used that G_n does not converge in the sense of Benjamini and Schramm [1], which in turn is necessary for convergence of the free energies, as proved in [5]).

4. Convergence without the assumption of upper regularity

In this section, we consider general sequences of weighted graphs G_n without any additional assumptions (except that G_n has at least one edge with nonzero edge weight). We will prove Lemma 3.2, Theorem 3.3, and Theorem 2.9.

4.1. Free energies and ground state energies.

In this section, we prove Lemma 3.2. We start with the proof of (i). To this end, we note that for all $\mathbf{a} \in \triangle_q$ we have the lower bound

$$Z(G_n, J, h)^{1/|V(G_n)|} \geq e^{\langle \mathbf{a}, h\rangle - \varepsilon \|h\|_1} \big(Z_{G,J}^{(\mathbf{a},\varepsilon)}\big)^{1/|V(G_n)|},$$

from which we conclude that

$$\limsup_{n\to\infty} F(G_n, J, h) \leq F_{\mathbf{a}}(J) - \langle \mathbf{a}, h\rangle.$$

Since $\mathbf{a} \in \triangle_q$ was arbitrary, this gives

$$\limsup_{n\to\infty} F(G_n, J, h) \leq \inf_{\mathbf{a}\in\triangle_q} \big(F_{\mathbf{a}}(J) - \langle \mathbf{a}, h\rangle\big).$$

To get a matching lower bound, we use the fact that $\triangle_q$ can be covered by $\lceil 1/(2\varepsilon)\rceil^q \leq \varepsilon^{-q}$ cubes of the form $\prod_{i=1}^q [a_i - \varepsilon, a_i + \varepsilon]$. Explicitly, let $\triangle_q^{(\varepsilon)}$ be the set of points $\mathbf{a}$ where each coordinate is an odd multiple of ε. Then

$$Z(G_n, J, h)^{1/|V(G_n)|} \leq \varepsilon^{-q/|V(G_n)|} \max_{\mathbf{a}\in\triangle_q^{(\varepsilon)}} e^{\langle \mathbf{a}, h\rangle + \varepsilon\|h\|_1} Z_{\mathbf{a},\varepsilon}(G_n, J)^{1/|V(G_n)|},$$

implying that

$$\begin{aligned}\liminf_{n\to\infty} F(G_n, J, h) &\geq -\varepsilon\|h\|_1 + \liminf_{n\to\infty} \min_{\mathbf{a}\in\triangle_q^{(\varepsilon)}} \big(F_{\mathbf{a},\varepsilon}(G_n, J) - \langle \mathbf{a}, h\rangle\big)\\ &= -\varepsilon\|h\|_1 + \min_{\mathbf{a}\in\triangle_q^{(\varepsilon)}} \liminf_{n\to\infty} \big(F_{\mathbf{a},\varepsilon}(G_n, J) - \langle \mathbf{a}, h\rangle\big)\end{aligned}$$

where in the second step, we used that the minimum is over a finite set. Let ε_k be a sequence going to zero, let $\mathbf{a}_k$ be the minimizer on the right hand side, and assume (by taking a subsequence, if necessary) that $\mathbf{a}_k$ converges to some $\mathbf{a}$. Let $\tilde\varepsilon_k = \varepsilon_k + \|\mathbf{a} - \mathbf{a}_k\|_\infty$. Since $F_{\mathbf{a}_k,\varepsilon_k}(G_n, J) \geq F_{\mathbf{a},\tilde\varepsilon_k}(G_n, J)$,

$$\liminf_{n\to\infty} F(G_n, J, h) \geq -\tilde\varepsilon_k\|h\|_1 + \liminf_{n\to\infty} \big(F_{\mathbf{a},\tilde\varepsilon_k}(G_n, J) - \langle \mathbf{a}, h\rangle\big).$$

Sending $k \to \infty$, we conclude that

$$\begin{aligned}\liminf_{n\to\infty} F(G_n, J, h) &\geq \lim_{\varepsilon\to 0} \liminf_{n\to\infty} \big(F_{\mathbf{a},\tilde\varepsilon}(G_n, J) - \langle \mathbf{a}, h\rangle\big)\\ &= \big(F_{\mathbf{a}}(J) - \langle \mathbf{a}, h\rangle\big) \geq \min_{\mathbf{a}\in\triangle_q} \big(F_{\mathbf{a}}(J) - \langle \mathbf{a}, h\rangle\big)\end{aligned}$$

as desired.

The proof of (ii) starts from the observations that

$$\begin{aligned}E_{\mathbf{a}',\varepsilon}(G_n, h) - \langle \mathbf{a}', h\rangle + \varepsilon\,\|h\|_1 &\geq E(G_n, J, h)\\ &\geq \min_{\mathbf{a}\in\triangle_q^{(\varepsilon)}} \big(E_{\mathbf{a},\varepsilon}(G_n, J) - \langle \mathbf{a}, h\rangle\big) - \varepsilon\,\|h\|_1\end{aligned}$$

for all $\mathbf{a}' \in \triangle$ and all $\varepsilon > 0$. Using these two bounds, the proof of (ii) is now identical to the proof of (i).

To prove (iii) and (iv), we note that the number of terms in (3.1) and (2.10) is at most $q^{|V(G)|}$, implying that

$$F(G, J, h) \le E(G, J, h) \le F(G, J, h) + \log q$$

and

$$F_{\mathbf{a},\varepsilon}(G, J) \le E_{\mathbf{a},\varepsilon}(G, J) \le F_{\mathbf{a},\varepsilon}(G, J) + \log q.$$

Rescaling J and h by a factor $\lambda \to \infty$, and using that both the energies and microcanonical energies are linear in λ, we obtain the claimed implications. □

4.2. **Convergence of quotients implies convergence of microcanonical ground state energies.** In this section, we prove Theorem 3.3.

To this end, we use (2.15) to express the microcanonical ground state energies as

$$E_{\mathbf{a},\varepsilon}(G, J) = -\max_{(\alpha,\beta)\in\mathcal{S}_{\mathbf{a},\varepsilon}(G)} \langle \beta, J\rangle.$$

where

$$\mathcal{S}_{\mathbf{a},\varepsilon}(G) = \mathcal{S}_q(G) \cap \mathcal{S}_{\mathbf{a},\varepsilon} \quad \text{with} \quad \mathcal{S}_{\mathbf{a},\varepsilon} = \{(\alpha, \beta) \in \mathcal{S}_q : \|\alpha - \mathbf{a}\|_\infty \le \varepsilon\}.$$

Proof of Theorem 3.3. In view of Lemma 3.2 it is enough to prove convergence of the microcanonical ground state energies.

Let $\varepsilon > 0$. Since $\mathcal{S}_q(G_n)$ is assumed to converge to $\mathcal{S}_q^\infty$, we can find an $n_0 \in \mathbb{N}$ such that

$$d_1^{\mathrm{Hf}}(\mathcal{S}_q(G_n), \mathcal{S}_q^\infty) \le \varepsilon \quad \text{for all} \quad n \ge n_0.$$

For $n \ge n_0$, choose $(\alpha^{(n)}, \beta^{(n)}) \in \mathcal{S}_{\mathbf{a},\varepsilon}(G_n) \subseteq \mathcal{S}_q(G_n)$ such that

$$E_{\mathbf{a},\varepsilon}(G_n, J) = -\langle \beta^{(n)}, J\rangle,$$

and choose $(\tilde\alpha^{(n)}, \tilde\beta^{(n)}) \in \mathcal{S}_q^\infty$ such that $d_1((\tilde\alpha^{(n)}, \tilde\beta^{(n)}), (\alpha^{(n)}, \beta^{(n)})) \le \varepsilon$. Then

$$E_{\mathbf{a},\varepsilon}(G_n, J) \ge -\langle \tilde\beta^{(n)}, J\rangle - \varepsilon\|J\|_\infty.$$

Since $|\tilde\alpha_i^{(n)} - a_i| \le |\alpha_i^{(n)} - a_i| + d_1((\tilde\alpha^{(n)}, \tilde\beta^{(n)}), (\alpha^{(n)}, \beta^{(n)})) \le 2\varepsilon$, we have that $(\tilde\alpha^{(n)}, \tilde\beta^{(n)}) \in \mathcal{S}_{\mathbf{a},2\varepsilon}$, proving in particular that

$$E_{\mathbf{a},\varepsilon}(G_n, J) \ge -\langle \tilde\beta^{(n)}, J\rangle - \varepsilon\|J\|_1 \ge -\sup_{(\alpha,\beta)\in\mathcal{S}_q^\infty\cap\mathcal{S}_{\mathbf{a},2\varepsilon}} \langle \beta, J\rangle - \varepsilon\|J\|_\infty.$$

Taking first the lim inf as $n \to \infty$ and then the limit $\varepsilon \to 0$, this shows that

$$\lim_{\varepsilon\to 0} \liminf_{n\to\infty} E_{\mathbf{a},\varepsilon}(G_n, J) \ge -\lim_{\varepsilon\to 0} \sup_{(\alpha,\beta)\in\mathcal{S}_q^\infty\cap\mathcal{S}_{\mathbf{a},\varepsilon}} \langle \beta, J\rangle = -\max_{(\alpha,\beta)\in\mathcal{S}_q^\infty\cap\mathcal{S}_{\mathbf{a}}} \langle \beta, J\rangle,$$

where the final step is due to compactness. The proof of the matching upper bound

$$\lim_{\varepsilon\to 0} \limsup_{n\to\infty} E_{\mathbf{a},\varepsilon}(G_n, J) \le -\lim_{\varepsilon\to 0} \sup_{(\alpha,\beta)\in\mathcal{S}_q^\infty\cap\mathcal{S}_{\mathbf{a},\varepsilon}} \langle \beta, J\rangle = -\max_{(\alpha,\beta)\in\mathcal{S}_q^\infty\cap\mathcal{S}_{\mathbf{a}}} \langle \beta, J\rangle,$$

proceeds along the same lines, now using that for any $(\alpha, \beta) \in \mathcal{S}_q^\infty$ with $\|\alpha - \mathbf{a}\|_\infty \le \varepsilon$ we can find $(\alpha^{(n)}, \beta^{(n)}) \in \mathcal{S}_{\mathbf{a},2\varepsilon}(G_n)$ with $d_1((\alpha, \beta), (\alpha^{(n)}, \beta^{(n)})) \le \varepsilon$. □

4.3. **LD convergence implies convergence of quotients.** In this section, we prove part (i) of Theorem 2.9, which is statement (iii) of the following lemma.

Lemma 4.1. *Let $q \in \mathbb{N}$, assume that $(G_n)_{n\geq 0}$ is a q-LD convergent sequence of weighted graphs with rate function I_q, and let $\mathcal{S}_q(I_q) = \{(\alpha,\beta) \in \mathcal{S}_q : I_q((\alpha,\beta)) < \infty\}$. Then the following are true:*

(i) *The set $\mathcal{S}_q(I_q)$ is closed with respect to the metric d_1.*
(ii) *The set $\mathcal{S}_q(I_q)$ is equal to the set $\mathcal{S}_q^\infty = \Big\{(\alpha,\beta) : d_1\Big((\alpha,\beta), \mathcal{S}_q(G_n)\Big) \to 0\Big\}$.*
(iii) *$\mathcal{S}_q(G_n)$ converges to $\mathcal{S}_q(I_q)$ in the Hausdorff distance.*

Proof. (i) For each $a \in \mathbb{R}$, the set $\{(\alpha,\beta) \in \mathcal{S}_q : I_q((\alpha,\beta)) \leq a\}$ is closed by the lower semicontinuity of I_q. To prove closedness of the set $\mathcal{S}_q(I_q)$, we observe that

$$I_{q,\varepsilon,n}((\alpha,\beta)) = -\frac{\log \mathcal{P}_{q,G_n}\left[d_1((\alpha,\beta), G_n/\phi) \leq \varepsilon\right]}{|V(G_n)|}$$

takes values in $[0, \log q] \cup \{\infty\}$, which in turn implies that I_q takes values in $[0, \log q] \cup \{\infty\}$ and shows that $\mathcal{S}_q(I_q) = \{(\alpha,\beta) : I_q((\alpha,\beta)) \leq \log q\}$.

(ii) Let us first assume that $(\alpha,\beta) \in \mathcal{S}_q(I_q)$. Then

$$\limsup_{n\to\infty} I_{q,\varepsilon,n}((\alpha,\beta)) \leq I_q((\alpha,\beta)) \leq \log q \quad \text{for all} \quad \varepsilon > 0,$$

because $I_{q,\varepsilon,n}((\alpha,\beta))$ is non-increasing in ε. Since $I_{q,\varepsilon,n}$ takes values in $[0, \log q] \cup \{\infty\}$, this implies that for all $\varepsilon > 0$ we can find an n_0 such that

$$I_{q,\varepsilon,n}((\alpha,\beta)) \leq \log q \quad \text{if} \quad n \geq n_0,$$

which in turn implies that

$$d_1\Big((\alpha,\beta), \mathcal{S}_q(G_n)\Big) \leq \varepsilon \quad \text{if} \quad n \geq n_0.$$

This proves that $(\alpha,\beta) \in \mathcal{S}_q(I_q)$ implies $d_1\Big((\alpha,\beta), \mathcal{S}_q(G_n)\Big) \to 0$.

Assume on the other hand that $d_1\Big((\alpha,\beta), \mathcal{S}_q(G_n)\Big) \to 0$, and by contradiction, assume further that $(\alpha,\beta) \notin \mathcal{S}_q(I_q)$, i.e., assume that $I_q((\alpha,\beta)) = \infty$. Since $I_{q,\varepsilon,n}((\alpha,\beta))$ takes values in $[0, \log q] \cup \{\infty\}$, this implies that there exists an $\varepsilon > 0$ such that

$$\liminf_{n\to\infty} I_{q,\varepsilon,n}((\alpha,\beta)) = \infty.$$

which in turn implies that there exists an $n_0 < \infty$ such that $I_{q,\varepsilon,n}((\alpha,\beta)) = \infty$ for all $n \geq n_0$. As a consequence,

$$d_1\Big((\alpha,\beta), \mathcal{S}_q(G_n)\Big) > \varepsilon \quad \text{if} \quad n \geq n_0,$$

contradicting the assumption that $d_1\Big((\alpha,\beta), \mathcal{S}_q(G_n)\Big) \to 0$.

(iii) Using the fact that $\mathcal{S}_q(I_q)$ is compact, one easily transforms the statement that $d_1\Big((\alpha,\beta), \mathcal{S}_q(G_n)\Big) \to 0$ for all $(\alpha,\beta) \in \mathcal{S}_q(I_q)$ into the uniform statement that

$$\sup_{(\alpha,\beta)\in\mathcal{S}_q(I_q)} d_1\Big((\alpha,\beta), \mathcal{S}_q(G_n)\Big) \to 0.$$

To prove convergence in the Hausdorff distance we have to prove the matching bound

$$\sup_{(\alpha,\beta)\in\mathcal{S}_q(G_n)} d_1\Big((\alpha,\beta), \mathcal{S}_q(I_q)\Big) \to 0.$$

Fix $\varepsilon > 0$, and let S be the set $S = \{(\alpha,\beta) \in \mathcal{S}_q : d_1((\alpha,\beta), \mathcal{S}_q(I_q)) \geq \varepsilon\}$. Since S is closed, we may use (3.4) to conclude that

$$\limsup_{n\to\infty} \frac{\log \mathcal{P}_{q,G_n}[G_n/\phi \in S]}{|V(G_n)|} \leq - \inf_{(\alpha,\beta)\in S} I_q((\alpha,\beta)) = -\infty.$$

Since the probability on the left hand side takes values in $\{0\} \cup [q^{-|V(G_n)|}, 1]$ this shows there must exists an $n_0 = n_0(\varepsilon, q)$ such that $\mathcal{P}_{q,G_n}[G_n/\phi \in S] = 0$ if $n \geq n_0$, showing that $\mathcal{S}_q(G_n) \cap S = \emptyset$ when $n \geq n_0$. Expressed differently, for all $\varepsilon > 0$ we can find an n_0 such that for $n \geq n_0$,

$$\mathcal{S}_q(G_n) \subseteq \{(\alpha,\beta) \in \mathcal{S}_q : d_1((\alpha,\beta), \mathcal{S}_q(I_q)) < \varepsilon\}.$$

Or still expressed differently, we can find a sequence $\varepsilon_n \to 0$ as $n \to \infty$ such that

$$d_1((\alpha,\beta), \mathcal{S}_q(I_q)) \leq \varepsilon_n \quad \text{for all} \quad (\alpha,\beta) \in \mathcal{S}_q(G_n). \qquad \square$$

4.4. **LD convergence implies convergence of free energies.** In this section, we prove part (ii) of Theorem 2.9.

(ii) Given $\delta, \varepsilon > 0$, chose an arbitrary $(\alpha,\beta) \in \mathcal{S}_{\mathbf{a},\delta}$, and let $\Omega_{(\alpha,\beta),\varepsilon}$ be the set of configurations $\phi \colon V(G_n) \to [q]$ such that $d_1((\alpha,\beta), G_n/\phi) \leq \varepsilon$. If $\phi \in \Omega_{(\alpha,\beta),\varepsilon}$, then $|a_i - \alpha_i(G_n/\phi)| \leq \varepsilon + \delta$, implying that $\Omega_{(\alpha,\beta),\varepsilon} \subseteq \Omega_{\mathbf{a},\varepsilon+\delta}(G_n)$. Using further that $d_1((\alpha,\beta), G_n/\phi) \leq \varepsilon$ implies that $|\langle\beta, J\rangle - \langle\beta(G_n/\phi), J\rangle| \leq \varepsilon\|J\|_\infty$, and we then bound

$$\begin{aligned} &q^{|V(G_n)|} e^{\langle\beta,J\rangle|V(G_n)|} \mathcal{P}_{q,G_n}\left[d_1((\alpha,\beta), G_n/\phi) \leq \varepsilon\right] \\ &\qquad \leq \sum_{\phi\in\Omega_{\mathbf{a},\varepsilon+\delta}(G_n)} e^{\left(\langle\beta(G_n/\phi),J\rangle + \varepsilon\|J\|_\infty\right)|V(G_n)|} \\ &\qquad = Z_{G_n,J}^{(\mathbf{a},\varepsilon+\delta)} e^{\varepsilon\|J\|_\infty|V(G_n)|}, \end{aligned}$$

where the last step follows from the definition (2.10) and the fact that $-E_\phi(G_n, J)$ can be expressed as $\langle\beta(G_n/\phi), J\rangle$. Using (2.16) plus monotonicity in ε to guarantee the existence of the limit $\varepsilon \to 0$, this implies that

$$\lim_{\varepsilon\to 0} \liminf_{n\to\infty} \frac{\log Z_{G_n,J}^{(\mathbf{a},\varepsilon+\delta)}}{|V(G_n)|} \geq \log q + \langle\beta, J\rangle - I_q((\alpha,\beta)).$$

Since $\delta > 0$ and $(\alpha,\beta) \in \mathcal{S}_{\mathbf{a},\delta}$ were arbitrary, this shows that

$$\begin{aligned} \lim_{\varepsilon\to 0} \liminf_{n\to\infty} \frac{\log Z_{G_n,J}^{(\mathbf{a},\varepsilon)}}{|V(G_n)|} &\geq \lim_{\delta\to 0} \sup_{(\alpha,\beta)\in\mathcal{S}_{\mathbf{a},\delta}} \Big(\log q + \langle\beta, J\rangle - I_q((\alpha,\beta))\Big) \\ &\geq \sup_{(\alpha,\beta)\in\mathcal{S}_{\mathbf{a}}} \Big(\log q + \langle\beta, J\rangle - I_q((\alpha,\beta))\Big) = -F_{\mathbf{a}}(I_q, J). \end{aligned}$$

To get a matching upper bound we again fix $\mathbf{a}$ and $\varepsilon, \delta > 0$. Since $\mathcal{S}_{\mathbf{a},\varepsilon}$ is closed and hence compact, we can find a finite set $S_\delta \subset \mathcal{S}_{\mathbf{a},\varepsilon}$ such that $d_1^{\mathrm{Hf}}(S_\delta, \mathcal{S}_{\mathbf{a},\varepsilon}) \leq \delta$. For $s \in S_\delta$, let $B_\delta(s)$ be the set of pairs $(\alpha,\beta) \in \mathcal{S}_{\mathbf{a},\varepsilon}$ such that $d_1(s, (\alpha,\beta)) \leq \delta$. Then $\mathcal{S}_{\mathbf{a},\varepsilon} = \bigcup_{s\in S_\delta} B_\delta(s)$. As a consequence,

$$\begin{aligned} Z_{G_n,J}^{(\mathbf{a},\varepsilon)} &= \sum_{\phi:V(G_n)\to[q]} e^{\langle G_n/\phi, J\rangle|V(G_n)|} \mathbf{1}_{G_n/\phi\in\mathcal{S}_{\mathbf{a},\varepsilon}} \\ &\leq q^{|V(G_n)|} \sum_{s\in S_\delta} \left(\sup_{(\alpha,\beta)\in B_\delta(s)} e^{\langle\beta/\phi,J\rangle|V(G_n)|} \mathcal{P}_{q,G_n}\left[G_n/\phi \in B_\delta(s)\right] \right). \end{aligned}$$

Since S_δ is finite and does not depend on n, we find that

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{\log Z^{(\mathbf{a},\varepsilon)}_{G_n,J}}{|V(G_n)|} \\
&\qquad \le \log q + \limsup_{n\to\infty} \max_{s\in S_\delta} \left(\sup_{(\alpha,\beta)\in B_\delta(s)} \langle \beta, J\rangle + \frac{\log \mathcal{P}_{q,G_n}\left[G_n/\phi \in B_\delta(s)\right]}{|V(G_n)|} \right) \\
&\qquad = \log q + \max_{s\in S_\delta} \left(\sup_{(\alpha,\beta)\in B_\delta(s)} \langle \beta, J\rangle + \limsup_{n\to\infty} \frac{\log \mathcal{P}_{q,G_n}\left[G_n/\phi \in B_\delta(s)\right]}{|V(G_n)|} \right) \\
&\qquad \le \log q + \max_{s\in S_\delta} \left(\sup_{(\alpha,\beta)\in B_\delta(s)} \langle \beta, J\rangle - \inf_{(\alpha,\beta)\in B_\delta(s)} I_q((\alpha,\beta)) \right),
\end{aligned}$$

where we used (3.4) in the last step.

Since $\sup_{(\alpha,\beta)\in B_\delta(s)}\langle \beta, J\rangle \le \langle \beta', J\rangle + 2\|J\|_\infty \delta$ for all $(\alpha',\beta') \in B_\delta(s)$, we conclude that

$$\begin{aligned}
\limsup_{n\to\infty} \frac{\log Z^{(\mathbf{a},\varepsilon)}_{G_n,J}}{|V(G_n)|} &\le \log q + \max_{s\in S_\delta} \sup_{(\alpha,\beta)\in B_\delta(s)} \Big(\langle \beta, J\rangle - I_q((\alpha,\beta)) \Big) + 2\delta \|J\|_\infty \\
&= \log q + \sup_{(\alpha,\beta)\in \mathcal{S}_{\mathbf{a},\varepsilon}} \Big(\langle \beta, J\rangle - I_q((\alpha,\beta)) \Big) + 2\delta \|J\|_\infty.
\end{aligned}$$

Sending $\delta \to 0$ and using the fact that $Z^{(\mathbf{a},\varepsilon)}_{G_n,J}$ is monotone in ε, this gives

$$\lim_{\varepsilon\to 0} \limsup_{n\to\infty} \frac{\log Z^{(\mathbf{a},\varepsilon)}_{G_n,J}}{|V(G_n)|} \le \sup_{(\alpha,\beta)\in \mathcal{S}_{\mathbf{a},\varepsilon}} \Big(\log q + \langle \beta, J\rangle - I_q((\alpha,\beta)) \Big).$$

Choose an arbitrary sequence ε_k going to zero, and choose $(\alpha_k,\beta_k) \in \mathcal{S}_{\mathbf{a},\varepsilon}$ such that the supremum on the right hand side is bounded by $\log q + \langle \beta_k, J\rangle - I_q((\alpha_k,\beta_k)) + \varepsilon_k$. Going to a subsequence if needed, assume that (α_k,β_k) converges to some $(\alpha,\beta) \in \mathcal{S}_{\mathbf{a}}$ in the d_1 distance. Then

$$\begin{aligned}
\lim_{\varepsilon\to 0} \limsup_{n\to\infty} \frac{\log Z^{(\mathbf{a},\varepsilon)}_{G_n,J}}{|V(G_n)|} &\le \log q + \langle \beta, J\rangle - \liminf_{k\to\infty} I_q((\alpha_k,\beta_k)) \\
&\le \log q + \langle \beta, J\rangle - I_q((\alpha,\beta))
\end{aligned}$$

where in the last step we used that I_q is lower semi-continuous. Since $(\alpha,\beta) \in \mathcal{S}_{\mathbf{a}}$, the right hand side is bounded by

$$\sup_{(\alpha,\beta)\in \mathcal{S}_{\mathbf{a}}} \Big(\log q + \langle \beta, J\rangle - I_q((\alpha,\beta)) \Big) = -F_{\mathbf{a}}(I_q, J),$$

as desired. □

5. Convergent sequences of graphons

In this section, we formulate and prove our main results in the language of graphons. Several of these results are generalizations of the corresponding results for L^∞ graphons proved in [8]; the exceptions are those involving LD convergence, which was not considered in [8]. It turns out, however, that most of our proofs are quite different from those of [8], most notably the proof that convergence of ground state energies implies convergence in metric (Section 5.5), which involves some new ideas not present in [8] such as the use of rearrangement inequalities.

5.1. **Upper regularity for graphons.** First we review the notion of upper regularity for graphons from [3].

Given a graphon W and a partition $\mathcal{P} = (Y_1, \ldots, Y_m)$ of the interval $[0,1]$ into finitely many measurable sets, we define $W_{\mathcal{P}}$ to be the step function whose value on $Y_i \times Y_j$ equals to the average of W over $Y_i \times Y_j$, i.e.,

$$W_{\mathcal{P}} = \frac{1}{\lambda(Y_i)\lambda(Y_j)} \int_{Y_i \times Y_j} W(x,y)\,dx\,dy \quad \text{on } Y_i \times Y_j,$$

where λ denotes the Lebesgue measure. An easy fact is that $W \mapsto W_{\mathcal{P}}$ is contractive with respect to the L^p norms $\|\cdot\|_p$ and the cut norm $\|\cdot\|_\square$, i.e.,

$$\|W_{\mathcal{P}}\|_\square \leq \|W\|_\square \qquad \text{and} \qquad \|W_{\mathcal{P}}\|_p \leq \|W\|_p \qquad \text{for all } p \geq 1. \tag{5.1}$$

Another standard fact is that up to a factor of 2, $W_{\mathcal{P}}$ is the best step function approximation to W with steps in $\mathcal{P}$, in the sense that

$$\|W - W_{\mathcal{P}}\|_\square \leq 2\|W - U_{\mathcal{P}}\|_\square$$

for all graphons U. To see why, note that

$$\begin{aligned}\|W - W_{\mathcal{P}}\|_\square &\leq \|W - U_{\mathcal{P}}\|_\square + \|U_{\mathcal{P}} - W_{\mathcal{P}}\|_\square \\ &= \|W - U_{\mathcal{P}}\|_\square + \|(W - U_{\mathcal{P}})_{\mathcal{P}}\|_\square \\ &\leq 2\|W - U_{\mathcal{P}}\|_\square.\end{aligned}$$

Definition 5.1. Let $K\colon (0,\infty) \to (0,\infty)$ be any function. We say that a graphon W has K-*bounded tails* if for each $\varepsilon > 0$,

$$\int_{[0,1]^2} |W(x,y)| \mathbf{1}_{|W(x,y)| \geq K(\varepsilon)}\,dx\,dy \leq \varepsilon. \tag{5.2}$$

A graphon W is (K,η)-*upper regular* if $W_{\mathcal{P}}$ has K-bounded tails whenever $\mathcal{P}$ is a partition of the interval $[0,1]$ into sets of measure at least η. A sequence $(W_n)_{n\geq 0}$ of graphons is *uniformly upper regular* if there exist $K\colon (0,\infty) \to (0,\infty)$ and a sequence $\eta_n \to 0$ such that W_n is (K,η_n)-upper regular for all n.

A key result from [3] is that every uniformly upper regular sequence of graphons contains a subsequent that converges in cut distance to some graphon. This is stated below.

Theorem 5.2 (Theorem C.7 in [3])**.** *If $(W_n)_{n\geq 0}$ is a sequence of uniformly upper regular graphons, then there exists a graphon W and a subsequence $(W_n')_{n\geq 0}$ of $(W_n)_{n\geq 0}$ such that $\delta_\square(W_n', W) \to 0$.*

5.2. **Equivalent notions of convergence for graphons.** The main theorem of this section, Theorem 5.3 below, is the analogue of the first four statements of Theorem 2.15. To state it, we need the analogue of the microcanonical ground state energies and microcanonical free energies define in (2.12) and (2.11), namely the quantities

$$\mathcal{E}_{\mathbf{a},\varepsilon}(W,J) = \inf_{\substack{\rho \in \mathsf{FP}_q \\ \|\alpha(\rho) - \mathbf{a}\|_\infty \leq \varepsilon}} \mathcal{E}_\rho(W,J),$$

and

$$\mathcal{F}_{\mathbf{a},\varepsilon}(W,J) = \inf_{\substack{\rho \in \mathsf{FP}_q \\ \|\alpha(\rho) - \mathbf{a}\|_\infty \leq \varepsilon}} \Big(\mathcal{E}_\rho(W,J) - \mathsf{Ent}(\rho)\Big). \tag{5.3}$$

The following theorem is the main theorem of this section, and will be proved in Sections 5.4, 5.5, and 5.6 below.

Theorem 5.3. *Let $(W_n)_{n\ge 0}$ be a sequence of uniformly upper regular graphons. Then the following statements are equivalent:*

(i) *$(W_n)_{n\ge0}$ is a Cauchy sequence in the cut metric $\delta_\square$.*
(ii) *For every $q \in \mathbb{N}$, the sequence $(\widehat{\mathcal{S}}_q(W_n))_{n\ge0}$ is a Cauchy sequence under the Hausdorff distance d_1^{Hf}.*
(iii) *For every $q \in \mathbb{N}$, $\mathbf{a} \in \triangle_q$, and symmetric matrix $J \in \mathbb{R}^{q\times q}$,*

$$\lim_{\varepsilon\to0} \liminf_{n\to\infty} \mathcal{E}_{\mathbf{a},\varepsilon}(W_n, J) = \lim_{\varepsilon\to0} \limsup_{n\to\infty} \mathcal{E}_{\mathbf{a},\varepsilon}(W_n, J).$$

(iv) *For every $q \in \mathbb{N}$, $\mathbf{a} \in \triangle_q$, and symmetric matrix $J \in \mathbb{R}^{q\times q}$,*

$$\lim_{\varepsilon\to0} \liminf_{n\to\infty} \mathcal{F}_{\mathbf{a},\varepsilon}(W_n, J) = \lim_{\varepsilon\to0} \limsup_{n\to\infty} \mathcal{F}_{\mathbf{a},\varepsilon}(W_n, J).$$

Remark 5.4.

(i) We will prove the theorem by showing that (i) ⇒ (ii) ⇒ (iii) ⇒ (i) and that (i) ⇒ (iv) ⇒ (iii). It turns out that the assumption of uniform upper regularity is only needed for the proof that (iii) ⇒ (i). All other implications hold for arbitrary sequences of graphons.
(ii) Under the assumption (ii) of the above theorem, $\widehat{\mathcal{S}}_q(W_n)$ converges to the compact set[5] $\mathcal{S}_q^\infty = \{(\alpha,\beta) \colon d_1((\alpha,\beta), \widehat{\mathcal{S}}_q(W_n)) \to 0\}$. Proceeding as in the proof of Theorem 3.3, this in turn implies that (iii) holds with the limit given as

$$E_{\mathbf{a}}(J) = -\max_{\substack{(\alpha,\beta)\in\mathcal{S}_q^\infty \\ \alpha=\mathbf{a}}} \langle \beta, J\rangle,$$

again without the assumption of uniform upper regularity.
(iii) Under the assumption (i) of the above theorem, the sequences $(\mathcal{E}_{\mathbf{a}}(W_n, J))_{n\ge0}$ and $(\mathcal{F}_{\mathbf{a}}(W_n, J))_{n\ge0}$ are convergent for all $q, \mathbf{a}, J$. (In particular, the use of ε in the theorem statement is just for comparison with the case of graphs, and not because it is truly needed.) Finally, under the assumption of uniform upper regularity, each of these two statements is not only necessary but also sufficient for convergence in metric to hold, as we will show in Sections 5.5 and 5.6.

5.3. **Compactness of quotient space.** Before jumping into the proof of Theorem 5.3, we prove some compactness results about quotients of W, thereby shedding light on the quantities $\mathcal{E}_{\mathbf{a}}(W, J)$, $\mathcal{F}_{\mathbf{a}}(W, J)$, and $I_q((\alpha,\beta), W)$.

Recall the ℓ_1 distance d_1 from (2.7) as well as the definitions of fractional graphon quotients from Section 2.6.

Proposition 5.5. *Let W be a graphon, let $q \in \mathbb{N}$, and let $\mathbf{a} \in \triangle_q$. Then $\widehat{\mathcal{S}}_q(W)$ and $\widehat{\mathcal{S}}_{\mathbf{a}}(W)$ are compact under the metric d_1.*

We will prove this proposition after we develop a few preliminaries.

In (2.17)–(2.19), $\mathcal{E}_{\mathbf{a}}(W, J)$, $\mathcal{F}_{\mathbf{a}}(W, J)$, and $I_q((\alpha,\beta), W)$ were originally defined as infima over some subset of fractional partitions. We will see that the infima are attained by some fractional partitions, so that the "inf" can be replaced by "min".

[5] The nonempty compact subsets of $\widehat{\mathcal{S}}_q$ form a complete metric space under d_1^{Hf} (see [13]).

Furthermore, we will see that

$$\mathcal{E}_{\mathbf{a}}(W,J) = \lim_{\varepsilon\to 0} \mathcal{E}_{\mathbf{a},\varepsilon}(W,J), \tag{5.4}$$

$$\mathcal{F}_{\mathbf{a}}(W,J) = \lim_{\varepsilon\to 0} \mathcal{F}_{\mathbf{a},\varepsilon}(W,J), \qquad \text{and} \tag{5.5}$$

$$I_q((\alpha,\beta),W) = \lim_{\varepsilon\to 0} \inf_{\substack{\rho\in\mathsf{FP}_q \\ d_1(W/\rho,(\alpha,\beta))\le\varepsilon}} (\log q - \mathsf{Ent}(\rho)). \tag{5.6}$$

The quantities $\mathcal{E}_{\mathbf{a}}(W,J)$ and $\mathcal{F}_{\mathbf{a}}(W,J)$ are both continuous with respect to $\mathbf{a}, W, J$. On the other hand, $I_q((\alpha,\beta),W)$ is lower semicontinuous in its arguments (Proposition 5.10), and it is not continuous (it takes values in $[0,\log q]$ when $(\alpha,\beta)\in\widehat{\mathcal{S}}_q(W)$ and is infinite otherwise).

It follows as an immediate corollary of Proposition 5.5 that

$$\mathcal{E}(W,J,h) = -\max_{(\alpha,\beta)\in\widehat{\mathcal{S}}_q(W)} \big(\langle\beta,J\rangle + \langle\alpha,h\rangle\big)$$

and

$$\mathcal{E}_{\mathbf{a}}(W,J) = -\max_{(\alpha,\beta)\in\widehat{\mathcal{S}}_{\mathbf{a}}(W)} \langle\beta,J\rangle, \tag{5.7}$$

since $(\alpha,\beta)\mapsto\langle\beta,J\rangle$ and $(\alpha,\beta)\mapsto\langle\alpha,h\rangle$ are continuous in the d_1 metric. This gives an alternate representation of ground state energies of W in terms of its quotients.

5.3.1. *Approximations by step functions.* One way to approximate a graphon W by step functions is given by the following lemma, which is an immediate consequence of the almost everywhere differentiability of the integral function.

Lemma 5.6. *Let $p\ge 1$. For a positive integer n, let $\mathcal{P}_n$ be the partition of $[0,1]$ into consecutive intervals of length $1/n$. If W is a graphon, then $W_{\mathcal{P}_n}\to W$ almost everywhere. In addition, $W_{\mathcal{P}_n}\to W$ in L^p whenever W is an L^p graphon.*

Proof. Almost everywhere convergence follows the Lebesgue differentiation theorem. To get convergence in L^p we approximate W by the bounded graphon $W_M = W\mathbf{1}_{|W|\le M}$, where $M>0$. By triangle inequality and (5.1),

$$\begin{aligned}\|W - W_{\mathcal{P}_n}\|_p &\le \|W_M - (W_M)_{\mathcal{P}_n}\|_p + \|W - W_M\|_p + \|(W-W_M)_{\mathcal{P}_n}\|_p \\ &\le \|W_M - (W_M)_{\mathcal{P}_n}\|_p + 2\|W - W_M\|_p.\end{aligned}$$

The second term on the right can be made arbitrarily small by setting M to be sufficiently large, and for any fixed M, the first term on the right goes to zero as $n\to\infty$. This shows that $\|W - W_{\mathcal{P}_n}\|_p \to 0$. ☐

The lemma does not, however, give any information on the speed of convergence. If instead of almost everywhere convergence we content ourselves with convergence in the cut metric, the situation is different, as is well known in the case of L^2 graphons, where one can apply the weak version of the regularity lemma first established in [10]. This lemma can be generalized to L^p graphons for $p>1$ and more generally to any graphon with K-bounded tails (see [3]), but we will not need this here, where we use only the corresponding version for uniformly upper regular sequences of graphs (see Theorem 6.1 in Section 6).

5.3.2. *Limits of fractional partitions.* We say that a sequence $\rho^{(n)} \in \mathsf{FP}_q$ of fractional partitions converges to $\rho \in \mathsf{FP}_q$ *over rational intervals* if $\int_D \rho_i^{(n)}(x)\,dx \to \int_D \rho_i(x)\,dx$ for every interval $D \subseteq [0,1]$ with rational endpoints.

Lemma 5.7. *Fix $q \in \mathbb{N}$. Every sequence $\rho^{(n)} \in \mathsf{FP}_q$ of fractional partitions contains a subsequence that converges to some $\rho \in \mathsf{FP}_q$ over rational intervals.*

Proof. By restricting to a subsequence, we may assume that $\int_D \rho_i^{(n)}(x)\,dx$ converges for every interval $D \subseteq [0,1]$ with rational endpoints, and let us denote this limiting value by $\mu_i(D)$. Then, by the extension theorem for measures (Proposition 5.5 in [11]), μ_i can be extended to a measure on $[0,1]$ such that $\sum_{i\in[q]} \mu_i(D) = \lambda(D)$ for every measurable $D \subseteq [0,1]$, where λ is the Lebesgue measure. It is easy to see that μ_i is absolutely continuous with respect to λ. Defining ρ_i to be the density of μ_i with respect to λ and changing ρ_i on a set of measure zero, we obtain the desired fractional partition $\rho = (\rho_i)_{i\in[q]}$. $\square$

Lemma 5.8. *If $\rho^{(n)} \in \mathsf{FP}_q$ converges to $\rho \in \mathsf{FP}_q$ over rational intervals and $\|W_n - W\|_\square \to 0$, then $d_1(W_n/\rho^{(n)}, W/\rho) \to 0$.*

Proof. We have $\alpha_i(\rho^{(n)}) = \int_{[0,1]} \rho_i^{(n)} \to \int_{[0,1]} \rho_i = \alpha_i(\rho)$. Next we have

$$\left|\beta_{ij}(W_n/\rho^{(n)}) - \beta_{ij}(W/\rho^{(n)})\right| = \left|\int \rho_i^{(n)}(x)\rho_j^{(n)}(y)(W_n(x,y) - W(x,y))\,dx\,dy\right| \\ \leq \|W_n - W\|_\square .$$

It remains to show that $\beta_{ij}(W/\rho^{(n)}) \to \beta_{ij}(W/\rho)$, i.e.,

$$\int \rho_i^{(n)}(x)\rho_j^{(n)}(y)W(x,y)\,dx\,dy \to \int \rho_i(x)\rho_j(y)W(x,y)\,dx\,dy,$$

which follows from Lemma 5.6 as we can approximate W arbitrarily well in L^1 using step functions with rational steps, and $\rho^{(n)}$ converges to ρ over rational intervals. $\square$

Now we can prove the compactness of the set of quotients.

Proof of Proposition 5.5. Let $(W/\rho^{(n)})_{n\geq 1}$ be a sequence of quotients in $\widehat{\mathcal{S}}_q(W)$ (or $\widehat{\mathcal{S}}_{\mathbf{a}}(W)$). By Lemma 5.7 we can restrict to a subsequence so that $\rho^{(n)}$ converges to some $\rho \in \mathsf{FP}_q$ over rational intervals. By Lemma 5.8, we have $d_1(W/\rho^{(n)}, W/\rho) \to 0$, thereby proving that the space of quotients is closed and hence compact. $\square$

The claim (5.4) has a similar proof: we have $\mathcal{E}_\rho(W,J) = -\langle \beta(W/\rho), J\rangle$, so that $d_1(W/\rho^{(n)}, W/\rho) \to 0$ implies $\mathcal{E}_{\rho^{(n)}}(W,J) \to \mathcal{E}_\rho(W,J)$.

5.3.3. *Entropy and lower semicontinuity.* Now we prove (5.5) and (5.6), and furthermore the claim that in the definitions (2.18) and (2.19) for $\mathcal{F}_{\mathbf{a}}(W,J)$ and $I_q((\alpha,\beta),W)$ the infimum is attained by some fractional partition. In fact, they are all immediate consequences of Lemma 5.7 along with the following lemma.

Lemma 5.9. *If $\rho^{(n)} \in \mathsf{FP}_q$ converges to $\rho \in \mathsf{FP}_q$ over rational intervals, then*

$$\limsup_{n\to\infty} \mathsf{Ent}(\rho^{(n)}) \leq \mathsf{Ent}(\rho).$$

Proof. For any positive integer k, let $\rho_{[k]} \in \mathsf{FP}_q$ (and similarly $\rho^{(n)}_{[k]}$) denote the fractional partition obtained from $\rho_{[k]}$ by averaging over the interval $[(j-1)/k, j/k)$ for each integer $j \in [k]$. Specifically, we set the value of $\rho_{[k],i}$ on $[(j-1)/k, j/k)$ to be $k\int_{(j-1)/k}^{j/k} \rho_i(x)\,dx$.

Since $-x\log x$ is concave, $\mathsf{Ent}(\rho^{(n)}) \le \mathsf{Ent}(\rho^{(n)}_{[k]})$ by Jensen's inequality. For a fixed k we have

$$\limsup_{n\to\infty} \mathsf{Ent}(\rho^{(n)}) \le \limsup_{n\to\infty} \mathsf{Ent}(\rho^{(n)}_{[k]}) = \mathsf{Ent}(\rho_{[k]})$$

where the last equality follows from $\rho^{(n)}$ converging to ρ on rational intervals. Finally, we have $\rho_{[k],i} \to \rho_i$ almost everywhere as $k \to \infty$ by the Lebesgue differentiation theorem, and thus $\mathsf{Ent}(\rho_{[k]}) \to \mathsf{Ent}(\rho)$ as $k \to \infty$ by the bounded convergence theorem. This proves the lemma. □

Proposition 5.10. *Let $q \in \mathbb{N}$. The function $I_q((\alpha,\beta), W)$ is lower semicontinuous (with the metric d_1 on the first argument and $\delta_\square$ on the second).*

Proof. We need to show that if $(\alpha^{(n)}, \beta^{(n)}) \to (\alpha,\beta)$ in d_1 and $W_n \to W$ in $\delta_\square$ as $n \to \infty$, then

$$\liminf_{n\to\infty} I_q((\alpha^{(n)}, \beta^{(n)}), W_n) \ge I_q((\alpha,\beta), W).$$

We may restrict to a subsequence so that $I_q((\alpha^{(n)}, \beta^{(n)}), W_n)$ converges to the original $\liminf$. Since I_q is invariant under measure preserving bijections for the graphon, we may assume that $\|W_n - W\|_\square \to 0$. The result is automatic if the limit is infinity, so we might as well assume that $I_q((\alpha^{(n)}, \beta^{(n)}), W_n) < \infty$ (and hence at most $\log q$) for all n, so that there is some $\rho^{(n)} \in \mathsf{FP}_q$ with $W/\rho^{(n)} = (\alpha^{(n)}, \beta^{(n)})$ and $I_q((\alpha^{(n)}, \beta^{(n)}), W_n) = \log q - \mathsf{Ent}(\rho^{(n)})$. By Lemma 5.7 we can further restrict to a subsequence so that $\rho^{(n)}$ converges to some $\rho \in \mathsf{FP}_q$ over rational intervals. By Lemma 5.8 we have $W/\rho = \lim_{n\to\infty} W_n/\rho^{(n)} = \lim_{n\to\infty}(\alpha^{(n)}, \beta^{(n)}) = (\alpha, \beta)$, so $I_q((\alpha,\beta), W) \le \log q - \mathsf{Ent}(\rho)$. By Lemma 5.9,

$$\begin{aligned}\liminf_{n\to\infty} I_q((\alpha^{(n)}, \beta^{(n)}), W_n) &= \liminf_{n\to\infty}(\log q - \mathsf{Ent}(\rho_n)) \\ &\ge \log q - \mathsf{Ent}(\rho) \ge I_q((\alpha,\beta), W),\end{aligned}$$

as desired. □

5.4. **Proof of (i)⇒(ii)⇒(iii) in Theorem 5.3.** The claim that (i) implies (ii) follows from Lemma 5.11 below. The claim that (ii) implies (iii)—with the limit expressed as described in Remark 5.4(ii)—is essentially identical to the proof of Theorem 3.3 from Section 4.2 and is left to the reader.

Lemma 5.11. *Let $q \in \mathbb{N}$, U and W be graphons, and $\rho \in \mathsf{FP}_q$. Then $d_1(U/\rho, W/\rho) \le q^2\|U - W\|_\square$ and hence*

$$d_1^{\mathrm{Hf}}(\widehat{\mathcal{S}}_q(U), \widehat{\mathcal{S}}_q(W)) \le q^2 \delta_\square(U, W). \tag{5.8}$$

Proof. We have $\alpha(U/\rho) = \alpha(\rho) = \alpha(W/\rho)$. Also for $i, j \in [q]$ we have

$$\begin{aligned}|\beta_{ij}(U/\rho) - \beta_{ij}(W/\rho)| &= \left|\int_{[0,1]^2} (U - W)(x,y)\rho_i(x)\rho_j(y)\,dx\,dy\right| \\ &\le \|U - W\|_\square .\end{aligned} \tag{5.9}$$

Summing over all $i, j \in [q]$ gives $d_1(U/\rho, W/\rho) \leq q^2 \|U - W\|_\square$. The claim (5.8) follows immediately. □

We next observe that microcanonical ground state energies are continuous in the cut metric. To state this result, we define the norm $\|J\|_1$ of a matrix $J \in \mathbb{R}^{q\times q}$ to be $\sum_{i,j\in[q]} |J_{ij}|$.

Proposition 5.12. *Let $q \in \mathbb{N}$, $\mathbf{a} \in \triangle_q$, and $h \in \mathbb{R}^q$, and let $J \in \mathbb{R}^{q\times q}$ be a symmetric matrix. If U and W are arbitrary graphons, then*

$$|\mathcal{E}_{\mathbf{a}}(U, J) - \mathcal{E}_{\mathbf{a}}(W, J)| \leq \|J\|_1 \delta_\square(W, U). \tag{5.10}$$

Proof. Since the left side of this bound does not change if we replace U by U^ϕ for some measure preserving bijection $\phi\colon [0, 1] \to [0, 1]$, it is enough to prove it in terms of $\|W - U\|_\square$ instead of $\delta_\square(W, U)$. Let $\rho \in \mathsf{FP}_q$. Using (5.9) we obtain

$$\begin{aligned} \left|\mathcal{E}_\rho(U, J, h) - \mathcal{E}_\rho(W, J, h)\right| &= \Big|\sum_{i,j\in[q]} (\beta_{ij}(U/\rho) - \beta_{ij}(W/\rho))J_{ij}\Big| \\ &\leq \|J\|_1\|U - W\|_\square, \end{aligned} \tag{5.11}$$

as desired. □

5.5. **Proof of (iii)⇒(i) in Theorem 5.3.** By the bound (5.10), convergence in metric implies convergence of the microcanonical ground state energies. To prove the converse, we will establish the following proposition, one of the main results of this section.

Theorem 5.13. *Let W and U be two graphons. If*

$$\mathcal{E}_{\mathbf{a}}(U, J) = \mathcal{E}_{\mathbf{a}}(W, J) \tag{5.12}$$

for all $q \in \mathbb{N}$, every symmetric matrix $J \in \mathbb{R}^{q\times q}$, and all $\mathbf{a}$ of the form

$$\mathbf{a}_q = (1/q, \dots, 1/q), \tag{5.13}$$

then $\delta_\square(W, U) = 0$.

This theorem proves the implication (iii)⇒(i) in Theorem 5.3, as well as the fact that convergence of $\mathcal{E}_{\mathbf{a}}(W_n, J)$ is sufficient for convergence in metric (see Remark 5.4(iii)). Indeed, for the second of these assertions, assume first that the ground state energies $\mathcal{E}_{\mathbf{a}}(W_n, J)$ converge for all $q \in \mathbb{N}$, all $\mathbf{a}$ of the form (5.13), and all J, while W_n does not converge in the cut metric. Since $(W_n)_{n\geq 0}$ is assumed to be uniformly upper regular, we may use Theorem 5.2 to find two subsequences W_n' and W_n'' of W_n that converge to two graphons W and U in the cut distance $\delta_\square$, while $\delta_\square(W, U) > 0$. But convergence in the cut distance implies convergence of the ground state energies by (5.10), which means that U and W have identical ground state energies, a contradiction. The proof of (iii)⇒(i) in Theorem 5.3 is similar, since convergence of W_n' to W in metric implies that $\widehat{\mathcal{S}}_q(W_n') \to \widehat{\mathcal{S}}_q(W)$ in the Hausdorff distance, which in turn can easily be seen to give convergence of the quantities in (iii) to $\mathcal{E}_{\mathbf{a}}(W, J)$ (the proof is the same as that of Theorem 3.3), and similarly for the convergence along the subsequence W_n'' to $\mathcal{E}_{\mathbf{a}}(U, J)$.

To prove Theorem 5.13, we will work with the quasi-inner product

$$\mathcal{C}(W, Y) = \sup_\phi \mathsf{E}[WY^\phi],$$

where the supremum is taken over all measure preserving bijections $\phi\colon [0,1] \to [0,1]$ and the expectation $\mathsf{E}[\cdot]$ is with respect to the Lebesgue measure on $[0,1]^2$, i.e.,

$$\mathsf{E}[WY^\phi] = \int_{[0,1]^2} W(x,y)Y^\phi(x,y)\,dx\,dy = \int_{[0,1]^2} W(x,y)Y(\phi(x),\phi(y))\,dx\,dy.$$

This quantity was defined in [8], where it was assumed that both W and Y are in L^∞. But the definition makes sense in our more general context, where we will assume that W is an arbitrary graphon and Y is bounded.

Lemma 5.14. *Let W and U be two graphons such that* (5.12) *holds for all $q \in \mathbb{N}$ and all $\mathbf{a}$ of the form* (5.13)*. Then*

$$\mathcal{C}(W,Y) = \mathcal{C}(U,Y) \tag{5.14}$$

for all bounded graphons Y.

Proof. If $Y = W^H$ for a weighted graph H on q nodes, where H has edge weights $J \in \mathbb{R}^{q\times q}$ and vertex weights $\mathbf{a}$ of the form (5.13), then

$$-\mathcal{E}_{\mathbf{a}}(W,J) = \mathcal{C}(W,W^H),$$

and the claim follows directly from the assumption (5.12). For general Y, we use Lemma 5.6 to approximate Y by step functions. More explicitly, let P_n be as in Lemma 5.6, and let $Y_n = Y_{P_n}$. Then $Y_n \to Y$ in L^1, and $\|Y_n\|_\infty \le \|Y\|_\infty$. Hence

$$\begin{aligned}\Big|\mathsf{E}[W\,Y_n^\phi] - \mathsf{E}[W\,Y^\phi]\Big| &\le 2\|Y\|_\infty \|W\mathbf{1}_{|W|\ge K}\|_1 + \|W\mathbf{1}_{|W|\le K}\|_\infty \|Y^\phi - Y_n^\phi\|_1 \\ &\le 2\|Y\|_\infty \|W\mathbf{1}_{|W|\ge K}\|_1 + K\|Y - Y_n\|_1.\end{aligned}$$

The right side can be made as small as desired by first choosing K large enough and then n large enough. Observing that the resulting convergence is uniform in ϕ, this implies that $\mathcal{C}(W,Y_n) \to \mathcal{C}(W,Y)$ and similarly for $\mathcal{C}(U,Y_n)$. From this, the claim follows. □

Lemma 5.15. *Let W and U be two graphons such that* (5.14) *holds for all bounded graphons Y. Consider the real valued random variables $\widehat{W} = W(x,y)$ and $\widehat{U} = U(x,y)$ where x,y are chosen independently uniformly at random from* $[0,1]$*. Then $\widehat{W}$ and $\widehat{U}$ have the same distribution.*

To prove this lemma, we will use some notions and results from the theory of monotone rearrangement.

5.5.1. *Monotone rearrangements and proof of Lemma 5.15.* Throughout this section, we identify graphons W with the real-valued random variables $W(x,y)$ obtained by choosing x,y independently uniformly at random from $[0,1]$; if W is such a random variable, we use $E[W]$ to denote its expectation. For $s \in \mathbb{R}$ we use $\{W > s\}$ to denote the event that $W > s$, namely $\{W > s\} = \{(x,y) : W(x,y) > s\}$, and $\Pr[W > s]$ to denote the probability of this event.

For a graphon W, we define the monotone rearrangement as the function

$$W^*(x_1,x_2) = \sup\{t \in \mathbb{R} : \Pr[W > t] > \|x\|_\infty^2\},$$

where $\|x\|_\infty = \max\{x_1,x_2\}$. Then $W^*(x_1,x_2)$ is a weakly decreasing function of $\|x\|_\infty$, and it has the same distribution as W; i.e.,

$$\Pr[W > t] = \Pr[W^* > t].$$

To see why, note that

$$\begin{aligned}\Pr[W^* > t_0] &= \Pr[\sup\{t \in \mathbb{R} : \Pr[W > t] > \|x\|_\infty^2\} > t_0] \\ &= \Pr[\text{there exists } t > t_0 \text{ such that } \Pr[W > t] > \|x\|_\infty^2] \\ &= \Pr[\Pr[W > t_0] > \|x\|_\infty^2] \\ &= \Pr\left[x_1 < \sqrt{\Pr[W > t_0]} \text{ and } x_2 < \sqrt{\Pr[W > t_0]}\right] \\ &= \Pr[W > t_0],\end{aligned}$$

where the third line follows from the fact that $t \mapsto Pr[W > t]$ is right-continuous.

Define two graphons W and U to be *aligned* if their level sets are nested, in the sense that for all $s, t \in \mathbb{R}$, either $\{W > s\} \subseteq \{U > t\}$ or $\{U > s\} \subseteq \{W > s\}$. It is easy to see that for any two graphons U, W, the monotone rearrangements U^* and W^* are aligned.

Let W be an L^1 graphon, and let Y be a bounded graphon. Then we have the *rearrangement inequality*

$$\mathsf{E}[W\,Y] \le \mathsf{E}[W^*\,Y^*].$$

See Appendix A for a proof. The proof also tells us that

$$\mathsf{E}[W\,Y] = \mathsf{E}[W^*\,Y^*]$$

if Y and W are aligned. Before delving into the proof of Lemma 5.15 we observe that

$$\mathsf{E}[WY] \le \mathcal{C}(W, Y) \le \mathsf{E}[W^*Y^*]. \tag{5.15}$$

The first inequality follows immediately from the definition of $\mathcal{C}(W, Y)$, and the second follows from the definition and the fact that Y and Y^ϕ have the same distribution, which in turn implies that $Y^* = (Y^\phi)^*$.

Proof of Lemma 5.15. Define $\operatorname{top}_\lambda(W) \subseteq [0, 1]^2$ in such a way that the Lebesgue measure of $\operatorname{top}_\lambda(W)$ is λ, and $W(u, v) \le \inf_{(x,y)\in\operatorname{top}_\lambda(W)} W(x, y)$ whenever $(u, v) \notin \operatorname{top}_\lambda(W)$. Explicitly, let $M = \sup\{s \in \mathbb{R} : \Pr[W > s] \ge \lambda\}$. If $\Pr\{W = M\} = 0$, then we have that $\Pr\{W > M\} = \Pr\{W \ge M\} = \lambda$, and we define $\operatorname{top}_\lambda(W) = \{W > M\}$. Otherwise, $\Pr\{W > M\} \le \lambda \le \Pr\{W \ge M\}$, in which case we chose $\operatorname{top}_\lambda(W)$ in such a way that $\{W > M\} \subseteq \operatorname{top}_\lambda(W) \subseteq \{W \ge M\}$ and $\mu[\operatorname{top}_\lambda(W)] = \lambda$, where μ denotes the Lebesgue measure on $[0, 1]^2$. In either case, we have $M = \inf_{(x,y)\in\operatorname{top}_\lambda(W)} W(x, y)$ and $W(u, v) \le M$ whenever $(u, v) \notin \operatorname{top}_\lambda(W)$.

It is easy to see that W and the indicator function $\mathbf{1}_{\operatorname{top}_\lambda(W)}$ are aligned, implying that $\mathsf{E}\Big[W\mathbf{1}_{\operatorname{top}_\lambda(W)}\Big] = \mathsf{E}\Big[W^*\mathbf{1}_{\operatorname{top}_\lambda}(W)^*\Big]$. Consider now the L^1 graphon $Y = \mathbf{1}_{\operatorname{top}_\lambda(W)}$. With the help of (5.15) and the fact that $\mathsf{E}[WY] = E[W^*Y^*]$ we have

$$\mathsf{E}[WY] = \mathcal{C}(W, Y) = \mathcal{C}(U, Y) \le \mathsf{E}[U^*Y^*].$$

Let $\tilde{Y} = \mathbf{1}_{\operatorname{top}_\lambda(U)}$. Then Y and $\tilde{Y}$ have the same distribution, implying that $\tilde{Y}^* = Y^*$. On the other hand, $\tilde{Y}$ and U are aligned, implying that $\mathsf{E}[U\tilde{Y}] = \mathsf{E}[U^*\tilde{Y}^*]$. Putting everything together, we conclude that

$$\mathsf{E}[WY] \le \mathsf{E}[U\tilde{Y}].$$

In a similar way, we show that $\mathsf{E}[U\tilde{Y}] \le \mathsf{E}[WY]$. We thus have shown that for all $\lambda \in [0, 1]$,

$$\mathsf{E}[W\mathbf{1}_{\operatorname{top}_\lambda(W)}] = \mathsf{E}[U\mathbf{1}_{\operatorname{top}_\lambda(U)}].$$

This in turn implies that W and U have the same distribution. □

5.5.2. *Proof of Theorem 5.13.* Theorem 5.13 follows immediately from Lemma 5.14 and the following proposition. We remark that when W is bounded, or even in L^2, the proof of the proposition is much easier as one can consider $\mathcal{C}(W,W) = \mathsf{E}[W^2]$. This does not work when W is only assumed to be in L^1. The proof begins by transforming W into a bounded graphon. In what follows, we use the metric

$$\delta_p(U,W) = \inf_{\phi} \|U - W^{\phi}\|_p,$$

where the infimum is over all measure-preserving bijections $\phi\colon [0,1] \to [0,1]$. It clearly satisfies $\delta_{\square}(U,W) \le \delta_1(U,W)$.

Proposition 5.16. *If U and W are graphons such that $\mathcal{C}(U,Y) = \mathcal{C}(W,Y)$ for all L^∞ graphons Y, then $\delta_1(U,W) = 0$.*

Proof. We know that U and W have the same distribution. Let $\widetilde{W} = \arctan W$ and $\widetilde{U} = \arctan U$ (note that $\widetilde{W}$ and $\widetilde{U}$ are both bounded). Since both $\arctan x$ and $x - \arctan x$ are increasing in x, for every measure preserving bijection $\sigma\colon [0,1] \to [0,1]$ the rearrangement inequality implies that

$$\mathsf{E}[(W - \widetilde{W})\widetilde{W}] \ge \mathsf{E}[(U - \widetilde{U})^{\sigma}\widetilde{W}]$$

and

$$\mathsf{E}[\widetilde{W}^2] \ge \mathsf{E}[\widetilde{U}^{\sigma}\widetilde{W}].$$

Thus

$$\mathsf{E}[W\widetilde{W}] - \mathsf{E}[U^{\sigma}\widetilde{W}] \ge \mathsf{E}[\widetilde{W}^2] - \mathsf{E}[\widetilde{U}^{\sigma}\widetilde{W}] \ge 0, \tag{5.16}$$

By assumption we have $\mathcal{C}(U,\widetilde{W}) = \mathcal{C}(W,\widetilde{W})$, which must equal $\mathsf{E}[W\widetilde{W}]$ by the rearrangement inequality. Taking the infimum over all σ in (5.16) and using the facts from the previous sentence yields $\mathcal{C}(\widetilde{W},\widetilde{W}) = \mathcal{C}(\widetilde{U},\widetilde{W})$. A similar argument shows that $\mathcal{C}(\widetilde{U},\widetilde{U}) = \mathcal{C}(\widetilde{U},\widetilde{W})$. Therefore

$$\begin{aligned}\delta_2(\widetilde{U},\widetilde{W})^2 = \inf_{\sigma} \mathsf{E}[(\widetilde{U}^{\sigma} - \widetilde{W})^2] &= \mathsf{E}[\widetilde{U}^2] + \mathsf{E}[\widetilde{W}^2] - 2\sup_{\sigma} \mathsf{E}[\widetilde{U}^{\sigma}\widetilde{W}] \\ &= \mathcal{C}(\widetilde{U},\widetilde{U}) + \mathcal{C}(\widetilde{W},\widetilde{W}) - 2\mathcal{C}(\widetilde{U},\widetilde{W}) = 0.\end{aligned}$$

Since $\delta_1(\widetilde{U},\widetilde{W}) \le \delta_2(\widetilde{U},\widetilde{W})$, we must have $\delta_1(\widetilde{U},\widetilde{W}) = 0$ as well.

Finally we need to deduce that $\delta_1(U,W) = 0$. Let $K > 0$. From the mean value theorem, we know that

$$|x - y| \le (1 + K^2)\, |\arctan x - \arctan y|$$

whenever $x, y \in [-K, K]$. It follows that for every σ,

$$\|(U^{\sigma} - W)\mathbf{1}_{|U^{\sigma}| \le K, |W| \le K}\|_1 \le (1 + K^2)\|\widetilde{U}^{\sigma} - \widetilde{W}\|_1.$$

The left side differs from $\|U^{\sigma} - W\|_1$ by at most $4\left\|U\mathbf{1}_{|U|>K}\right\|_1$ (here we use the triangle inequality, and the fact that U and W have the same distribution, so that $\left\|W\mathbf{1}_{|U^{\sigma}|>K}\right\|_1 \le \left\|U\mathbf{1}_{|U|>K}\right\|_1$ and $\left\|U^{\sigma}\mathbf{1}_{|W|>K}\right\|_1 \le \left\|U\mathbf{1}_{|U|>K}\right\|_1$ by the rearrangement inequality). Thus

$$\|U^{\sigma} - W\|_1 \le (1 + K^2)\|\widetilde{U}^{\sigma} - \widetilde{W}\|_1 + 4\left\|U\mathbf{1}_{|U|>K}\right\|_1.$$

Taking the infimum over σ and using $\delta_1(\widetilde{U},\widetilde{W}) = 0$, we find that $\delta_1(U,W) \le 4\left\|U\mathbf{1}_{|U|>K}\right\|$. Since K can be made arbitrarily large, $\delta_1(U,W) = 0$. □

5.6. **Proof of (i)⇒(iv)⇒(iii) in Theorem 5.3.** The following result gives the implication (i)⇒(iv) in Theorem 5.3 as well as the statement that $\mathcal{F}_{\mathbf{a}}(W_n, J)$ converges whenever W_n converges in metric.

Proposition 5.17. *Let $q \in \mathbb{N}$, $\mathbf{a} \in \triangle_q$, $\varepsilon > 0$, and $h \in \mathbb{R}^q$, and let $J \in \mathbb{R}^{q\times q}$ be a symmetric matrix. For any two graphons U and W,*

$$\Big|\mathcal{F}_{\mathbf{a}}(U, J) - \mathcal{F}_{\mathbf{a}}(W, J)\Big| \leq \|J\|_1 \delta_\square(U, W)$$

and

$$\Big|\mathcal{F}_{\mathbf{a},\varepsilon}(U, J) - \mathcal{F}_{\mathbf{a},\varepsilon}(W, J)\Big| \leq \|J\|_1 \delta_\square(U, W).$$

Proof. Since the left sides of the above bounds do not change if we replace U by U^ϕ for a measure preserving bijection $\phi\colon [0, 1] \to [0, 1]$, it is enough to prove the lemma with a bound in terms of $\|U - W\|_\square$ instead of $\delta_\square(U, W)$. The result then follows immediately from (5.11) and the definitions (2.18) and (5.3). □

Next we show that convergence of $\mathcal{F}_{\mathbf{a}}(W_n, J)$ for all J implies convergence of $\mathcal{E}_{\mathbf{a}}(W_n, J)$. Together with our results from the last section, this shows that convergence of $\mathcal{F}_{\mathbf{a}}(W_n, J)$ for all $q \in \mathbb{N}$, $\mathbf{a} \in \triangle_q$, and symmetric $J \in \mathbb{R}^{q\times q}$ is sufficient for metric convergence, which concludes the proof of Remark 5.4(ii).

Lemma 5.18. *Let $q \in \mathbb{N}$, $\mathbf{a} \in \triangle_q$, and $c > 0$, let $J \in \mathbb{R}^{q\times q}$ be symmetric, and let U and W be two graphons. Then*

$$\big|\mathcal{E}_{\mathbf{a}}(W, J) - \mathcal{E}_{\mathbf{a}}(U, J)\big| \leq \frac{1}{c}\big|\mathcal{F}_{\mathbf{a}}(W, cJ) - \mathcal{F}_{\mathbf{a}}(U, cJ)\big| + \frac{2\log q}{c}.$$

Proof. Using the fact that $\mathsf{Ent}(\rho) \leq \log q$, we get by (2.17) and (2.18)

$$\big|\mathcal{E}_{\mathbf{a}}(W, J) - \mathcal{F}_{\mathbf{a}}(W, J)\big| \leq \log q,$$

for all J, and similarly for U. Hence

$$\begin{aligned}\big|\mathcal{E}_{\mathbf{a}}(W, J) - \mathcal{E}_{\mathbf{a}}(U, J)\big| &= \frac{1}{c}\big|\mathcal{E}_{\mathbf{a}}(W, cJ) - \mathcal{E}_{\mathbf{a}}(U, cJ)\big| \\ &\leq \frac{1}{c}\Big(\big|\mathcal{F}_{\mathbf{a}}(W, cJ) - \mathcal{F}_{\mathbf{a}}(U, cJ)\big| + 2\log q\Big),\end{aligned}$$

which proves the claim. □

Proof of (iv)⇒(iii) in Theorem 5.3. As in the proof above, one sees that

$$\frac{1}{c}\mathcal{F}_{\mathbf{a},\varepsilon}(W_n, cJ) - \frac{\log q}{c} \leq \mathcal{E}_{\mathbf{a},\varepsilon}(W_n, J) \leq \frac{1}{c}\mathcal{F}_{\mathbf{a},\varepsilon}(W_n, cJ) + \frac{\log q}{c}.$$

But this clearly shows that (iv)⇒(iii) in Theorem 5.3. □

5.7. **Quantitative bounds on distance between fractional quotients.** In this section, we prove a quantitative bound on the distance between two different quotients of the same graphon, which will be used in the next section.

We define the L^1 distance in FP_q to be

$$d_1(\rho, \rho') = \sum_{i\in[q]} \int_0^1 |\rho_i(x) - \rho'_i(x)|\, dx; \tag{5.17}$$

note that $d_1(\rho, \rho') \leq 2$. We also need the definition of K-bounded tails from (5.2).

Lemma 5.19. *Let $q \in \mathbb{N}$, and let W be a graphon with K-bounded tails for some function $K\colon (0,\infty) \to (0,\infty)$. Then there exists a weakly increasing function $\varepsilon_K : [0,2] \to [0,\infty)$ such that $\varepsilon_K(x) \to 0$ as $x \to 0$ and*

$$d_1(W/\rho, W/\rho') \le \varepsilon_K(d_1(\rho,\rho'))$$

for all $\rho, \rho' \in \mathsf{FP}_q$. For $\mathbf{a}, \mathbf{b} \in \triangle_q$,

$$d_1^{\mathrm{Hf}}\Big(\widehat{\mathcal{S}}_{\mathbf{a}}(W), \widehat{\mathcal{S}}_{\mathbf{b}}(W)\Big) \le \varepsilon_K(\|\mathbf{a}-\mathbf{b}\|_1), \tag{5.18}$$

where $\|\mathbf{a}-\mathbf{b}\|_1 = \sum_i |a_i - b_i|$.

Proof. Fix $\varepsilon > 0$. Clearly

$$\sum_i |\alpha_i(\rho) - \alpha_i(\rho')| \le \sum_i \int_0^1 |\rho_i(x) - \rho_i'(x)|\,dx = d_1(\rho,\rho').$$

On the other hand, using the fact that $\sum_i \rho_i(x) = \sum_j \rho_i'(x) = 1$ for all $x \in [0,1]$, we have

$$\begin{aligned}
\sum_{i,j}\Big|\beta_{ij}(W/\rho) - \beta_{ij}(W/\rho')\Big| &= \sum_{i,j}\Big|\int_{[0,1]^2} W(x,y)\Big(\rho_i(x)\rho_j(y) - \rho_i'(x)\rho_j'(y)\Big)\,dx\,dy\Big| \\
&\le \int_{[0,1]^2} |W(x,y)| \sum_{i,j}\Big|\rho_i(x)\rho_j(y) - \rho_i(x)\rho_j'(y)\Big|\,dx\,dy \\
&\quad + \int_{[0,1]^2} |W(x,y)| \sum_{i,j}\Big|\rho_i(x)\rho_j'(y) - \rho_i'(x)\rho_j'(y)\Big|\,dx\,dy \\
&= 2\int_{[0,1]^2} |W(x,y)| \sum_i \Big|\rho_i(x) - \rho_i'(x)\Big|\,dx\,dy \\
&\le 2\int_{[0,1]^2} K(\varepsilon/8) \sum_i \Big|\rho_i(x) - \rho_i'(x)\Big|\,dx\,dy \\
&\quad + 4\int_{[0,1]^2} |W(x,y)| \mathbf{1}_{|W(x,y)| \ge K(\varepsilon/8)}\,dx\,dy \\
&\le 2K(\varepsilon/8) d_1(\rho,\rho') + \varepsilon/2,
\end{aligned}$$

showing that

$$d_1(W/\rho, W/\rho') \le (1 + 2K(\varepsilon/8)) d_1(\rho,\rho') + \varepsilon/2,$$

which is at most ε provided that $d_1(\rho,\rho') \le \varepsilon/(2 + 4K(\varepsilon/8))$. Since $\varepsilon > 0$ was arbitrary, this immediately implies the existence of the desired function ε_K.

The claim (5.18) follows from noting that for any $\mathbf{a}, \mathbf{b} \in \triangle_q$ and $\rho \in \mathsf{FP}_q$ with $\alpha(\rho) = \mathbf{a}$, we can find a $\rho' \in \mathsf{FP}_q$ with $\alpha(\rho') = \mathbf{b}$ such that $d_1(\rho,\rho') = \|\mathbf{a}-\mathbf{b}\|_1$. (In fact, we can choose ρ' so that $\rho_i \le \rho_i'$ if and only if $a_i \le b_i$.) □

The above lemma can be used to show that $\mathcal{E}_{\mathbf{a}}(W,J)$ is continuous in $\mathbf{a}$. The continuity of $\mathcal{F}_{\mathbf{a}}(W,J)$ then follows from noting that $\mathsf{Ent}(\rho)$ is uniformly continuous in $\rho \in \mathsf{FP}_q$ with respect to d_1 (as $-x\log x$ is continuous on $[0,1]$). More explicitly, we have the following lemma.

Lemma 5.20. *Let $q \in \mathbb{N}$. Then the function $\mathsf{Ent}\colon \mathsf{FP}_q \to [0,\log q]$ with $\rho \mapsto \mathsf{Ent}(\rho)$ is uniformly continuous in the metric d_1 defined in (5.17). Explicitly,*

$$|\,\mathsf{Ent}(\rho) - \mathsf{Ent}(\rho')| \le q\tilde{f}\Big(\frac{1}{q} d_1(\rho,\rho')\Big),$$

where $\tilde{f}(x) = x(1 - \log x)$.

Proof. Because the function $f(x) = -x \log x$ is concave, for each $t > 0$ the function $x \mapsto f(x) - f(x+t)$ is weakly increasing. It follows from this and $f(0) = f(1) = 0$ that for any $x, y \in [0,1]$,

$$\begin{aligned}
|f(x) - f(y)| &\leq \max\{|f(0) - f(|x-y|)|, |f(1 - |x-y|) - f(1)|\} \\
&= \max\{f(|x-y|), f(1 - |x-y|)\} \\
&\leq f(|x-y|) + f(1 - |x-y|) \\
&\leq f(|x-y|) + |x-y| \\
&= \tilde{f}(|x-y|),
\end{aligned}$$

where the last inequality holds because $f(1-t) \leq t$ for all $t \in [0,1]$. Since $d_1(\rho, \rho') = \sum_i \int_{[0,1]} |\rho_i(x) - \rho_i'(x)|\, dx$ and $\tilde{f}$ is concave, we have

$$\begin{aligned}
\frac{1}{q}\Big|\mathsf{Ent}(\rho) - \mathsf{Ent}(\rho')\Big| &= \frac{1}{q}\Big|\sum_{i=1}^{q} \int_0^1 \Big(f(\rho_i(x)) - f(\rho_i'(x))\Big)\, dx\Big| \\
&\leq \frac{1}{q} \sum_{i=1}^{q} \int_0^1 \tilde{f}(|\rho_i(x) - \rho_i'(x)|)\, dx \\
&\leq \tilde{f}\Big(\frac{1}{q} d_1(\rho, \rho')\Big).
\end{aligned}$$

Because $\tilde{f}$ is continuous at 0, this completes the proof. □

6. Convergent sequences of uniformly upper regular graphs

In this section we prove Theorem 2.15. Theorem 2.10 will follow from this theorem and Theorem 2.17, which we prove in Section 7.

6.1. **Preliminaries.** We start by stating some of the results from [3], which will allow us to replace uniformly upper regular sequences of weighted graphs by sequences of weighted graphs with K-bounded tails. To state them, we will use the cut distance between two weighted graphs G, G' with identical node sets $V(G) = V(G') = V$ and identical node weights $\alpha_x = \alpha_x(G) = \alpha_x(G')$, defined as

$$d_\square(G, G') = \max_{S,T \subseteq V} \Big| \sum_{(x,y) \in S \times T} \frac{\alpha_x \alpha_y}{\alpha_G^2} \Big(\beta_{xy}(G) - \beta_{xy}(G')\Big) \Big|. \tag{6.1}$$

Note that this distance is equal to $\|W^G - W^{G'}\|_\square$, where W^G and $W^{G'}$ are the step functions defined in (2.2) and $\|\cdot\|_\square$ is the cut norm defined in (2.1). Indeed, for $W = W^G$, the supremum in (2.1) can easily be shown to be a maximum that is attained for sets S and T which are both unions of the intervals I_i.

We will also use the notion of an equipartition of the vertex set $V(G)$ of a weighted graph G, defined by requiring that the weights of the parts of the partition differ from an equal distribution by at most $\alpha_{\max}(G)$. Explicitly, a partition $\mathcal{P} = (V_1, \dots, V_k)$ of $V(G)$ is called an *equipartition* if $|\alpha_{V_i} - \frac{1}{k}\alpha_G| \leq \alpha_{\max}(G)$ for all $i \in [k]$. The following version of the weak regularity lemma was proved in [3].

Theorem 6.1 (Theorem C.12 in [3]). *Let $K\colon (0,\infty) \to (0,\infty)$ and $0 < \varepsilon < 1$. Then there exist constants $N = N(K,\varepsilon)$ and $\eta_0 = \eta_0(K,\varepsilon)$ such that the following holds* *for all $\eta \le \eta_0$: for every (K,η)-upper regular graph G and each natural number $k \ge N$, there exists a equipartition $\mathcal{P} = (V_1,\dots,V_k)$ of $V(G)$ into k parts, such that*

$$d_\square(G, G_\mathcal{P}) \le \varepsilon \|G\|_1.$$

As a consequence, given a K-upper regular sequence of weighted graphs G_n with no dominant nodes, we can find a sequence of equipartitions $\mathcal{P}_n$ of $V(G_n)$ into k_n classes such that the sequence of weighted graphs[6] $\widehat{G}_n = \frac{1}{\|G_n\|_1}(G_n)_{\mathcal{P}_n}$ satisfies

$$d_\square\Big(\frac{1}{\|G_n\|_1} G_n, \widehat{G}_n\Big) \to 0, \qquad k_n \frac{\alpha_{\max}(G_n)}{\alpha_{G_n}} \to 0, \qquad \|\widehat{G}_n\|_1 \le 1 \tag{6.2}$$

and

$$W^{\widehat{G}_n} \text{ has } K\text{-bounded tails.} \tag{6.3}$$

We call a sequence $(\widehat{G}_n)_{n\ge0}$ with these properties a *regularized version* of $(G_n)_{n\ge0}$, and $\mathcal{P}_n$ a *regularizing partition* for G_n.

To see that all these conditions can be simultaneously achieved, let $\eta_n \to 0$ be such that G_n is (K,η_n)-upper regular. Assume that ε_n goes to zero slowly enough that $\eta_n \le \eta_0(K,\varepsilon_n)$ and $\eta_n N(K,\varepsilon_n) \to 0$ in Theorem 6.1. Choosing $k_n = N(K,\varepsilon_n)$, the theorem then gives a sequence of equipartitions $\mathcal{P}_n$ of $V(G_n)$ into k_n classes such that (6.2) holds. The weight of each class of $\mathcal{P}_n$ is bounded from below by $\alpha_{G_n}/k_n - \alpha_{\max}(G_n)$, which is asymptotically greater than $\eta_n\alpha_{G_n}$ because $\eta_n k_n \to 0$ and $k_n\alpha_{\max}(G_n)/\alpha_{G_n} \to 0$. Thus, the bound (2.21) holds for $\mathcal{P} = \mathcal{P}_n$, establishing that

$$\sum_{x,y\in V(G_n)} \frac{\alpha_x(G_n)\alpha_y(G_n)}{\alpha_{G_n}^2} |\beta_{xy}(\hat{G}_n)| \mathbf{1}_{|\beta_{xy}(\hat{G}_n)|\ge K(\varepsilon)} \le \varepsilon$$

for all $\varepsilon > 0$. In other words, $W^{\widehat{G}_n}$ has K-bounded tails.

6.2. **Comparing sequence of graphs to sequences of graphons.** In this section we prove three lemmas, which are the main technical lemmas used to reduce many statements in Section 2 to those in Section 5. We define

$$\widehat{\mathcal{S}}_{\mathbf{a},\varepsilon}(W) = \{(\alpha,\beta) \in \widehat{\mathcal{S}}_q(W) : \|\alpha - \mathbf{a}\|_\infty \le \varepsilon\}.$$

Lemma 6.2. *Let G be a weighted graph, let $\mathcal{P}$ be an equipartition of $V(G)$ into k classes such that $G = G_\mathcal{P}$, and let $q \in \mathbb{N}$. Then there exist two maps $\phi \mapsto \rho_\phi$ and $\rho \to \bar\rho$ from the set of configurations $\phi\colon V(G) \to [q]$ into the set of fractional partitions FP_q and from FP_q to FP_q, respectively, such that the following hold:*

(i) *$W^G/\rho = W^G/\bar\rho$ for all $\rho \in \mathsf{FP}_q$.*
(ii) *$\|G\|_1(G/\phi) = W^G/\rho_\phi$ for all $\phi\colon V(G) \to [q]$.*
(iii) *For each $\rho \in \mathsf{FP}_q$ there exists a $\phi\colon V(G) \to \mathsf{FP}_q$ such that*

$$d_1(\rho_\phi, \bar\rho) \le qk\frac{\alpha_{\max(G)}}{\alpha_G}. \tag{6.4}$$

Proof. Let $\mathcal{P} = (V_1,\dots,V_k)$, and assume that the vertices in G are ordered in such a way that $V_1 = \{1,2,\dots,|V_1|\}$, $V_2 = \{|V_1|+1,\dots,|V_1|+|V_2|\}$, etc. Let $x_\mu = \alpha_{V_\mu}(G)/\alpha_G$ for $\mu \in [k]$, and let $I_1,\dots,I_k \subseteq [0,1]$ be consecutive intervals of length $x_1, \dots, x_k$.

[6] If $\|G_n\|_1 = 0$, we choose $\widehat{G}_n$ to have edge weights 0.

For $\rho \in \mathsf{FP}_q$ define $\bar\rho$ by averaging ρ over the intervals I_μ, i.e., $\bar\rho_i(x) = \frac{1}{x_\mu}\int_{I_\mu} \rho_i(y)\,dy$ if $x \in I_\mu$. Since W^G is constant on sets of the form $I_\mu \times I_\nu$ for $\mu, \nu \in [k]$, we clearly have $W^G/\rho = W^G/\bar\rho$, proving (i).

Next, given $\phi\colon V(G) \to [q]$, let ρ_ϕ be the fractional q-partition defined by

$$(\rho_\phi)_i(x) = \frac{1}{\alpha_{V_\mu(G)}} \sum_{u \in V_\mu(G)} \alpha_u(G)\mathbf{1}_{\phi(u)=i} \qquad \text{when } x \in I_\mu. \tag{6.5}$$

Using the fact that $\beta_{uv}(G)$ is constant on sets of the form $V_\mu \times V_\nu$, it is then easy to check that $\|G\|_1 G/\phi = W^G/\rho_\phi$, proving (ii).

To prove (iii), consider $\rho \in \mathsf{FP}_q$, and let $\alpha_{\mu,i} = \int_{I_\mu} \rho_i(x)\,dx$. We then decompose V_μ into q sets $V_{\mu,i}$ such that

$$|\alpha_{\mu,i} - \alpha_{V_{\mu,i}}(G)/\alpha_G| \le \frac{\alpha_{\max}(G)}{\alpha_G}$$

for all i and μ. Setting $\phi = i$ on $\bigcup_\mu V_{\mu,i}$, we then get a map $\phi\colon V(G) \to [q]$ such that (6.4) holds. □

Lemma 6.3. *Let $q \in \mathbb{N}$ and let $(G_n)_{n\ge 0}$ be a uniformly upper regular sequence of weighted graphs. If $\widehat{G}_n$ is a regularized version of G_n, then*

$$d_1^{\mathrm{Hf}}\Big(\mathcal{S}_q(G_n), \widehat{\mathcal{S}}_q(W^{\widehat{G}_n})\Big) \to 0. \tag{6.6}$$

Proof of Lemma 6.3. We start by showing that

$$d_1(G_n/\phi, \|\widehat{G}_n\|_1(\widehat{G}_n/\phi)) \le q^2 d_\square\left(\frac{1}{\|G_n\|_1}G_n, \widehat{G}_n\right) \tag{6.7}$$

whenever $\widehat{G}_n$ is a regularized version of G_n. Indeed, from the definition of the cut distance (6.1) it is easy to see that $d_1(G/\phi, G'/\phi) \le q^2 d_\square(\frac{1}{\|G\|_1}G, \frac{1}{\|G'\|_1}G')$ whenever G and G' have identical node sets and node weights and for any $\phi\colon V(G) \to [q]$. More generally, for any $\lambda \ge 0$, we have $d_1(G/\phi, \lambda(G'/\phi)) \le q^2 d_\square(\frac{1}{\|G\|_1}G, \frac{\lambda}{\|G'\|_1}G')$. Setting $\lambda_n = \|\widehat{G}_n\|_1$ proves (6.7).

Next we observe that by Lemma 6.2(ii),

$$\|\widehat{G}_n\|_1 \mathcal{S}_q(\widehat{G}_n) \subseteq \widehat{\mathcal{S}}_q(W^{\widehat{G}_n}). \tag{6.8}$$

On the other hand, given $W^{\widehat{G}_n}/\rho \in \widehat{\mathcal{S}}_q(W^{\widehat{G}_n})$, we may use Lemma 6.2 and Lemma 5.19 to find a quotient $\widehat{G}_n/\phi \in \mathcal{S}_q(\widehat{G}_n)$ such that

$$\begin{aligned} d_1\Big(W^{\widehat{G}_n}/\rho, \|\widehat{G}_n\|_1\widehat{G}_n/\phi\Big) &= d_1(W^{\widehat{G}_n}/\bar\rho, W^{\widehat{G}_n}/\rho_\phi) \\ &\le \varepsilon_K\left(qk_n\frac{\alpha_{\max(G_n)}}{\alpha_{G_n}}\right), \end{aligned} \tag{6.9}$$

where k_n is the number of classes in the regularizing partition corresponding to $\widehat{G}_n$ and K is the function from (6.3).

Taking into account the second bound in (6.2), the bound (6.9) together with (6.8) implies that

$$d_1^{\mathrm{Hf}}\Big(\|\widehat{G}_n\|_1\mathcal{S}_q(\widehat{G}_n), \widehat{\mathcal{S}}_q\Big(W^{\widehat{G}_n}\Big)\Big) \to 0,$$

while (6.7) and the fact that $d_\square(\frac{1}{\|G_n\|_1}G_n, \widehat{G}_n) \to 0$ show that

$$d_1^{\mathrm{Hf}}(\mathcal{S}_q(G_n), \|\widehat{G}_n\|_1 \mathcal{S}_q(\widehat{G}_n)) \to 0.$$

Together, these two bounds imply (6.6). □

6.3. **Proof of (i)⇒(ii) and (ii)⇒(iii) in Theorem 2.10.** Recall that by Theorem 2.14, convergence in metric implies uniform upper regularity. As a consequence, (i)⇒(ii) follows immediately from Lemma 6.3. Indeed, $\delta_\square\Big(\frac{1}{\|G_n\|_1}W^{G_n}, W\Big) \to 0$ by (6.2), which implies that $\delta_\square(W^{\widehat{G}_n}, W) \to 0$. We then use the bound (6.6) from Lemma 6.3 in conjunction with (5.8) from Lemma 5.11 to conclude that

$$d_1^{\mathrm{Hf}}\Big(\mathcal{S}_q(G_n), \widehat{\mathcal{S}}_q(W)\Big) \to 0,$$

completing the proof of (i)⇒(ii) in Theorem 2.10.

The implication (ii)⇒(iii) follows with the help of Theorem 3.3, the closedness of $\widehat{\mathcal{S}}_q(W)$, and the representation (5.7).

6.4. **Proof of the equivalence (i)⇔(ii)⇔(iii) in Theorem 2.15.** Recall that by Theorem 2.13, for a uniformly upper regular sequence convergence in metric implies convergence in metric to some graphon W, so the above proof of the implication (i)⇒(ii) for Theorem 2.10 also proves it for Theorem 2.15. The implication (ii)⇒(iii) follows from Theorem 3.3.

It remains to show (iii)⇒(i); i.e., under the assumption of uniform upper regularity, convergence of the microcanonical ground state energies implies convergence in metric. Assume for the sake of contradiction that G_n does not converge in metric. By Theorem 2.13 this implies that there are two subsequences G'_n and G''_n and two graphons U and W such that $G'_n \to U$ and $G''_n \to W$ in metric while $\delta_\square(U, W) > 0$. By the already proved (i)⇒(iii) in Theorem 2.10, the microcanonical ground state energies of G'_n and G''_n converge to those of U and W, and by our assumption that G_n has convergent ground state energies, this implies that U and W have identical ground state energies, contradicting Theorem 5.13.

6.5. **Convergence in metric implies LD convergence.** Our main result in this section is the following theorem, which by (5.6) proves the implication (i)⇒(iv) in Theorem 2.10 and hence also Theorem 2.15.

Theorem 6.4. *Let $q \in \mathbb{N}$ and let $(G_n)_{n\geq 0}$ be a uniformly upper regular sequence of weighted graphs such that G_n converges to a graphon W in metric and G_n has vertex weights one. Then the limit* (2.16) *exists with*

$$I_q((\alpha,\beta)) = \log q - \lim_{\varepsilon\to 0} \sup_{\substack{\rho \in \mathsf{FP}_q \\ d_1(W/\rho,(\alpha,\beta)) \leq \varepsilon}} \mathsf{Ent}(\rho).$$

Proof. To prove the theorem, we will need to calculate probabilities of the form $\mathcal{P}_{q,G}\big[d_1((\alpha,\beta), G/\phi) \leq \varepsilon\big]$ for graphs G that are near to W in the normalized cut distance. Assume for the moment that G is well approximated by a weighted graph whose edge weights are constant over large blocks. More precisely, assume that there exists an equipartition $\mathcal{P} = (V_1, \dots, V_k)$ such that G and $G_\mathcal{P}$ are close in the cut norm. Under such a condition, the quotients of G are close to those of $G_\mathcal{P}$, provided they are suitably normalized. More precisely, if we define $\widehat{G}$ to be the normalized weighted graph $\widehat{G} = \frac{1}{\|G\|_1}(G)_\mathcal{P}$, the quotients of G are close to those of $\widehat{G}$ multiplied by $\|G\|_1$ (see (6.7)). Consider thus the probabilities $\mathcal{P}_{q,G}\big[d_1((\alpha,\beta),\|G\|_1(\widehat{G}/\phi)) \le \varepsilon\big]$.

Since the edge weights of $\widehat{G}$ are constant on sets of the form $V_\mu \times V_\nu$, we have that $\widehat{G}/\phi = \widehat{G}/\phi'$ if for all $i \in [q]$ and all $\mu \in [k]$, the number of vertices in V_μ that are mapped to $i \in [q]$ is the same in ϕ and ϕ'. Denote this number by $k_{i,\mu} = k_{i,\mu}(\phi)$. The number of configurations $\phi\colon V(G) \to [q]$ with given $k_{i,\mu}$ is

$$N(\{k_{i,\mu}\}) = \prod_{\mu=1}^{k} \frac{n_\mu!}{k_{1,\mu}! \dots k_{q,\mu}!}, \tag{6.10}$$

where $n_\mu = |V_\mu|$. Approximating $k!$ as $(k/e)^k$ (we analyze the error term below), we have

$$N(\{k_{i,\mu}\}) \approx \exp\Big(-\sum_{\mu=1}^{k}\sum_{i=1}^{q} -k_{i,\mu}\log\Big(\frac{k_{i,\mu}}{n_\mu}\Big)\Big) = e^{|V(G)|\,\mathsf{Ent}(\rho_\phi)},$$

where ρ_ϕ is the fractional partition defined in (6.5). Observing that the number of choices for $\{k_{\mu,i}\}$ is polynomial in $|V(G)|$, and hence will not contribute to I_q (again we bound the error later), and noting further that $\|\widehat{G}\|_1(\widehat{G}/\phi) = W^{\widehat{G}}/\rho_\phi$ by Lemma 6.2, we then approximate $\mathcal{P}_{q,G}\big[d_1((\alpha,\beta),\|G\|_1(\widehat{G}/\phi)) \le \varepsilon\big]$ by

$$q^{-|V(G)|} \max_{\substack{\phi\colon V(G)\to[q]\\ d_1((\alpha,\beta),W^{\widehat{G}}/\rho_\phi)\le\varepsilon}} e^{|V(G)|\,\mathsf{Ent}(\rho_\phi)}.$$

Taking into account that (again by Lemma 6.2) any fractional quotient $W^{\widehat{G}}/\rho$ can be well approximated by a quotient of the form $W^{\widehat{G}}/\rho_\phi$, we obtain the theorem.

The formal proof proceeds as follows. Let $K : (0,\infty) \to (0,\infty)$ and $(\eta_n)_{n\ge0}$ be such that $\lim_{n\to\infty}\eta_n = 0$ and G_n is (K,η_n)-upper regular. Fix $\varepsilon > 0$. By Theorem 6.1 and Definition 2.12, there are constants $k \in \mathbb{N}$ and $n_0 < \infty$ such that for each $n \ge n_0$ there exists an equipartition $\mathcal{P} = \mathcal{P}_n = (V_1,\dots,V_k)$ of $V(G_n)$ into k parts such that

$$q^2 d_\square\Big(G_n,(G_n)_{\mathcal{P}}\Big) \le \frac{\varepsilon}{2}\|G_n\|_1$$

and $W^{\widehat{G}_n}$ with $\widehat{G}_n = \frac{1}{\|G_n\|_1}(G_n)_{\mathcal{P}}$ has K-bounded tails.

Let $G = G_n$ and $\widehat{G} = \widehat{G}_n$. By the bound (6.7),

$$d_1(G/\phi,\|\widehat{G}\|_1(\widehat{G}/\phi)) \le q^2 d_\square\left(\frac{1}{\|G\|_1}G,\widehat{G}\right) \le \frac{\varepsilon}{2},$$

implying that

$$\begin{aligned}\mathcal{P}_{q,G}\big[d_1((\alpha,\beta),\|\widehat{G}\|_1(\widehat{G}/\phi)) \le \varepsilon/2\big] &\le \mathcal{P}_{q,G}\big[d_1((\alpha,\beta),G/\phi) \le \varepsilon\big]\\ &\le \mathcal{P}_{q,G}\big[d_1((\alpha,\beta),\|\widehat{G}\|_1(\widehat{G}/\phi)) \le 3\varepsilon/2\big].\end{aligned}$$

Given a configuration $\phi\colon V(G) \to [q]$, let $k_{i,\mu}(\phi)$ be the number of vertices $v \in V_\mu$ such that $\phi(v) = i$, and let $N(\{k_{i,\mu}\})$ be the number of maps ϕ leading to the same $k_{i,\mu}$; see (6.10) above. Bounding the number of choices for $\{k_{\mu,i}\}$ by $|V(G)|^{qk}$ and observing that $\|\widehat{G}\|_1(\widehat{G}/\phi) = W^{\widehat{G}}/\rho_\phi$, we then bound

$$\max_{\phi: d_1((\alpha,\beta), W^{\widehat{G}}/\rho_\phi)\le\varepsilon/2} N(\{k_{i,\mu}(\phi)\}) \le q^{|V(G)|}\mathcal{P}_{q,G}\big[d_1((\alpha,\beta),G/\phi)\le\varepsilon\big]$$
$$\le |V(G)|^{qk} \max_{\phi: d_1((\alpha,\beta), W^{\widehat{G}}/\rho_\phi)\le 3\varepsilon/2} N(\{k_{i,\mu}(\phi)\}).$$

Since $(k/e)^k \le k! \le ek(k/e)^k$,

$$\Big(\frac{1}{e|V(G)|}\Big)^{kq} e^{\mathsf{Ent}(\rho_\phi)|V(G)|} \le N(\{k_{i,\mu}\}) \le (e|V(G)|)^k e^{\mathsf{Ent}(\rho_\phi)|V(G)|},$$

implying

$$\begin{aligned} q^{|V(G)|}&\mathcal{P}_{q,G}\big[d_1((\alpha,\beta),G/\phi)\le\varepsilon\big] \\ &\le (e|V(G)|)^{k(q+1)} \max_{\substack{\phi:\, V(G)\to[q]\\ d_1((\alpha,\beta),W^{\widehat{G}}/\rho_\phi)\le 3\varepsilon/2}} e^{|V(G)|\,\mathsf{Ent}(\rho_\phi)} \\ &\le (e|V(G)|)^{k(q+1)} \sup_{\substack{\rho\in\mathsf{FP}_q\\ d_1((\alpha,\beta),W^{\widehat{G}}/\rho)\le 3\varepsilon/2}} e^{|V(G)|\,\mathsf{Ent}(\rho)} \end{aligned}$$

and

$$\begin{aligned} q^{|V(G)|}&\mathcal{P}_{q,G}\big[d_1((\alpha,\beta),G/\phi)\le\varepsilon\big] \\ &\ge (e|V(G)|)^{-kq} \max_{\substack{\phi:\, V(G)\to[q]\\ d_1((\alpha,\beta),W^{\widehat{G}}/\rho_\phi)\le \varepsilon/2}} e^{|V(G)|\,\mathsf{Ent}(\rho_\phi)}. \end{aligned}$$

Next we use Lemma 6.2 to approximate an arbitrary fractional partition $\rho \in \mathsf{FP}_q$ by a fractional partition of the form ρ_ϕ, with a error of $qk/|V(G)|$ in the d_1 distance. With the help of Lemmas 5.19 and 5.20, we can ensure that for n (and hence $|V(G)| = |V(G_n)|$) large enough, the resulting errors in $d_1((\alpha,\beta), W^{\widehat{G}}/\rho)$ and $\mathsf{Ent}(\rho)$ are bounded by $\varepsilon/4$ and ε, respectively, leading to the lower bound

$$\begin{aligned} q^{|V(G)|}&\mathcal{P}_{q,G}\big[d_1((\alpha,\beta),G/\phi)\le\varepsilon\big] \\ &\ge (e|V(G)|)^{-kq} \sup_{\substack{\rho\in\mathsf{FP}_q\\ d_1((\alpha,\beta),W^{\widehat{G}}/\rho)\le \varepsilon/4}} e^{|V(G)|(\mathsf{Ent}(\rho)-\varepsilon)}. \end{aligned}$$

To conclude the proof, we note that if $G_n \to W$ in metric, then $\delta_\square(W^{\widehat{G}_n}, W) \to 0$. Taking into account Lemma 5.19 and the fact that the entropy $\mathsf{Ent}(\rho)$ is invariant under measure preserving transformations, we get that for n sufficiently large

$$\begin{aligned} (e|V(G_n)|)^{-kq} &\sup_{\substack{\rho\in\mathsf{FP}_q\\ d_1((\alpha,\beta),W/\rho)\le \varepsilon/8}} e^{|V(G_n)|(\mathsf{Ent}(\rho)-\varepsilon)} \\ &\le q^{|V(G_n)|}\mathcal{P}_{q,G_n}\big[d_1((\alpha,\beta),G_n/\phi)\le\varepsilon\big] \\ &\le (e|V(G)|)^{k(q+1)} \sup_{\substack{\rho\in\mathsf{FP}_q\\ d_1((\alpha,\beta),W/\rho)\le 2\varepsilon}} e^{|V(G)|\,\mathsf{Ent}(\rho)}. \end{aligned}$$

As a consequence

$$\begin{aligned}
-\log q - \varepsilon + \sup_{\substack{\rho\in\mathsf{FP}_q\\ d_1((\alpha,\beta),W/\rho)\le\varepsilon/8}} \mathsf{Ent}(\rho) &\le \liminf_{n\to\infty} \frac{\log \mathcal{P}_{q,G_n}\big[d_1((\alpha,\beta),G_n/\phi)\le\varepsilon\big]}{|V(G_n)|}\\
&\le \limsup_{n\to\infty} \frac{\log \mathcal{P}_{q,G_n}\big[d_1((\alpha,\beta),G_n/\phi)\le\varepsilon\big]}{|V(G_n)|}\\
&\le -\log q + \sup_{\substack{\rho\in\mathsf{FP}_q\\ d_1((\alpha,\beta),W/\rho)\le 2\varepsilon}} \mathsf{Ent}(\rho).
\end{aligned}$$

Sending $\varepsilon\to 0$ completes the proof. □

6.6. **Completion of the proofs of Theorems 2.15 and 2.10.** To complete the proof of Theorem 2.15, we still need to show for graphs with node weights one, statements (iv) and (v) are equivalent to the other statements of the theorem. We will also have to establish the limit expressions given in Theorem 2.10.

By Theorem 6.4, we know the implication (i)⇒(iv) in Theorem 2.15, and we also know that the rate function is given by (2.19), as claimed in Theorem 2.10(iv). Finally, by Theorem 2.9(ii), in Theorem 2.15 statement (iv) implies statement (v), and by Lemma 3.2(iii), this in turn implies statement (iii) of Theorem 2.15, completing the proof of Theorem 2.15.

Theorem 2.17, which we prove in the next section, is nearly enough to deduce Theorem 2.10 from Theorem 2.15. The only missing piece is the explicit limit expressions stated in Theorem 2.10 for how limiting quotients, ground state energies, free energies, and large deviations rate function depend on W. So far, we have dealt with all of them except for the microcanonical free energies. Since we already have shown that convergence in metric for graphs with node weights one implies LD convergence with rate function given by (2.19), this follows from the following lemma.

Lemma 6.5. *Let W be a graphon, and let $(G_n)_{n\ge0}$ be a sequence of weighted graphs. If $(G_n)_{n\ge0}$ is LD convergent with rate function $I_q = I_q(\cdot, W)$ as defined in* (2.19)*, then the microcanonical free energies of $(G_n)_{n\ge0}$ converge to those of W, as defined in* (2.18)*.*

Proof. By Theorem 2.9(ii), the assumption implies convergence of the microcanonical free energies, with the limiting free energies given by

$$\begin{aligned}
F_{\mathbf{a}}(J) &= \inf_{(\alpha,\beta)\in\widehat{\mathcal{S}}_{\mathbf{a}}} \big(-\langle\beta,J\rangle + I_q((\alpha,\beta),W)\big) - \log q.\\
&= \inf_{(\alpha,\beta)\in\widehat{\mathcal{S}}_{\mathbf{a}}} \inf_{\substack{\rho\in\mathsf{FP}_q\\ W/\rho=(\alpha,\beta)}} \big(-\langle\beta,J\rangle - \mathsf{Ent}(\rho)\big)\\
&= \inf_{\substack{\rho\in\mathsf{FP}_q\\ \alpha(\rho)=\mathbf{a}}} \big(-\langle\beta(W/\rho),J\rangle - \mathsf{Ent}(\rho)\big)\\
&= \mathcal{F}_{\mathbf{a}}(W,J),
\end{aligned}$$

as desired. □

7. Inferring uniform upper regularity

In this section we prove Theorem 2.17. We have already proved a number of implications between the four conditions in the theorem statement. Specifically, from

Lemma 6.5 we know that (iv) implies (iii). From Lemma 3.2(iii) (and the analogous assertion that $\mathcal{E}_{\mathbf{a}}(W, J) = \lim_{\lambda\to\infty} \lambda^{-1}\mathcal{F}_{\mathbf{a}}(\lambda J)$ with an essentially identical proof) we know that (iii) implies (i). From Theorem 3.3 we deduce that (ii) implies (i). Thus, it remains to show that (i) implies that $(G_n)_{n\geq 0}$ is uniformly upper regular, which is the statement of the following proposition.

Proposition 7.1. *Let $(G_n)_{n\geq 0}$ be a sequence of weighted graphs with no dominant nodes and W a graphon. If the microcanonical ground state energies of G_n converge to those of W, then $(G_n)_{n\geq 0}$ is uniformly upper regular.*

In order to prove this proposition, we introduce a notion of *equipartition upper regularity*, where instead of considering all partitions of the vertex set with no part having weight smaller than $\eta\alpha_G$, we consider all equipartitions into q parts. Following the definition, we prove a lemma which says that the two notions of upper regularity are qualitatively equivalent.

Definition 7.2. Let $K\colon (0,\infty) \to (0,\infty)$ be any function and let $q \in \mathbb{N}$. A weighted graph G is (K, q)-*equipartition upper regular* if $\alpha_{\max}(G) \leq \alpha_G/(2q)$ and for every $\varepsilon > 0$ and equipartition $\mathcal{P}$ of $V(G)$ into q parts,

$$\sum_{i,j\in[q]} |\beta_{ij}(G/\mathcal{P})|\mathbf{1}_{|\beta_{ij}(G/\mathcal{P})|\geq K(\varepsilon)\alpha_i(G/\mathcal{P})\alpha_j(G/\mathcal{P})} \leq \varepsilon.$$

Equipartitions are defined in Section 6.1. We use $\beta_{ij}(G/\mathcal{P})$ and $\beta_{ij}(G/\phi)$ as synonyms (see (2.5) for the definition), where the function $\phi\colon V(G) \to [q]$ defines the partition $\mathcal{P}$ as the preimages of the points in $[q]$, but recall from (2.20) that $\beta_{ij}(G_{\mathcal{P}})$ is normalized differently from $\beta_{ij}(G/\mathcal{P})$.

Using (2.4), (2.5), (2.20), and (2.21), it is clear that if G is (K, η)-upper regular, then it is (K, q)-equipartition upper regular for $q \leq 1/(2\eta)$, since in every equipartition of $V(G)$ into q parts, the weight of each part is at least

$$\alpha_G/q - \alpha_{\max}(G) \geq \alpha_G/q - \eta\alpha_G \geq \eta\alpha_G.$$

Conversely, we also have the following.

Lemma 7.3. *Let $K'\colon (0,\infty) \to (0,\infty)$ be any function, and let*

$$K(\varepsilon) = \max\{4\varepsilon^{-1}K'(\varepsilon/4), 16\varepsilon^{-2}\}.$$

Let $\eta > 0$ and $q_0 = \eta^{-2}$. If a weighted graph G is (K', q')-equipartition upper regular for some $q' \geq q_0$, then G is (K, η)-upper regular.

Proof. By scaling the vertex weights, we may assume without loss of generality that $\alpha_G = 1$. Let $\mathcal{P} = (V_1, \ldots, V_q)$ be a partition of $V(G)$ into q classes, where $\alpha_{V_i} \geq \eta$ for each $i \in [q]$, and let $\varepsilon > 0$. Define

$$S := \{(i,j) \in [q]\times[q] : |\beta_{ij}(G/\mathcal{P})| \geq K(\varepsilon)\alpha_i(G/\mathcal{P})\alpha_j(G/\mathcal{P})\}.$$

We need to prove that

$$\sum_{(i,j)\in S} |\beta_{ij}(G/\mathcal{P})| \leq \varepsilon. \tag{7.1}$$

Since $\alpha_i(G/\mathcal{P}) \geq \eta$ for all $i \in [q]$ and $\sum_{i,j\in[q]} |\beta_{ij}(G/\mathcal{P})| \leq 1$, we have

$$\eta^2|S| \leq \sum_{(i,j)\in S} \alpha_i(G/\mathcal{P})\alpha_j(G/\mathcal{P}) \leq \frac{1}{K(\varepsilon)} \sum_{(i,j)\in S} |\beta_{ij}(G/\mathcal{P})| \leq \frac{1}{K(\varepsilon)}. \tag{7.2}$$

Thus if $\eta^2 K(\varepsilon) > 1$, then $S = \emptyset$ and (7.1) trivially holds. So assume from now on that $\eta^2 K(\varepsilon) \le 1$. Since $K(\varepsilon) \ge 16\varepsilon^{-2}$ by assumption, we have

$$\varepsilon \ge 4\eta. \tag{7.3}$$

For each $x \in V(G)$, define the (weighted) degree of x to be

$$\deg_x(G) = \sum_{y \in V(G)} \alpha_y(G) \left|\beta_{xy}(G)\right|.$$

We construct an equipartition $\mathcal{P}'$ of $V(G)$ as follows. For each $i \in [q]$, we partition V_i into subsets $V_{i,0}, V_{i,1}, \dots, V_{i,k_i}$ such that

$$\alpha_{V_{i,k}}(G) \in \Big[\frac{1}{q'} - \alpha_{\max}(G), \frac{1}{q'} + \alpha_{\max}(G)\Big] \qquad \text{for } 1 \le k \le k_i,$$

$\alpha_{V_{i,0}}(G) < 1/q'$, and the vertices in $V_{i,0}$ all have the lowest degree present in V_i. We will do this in such a way that for all $i = 1\dots,q$ and all $j = 1, \dots, k_i$,

$$\frac{1}{q'}\left(\sum_{i'=1}^{i-1} k_{i'} + j\right) \le \sum_{i'=1}^{i-1}\sum_{j'=1}^{k_{i'}} \alpha_{V_{i',j'}(G)} + \sum_{j'=1}^{j} \alpha_{V_{i,j'}(G)} \le \frac{1}{q'}\left(\sum_{i'=1}^{i-1} k_{i'} + j\right) + \alpha_{\max}(G).$$

For example, we can do this greedily by sorting all the vertices in V_i according to their degrees and then placing them into $V_{i,j}$ for $j = 1, 2, \dots$ in decreasing order by degree until the lower bound in the above inequality is satisfied. Since the last vertex added contributed at most $\alpha_{\max}(G)$ to the sum, we are guaranteed to have the upper bound as well. When the total weight left in V_i is too small to fill another $V_{i,j}$, we are necessarily left with a remainder $V_{i,0}$ that has weight less than $1/q'$ and contains only vertices with the lowest degree from V_i.

Next, consider the remainder sets $V_{1,0}, \dots, V_{q,0}$, whose union we denote by V_0. By construction, either V_0 is empty, in which case we do nothing, or α_{V_0} lies between $k_0/q' - \alpha_{\max}(G)$ and k_0/q', where $k_0 = q' - \sum_{i \ge 1} k_i$. Proceeding again greedily (this time ignoring the degrees), we decompose V_0 into k_0 sets $V_{0,1}, \dots, V_{0,k_0}$ with weights between $1/q' - \alpha_{\max}(G)$ and $1/q' + \alpha_{\max}(G)$. The sets $V_{i,j}$ with $0 \le i \le q$ and $j \ge 1$ then form an equipartition $\mathcal{P}'$ of $V(G)$ into q' sets.

Define S' to be the set of pairs $(u, v) \in [q'] \times [q']$ for which we can find an $(i, j) \in S$ such that u is the label of a subclass of V_i and v is the label of a subclass of V_j. In other words, S' refines the set S from $[q] \times [q]$ to $[q'] \times [q']$, except that it does not necessarily contain pairs (u, v) for which u or v is in V_0 (since the remainder sets used to form V_0 do not necessarily come from a single part of $\mathcal{P}$). Thus, we have

$$\begin{aligned}
\sum_{(i,j)\in S} |\beta_{ij}(G/\mathcal{P})| &\le \sum_{(u,v)\in S'} |\beta_{uv}(G/\mathcal{P}')| + \frac{1}{\|G\|_1} \sum_{\substack{x,y \in V(G)\\ x\in V_0 \text{ or } y \in V_0}} \alpha_x(G)\alpha_y(G)|\beta_{xy}(G)| \\
&\le \sum_{(u,v)\in S'} |\beta_{uv}(G/\mathcal{P}')| + \frac{2}{\|G\|_1} \sum_{x\in V_0} \alpha_x(G)\deg_x(G). \qquad (7.4)
\end{aligned}$$

It remains to prove that (7.4) is at most ε.

We begin with the second term. For each i, since the vertices in $V_{i,0}$ are among the lowest degree vertices of V_i, $\alpha_{V_{i,0}}(G) < 1/q'$, and $\alpha_{V_i}(G) \ge \eta$, we have

$$\sum_{x \in V_{i,0}} \alpha_x(G)\deg_x(G) \le \frac{\alpha_{V_{i,0}}(G)}{\alpha_{V_i}(G)} \sum_{x\in V_i} \alpha_x(G)\deg_x(G) \le \frac{1}{q'\eta} \sum_{x\in V_i} \alpha_x(G)\deg_x(G).$$

Summing over $i \in [q]$ and using $\sum_{x\in V(G)} \alpha_x(G)\deg_x(G) = \|G\|_1$, we see that the second term in (7.4) is at most

$$\frac{2}{q'\eta} \le \frac{2}{q_0\eta} = 2\eta \le \frac{\varepsilon}{2}$$

by (7.3).

To bound the first term in (7.4), we decompose the sum into a sum of those terms for which $|\beta_{u,v}(G/\mathcal{P}')|$ is larger than $K'(\varepsilon/4)\alpha_u(G/\mathcal{P}')\alpha_v(G/\mathcal{P}')$ and a sum of those for which it is at most this large. Since G is (K', q')-equipartition upper regular, we can bound the first sum by $\varepsilon/4$, while the second is clearly bounded by the sum of $K'(\varepsilon/4)\alpha_u(G/\mathcal{P}')\alpha_v(G/\mathcal{P}')$ over $(u, v) \in S'$. Taking into account that $K'(\varepsilon/4) \le \frac{\varepsilon}{4}K(\varepsilon)$, this proves that

$$\sum_{(u,v)\in S'} |\beta_{uv}(G/\mathcal{P}')| \le \frac{\varepsilon}{4} + \frac{\varepsilon K(\varepsilon)}{4} \sum_{(u,v)\in S'} \alpha_u(G/\mathcal{P}')\alpha_v(G/\mathcal{P}').$$

Since

$$\sum_{(u,v)\in S'} \alpha_u(G/\mathcal{P}')\alpha_v(G/\mathcal{P}') \le \sum_{(i,j)\in S} \alpha_i(G/\mathcal{P})\alpha_j(G/\mathcal{P}) \le \frac{1}{K(\varepsilon)}$$

by (7.2), we have shown that the first term in (7.4) is bounded by $\varepsilon/2$, which completes our proof. □

Proof of Proposition 7.1. To show that $(G_n)_{n\ge 0}$ is uniformly upper regular by Lemma 7.3, it suffices to show that there is some $K\colon (0,\infty) \to (0,\infty)$ and some sequence of integers $q_n \to \infty$ such that G_n is (K, q_n)-equipartition upper regular.

Equivalently, this amounts to showing that we can find some $q_n \to \infty$ so that $\alpha_{\max}(G_n)/\alpha_{G_n} \le 1/(2q_n)$ and for every $\varepsilon > 0$, there is some real $K > 0$ so that for sufficiently large[7] n, we have

$$\max_{\substack{\phi\colon V(G_n)\to[q_n] \\ \text{equipartition}}} \sum_{i,j\in[q_n]} |\beta_{ij}(G_n/\phi)| \mathbf{1}_{|\beta_{ij}(G_n/\phi)| \ge K\alpha_i(G_n/\phi)\alpha_j(G_n/\phi)} \le \varepsilon. \tag{7.5}$$

For any weighted graph G with $\alpha_{\max}(G)/\alpha_G \le 1/(2q)$ and equipartition $\phi\colon V(G) \to [q]$, we have

$$\sum_{i,j\in[q]} |\beta_{ij}(G/\phi)| \mathbf{1}_{|\beta_{ij}(G/\phi)| \ge K\alpha_i(G/\phi)\alpha_j(G/\phi)} = -E_\phi(G, J)$$

by the definition (2.9) of $E_\phi(G, J)$, where $J \in \{-1, 0, 1\}^{q\times q}$ is given by

$$J_{ij} = \operatorname{sign}(\beta_{ij}(G/\phi)) \mathbf{1}_{|\beta_{ij}(G/\phi)| \ge K\alpha_i(G/\phi)\alpha_j(G/\phi)}.$$

[7]This is equivalent to the same claim for all n since we can increase K to account for the first finitely many values of n.

Using $\alpha_i(G/\phi) \geq 1/q - \alpha_{\max}(G)/\alpha_G \geq 1/(2q)$, we obtain

$$\begin{aligned}\sum_{i,j\in[q]} |J_{ij}| &= \sum_{i,j\in[q]} \mathbf{1}_{|\beta_{ij}(G/\phi)|\geq K\alpha_i(G/\phi)\alpha_j(G/\phi)} \\ &\leq \sum_{i,j\in[q]} \frac{|\beta_{ij}(G/\phi)|}{K\alpha_i(G/\phi)\alpha_j(G/\phi)} \\ &\leq \frac{4q^2}{K}\sum_{i,j\in[q]} |\beta_{ij}(G/\phi)| \leq \frac{4q^2}{K}.\end{aligned}$$

It follows from the definition (2.12) of $E_{\mathbf{a},\varepsilon}(G,J)$ that

$$\begin{aligned}\text{left side of (7.5)} &\leq \max_{\substack{\phi\colon V(G_n)\to[q_n] \\ \text{equipartition}}} \max_{\substack{J\in\{-1,0,1\}^{q_n\times q_n} \\ \text{symmetric} \\ \sum_{i,j}|J_{ij}|\leq 4q_n^2/K}} (-E_\phi(G_n,J)) \\ &= \max_{\substack{J\in\{-1,0,1\}^{q_n\times q_n} \\ \text{symmetric} \\ \sum_{i,j}|J_{ij}|\leq 4q_n^2/K}} (-E_{\mathbf{q}_n,\alpha_{\max}(G_n)/\alpha_{G_n}}(G_n,J)) \qquad (7.6)\end{aligned}$$

(here $\mathbf{q}_n = (1/q_n,\dots,1/q_n) \in \triangle^{q_n}$, and we also write $\mathbf{q} = (1/q,\dots,1/q) \in \triangle^q$ below).

Since the microcanonical ground state energies of G_n converge to those of W, we know that for every $q \in \mathbb{N}$, we can find $\varepsilon_0(q) > 0$ and $n_0(q)$ so that

$$-E_{\mathbf{q},\varepsilon}(G_n,J) \leq -\mathcal{E}_{\mathbf{q}}(W,J) + 1/q$$

for all $0 < \varepsilon < \varepsilon_0(q)$, all $n > n_0(q)$, and every symmetric matrix $J \in \{-1,0,1\}^{q\times q}$ (as there are only finitely many such J for each q). Set $\varepsilon_n = \alpha_{\max}(G_n)/\alpha_{G_n}$, so that $\varepsilon_n \to 0$ because there are no dominant nodes. It follows that we can find a slowly growing sequence $q_n \to \infty$ so that (for sufficiently large n) we have $\varepsilon_n < \varepsilon_0(q_n)$, $n > n_0(q_n)$, and $q_n\varepsilon_n \leq 1/2$, from which it follows that

$$-E_{\mathbf{q}_n,\varepsilon_n}(G_n,J) \leq -\mathcal{E}_{\mathbf{q}_n}(W,J) + 1/q_n$$

for all symmetric matrices $J \in \{-1,0,1\}^{q_n\times q_n}$. Hence for sufficiently large n

$$\begin{aligned}(7.6) &\leq \max_{\substack{J\in\{-1,0,1\}^{q_n\times q_n} \\ \text{symmetric} \\ \sum_{i,j}|J_{ij}|\leq 4q_n^2/K}} \big(-\mathcal{E}_{\mathbf{q}_n}(W,J) + 1/q_n\big) \\ &\leq \sup_{\substack{S\subseteq[0,1]^2 \\ \lambda(S)\leq 4/K}} \int_S |W(x,y)|\,dx\,dy + 1/q_n.\end{aligned}$$

We can choose K large enough that the first term in the final bound above is at most $\varepsilon/2$. Since $q_n \to \infty$, the second term is also at most $\varepsilon/2$ for sufficiently large n. This proves (7.5), showing that $(G_n)_{n\geq 0}$ is uniformly upper regular. □

Appendix A. Proof of the rearrangement inequality

In this appendix, we prove that

$$\mathsf{E}[W\,U] \leq \mathsf{E}[W^*\,U^*].$$

when W is an L^p graphon and U is an $L^{p'}$ graphon with $\frac{1}{p} + \frac{1}{p'} = 1$, with equality holding whenever U and W are aligned.

If $W, U \geq 0$, the proof of the rearrangement inequality is standard, and can, e.g., be deduced from the following level-set representations

$$U(x,y) = \int_0^\infty dt\, \mathbf{1}[U(x,y) > t] \quad \text{and} \quad W(x,y) = \int_0^\infty ds\, \mathbf{1}[U(x,y) > s].$$

Indeed, with the help of this representation, we get

$$\begin{aligned} \mathsf{E}[W\,U] &= \mathsf{E}\Big[\int_0^\infty ds\, \mathbf{1}[W > s] \int_0^\infty dt\, \mathbf{1}[U > t]\Big] \\ &= \int_0^\infty ds \int_0^\infty dt\ \Pr\Big[W > s \text{ and } U > t\Big] \\ &\leq \int_0^\infty ds \int_0^\infty dt\ \min\Big\{\Pr[W > s], \Pr[U > t]\Big\} \\ &= \int_0^\infty ds \int_0^\infty dt\ \min\Big\{\Pr[W^* > s], \Pr[U^* > t]\Big\}, \end{aligned}$$

where in the last step we used that U and U^* as well as W and W^* have the same distribution. Since the U^* and W^* have nested level sets, the expression in the last line is equal to

$$\begin{aligned} &\int_0^\infty ds \int_0^\infty dt\ \Pr\Big[W^* > s \text{ and } U^* > t\Big] \\ &\qquad = \mathsf{E}\Big[\int_0^\infty ds\, \mathbf{1}[W^* > s] \int_0^\infty dt\, \mathbf{1}[U^* > t]\Big] \\ &\qquad = \mathsf{E}[W^*\,U^*]. \end{aligned}$$

If W and U are aligned themselves, the only inequality in the above proof becomes an equality, showing that $\mathsf{E}[W\,U] = \mathsf{E}[W^*\,U^*]$ if W and U are aligned.

If W and U are bounded below, say by $W \geq -M$ and $U \geq -M$ for some $M < \infty$, then we just use that $\mathsf{E}[W^*\,U^*] - \mathsf{E}[W\,U] = \mathsf{E}[(W+M)^*\,(U+M)^*] - \mathsf{E}[(W+M)\,(U+M)]$, which follows from linearity of expectations and the fact that $\mathsf{E}[W] = \mathsf{E}[W^*]$ and $\mathsf{E}[U] = \mathsf{E}[U^*]$. Finally, to control the tails as $M \to \infty$, we bound

$$\begin{aligned} &\Big|\mathsf{E}[WU] - \mathsf{E}\Big[W\mathbf{1}[W \geq -M]\,U\mathbf{1}[U \geq -M]\Big]\Big| \\ &\qquad \leq \mathsf{E}\Big[|WU|\mathbf{1}[|U| \geq M]\Big] + \mathsf{E}\Big[|WU|\mathbf{1}[|W| \geq M]\Big] \\ &\qquad \leq \|W\|_p \Big\|U\mathbf{1}[|U| \geq M]\Big\|_{p'} + \|U|\|_{p'} \Big\|W\mathbf{1}[|W| \geq M]\Big\|_p. \end{aligned}$$

Now the right side goes to zero as $M \to \infty$ by our assumption that $W \in L^p$ and $U \in L^{p'}$.

References

[1] I. Benjamini and O. Schramm, *Recurrence of distributional limits of finite planar graphs*, Electron. J. Probab. **6** (2001), no. 23, 13 pp. arXiv:math/0011019 doi:10.1214/EJP.v6-96 MR1873300

[2] B. Bollobás and O. Riordan, *Metrics for sparse graphs*, in S. Huczynska, J. D. Mitchell, and C. M. Roney-Dougal, eds., *Surveys in combinatorics 2009*, pages 211–287, London Math. Soc. Lecture Note Ser. **365**, Cambridge University Press, Cambridge, 2009. arXiv:0708.1919 MR2588543

[3] C. Borgs, J. T. Chayes, H. Cohn, and Y. Zhao, *An L^p theory of sparse graph convergence I: limits, sparse random graph models, and power law distributions*, preprint, 2014. arXiv:1401.2906

[4] C. Borgs, J. Chayes, and D. Gamarnik, *Convergent sequences of sparse graphs: a large deviations approach*, preprint, 2013. arXiv:1302.4615

[5] C. Borgs, J. Chayes, J. Kahn, and L. Lovasz, *Left and right convergence of graphs with bounded degree*, Random Struct. Alg. **42** (2013), 1–28. arXiv:1002.0115 doi:10.1002/rsa.20414 MR2999210

[6] C. Borgs, J. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi, *Counting graph homomorphisms*, in M. Klazar, J. Kratochvíl, M. Loebl, J. Matoušek, R. Thomas, and P. Valtr, eds., *Topics in discrete mathematics*, pages 315–371, Algorithms Combin. **26**, Springer, Berlin, 2006. doi:10.1007/3-540-33700-8_18 MR2249277

[7] 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** (2008), 1801–1851. arXiv:math/0702004 doi:10.1016/j.aim.2008.07.008 MR2455626

[8] 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** (2012), 151–219. doi:10.4007/annals.2012.176.1.2 MR2925382

[9] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, corrected reprint of the second (1998) edition, Stochastic Modelling and Applied Probability **38**, Springer-Verlag, Berlin, 2010. doi:10.1007/978-3-642-03311-7 MR2571413

[10] A. Frieze and R. Kannan, *Quick approximation to matrices and applications*, Combinatorica **19** (1999), 175–220. doi:10.1007/s004930050052 MR1723039

[11] A. W. Knapp, *Basic real analysis*, Cornerstones, Birkhäuser Boston, Inc., Boston, MA, 2005. doi:10.1007/0-8176-4441-5 MR2155259

[12] L. Lovász and B. Szegedy, *Szemerédi's lemma for the analyst*, Geom. Funct. Anal. **17** (2007), 252–270. doi:10.1007/s00039-007-0599-6 MR2306658

[13] G. B. Price, *On the completeness of a certain metric space with an application to Blaschke's selection theorem*, Bull. Amer. Math. Soc. **46** (1940), 278–280. doi:10.1090/S0002-9904-1940-07195-2 MR0002010

Microsoft Research, One Memorial Drive, Cambridge, MA 02142
E-mail address: borgs@microsoft.com

Microsoft Research, One Memorial Drive, Cambridge, MA 02142
E-mail address: jchayes@microsoft.com

Microsoft Research, One Memorial Drive, Cambridge, MA 02142
E-mail address: cohn@microsoft.com

Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139
E-mail address: yufeiz@mit.edu