Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Advances in Mathematics 219 (2008) 1801–1851
ADVANCES IN Mathematics
www.elsevier.com/locate/aim

Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing

C. Borgs [a], J.T. Chayes [a,*], L. Lovász [a,1], V.T. Sós [b,2], K. Vesztergombi [c]

[a] *Microsoft Research, One Microsoft Way, Redmond, WA 98052, USA*
[b] *Alfréd Rényi Institute of Mathematics, POB 127, H-1364 Budapest, Hungary*
[c] *Eötvös Loránd University, Pázmány Péter sétani 1/C, H-1117 Budapest, Hungary*

Received 31 January 2007; accepted 21 July 2008

Available online 30 August 2008

Communicated by Sara Billey

Abstract

We consider sequences of graphs (G_n) and define various notions of convergence related to these sequences: "left convergence" defined in terms of the densities of homomorphisms from small graphs into G_n; "right convergence" defined in terms of the densities of homomorphisms from G_n into small graphs; and convergence in a suitably defined metric.

In Part I of this series, we show that left convergence is equivalent to convergence in metric, both for simple graphs G_n, and for graphs G_n with nodeweights and edgeweights. One of the main steps here is the introduction of a cut-distance comparing graphs, not necessarily of the same size. We also show how these notions of convergence provide natural formulations of Szemerédi partitions, sampling and testing of large graphs.

Keywords: Graph sequences; Cut metric; Sampling; Parameter testing; Szemerédi partitions; Partition functions; Free energies

* Corresponding author.
E-mail address: jchayes@microsoft.com (J.T. Chayes).
1 Current address: Eötvös Loránd University, Pázmány Péter sétány 1/C, H-1117 Budapest, Hungary.
2 Research supported in part by OTKA grants T032236, T038210, T042750.

doi:10.1016/j.aim.2008.07.008

1. Introduction

In this and accompanying papers, we define a natural notion of convergence of a sequence of graphs, and show that other useful notions of convergence are equivalent to it. We are motivated by the fact that, in many subfields of mathematics, computer science and physics, one studies properties of very large graphs, or properties of graph sequences that grow beyond all limits. Let us give a few examples:

Random networks. There is a large literature of graph models of the Internet, the WWW, and other so-called "scale-free" technological networks, first modeled in this context by Barabási and Albert [7]. Among the technological networks that are modeled are the graph of computers and physical links between them, the graph of so-called Autonomous Systems such as Internet Service Providers, the graph of webpages with hyperlinks, etc. These graphs are often similar to various graphs of social networks: acquaintances, co-publications, the spreading of certain diseases, etc.

These networks are formed by random processes, but their properties are quite different from the traditional Erdős–Rényi random graphs: their degree distribution has a "heavy tail," they tend to be clustered, the neighborhoods of their nodes are denser than the average edge density, etc. Several models of random scale-free graphs have been proposed and studied. For rigorous work, see [12,13] for undirected models, [25] for "copying models," [10] for a directed model, [9] for the spread of viruses on these networks, and [11] for a survey of rigorous work with more complete references.

Quasirandom graphs. Quasirandom (also called pseudorandom) graphs were introduced by Thomason [31] and Chung, Graham and Wilson [20]. These graph sequences can be deterministic, but have many properties of true random graphs. A nice example is the sequence of Paley graphs (quadratic residue graphs). These graphs are remarkably similar to a random graph with edge-probability $1/2$ on the same number of nodes in many ways. The most relevant for us is that they contain (asymptotically) the same number of copies of each fixed graph F as the random graph—this is one of the many equivalent ways to define quasirandom graphs. Many other questions in graph theory, in particular in extremal graph theory, also involve asymptotic counting of small graphs.

Property testing of large graphs. Say we are given a large graph and we want to determine certain numerical parameters, e.g., the edge density, of that graph by sampling a bounded number of nodes. Or perhaps we want to determine whether the large graph has a given property, e.g., is it 3-colorable? In particular, which parameters can be accurately estimated and which properties can be tested with high probability by looking only at subgraphs on small randomly chosen subsets of the nodes? A precise definition of property testing was given by Goldreich, Goldwasser and Ron [24], who also proved several fundamental results about this problem.

Statistical mechanics. Many models in physics are described by a weighted coloring of some large graph G. The graph G typically represents underlying geometric structure of the model under consideration, e.g. a crystal lattice and its nearest neighbor structure, while the color of a given node represents the local state. In the simplest case of two colors, the two colors can represent quantities like the two possible orientations of a spin variable, or the presence or absence of a molecule at a given position. The interactions between different local states can then be described by a weighted "interaction graph" H, with smaller edgeweights corresponding to weaker interactions, and larger edgeweights representing stronger interactions. In this context, the weighted number of colorings represents the so-called partition function of the model.

Combinatorial optimization. Many optimization problems can be described as weighted coloring problems. A simple example is the max-cut problem, where our task is to find the maximal cut in a large graph G. If we consider a coloring of G with two colors, 1 and 2, and weight a coloring by the number of edges with two differently colored endnodes, then the maximum cut is just given by the maximum weight coloring.

In this and two accompanying papers we develop a theory of convergence of graph sequences, which works best in two extreme cases: dense graphs (the subject of this paper and [17]) and graphs with bounded degree (the subject of [18]). Convergence of graph sequences was defined by Benjamini and Schramm [8] for graphs with bounded degree. For dense graphs, convergence was defined by Erdös, Lovász and Spencer [21] implicitly, and by the authors of this paper [15] explicitly.

Our general setup will be the following. We have a "large" graph G with node set $V(G)$ and edge set $E(G)$. There are (at least) two ways of studying G using homomorphisms. First, we can count the number of copies of various "small" graphs F in G, more precisely, we count the number of homomorphisms from F to G; this way of looking at G allows us to treat many problems in, e.g., extremal graph theory. Second, we can count homomorphisms from G into various small graphs H; this includes many models in statistical physics and many problems on graph coloring.

These two notions of probing a large graph with a small graph lead to two different notions of convergence of a sequence of graphs (G_n): convergence from the left, corresponding to graphs which look more and more similar when probed with homomorphisms from small graphs into G_n, and convergence from the right, corresponding to graph sequences whose elements look more and more similar when probed with homomorphism from G_n into a small graphs.

This theory can also be viewed as a substantial generalization of the theory of quasirandom graphs. In fact, most of the equivalent characterizations of quasirandom graphs are immediate corollaries of the general theory developed here and in our companion paper [17].

In this paper we study convergence from the left, both for sequences of simple graphs and sequences of weighted graphs, and its relations to sampling and testing. Since this paper focuses on convergence from the left, we will often omit the phrase "from the left." We will also show that convergence from the left is equivalent to convergence in metric for a suitable notion of distance between two weighted graphs. Finally, we will show that convergence from the left is equivalent to the property that the graphs in the sequence have asymptotically the same Szemerédi partitions.

Convergence from the right will be the subject matter of the sequel of this paper [17].

Convergence in metric clearly allows for a completion by the usual abstract identification of Cauchy sequences of distance zero. But it turns out (Lovász and Szegedy [26]) that the limit object of a convergent graph sequence has a much more natural representation in terms of a measurable symmetric function $W : [0,1]^2 \to \mathbb{R}$ (we call these functions *graphons*). In fact, it is often useful to represent a finite graph G in terms of a suitable function W_G on $[0,1]^2$, defined as step function with steps of length $1/|V(G)|$ and values 0 and 1, see below for the precise definition. While the introduction of graphons requires some basic notions of measure theory, it will simplify many proofs in this paper.

The organization of this paper is as follows: In the next section, we introduce our definitions: in addition to left-convergence, we define a suitable distance between weighted graphs, and state our main results for weighted graphs. In Section 3, we generalize these definitions and results to graphons. The following section, Section 4, is devoted to sampling, and contains the proofs of the main results of this paper, including the equivalence of left-convergence and convergence in metric. Section 5 relates convergence in metric to an *a priori* weaker form of "convergence in norm" and to Szemerédi partitions, and Section 6 proves our results on testing. We close this paper with a section on miscellaneous results and an outlook on right-convergence. In Appendix A, we describe a few details of proofs which are omitted in the main body of the paper.

2. Weighted and unweighted graphs

2.1. Notation

We consider both unweighted, simple graphs and weighted graphs, where, as usual, a simple graph G is a graph without loops or multiple edges. We denote the node and edge set of G by $V(G)$ and $E(G)$, respectively.

A *weighted graph* G is a graph with a weight $\alpha_i = \alpha_i(G) > 0$ associated with each node and a weight $\beta_{ij} = \beta_{ij}(G) \in \mathbb{R}$ associated with each edge ij, including possible loops with $i = j$. For convenience, we set $\beta_{ij} = 0$ if $ij \notin E(G)$. We set

$$\alpha_G = \sum_i \alpha_i(G), \qquad \|G\|_\infty = \max_{i,j} \bigl|\beta_{ij}(G)\bigr|, \quad \text{and} \quad \|G\|_2 = \biggl(\sum_{i,j} \frac{\alpha_i \alpha_j}{\alpha_G^2} \beta_{ij}^2\biggr)^{1/2},$$

and for $S, T \subset V(G)$, we define

$$e_G(S,T) = \sum_{\substack{i \in S\\ j \in T}} \alpha_i(G)\alpha_j(G)\beta_{ij}(G). \tag{2.1}$$

A weighted graph G is called *soft-core* if it is a complete graph with loops at each node, and every edgeweight is positive. An *unweighted graph* is a weighted graph where all the node- and edgeweights are 1. Note that $e_G(S,T)$ reduces to the number of edges in G with one endnode in S and the other in T if G is unweighted.

Let G be a graph and $k \geqslant 1$. The *k-fold blow-up* of G is the graph $G[k]$ obtained from G by replacing each node by k independent nodes, and connecting two new nodes if and only if their originals were connected. If G is weighted, we define $G[k]$ to be the graph on nk nodes labeled by pairs iu, $i \in V(G)$, $u = 1, \dots, k$, with edgeweights $\beta_{iu,jv}(G[k]) = \beta_{ij}(G)$ and nodeweights $\alpha_{iu}(G[k]) = \alpha_i(G)$. A related notion is the notion of *splitting nodes*. Here a node i with nodeweight α_i is replaced by k nodes $i_1, \dots, i_k$ with nodeweights $\alpha_{i_1}, \dots, \alpha_{i_k}$ adding up to α_i, with new edgeweights $\beta_{i_u,j_v} = \beta_{i,j}$. Up to a global rescaling of all nodeweights, blowing up a graph by a factor k is thus the same as splitting all its nodes evenly into k nodes, so that the new weights α_{i_t} are equal to the old weights α_i divided by k.

As usual, a function from the set of simple graphs into the reals is called a *simple graph parameter* if it is invariant under relabeling of the nodes. Finally, we write $G \cong G'$ if G and G' are isomorphic, i.e., if G' can be obtained from G by a relabeling of its nodes.

2.2. Homomorphism numbers and left convergence

Let F and G be two simple graphs. We define $\hom(F,G)$ as the number of homomorphisms from F to G, i.e., the number of adjacency preserving maps $V(F)\to V(G)$, and the *homomorphism density* of F in G as

$$t(F,G)=\frac{1}{|V(G)|^{|V(F)|}}\hom(F,G).$$

The homomorphism density $t(F,G)$ is thus the probability that a random map from $V(F)$ to $V(G)$ is a homomorphism.

Alternatively, one might want to consider the probability $t_{\mathrm{inj}}(F,G)$ that a random injective map from $V(F)$ to $V(G)$ is adjacency preserving, or the probability $t_{\mathrm{ind}}(F,G)$ that such a map leads to an induced subgraph. Since most maps into a large graph G are injective, there is not much of a difference between $t(F,G)$ and $t_{\mathrm{inj}}(F,G)$. As for $t_{\mathrm{inj}}(\cdot,G)$ and $t_{\mathrm{ind}}(\cdot,G)$, they can be quite different even for large graphs G, but by inclusion–exclusion, the information contained in the two is strictly equivalent. We therefore incur no loss of generality if we restrict ourselves to the densities $t(\cdot,G)$.

We extend the notion of homomorphism numbers to weighted graphs G by setting

$$\hom(F,G)=\sum_{\phi:\,V(F)\to V(G)}\prod_{i\in V(F)}\alpha_{\phi(i)}(G)\prod_{ij\in E(F)}\beta_{\phi(i),\phi(j)}(G) \tag{2.2}$$

where the sum runs over all maps from $V(F)$ to $V(G)$, and define

$$t(F,G)=\frac{\hom(F,G)}{\alpha_G^k}, \tag{2.3}$$

where k is the number of nodes in F.

It seems natural to think of two graphs G and G' as similar if they have similar homomorphism densities. This leads to the following definition.

Definition 2.1. Let (G_n) be a sequence of weighted graphs with uniformly bounded edgeweights. We say that (G_n) is *convergent from the left*, or simply *convergent*, if $t(F,G_n)$ converges for any simple graph F.

In [19], the definition of convergence was restricted to sequences of graphs (G_n) with $|V(G_n)|\to\infty$. As long as we deal with simple graphs, this is reasonable, since there are only a finite number of graphs with bounded size. But in this paper we also want to cover sequences of weighted graphs on, say, the same set of nodes, but with the node- or edgeweights converging; so we do not assume that the number of nodes in convergent graph sequences tends to infinity.

A simple example of a convergent graph sequence is a sequence of random graphs $(G_{n,p})$, for which $t(F,G_{n,p})$ is convergent with probability one, with $t(F,G_{n,p})\to p^{|E(F)|}$ as $n\to\infty$. Other examples are quasirandom graph sequences (G_n) for which $t(F,G_{n,p})\to p^{|E(F)|}$ by definition, and the sequence of half-graphs $(H_{n,n})$, with $H_{n,n}$ defined as the bipartite graph on $[2n]$ with an edge between i and j if $j\geqslant n+i$ (see Examples 3.11–3.13 in Section 3.1).

2.3. Cut-distance

We define a notion of distance between two graphs, which will play a central role throughout this paper. Among other equivalences, convergence from the left will be equivalent to convergence in this metric.

To illuminate the rather technical definition, we first define a distance of two graphs G and G' in a special case, and then extend it in two further steps. In all these definitions, it is not hard to verify that the triangle inequality is satisfied.

(1) *G and G' are labeled graphs with the same set of unweighted nodes V*. Several notions of distance appear in the literature, but for our purpose the most useful is the cut or rectangle distance introduced by Frieze and Kannan [23]:

$$d_\square(G,G') = \max_{S,T\subset V} \frac{1}{|V|^2}\left|e_G(S,T) - e_{G'}(S,T)\right|. \tag{2.4}$$

The cut distance between two labeled graphs thus measures how different two graphs are when considering the size of various cuts. This definition can easily be generalized to weighted graphs G and G' on the same set V, *and with the same nodeweights $\alpha_i = \alpha_i(G) = \alpha_i(G')$*:

$$d_\square(G,G') = \max_{S,T\subset V} \frac{1}{\alpha_G^2}\left|e_G(S,T) - e_{G'}(S,T)\right|. \tag{2.5}$$

As a motivation of this notion, consider two independent random graphs on n nodes with edge density $1/2$. If we measure their distance, say, by the number of edges we need to change to get one from the other (edit distance), then their distance is very large (with large probability). But the theory of random graphs teaches us that these two graphs are virtually indistinguishable, which is reflected by the fact that their $d_\square$ distance is only $O(1/n)$ with large probability.

There are many other ways of defining or approximating the cut distance; see Section 7.1.

(2) *G and G' are unlabeled graphs with the same number of unweighted nodes.* The cut-metric $d_\square(G,G')$ is not invariant under relabeling of the nodes of G and G'. For graphs without nodeweights, this can easily be cured by defining a distance $\widehat{\delta}_\square(G,G')$ as the minimum over all "overlays" of G and G', i.e.,

$$\widehat{\delta}_\square(G,G') = \min_{\widetilde{G}\cong G} d_\square(\widetilde{G},G'). \tag{2.6}$$

(3) *G and G' are unlabeled graphs with different number of nodes or with weighted nodes.* The distance notion (2.6) does not extend in a natural way to graphs with nodeweights, since it would not make much sense to overlay nodes with different nodeweights, and even less sense for graphs with different number of nodes.

To motivate the definition that follows, consider two graphs G and G', where G has three nodes with nodeweights equal to $1/3$, and G' has two nodes and nodeweights $1/3$ and $2/3$. Here a natural procedure would be the following: first split the second node into two nodes of weight $1/3$ and then calculated the optimal overlay of the resulting two graphs on three nodes.

This idea naturally leads to the following notion of "fractional overlays" of two weighted graphs G and G' on n and n' nodes, respectively. Let us first assume that both G and G' have total nodeweight 1. Viewing $\alpha(G)$ and $\alpha(G')$ as probability distributions, we then define a fractional overlay to be a coupling between these two distributions. More explicitly, a *fractional overlay* of G and G' is defined to be a non-negative $n \times n'$ matrix X such that

$$\sum_{u=1}^{n'} X_{iu} = \alpha_i(G) \quad \text{and} \quad \sum_{i=1}^{n} X_{iu} = \alpha_u(G').$$

We denote the set of all fractional overlays by $\mathcal{X}(G, G')$.

Let $X \in \mathcal{X}(G, G')$. Thinking of X_{iu} as the portion of node i that is mapped onto node u, we introduce the following "overlaid graphs" $G[X]$ and $G'[X^\top]$ on $[n] \times [n']$: in both $G[X]$ and $G'[X^\top]$, the weight of a node $(i, u) \in [n] \times [n']$ is X_{iu}; in $G[X]$, the weight of an edge $((i, u), (j, v))$ is β_{ij}, and in $G'[X^\top]$, the weight of an edge $((i, u), (j, v))$ is β'_{uv}. Since $G[X]$ and $G'[X^\top]$ have the same nodeset, the distance $d_\square(G[X], G'[X^\top])$ is now well defined. Taking the minimum over all fractional overlays, this gives:

Definition 2.2. For two weighted graphs G, G' with total nodeweight $\alpha_G = \alpha_{G'} = 1$, we set

$$\delta_\square(G, G') = \min_{X \in \mathcal{X}(G, G')} d_\square\big(G[X], G'\big[X^\top\big]\big). \tag{2.7}$$

If the total nodeweight of G or G' is different from 1, we define the distance between G and G' by the above formulas, applied to the graphs $\tilde{G}$ and $\tilde{G}'$ obtained from G and G' by dividing all nodeweights by α_G and $\alpha_{G'}$, respectively.

Fractional overlays can be understood as integer overlays of suitably blown up versions of G and G', at least if the entries of X are rational (otherwise, one has to take a limit of blowups). This observation shows that for two graphs G and G' with nodeweights one, we have

$$\delta_\square(G, G') = \lim_{\substack{k, k' \to \infty \\ k/k' = n'/n}} \widehat{\delta}_\square\big(G[k], G'[k']\big). \tag{2.8}$$

Note that $\delta_\square(G, G')$ can be 0 for non-isomorphic graphs G and G'; for example,

$$\delta_\square\big(G, G[k]\big) = 0 \tag{2.9}$$

for all $k \geqslant 1$. So δ_G is only a pre-metric; but we will call it, informally, a metric.

Of course, the definition of $\delta_\square(G, G')$ also applies if G and G' have the same number of nodes, and it may give a value different from $\widehat{\delta}_\square(G, G')$. The following theorem relates these two values.

Theorem 2.3. *Let G_1 and G_2 be two weighted graphs with edgeweights in $[-1, 1]$ and with the same number of unweighted nodes. Then*

$$\delta_\square(G_1, G_2) \leqslant \widehat{\delta}_\square(G_1, G_2) \leqslant 32\delta_\square(G_1, G_2)^{1/67}.$$

The first inequality is trivial, but the proof of the second is quite involved, and will be given in Section 5.1.

2.4. Szemerédi partitions of graphs

The Regularity Lemma of Szemerédi is a fundamental tool in graph theory, which has a natural formulation in our framework, as a result about approximating large graphs by small graphs. Here we mostly use the so-called weak version due to Frieze and Kannan [23].

We need some notation: for a weighted graph G and a partition $\mathcal{P} = \{V_1, \dots, V_k\}$ of $V(G)$, we define two weighted graphs $G/\mathcal{P}$ and $G_\mathcal{P}$ as follows: Let $\alpha_{V_i} = \sum_{x \in V_i} \alpha_x(G)$. The *quotient graph* $G/\mathcal{P}$ is a weighted graph on $[k]$, with nodeweights $\alpha_i(G/\mathcal{P}) = \alpha_{V_i}/\alpha_G$ and edgeweights $\beta_{ij}(G/\mathcal{P}) = \frac{e_G(V_i,V_j)}{\alpha_{V_i}\alpha_{V_j}}$, while $G_\mathcal{P}$ is a weighted graph on $V(G)$, with nodeweights $\alpha_x(G_\mathcal{P}) = \alpha_x(G)$ and edgeweights $\beta_{xy}(G_\mathcal{P}) = \beta_{ij}(G/\mathcal{P})$ for $x \in V_i$ and $y \in V_j$. These two graphs have different number of nodes, but they are similar in the sense that $\delta_\square(G/\mathcal{P}, G_\mathcal{P}) = 0$.

In our language, the Weak Regularity Lemma of Frieze and Kannan states that given a weighted graph G, one can find a partition $\mathcal{P}$ such that the graph $G_\mathcal{P}$ is near to the original graph G in the distance $d_\square$. We call the partition $\mathcal{P}$ *weakly ε-regular* if $d_\square(G, G_\mathcal{P}) \leqslant \varepsilon$. But for many purposes, all that is needed is the quotient graph $G/\mathcal{P}$. Since $\delta_\square(G, G_\mathcal{P}) \leqslant d_\square(G, G_\mathcal{P})$ and $\delta_\square(G, G_\mathcal{P}) = \delta_\square(G, G/\mathcal{P})$, the Weak Regularity Lemma also guarantees a good approximation of the original graph by a small weighted graph, the graph $H = G/\mathcal{P}$. We summarize these facts in the following lemma, which is essentially a reformulation of the Weak Regularity Lemma of [23] in the language developed in this paper.

Lemma 2.4 *(Weak Regularity Lemma [23]). For every $\varepsilon > 0$, every weighted graph G has a partition $\mathcal{P}$ into at most $4^{1/\varepsilon^2}$ classes such that*

$$d_\square(G, G_\mathcal{P}) \leqslant \varepsilon \|G\|_2, \tag{2.10}$$

so, in particular,

$$\delta_\square(G, G/\mathcal{P}) \leqslant \varepsilon \|G\|_2. \tag{2.11}$$

In Lemma 2.4 we approximate the graph G by a small weighted graph $H = G/\mathcal{P}$. If G is simple, or more generally, has edgeweights in $[0, 1]$, it is possible to strengthen this by requiring simple graphs H. Indeed, starting from a standard strengthening of the Weak Regularity Lemma (Corollary 3.4(ii) below) to obtain a weighted graph H with nodeweights one, and then applying a simple randomization procedure (Lemma 4.3 below) to the edges of H to convert this graph into a simple graph, one gets the following lemma, see appendix for details.

Lemma 2.5. *Let $\varepsilon > 0$, let $q \geqslant 2^{20/\varepsilon^2}$, and let G be a weighted graph with edgeweights in* $[0, 1]$. *Then there exists a simple graph H on q nodes such that $\delta_\square(G, H) \leqslant \varepsilon$.*

2.5. Main results

2.5.1. Left convergence versus convergence in metric

Here we state one of the main results of this paper, namely, that convergence from the left is equivalent to convergence in the metric $\delta_\square$.

Theorem 2.6. *Let (G_n) be a sequence of weighted graphs with uniformly bounded edgeweights. Then (G_n) is left convergent if and only if it is a Cauchy sequence in the metric $\delta_\square$.*

In fact, we have the following quantitative version. To simplify our notation, we only give this quantitative version for graphs with edgeweights in $[-1, 1]$; the general case follows by simply scaling all edgeweights appropriately.

Theorem 2.7. *Let G_1, G_2 be weighted graphs with edgeweights in $[-1, 1]$.*

(a) *Let F be a simple graph, then*

$$\left|t(F, G_1) - t(F, G_2)\right| \leqslant 4\left|E(F)\right|\delta_\square(G_1, G_2).$$

(b) *Let $k \geqslant 1$, and assume that $|t(F, G_1) - t(F, G_2)| \leqslant 3^{-k^2}$ for every simple graph F on k nodes. Then*

$$\delta_\square(G_1, G_2) \leqslant \frac{22}{\sqrt{\log_2 k}}.$$

The first part of this theorem is closely related to the "Counting Lemma" in the theory of Szemerédi partitions. Theorems 2.6 and 2.7 will follow from the analogous facts for graphons, Theorems 3.8 and 3.7, see Section 3.5.

2.5.2. Szemerédi partitions for graph sequences

Convergent graph sequences can also be characterized by the fact that (for any fixed error) they have Szemerédi partitions which become more and more similar.

Theorem 2.8. *Let (G_n) be a sequence of weighted graphs with nodeweights 1, edgeweights in $[-1, 1]$, and $|V(G_n)| \to \infty$ as $n \to \infty$. Then (G_n) is left-convergent if and only if for every $\varepsilon > 0$ we can find an integer $q \leqslant 2^{10/\varepsilon^2}$, and a sequence of partitions, $(\mathcal{P}_n)$ such that the following two conditions hold.*

(i) *If $|V(G_n)| \geqslant q$, then $\mathcal{P}_n$ is a weakly ε-regular partition of G_n into q classes.*
(ii) *As $n \to \infty$, the quotient graphs $G_n/\mathcal{P}_n$ converge to a weighted graph H_ε on q nodes.*

Note that the graphs in (ii) have the same node set $[q]$, so their convergence to H_ε means simply that corresponding nodeweights and edgeweights converge.

Let G_n be a convergent sequence of weighted graphs obeying the assumptions of this theorem. For n sufficiently large, the quotient graphs $G_n/\mathcal{P}_n$ are then near to both the original graph G_n and the graph H_ε, implying that $\delta_\square(G_n, H_\varepsilon) \leqslant 2\varepsilon$ whenever n is large enough. Since G_n is convergent, this implies by Theorem 2.6 that the graphs H_ε form a convergent sequence as $\varepsilon \to 0$.

The theorem can be generalized in several directions. First, using the results of Section 5.2, we can relax the condition on the nodeweights of G_n to require only that G_n has no dominant nodeweights in the sense that the maximum nodeweight of G_n divided by the total nodeweight α_{G_n} goes to zero. Second, we can strengthen the statement, to obtain a sequence of partitions which satisfy the stronger regularity conditions of the original Szemerédi Regularity Lemma [30]. We leave the details to the interested reader, who will easily see how to modify the proof in Section 5.3 to obtain these generalizations. Alon, Fisher, Krivelevich and Szegedy [3] obtain a superstrong version, which can also be derived in the framework of convergent graph sequences, see [28].

2.5.3. Sampling

Our above versions of Szemerédi's lemma (Lemmas 2.4 and 2.5) state that any graph G can be well approximated by a small graph H in the $\delta_\square$ distance. While the proofs are constructive, it will be very useful to know that such a small graph can obtained by straightforward sampling. For simplicity, we state the results for graphs with edgeweights in $[-1,1]$. For a graph G and positive integer n, let $\mathbb{G}(n,G)$ denote the (random) induced subgraph $G[S]$, where S is chosen uniformly from all subsets of $V(G)$ of size n.

Theorem 2.9. *Let G be a weighted graph with nodeweights* 1 *and edgeweights in* $[-1,1]$. *Let $k \leqslant |V(G)|$. Then*

$$\delta_\square\big(G, \mathbb{G}(k,G)\big) \leqslant \frac{10}{\sqrt{\log_2 k}} \tag{2.12}$$

with probability at least $1 - e^{-k^2/(2\log_2 k)}$.

In order to prove this theorem, we will need a theorem which allows us to compare samples from two weighted graphs on the same set of nodes. This extends a result by Alon, Fernandez de la Vega, Kannan and Karpinski [1,2]; in particular, our result concerns two graphs at arbitrary distance $d_\square(G_1, G_2)$, and also gives an improvement in the error bound.

Theorem 2.10. *Let G_1 and G_2 be weighted graphs on a common vertex set V, with nodeweights one and edgeweights in* $[-1,1]$. *Let $k \leqslant |V|$. If S is chosen uniformly from all subsets of V of size k, then*

$$\big|d_\square\big(G_1[S], G_2[S]\big) - d_\square(G_1, G_2)\big| \leqslant \frac{20}{k^{1/4}} \tag{2.13}$$

with probability at least $1 - 2e^{-\sqrt{k}/8}$.

2.5.4. Testing

The above theorem allows us to prove several results for testing graph parameters and graph properties in a straightforward way.

In this paper, we only consider parameter testing. We may want to determine some parameter of G. For example, what is the edge density? Or how large is the density of the maximum cut? Of course, we will not be able to determine the exact value of the parameter; the best we can hope for is that if we take a sufficiently large sample, we can find the approximate value of the parameter with large probability.

Definition 2.11. A graph parameter f is *testable* if for every $\varepsilon > 0$ there is a positive integer k such that if G is a graph with at least k nodes, then from the random subgraph $\mathbb{G}(k,G)$ we can compute an estimate $\widetilde{f}$ of f such that

$$\Pr\big(\big|f(G) - \widetilde{f}\big| > \varepsilon\big) \leqslant \varepsilon.$$

It is an easy observation that we can always use $\widetilde{f} = f(\mathbb{G}(k,G))$.

Testability is related to our framework through the following observation:

Proposition 2.12.

(a) *A simple graph parameter f is testable if and only if $f(G_n)$ converges for every convergent graph sequence (G_n) with $|V(G_n)| \to \infty$.*
(b) *A sequence (G_n) of simple graphs with $|V(G_n)| \to \infty$ is convergent if and only if $f(G_n)$ converges for every testable simple graph parameter f.*

Using the notions and results concerning graph distance and convergence above, we can give several characterizations of testable parameters, see Section 6.

Property testing, mentioned in the introduction, is related to parameter testing in many ways. For example, Fischer and Newman [22] proved that the edit distance (see Section 4.3) from the set of graphs exhibiting a testable property $\mathcal{P}$ is a testable parameter. A characterization of testable graph properties was given by Alon, Fischer, Newman and Shapira [4], and an elegant sufficient condition by Alon and Shapira [6]. See [27] for more connections with graph distances and graph limits.

3. Graphons

In [26], Lovász and Szegedy introduced graphons as limits of left-convergent graph sequences. Here we will first study the space of graphons in its own right, defining in particular a generalization of the distance $\delta_\square$ to graphons, and state the analogue of Theorem 2.7 for graphons. The discussion of graphons as limit objects of left-convergent graph sequences will be postponed to the last subsection of this section.

3.1. Homomorphism densities

Let $\mathcal{W}$ denote the space of all bounded measurable functions $W : [0,1]^2 \to \mathbb{R}$ that are symmetric, i.e., $W(x,y) = W(y,x)$ for all $x, y \in [0,1]$. Let $\mathcal{W}_{[0,1]}$ be the set of functions $W \in \mathcal{W}$ with $0 \leqslant W(x,y) \leqslant 1$. More generally, for a bounded interval $I \subset \mathbb{R}$, let $\mathcal{W}_I$ be the set of all functions $W \in \mathcal{W}$ with $W(x,y) \in I$. Given a function $W \in \mathcal{W}$, we can think of the interval $[0,1]$ as the set of nodes, and of the value $W(x,y)$ as the weight of the edge xy. We call the functions in $\mathcal{W}$ *graphons*.

We call a partition $\mathcal{P}$ of $[0,1]$ *measurable* if all the partition classes are (Lebesgue) measurable. The partition is an *equipartition* if all of its classes have the same Lebesgue measure. A *step function* is a function $W \in \mathcal{W}$ for which there is a partition of $[0,1]$ into a finite number of measurable sets $V_1, \dots, V_k$ so that W is constant on every product set $V_i \times V_j$. We call the sets V_i the *steps* of the step function. Often, but not always, we consider step functions whose steps are intervals; we call these *interval step functions*. If all steps of a step function have the same measure $1/k$, we say that it has *equal steps*.

Every graphon W defines a simple graph parameter as follows [26]: If F is a simple graph with $V(F) = \{1, \dots, k\}$, then let

$$t(F,W) = \int_{[0,1]^k} \prod_{ij \in E(F)} W(x_i, x_j)\, dx. \tag{3.1}$$

Every weighted graph G with nodes labeled $1, \dots, n$ defines an interval step function W_G such that $t(F, G) = t(F, W_G)$: We scale the nodeweights of G to sum to 1. Let $I_1 = [0, \alpha_1(G)]$, $I_2 = (\alpha_1(G), \alpha_1(G) + \alpha_2(G)]$, ..., and $I_n = [\alpha_1(G) + \cdots + \alpha_{n-1}(G), 1]$. We then set

$$W_G(x, y) = \beta_{v(x)v(y)}(G)$$

where $v(x) = i$ whenever $x \in I_i$. (Informally, we consider the adjacency matrix of G, and replace each entry (i, j) by a square of size $\alpha_i \times \alpha_j$ with the constant function β_{ij} on this square.) If G is unweighted, then the corresponding interval step function is a 0–1-function with equal steps.

With this definition, we clearly have $W_G \in \mathcal{W}$, and

$$t(F, G) = t(F, W_G) \tag{3.2}$$

for every finite graph F. The definition (3.1) therefore gives a natural generalization of the homomorphism densities defined in (2.2) and (2.3).

Using this notation, we can state the main result of [26]:

Theorem 3.1. *For every left-convergent sequence* (G_n) *of simple graphs there is a graphon* W *with values in* $[0, 1]$ *such that*

$$t(F, G_n) \to t(F, W)$$

for every simple graph F. *Moreover, for every graphon* W *with values in* $[0, 1]$ *there is a left-convergent sequence of graphs satisfying this relation.*

3.2. The cut norm for graphons

The distance $d_\square$ of two graphs introduced in Section 2.3 was extended to graphons by Frieze and Kannan [23]. It will be given in terms of a norm on the space $\mathcal{W}$, the *rectangle* or *cut norm*

$$\|W\|_\square = \sup_{S,T \subseteq [0,1]} \left| \int_{S\times T} W(x, y)\,dx\,dy \right| = \sup_{f,g:[0,1]\to[0,1]} \left| \int W(x, y) f(x) g(y)\,dx\,dy \right| \tag{3.3}$$

where the suprema go over all pairs of measurable subsets and functions, respectively. The cut norm is closely related to $L_\infty \to L_1$ norm of W, considered as an operator on $L^2([0, 1])$:

$$\|W\|_{\infty\to 1} = \sup_{f,g:[0,1]\to[-1,1]} \int_{[0,1]^2} W(x, y) f(x) g(y)\,dx\,dy. \tag{3.4}$$

Indeed, the two norms are equivalent:

$$\frac{1}{4}\|W\|_{\infty\to 1} \leqslant \|W\|_\square \leqslant \|W\|_{\infty\to 1}. \tag{3.5}$$

See Section 7.1 for more on connections between the cut norm and other norms.

It is not hard to see that for any two weighted graphs G and G' on the same set of nodes and with the same nodeweights,

$$d_\square(G, G') = \|W_G - W_{G'}\|_\square, \tag{3.6}$$

where W_G denotes the step function introduced in Section 3.1. The cut-norm therefore extends the distance $d_\square$ from weighted graphs to graphons.

We will also need the usual norms of W as a function from $[0,1]^2 \to \mathbb{R}$; we denote the corresponding L_1 and L_2 norms (with respect to the Lebesgue measure) by $\|W\|_1$ and $\|W\|_2$. The norm $\|.\|_2$ defines a Hilbert space, with inner product

$$\langle U, W\rangle = \int\limits_{[0,1]^2} U(x,y)W(y,x)\,dx\,dy.$$

3.3. Approximation by step functions

We need to extend two averaging operations from graphs to graphons. For $W \in \mathcal{W}$ and every partition $\mathcal{P} = (V_1, \ldots, V_q)$ of $[0,1]$ into measurable sets, we define a weighted graph on q nodes, denoted by $W/\mathcal{P}$, and called the quotient of W and $\mathcal{P}$, by setting $\alpha_i(W/\mathcal{P}) = \lambda(V_i)$ (where λ denotes the Lebesgue measure) and

$$\beta_{ij}(W/\mathcal{P}) = \frac{1}{\lambda(V_i)\lambda(V_j)} \int\limits_{V_i \times V_j} W(x,y)\,dx\,dy. \tag{3.7}$$

In addition to the quotient $W/\mathcal{P}$, we also consider the graphon $W_\mathcal{P}$ defined by

$$W_\mathcal{P}(x,y) = \beta_{ij}(W/\mathcal{P}) \quad \text{whenever } x \in V_i \text{ and } y \in V_j. \tag{3.8}$$

It is not hard to check that the averaging operation $W \mapsto W_\mathcal{P}$ is contractive with respect to the norms introduced above:

$$\|W_\mathcal{P}\|_\square \leqslant \|W\|_\square, \qquad \|W_\mathcal{P}\|_1 \leqslant \|W\|_1 \quad \text{and} \quad \|W_\mathcal{P}\|_2 \leqslant \|W\|_2. \tag{3.9}$$

The graphon $W_\mathcal{P}$ is an approximation of W by a step function with steps $\mathcal{P}$. Indeed, it is the best such approximation, at least in the L_2-norm:

$$\|W - W_\mathcal{P}\|_2 = \min_{U_\mathcal{P}} \|W - U_\mathcal{P}\|_2 \tag{3.10}$$

where the minimum runs over all step functions with steps $\mathcal{P}$ (this bound can easily be verified by varying the height of the steps in $U_\mathcal{P}$). While it is not true that $W_\mathcal{P}$ is the best approximation of W with step in $\mathcal{P}$ in the cut-norm, it is not off by more than a factor of two, as observed in [2]:

$$\|W - W_\mathcal{P}\|_\square \leqslant 2 \min_{U_\mathcal{P}} \|W - U_\mathcal{P}\|_\square \tag{3.11}$$

where the minimum runs over all step functions with steps $\mathcal{P}$. Indeed, combining the triangle inequality with the second bound in (3.9) and the fact that $(U_{\mathcal{P}})_{\mathcal{P}} = U_{\mathcal{P}}$, we conclude that $\|W - W_{\mathcal{P}}\|_\square \leqslant \|W - U_{\mathcal{P}}\|_\square + \|U_{\mathcal{P}} - W_{\mathcal{P}}\|_\square \leqslant \|W - U_{\mathcal{P}}\|_\square + \|U_{\mathcal{P}} - W\|_\square$, as required.

The definition of $W_{\mathcal{P}}$ raises the question on how well $W_{\mathcal{P}}$ approximates W. One answer to this question is provided by following lemma, which shows that W can be approximated arbitrarily well (pointwise almost everywhere) by interval step functions with equal steps. (The lemma is an immediate consequence of the almost everywhere differentiability of the integral function, see e.g. Theorem 7.10 of [29].)

Lemma 3.2. *For a positive integer n, let $\mathcal{P}_n$ be the partition of $[0, 1]$ into consecutive intervals of length $1/n$. For any $W \in \mathcal{W}$, we have $W_{\mathcal{P}_n} \to W$ almost everywhere.*

While the previous lemma gives the strong notion of almost everywhere convergence, it does not give any bounds on the rate of convergence. In this respect, the following lemma extending the weak Regularity Lemma from graphs to graphons, is much better. In particular, it gives a bound on the convergence rate which is independent of the graphon W.

Lemma 3.3. *(See [23].) For every graphon W and every $\varepsilon > 0$, there exists a partition $\mathcal{P}$ of $[0, 1]$ into measurable sets with at most $4^{1/\varepsilon^2}$ classes such that*

$$\|W - W_{\mathcal{P}}\|_\square \leqslant \varepsilon \|W\|_2.$$

With a slightly weaker bound for the number of classes, the lemma follows from Theorem 12 of [23] and the bound (3.11), or the results of [28]. As stated, it follows from Lemma 7.3 in Section 7.2, which generalizes both Lemma 3.3 and the analogous statement for graphs, Lemma 2.4. The Szemerédi Regularity Lemma [30] also extends to graphons in a straightforward way. See [28] for this and further extensions.

At the cost of increasing the bound on the number of classes, Lemmas 2.4 and 3.3 can be strengthened in several directions. In this paper, we need the following form, which immediately follows from Lemmas 2.4 and 3.3 by standard arguments, see Appendix A for the details.

Corollary 3.4. *Let $\varepsilon > 0$, and $q \geqslant 2^{20/\varepsilon^2}$. Then the following holds:*

(i) *For all graphons W, there is an equipartition $\mathcal{P}$ of $[0, 1]$ into q measurable sets such that*

$$\|W - W_{\mathcal{P}}\|_\square \leqslant \varepsilon \|W\|_2.$$

If we impose the additional constraint that $\mathcal{P}$ refines a given equipartition $\tilde{\mathcal{P}}$ of $[0, 1]$ into k measurable sets, it is possible to achieve this bound provided q is an integer multiple of k and $q/k \geqslant 2^{20/\varepsilon^2}$.

(ii) *For all weighted graphs G on at least q nodes there exists a partition $\mathcal{P} = (V_1, \dots, V_q)$ of $V(G)$ such that*

$$d_\square(G, G_{\mathcal{P}}) \leqslant \varepsilon \|G\|_2$$

and

$$\left|\sum_{u\in V_i}\alpha_u(G)-\frac{\alpha_G}{q}\right|<\alpha_{\max}(G)\quad \textit{for all } i=1,\dots,q. \tag{3.12}$$

3.4. The metric space of graphons

We now generalize the definition of the cut-distance (2.7) from graphs to graphons.

Let $\mathcal{M}$ denote the set of couplings of the uniform distribution on [0, 1] with itself, i.e., the set of probability measures on $[0,1]^2$ for which both marginals are the Lebesgue measure. (This is the natural generalization of overlays from graphs to graphons.) We then define:

$$\delta_\square(W,W')=\inf_{\mu\in\mathcal{M}}\sup_{S,T\subseteq[0,1]^2}\left|\int\limits_{\substack{(x,u)\in S\\(y,v)\in T}}\big(W(x,y)-W'(u,v)\big)\,d\mu(x,u)\,d\mu(y,v)\right|.$$

For two step functions, finding the optimal "overlay" can be described by specifying what fraction of each step of one function goes onto each step of the other function, which amounts to fractional overlay of the corresponding graphs. Hence the distances of two unlabeled weighted graphs and the corresponding interval step functions are the same:

$$\delta_\square(G,G')=\delta_\square(W_G,W_{G'}). \tag{3.13}$$

It will be convenient to use the hybrid notation $\delta_\square(U,G)=\delta_\square(U,W_G)$.

The next lemma gives an alternate representation of the distance $\delta_\square(W,W')$, in which "overlay" is interpreted in terms of measure-preserving maps rather than couplings. We need some definitions. Recall that a map $\phi:[0,1]\to[0,1]$ is *measure-preserving*, if the pre-image $\phi^{-1}(X)$ is measurable for every measurable set X, and $\lambda(\phi^{-1}(X))=\lambda(X)$. A *measure-preserving bijection* is a measure-preserving map whose inverse map exists and is also measurable (and then also measure-preserving). Finally, we consider certain very special measure-preserving maps defined as follows: Let us consider the partition $\mathcal{P}_n=(V_1,\dots,V_n)$ of [0, 1] into consecutive intervals of length $1/n$, and let π be a permutation of $[n]$. Let us map each V_i onto $V_{\pi(i)}$ by translation, to obtain a piecewise linear measure-preserving map $\tilde\pi:[0,1]\to[0,1]$. We call $\tilde\pi$ an *n-step interval permutation*.

For $W\in\mathcal{W}$ and $\phi:[0,1]\to[0,1]$, we define W^ϕ by $W^\phi(x,y)=W(\phi(x),\phi(y))$.

Lemma 3.5. *Let $U,W\in\mathcal{W}$. Then*

$$\delta_\square(U,W)=\inf_{\phi,\psi}\big\|U^\phi-W^\psi\big\|_\square \tag{3.14}$$

$\big($*where the infimum is over all measure-preserving maps* $\phi,\psi:[0,1]\to[0,1]\big)$

$$=\inf_{\psi}\big\|U-W^\psi\big\|_\square \tag{3.15}$$

$\big($*where the infimum is over all measure-preserving bijections* $\psi:[0,1]\to[0,1]\big)$

$$= \lim_{n\to\infty} \min_{\pi} \big\| U - W^{\tilde{\pi}} \big\|_{\square} \tag{3.16}$$

$\big($*where the minimum is over all permutations* π *of* $[n]\big)$.

The proof of the lemma is somewhat tedious, but straightforward, see Appendix A for details.

Note that, for $W \in \mathcal{W}$, for an n-step interval permutation $\tilde{\pi}$, and for the partition $\mathcal{P}_n$ of $[0,1]$ into consecutive intervals of lengths $1/n$, the graph $W^{\tilde{\pi}}/\mathcal{P}_n$ is obtained from $W/\mathcal{P}_n$ by a permutation of the nodes of $W/\mathcal{P}_n$. As a consequence, the identity (3.16) is equivalent to the following analogue of (2.8) for graphons:

$$\delta_{\square}(U, W) = \lim_{n\to\infty} \widehat{\delta}_{\square}(U/\mathcal{P}_n, W/\mathcal{P}_n). \tag{3.17}$$

Using Lemma 3.5, it is easy to verify that $\delta_{\square}$ satisfies the triangle inequality. Strictly speaking, the function $\delta_{\square}$ is just a pre-metric, and not a metric: Formula (3.15) implies that $\delta_{\square}(W, W') = 0$ whenever $W = W^{\phi}$ for some measure preserving transformation $\phi : [0,1] \to [0,1]$. Nevertheless, for the sake of linguistic simplicity, we will often refer to it as a distance or metric, taking the implicit identification of graphs or graphons with distance zero for granted. Note that the fact that there are graphons $W, W' \in \mathcal{W}$ which are different but have "distance" zero is not just a peculiarity of the limit: for simple graphs $\delta_{\square}(G, G')$ is zero if, e.g., G' is a blow-up of G (cf. (2.9)). We will say more about graphons with distance 0 in the next section.

It was proved in [28] that the metric space $(\mathcal{W}_{[0,1]}, \delta_{\square})$ is compact. Since $\mathcal{W}_I$ and $\mathcal{W}_{[0,1]}$ are linear images of each other, this immediately implies the following proposition.

Proposition 3.6. *Let I be a finite interval, and let $\mathcal{W}_I$ be the set of graphons with values in I. After identifying graphons with $\delta_{\square}$ distance zero, the metric space $(\mathcal{W}_I, \delta_{\square})$ is compact.*

3.5. Left versus metric convergence

We are now ready to state the analogue of Theorems 2.6 and 2.7 for graphons. We start with the analogue of Theorem 2.7.

Theorem 3.7. *Let $W, W' \in \mathcal{W}$, let $C = \max\{1, \|W\|_\infty, \|W'\|_\infty\}$, and let $k \geqslant 1$.*

(a) *If F is a simple graph with m edges, then*

$$\big| t(F, W) - t(F, W') \big| \leqslant 4mC^{m-1} \delta_{\square}(W, W'). \tag{3.18}$$

(b) *If $|t(F, W) - t(F, W')| \leqslant 3^{-k^2}$ for every simple graph F on k nodes, then*

$$\delta_{\square}(W, W') \leqslant \frac{22C}{\sqrt{\log_2 k}}.$$

The first statement of this theorem is closely related to the "Counting Lemma" in the theory of Szemerédi partitions, and gives an extension of a similar result of [26] for functions in $\mathcal{W}_{[0,1]}$ to general graphons. It shows that for any simple graph F, the function $W \mapsto t(F, W)$ is Lipschitz-continuous in the metric $\delta_{\square}$, and is reasonably easy to prove. By contrast, the proof of the second one is more involved and relies on our results on sampling.

In particular, we will need an analogue of Theorems 2.9 and 2.10 to sampling from graphons. These theorems are stated and proved in Section 4.5 (Theorems 4.7 and 4.6). Using Theorem 4.7, we then prove Theorem 3.7 in Section 4.6.

Theorem 3.7 immediately implies the analogue of Theorem 2.6 for graphons:

Theorem 3.8. *Let I be a finite interval and let (W_n) be a sequence of graphons in $\mathcal{W}_I$. Then the following are equivalent:*

(a) *$t(F, W_n)$ converges for all finite simple graphs F;*
(b) *W_n is a Cauchy sequence in the $\delta_\square$ metric;*
(c) *there exists a $W \in \mathcal{W}_I$ such that $t(F, W_n) \to t(F, W)$ for all finite simple graphs F.*

Furthermore, $t(F, W_n) \to t(F, W)$ for all finite simple graphs F for some $W \in \mathcal{W}$ if and only if $\delta_\square(W_n, W) \to 0$.

Note that by Eqs. (3.2) and (3.13), Theorems 2.6 and 2.7 immediately follow from Theorems 3.8 and 3.7. Together with Theorem 3.1, these results imply that after identifying graphons with distance zero, the set of graphons $\mathcal{W}_{[0,1]}$ is the completion of the metric space of simple graphs. Proposition 3.6, Eq. (3.13) and Lemma 3.2 easily imply that the existence of the limit object (Theorem 3.1) can be extended to convergent sequences of weighted graphs:

Corollary 3.9. *For any convergent sequence (G_n) of weighted graphs with uniformly bounded edgeweights there exists a graphon W such that $\delta_\square(W_{G_n}, W) \to 0$. Conversely, any graphon W can be obtained as the limit of a sequence of weighted graphs with uniformly bounded edgeweights. The limit of a convergent graph sequence is essentially unique: If $G_n \to W$, then also $G_n \to W'$ for precisely those graphons W' for which $\delta_\square(W, W') = 0$.*

As another consequence of Theorem 3.8, we get a characterization of graphons of distance 0:

Corollary 3.10. *For two graphons W and W' we have $\delta_\square(W, W') = 0$ if and only if $t(F, W) = t(F, W')$ for every simple graph F.*

Another characterization of such pairs is given in [14]: $\delta_\square(W, W') = 0$ if and only if there exists a third graphon U such that $W = U^\phi$ and $W' = U^\psi$ for two measure-preserving functions $\phi, \psi : [0, 1] \to [0, 1]$ (in other words, the infimum in (3.14) is a minimum if the distance is 0).

3.6. Examples

Example 3.11 *(Random graphs).* Let $\mathbf{G}(n, p)$ be a random graph on n nodes with edge density $0 \leqslant p \leqslant 1$; then it is not hard to prove (using high concentration results) that the sequence $(\mathbf{G}(n, p), n = 1, 2, \ldots)$ is convergent with probability 1. In fact, $t(F, \mathbf{G}(n, p))$ converges to $p^{|E(F)|}$ with probability 1, and so (with probability 1) $\mathbf{G}(n, p)$ converges to the constant function $W = p$.

Example 3.12 *(Quasirandom graphs).* A graph sequence is quasirandom with density p if and only if it converges to the constant function p. Quasirandom graph sequences have many other interesting characterizations in terms of edge densities of cuts, subgraphs, etc. [20]. In the second part of this paper we will discuss how most of these characterizations extend to convergent graphs sequences.

Example 3.13 *(Half-graphs).* Let $H_{n,n}$ denote the bipartite graph on $2n$ nodes $\{1, \dots, n, 1', \dots, n'\}$, where i is connected to j' if and only if $i \leqslant j$. It is easy to see that this sequence is convergent, and its limit is the function

$$W(x, y) = \begin{cases} 1, & \text{if } |x - y| \geqslant 1/2, \\ 0, & \text{otherwise.} \end{cases}$$

Example 3.14 *(Uniform attachment).* Various sequences of growing graphs, motivated by (but different from) internet models, are also convergent. We define a (dense) *uniform attachment graph sequence* as follows: if we have a current graph G_n with n nodes, then we create a new isolated node, and then for every pair of previously non-adjacent nodes, we connect them with probability $1/n$.

One can prove that with probability 1, the sequence (G_n) has a limit, which is the function $W(x, y) = \min(x, y)$. From this, it is easy to calculate that with probability 1, the edge density of G_n tends to $\int W = 1/3$. More generally, the density of copies of any fixed graph F in $G(n)$ tends (with probability 1) to $t(F, W)$, which can be evaluated by a simple integration.

4. Sampling

4.1. Injective and induced homomorphisms

In order to discuss sampling, we will consider not only the number of homomorphisms defined earlier, but also the number of injective and induced homomorphisms between two simple graphs F and G. We use $\operatorname{inj}(F, G)$ to denote the number of *injective* homomorphisms from F to G, and $\operatorname{ind}(F, G)$ to denote the number of those injective homomorphisms that also preserve non-adjacency (equivalently, the number of embeddings of F into G as an induced subgraph).

We will need to generalize these notions to the case where G is a weighted graph with nodeweights one and edgeweights $\beta_{ij}(G) \in \mathbb{R}$, where we define

$$\operatorname{inj}(F, G) = \sum_{\phi \in \operatorname{Inj}(F,G)} \prod_{uv \in E(F)} \beta_{\phi(u),\phi(v)}(G) \tag{4.1}$$

and

$$\operatorname{ind}(F, G) = \sum_{\phi \in \operatorname{Inj}(F,G)} \prod_{uv \in E(F)} \beta_{\phi(u),\phi(v)}(G) \prod_{uv \in E(\overline{F})} \bigl(1 - \beta_{\phi(u),\phi(v)}(G)\bigr). \tag{4.2}$$

Here $\operatorname{Inj}(F, G)$ denotes the set of injective maps from $V(F)$ to $V(G)$, and $E(\overline{F})$ consists of all pairs $\{u, v\}$ of distinct nodes such that $uv \notin E(F)$. We also introduce the densities

$$t_{\operatorname{inj}}(F, G) = \frac{\operatorname{inj}(F, G)}{(|V(G)|)_{|V(F)|}} \quad \text{and} \quad t_{\operatorname{ind}}(F, G) = \frac{\operatorname{ind}(F, G)}{(|V(G)|)_{|V(F)|}} \tag{4.3}$$

where $(n)_k = n(n-1)\cdots(n-k+1)$.

The quantities $t(F,G)$, $t_{\mathrm{inj}}(F,G)$ and $t_{\mathrm{ind}}(F,G)$ are closely related. It is easy to see that

$$t_{\mathrm{inj}}(F,G)=\sum_{F'\supset F} t_{\mathrm{ind}}(F',G) \quad \text{and} \quad t_{\mathrm{ind}}(F,G)=\sum_{F'\supset F}(-1)^{|E(F')\setminus E(F)|}t_{\mathrm{inj}}(F',G) \tag{4.4}$$

whenever F is simple and G is a weighted graph with nodeweights $\alpha_i(G)=1$. The quantity $t(F,G)$ is not expressible as a function of the values $t_{\mathrm{inj}}(F,G)$ (or $t_{\mathrm{ind}}(F,G)$), but for large graphs, they are essentially the same. Indeed, bounding the number of non-injective homomorphisms from $V(F)$ to $V(G)$ by $\binom{|V(F)|}{2}|V(G)|^{|V(F)|-1}$, one easily proves that

$$\left|t(F,G)-t_{\mathrm{inj}}(F,G)\right|<\frac{2}{|V(G)|}\binom{|V(F)|}{2}\|G\|_\infty^{|E(F)|}. \tag{4.5}$$

If all edgeweights of G lie in the interval [0, 1], this bound can be strengthened to

$$\left|t(F,G)-t_{\mathrm{inj}}(F,G)\right|<\frac{1}{|V(G)|}\binom{|V(F)|}{2}; \tag{4.6}$$

see [26] for a proof.

4.2. Sampling concentration

We will repeatedly use the following consequences of Azuma's Inequality:

Lemma 4.1. *Let $(\Omega,\mathcal{A},\mathsf{P})$ be a probability space, let k be a positive integer, and let $C>0$.*

(i) *Let $Z=(Z_1,\dots,Z_k)$, where $Z_1,\dots,Z_k$ are independent random variables, and Z_i takes values in some measure space $(\Omega_i,\mathcal{A}_i)$. Let $f:\Omega_1\times\cdots\times\Omega_k\to\mathbb{R}$ be a measurable function. Suppose that $|f(x)-f(y)|\leqslant C$ whenever $x=(x_1,\dots,x_k)$ and $y=(y_1,\dots,y_k)$ differ only in one coordinate. Then*

$$\mathsf{P}\big(f(Z)>\mathsf{E}\big[f(Z)\big]+\lambda C\big)<e^{-\lambda^2/2k} \quad \textit{and}$$
$$\mathsf{P}\big(\big|f(Z)-\mathsf{E}\big[f(Z)\big]\big|>\lambda C\big)<2e^{-\lambda^2/2k}. \tag{4.7}$$

(ii) *The bounds* (4.7) *also hold if $Z_1,\dots,Z_k$ are chosen uniformly without replacement from some finite set V and $|f(x)-f(y)|\leqslant C$ for all x and y which either differ in at most one component, or can be obtained from each other by permuting two components.*

Proof. Let $M_j=\mathsf{E}[f(Z)\mid Z_1,\dots,Z_j]$. Then $M_0,\dots,M_k$ is a martingale with bounded martingale differences (for the case (ii) this requires a little calculation which we leave to the reader). The statement now follows from Azuma's inequality for bounded martingales. □

4.3. Randomizing weighted graphs

Given a weighted graph H with nodeweights 1 and edgeweights in [0, 1], let $\mathbf{G}(H)$ denote the random simple graph with $V(G)=V(H)$ obtained by joining nodes i and j with probability $\beta_{ij}(H)$ (making an independent decision for every pair ij, and ignoring the loops in H).

We need two simple properties of this well-known construction. To state the first, we define the edit distance d_1 of two weighted graphs with the same node set $[n]$ and nodeweights 1 as

$$d_1(H_1, H_2) = \frac{1}{n^2} \sum_{i,j=1}^{n} \left|\beta_{ij}(H_1) - \beta_{ij}(H_2)\right|.$$

Lemma 4.2. *Let H_1 and H_2 be two weighted graphs on the same set of nodes with nodeweights* 1 *and with edgeweights in* $[0, 1]$. *Then $\mathbf{G}(H_1)$ and $\mathbf{G}(H_2)$ can be coupled so that*

$$\mathsf{E}\big(d_1\big(\mathbf{G}(H_1), \mathbf{G}(H_2)\big)\big) = d_1(H_1, H_2).$$

Proof. For every edge ij, we couple the decisions about the edge so that in both graphs this edge is inserted with probability $\min(\beta_{ij}(H_1), \beta_{ij}(H_2))$ and missing with probability $1 - \max(\beta_{ij}(H_1), \beta_{ij}(H_2))$. So the probability that the edge is present in exactly one of $\mathbf{G}(H_1)$ and $\mathbf{G}(H_2)$ is $|\beta_{ij}(H_1) - \beta_{ij}(H_2)|$, which proves the lemma. □

Lemma 4.3. *Let H be a weighted graph on n nodes with nodeweights* 1 *and with edgeweights in* $[0, 1]$. *Then*

$$\Pr\left(d_\square\big(H, \mathbf{G}(H)\big) < \frac{4}{\sqrt{n}}\right) > 1 - 2^{-n}.$$

Proof. Let $V(H) = V(\mathbf{G}(H)) = V$, let $\mu = 3/\sqrt{n}$, and let $\widetilde{H}$ be the graph obtained from H by deleting all diagonal entries in $\beta(H)$. Fix two sets $S, T \subseteq V$. For $i \neq j \in V$, let

$$X_{ij} = \begin{cases} 1 & \text{if } ij \in E(\mathbf{G}(H)), \\ 0 & \text{otherwise.} \end{cases}$$

Observe that the expectation of $e_{\mathbf{G}(H)}(S, T)$ is $e_{\widetilde{H}}(S, T)$. Since $e_{\mathbf{G}(H)}(S, T)$ is a function of the $\frac{n(n-1)}{2}$ independent random variables $(X_{ij})_{i<j}$ that changes by at most 2 if we change one of these variables, we may apply Lemma 4.1 to conclude that

$$\Pr\left(\left|e_{\mathbf{G}(H)}(S, T) - e_{\widetilde{H}}(S, T)\right| \geqslant \mu n^2\right) \leqslant 2\exp\left(-\frac{\mu^2 n^4}{4n(n-1)}\right) < \exp\left(-\frac{\mu^2 n^2}{4}\right).$$

(Here we used that $e^{-\mu^2 n/4} < 1/2$ in the last step.) Taking into account that the number of pairs (S, T) is 4^n, we concluded that the probability that $d_\square(\widetilde{H}, \mathbf{G}(H)) < \mu = 3/\sqrt{n}$ is larger than $1 - 4^n e^{-\mu^2 n^2/4} \geqslant 1 - 2^{-n}$. Since $d_\square(\widetilde{H}, H) \leqslant 1/n \leqslant 1/\sqrt{n}$, this completes the proof. □

4.4. W-random graphs

Given a graphon $W \in \mathcal{W}$ and a subset $S \subseteq [0, 1]$, we define the weighted graph $W[S]$ on node set S, all nodes with weight 1, in which $\beta_{xy}(W[S]) = W(x, y)$. If $W \in \mathcal{W}_{[0,1]}$, then we can construct a random *simple* graph $\widehat{W}[S]$ on S by connecting nodes X_i and X_j with probability $W(X_i, X_j)$ (making an independent decision for every pair).

This construction gives rise to two random graph models defined by the graphon W. For every integer $n > 0$, we generate a *W-random weighted graph* $\mathbf{H}(n, W)$ on nodes $\{1, \ldots, n\}$ as follows: We generate n independent samples $X_1, \dots, X_n$ from the uniform distribution on [0, 1], and consider $W[\{X_1, \dots, X_n\}]$ (renaming i the node X_i). If $W \in \mathcal{W}_{[0,1]}$, then we also define the *W-random (simple) graph* $\mathbf{G}(n, W) \cong \widehat{W}[\{X_1, \dots, X_n\}]$.

When proving concentration, it will often be useful to generate $\mathbf{G}(n, W)$ by first independently choosing n random variables $X_1, \dots, X_n$ and $n(n+1)/2$ random variables Y_{ij} $(i \leqslant j)$ uniformly at random from [0, 1], and then defining $\mathbf{G}(n, W)$ to be the graph with an edge between i and j whenever $Y_{ij} \leqslant W(X_i, X_j)$. This allows us to express the adjacency matrix of $\mathbf{G}(n, W)$ as a function of the *independent* random variables $Z_1 = (X_1, Y_{11})$, $Z_2 = (X_2, Y_{12}, Y_{22}), \dots, Z_n = (X_n, Y_{1n}, Y_{2n}, \dots, Y_{nn})$, as required for the application of Lemma 4.1(i).

It is easy to see that for every simple graph F with k nodes

$$\mathsf{E}\big(t_{\mathrm{inj}}\big(F, \mathbf{G}(n, W)\big)\big) = \mathsf{E}\big(t_{\mathrm{inj}}\big(F, \mathbf{H}(n, W)\big)\big) = t(F, W), \tag{4.8}$$

where the second equality holds for all W while the first requires $W \in \mathcal{W}_{[0,1]}$. From this we get that

$$\big|\mathsf{E}\big(t\big(F, \mathbf{H}(n, W)\big)\big) - t(F, W)\big| < \frac{2}{n}\binom{k}{2} \quad \text{if } \|W\|_\infty \leqslant 1, \tag{4.9}$$

and

$$\big|\mathsf{E}\big(t\big(F, \mathbf{G}(n, W)\big)\big) - t(F, W)\big| < \frac{1}{n}\binom{k}{2} \quad \text{if } W \in \mathcal{W}_{[0,1]}. \tag{4.10}$$

Concentration for the W-random graph $\mathbf{G}(n, W)$ was established in Theorem 2.5 of [26]. To get concentration for W-weighted random graphs, we use Lemma 4.1. This gives the following lemma, which also slightly improves the bound of Theorem 2.5 of [26] for W-random graphs. See Appendix A for details of the proof.

Lemma 4.4. *Let F be a simple graph on k nodes, let $0 < \varepsilon < 1$ and let $W \in \mathcal{W}$. Then*

$$\mathsf{P}\big(\big|t\big(F, \mathbf{H}(n, W)\big) - t(F, W)\big| > \varepsilon\big) \leqslant 2\exp\left(-\frac{\varepsilon^2}{11k^2}n\right) \quad \textit{if } \|W\|_\infty \leqslant 1, \tag{4.11}$$

and

$$\mathsf{P}\big(\big|t\big(F, \mathbf{G}(n, W)\big) - t(F, W)\big| > \varepsilon\big) \leqslant 2\exp\left(-\frac{\varepsilon^2}{4k^2}n\right) \quad \textit{if } W \in \mathcal{W}_{[0,1]}. \tag{4.12}$$

From this lemma we immediately get the following:

Theorem 4.5.

(a) *For any $W \in \mathcal{W}$, the graph sequence $\mathbf{H}(n, W)$ is convergent with probability* 1*, and its limit is the graphon W.*
(b) *For any $W \in \mathcal{W}_{[0,1]}$, the graph sequence $\mathbf{G}(n, W)$ is convergent with probability* 1*, and its limit is the graphon W.*

4.5. The distance of samples

The closeness of a sample to the original graph lies at the heart of many results in this and the companion paper [17]. We will prove these results starting with an extension of Theorem 2.10, which is an improvement of a theorem of Alon, Fernandez de la Vega, Kannan and Karpinski [1,2], as discussed earlier.

Theorem 4.6. *Let k be a positive integer.*

(i) *If $U \in \mathcal{W}$, then*

$$\big|\|\mathbf{H}(k,U)\|_\square - \|U\|_\square\big| \leqslant \frac{10}{k^{1/4}}\|U\|_\infty \tag{4.13}$$

with probability at least $1 - 2e^{-\sqrt{k}/8}$.

(ii) *If $U_1, U_2 \in \mathcal{W}_{[0,1]}$, then $\mathbf{G}(k,U_1)$ and $\mathbf{G}(k,U_2)$ can be coupled in such a way that*

$$\big|d_\square\big(\mathbf{G}(k,U_1),\mathbf{G}(k,U_2)\big) - \|U_1 - U_2\|_\square\big| \leqslant \frac{10}{k^{1/4}}, \tag{4.14}$$

with probability at least $1 - e^{-\sqrt{k}/8}$.

Proof. (i) Since the assertion is homogeneous in U, we may assume without loss of generality that $\|U\|_\infty = 1$. The proof proceeds in two steps: first we prove that for $k \geqslant 10^4$, we have

$$-\frac{2}{k} \leqslant \mathsf{E}\big(\|\mathbf{H}(k,U)\|_\square\big) - \|U\|_\square < \frac{8}{k^{1/4}}, \tag{4.15}$$

and then use Lemma 4.1 to prove concentration.

It turns out that the most difficult step is the proof of an upper bound on the expectation of $\|\mathbf{H}(k,U)\|_\square$. To obtain this bound, we use a refinement of the proof strategy of [2], using Lemma 3 from that paper as our starting point. The main difference between their and our proofs is that we first bound the expectation of $\|\mathbf{H}(k,U)\|_\square$, and only use Lemma 4.1 in the very end. This allows us to simplify their proof, giving at the same time a better dependence of our errors on k.

Let $X_1,\dots,X_k$ be i.i.d. random variables, distributed uniformly in [0, 1], let $B = B(\mathbf{X})$ be the $k \times k$ matrix with entries $B_{ij} = U(X_i, X_j)$, and for $S_1, S_2 \subset [k]$, let $B(S_1,S_2) = \sum_{i\in S_1,\, j\in S_2} B_{ij}$. Finally, given a set $S \subset [k]$, let $P(S)$ be the set of nodes $i \in [k]$ such that $B(\{i\},S) > 0$, and $N(S)$ be the set of nodes for which $B(\{i\},S) < 0$. We will prove upper and lower bounds on the expectation of $\|B\|_\square = \max_{S_1,S_2\subset[k]} |B(S_1,S_2)|$.

The lower bound is a simple sampling argument: consider two measurable subsets $T_1, T_2 \subset [0,1]$, and let $S_1 = \{i \in [k]\colon X_i \in T_1\}$, and similarly for S_2. Then

$$\begin{aligned}\mathsf{E}\|\mathbf{H}(k,U)\|_\square &\geqslant \frac{1}{k^2}\big|\mathsf{E}\big[B(S_1,S_2)\big]\big| = \left|\frac{k-1}{k}\int_{T_1\times T_2} U(x,y)\,dx\,dy + \frac{1}{k}\int_{T_1\cap T_2} U(x,x)\,dx\right| \\ &\geqslant \left|\int_{T_1\times T_2} U(x,y)\,dx\,dy\right| - \frac{2}{k}.\end{aligned}$$

Taking the supremum over all measurable sets $S_1, S_2 \subset [0,1]$, this proves the lower bound in (4.15).

To prove an upper bound on the expectation of $\|\mathbf{H}(k,U)\|_\square$, we start from Lemma 3 of [2]. In our context, it states that for a random subset $Q \subset [k]$ of size p, we have

$$B(S_1,S_2) \leqslant \mathsf{E}_Q\big[B\big(P(Q\cap S_2),S_2\big)\big] + \frac{k}{\sqrt{p}}\|B\|_F,$$

where $\|B\|_F = \sqrt{\sum_{i,j} B_{ij}^2} \leqslant k$. Inserting this inequality into itself, we obtain that

$$B(S_1,S_2) \leqslant \mathsf{E}_{Q_1,Q_2}\Big[\max_{T_i\subseteq Q_i} B\big(P(T_1),P(T_2)\big)\Big] + \frac{2k^2}{\sqrt{p}}.$$

In order to take the expectation over the random variables $X_1,\dots,X_k$, it will be convenient to decompose $P(T_1)$ and $P(T_2)$ into the parts which intersect $Q = Q_1 \cup Q_2$ and the parts which intersect $Q^c = [k] \setminus (Q_1 \cup Q_2)$. Let $P_Q(T) = P(T) \cap Q$ and $P_{Q^c}(T) = P(T) \setminus Q$. Since $|P(T)| \leqslant k$ and $|P(T) \setminus P_{Q^c}(T)| = |P_Q(T)| \leqslant 2p$ we have that $B(P(T_1),P(T_2)) \leqslant B(P_{Q^c}(T_1),P_{Q^c}(T_2)) + 4pk$, implying that

$$B(S_1,S_2) \leqslant \mathsf{E}_{Q_1,Q_2}\Big[\max_{T_i\subseteq Q_i} B\big(P_{Q^c}(T_1),P_{Q^c}(T_2)\big)\Big] + \frac{2k^2}{\sqrt{p}} + 4pk.$$

Applying this estimate to $-B$ and taking the maximum of the two bounds, this gives

$$\|B\|_\square \leqslant \mathsf{E}_{Q_1,Q_2}\Big[\max_{T_i\subseteq Q_i} \max\big\{B\big(P_{Q^c}(T_1),P_{Q^c}(T_2)\big), -B\big(N_{Q^c}(T_1),N_{Q^c}(T_2)\big)\big\}\Big] + \frac{2k^2}{\sqrt{p}} + 4pk, \tag{4.16}$$

where $N_{Q^c}(T) = N(T) \setminus Q$.

Consider a fixed pair of subsets $Q_1, Q_2 \subset [k]$. Fixing, for the moment, the random variables $\mathbf{X}_Q$, let us consider the expectation of, say, $B(P_{Q^c}(T_1),P_{Q^c}(T_2))$. For $T \subset [k]$, let $\mathbf{X}_T$ be the collection of random variables X_i with $i \in T$, and let

$$\mathcal{P}(\mathbf{X}_T) = \Big\{x \in [0,1]\colon \sum_{i\in T} U(x,X_i) > 0\Big\}.$$

Then $P_{Q^c}(T_i) = \{j \in Q^c\colon X_j \in \mathcal{P}(\mathbf{X}_{T_i})\}$, and

$$\begin{aligned}
&\mathsf{E}_{\mathbf{X}_{Q^c}}\big[B\big(P_{Q^c}(T_1),P_{Q^c}(T_2)\big)\big]\\
&\quad= \big|Q^c\big|\big(\big|Q^c\big|-1\big)\int_{\mathcal{P}(\mathbf{X}_{T_1})\times\mathcal{P}(\mathbf{X}_{T_2})} U(x,y)\,dx\,dy + \big|Q^c\big|\int_{\mathcal{P}(\mathbf{X}_{T_1})\cap\mathcal{P}(\mathbf{X}_{T_2})} U(x,x)\,dx\\
&\quad\leqslant k^2\|U\|_\square + k.
\end{aligned}$$

It is not hard to see that the random variable $Y = B(P_{Q^c}(T_1), P_{Q^c}(T_2))$ is highly concentrated. Indeed, consider Y as a function of the random variables in $\mathbf{X}_{Q^c}$. If we change one of these variables, then Y changes by at most $4k$, implying that with probability at least $1 - e^{-k\rho^2/32}$, $Y \leqslant \mathsf{E}(Y) + \rho k^2$. The same bound holds for the random variable $\tilde{Y} = -B(N_{Q^c}(T_1), N_{Q^c}(T_2))$. As a consequence, the maximum in (4.16) obeys this bound with probability at least $1 - 2^{2p+1}e^{-k\rho^2/32}$. Since $Y, \tilde{Y} \leqslant k^2$ for all $\mathbf{X}$, we conclude that

$$\mathsf{E}\big[\big\|\mathbf{H}(k,U)\big\|_\square\big] = \frac{1}{k^2}\mathsf{E}\big[\|B\|_\square\big] \leqslant \|U\|_\square + \frac{2}{\sqrt{p}} + \frac{4p}{k} + \frac{1}{k} + \rho + 2^{2p+1}e^{-k\rho^2/32}.$$

Choosing p and ρ of the form $p = \lceil \alpha\sqrt{k}\rceil$ and $\rho = \beta k^{-1/4}$ with $\alpha = (4\sqrt{\log 2}\,)^{-1}$ and $\beta = 4(\log 2)^{1/4} + 4/10$, this implies that for $k \geqslant 10^4$, we have

$$\mathsf{E}\big[\big\|\mathbf{H}(k,U)\big\|_\square\big] \leqslant \|U\|_\square + \frac{1}{k^{1/4}}\big(8(\log 2)^{1/4} + 0.534\ldots\big) \leqslant \|U\|_\square + \frac{8}{k^{1/4}}, \tag{4.17}$$

which is the upper bound in (4.15).

To prove concentration, we use that $\|\mathbf{H}(k,U)\|_\square$ changes by at most $\frac{4}{k}\|U\|_\infty$ if we change one of the random variables $X_1, \ldots, X_k$, so applying Lemma 4.1(i) we get that its values are highly concentrated around its expectation:

$$\mathsf{P}\bigg(\big|\big\|\mathbf{H}(k,U)\big\|_\square - \mathsf{E}\big[\big\|\mathbf{H}(k,U)\big\|_\square\big]\big| > \frac{2}{k^{1/4}}\bigg) < 2e^{-\sqrt{k}/8}.$$

This completes the proof of (i).

(ii) We couple $\mathbf{G}(k,U_1)$ and $\mathbf{G}(k,U_2)$ as follows: as in the proof of (i) we chose $X_1, \ldots, X_k$ to be i.i.d., distributed uniformly in $[0,1]$. In addition, we independently choose $k(k+1)/2$ random variables $Y_{ij} = Y_{ji}$ uniformly at random from $[0,1]$. In terms of these random variables, we then define G_1 to be the graph on $[k]$ which has an edge between i and j whenever $U_1(X_i, X_j) < Y_{ij}$, and G_2 to be the graph which has an edge between i and j whenever $U_2(X_i, X_j) < Y_{ij}$. Then G_1, G_2 is a coupling of $\mathbf{G}(k,U_1)$ and $\mathbf{G}(k,U_2)$, and

$$d_\square(G_1, G_2) = \frac{1}{k^2}\max_{S_1,S_2\subset[k]}\big|B(S_1,S_2)\big|,$$

where B is the matrix with entries $B_{ij} = \mathbf{1}_{Y_{ij}<U_1(X_i,X_j)} - \mathbf{1}_{Y_{ij}<U_2(X_i,X_j)}$. We again have to bound the expectation of $B(P_{Q^c}(T_1), P_{Q^c}(T_2))$. As before, we fix the variables X_i for $i \in Q$, but now we also fix all random variables Y_{ij} for which $\{i,j\}$ intersects Q. In order to calculate expectations with respect to the remaining random variables, we express $B(P_{Q^c}(T_1), P_{Q^c}(T_2))$ as the sum $\sum_{i,j\in Q^c} B^+_{i,j}(T_1,T_2)$, where

$$B^+_{i,j}(T_1,T_2) = B_{ij}\mathbf{1}_{i\in P(T_1)}\mathbf{1}_{j\in P(T_1)}.$$

For $i \in Q^c$, the event that $i \in P(T)$ for some $T \subset Q$ can then be re-expressed as the event that X_i lies in the set

$$\bigg\{x \in [0,1]\colon \sum_{j'\in T}\mathbf{1}_{Y_{ij'}<U_1(x,X_{j'})} > \sum_{j'\in T}\mathbf{1}_{Y_{ij'}<U_2(x,X_{j'})}\bigg\}.$$

Observing that this set only depends on the fixed random variables, we can now proceed as before to calculate the expectation of $B^{+}_{i,j}(T_1, T_2)$, and hence the expectation of $B(P_{Q^c}(T_1), P_{Q^c}(T_2))$. This, together with a similar (again much easier) calculation for the lower bound, leads to the estimate

$$-\frac{2}{k} \leqslant \mathsf{E}\bigl[d(G_1, G_2)_\square\bigr] - \|U_1 - U_2\|_\square \leqslant \frac{8}{k^{1/4}}, \tag{4.18}$$

as before valid for $k \geqslant 10^4$. Concentration is again proved with the help of Lemma 4.1. □

Essentially the same proof also gives Theorem 2.10:

Proof of Theorem 2.10. We generate the set S by choosing $X_1, \dots, X_k$ uniformly without replacement from V. In this way, we rewrite $d_\square(G_1[S], G_2[S])$ in terms of the matrix $B = B(\mathbf{X})$ with entries $B_{ij} = \beta_{X_i X_j}(G_1) - \beta_{X_i X_j}(G_2)$. Observing that

$$|B_{ij}| \leqslant C = 2\max\bigl\{\|G_1\|_\infty, \|G_2\|_\infty\bigr\},$$

we may proceed exactly as in the proof of Theorem 4.6(i), leading to the bound

$$-\frac{2C}{k} \leqslant \mathsf{E}\bigl[d_\square\bigl(G_1[S], G_2[S]\bigr)\bigr] - d_\square(G_1, G_2) \leqslant \frac{8C}{k^{1/4}}. \tag{4.19}$$

Using Lemma 4.1(ii) to prove concentration, we get the bound (2.13). □

We now come to the main theorem about sampling.

Theorem 4.7. *Let k be a positive integer.*

(i) *If $U \in \mathcal{W}$, then with probability at least $1 - e^{-k^2/(2\log_2 k)}$, we have*

$$\delta_\square\bigl(U, \mathbf{H}(k, U)\bigr) \leqslant \frac{10}{\sqrt{\log_2 k}}\|U\|_\infty.$$

(ii) *If $U \in \mathcal{W}_{[0,1]}$, then with probability at least $1 - e^{-k^2/(2\log_2 k)}$, we have*

$$\delta_\square\bigl(U, \mathbf{G}(k, U)\bigr) \leqslant \frac{10}{\sqrt{\log_2 k}}.$$

Proof. We again first bound expectations and then prove concentration, and as before, we assume without loss of generality that $\|U\|_\infty \leqslant 1$. Finally, we may assume that $k \geqslant 2^{25} \geqslant 10^4$, since otherwise the bounds of the theorem are trivial.

In a first step, we use the weak Szemerédi approximation in Lemma 3.3 and the sampling bound (4.15) to show that it is enough to consider the case where U is a step function. Indeed, given $\varepsilon > 0$, let U_1 be a step function with $q \leqslant 4^{1/\varepsilon^2}$ steps such that

$$\|U - U_1\|_\square \leqslant \varepsilon. \tag{4.20}$$

By the bound (4.15), we have that

$$\mathsf{E}\big[\delta_\square\big(\mathbf{H}(k,U_1),\mathbf{H}(k,U)\big)\big] \leqslant \mathsf{E}\big[\big\|\mathbf{H}(k,U-U_1)\big\|_\square\big] \leqslant \varepsilon + \frac{16}{k^{1/4}}.$$

As a consequence,

$$\begin{aligned}\mathsf{E}\big[\delta_\square\big(U,\mathbf{H}(k,U)\big)\big] &\leqslant \delta_\square(U,U_1) + \mathsf{E}\big[\delta_\square\big(U_1,\mathbf{H}(k,U_1)\big)\big] + \mathsf{E}\big[\delta_\square\big(\mathbf{H}(k,U_1),\mathbf{H}(k,U)\big)\big] \\ &\leqslant 2\varepsilon + \frac{16}{k^{1/4}} + \mathsf{E}\big[\delta_\square\big(U_1,\mathbf{H}(k,U_1)\big)\big]. \end{aligned} \tag{4.21}$$

We are thus left with the problem of sampling from the step function U_1. Let U_1 have steps $J_1,\dots,J_q \subseteq [0,1]$, and $\lambda(J_i)=\alpha_i$. Let $X_1,\dots,X_k$ be independent random variables that are uniformly distributed on [0, 1], and let Z_i be the number of points X_j that fall into the set J_i. It is easy to compute that

$$\mathsf{E}(Z_i)=\alpha_i k, \qquad \mathsf{Var}(Z_i)=\big(\alpha_i-\alpha_i^2\big)k<\alpha_i k.$$

Construct a partition of [0, 1] into measurable sets $J_1',\dots,J_q'$ such that $\lambda(J_i')=Z_i/k$ and

$$\lambda\big(J_i\cap J_i'\big)=\min(\alpha_i, Z_i/k),$$

and also construct a symmetric function $U'\in\mathcal{W}$ such that the value of U' on $J_i'\times J_j'$ is the same as the value of U_1 on $J_i\times J_j$. Then U' is a step function representation of $\mathbf{H}(k,U_1)$, and it agrees with U_1 on the set $Q=\bigcup_{i,j=1}^q (J_i\cap J_i')\times(J_j\cap J_j')$. Thus

$$\begin{aligned}\delta_\square\big(U_1,\mathbf{H}(k,U_1)\big) &\leqslant \|U_1-U'\|_\square \leqslant \|U_1-U'\|_1 \leqslant 2\big(1-\lambda(Q)\big) \\ &= 2\left(1-\left(\sum_i \min\left(\alpha_i,\frac{Z_i}{k}\right)\right)^2\right) \leqslant 4\left(1-\sum_i \min\left(\alpha_i,\frac{Z_i}{k}\right)\right) \\ &= 2\sum_i\left|\alpha_i-\frac{Z_i}{k}\right| \leqslant 2\left(q\sum_i\left(\alpha_i-\frac{Z_i}{k}\right)^2\right)^{1/2},\end{aligned}$$

which we rewrite as

$$\delta_\square\big(U_1,\mathbf{H}(k,U_1)\big)^2 \leqslant 4q\sum_i\left(\alpha_i-\frac{Z_i}{k}\right)^2.$$

The expectation of the right-hand side is

$$\frac{4q}{k^2}\sum_i \mathsf{Var}(Z_i) < \frac{4q}{k},$$

so by Cauchy–Schwarz

$$\mathsf{E}\big[\delta_\square\big(U_1,\mathbf{H}(k,U_1)\big)\big] \leqslant \sqrt{\frac{4q}{k}}. \tag{4.22}$$

Inserted into (4.21), this gives

$$\mathsf{E}\big[\delta_\square\big(U,\mathbf{H}(k,U)\big)\big] \leqslant 2\varepsilon + \frac{16}{k^{1/4}} + \sqrt{\frac{4n}{k}} \leqslant 2\varepsilon + \frac{16}{k^{1/4}} + \frac{2}{k^{1/2}} 2^{1/\varepsilon^2}.$$

Choosing $\varepsilon = 2/(\log_2 k)$ and recalling that $k \geqslant 2^{25}$, this gives the upper bound

$$\mathsf{E}\big[\delta_\square\big(U,\mathbf{H}(k,U)\big)\big] \leqslant \frac{1}{\sqrt{\log_2 k}}\left(4 + (16+2)\frac{\sqrt{\log_2 k}}{k^{1/4}}\right) \leqslant \frac{6}{\sqrt{\log_2 k}}. \tag{4.23}$$

Observing that $\delta_\square(U,\mathbf{H}(k,U))$ changes by at most $4/k$ if one of the random variables X_i changes its value, we finally use Lemma 4.1(i) to complete the proof of the first statement.

(ii) The proof of the second statement is completely analogous: we first show that

$$\mathsf{E}\big[\delta_\square\big(U,\mathbf{G}(k,U)\big)\big] \leqslant \frac{1}{\sqrt{\log_2 k}}\left(4 + (8+2)\frac{\sqrt{\log_2 k}}{k^{1/4}}\right) \leqslant \frac{5}{\sqrt{\log_2 k}}, \tag{4.24}$$

and then use Lemma 4.1(i) to prove concentration. □

The proof again generalizes to samples from weighted graphs, this time leading to Theorem 2.9.

Proof of Theorem 2.9. Again, the proof is analogous to the proof of statement (i) above, except that we now use the original Frieze–Kannan Lemma (Lemma 2.4) for graphs instead of our Weak Szemerédi Lemma (Lemma 3.3) for graphons. One also needs to generalize the bound (4.22) to samples $X_1,\dots,X_k$ chosen uniformly without replacement from $V(G)$, but this is again straightforward: the random variable Z_i is now the number of points that fall into the class V_i of a weak Szemerédi partition $\mathcal{P}=(V_1,\dots,V_q)$, and its variance is bounded by $\alpha_i k$, where $\alpha_i = |V_i|/|V(G)|$. Continuing as in the above proof, these considerations now lead to the bound

$$\mathsf{E}\big[\delta_\square\big(G,\mathbb{G}(k,G)\big)\big] \leqslant \frac{1}{\sqrt{\log_2 k}}\left(4 + (16+2)\frac{\sqrt{\log_2 k}}{k^{1/4}}\right) \leqslant \frac{6}{\sqrt{\log_2 k}}, \tag{4.25}$$

where we again assume that the weights have been rescaled so that $\|G\|_\infty = 1$. Using Lemma 4.1(ii) to show concentration, this gives the bound (2.12). □

4.6. *Proof of Theorem 3.7*

4.6.1. *Proof of Theorem 3.7(a)*

Let $V(F)=[n]$ and $E(F)=\{e_1,\dots,e_m\}$. For $t=1,\dots,m$, let i_t, j_t be the endpoints of the edge e_t, and let $E_t=\{e_1,\dots,e_t\}\subset E(F)$. Then $t(F,W)-t(F,W')$ can be rewritten as

$$t(F,W)-t(F,W') = \int\limits_{[0,1]^n} \left(\prod_{ij\in E(F)} W(x_i,x_j) - \prod_{ij\in E(F)} W'(x_i,x_j)\right) dx_1\dots dx_n$$

$$= \sum_{t=1}^{m} \int_{[0,1]^n} \prod_{s<t} W(x_{i_s}, x_{j_s}) \times \prod_{s>t} W'(x_{i_s}, x_{j_s}) \big(W(x_{i_t}, x_{j_t}) - W'(x_{i_t}, x_{j_t}) \big) \, dx_1 \ldots dx_n. \quad (4.26)$$

Take any term in this sum, and for notational convenience, assume that $i_t = 1$ and $j_t = 2$. Let $X(x_1, x_3, \ldots, x_n)$ be the product of those factors in $\prod_{s<t} W(x_{i_s}, x_{j_s}) \prod_{s>t} W'(x_{i_s}, x_{j_s})$ that contain x_1, and let $Y(x_2, \ldots, x_n)$ be the product of the rest. Then we have

$$\int_{[0,1]^n} \prod_{s<t} W(x_{i_s}, x_{j_s}) \prod_{s>t} W'(x_{i_s}, x_{j_s}) \big(W(x_{i_t}, x_{j_t}) - W'(x_{i_t}, x_{j_t}) \big) \, dx_1 \ldots dx_n$$
$$= \int_{[0,1]^{n-2}} \Bigg(\int_{[0,1]^2} X(x_1, x_3, \ldots, x_n) Y(x_2, \ldots, x_n) \big(W(x_1, x_2) - W'(x_1, x_2) \big) \, dx_1 \, dx_2 \Bigg) dx_3 \ldots dx_n.$$

Here the interior integral is bounded by

$$\Bigg| \int_{[0,1]^2} X(x_1, x_3, \ldots, x_n) Y(x_2, \ldots, x_n) \big(W(x_1, x_2) - W'(x_1, x_2) \big) \, dx_1 \, dx_2 \Bigg| \leqslant \|X\|_\infty \|Y\|_\infty \|W - W'\|_{\infty \to 1}.$$

Substituting into (4.26), and using that

$$\|X\|_\infty \|Y\|_\infty \leqslant \|W\|_\infty^{t-1} \|W'\|_\infty^{m-t} \leqslant C^{m-1}$$

and (by (3.5))

$$\|W - W'\|_{\infty \to 1} \leqslant 4 \|W - W'\|_\square,$$

we get

$$\big| t(F, W) - t(F, W') \big| \leqslant 4mC^{m-1} \|W - W'\|_\square.$$

Using the representation in Lemma 3.5 for the $\delta_\square$ distance and the fact that $t(F, W) = t(F, W^\phi)$ whenever ϕ is a measure-preserving function from $[0, 1]$ to $[0, 1]$, this bound implies the bound (3.18). $\square$

Remark 4.8. The above proof can easily be generalized to show that

$$\big| t(F, W) \big| \leqslant 4 \|W\|_\infty^{|E(F)|-1} \|W\|_\square \quad (4.27)$$

for all $W \in \mathcal{W}$ and all simple graphs F. Also, it is not hard to show that the factor 4 in (3.18) and (4.27) is not needed if W and W' are non-negative.

4.6.2. Proof of Theorem 3.7(b)

Without loss of generality, we may assume that $\|W\|_\infty, \|W'\|_\infty \leqslant 1$ (otherwise we just rescale both W and W' by the maximum of these two numbers). Let $U, U' \in \mathcal{W}_{[0,1]}$ be the graphons $U = \frac{1}{2}W + \frac{1}{2}$ and $U' = \frac{1}{2}W' + \frac{1}{2}$. Then $\delta_\square(W, W') = 2\delta_\square(U, U')$, so it is enough to prove that $\delta_\square(U, U') \leqslant 11/\sqrt{\log_2 k}$. We will prove this bound by relating the distance of U and U' to the distance of the random graphs $\mathbf{G}(k, U)$ and $\mathbf{G}(k, U')$. To this end, we need an expression for the probability that $\mathbf{G}(k, U)$ is equal to some given graph F on k nodes. We first use the relations (4.8) to express $t(F, U)$ as a sum over graphs G on k nodes. Combined with (4.4) and the fact that for all graphs F' and G on k nodes, $t_{\text{ind}}(F', G) = 0$ unless G is isomorphic to F', we have that

$$\begin{aligned} t(F, U) &= \sum_G \mathsf{P}\big(\mathbf{G}(k, U) = G\big) t_{\text{inj}}(F, G) \\ &= \sum_G \sum_{F' \supset F} \mathsf{P}\big(\mathbf{G}(k, U) = G\big) t_{\text{ind}}(F', G) \\ &= \frac{1}{k!} \sum_{F' \supset F} \mathsf{P}\big(\mathbf{G}(k, U) = F'\big)\big(\text{ind}(F', F')\big)^2, \end{aligned}$$

where $\text{ind}(\cdot,\cdot)$ is defined in (4.2). With the help of inclusion–exclusion, this leads to

$$\mathsf{P}\big(\mathbf{G}(k, U) = F\big) = k! \sum_{F' \supset F} (-1)^{|E(F') \setminus E(F)|} \big(\text{ind}(F', F')\big)^{-2} t(F', U),$$

which in turn implies that

$$\sum_F \big|\mathsf{P}\big(\mathbf{G}(k, U)) = F\big) - \mathsf{P}\big(\mathbf{G}(k, U') = F\big)\big| \leqslant k! \sum_{\substack{F, F': \\ E(F') \supset E(F)}} \big|t(F', U) - t(F', U')\big|$$

where the sum runs over graphs F and F' on k nodes. To continue, we need to relate the homomorphism densities of U and U' to those of W and W'. To this end, we insert the relation $U = \frac{1}{2}(W + 1)$ into the definition of $t(F', U)$. For a graph F' on k nodes, this leads to the identity

$$t(F', U) = 2^{-|E(F')|} \sum_{F'' \subset F'} t(F'', W)$$

where the sum goes over all subgraphs that have the same node set as F'. Using the assumption of the theorem, we thus obtain the bound $|t(F', U) - t(F', U')| \leqslant 3^{-k^2}$ which in turn implies that

$$\sum_F \big|\mathsf{P}\big(\mathbf{G}(k, U) = F\big) - \mathsf{P}\big(\mathbf{G}(k, U') = F\big)\big| \leqslant k! \sum_{\substack{F, F': \\ E(F) \subset E(F')}} 3^{-k^2} = k! 3^{-k^2} 3^{k(k-1)/2}.$$

Bounding $k!$ (rather crudely) by $3^{k^2/2}$, we note that the right-hand side is smaller than $3^{-k/2}$.

As a consequence, $\mathbf{G}(k,U)$ and $\mathbf{G}(k,U')$ can be coupled in such a way that $\mathbf{G}(k,U) = \mathbf{G}(k,U')$ with probability at least $1-3^{-k/2}$. Combined with the triangle inequality and the bound (4.24), we obtain that

$$\delta_\square(U,U') \leqslant \mathsf{E}\delta_\square\big(\mathbf{G}(k,U),\mathbf{G}(k,U')\big) + \mathsf{E}\big[\delta_\square\big(\mathbf{G}(k,U),U\big)\big] + \mathsf{E}\big[\delta_\square\big(\mathbf{G}(k,U'),U'\big)\big]$$
$$\leqslant 3^{-k/2} + \frac{10}{\sqrt{\log_2 k}} \leqslant \frac{11}{\sqrt{\log_2 k}}. \qquad \square$$

5. Convergence in norm and uniform Szemerédi partitions

5.1. Comparison of fractional and non-fractional overlays

Let G and G' be two graphs on n nodes, both with nodeweights $1/n$. Consider any labeling that attains the minimum in the definition (2.6) of $\widehat{\delta}_\square$, and identify the nodes of G and G' with the same label. In this case, we say that G and G' are *optimally overlaid.*

In addition to this distance we also defined the distance $\delta_\square$, given in terms of fractional overlays $X \in \mathcal{X}(G,G')$, see (2.7). Since every bijection between the nodes of G_1 and G_2 defines a fractional overlay X, we trivially have that

$$\delta_\square(G,G') \leqslant \widehat{\delta}_\square(G,G').$$

This inequality can be strict. Let $G = K_2$, and let G' be a graph with two non-adjacent nodes but with a loop at each node. It is easy to see that $\widehat{\delta}_\square(G,G') = 1/4$, but $\delta_\square(G,G') = 1/8$ (the best fractional overlay is $X_{iu} = 1/4$ for all $i \in V(G)$ and $u \in V(G')$). To give an example without loops, let $G = K_{3,3}$, and let G' consist of two disjoint triangles Δ_1 and Δ_2. There are only two essentially different ways to overlay these graphs; the better one maps two nodes of Δ_i into one color class of $K_{3,3}$ and the third one into the other color class. The number of edges in G' between Δ_1 and Δ_2 is 5, whence $\widehat{\delta}_\square(G,G') \geqslant 5/36$. (One can check that equality holds.) On the other hand, let us double each node in both graphs, to get $G(2) = K_{6,6}$ and $G'(2) = \Delta_1(2) \cup \Delta_2(2)$. Let us map one copy of each twin node of $G'(2)$ into one color class of $K_{6,6}$, and its pair into the other color class. Case distinction shows that the worst choice for S and T in the definition of $d_\square$ is $S = V(\Delta_1(2))$ and $T = V(\Delta_2(2))$, showing that

$$\delta_\square(G,G') \leqslant \widehat{\delta}_\square\big(G(2),G'(2)\big) = \frac{18}{144} = \frac{1}{8} < \frac{5}{36}.$$

We have no example disproving the possibility that $\widehat{\delta}_\square(G_1,G_2) = O(\delta_\square(G_1,G_2))$, but we are only able to prove a much weaker inequality given in Theorem 2.3. We start with a simple but very weak bound we will need.

Lemma 5.1. *Let G_1 and G_2 be weighted graphs on n nodes. If both G_1 and G_2 have nodeweights one, then*

$$\widehat{\delta}_\square(G_1,G_2) \leqslant n^6 \delta_\square(G_1,G_2).$$

Proof. Let (X_{ui}) be an optimal fractional overlay of G_1 and G_2, normalized in such a way that $\sum_i X_{ui} = \sum_v X_{vj} = 1/n$ for all n. We claim that there is a bijection $\pi : V(G_1) \to V(G_2)$ such that $X_{u\pi(u)} \geqslant 1/n^3$ for all $u \in V(G_1)$. This follows from a routine application of the Marriage Theorem: if there is no such bijection, then there are two sets $S \subseteq V(G_1)$ and $T \subseteq V(G_2)$ such that $|S| + |T| > n$ and $X_{st} < 1/n^3$ for all $s \in S$ and all $t \in T$. But then

$$\frac{|S|}{n} = \sum_{u\in S} \sum_{i\in V(G_2)} X_{iu} = \sum_{u\in S} \sum_{i\in T} X_{iu} + \sum_{u\in S} \sum_{i\in V(G_2)\setminus T} X_{iu}$$
$$\leqslant \frac{1}{n^3}|S| \cdot |T| + \frac{|V(G_2) \setminus T|}{n} < \frac{1}{n} + \frac{|S|-1}{n},$$

a contradiction. Thus a map π with the desired properties exists.

Let G_1' be the image of G_1 under this map. Then

$$\begin{aligned}\widehat{\delta}_\square(G_1, G_2) &\leqslant \max_{s\in V(G_1)} \max_{t\in V(G_2)} \big|\beta_{st}(G_1) - \beta_{\pi(s)\pi(t)}(G_2)\big| \\ &\leqslant \max_{s\in V(G_1)} \max_{t\in V(G_2)} n^6 X_{s\pi(s)} X_{t\pi(t)} \big|\beta_{st}(G_1) - \beta_{\pi(s)\pi(t)}(G_2)\big| \\ &\leqslant n^6 \delta_\square(G_1, G_2),\end{aligned}$$

which proves the lemma. □

Now we are able to prove Theorem 2.3, which shows that the two distances δ and $\widehat{\delta}$ define the same topology. As the reader may easily verify, the proof below gives an exponent of 1/3 instead of the exponent of 1/67 from Theorem 2.3 if the number of nodes is large enough. However, even under this assumption, we could not obtain a linear bound, i.e., bi-Lipschitz equivalence.

Proof of Theorem 2.3. The first inequality, as remarked before, is trivial. To prove the second, write $\delta_\square(G_1, G_2)^{1/67} = \varepsilon$. If $n \leqslant \varepsilon^{-11}$, then the bound follows by Lemma 5.1, and if $\varepsilon \geqslant 2/36$, the bound is trivial, so we may assume that

$$n > \varepsilon^{-11} \quad \text{and} \quad \varepsilon \leqslant \frac{1}{16}. \tag{5.1}$$

Consider an optimal overlay of W_{G_1} and W_{G_2} (in other words, consider a measure-preserving bijection $\phi : [0, 1] \to [0, 1]$) such that

$$\big\| W_{G_1} - W_{G_2}^{\phi} \big\|_\square = \delta_\square(G_1, G_2) = \varepsilon^{67}.$$

Let us select a set Z of $k = \lceil n/\varepsilon \rceil$ random points from [0, 1]. Let $H_1 = W_{G_1}[Z]$ and $H_2 = W_{G_2^\phi}[Z]$. Then by Theorem 4.6, we get that with probability at least $1 - e^{-\frac{1}{8}\sqrt{n/\varepsilon}} \geqslant 1 - 8\varepsilon^6$,

$$d_\square(H_1, H_2) \leqslant \big\| W_{G_1} - W_{G_2}^{\phi} \big\|_\square + \frac{20}{k^{1/4}} = \varepsilon^{67} + \frac{20}{k^{1/4}}.$$

Each element $z \in Z$ corresponds to a node $i_z \in V(G_1)$ and a node $j_z \in V(G_2)$. These pairs (i_z, j_z) form a bipartite graph F with color classes $V(G_1)$ and $V(G_2)$ (which we assume are disjoint).

Claim 5.2. *With probability at least* $1-(2/e)^{-2n}$, *the bipartite graph* F *has a matching of size at least* $(1-2\varepsilon)n$.

To prove this claim, we use König's Theorem: if F does not contain a matching of size $(1-2\varepsilon)n$, then its edges can be covered by a set X of nodes with $|X|<(1-2\varepsilon)n$. Let $Y_i = V(G_i)\setminus X$. Then there is no edge of F between Y_1 and Y_2.

On the other hand, $|Y_1|+|Y_2|\geqslant(1+2\varepsilon)n$. Let $J_i\subseteq[0,1]$ be the union of intervals representing Y_i in W_{G_i}, so that $\lambda(J_1)+\lambda(J_2)\geqslant 1+2\varepsilon$, and hence also $\lambda(J_1)+\lambda(\phi(J_2))\geqslant 1+2\varepsilon$. This implies that $\lambda(J_1\cap\phi(J_2))\geqslant 2\varepsilon$. The random set Z avoided this intersection; the probability of this happening is at most

$$(1-2\varepsilon)^k < e^{-2\varepsilon k}\leqslant e^{-2n}.$$

Since there are at most 4^n pairs of sets (Y_1,Y_2), the probability that F does not have a matching of cardinality at least $(1-2\varepsilon)n$ is less than $4^n e^{-2n}$. This proves the claim.

Now let $(i_1,j_1),\dots,(i_m,j_m)$ be a maximum matching in F. With positive probability, we have both $d_\square(H_1,H_2)\leqslant\varepsilon^{67}+20/k^{1/4}$ and $m\geqslant(1-\frac{2n}{k})n$. Fix a choice of Z for which this happens. Let $(i_{m+1},j_{m+1}),\dots,(i_n,j_n)$ be an arbitrary pairing of the remaining nodes of $V(G_1)$ and $V(G_2)$. We claim that the pairing $\pi: i_r\mapsto j_r$ gives an overlay of G_1 and G_2 with small $d_\square$ distance.

Let $S,T\subseteq V(G_2)$, and let $S'=S\cap\{j_1,\dots,j_m\}$, $T'=T\cap\{j_1,\dots,j_m\}$. Then $|S\setminus S'|\leqslant\frac{2n^2}{k}$, and $|T\setminus T'|\leqslant\frac{2n^2}{k}$, and hence

$$\bigl|e_{G_1'}(S,T)-e_{G_2}(S,T)\bigr|\leqslant\bigl|e_{G_1'}(S',T')-e_{G_2}(S',T')\bigr|+2\biggl(2n\frac{2n^2}{k}+\biggl(\frac{2n^2}{k}\biggr)^2\biggr).$$

Here

$$\bigl|e_{G_1'}(S',T')-e_{G_2}(S',T')\bigr|=\bigl|e_{H_1}(S',T')-e_{H_2}(S',T')\bigr|\leqslant d_\square(H_1,H_2)k^2\leqslant\biggl(\varepsilon^{67}+\frac{20}{k^{1/4}}\biggr)k^2,$$

and hence

$$\begin{aligned}\frac{|e_{G_1'}(S,T)-eG_2(S,T)|}{n^2}&\leqslant\frac{8n}{k}+\frac{8n^2}{k^2}+\frac{\varepsilon^{67}k^2}{n^2}+\frac{20k^{\frac{7}{4}}}{n^2}\\&\leqslant 8\varepsilon+8\varepsilon^2+\varepsilon^{67}\bigl(1+\varepsilon^{-1}\bigr)^2+20\bigl(1+\varepsilon^{-1}\bigr)^{7/4}n^{-1/4}\\&\leqslant 8\varepsilon+8\varepsilon^2+\varepsilon^{65}(1+\varepsilon)^2+20\varepsilon(1+\varepsilon)^{7/4}\leqslant 32\varepsilon\end{aligned}$$

where we used (5.1) in the last two bounds. □

5.2. Convergence in norm

Let (G_n) be a convergent sequence of weighted graphs. Theorem 3.8 then implies that there exists a graphon $W\in\mathcal{W}$ such that $W_{G_n}\to W$ in the $\delta_\square$ distance. This does not imply, however, that $W_{G_n}\to W$ in the $\|\cdot\|_\square$ norm. It turns out, however, that the graphs in the sequence (G_n)

can be relabeled in such a way that this becomes true, provided (G_n) has *no dominant nodes* in the sense that

$$\frac{\max_i \alpha_i(G_n)}{\alpha_{G_n}} \to 0$$

as $n \to \infty$. This is the content of the following lemma, which will be useful when discussing testability.

Lemma 5.3. *Let (G_n) be a sequence of weighted graphs with uniformly bounded edgeweights, and no dominant nodes. If*

$$\delta_\square(U, W_{G_n}) \to 0$$

for some $U \in \mathcal{W}$, then the graphs in the sequence (G_n) can be relabeled in such a way that the resulting sequence (G'_n) of labeled graphs converges to U in the cut-norm:

$$\|U - W_{G'_n}\|_\square \to 0.$$

Proof. We first prove the lemma for graphs with nodeweights one. Let $m(n) = |V(G_n)|$, and let $\mathcal{P}_{m(n)}$ be a partition of $[0,1]$ into consecutive intervals of length $1/m(n)$. By Lemma 3.2, we have that $\|U - U_{\mathcal{P}_{m(n)}}\|_\square \to 0$, so combined with the assumption that $\delta_\square(U, W_{G_n}) \to 0$ we conclude that $\delta_\square(U_{\mathcal{P}_{m(n)}}, W_{G_n}) \to 0$. But the left-hand side can be expressed as the distance of two weighted graphs on the same number of nodes, $\delta_\square(U_{\mathcal{P}_{m(n)}}, W_{G_n}) = \delta_\square(U/\mathcal{P}_{m(n)}, G_n)$, so by Theorem 2.3, we get that

$$\hat{\delta}_\square(U/\mathcal{P}_{m(n)}, G_n) \to 0.$$

This means that the graphs in the sequence (G_n) can be relabeled in such a way that for the resulting graph sequence, (G'_n), we have

$$\|U_{\mathcal{P}_{m(n)}} - W_{G'_n}\|_\square = d_\square\big(U/\mathcal{P}_{m(n)}, G'_n\big) \to 0.$$

Combined with the fact that $\|U - U_{\mathcal{P}_{m(n)}}\|_\square \to 0$, this gives the statement of the lemma for graphs with nodeweights one.

To prove the lemma for general sequences without dominant nodes, we use the Weak Regularity Lemma to approximate (G_n) by a sequence of graphs $(\tilde{G}_n)$ with nodeweights one. Indeed, let us assume without loss of generality that all graphs in the sequence (G_n) have total nodeweight $\alpha_{G_n} = 1$. Define $a_n = \max_i \alpha_i(G_n)$, and choose ε_n in such a way that $\varepsilon_n \to 0$ and $a_n 2^{40/\varepsilon_n^2} \to 0$ as $n \to \infty$. With the help of Corollary 3.4(ii), we then construct a partition $\mathcal{P}_n$ of $V(G_n)$ into $q_n \leqslant 2^{20/\varepsilon_n^2}$ classes such that $d_\square(G_n, (G_n)_{\mathcal{P}_n}) \to 0$ and the classes in $\mathcal{P}_n$ have almost equal weights (in the sense that $|\sum_{x \in V_i} \alpha_x(G_n) - \sum_{y \in V_j} \alpha_y(G_n)| \leqslant a_n$ for all $i, j \in [q_n]$). Consider the sequence of graphs $\tilde{G}_n$ that are obtained from $G_n/\mathcal{P}_n$ by changing all nodeweights to 1. Since the classes of $\mathcal{P}_n$ have almost equal weights, we have that $\|W_{\tilde{G}_n} - W_{G_n/\mathcal{P}_n}\|_\square \leqslant q_n^2 a_n \to 0$ which in turn implies that $\delta_\square(G_n, \tilde{G}_n) \to 0$. Thus $(\tilde{G}_n)$ is a sequence of weighted graphs with nodeweights one which converges to U, implying that it can be reordered in such a way that $\|W_{\tilde{G}_n} - U\|_\square \to 0$. But this means that $G_n/\mathcal{P}_n$ can be relabeled in such a way that $\|W_{G_n/\mathcal{P}_n} - U\|_\square \to 0$, which in turn implies that G_n itself can be relabeled so that $\|W_{G_n} - U\|_\square \to 0$, as desired. $\square$

The previous lemma suggests that we extend the definition of the distance $\hat{\delta}$ to the case when one of the arguments is a graphon U:

$$\hat{\delta}_\square(U, G) = \min_{G'} \|U - W_{G'}\|_\square \tag{5.2}$$

where the minimum goes over all relabelings G' of G. Then the lemma asserts that if G_n is a convergent graph sequence with uniformly bounded edgeweights and no dominant nodes, then $\hat{\delta}_\square(U, G_n) \to 0$.

In the special case of nodeweights one, Theorem 2.3 and Lemma 5.3 naively suggest the stronger statement that $\hat{\delta}_\square(U, G)$ can be bounded by a function $f(\delta_\square(U, W_G), |V(G)|)$ such that $f(x, n) \to 0$ if $x \to 0$ and $n \to \infty$. However, this is false, as the following example shows.

Example 5.4. Let $G = K_{n,n}$ be the complete bipartite graph on $2n$ nodes, and let $H = K_{nm,nm}$ be the complete bipartite graph on $2nm$ nodes, where H has randomly labeled nodes, and let $U = W_H$. Then $\delta_\square(U, W_G) = 0$. But it is not hard to show that for every n, if m is sufficiently large, then with large probability, $\|W_H - W_{G'}\|_\square \geqslant 1/10$ for every relabeling G' of G, implying that $\hat{\delta}_\square(U, G) \geqslant 1/10$, see Appendix A for details.

5.3. Convergent Szemerédi partitions

Given Lemma 5.3, we now are ready to prove Theorem 2.8 about the convergence of the quotient graphs of suitably chosen Szemerédi partitions for a convergent graph sequence (G_n).

We start by proving the easier direction, namely that convergent quotients imply convergence of the sequence (G_n).

Let $\varepsilon > 0$, and let q be such that the conditions (i) and (ii) of Theorem 2.8 hold. For a fixed q, convergence of $G_n/\mathcal{P}_n$ is equivalent to convergence of all edgeweights and nodeweights, which in turn implies convergence in the $\delta_\square$ distance. Let n_0 be such that $\delta_\square(G_n/\mathcal{P}_n, G_m/\mathcal{P}_m) \leqslant \varepsilon$ whenever $n, m \geqslant n_0$, and $|V(G_n)| \geqslant q$ whenever $n \geqslant n_0$. Then $\delta_\square(G_n, G_m) \leqslant 3\varepsilon$ for all $n, m \geqslant n_0$ by the triangle inequality, the property (i) and the fact that $\delta_\square(G_n, G_n/\mathcal{P}_n) \leqslant d_\square(G_n, (G_n)_{\mathcal{P}_n})$. This proves that (G_n) is a Cauchy sequence in the metric $\delta_\square$, and hence left-convergent.

To prove the necessity of the conditions (i) and (ii), consider a convergent sequence (G_n), a graphon U' such that $\delta_\square(G_n, U') \to 0$, and a constant $\varepsilon > 0$. With the help of the Weak Regularity Lemma for graphons, Lemma 3.3, we can find a partition $\mathcal{P}'$ of $[0, 1]$ into $q_0 \leqslant 2^{10/\varepsilon^2}$ classes such that $\|U' - U'_{\mathcal{P}'}\|_\square < \varepsilon/\sqrt{5}$. Applying a measure-preserving map to both U' and $\mathcal{P}'$, this allows us to find a graphon U and a partition $\mathcal{P}''$ of $[0, 1]$ into q_0 consecutive intervals such that $\|U - U_{\mathcal{P}''}\|_\square \leqslant \varepsilon/\sqrt{5}$ and $\delta_\square(G_n, U) \to 0$. Appealing to Lemma 5.3 we finally relabel the graphs in the sequence (G_n) in such a way that for the relabeled sequence (which we again denote by (G_n)), we get convergence in norm,

$$\|W_{G_n} - U\|_\square \to 0$$

as $n \to \infty$.

On the other hand, by Lemma 2.4, we can find a sequence of partitions $(\mathcal{P}'_n)$ such that $\mathcal{P}'_n$ is a weakly $(\varepsilon/\sqrt{5})$-regular partition of G_n with not more than $2^{10/\varepsilon^2}$ classes. Let q_n be the number of classes in $\mathcal{P}'_n$, and let $q = \max_{n\geqslant 0} q_n$. By the bound (3.11), we can refine the partitions $\mathcal{P}'_n$ to obtain weakly $(2\varepsilon/\sqrt{5} \leqslant \varepsilon)$-regular partitions $\mathcal{P}''_n$ with exactly q classes whenever $|V(G_n)| \geqslant q$. In a similar way, we can refine the partition $\mathcal{P}''$ to obtain a partition $\mathcal{P}$ of [0, 1] into q consecutive intervals such that

$$\|U - U_{\mathcal{P}}\|_\square \leqslant \frac{2}{\sqrt{5}}\varepsilon \leqslant \frac{9}{10}\varepsilon.$$

Let n_0 be such that for $n \geqslant n_0$

$$\|W_{G_n} - U\|_\square \leqslant \frac{\varepsilon}{30} \quad \text{and} \quad \frac{q}{|V(G_n)|} \leqslant \frac{\varepsilon}{30},$$

and let α_i be the Lebesgue measure of the ith partition class of $\mathcal{P}$. For $n < n_0$, we then set $\mathcal{P}_n = \mathcal{P}''_n$, and for $n \geqslant n_0$ we define $\mathcal{P}_n$ to be the partition into the classes $V_1^{(n)} = \{1, \dots, k_1^{(n)}\}$, $V_2^{(n)} = \{k_1^{(n)} + 1, \dots, k_1^{(n)} + k_2^{(n)}\}$, etc., where the integers $k_i^{(n)}$ are chosen in such a way that $\lfloor \alpha_i |V(G_n)| \rfloor \leqslant k_i^{(n)} \leqslant \lceil \alpha_i |V(G_n)| \rceil$. With this definition, we have that

$$\big\|(W_{G_n})_{\mathcal{P}} - W_{(G_n)_{\mathcal{P}_n}}\big\|_\square \leqslant \big\|(W_{G_n})_{\mathcal{P}} - W_{(G_n)_{\mathcal{P}_n}}\big\|_1 \leqslant \frac{q}{|V(G_n)|} \leqslant \frac{\varepsilon}{30}$$

for all $n \geqslant n_0$. Combined with the triangle inequality and the bound (3.9), this gives

$$\begin{aligned}
d_\square\big(G_n, (G_n)_{\mathcal{P}_n}\big) &= \|W_{G_n} - W_{(G_n)_{\mathcal{P}_n}}\|_\square \\
&\leqslant \|W_{G_n} - U\|_\square + \|U - U_{\mathcal{P}}\|_\square + \big\|U_{\mathcal{P}} - (W_{G_n})_{\mathcal{P}}\big\|_\square \\
&\quad + \big\|(W_{G_n})_{\mathcal{P}} - W_{(G_n)_{\mathcal{P}_n}}\big\|_\square \\
&\leqslant 2\|W_{G_n} - U\|_\square + \|U - U_{\mathcal{P}}\|_\square + \big\|(W_{G_n})_{\mathcal{P}} - W_{(G_n)_{\mathcal{P}_n}}\big\|_\square \\
&\leqslant 2\frac{\varepsilon}{30} + \frac{9}{10}\varepsilon + \frac{\varepsilon}{30} = \varepsilon
\end{aligned}$$

whenever $n \geqslant n_0$. Thus $\mathcal{P}_n$ is a weakly ε-regular partition of G_n, whether $n < n_0$, or $n \geqslant n_0$. By definition, it also is a partition into exactly q classes whenever $|V(G_n)| \geqslant q$. This proves (i).

To prove (ii), we note that for $n \geqslant n_0$, we have

$$\begin{aligned}
d_\square(G_n/\mathcal{P}_n, U/\mathcal{P}) &\leqslant \big\|W_{(G_n)_{\mathcal{P}_n}} - (W_{G_n})_{\mathcal{P}}\big\|_\square + \big\|(W_{G_n})_{\mathcal{P}} - U_{\mathcal{P}}\big\|_\square \\
&\leqslant \frac{q}{|V(G_n)|} + \|W_{G_n} - U\|_\square.
\end{aligned}$$

As $n \to 0$, the right-hand side goes to 0, as required.

6. Parameter testing

6.1. Definitions and statements of results

In this section we discuss the notion of continuous graph parameters, and the closely related notion of parameter testing. In parameter testing, one wants to determine some parameter of a large, simple graph G, e.g., the edge density or the density of the maximum cut. It is usually difficult to determine the exact value of such a parameter, but using sufficiently large samples, one might hope to approximate the value of parameter with large probability at a much lower computation cost; recall Definition 2.11.

Throughout this section, G will be a simple graph; this will be made explicit only in the statements of theorems and supplements. Our principal theorem gives several useful characterizations of testable graph parameters.

Theorem 6.1. *Let f be a bounded simple graph parameter. Then the following are equivalent:*

(a) *f is testable.*
(b) *For every $\varepsilon > 0$, there is an integer k such that for every simple graph G on at least k nodes,*

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

(c) *For every convergent sequence (G_n) of simple graphs with $|V(G_n)| \to \infty$, the limit of $f(G_n)$ exists as $n \to \infty$.*
(d) *There exists a functional $\check{f}(W)$ on $\mathcal{W}$ that is continuous in the rectangle norm, and $\check{f}(W_G) - f(G) \to 0$ if $|V(G)| \to \infty$.*
(e) *For every $\varepsilon > 0$ there is an $\varepsilon_0 > 0$ and an $n_0 \in \mathbb{Z}_+$ such that if G_1, G_2 are two simple graphs with $|V(G_1)|, |V(G_2)| \geqslant n_0$ and $\delta_\square(G_1, G_2) < \varepsilon_0$, then $|f(G_1) - f(G_2)| < \varepsilon$.*

If we want to use (e) to prove that a certain invariant is testable, then the complicated definition of the $\delta_\square$ distance may cause a difficulty. So it is useful to show that (e) can be replaced by a weaker condition, which consists of three special cases of (e). For this purpose, we define the disjoint union $G \cup G'$ of two graphs G and G' as the graph whose node set is the disjoint union of $V(G)$ and $V(G')$, and whose edge set is $E(G) \cup E(G')$.

Supplement 6.2. *The following three conditions together are also equivalent to testability:*

(e.1) *For every $\varepsilon > 0$ there is an $\varepsilon' > 0$ such that if G and G' are two simple graphs on the same node set and $d_\square(G, G') \leqslant \varepsilon'$, then $|f(G) - f(G')| < \varepsilon$.*
(e.2) *For every simple graph G, $f(G[m])$ has a limit as $m \to \infty$.*
(e.3) *$f(G \cup K_1) - f(G) \to 0$ if $|V(G)| \to \infty$.*

We formulate two further conditions for testability, in terms of Szemerédi partitions. Recall that a partition $\{V_1, \dots, V_k\}$ of a finite set V is an *equitable partition* if $\lfloor |V|/k \rfloor \leqslant |V_i| \leqslant \lceil |V|/k \rceil$ for every $1 \leqslant i \leqslant k$. Let $\mathcal{P} = \{V_1, \dots, V_k\}$ be an equitable partition of the node set of a simple graph G. A pair (V_i, V_j) if partition classes is called an *ε-regular pair*, if for all $X \subseteq V_i$ and $Y \subseteq V_j$ with $|X|, |Y| \geqslant \varepsilon |V(G)|/k$, we have

$$\left| \frac{e_G(X,Y)}{|X| \cdot |Y|} - \frac{e_G(V_i,V_j)}{|V_i| \cdot |V_j|} \right| \leqslant \varepsilon.$$

The partition $\mathcal{P}$ is *ε-regular* if all but at most εk^2 pairs (V_i, V_j) are ε-regular.

Every ε-regular partition is weakly (7ε)-regular, but in the reverse direction only a much weaker implication holds: a weakly ε-regular partition with k classes is $\tilde{\varepsilon}$-regular with $\tilde{\varepsilon} = \sqrt[3]{k^2 \varepsilon}$.

The "original" Szemerédi Lemma can be stated as follows.

Lemma 6.3 *(Szemerédi Regularity Lemma [30]). For every $\varepsilon > 0$ and $l > 0$ there is a $k(\varepsilon, l) > 0$ such that every simple graph $G = (V, E)$ with at least l nodes has an ε-regular partition into at least l and at most $k(\varepsilon, l)$ classes.*

Let G be a graph and let $\mathcal{P}$ be an equitable partition of $V(G)$. Then $G/\mathcal{P}$ is a weighted graph with almost equal nodeweights. We modify this graph by making all nodeweights equal to 1. This way we get a weighted graph $G \div \mathcal{P}$ with nodeweights 1 and edgeweights in [0, 1].

For every bounded, simple graph parameter f and every weighted graph H with nodeweights 1 and edgeweights in [0, 1], define

$$\widehat{f}(H) = \mathsf{E}\big(f\big(\mathbf{G}(H)\big)\big),$$

where $\mathbf{G}(H)$ is the graph obtained by the randomizing procedure described in Section 4.3. Clearly $\widehat{f}$ is an extension of f.

Supplement 6.4. *Either one of the following conditions is also equivalent to testability:*

(f) *For every $\varepsilon > 0$ there is an $\varepsilon' > 0$ and an $n_0 \in \mathbb{Z}_+$ such that if G is any graph with $|V(G)| \geqslant n_0$ and $\mathcal{P}$ is an equitable weakly ε'-regular partition of G with $n_0 \leqslant |\mathcal{P}| \leqslant \varepsilon' |V(G)|$, then $|f(G) - \widehat{f}(G \div \mathcal{P})| \leqslant \varepsilon$.*

(g) *The parameter f has an extension $\check{f}$ to (finite) weighted graphs with nodeweights* 1 *and edgeweights in* [0, 1] *such that*
   (g.1) *for every fixed n, $\check{f}$ is a continuous function of the edgeweights on n-node graphs, and*
   (g.2) *for every $\varepsilon > 0$ there is an $\varepsilon' > 0$ and an $n_0 \in \mathbb{Z}_+$ such that if G is any graph with $|V(G)| \geqslant n_0$ and $\mathcal{P}$ is an ε'-regular partition of G with $n_0 \leqslant |\mathcal{P}| \leqslant \varepsilon' |V(G)|$, then $|f(G) - \check{f}(G \div \mathcal{P})| \leqslant \varepsilon$.*

Condition (g) is *a priori* weaker than (f) on two counts: First, (g) allows an arbitrary extension of f to weighted graphs, while (f) makes assumptions about a specific extension $\widehat{f}$. Second, (g) states a condition about ε-regular partitions, for which the condition may be easier to verify than for weakly ε-regular partitions. Of course, we could formulate two "intermediate" conditions, in which only one of these relaxations is used.

Condition (g.1) concerns graphs on a fixed finite set, so when we say that $\check{f}$ should be "continuous," we do not have to specify in which metric we mean this. But to be concrete, we will use the $d_\square$ distance as a metric on these graphs.

6.2. Proofs

Before proving the equivalence of the above conditions, we state and prove a simple lemma.

Lemma 6.5. *Let G, G' be weighted graphs with edgeweights in some interval I of length $|I|$ and total nodeweight one. If G and G' have equal edgeweights but different nodeweights, then*

$$\delta_\square(G, G') \leqslant |I| \sum_i \bigl|\alpha_i(G) - \alpha_i(G')\bigr|.$$

Proof. Let α_i be the nodeweights of G, and α_i' be those of G'. Consider a coupling X such that $X_{ii} = \min\{\alpha_i, \alpha_i'\}$. Then

$$\delta_\square\bigl((G, G') \leqslant d_\square\bigl(G[X], G'\bigl[X^\top\bigr]\bigr) \leqslant \sum_{i,j,k,l} X_{ik}X_{jl}\bigl|\beta_{ij}(G) - \beta_{kl}(G')\bigr| \leqslant |I| \sum_{\substack{i\neq j \text{ or} \\ k\neq l}} X_{ik}X_{jl}$$

$$= |I|\biggl(1 - \biggl(\sum_i \min\bigl\{\alpha_i, \alpha_i'\bigr\}\biggr)^2\biggr) = |I|\biggl(1 - \biggl(1 - \frac{1}{2}\sum_i\bigl|\alpha_i - \alpha_i'\bigr|\biggr)^2\biggr)$$

$$\leqslant |I| \sum_i \bigl|\alpha_i - \alpha_i'\bigr|. \qquad \square$$

After these preparations, we are ready to prove the results of the last section.

Proof of Theorem 6.1. We first prove that (a), (b), (c) and (e) are equivalent:

(a) ⇒ (b). The definition of testability is very similar to condition (b): it says, in this language, that a random set S_k on k nodes of G as in (b) satisfies

$$\bigl|f(G) - f\bigl(\mathbb{G}(k, G)\bigr)\bigr| \leqslant \varepsilon$$

with large probability. This clearly implies that this difference is small on average.

(b) ⇒ (c). Let (G_n) be a convergent sequence with $|V(G_n)| \to \infty$. Given $\varepsilon > 0$, let k be such that for every graph G on at least k nodes, $|f(G) - \mathsf{E}(f(\mathbb{G}(k, G)))| \leqslant \varepsilon$. Since G_n is convergent, $t(F, G_n)$ tends to a limit for all graphs F on k nodes, from which we get that $t_{\mathrm{ind}}(F, G_n)$ tends to a limit $t_{\mathrm{ind}}(F)$ for all graphs on k nodes. This means that $\Pr(\mathbb{G}(k, G_n) = F) \to t_{\mathrm{ind}}(F)$, and so

$$\mathsf{E}\bigl(f\bigl(\mathbb{G}(k, G_n)\bigr)\bigr) \to \sum_F t_{\mathrm{ind}}(F) f(F) = a_k.$$

As a consequence, we have that for all sufficiently large n,

$$\bigl|f(G_n) - a_k\bigr| \leqslant \bigl|f(G) - \mathsf{E}\bigl(f\bigl(\mathbb{G}(k, G_n)\bigr)\bigr)\bigr| + \varepsilon \leqslant 2\varepsilon,$$

so $f(G_n)$ oscillates less than 4ε if n is large enough. This proves that the sequence $(f(G_n))$ is convergent.

(c) $\Rightarrow$ (e). Suppose that (e) does not hold for some $\varepsilon > 0$; then there are two sequences (G_n) and (G'_n) of graphs such that $|V(G_n)|, |V(G'_n)| \to \infty$, $\delta_\square(G_n, G'_n) \to 0$, but $|f(G_n) - f(G'_n)| > \varepsilon$. We may assume that both graph sequences are convergent; but then $\delta_\square(G_n, G'_n) \to 0$ implies that the merged sequence $(G_1, G'_1, G_2, G'_2, \ldots)$ is also convergent, so by (c), the sequence of numbers $(f(G_1), f(G'_1), f(G_2), f(G'_2), \ldots)$ is also convergent, a contradiction.

(e) $\Rightarrow$ (a). Suppose (a) does not hold. Then there is an $\varepsilon > 0$ and a sequence (G_n) of graphs with $|V(G_n)| \geqslant n$ such that with probability at least ε we have that $|f(G_n) - f(\mathbb{G}(n, G_n))| > \varepsilon$ for all n. Now choose $\varepsilon_0 > 0$ and an $n_0 \in \mathbb{Z}_+$ in such a way that $|f(G_1) - f(G_2)| < \varepsilon$ whenever G_1, G_2 are two graphs with $|V(G_1)|, |V(G_2)| \geqslant n_0$ and $\delta_\square(G_1, G_2) < \varepsilon_0$ (this is possible by (e)). But by Theorem 4.7(iii), we have $\delta_\square(G_n, \mathbb{G}(k, G_n)) \to 0$ in probability, implying in particular that for n large enough, $\delta_\square(G_n, \mathbb{G}(k, G_n)) < \varepsilon_0$ with probability at least $1 - \varepsilon/2$. By our choice of ε_0, this implies that for n large enough, $|f(G_n) - f(\mathbb{G}(k, G_n))| < \varepsilon$ with probability at least $1 - \varepsilon/2$, a contradiction.

We continue with proving that (d) is equivalent to (a), (b), (c) and (e).

(e) $\Rightarrow$ (d). For $W \in \mathcal{W}_{[0,1]}$, define $\check{f}(W) = \lim_{n\to\infty} f(G_n)$, where (G_n) is any sequence of graphs such that $G_n \to W$ and $|V(G_n)| \to \infty$ (by (e), $\check{f}(W)$ does not depend on the choice of the sequence G_n as long as $G_n \to W$). We prove that this functional has the desired properties.

Let $\varepsilon > 0$, and let ε' and n_0 be as given by (e) with ε replaced by $\varepsilon/3$. To prove continuity, we will prove that $|\check{f}(W) - \check{f}(W')| \leqslant \varepsilon$ whenever $\|W - W'\|_\square \leqslant \varepsilon'/3$. Considering a graph sequence tending to W, we can choose a simple graph G such that $|V(G)| \geqslant n_0$, $\delta_\square(G, W) < \varepsilon'/3$ and $|f(G) - \check{f}(W)| \leqslant \varepsilon/3$; similarly, we can choose a simple graph G' such that $|V(G')| \geqslant n_0$, $\delta_\square(G', W') < \varepsilon'/3$ and $|f(G') - \check{f}(W')| < \varepsilon/3$. Then

$$\delta_\square(G, G') \leqslant \delta_\square(G, W) + \delta_\square(W, W') + \delta_\square(W', G') \leqslant \varepsilon',$$

and hence by (e), $|f(G) - f(G')| \leqslant \varepsilon/3$. But then

$$\left|\check{f}(W) - \check{f}(W')\right| \leqslant \left|\check{f}(W) - f(G)\right| + \left|f(G) - f(G')\right| + \left|f(G') - \check{f}(W)\right| \leqslant \varepsilon$$

as claimed. Note that our proof shows that, in fact, $\check{f}$ is continuous in the $\delta_\square$ metric.

The fact that $\check{f}(W_{G_n}) - f(G_n) \to 0$ whenever $|V(G_n)| \to \infty$ can now be easily verified by contradiction. Indeed, assume that this is not the case. Using compactness, we may choose a subsequence such that $G_n \to W$ for some W. But this implies that $f(G_n) \to \check{f}(W)$, and by the continuity of $\check{f}$, also that $\check{f}(W_{G_n}) \to \check{f}(W)$, a contradiction.

(d) $\Rightarrow$ (c). Consider a convergent graph sequence (G_n) with $|V(G_n)| \to \infty$, and let $W \in \mathcal{W}$ be its limit. Then $\delta_\square(W_{G_n}, W) \to 0$, and by Lemma 5.3, $\|W_{G'_n} - W\|_\square \to 0$ for a relabeling of G_n. By the continuity of $\check{f}$, we have $\check{f}(W_{G'_n}) - \check{f}(W) \to 0$. By assumption, $f(G_n) - \check{f}(W_{G'_n}) = f(G_n) - \check{f}(W_{G_n}) \to 0$, and so $f(G_n) - \check{f}(W) \to 0$. This proves that $(f(G_n))$ is convergent. $\square$

Proof of Supplement 6.2. (e) $\Rightarrow$ (e.1), (e.2), (e.3). To see that (e.1) is a special case of (e), it suffices to note that if G_1 and G_2 are two different graphs on the same set of n nodes, and $d_\square(G, G') \leqslant \varepsilon'$, then $n \geqslant 1/\sqrt{\varepsilon'}$. For (e.2), note that $\delta_\square(G[m], G[m']) = 0$ for all m, m'. For (e.3), it suffices to verify that $\delta_\square(G, GK_1) \leqslant 1/|V(G)|$.

(e.1), (e.2), (e.3) $\Rightarrow$ (c). Suppose that (c) does not hold. Then there exist two graph sequences (G_n) and (G'_n) with $|V(G_n)| \to \infty$ and $|V(G'_n)| \to \infty$ such that $G_n, G'_n \to W$, $f(G_n) \to a$ and $f(G'_n) \to b$ as $n \to \infty$, but $a \neq b$.

By (e.1), there exists an $\varepsilon > 0$ such that if G and G' are two graphs on the same node set, and $d_\square(G, G') \leqslant \varepsilon$, then $|f(G) - f(G')| \leqslant |a-b|/4$. Let $\varepsilon_1 = (\varepsilon/32)^{67}$. By Lemma 2.5, for every n there is a simple graph H_n whose number of nodes k depends only on ε such that $\delta_\square(G_n, H_n) \leqslant \varepsilon_1/2$. By selecting an appropriate subsequence, we may assume that $H_n = H$ is the same graph for all n. Since (G_n) and (G'_n) have the same limit, it follows that $\delta_\square(G'_n, H) \leqslant \varepsilon_1$ for all n that are large enough.

Let us add to each G_n at most $k-1$ isolated nodes so that the resulting graph G_n^* has km_n nodes for some integer m_n. The m_n-fold blow-up $H[m_n]$ of H then satisfies $\delta_\square(G_n^*, H[m_n]) \leqslant \varepsilon_1$, and so by Theorem 2.3, for a suitable overlay of G_n and $H[m_n]$, we have $d_\square(G_n^*, H[m_n]) \leqslant \varepsilon$, and so, by the definition of ε,

$$\left|f\left(G_n^*\right) - f\left(H[m_n]\right)\right| \leqslant \frac{|a-b|}{4}.$$

Using (e.3), we see that $f(G_n^*) - f(G_n) \to 0$, and hence

$$\left|f(G_n) - f\left(H[m_n]\right)\right| \leqslant \frac{|a-b|}{3}$$

if n is large enough. Similarly, H has a m'_n-node blow-up $H[m'_n]$ such that

$$\left|f\left(G'_n\right) - f\left(H\left[m'_n\right]\right)\right| \leqslant \frac{|a-b|}{3}.$$

But since $H[m_n]$ and $H[m'_n]$ are blow-ups of the same graph H, (e.2) implies that $f(H[m_n]) - f(H[m'_n]) \to 0$, a contradiction. $\square$

Proof of Supplement 6.4. We will assume (without loss of generality) that $|f| \leqslant 1$.

(e) $\Rightarrow$ (f). Choose ε_0 and n_0 so that (i) for $n \geqslant n_0$ the bound in Lemma 4.3 is at most ε_0 with probability at least $1 - \varepsilon/4$ and (ii) for two graphs G_1 and G_2 with $|V(G_1)|, |V(G_2)| \geqslant n_0$ and $\delta_\square(G_1, G_2) \leqslant 3\varepsilon_0$, we have $|f(G_1) - f(G_2)| < \varepsilon/4$. Suppose that $|V(G)| \geqslant n_0$ and $n_0 \leqslant |\mathcal{P}| \leqslant \varepsilon_0 |V(G)|$. By definition of ε_0-regular partitions and by Lemma 6.5 we have $\delta_\square(G, G \div \mathcal{P}) \leqslant 2\varepsilon_0$. Furthermore, by Lemma 4.3, we have $\delta_\square(G \div \mathcal{P}, \mathbf{H}(G \div \mathcal{P})) \leqslant \varepsilon_0$ with probability at least $1 - \varepsilon/4$. If this occurs, then $\delta_\square(G, \mathbf{H}(G \div \mathcal{P})) \leqslant 3\varepsilon_0$, and by the choice of ε_0 and n_0, it follows that $|f(G) - f(\mathbf{H}(G \div \mathcal{P}))| \leqslant \varepsilon/2$. Hence

$$\left|f(G) - \widehat{f}(G)\right| = \left|\mathsf{E}\left(f(G) - f\left(\mathbf{H}(G \div \mathcal{P})\right)\right)\right| \leqslant \frac{\varepsilon}{2}(1 - \varepsilon/4) + 2(\varepsilon/4) < \varepsilon.$$

(f) $\Rightarrow$ (g). All we have to verify is that for weighted graphs on a fixed node set (say $[n]$), $\widehat{f}$ is a continuous function of the edgeweights. Consider two weighted graphs H_1, H_2 with $V(H_1) = V(H_2) = [n]$ and $d_1(H_1, H_2) \leqslant \varepsilon/n^2$. By Lemma 4.2, the randomized simple graphs $\mathbf{G}(H_1)$ and $\mathbf{G}(H_2)$ can be coupled so that $\mathsf{E}[d_1(\mathbf{G}(H_1), \mathbf{G}(H_2))] \leqslant \varepsilon/n^2$, which by Markov's inequality implies that $\mathbf{G}(H_1) = \mathbf{G}(H_2)$ with probability at least $1 - \varepsilon$. This implies that $|\widehat{f}(\mathbf{G}(H_1)) - \widehat{f}(\mathbf{G}(H_2))| \leqslant \varepsilon$. (Note that this argument did not require that f is testable.)

(g) $\Rightarrow$ (e.1), (e.2), (e.3). Let $\varepsilon > 0$, and consider the extension $\check{f}$ postulated in (g). By property (g.2), we can choose an $\varepsilon_1 > 0$ and an $n_1 \in \mathbb{Z}_+$ such that if $|V(G)| \geqslant n_1$ and $\mathcal{P}$ is an ε_1-regular partition of G with $n_1 \leqslant |\mathcal{P}| \leqslant \varepsilon_1 |V(G)|$, then $|f(G) - \check{f}(G \div \mathcal{P})| \leqslant \varepsilon/3$. Using the Regularity Lemma 6.3, fix $k = k(\varepsilon_1/2, n_1) > n_1$ so that every graph G with at least n_1 nodes has an $(\varepsilon_1/2)$-regular partition with at least n_1 and at most k classes. Let $n_0 = \max\{n_1, \lceil k/\varepsilon_1\rceil\}$. Using condition (g.1), choose an $\varepsilon_2 > 0$ such that for any two graphs H_1 and H_2 on the same set of $m \leqslant n_0$ nodes with $d_\square(H_1, H_2) \leqslant \varepsilon_2$ we have $|\check{f}(H_1) - \check{f}(H_2)| \leqslant \varepsilon/3$. Finally, choose $\varepsilon' = \min\{\varepsilon_2/4, \varepsilon_1^3/(8k^2)\}$.

To prove (e.1), let G and G' be two simple graphs on the same node set with $d_\square(G, G') \leqslant \varepsilon'$. If $n = |V(G)| \leqslant n_0$, then $|f(G_1) - f(G_2)| = |\check{f}(G_1) - \check{f}(G_2)| \leqslant \varepsilon/3$, so we can assume $|V(G)| > n_0$. Let $\mathcal{P} = (V_1, \dots, V_l)$ be an $(\varepsilon_1/2)$-regular partition of G and $n_1 \leqslant l \leqslant k$. From the assumption that $d_\square(G, G') \leqslant \varepsilon' \leqslant \varepsilon_1^3/(8k^2)$, it follows that $\mathcal{P}$ is an ε_1-regular partition of G'. By the choice of ε_1 and n_1 it follows that

$$\bigl|f(G) - \check{f}(G \div \mathcal{P})\bigr| \leqslant \frac{\varepsilon}{3} \quad \text{and} \quad \bigl|f(G') - \check{f}(G' \div \mathcal{P})\bigr| \leqslant \frac{\varepsilon}{3}.$$

Furthermore, we have

$$n^2 d_\square(G, G') \geqslant l^2 d_\square(G \div \mathcal{P}, G' \div \mathcal{P}) \left\lfloor \frac{n}{l} \right\rfloor^2,$$

and so

$$d_\square(G \div \mathcal{P}, G' \div \mathcal{P}) \leqslant \frac{n^2}{l^2} \left\lfloor \frac{n}{l} \right\rfloor^{-2} d_\square(G, G') \leqslant 4\varepsilon' \leqslant \varepsilon_2.$$

Hence by the choice of ε_2, we have

$$\bigl|\check{f}(G \div \mathcal{P}) - \check{f}(G' \div \mathcal{P})\bigr| \leqslant \frac{\varepsilon}{3}.$$

Summing up,

$$\begin{aligned}\bigl|f(G) - f(G')\bigr| &\leqslant \bigl|f(G) - \check{f}(G \div \mathcal{P})\bigr| + \bigl|\check{f}(G \div \mathcal{P}) - \check{f}(G' \div \mathcal{P})\bigr| + \bigl|\check{f}(G' \div \mathcal{P}) - f(G')\bigr| \\ &\leqslant \varepsilon.\end{aligned}$$

The proof of (e.2) is similar, but divisibility concerns cause some complications. We have to show that $f(G[q])$ is a Cauchy sequence, i.e., we have to show that given G and $\varepsilon > 0$, we can find a q_0 such that $|f(G[q]) - f(G[q'])| \leqslant 2\varepsilon$ whenever $q, q' \geqslant q_0$. Let $n = |V(G)|$, $p > 1$ large enough, and let $\mathcal{P} = \{V_1, \dots, V_l\}$ be an $(\varepsilon_1/2)$-regular partition of $G[p]$ with $n_1 \leqslant l \leqslant k$. By our choice of p we have, in particular, that $l \leqslant \frac{\varepsilon_1}{2} np$, so we may apply (g.2) to get

$$\bigl|f\bigl(G[p]\bigr) - \check{f}\bigl(G[p] \div \mathcal{P}\bigr)\bigr| \leqslant \frac{\varepsilon}{3}. \tag{6.1}$$

Let r be sufficiently large, and let $q = pr + s$ with $0 \leqslant s < p$. The graph $G[pr]$ arises from $G[p]$ by blowing up each node into r nodes, and then $G[q]$ arises from $G[pr]$ adding sn further nodes.

First we consider the graph $G[pr]$. The partition $\mathcal{P}$ determines a partition $\mathcal{Q} = \{U_1, \dots, U_l\}$ of $V(G[pr])$. Let

$$p_{ij} = \frac{e_{G[p]}(V_i, V_j)}{|V_i| \cdot |V_j|} = \frac{e_{G[pr]}(U_i, U_j)}{|U_i| \cdot |U_j|}$$

denote the edge density between V_i and V_j in $G[p]$ (which is the same as the edge density between U_i and U_j in $G[pr]$).

We claim that if (V_i, V_j) is an $(\varepsilon_1/2)$-regular pair in $\mathcal{P}$, then for all $X \subseteq U_i$ and $Y \subseteq U_j$ with $|X|, |Y| \geqslant 2\varepsilon_1 npr/(3l)$, we have

$$p_{ij} - \frac{\varepsilon_1}{2} \leqslant \frac{e_{G[pr]}(X, Y)}{|X| \cdot |Y|} \leqslant p_{ij} + \frac{\varepsilon_1}{2}. \tag{6.2}$$

For $u \in V_i$, let x'_u denote the number of elements in X among the r copies of u in $G[pr]$, and let $x_u = x'_u/r$. Clearly $0 \leqslant x_u \leqslant 1$ and $\sum_u x_u \geqslant \lfloor 2\varepsilon_1 npr/(3l) \rfloor$. We define y_v for $v \in V_j$ analogously. Then we have

$$e_{G[pr]}(X, Y) - |X| \cdot |Y|(p_{ij} - \varepsilon_1) = \sum_{u \in V_i} \sum_{v \in V_j} r^2 x_u y_v \left(a_{uv} - p_{ij} - \frac{\varepsilon_1}{2} \right). \tag{6.3}$$

Since the right-hand side is linear in each x_u, it attains its minimum over $0 \leqslant x_u \leqslant 1$, $\sum_u x_u \geqslant \lfloor 2\varepsilon_1 npr/(3l) \rfloor$, at a vertex of this domain, which is a 0–1 vector. Similarly, the minimizing choice of the y_v is a 0–1 vector. Let $S = \{u : x_u = 1\}$ and $T = \{v : y_v = 1\}$, then $|S|, |T| \geqslant \lfloor 2\varepsilon_1 npr/(3l) \rfloor \geqslant \frac{\varepsilon_1}{2} |V(G[p]|/l$. The right-hand side of (6.3) is equal to

$$r^2 \left(e_{G[p]}(S, T) - |S| \cdot |T| \left(p_{ij} - \frac{\varepsilon_1}{2} \right) \right) \geqslant 0,$$

since (V_i, V_j) is $(\varepsilon_1/2)$-regular. This proves the first inequality in (6.2); the proof of the second is similar.

Now $\mathcal{Q}$ is not necessarily an equitable partition; the largest and smallest class sizes may differ by r, not by 1. Let us remove r nodes from those classes of $\mathcal{Q}$ that are too big. To get a partition of $V(G[q])$, we have to add back these nodes ($t \leqslant l$) and sn further nodes. Let us distribute these nodes as equally as possible between the classes, to get an equitable partition $\mathcal{Q}' = \{U'_1, \ldots, U'_l\}$ of $G[q]$.

We claim that $\mathcal{Q}'$ is ε_1-regular. Consider a pair (V_i, V_j) that is $(\varepsilon_1/2)$-regular in $G[p]$, and subsets $X' \subseteq U'_i$ and $Y' \subseteq U'_j$ with $|X|, |Y| \geqslant \varepsilon_1 nq/l$. Remove the new nodes from X' and Y' to get $X \subseteq U_i$ and $Y \subseteq U_j$. Clearly $|X|, |Y| \geqslant \varepsilon_1 nq/l - r - \lceil sn/l \rceil \geqslant 2\varepsilon_1 npr/(3l)$, and so (6.2) is satisfied. Now it is easy to check that

$$\left| \frac{e_{G[q]}(U'_i, U'_j)}{|U'_i| \cdot |U'_j|} - \frac{e_{G[pr]}(U_i, U_j)}{|U_i| \cdot |U_j|} \right| \leqslant \frac{\varepsilon_1}{4} \quad \text{and} \quad \left| \frac{e_{G[q]}(X', Y')}{|X'| \cdot |Y'|} - \frac{e_{G[pr]}(X, Y)}{|X| \cdot |Y|} \right| \leqslant \frac{\varepsilon_1}{4}$$

if $p > 32l/\varepsilon_1^2$ and $q > 32p/\varepsilon_1^2$, which proves that

$$\left| \frac{e_{G[q]}(U'_i, U'_j)}{|U'_i| \cdot |U'_j|} - e_{G[q]}(X', Y')|X'| \cdot |Y'| \right| \leqslant \varepsilon_1.$$

Thus $\mathcal{Q}'$ is an ε_1-regular partition of $G[q]$. It follows by (g.2) and the choice of ε_1 that

$$\left| f\big(G[q]\big) - \check{f}\big(G[q] \div \mathcal{Q}'\big) \right| \leqslant \frac{\varepsilon}{3}. \tag{6.4}$$

The weighted graphs $G[p] \div \mathcal{P}$ and $G[q] \div \mathcal{Q}'$ are very close to each other. In fact, $G[p] \div \mathcal{P} \cong G[pr] \div \mathcal{Q}$, while it is easy to check that $d_\square(G[pr] \div \mathcal{P}, G[q] \div \mathcal{Q}) < \varepsilon_2$ if $q > 4p/\varepsilon_2$, and so (g.1) implies that

$$\left|\check{f}\big(G[p] \div \mathcal{P}\big) - \check{f}\big(G[q] \div \mathcal{Q}'\big)\right| \leqslant \frac{\varepsilon}{3}. \tag{6.5}$$

Now (6.1), (6.4) and (6.5) imply that $|f(G[p]) - f(G[q])| \leqslant \varepsilon$. So if $q, q' > \max(32p/\varepsilon_1^2, 4p/\varepsilon_2\}$, then

$$\left|f\big(G[q]\big) - f\big(G[q']\big)\right| \leqslant \left|f\big(G[q]\big) - f\big(G[p]\big)\right| + \left|f\big(G[p]\big) - f\big(G[q']\big)\right| \leqslant 2\varepsilon.$$

Thus $f(G[q])$ is a Cauchy sequence, which proves (e.2).

The proof of (e.3) is similar but easier, and is left to the reader. $\square$

7. Concluding remarks

7.1. Norms related to the cut-norm

The cut-norm (and the cut-distance of graphs) is closely related to several other norms that are often used. We formulate these connections for the case of graphons, but similar remarks would apply to the $d_\square$ distance of graphs.

We start with the remark that we could restrict the sets S, T in the definition of the cut-norm in (3.3); it turns out that "reasonable" restrictions only change the supremum value by a constant factor. In particular, it is not hard to see that

$$\frac{1}{2}\|W\|_\square \leqslant \sup_{S=T}\left|\int_{S\times T} W\right| \leqslant \|W\|_\square, \tag{7.1}$$

$$\frac{1}{4}\|W\|_\square \leqslant \sup_{S\cap T=\emptyset}\left|\int_{S\times T} W\right| \leqslant \|W\|_\square, \tag{7.2}$$

and

$$\frac{2}{3}\sup_{S\cap T=\emptyset}\left|\int_{S\times T} W\right| \leqslant \sup_{S=[0,1]\setminus T}\left|\int_{S\times T} W\right| \leqslant \sup_{S\cap T=\emptyset}\left|\int_{S\times T} W\right|, \tag{7.3}$$

see Appendix A for details.

To relate the cut norm to homomorphisms, we start with noticing that

$$t(C_4, U)^{1/4} = \left(\operatorname{Tr} T_U^4\right)^{1/4}, \tag{7.4}$$

and hence the functional $t(C_4, .)^{1/4}$ defines a norm, called the *trace norm* or *Schatten norm*, on $\mathcal{W}$. (More generally, any even cycle leads to a norm in this way.) The following lemma shows that the norm in (7.4) is intimately related to the cut-norm.

Lemma 7.1. *For $U \in \mathcal{W}$ with $\|U\|_\infty \leqslant 1$, we have*

$$\frac{1}{4} t(C_4, U) \leqslant \|U\|_\square \leqslant t(C_4, U)^{1/4}.$$

Proof. The first inequality is a special case of (4.27). To prove the second inequality, we use (3.5) and (7.4):

$$\|U\|_\square \leqslant \|U\|_{\infty\to 1} = \sup_{|f|,|g|\leqslant 1} \langle f, T_U g\rangle,$$

and here

$$\begin{aligned}\langle f, T_U g\rangle &\leqslant \|f\|_2 \cdot \|T_U g\|_2 = \|f\|_2 \cdot \langle T_U g, T_U g\rangle^{1/2} = \|f\|_2 \cdot \langle g, T_U^2 g\rangle^{1/2} \\ &\leqslant \|f\|_2 \cdot \|g\|_2 \cdot \|T_U^2\|_{2\to 2}^{1/2} \leqslant \|T_U^2\|_2^{1/2} = \left(\operatorname{Tr} T_U^4\right)^{1/4} = t(C_4, U)^{1/4}. \qquad \square\end{aligned}$$

Lemma 7.1 allows for a significant simplification of the proof that the sample from a weighted graphs is close to the original graph. Indeed, it is possible to use the following easy lemma instead of the quite difficult Theorem 4.6 (or the equally difficult results of [2]) to establish a slightly weaker version of Theorem 4.7, which is still strong enough to prove the equivalence of left-convergence and convergence in metric (see [16] for details).

Lemma 7.2. *Let $0 < \varepsilon < 1$, $0 < \delta < 1$, and let $U \in \mathcal{W}$, $\|U\|_\infty \leqslant 1$. If*

$$\|U\|_\square \leqslant \frac{1}{8}\varepsilon^{1/4} \quad \textit{and} \quad k \geqslant \frac{352}{\varepsilon^8} \ln\left(\frac{2}{\delta}\right),$$

then

$$\mathsf{P}\left(\left\|\mathbf{H}(k, U)\right\|_\square \leqslant \varepsilon\right) \geqslant 1 - \delta.$$

Proof. Applying first Lemma 7.1 and then Lemma 4.4, we have

$$\left\|\mathbf{H}(k, U)\right\|_\square \leqslant t\left(C_4, \mathbf{H}(k, U)\right)^{1/4} \leqslant \left(t(C_4, U) + \frac{\varepsilon^4}{2}\right)^{1/4}$$

with probability at least $1 - \delta$. Thus, using Lemma 7.1 again, we get

$$\left\|\mathbf{H}(k, U)\right\|_\square \leqslant \left(t(C_4, U) + \frac{\varepsilon^4}{2}\right)^{1/4} \leqslant \left(4\|U\|_\square + \frac{\varepsilon^4}{2}\right)^{1/4} \leqslant \varepsilon,$$

as claimed. $\square$

A substantial advantage of the norm $t(C_4, W)$ over the cut-norm is that for weighted graphs (and for many other types of graphons W, for example, for polynomials), it is polynomial-time computable. A better polynomial-time computable approximation is the *Grothendieck norm*, which approximates the cut-norm within a constant factor (see [5]).

7.2. A common generalization of Lemmas 2.4 and 3.3

As stated, Lemma 3.3 does not generalize Lemma 2.4 since the partition in Lemma 3.3 is not necessarily aligned with the steps of the stepfunction W_G. But the following lemma implies both Lemmas 2.4 and 3.3.

Lemma 7.3. *Let $\mathcal{A}$ be an algebra of measurable subsets of* $[0,1]$*, and let*

$$\|W\|_{\mathcal{A},\square} = \sup_{S,T\in\mathcal{A}} \left| \int_{S\times T} W \right|.$$

Then for every graphon W and every $\varepsilon > 0$, there exists a partition $\mathcal{P}$ of $[0,1]$ *into sets in $\mathcal{A}$ with at most $4^{\lceil 1/\varepsilon^2\rceil - 1}$ classes such that*

$$\|W - W_{\mathcal{P}}\|_{\mathcal{A},\square} \leqslant \varepsilon \|W\|_2.$$

Proof. Let $\mathcal{P}$ be a partition of $[0,1]$ into q classes in $\mathcal{A}$, let $S, T \in \mathcal{A}$, and let $\mathcal{P}'$ be the partition generated by S, T and $\mathcal{P}$. Clearly $\mathcal{P}'$ has at most $4q$ classes, all of which lie in $\mathcal{A}$. Since $W_{\mathcal{P}'}$ gives the best L_2-approximation of W among all step functions with steps $\mathcal{P}'$, we conclude that for every real number t, we have

$$\|W - W_{\mathcal{P}'}\|_2^2 \leqslant \|W - W_{\mathcal{P}} - t\mathbf{1}_{S\times T}\|_2^2.$$

Bounding the right-hand side by $\|W - W_{\mathcal{P}}\|_2^2 - 2t\langle \mathbf{1}_{S\times T}, W - W_{\mathcal{P}}\rangle + t^2$ and choosing $t = \langle \mathbf{1}_{S\times T}, W - W_{\mathcal{P}}\rangle$, this gives

$$\langle \mathbf{1}_{S\times T}, W - W_{\mathcal{P}}\rangle^2 \leqslant \|W - W_{\mathcal{P}}\|_2^2 - \|W - W_{\mathcal{P}'}\|_2^2 = \|W_{\mathcal{P}'}\|_2^2 - \|W_{\mathcal{P}}\|_2^2.$$

Taking the supremum over all sets $S, T \in \mathcal{A}$, this gives

$$\|W - W_{\mathcal{P}}\|_{\mathcal{A},\square}^2 \leqslant \sup_{\mathcal{P}'} \|W_{\mathcal{P}'}\|_2^2 - \|W_{\mathcal{P}}\|_2^2, \tag{7.5}$$

where the supremum goes over all partitions of $[0,1]$ into $4q$ classes in $\mathcal{A}$. From this bound, the lemma then follows by standard arguments. □

7.3. Right convergence

When studying homomorphisms from G into a small graph H it will be convenient to consider graphs H with nodeweights, $\alpha_i(H) > 0$, and edgeweights $\beta_{ij}(H) \in \mathbb{R}$ (with i running over all nodes in H, and ij running over all edges of F). For such a graph, we define

$$\hom(G,H) = \sum_{\phi} \prod_{i\in V(G)} \alpha_{\phi(i)}(H) \prod_{ij\in E(G)} \beta_{\phi(i),\phi(j)}(H)$$

where the sum runs over all maps ϕ from $V(G)$ to $V(H)$ and $\beta_{\phi\phi'}(H)$ is set to zero if $\phi\phi'$ is not an edge in H. We call $\hom(G,H)$ the weighted number of H-colorings of G.

For dense graphs G, the weighted H-coloring numbers $\hom(G, H)$ turn out to be most interesting when all edgeweights of H are strictly positive (we say that H is a soft-core graph if this is the case). Under this assumption, the homomorphism numbers $\hom(G, H)$ typically grow exponentially in the number of edges in G. For homomorphism into small graphs H, it is therefore natural to consider the logarithm of $\hom(G, H)$ divided by the number of nodes in G to the power two. We will also consider the "microcanonical ensemble," where the number of nodes in $V(G)$ colored by a given color $c \in V(H)$ is fixed to be some constant proportion a_c of all nodes. We denote these homomorphism densities by $\hom_{\mathbf{a}}(G, H)$, where $\mathbf{a}$ is the vector with components a_c, $c \in V(H)$.

The above perspective leads to two *a priori* different notions of convergence: A sequence of weighted graphs (G_n) will be called *convergent from the left* if the homomorphism densities $t(F, G_n)$ converge for all finite graphs F, and it will be called *convergent from the right* if the quantity $|V(G_n)|^{-2} \log \hom_{\mathbf{a}}(G_n, H)$ converges for all $\mathbf{a}$ and all soft-core graphs H. It will turn out that these two notions are closely related. Convergence from the left implies convergence from the right (both for the standard homomorphism numbers and the microcanonical ones), and convergence from the right for the microcanonical homomorphism numbers implies convergence from the left. This will be discussed in more detail in the continuation of this paper [17].

Appendix A

In this appendix, we collect various proofs which were omitted in the body of the paper.

A.1. Proof of Corollary 3.4

Proof of (i). Let $\mathcal{P}'$ be a partition of $[0,1]$ into $q' \leqslant 2^{81/(8\varepsilon^2)}$ classes such that $\|W - W_{\mathcal{P}'}\|_\square \leqslant \frac{4\varepsilon}{9}\|W\|_2$. Then there exists an equipartition $\mathcal{P}$ of $[0,1]$ into q classes such that at most q' of its classes intersect more than one class of $\mathcal{P}'$. Let R be the union of these exceptional classes, and let U be the step function which is equal to $W_{\mathcal{P}'}$ on $([0,1] \setminus R)^2$, and 0 on the complement. Then $\|W - U\|_\square$ can easily be bounded by decomposing the sets S, T in the definition of the cut-norm into the part contained in $[0,1] \setminus R$ and its complement. Using the fact that $\lambda(R) \leqslant \frac{q'}{q} \leqslant 2^{-79/(8\varepsilon^2)} \leqslant \varepsilon^2 2^{-79/8}$, this leads to the estimate

$$\|W - U\|_\square \leqslant \left(\frac{4\varepsilon}{9} + \sqrt{2\lambda(R)}\right)\|W\|_2 \leqslant \frac{\varepsilon}{2}\left(\frac{8}{9} + \sqrt{8 \cdot 2^{-79/8}}\right)\|W\|_2 \leqslant \frac{\varepsilon}{2}\|W\|_2.$$

By construction, U is a step function with steps in $\mathcal{P}$. Using the bound (3.11), we conclude that $\|W - W_{\mathcal{P}}\|_\square \leqslant 2\|W - U\|_\square \leqslant \varepsilon\|W\|_2$, which gives the first statement of the corollary. The second assertion in statement (i) is proved in a similar way, starting from a common refinement of $\mathcal{P}'$ and $\tilde{\mathcal{P}}$.

Proof of (ii). Let $V = V(G)$, let $n = |V|$, and assume without loss of generality that $\alpha_G = 1$. Choosing a partition $\mathcal{P}' = (V_1', \dots, V_{q'}')$ of V with $q' \leqslant 2^{81/(8\varepsilon^2)}$ classes such that $d_\square(G, G_{\mathcal{P}'}) \leqslant \frac{4\varepsilon}{9}\|G\|_2$, we would like to divide each class in $\mathcal{P}'$ into subclasses V_i such that all of them obey the condition (3.12). To this end, we proceed as follows: Starting with V_1', we successively remove sets $V_1, V_2, \dots$ from first V_1', then V_2', etc. so that

$$\left|\alpha_{G[V_i]} - \frac{1}{q}\right| < \alpha_{\max}(G) \quad \text{for all } i = 1, \dots, q \tag{A.1}$$

and after each step

$$\left|\sum_{i=1}^{t} \alpha_{G[V_i]} - \frac{t}{q}\right| \leqslant \frac{\alpha_{\max}(G)}{2}.$$

When it is not possible to further remove a set V_i from V_1' while maintaining these constraints, we are left with a (possibly empty) remainder R_1 that has weight $\alpha_{G[R_1]} < 1/q$ (otherwise, we could have continued for at least one more step). Continuing with $V_2', \dots, V_{q'}'$, we will eventually end up with $q - q' \leqslant t \leqslant q$ disjoint sets $V_1, \dots, V_t$ obeying the condition (A.1), and $r \leqslant q'$ non-empty remainders R_i such that the total weight of their union, $R = \bigcup_i R_i$, obeys the condition $|\alpha_{G[R]} - (q - t)/q| \leqslant \alpha_{\max}/2$. Using this condition, it is not hard to see that R can be split into $q - t$ final sets $V_{t+1}, \dots, V_q$ obeying the condition (A.1). Since each of the remainders had weight at most $1/q$, the total weight of R is at most $q'/q \leqslant 2^{-79/(8\varepsilon^2)} \leqslant \varepsilon^2 2^{-79/8}$.

From here on the proof is completely analogous to the proof of (i).

A.2. Proof of Lemma 2.5

We start with the observation that the proof of the last section actually gives the stronger bound

$$\|W - W_{\mathcal{P}}\|_{\square} \leqslant \varepsilon\left(\frac{8}{9} + \sqrt{8 \cdot 2^{-79/8}}\right)\|W\|_2 \leqslant 0.982\varepsilon\|W\|.$$

Applying this bound to W_G, this gives a weighted graph $H = W_G/\mathcal{P}$ on $q \geqslant 2^{20/\varepsilon^2} \geqslant 2^{20}/\varepsilon^2$ nodes such that H has nodeweights one and $\delta_{\square}(G, H) \leqslant 0.982\varepsilon$. Combined with Lemma 4.3, this gives the existence of a weighted graph $\widetilde{H}$ on $[q]$ such that

$$\delta_{\square}(G, \widetilde{H}) \leqslant 0.982\varepsilon + \frac{4}{\sqrt{q}} \leqslant \varepsilon,$$

as required.

A.3. Proof of Lemma 3.5

We notice that $\delta_{\square}(U, W)$ as well as the infima and limits in (3.14), (3.15) and (3.16) are continuous in both U and W, with respect to the $\|.\|_{\square}$ norm. This fact and Lemma 3.2 imply that it is enough to prove the lemma for graphons U and W that are interval step functions with equal steps, corresponding to some finite graphs G and G'.

Furthermore, the inequalities

$$\inf_{\phi,\psi}\|U^{\phi} - W^{\psi}\|_{\square} \leqslant \inf_{\psi}\|U - W^{\psi}\|_{\square} \leqslant \liminf_{n\to\infty}\min_{\pi}\|U - W^{\tilde{\pi}}\|_{\square}$$

are trivial, so it suffices to prove that

$$\delta_\square(U,W) \leqslant \left\| U^\phi - W^\psi \right\|_\square \tag{A.2}$$

for all measure-preserving maps $\phi, \psi : [0,1] \to [0,1]$, and

$$\limsup_{n\to\infty} \min_\pi \left\| U - W^{\tilde\pi} \right\|_\square \leqslant \delta_\square(U,W), \tag{A.3}$$

where the infimum is over all permutations of $[n]$.

To prove (A.2), we consider two weighted graphs G and G' with $\alpha_G = \alpha'_G = 1$, together with the step functions $U = W_G$ and $W = W_{G'}$. Let $I_1, \dots, I_n$ be the intervals $[0, \alpha_1(G)]$, $(\alpha_1(G), \alpha_1(G) + \alpha_2(G)]$, $\dots$, $(\alpha_1(G) + \cdots + \alpha_{n-1}(G), 1]$, and similarly for $I'_1, \dots, I'_{n'}$. For two measure-preserving maps $\phi, \psi : [0,1] \to [0,1]$, we then rewrite the norm on the right-hand side of (A.2) in the explicit form

$$\left\| U^\phi - W^\psi \right\|_\square = \sup_{S,T\subset[0,1]} \left| \int\limits_{S\times T} \bigl(W_G\bigl(\phi(x),\phi(y)\bigr) - W_{G'}\bigl(\psi(x),\psi(y)\bigr)\bigr)\,dx\,dy \right|.$$

The supremum on the right-hand side is attained when S and T are unions of sets of the form $V_{iu} = \phi^{-1}(I_i) \cap \psi^{-1}(I'_u)$, $i \in V(G)$, $u \in V(G')$. But for these sets, the integral on the right is a sum of terms of the form

$$\int\limits_{V_{iu}\times V_{jv}} \bigl(W_G\bigl(\phi(x),\phi(y)\bigr) - W_{G'}\bigl(\psi(x),\psi(y)\bigr)\bigr)\,dx\,dy = \beta_{ij}(G)\beta_{uv}(G')X_{iu}X_{jv}$$

where X_{iu} is the Lebesgue measure of the set V_{iu}. As a consequence, we have that

$$\left\| U^\phi - W^\psi \right\|_\square = d_\square\bigl(G[X], G'\bigl[X^\top\bigr]\bigr).$$

Using the fact that ϕ and ψ are measure-preserving, it is not hard to check that X is a coupling of the distributions $(\alpha_i(G))_{i\in V(G)}$ and $(\alpha_u(G'))_{u\in V(G')}$, implying that

$$\left\| U^\phi - W^\psi \right\|_\square = d_\square\bigl(G[X], G'\bigl[X^\top\bigr]\bigr) \geqslant \delta_\square(U,W). \tag{A.4}$$

To show (A.3), let us consider a coupling $X \in \mathcal{X}(G,G')$. First construct a special measure-preserving bijection $\psi : [0,1] \to [0,1]$ such that $\|W_G - W_{G'}^\psi\|_\square = d_\square(G[X], G'[X^\top])$. As shown above, this is equivalent to finding a measure-preserving map such that X_{iu} is the Lebesgue measure of $I_i \cap \psi^{-1}(I'_u)$. But the construction of such a map is straightforward. Indeed, for $iu \in V(G) \times V(G')$, let $b_i = \alpha_1(G) + \cdots + \alpha_{i-1}(G)$, $b'_u = \alpha_1(G') + \cdots + \alpha_{u-1}(G')$, $c_{iu} = b_i + X_{i1} + X_{i2} + \cdots + X_{i(u-1)}$ and $c'_{iu} = b'_i + X_{1u} + X_{2u} + \cdots + X_{nu}$. Let I_{iu} and I'_{iu} be the intervals $I_{iu} = (c_{iu}, c_{iu} + X_{iu}]$ and $I'_{iu} = (c'_{iu}, c'_{iu} + X_{iu}]$. We then choose ψ to be the translation that maps I_{iu} into I'_{iu}. Then $\psi^{-1}(I'_u) = \bigcup_i \psi^{-1}(I'_{iu}) = \bigcup_i I_{iu}$ and $I_i \cap \psi^{-1}(I'_u) = I_{iu}$, implying in particular that this set has measure X_{iu}, as required.

So we have two partitions $\{I_1, \dots, I_m\}$ and $\{J_1, \dots, J_m\}$ of $[0,1]$ into intervals, and ψ maps each I_k onto a $J_{f(k)}$ by translation. Furthermore, we also know that both U and W are constant on each rectangle $I_k \times I_l$ as well as on $J_k \times J_l$.

Let N be a large integer, and consider the partition $\{L_1, \dots, L_N\}$ of $[0,1]$ into intervals of size $1/N$. We define a permutation π of $[N]$. For every $k \leqslant m$, the intervals I_k and $J_{f(k)}$ have the same length, and so the numbers of intervals L_i contained in them can differ by at most one. Let π match the indices i such that $L_i \subseteq I_k$ with the indices j such that $L_j \subseteq J_{f(k)}$, with at most one exception. This way $\pi(i)$ is defined for at least $N - 3m$ integers $i \in [N]$. Call the corresponding intervals K_i *well-matched*. We extend π to a permutation of $[N]$ arbitrarily.

We see that $W^{\psi}(x,y) = W^{\tilde{\pi}}(x,y)$ whenever both x and y belong to well-matched intervals. Hence

$$\|W^{\psi} - W^{\tilde{\pi}}\|_{\square} \leqslant \|W^{\psi} - W^{\tilde{\pi}}\|_1 \leqslant \frac{6m}{N}\|W\|_{\infty},$$

and so

$$\|U - W^{\tilde{\pi}}\|_{\square} \leqslant \|U - W^{\psi}\|_{\square} + \|W^{\psi} - W^{\tilde{\pi}}\|_{\square} \leqslant \delta_{\square}(U,W) + \frac{6m}{N}\|W\|_{\infty}.$$

This implies (A.3).

A.4. Proof of Lemma 4.4

Starting with the proof of (4.11), let us assume that $k^2/n \leqslant \varepsilon/(11\log 2)$ (otherwise the bound (4.11) is trivial). Using the bound (4.9), we then estimate the probability on the left-hand side of (4.11) by

$$\begin{aligned}&\mathsf{P}\Big(\big|t\big(F, \mathbf{H}(n,W)\big) - \mathsf{E}\big[t\big(F, \mathbf{H}(n,W)\big)\big]\big| > \Big(\varepsilon - \frac{2k^2}{n}\Big)\Big)\\ &\quad\leqslant \mathsf{P}\Big(\big|t\big(F, \mathbf{H}(n,W)\big) - \mathsf{E}\big[t\big(F, \mathbf{H}(n,W)\big)\big]\big| > \varepsilon\Big(1 - \frac{1}{11\log 2}\Big)\Big).\end{aligned}$$

Applying Lemma 4.1(i) and the observation that $t(F, W[\{z_1, \dots, z_n\}])$ changes by at most $2k/n$ if we change one of the variables z_i we immediately obtain the bound (4.11).

Expressing $\mathbf{G}(n,W)$ as a function of the random variables $Z_1, \dots, Z_n$ introduced above, and observing that $t(F, \mathbf{G}(n,W))$ changes by at most k/n if we change one of the variables Z_i, the proof of (4.12) is virtually identical to that of (4.11). We leave the details to the reader.

A.5. Detail concerning Example 5.4

Let $H = K_{nm,nm}$ and $G = K_{n,n}$ be the graphs considered in Example 5.4. Here we show that for every n, there exists an m such that with large probability, $\|W_H - W_{G'}\|_{\square} \geqslant 1/10$ for every relabeling G' of G.

Indeed, let S' and T' be the color classes of G', and let I and J be the subsets of $[0,1]$ corresponding to S' and T'. Let X and Y be the subsets of $V(H)$ that correspond to I and J in W_H. Then

$$\|W_H - W_{G'}\|_{\square} \geqslant \int_{I \times J} (W_{G'} - W_H) = \frac{e_{G'}(S,T)}{n^2} - \frac{e_H(X,Y)}{n^2m^2} = \frac{1}{4} - \frac{e_H(X,Y)}{n^2m^2}.$$

Now because of the random labeling of $V(H)$, the expectation of the second term is $1/8$, and the value of the second term will be arbitrarily highly concentrated around this value if m is sufficiently large. There are only $n!$ possible relabelings of G, so with large probability the second term will be less than $3/20$ for all of these, which proves our claim.

A.6. Proofs of (7.1), (7.2) and (7.3)

The upper bound in each of the three equations is trivial, so we only need to prove the lower bounds. To prove these, we introduce the notation

$$W(S,T)=\int_{S\times T} W(x_1,x_2)\,dx_1\,dx_2$$

for the value of the "cut between S and T."

Proof of the lower bound in (7.1). This follows easily by inclusion–exclusion applied to $W(S\cup T, S\cup T)$. □

Proof of the lower bound in (7.2). To this end, we first approximate W by a weighted finite graph without loops. Next, we discard each point in $S\setminus T$ and $T\setminus S$ with probability $1/2$, and then we add each point in $S\cap T$ uniformly at random either to the remaining points in $S\setminus T$, or to the remaining points in $T\setminus S$. In expectation, the weighted number of edges between the resulting, disjoint sets is off by a factor of $1/4$, giving the first bound in (7.2). □

Proof of the lower bound in (7.3). To prove this bound, we consider two disjoint sets S and T and their complement $R=[0,1]\setminus(S\cup T)$. We then express each of the three cuts $W(S,T\cup R)$, $W(T,S\cup R)$, and $W(S\cup T,R)$, in terms of $W(S,T)$, $W(S,R)$ and $W(T,R)$. Combining these three relations leads to the desired bound in (7.3). □

References

[1] N. Alon, W. Fernandez de la Vega, R. Kannan, M. Karpinski, Random sampling and approximation of MAX-CSP problems, in: Proc. of the 34th ACM Symp. Theor. of Comp. (STOC), 2002, pp. 232–239.

[2] N. Alon, W. Fernandez de la Vega, R. Kannan, M. Karpinski, Random sampling and approximation of MAX-CSPs, J. Comput. System Sci. 67 (2003) 212–243.

[3] N. Alon, E. Fisher, M. Krivelevich, M. Szegedy, Efficient testing of large graphs, Combinatorica 20 (2000) 451–476.

[4] N. Alon, E. Fischer, I. Newman, A. Shapira, A combinatorial characterization of the testable graph properties: It's all about regularity, in: Proc. of the 38th ACM Symp. Theor. of Comp. (STOC), 2006, pp. 251–260.

[5] N. Alon, A. Naor, Approximating the cut-norm via Grothendieck's inequality, in: Proc. of the 36 ACM STOC, ACM Press, Chicago, 2004, pp. 72–80; journal version: SIAM J. Comput. 35 (2006) 787–803.

[6] N. Alon, A. Shapira, Every monotone graph property is testable, in: Proc. of the 37th ACM Symp. Theor. of Comp. (STOC), 2005, pp. 128–137.

[7] A. Barabasi, R. Albert, Emergence of scaling in random networks, Science 286 (1999) 509–512.

[8] I. Benjamini, O. Schramm, Recurrence of distributional limits of finite planar graphs, Electron. J. Probab. 6 (23) (2001) 1–13.

[9] N. Berger, C. Borgs, J.T. Chayes, A. Saberi, On the spread of viruses on the Internet, in: Proc. of the 16th ACM–SIAM Symp. on Discr. Alg. (SODA), 2005, pp. 301–310.

[10] B. Bollobas, C. Borgs, J.T. Chayes, O. Riordan, Directed scale-free graphs, in: Proc of the 14th ACM–SIAM Symp. on Discr. Alg. (SODA), 2003, pp. 132–139.

[11] B. Bollobas, O. Riordan, Mathematical results on scale-free random graphs, in: Handbook of Graphs and Networks, Wiley, Weinheim, 2003, pp. 1–34.
[12] B. Bollobas, O. Riordan, The diameter of a scale-free random graph, Combinatorica 24 (2004) 5–34.
[13] B. Bollobas, O. Riordan, J. Spencer, G. Tusnady, The degree sequence of a scale-free random graph process, Random Structures Algorithms 18 (2001) 279–290.
[14] C. Borgs, J. Chayes, L. Lovász, Moments of two-variable functions and the uniqueness of graph limits, preprint, http://arxiv.org/abs/0803.1244.
[15] C. Borgs, J. Chayes, L. Lovász, V.T. Sós, K. Vesztergombi, unpublished, 2003.
[16] C. Borgs, J.T. Chayes, L. Lovász, V.T. Sós, K. Vesztergombi, Counting graph homomorphisms, in: M. Klazar, J. Kratochvil, M. Loebl, J. Matoušek, R. Thomas, P. Valtr (Eds.), Topics in Discrete Mathematics, Springer, 2006, pp. 315–371.
[17] C. Borgs, J.T. Chayes, L. Lovász, V.T. Sós, K. Vesztergombi, Convergent graph sequences II. Multiway cuts and statistical physics, preprint, http://research.microsoft.com/~borgs/ConRight.pdf, 2007.
[18] C. Borgs, J.T. Chayes, L. Lovász, V.T. Sós, K. Vesztergombi, Convergence of sparse graph sequences, in preparation.
[19] C. Borgs, J.T. Chayes, L. Lovász, V.T. Sós, B. Szegedy, K. Vesztergombi, Graph limits and parameter testing, in: Proc. of the 38th ACM Symp. Theory of Comp., 2006, pp. 261–270.
[20] F. Chung, R.L. Graham, R.M. Wilson, Quasi-random graphs, Combinatorica 9 (1989) 345–362.
[21] P. Erdös, L. Lovász, J. Spencer, Strong independence of graphcopy functions, in: Graph Theory and Related Topics, Academic Press, 1979, pp. 165–172.
[22] E. Fischer, I. Newman, Testing versus estimation of graph properties, in: Proc. 37th Ann. ACM Symp. on Theory of Computing, Baltimore, MD, 2005, pp. 138–146.
[23] A. Frieze, R. Kannan, Quick approximation to matrices and applications, Combinatorica 19 (1999) 175–220.
[24] O. Goldreich, S. Goldwasser, D. Ron, Property testing and its connection to learning and approximation, J. ACM 45 (1998) 653–750.
[25] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, E. Upfal, Stochastic models for the web graph, in: Proc. of the 41th ACM Found. of Comp. Sci. (FOCS), 2000, pp. 57–65.
[26] L. Lovász, B. Szegedy, Limits of dense graph sequences, J. Combin. Theory Ser. B 96 (2006) 933–957.
[27] L. Lovász, B. Szegedy, Graph limits and testing hereditary graph properties, J. Math., in press, http://arxiv.org/abs/0803.1248.
[28] L. Lovász, B. Szegedy, Szemerédi's lemma for the analyst, Geom. Func. Anal. 17 (2007) 252–270.
[29] W. Rudin, Real and Complex Analysis, third ed., McGraw–Hill, New York, 1987.
[30] E. Szemerédi, Regular partitions of graphs, in: J.-C. Bermond, J.-C. Fournier, M. Las Vergnas, D. Sotteau (Eds.), Colloque Inter. CNRS, 1978, pp. 399–401.
[31] A. Thomason, Pseudorandom graphs, in: Random Graphs '85, in: North-Holland Math. Stud., vol. 144, North-Holland, Amsterdam, 1987, pp. 307–331.