Action convergence of operators and graphs

Ágnes Backhausz and Balázs Szegedy

Abstract. We present a new approach to graph limit theory that unifies and generalizes the two most well-developed directions, namely dense graph limits (even the more general L^p limits) and Benjamini–Schramm limits (even in the stronger local-global setting). We illustrate by examples that this new framework provides a rich limit theory with natural limit objects for graphs of intermediate density. Moreover, it provides a limit theory for bounded operators (called P-operators) of the form $L^\infty(\Omega) \to L^1(\Omega)$ for probability spaces Ω. We introduce a metric to compare P-operators (for example, finite matrices) even if they act on different spaces. We prove a compactness result, which implies that, in appropriate norms, limits of uniformly bounded P-operators can again be represented by P-operators. We show that limits of operators, representing graphs, are self-adjoint, positivity-preserving P-operators called graphops. Graphons, L^p graphons, and graphings (known from graph limit theory) are special examples of graphops. We describe a new point of view on random matrix theory using our operator limit framework.

1 Introduction

A fundamental question posed in the emerging field of graph limit theory is the following: *How can we measure similarity of graphs?* Each branch of graph limit theory is based on a similarity metric [28]. Experience shows that, to be useful in applications, the similarity metric should satisfy a few natural properties.

(1) **(Expressive power)** *The similarity metric should be fine enough to provide a rich enough picture of graph theory.*

(2) **(Compactness)** *The similarity metric should be coarse enough to provide many interesting Cauchy convergent graph sequences.*

(3) **(Limit objects)** *Limits of Cauchy convergent sequences of graphs should be naturally represented by "graph-like" analytic objects.*

The tension between the first and the second requirement makes the search for useful similarity metrics especially interesting. The so-called dense graph limit theory is based on a set of equivalent metrics. One of them is the $\delta_\square$-distance [9, 28, 29]. Convergence in $\delta_\square$ is equivalent to the convergence of subgraph densities. The completion of the set of all graphs in this metric is compact, and thus every graph sequence has a convergent subsequence, which is a very useful property. A shortcoming of dense graph limit theory is that sparse graphs are considered to be similar to the empty graph and thus it has not enough expressive power to study graphs in which the number of edges is subquadratic in the number of vertices. Another similarity notion was introduced by Benjamini and Schramm [4] to study bounded degree graphs that are basically the sparsest graphs. This metric requires an absolute bound for the largest degree and hence it cannot be used for graphs with super-linear number of edges. Graph sequences in which the number of edges is super-linear and subquadratic in terms of the number of vertices are called graphs of *intermediate density*.

Finding useful similarity notions for graphs of intermediate density is a major research direction in graph limit theory. There are many promising nonequivalent approaches to this subject [6–8, 18, 23, 32, 33, 35]. However, none of them provides a real unification of the most well-developed branches: dense graph limit theory (together with its L^p extension [7, 8]), Benjamini–Schramm limit theory (together with the stronger local-global convergence, see, e.g., [11, 21]), and corresponding limit objects: graphons, L^p graphons, and graphings.

In this paper, we take a new point of view on the subject. Instead of considering graphs as static structures, we focus more on the action and dynamics generated by graphs. One can associate various operators with graphs. The most well-known examples are: adjacency operators, Laplace operators, and Markov kernel operators (related to random walks). We formulate a framework theory of operator convergence and apply it to graph theory through representing operators.

The dynamical aspect is present in many existing limit theories. However, it has not been exploited to unify them. Limit objects, such as graphons and graphings, act on L^2 spaces of probability spaces. (Even the so-called L^p graphons can be viewed as operators of the form $L^q(\Omega) \to L^p(\Omega)$, where Ω is a probability space.) While graphons are compact operators represented by measurable functions of the form $W : [0,1]^2 \to [0,1]$, graphings are noncompact and are represented by singular measures on $[0,1]^2$ concentrated on edge sets of bounded degree Borel graphs [14, 21]. *A common property of all of these objects is that they are bounded operators in an appropriate norm and they act on function spaces of random variables.* Graphons and graphings are bounded in the usual L^2 operator norm $\|.\|_{2\to2}$, and L^p graphons are bounded in the $\|.\|_{q\to p}$ norm, where $p^{-1} + q^{-1} = 1$.

In spite of the fact that existing convergence notions for graphons and graphings are intuitively similar, the exact connection has not yet been explained from a functional analytic point of view. In this paper, we introduce a general convergence notion for operators acting on functions on probability spaces. We show that graphon convergence, L^p graphon convergence, and local-global convergence of graphings are all special cases of this general convergence notion. Moreover, we obtain a very general framework for graph limit theory by studying the convergence of operator representations of graphs.

We also demonstrate that the new limit theory for operators has applications beyond graph theory through a new approach to random matrix theory. An important motivation for this paper comes from a previous result, by the authors, which proves Gaussianity for almost eigenvectors of random regular graphs using graph limit techniques (local-global limits) and information theory [2]. It is very natural to ask if similar limit techniques can be used to study dense random matrices, such as matrices with i.i.d. ±1 entries. Available graph limit techniques proved to be too weak for this problem. Dense random matrices (when regarded as weighted graphs) converge to trivial objects in dense graph limit theory. Note that an interesting connection between dense graph limits and random matrices was investigated in [31, 37].

We propose a new limit approach for matrices, graphs, and operators, which is based on the following quite simple and natural probabilistic view point on matrix actions. Let $A \in \mathbb{R}^{n \times n}$ be an arbitrary matrix and let $v \in \mathbb{R}^n$ be a vector. Let M denote the $2 \times n$ matrix, whose rows are v and vA. Each column of M is an element in $\mathbb{R}^2$, thus, by choosing a random column, we obtain a probability distribution μ_v on $\mathbb{R}^2$ (see Figure 1). The following interesting question arises:

How much do we learn about A if we know the set of all probability measures μ_v arising this way?

It is easy to see, for example, that A is the identity matrix if and only if μ_v is supported on the line $y = x$ in $\mathbb{R}^2$ for every $v \in \mathbb{R}^n$. The matrix A is degenerate if and only if there is a measure μ_v, which is not the Dirac measure $\delta_{(0,0)}$, but it is supported on the line $y = 0$.

Philosophically, we regard each measure μ_v as an observation associated with the action of A and we regard the set of all possible observations $\{\mu_v : v \in \mathbb{R}^n\}$ as the profile of A. A useful fact about profiles is that they allow us to compare matrices of different sizes, since they are sets of probability measures on $\mathbb{R}^2$ independently of the sizes of the matrices. Another nice fact is that the profile of A contains rather detailed information about the eigenvalues of A and the entry distributions of the corresponding eigenvectors. It is easy to see that v is an eigenvector with eigenvalue λ if and only if the measure μ_v is supported on the line $y = \lambda x$ in $\mathbb{R}^2$. The entry distribution of v is simply the distribution of the x coordinates in μ_v.

It is useful to extend this idea to the case when k vectors $v_1, v_2, \ldots, v_k$ are considered simultaneously. (For some technical reasons, we will assume that $v_1, v_2, \ldots, v_k$ are in $[-1, 1]^n$.) In this case, M is the $2k \times n$ matrix with rows $\{v_i\}_{i=1}^k$ and $\{v_iA\}_{i=1}^k$. A random column in M yields a probability distribution on $\mathbb{R}^{2k}$, and the k-profile $\mathcal{S}_k(A)$ of A is the set of all such probability measures. We regard A and B to be similar if for small natural numbers k their k-profiles are close in the Hausdorff metric d_H defined for sets of probability measures on $\mathbb{R}^{2k}$ based on the Lévy–Prokhorov metric for individual measures (precise definition will be given in Section 2.) This similarity can be metrized by the formula

$$d_M(A, B) := \sum_{k=1}^{\infty} 2^{-k} d_H(S_k(A), S_k(B)).$$

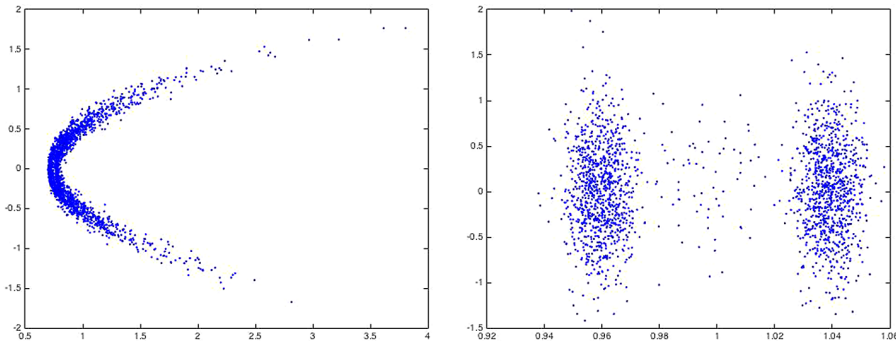A sequence of matrices converges in this metric if for every fixed k, their k-profiles converge in d_H.

Figure 1: Two probability measures of the form μ_v in the profile of a $2,000 \times 2,000$ random matrix.

The above ideas generalize naturally to the framework where $(\Omega, \mathcal{A}, \mu)$ is a probability space and A is an operator of the form $A : L^\infty(\Omega) \to L^1(\Omega)$. Such operators with an appropriate boundedness condition will be called *P-operators* (Definition 2.1). If $v \in L^\infty(\Omega)$, then both v and vA are random variables, and their joint distribution is a measure μ_v on $\mathbb{R}^2$. This allows us to define k-profiles, metric, and convergence for *P*-operators similarly as we defined them for matrices. Note that matrices are special *P*-operators, where the probability space is $[n] := \{1, 2, \ldots, n\}$ with the uniform distribution. In this case $L^\infty([n]) = L^1([n]) = \mathbb{R}^{[n]}$, and every matrix is a *P*-operator. Note that both graphons (symmetric measurable functions of the form $W : [0,1]^2 \to [0,1]$) and graphings (certain bounded degree Borel graphs on measure spaces) are special *P*-operators. We prove the next surprising result.

Theorem 1.1 *P-operator convergence (given by Definition 2.5) restricted to the set of graphons is the same as graphon convergence. Furthermore, P-operator convergence restricted to the set of graphings is equivalent to the local-global convergence of graphings.*

The proof of the above theorem relies on a recent result of the second author, which reformulates local-global convergence in terms of colored star metric [36]. Our main theorem (in an informal language) is the following.

Theorem 1.2 (Compactness and limit object) *Every sequence of P-operators with uniformly bounded $\|.\|_{\infty \to 1}$ norms has a Cauchy convergent subsequence with respect to d_M. Furthermore, if $p, q \in [1, \infty)$, then every Cauchy convergent sequence of $\|.\|_{p \to q}$ uniformly bounded P-operators has a limit, which is also a P-operator, and the same bound applies for its norm.*

We show that, under certain boundedness conditions, a number of important operator properties are closed with respect to *P*-operator convergence. This includes self-adjointness, positivity, and the positivity-preserving property. A *P*-operator $A : L^\infty(\Omega) \to L^1(\Omega)$ is called positivity-preserving if vA is a non-negative function on Ω whenever $v \in L^\infty(\Omega)$ is non-negative. The graph-like objects in the universe of *P*-operators are special *P*-operators called graphops.

Definition 1.1 A *graphop* is a positivity-preserving, self adjoint *P*-operator.

A particularly nice property of graphops is that they can be represented by symmetric finite measures ν on Ω^2 with absolutely continuous marginals (see Theorem 6.3).

Intuitively, the measure ν plays the role of the "edge set" of the graphop A. When scaled to a probability measure, ν can be used to sample a random element in $\Omega \times \Omega$, which is the analogue of a random directed edge in a finite graph. By disintegrating ν, we obtain measures ν_x for every $x \in \Omega$, describing "neighborhoods" in A.

Adjacency matrices of graphs (or positive weighted graphs), graphons, L^p-graphons, and graphings are all examples for graphops. A concrete example for a graphop (called spherical graphop), which is none of the previous classes, is explained in Figure 4.

Remark 1.3 **Action convergence vs. shape (quotient) convergence:** The representation of graph limits with measures on Ω^2 (or more specifically on $[0,1]^2$) is not new. It was proved in [26] that limits of shape convergent graph sequences can be represented by such measures. Roughly speaking, a graph sequence is shape convergent if the sets of possible bounded quotients of the graphs converge in the Hausdorff metric. Although shape convergence can be used to study intermediate density graph sequences, in general, it has less expressive power than action convergence. For many interesting graph sequences, action convergence is strictly finer than shape convergence. For example, shape convergence does not capture Benjamin–Schramm limits, whereas action convergence captures the even finer local-global limits. Another difference is that action convergence works for more general operators (for example, matrices with negative entries), but quotient convergence was not extended to this case due to various difficulties.

Remark 1.4 Graphops have *"edge densities"* and *"degrees."* If $A : L^\infty(\Omega) \to L^1(\Omega)$ is a graphop, then $1_\Omega A$ is a non-negative function. The expected value of $1_\Omega A$ is the edge density of A. The value of $1_\Omega A$ at a point $x \in \Omega$ is the "degree" of x. The distribution of the random variable $1_\Omega A$ is the "degree distribution" of A.

Remark 1.5 In dense graph limit theory, there is a rather natural answer to the question when two graphons are isomorphic, i.e., their $\delta_\square$ distance is 0 [10]. Similar statements would be useful for sparse graph limits. Unfortunately, it is well known that in the case of Benjamini–Schramm and local-global convergence graphing isomorphism is much more complicated and no natural description is known. This problem is also inherited by action convergence. However, on the positive side, there are several natural invariants such as eigenvalue gap or "degree distribution" that are isomorphism invariants, as the following remark shows.

Remark 1.6 It is a known phenomenon in the theory of dense graph limit theory that certain graphon parameters are not continuous with respect to the topology but they satisfy lower semicontinuity. An example for this is the entropy $\int -W \log W$ [27]. Similar phenomenon happens in the theory of local-global limits. While the second largest eigenvalue is continuous in dense graph limit theory, it is only lower semicontinuous in the theory of local-global convergence, in particular, the spectral gap of the limit object is greater than or equal to the limit of the spectral gap of the graphs. This can be seen by making the following argument more precise. First, after
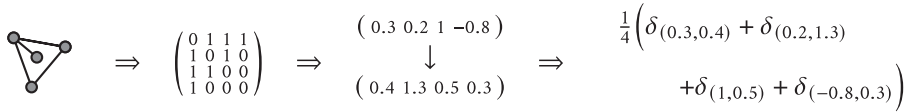
$$\begin{array}{ccc}
\begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} & \Rightarrow & \begin{array}{c} (\ 0.3\ 0.2\ 1\ -0.8\ ) \\ \downarrow \\ (\ 0.4\ 1.3\ 0.5\ 0.3\ ) \end{array}
\end{array}$$

$$\Rightarrow \quad \tfrac{1}{4}\Big(\delta_{(0.3,0.4)} + \delta_{(0.2,1.3)} + \delta_{(1,0.5)} + \delta_{(-0.8,0.3)}\Big)$$

Figure 2: Graph $\Rightarrow$ operator $\Rightarrow$ action $\Rightarrow$ measure (computing an element in the 1-profile of a graph).

an appropriate truncation, every eigenvector of the limit object can be approximated with almost eigenvectors on the finite graphs, and hence we can also find proper eigenvectors to an eigenvalue close to the original eigenvalue of the limit object. On the other hand, by considering disjoint union of a 3-regular random graph and a 3-regular graph without spectral gap, and looking at the eigenvectors of the first part with some additional zeros, we can decrease the spectral gap of the limit with this kind of perturbation of the graph. Hence spectral gap cannot be continuous with respect to action convergence. Other graph parameters happen to be continuous, for example, the size of the maximal cut normalized by the number of edges, which can be represented in the profile of the graph (Figure 2).

Adjacency operator convergence: We obtain a general graph convergence notion by considering the convergence of appropriately normalized adjacency matrices of graphs. For a graph G, let $A(G)$ denote the adjacency matrix of G. It turns out that for a bounded degree sequence of graphs $\{G_i\}_{i=1}^{\infty}$ the P-operator convergence of the sequence $\{A(G_i)\}_{i=1}^{\infty}$ is equivalent to local-global convergence (and thus it implies Benjamini–Schramm convergence). On the other hand, for a general graph sequence, the P-operator convergence of $A(G_i)/|V(G_i)|$ is equivalent to dense graph convergence. For graph sequences of intermediate growth, we normalize each operator $A(G_i)$ by a constant depending on G_i to obtain nontrivial convergence notion and limit object. A natural choice is the spectral radius given by $\|A(G_i)\|_{2\to 2}$ or, more generally, norms of the form $\|A(G_i)\|_{p\to q}$.

The convergence of normalized adjacency matrices leads to a rich limit theory for graphs of intermediate density. To demonstrate this, we give various examples for convergent sequences and limit objects. We calculate the limit object of hypercube graphs. The hypercube graph Q_n is the graph on $\{0,1\}^n$ in which two vectors are connected if they differ at exactly one coordinate. These graphs are very sparse and they are of intermediate density. The graph Q_n is vertex-transitive and can be represented as a Cayley graph of the elementary abelian group Z_2^n with respect to a minimal generating system. Quite surprisingly, the limiting P-operator turns out to be also a Cayley graph of the compact group Z_2^∞ with respect to a carefully chosen topological generating system. This illustrates that our limit objects give natural representations of convergent sequences. We calculate, similarly, natural representations for other convergent graph sequences such as increasing powers of regular graphs and incidence graphs of projective planes.

Random walk metric and convergence: A possible limitation for the use of adjacency operator convergence is that it may trivialize if the degree distribution is very uneven in a graph sequence. The simplest examples are stars and subdivisions of complete
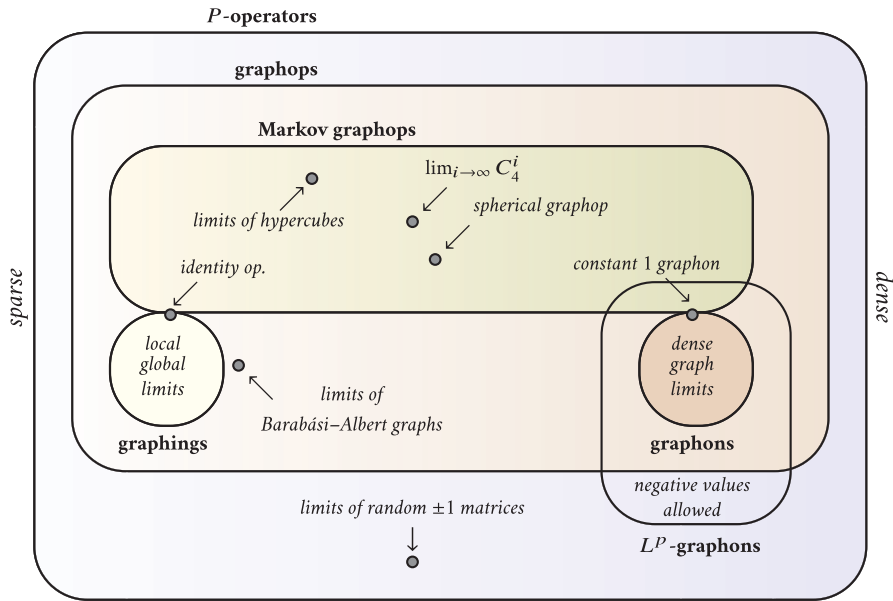
Figure 3: Universe of *P*-operators.

graphs. In the star graph S_n, there is one vertex with degree n and n vertices with degree 1. When normalized in any reasonable way, they converge to the 0 operator. The property that a graph has very uneven degree distribution is related to the property that a random walk on the graph spends a positive proportion of the time in a negligible fraction of the vertex set. A natural way to counterbalance this problem is to use Markov kernels of random walks instead of adjacency operators. (Such a modified limit was first used by Benjamini and Curien in the case of bounded degree graphs [3].) The *P*-operator language shows a nice advantage, in this case, to the plain matrix language (Figure 3). Even for finite graphs G, the corresponding Markov kernel is not just a matrix. The underlying probability space on $V(G)$ is modified from the uniform distribution to the stationary distribution ν_G of the random walk. Note that $\nu_G(i)$ is proportional to the degree $d(i)$ of $i \in V(G)$. The operator $M(G)$ is given by

$$(1.1) \qquad \big(\nu M(G)\big)(i) = d(i)^{-1} \sum_{(i,j) \in E(G)} \nu(j)$$

for $i \in \mathrm{supp}(\nu_G)$. Although $M(G)$ is not symmetric when viewed as a matrix, its action on $L^2(V(G), \nu_G)$ is self-adjoint. Thus, $M(G)$ is a positivity-preserving, self-adjoint *P*-operator with the property that $1_{V(G)} M(G) = 1_{V(G)}$. The last property is called 1-regularity.

The random walk metric d_{RW} on finite nonempty graphs is given by

$$d_{\mathrm{RW}}(G_1, G_2) := d_M(M(G_1), M(G_2)).$$

*The completion of the set of finite, nonempty graphs in d_{RW} is a compact space $\mathcal{G}_{\mathrm{RW}}$. Elements of $\mathcal{G}_{\mathrm{RW}}$ can be represented by **Markov graphops**, i.e., positivity-preserving, self-adjoint 1-regular P-operators.*

Note that if A is a Markov graphop, then $\|A\|_{2 \to 2} = 1$. We will see that Markov graphops can also be represented by symmetric self-couplings of probability spaces $(\Omega, \mathcal{A}, \mu)$. A symmetric self-coupling is a probability measure ν on $(\Omega \times \Omega, \mathcal{A} \otimes \mathcal{A})$ such that ν is symmetric with respect to interchanging the coordinates and both marginals of ν are equal to μ. A very pleasant property of the set of all Markov graphops is that it is compact in the metric d_M (see Theorem 3.3), and thus we do not need any extra conditions to guarantee convergent subsequences.

We will show in the examples section (Section 12) that stars and subdivisions of complete graphs converge to natural and nontrivial limit objects according to random walk convergence. Note that random walk convergence coincides with normalized adjacency operator convergence for regular graphs (graphs in which every degree is the same).

As we mentioned before, random walk convergence is very convenient. Every graph sequence $\{G_i\}_{i=1}^{\infty}$ with nonempty graphs has a convergent subsequence in d_{RW}, and the limit object is usually an interesting structured object independently of the sparsity of the sequence. The most trivial the limit object one can get is the quasi-random Markov graphop, which can be represented by the constant 1 graphon $W(x, y) := 1$. This occurs, for example, if the second largest eigenvalue in absolute value of $M(G_i)$ is $o(1)$.

Extended random walk convergence: Finally, we describe a general convergence notion that combines the advantages of adjacency operator convergence and random walk convergence. A feature of random walk convergence is that some information may be lost in the limit regarding degree distributions. It turns out that there is a rather natural way to solve this problem, using a mild extension of random walk convergence based on a simultaneous version of action convergence. The principle of action convergence allows us to introduce the convergence of pairs (A, f), where $A : L^{\infty}(\Omega) \to L^1(\Omega)$ is a P-operator and f is a measurable function on Ω. Roughly speaking, this goes by considering f as a reference function that is automatically included as the last function into every function system used in the definition of the k-profile of A. More precisely, we define $S_k(A, f)$ as the set of all possible joint distributions of the random variables $\{v_i\}_{i=1}^{k}$, $\{v_i A\}_{i=1}^{k}$ and f on Ω, where the values of v_i are in $[-1, 1]$. We have that $S_k(A, f)$ is a set of probability measures on $\mathbb{R}^{2k+1}$.

We can use the extra function to store information on the degrees of vertices in G. For a graph G, let d_G^* denote a function on $V(G)$ that is an appropriately normalized version of the degree function d_G. We can represent G by the pair $(M(G), d_G^*)$ (recall equation (1.1)). In the limit, we obtain a similar pair (A, f), where A is a Markov graphop. The non-negative function f^{-1} (which may also take the value ∞) can be used to "re-scale" the probability measure on Ω to a possibly infinite measure.

Pairs of the form (A, f) can also be used to represent generalized graphons of the form $W : \mathbb{R}^+ \times \mathbb{R}^+ \to [0, 1]$, where W is symmetric and $\|W\|_1 \leq \infty$. This construction will be described in Section 5. Note that these generalized graphons arise in the recently emerging theory of graphexes [6, 23].

Sparse random graphs: An important application of graph limit theory is in the study of random graph models. By useful limit theories, we can view large random graphs as approximations of a single idealistic infinite object. For example, in the framework of dense graph limit theory, growing Erdős–Rényi random graphs with edge probability 1/2 converge to the constant 1/2 graphon. Note that the existence of a single limit object requires a concentration of measure type result for the random model in the given graph metric. These concentration results are interesting even if an actual representation of the limit object is unknown. A good example for this is the fact that for fixed d random d-regular graphs on n points are concentrated in the local-global topology if n is large [1]. It is a far reaching open problem, however, that they concentrate around the same point for growing n. The case of preferential attachment trees is much better understood. They converge in the Benjamini–Schramm metric [5, 34], moreover they are trees and thus they are hyperfinite graphs. It follows that they also converge in the local-global topology. Action convergence for bounded degree graphs is equivalent to local-global convergence by Theorem 9.2, and thus this theory includes all these facts. The next natural question is what happens for sparse but not bounded degree (intermediate density) graphs. It turns out that the intermediate density Erdős–Rényi graphs (normalized by the average degree) all converge to the constant 1 graphon, as Remark 2.15 in Section 2 shows. In this sense, Erdős–Rényi graphs behave similar to the dense setting.

Applications to random matrix theory: As we mentioned earlier, the notion of P-operator convergence was partially motivated by efforts to find a fine enough convergence notion such that random matrices converge to a structured, nontrivial object. The study of this limiting object can help in describing approximate properties of random matrices, such as entry distributions of eigenvectors and almost eigenvectors. In dense graph limit theory, random matrices with i.i.d. ±1 entries converge, but the limit object is the constant 0 function. (In a refinement of dense limit theory [25] the limit object is the constant function on $[0,1]^2$ whose value is the uniform probability measure on $\{1, -1\}$.)

Our main observation about random matrices is that k-profiles of appropriately normalized random matrices are nontrivial rich objects and their study brings a new point of view on random matrix theory. Let $\mathbb{G}_n$ denote a random matrix whose entries are i.i.d. zero-mean $\pm 1/\sqrt{n}$-valued random variables. The normalizing factor $\sqrt{n}$ is needed to obtain bounded spectral radius. With probability close to 1, we have that $\|\mathbb{G}_n\|_{2\to 2}$ is close to 2, see, e.g., [19]. In this paper, we do ground work on the limiting properties of $\mathbb{G}_n$ according to action convergence. We prove a *concentration of measure* type statement for $\mathbb{G}_n$ with respect to the metric d_M. This means that for large n the matrix-valued random variable $\mathbb{G}_n$ is well concentrated in the metric space of P-operators together with distance d_M. This concentration result, together with our compactness results, implies that for certain good sequences $\{n_i\}_{i=1}^{\infty}$ of natural numbers, $\{\mathbb{G}_{n_i}\}_{i=1}^{\infty}$ is convergent with probability one and the limit object is represented by some P-operator L acting on $L^2([0,1])$ with bounded $\|.\|_{2\to 2}$ norm. In this paper, we leave the question open whether the sequence of all natural numbers is a good sequence. Note that a similar open problem is known for random regular graphs using local-global convergence [1] and it is known that a positive answer would imply the convergence of a great number of interesting graph parameters. For

our application, it will be enough that any sequence of natural numbers contains a good subsequence. Our general results in this paper prepare a follow-up paper, which focuses on the limiting properties of random matrices with a special emphasis on eigenvectors and almost eigenvectors.

2 Limits of matrices and operators

For k vectors $\{v_i \in \mathbb{R}^n\}_{i=1}^k$, let us define their joint empirical entry distribution, denoted by $\mathcal{D}(v_1, v_2, \ldots, v_k)$, as the probability measure on $\mathbb{R}^k$ given by

$$(2.1) \qquad \mathcal{D}(v_1, v_2, \ldots, v_k) := n^{-1} \sum_{j=1}^n \delta_{(v_{1,j}, v_{2,j}, \ldots, v_{k,j})},$$

where $v_{i,j}$ denotes the jth component of v_i and δ_x denotes the Dirac measure at $x \in \mathbb{R}^k$. A natural view on empirical entry distributions is the following. Consider $[n] := \{1, 2, \ldots, n\}$ as a probability space with the uniform distribution $\mu_{[n]}$ and vectors in $\mathbb{R}^n$ as functions of the form $v : [n] \to \mathbb{R}$. From this view point, vectors are random variables and matrices in $\mathbb{R}^{n \times n}$ are operators acting on the space of random variables on the probability space $([n], \mu_{[n]})$. The joint empirical entry distribution $\mathcal{D}(v_1, v_2, \ldots, v_k)$ is simply the joint entry distribution of the vectors $v_1, v_2, \ldots, v_k$ viewed as random variables.

Let $(\Omega, \mathcal{A}, \mu)$ (or shortly Ω) be a probability space and assume that $v_1, v_2, \ldots, v_k$ are $\mathbb{R}$-valued measurable functions on Ω. We denote by $\mathcal{D}(v_1, v_2, \ldots, v_k)$ the joint distribution of $v_1, v_2, \ldots, v_k$. In other words, it is the push-forward of the measure μ under the map $x \mapsto (v_1(x), v_2(x), \ldots, v_k(x))$, which is a Borel measure on $\mathbb{R}^k$.

Definition 2.1 A *P-operator* is a linear operator of the form $A : L^\infty(\Omega) \to L^1(\Omega)$ such that

$$\|A\|_{\infty \to 1} := \sup_{v \in L^\infty(\Omega)} \|vA\|_1 / \|v\|_\infty$$

is finite. We denote by $\mathcal{B}(\Omega)$ the set of all *P-operators* on Ω.

Remark 2.1 If $\Omega = [n]$ and $\mu = \mu_{[n]}$, then $L^1(\Omega) = L^\infty(\Omega) = \mathbb{R}^n$. In this case, $\mathcal{B}(\Omega)$ is the set of all $n \times n$ matrices. Thus, every matrix $A \in \mathbb{R}^{n \times n}$ is a *P-operator*.

For a set $S \subseteq \mathbb{R}$ we denote by $L_S^\infty(\Omega)$ the set of bounded measurable functions on Ω whose values are in S.

Definition 2.2 (*k*-profile of *P-operators*) For a *P-operator* $A \in \mathcal{B}(\Omega)$, we define the k-profile of A, denoted by $\mathcal{S}_k(A)$, as the set of all possible probability measures of the form

$$(2.2) \qquad \mathcal{D}(v_1, v_2, \ldots, v_k, v_1 A, v_2 A, \ldots, v_k A),$$

where $v_1, v_2, \ldots, v_k$ run through all possible k-tuples of functions in $L_{[-1,1]}^\infty(\Omega)$.

For joint distributions of the form (2.2), we will often use the shorthand notation

$$\mathcal{D}_A(v_1, v_2, \ldots, v_k) := \mathcal{D}(v_1, v_2, \ldots, v_k, v_1 A, v_2 A, \ldots, v_k A).$$

Let $\mathcal{P}(\mathbb{R}^k)$ denote the set of Borel probability measures on $\mathbb{R}^k$ for $k \in \mathbb{N}$.

Definition 2.3 (Lévy–Prokhorov metric)   The Lévy–Prokhorov metric d_{LP} on $\mathcal{P}(\mathbb{R}^k)$ is defined by

$$d_{\mathrm{LP}}(\eta_1, \eta_2) = \inf\{\varepsilon > 0 : \eta_1(U) \leq \eta_2(U^\varepsilon) + \varepsilon \text{ and } \eta_2(U) \leq \eta_1(U^\varepsilon) + \varepsilon \text{ for all } U \in \mathcal{B}_k\},$$

where $\mathcal{B}_k$ is the Borel σ-algebra on $\mathbb{R}^k$ and U^ε is the set of points that have distance smaller than ε from U.

Definition 2.4 (Hausdorff metric)   We measure the distance of subsets X, Y in $\mathcal{P}(\mathbb{R}^k)$ using the Hausdorff metric d_H.

$$d_H(X, Y) := \max\left\{ \sup_{x \in X} \inf_{y \in Y} d_{\mathrm{LP}}(x, y) , \ \sup_{y \in Y} \inf_{x \in X} d_{\mathrm{LP}}(x, y) \right\}.$$

Note that $d_H(X, Y) = 0$ if and only if $\mathrm{cl}(X) = \mathrm{cl}(Y)$, where cl is the closure in d_{LP}. It follows that d_H is a pseudometric for all subsets in $\mathcal{P}(\mathbb{R}^k)$ and it is a metric for closed sets.

Definition 2.5 (Action convergence of P-operators)   We say that a sequence of P-operators $\{A_i \in \mathcal{B}(\Omega_i)\}_{i=1}^\infty$ is action convergent if for every $k \in \mathbb{N}$ the sequence $\{\mathcal{S}_k(A_i)\}_{i=1}^\infty$ is a Cauchy sequence in d_H.

Remark 2.2   We will often use the following consequence of the definition. For an action convergent sequence of operators $\{A_i\}_{i=1}^\infty$ for every $v \in L^\infty_{[-1,1]}(\Omega)$, there are elements $v_i \in L^\infty_{[-1,1]}(\Omega_i)$ such that $\mathcal{D}_{A_i}(v_i)$ weakly converges to $\mathcal{D}_A(v)$ as i goes to infinity.

Remark 2.3   The completeness of $(\mathcal{P}(\mathbb{R}^k), d_{LP})$ implies that the induced Hausdorff topology is also complete [20]. Therefore, a sequence $\{A_i\}_{i=1}^\infty$ in the above definition is convergent if and only if for every $k \in \mathbb{N}$ there is a closed set X_k such that $\lim_{i \to \infty} d_H(\mathcal{S}_k(A_i), X_k) = 0$.

Definition 2.6 (Metrization of action convergence)   For two P-operators A, B, let

$$(2.3) \qquad\qquad d_M(A, B) := \sum_{k=1}^\infty 2^{-k} d_H(\mathcal{S}_k(A), \mathcal{S}_k(B)).$$

For a probability measure μ on $\mathbb{R}^k$, let $\tau(\mu) \in [0, \infty]$ denote the quantity

$$(2.4) \qquad\qquad \max_{1 \leq i \leq k} \int_{(x_1, x_2, \ldots, x_k) \in \mathbb{R}^k} |x_i| \, d\mu.$$

For $c \in \mathbb{R}^+$ and $k \in \mathbb{N}$, let

$$\mathcal{P}_c(\mathbb{R}^k) := \{\mu : \mu \in \mathcal{P}(\mathbb{R}^k), \tau(\mu) \leq c\}.$$

Furthermore, let $\mathcal{Q}_c(\mathbb{R}^k)$ denote the set of closed sets in the metric space $(\mathcal{P}_c(\mathbb{R}^k), d_{\mathrm{LP}})$. The following lemma is an easy consequence of classical results.

Lemma 2.4   *The metric spaces $(\mathcal{P}_c(\mathbb{R}^k), d_{\mathrm{LP}})$ and $(\mathcal{Q}_c(\mathbb{R}^k), d_H)$ are both compact and complete metric spaces.*

Proof   Markov's inequality gives uniform tightness in $\mathcal{P}_c(\mathbb{R}^k)$, which implies the compactness of $(\mathcal{P}_c(\mathbb{R}^k), d_{\mathrm{LP}})$. It is known that Hausdorff distance for the closed

subsets in a compact space is again compact (see, e.g., [20] or "Hausdorff metric" in [22]). ∎

Lemma 2.5   *Let $A \in \mathcal{B}(\Omega)$ and let $c := \max\{\|A\|_{\infty \to 1}, 1\}$. Then for every $k \in \mathbb{N}$, we have that $\mathcal{S}_k(A) \in \mathcal{Q}_c(\mathbb{R}^{2k})$.*

Proof   Let $\{v_i\}_{i=1}^k$ be a system of vectors in $L_{[-1,1]}^\infty(\Omega)$. We have that $\|v_i\|_1 \le \|v_i\|_\infty \le 1$ and $\|v_i A\|_1 \le \|A\|_{\infty \to 1}$ holds for $1 \le i \le k$. Since the first moments of the absolute values of the coordinates in (2.2) are given by $\{\|v_i\|_1\}_{i=1}^k$ and $\{\|v_i A\|_1\}_{i=1}^k$, it follows that $\mathcal{S}_k(A) \in \mathcal{P}_c(\mathbb{R}^k)$. Further using that $\mathcal{S}_k(A)$ is closed, the proof is complete.   ∎

Lemma 2.4 and Lemma 2.5 have the following corollary.

Lemma 2.6 (Sequential compactness)   *Let $\{A_i\}_{i=1}^\infty$ be a sequence of P-operators with uniformly bounded $\|.\|_{\infty \to 1}$ norms. Then, it has an action convergent subsequence.*

Proof   By the previous lemmas, for every k, we can extract a subsequence for which $\mathcal{S}_k(A)$ is convergent. By a diagonalization argument, we can also find a subsequence for which $\mathcal{S}_k(A)$ is convergent for every k at the same time, and this implies action convergence.   ∎

For a real number $p \in [1, \infty]$ and measurable function $v : \Omega \to \mathbb{R}$, we have that

$$\|v\|_p := \left(\int |v|^p \, d\mu \right)^{1/p} \in \mathbb{R}^+ \cup \{0, \infty\}.$$

Note that if $p = \infty$, then $\|v\|_\infty$ denotes the "essential maximum" of v. It is well known that for $p \le q$, we have $\|v\|_p \le \|v\|_q$ for any measurable function v on Ω. Let $p, q \in [1, \infty]$ be real numbers and let $A : L^\infty(\Omega) \to L^1(\Omega)$ be a linear operator. The operator norm $\|A\|_{p \to q}$ is defined by

$$\|A\|_{p \to q} := \sup_{v \in L^\infty(\Omega)} \|v A\|_q / \|v\|_p.$$

We say that A is (p, q)-bounded if $\|A\|_{p \to q}$ is finite. We have that if $p', q' \in [1, \infty]$ satisfy $p' \ge p$ and $q' \le q$ then $\|A\|_{p' \to q'} \le \|A\|_{p \to q}$. We denote by $\mathcal{B}_{p,q}(\Omega)$ the set of (p, q)-bounded linear operators from $L^\infty(\Omega)$ to $L^1(\Omega)$. If $p', q' \in [1, \infty]$ satisfy $p' \ge p$ and $q' \le q$, then $\mathcal{B}_{p,q}(\Omega) \subseteq \mathcal{B}_{p',q'}(\Omega)$. In particular, $\mathcal{B}_{p,q}(\Omega) \subseteq \mathcal{B}_{p,1}(\Omega)$ and $\mathcal{B}(\Omega) = \mathcal{B}_{\infty,1}(\Omega)$ contains $\mathcal{B}_{p,q}$ for every $p, q \in [1, \infty]$.

Remark 2.7 (L^2 theory)   If a P-operator A satisfies $\|A\|_{p \to q} < \infty$, then A extends uniquely to an operator of the form $L^p(\Omega) \to L^q(\Omega)$. In this sense, by slightly abusing the notation, we can identify the set $\mathcal{B}_{p,q}(\Omega)$ with the set of operators $L^p(\Omega) \to L^q(\Omega)$ with bounded $\|.\|_{p \to q}$ norm. An especially nice class of P-operators is the set $\mathcal{B}_{2,2}(\Omega)$. If Ω is fixed then these operators are closed with respect to composition and they form a so-called von Neumann algebra.

Remark 2.8 (L^p graphons as P-operators)   Let $p \in (0, \infty]$ and assume that $W : \Omega \times \Omega \to \mathbb{R}$ is a measurable function such that

$$\|W\|_p := \left(\int |W|^p \, d\mu^2 \right)^{1/p} < \infty.$$

Let $q := p/(p-1)$. We can associate an operator $A_W : L^q(\Omega) \to L^p(\Omega)$ with W, defined by $(fA_W)(x) := \int f(y) W(y,x) \, d\mu$. It is easy to see that $\|A_W\|_{q \to p} < \infty$ and thus $A_W \in \mathcal{B}_{q,p}$ is a P-operator representing the so-called L^p-graphon W. For a theory of L^p-graphon convergence, see [7]. It follows from our theory that for sequences of L^p-graphons $\{W_i\}_{i=1}^{\infty}$ with uniformly bounded L^p-norms, action convergence of the representing operators A_{W_i} is equivalent to L^p-graphon convergence (see Section 8, in particular, Theorem 8.2).

The following theorem is one of the main results in this paper. Its proof can be found in Section 4.

Theorem 2.9 (Existence of limit object)   *Let $p \in [1, \infty)$ and $q \in [1, \infty]$. Let $\{A_i\}_{i=1}^{\infty}$ be a convergent sequence of P-operators with uniformly bounded $\|.\|_{p \to q}$ norms. Then there is a P-operator A such that $\lim_{i \to \infty} d_M(A_i, A) = 0$, and $\|A\|_{p \to q} \leq \sup_{i \in \mathbb{N}} \|A_i\|_{p \to q}$.*

For a P-operator A and $k \in \mathbb{N}$ let $\mathcal{S}_k^*(A)$ denote the closure of $\mathcal{S}_k(A)$ in the space $(\mathcal{P}(\mathbb{R}^{2k}), d_{\mathrm{LP}})$.

Definition 2.7 (Weak equivalence and weak containment)   Let A and B be two P-operators. We say that A and B are *weakly equivalent* if $d_M(A, B) = 0$. We have that A and B are weakly equivalent if and only if $\mathcal{S}_k^*(A) = \mathcal{S}_k^*(B)$ holds for every $k \in \mathbb{N}$ (this follows from the characterization of $d_H(X, Y) = 0$ in Definition 2.4). We say that A is *weakly contained* in B if $\mathcal{S}_k^*(A) \subseteq \mathcal{S}_k^*(B)$ holds for every $k \in \mathbb{N}$. We denote weak containment by $A \prec B$.

It is easy to see that norms of the form $\|.\|_{p \to q}$ are invariant with respect to weak equivalence (these norms can be read off from the 1-profiles of P-operators). Let $\mathcal{X}$ denote the set of weak equivalence classes of P-operators and let $\mathcal{X}' \subset \mathcal{X}$ denote the set of equivalence classes of P-operators defined on atomless probability spaces. For $c \in \mathbb{R}^+, p \in [1, \infty), q \in (1, \infty]$ let $\mathcal{X}_{p,q,c} := \{A : A \in \mathcal{X}, \|A\|_{p \to q} \leq c\}$ and $\mathcal{X}'_{p,q,c} := \{A : A \in \mathcal{X}', \|A\|_{p \to q} \leq c\}$.

The next theorem follows from Lemma 2.6, Theorem 2.9, and Lemma 3.1.

Theorem 2.10 (Compactness)   *For every $c \in \mathbb{R}^+, p \in [1, \infty), q \in (1, \infty]$ the spaces $(\mathcal{X}_{p,q,c}, d_M)$ and $(\mathcal{X}'_{p,q,c}, d_M)$ are compact.*

Corollary 2.11   *Let $c \in \mathbb{R}^+$. We have that $\{W : W \in \mathcal{B}_{2 \to 2}([0,1]), \|W\|_2 \leq c\}$ is compact in the topology generated by d_M.*

Lemma 2.12   *Let $k \in \mathbb{N}$ and let A, B be P-operators both in $\mathcal{B}(\Omega)$ for some probability space $(\Omega, \mathcal{A}, \mu)$. Then,*

$$d_H(\mathcal{S}_k(A), \mathcal{S}_k(B)) \leq \|A - B\|_{\infty \to 1}^{1/2} (2k)^{3/4}.$$

Proof   Let $x \in \mathcal{S}_k(A)$ be arbitrary. We have that there are functions $v_1, v_2, \ldots, v_k \in L_{[-1,1]}^{\infty}(\Omega)$ such that x is equal to the probability measure $\mathcal{D}_A(\{v_i\}_{i=1}^k)$. Let $y := \mathcal{D}_B(\{v_i\}_{i=1}^k) \in \mathcal{S}_k(B)$. Since

$$\|v_i A - v_i B\|_1 \leq \|v_i\|_{\infty} \|A - B\|_{\infty \to 1} \leq \|A - B\|_{\infty \to 1}$$

holds for every $i \in [k]$, we have by Lemma 13.2 that $d_{\mathrm{LP}}(x, y) \leq \|A - B\|_{\infty \to 1}^{1/2}(2k)^{3/4}$. We obtained that

$$\sup_{x \in \mathcal{S}_k(A)} \inf_{z \in \mathcal{S}_k(B)} d_{\mathrm{LP}}(x, z) \leq \|A - B\|_{\infty \to 1}^{1/2}(2k)^{3/4}.$$

By switching the roles of A and B and repeating the same argument we get the above inequality with A and B switched. This implies the statement of the lemma. ∎

Lemma 2.13 (Norm distance vs. d_M distance)   *Assume that A, B are P-operators acting on the same space $L^\infty(\Omega)$. We have*

$$d_M(A, B) \leq 3\|A - B\|_{\infty \to 1}^{1/2} \leq 3\|A - B\|_{2 \to 2}^{1/2}.$$

Proof   Using Lemma 2.12, we obtain that

$$d_M(A, B) \leq \|A - B\|_{\infty \to 1}^{1/2} \sum_{k=1}^{\infty} 2^{-k}(2k)^{3/4} \leq 3\|A - B\|_{\infty \to 1}^{1/2}.$$

The last inequality is clear from the argument after Lemma 2.6. ∎

Let $A \in \mathcal{B}(\Omega, \mathcal{A}, \mu)$ be a P-operator. We define the bilinear form $(f, g)_A$ for functions $f, g \in L^\infty(\Omega)$ by

$$(2.5) \qquad (f, g)_A := \int_\Omega (fA)g \, d\mu = \mathbb{E}((fA)g).$$

Note that

$$\left|(f, g)_A\right| \leq \|f\|_\infty \|g\|_\infty \|A\|_{\infty \to 1},$$

and thus $(f, g)_A$ is finite. In general, if $\|A\|_{p \to q} < \infty$ holds for a conjugate pair with $1/p + 1/q = 1$, then we have

$$(2.6) \qquad \left|(f, g)_A\right| \leq \|f\|_p \|g\|_p \|A\|_{p \to q}.$$

We define the cut norm of A by

$$\|A\|_\square := \sup_{S, T \in \mathcal{A}} \left|(1_S, 1_T)_A\right|.$$

It is well known (see equation (3.6) in [9] or Lemma 8.11 in [27]) that

$$(2.7) \qquad \|A\|_\square \leq \|A\|_{\infty \to 1} \leq 4\|A\|_\square,$$

which means that $\|.\|_\square$ is equivalent to the norm $\|.\|_{\infty \to 1}$. Let $\psi : \Omega \to \Omega$ be an invertible measure-preserving transformation with measure-preserving inverse. The transformation ψ induces a natural, linear action on $L^\infty(\Omega)$, also denoted by ψ, defined by $f^\psi(x) := f(\psi(x))$. Furthermore, for $A \in \mathcal{B}(\Omega)$ let $A^\psi := \psi^{-1} \circ A \circ \psi$. It is easy to see that if $A \in \mathcal{B}(\Omega)$, then $A^\psi \in \mathcal{B}(\Omega)$ and that $d_M(A, A^\psi) = 0$. Let

$$\delta_\square(A, B) := \inf_{\psi, \phi} \|A^\psi - B^\phi\|_\square,$$

where ψ, ϕ run through all invertible measure, preserving transformations of Ω. The proof of the next lemma follows from Lemma 2.13 and inequality (2.7).

Lemma 2.14   *Assume that A, B are P-operators acting on the same space $L^\infty(\Omega)$. Then $d_M(A, B) \le 12\delta_\square(A, B)^{1/2}$.*

Remark 2.15   Let G_n denote the Erdős–Rényi graph with average degree $f(n)$ such that $\lim_{n\to\infty} f(n) = \infty$. Let A_n be adjacency operator of G_n and let B_n be the $n \times n$ matrix with every entry equal to $1/n$. Theorem 6.2 and 6.3 of [16] implies that the largest eigenvalue of $A_n/f(n) - B_n$ is $o(1)$ with probability close to 1. Recalling that the $2 \to 2$ norm is related to the spectrum, it follows from Lemma 2.13 that if n is large enough then $d_M(A_n/f(n), B_n) = o(1)$ and furthermore $d_M(B_n, W) = o(1)$, where W is the constant 1 graphon. It follows that the random graphs G_n converge in probability to W.

3 *P*-operators with special properties

The goal of this chapter is to show that various fundamental properties behave well with respect to P-operator convergence.

Definition 3.1   Let $A \in \mathcal{B}(\Omega)$ be a P-operator.

- A is *self-adjoint* if $(v, w)_A = (w, v)_A$ holds for every $v, w \in L^\infty(\Omega)$.
- A is *positive* if $(v, v)_A \ge 0$ holds for every $v \in L^\infty(\Omega)$.
- A is *positivity-preserving* if for every $v \in L^\infty(\Omega)$ with $v(x) \ge 0$ for almost every $x \in \Omega$, we have that $(vA)(x) \ge 0$ holds for almost every $x \in \Omega$.
- A is *c-regular* if $1_\Omega A = c1_\Omega$ for some $c \in \mathbb{R}$.
- A is a *graphop* if it is positivity-preserving and self-adjoint.
- A is a *Markov graphop* if A is a 1-regular graphop.
- A is *atomless* if Ω is atomless.

Lemma 3.1   *Atomless P-operators are closed with respect to d_M.*

Proof   Let $A \in \mathcal{B}(\Omega)$ be an atomless P-operator and let $B \in \mathcal{B}(\Omega_2)$ with $d_M(B, A) = \varepsilon$. We have that there is a function $v \in L^\infty_{[-1,1]}(\Omega)$ such that the distribution of v is uniform on $[-1, 1]$. Let $\alpha := \mathcal{D}_A(v)$. It follows from $d_H(\mathcal{S}_1(A), \mathcal{S}_1(B)) \le 2\varepsilon$ that there is $\beta = \mathcal{D}_B(w) \in \mathcal{S}_1(B)$ with $d_{LP}(\beta, \alpha) \le 3\varepsilon$ and thus $d_{LP}(\alpha_1, \beta_1) \le 3\varepsilon$, where $\alpha_1 = \mathcal{D}(v)$ and $\beta_1 = \mathcal{D}(w)$ are the marginals of α and β on the first coordinate. It follows that β_1 is at most 3ε far from the uniform distribution in d_{LP}, and thus the largest atom in β_1 is at most 10ε. Hence the largest atom in Ω_2 has weight at most $10\varepsilon = 10 d_M(B, A)$. We obtained that if B is the limit of atomless operators, then B is atomless.   ■

Proposition   *Let $p \in [1, \infty]$ and $q \in (1, \infty)$. Let $\{A_i \in \mathcal{B}(\Omega_i)\}_{i=1}^\infty$ be a sequence of uniformly (p, q)-bounded P-operators converging to a P-operator $A \in \mathcal{B}(\Omega)$. Then, we have the following two statements.*

(1) *If A_i is positive for every i, then A is also positive.*
(2) *If A_i is self-adjoint for every i, then A is also self-adjoint.*

Proof   To prove the first claim let $v \in L^\infty_{[-1,1]}(\Omega)$. We choose $v_i \in L^\infty_{[-1,1]}(\Omega_i)$ according to Remark 2.2, that is, such that $\mathcal{D}_{A_i}(v_i)$ weakly converges to $\mathcal{D}_A(v)$ as i goes to infinity. By Lemma 13.4, we have that $(v_i, v_i)_{A_i} = \mathbb{E}(v_i(A_i v_i))$ converges to $(v, v)_A =$

$\mathbb{E}(v(Av))$. We have by the assumption that $(v_i, v_i)_{A_i} \geq 0$ holds for every i, and thus $(v, v)_A \geq 0$ holds.

To prove the second claim let $v, w \in L^\infty_{[-1,1]}(\Omega)$ and let $\mu := \mathcal{D}_A(v, w)$. Again by Remark 2.2, we have that for every $i \in \mathbb{N}$ there exist functions $v_i, w_i \in L^\infty_{[-1,1]}(\Omega_i)$ such that $\mu_i := \mathcal{D}_{A_i}(v_i, w_i)$ weakly converges to μ. By Lemma 13.4, we obtain that $\mathbb{E}(v_i(w_i A_i))$ converges to $\mathbb{E}(v(wA))$ and $\mathbb{E}((v_i A_i)w_i)$ converges to $\mathbb{E}((vA)w)$ as i goes to infinity. On the other hand, we have that

$$\mathbb{E}(v_i(w_i A_i)) = (w_i, v_i)_{A_i}, \qquad \mathbb{E}(v(wA)) = (w, v)_A,$$

$$\mathbb{E}((v_i A_i)w_i) = (v_i, w_i)_{A_i}, \qquad \mathbb{E}((vA)w) = (v, w)_A.$$

Since by assumption, we have $(v_i, w_i)_{A_i} = (w_i, v_i)_{A_i}$ the proof is complete. ∎

Proposition *Let $p \in [1, \infty)$, $q \in [1, \infty]$, $c \in \mathbb{R}$ and let $\{A_i \in \mathcal{B}(\Omega_i)\}_{i=1}^\infty$ be a sequence of uniformly (p, q)-bounded P-operators converging to a P-operator $A \in \mathcal{B}(\Omega)$. Then, we have the following two statements.*

(1) *If A_i is positivity-preserving for every i, then A is also positivity-preserving.*
(2) *If A_i is c-regular for every i, then A is also c-regular.*

Proof Let $v \in L^\infty_{[0,1]}(\Omega)$. By Remark 2.2, there is a sequence $\{v_i \in L^\infty_{[-1,1]}(\Omega_i)\}_{i=1}^\infty$ such that $\mathcal{D}_{A_i}(v_i)$ weakly converges to $\mathcal{D}_A(v)$. The fact that $\mathcal{D}(v_i)$ weakly converges to the non-negative distribution $\mathcal{D}(v)$ implies that $\mathcal{D}(v_i - |v_i|)$ weakly converges to δ_0. Based on the uniform boundedness of the functions, it follows from Lemma 13.6 that $d_{\mathrm{LP}}(\mathcal{D}_{A_i}(v_i), \mathcal{D}_{A_i}(|v_i|))$ converges to 0 and so (v, vA) is the weak limit of $\mathcal{D}_{A_i}(|v_i|)$. Since $|v_i|A_i$ is non-negative for every i we obtain that vA is non-negative.

Let $\{v_i \in L^\infty_{[-1,1]}(\Omega_i)\}_{i=1}^\infty$ be a sequence of functions such that $\mathcal{D}_{A_i}(v_i)$ weakly converges to $\mathcal{D}_A(1_\Omega)$. We have that $\mathcal{D}(v_i - 1_{\Omega_i})$ weakly converges to δ_0 and hence by Lemma 13.6 we have that $d_{\mathrm{LP}}(\mathcal{D}_{A_i}(1_{\Omega_i}), \mathcal{D}_{A_i}(v_i))$ goes to 0 as i goes to infinity. It follows that $\mathcal{D}_{A_i}(1_{\Omega_i})$ weakly converges to $\mathcal{D}_A(1_\Omega)$. Since $1_{\Omega_i} A_i = c 1_{\Omega_i}$, the proof is complete. ∎

Lemma 3.2 *Let $A \in \mathcal{B}(\Omega)$ be a Markov graphop. Then $\|A\|_{2\to2} = 1$.*

Proof First, we show that $\|A\|_{\infty\to\infty} = 1$. Let $v \in L^\infty_{[0,1]}(\Omega)$. We have that $1_\Omega - v$ is non-negative and thus $(1_\Omega - v)A = 1_\Omega - vA$ is non-negative. It follows that vA is non-negative with $\|vA\|_\infty \leq 1$. Let $v \in L^\infty_{[-1,1]}(\Omega)$. We can write $v = v_1 - v_2$ such that $\|v_1\|_\infty, \|v_2\|_\infty \leq 1$ and both v_1, v_2 are non-negative. We have shown that the values of $v_1 A$ and $v_2 A$ are in $[0, 1]$ and so $vA = v_1 A - v_2 A$ takes values in $[-1, 1]$. It follows that $\|vA\|_\infty \leq 1$. On the other hand, the 1-regular property of A implies that $\|A\|_{\infty\to\infty} = 1$. In general, we have for $v \in L^\infty(\Omega)$ with $m := \|v\|_\infty$ that $v \in L^\infty_{[-m,m]}$ and thus by linearity and the previous statement we obtain that $\|vA\|_\infty \leq m$.

The fact that $1_\Omega A = 1_\Omega$ implies that $\|A\|_{2\to2} \geq 1$. Now, let $v \in L^\infty(\Omega)$ be arbitrary. Then, for every $k \in \mathbb{N}$ we have that $\|vA^k\|_2 \leq \|vA^k\|_\infty \leq \|v\|_\infty$, by the first part of the proof. Now, we use the spectral theorem for the bounded self-adjoint operator A to get a projection-valued measure E representing A. Suppose that this is not supported on $[-1, 1]$ and there exists $w \in L^2(\Omega)$ such that $E_{\mathbb{R}\setminus[-1,1]}w \neq 0$. Then, by continuity, we can find $v \in L^\infty(\Omega)$ such that $E_{\mathbb{R}\setminus[-1,1]}v \neq 0$. Furthermore, by the spectral representation,

we have $\|vA^k\|_2 = \int_{|\lambda|>1} \lambda^{2k} dE_\lambda v$. Since the image of v is not zero, this tends to infinity as $k \to \infty$. This contradicts to the fact that $\|vA^k\|_2$ is uniformly bounded for $v \in L^\infty(\Omega)$. It follows that $\|A\|_{2\to2} \le 1$, and this finishes the proof. ∎

Theorem 3.3 *Let $\mathcal{M}$ be the set of weak equivalence classes of Markov graphops. Then $(\mathcal{M}, d_M)$ is a compact metric space.*

Proof   Let $\{A_i\}_{i=1}^\infty$ be a sequence of Markov graphops with limit A. Lemma 3.2 guarantees that this sequence has uniformly bounded $(2,2)$-norms. Hence we can apply Proposition 3 and Proposition 3, and we get that A is also a Markov graphop. The compactness follows from Lemma 2.6 and Theorem 2.9. ∎

Remark 3.4   For the case of dense graph limits, it is known that every graphon can arise as limits of finite graphs [28]. The answer for the same question for graphops is more complicated. To start with, for the bounded degree case, one of the main open problems related to Benjamini–Schramm graph limits is the question whether every unimodular random graph is the Benjamini–Schramm limit of finite graphs. A positive answer to this question would imply that every group is sofic and it would solve several important open problems in group theory (see, e.g., [15]). Unfortunately, the analogue question for the stronger local-global convergence was refuted by Kun and Thom [24]. This implies that not every graphing is the local-global limit of a finite graph sequence. Since the theory of local-global limits is embedded into the theory of action convergence, we obtain examples for graphops that do not arise as finite graph limits. It is also natural to ask if there is some general invariance property that is common to all finite graph limits, in the same spirit as unimodularity for Benjamini–Schramm limits? This turns out to be true. Unimodularity, in this sense, is equivalent to the property that the representing graphing is a self-adjoint operator. Our Lemma 3 shows that, under mild conditions, action limits of finite graphs are always self-adjoint operators.

4   Construction of the limit object

In this section, we prove Theorem 2.9. Let $\{(\Omega_i, \mathcal{A}_i, \mu_i)\}_{i=1}^\infty$ be a sequence of probability spaces. Assume that $\{A_i \in \mathcal{B}_{p,q}(\Omega_i)\}_{i=1}^\infty$ is a convergent sequence of P-operators such that $\sup_{i\in\mathbb{N}} \|A_i\|_{p\to q} \le c$ for some $c \in \mathbb{R}^+$. For every $k \in \mathbb{N}$, we define

$$X_k := \lim_{i\to\infty} \mathcal{S}_k^*(A_i).$$

We wish to prove that there is a P-operator $A \in \mathcal{B}_{p,q}(\Omega)$ for some probability space $(\Omega, \mathcal{A}, \mu)$ such that for every $k \in \mathbb{N}$ we have that

$$\lim_{i\to\infty} \mathcal{S}_k^*(A_i) = \mathcal{S}_k^*(A).$$

Before the formal definitions, we explain the heuristics behind this proof. Recall that our goal is to construct a P-operator such that its k-profile is the limit of the k-profiles of the operators in a given convergent sequence of operators for every fixed k. Take countable dense sets of points in the limiting k-profiles of the operators for every k. Then we have that each point in this dense system can be approximated by

elements in the k-profiles of the converging operators A_i. On the other hand, every point in the k-profile of A_i involves $2k$ measurable functions on Ω_i. Recall that since Ω_i is a probability space, these functions are random variables. Very roughly speaking, the main idea is to take all the functions involved into an approximating profile point of A_i for our dense system of points in the limiting profiles. These are countably many functions for each i. By choosing a subsequence, we can assume that the joint distributions of these countably many functions (random variables) converge weakly and the limit is some probability distribution on $\Omega := \mathbb{R}^\infty$. As a first attempt, we could try to produce the limiting operator on the function space of this probability space. Observe that each coordinate function in the probability space on $\mathbb{R}^\infty$ corresponds to a function involved in a k-profile for some k. Since every k-profile comes from k functions and their k images, we obtain some information on a possible limiting operator. More precisely, we obtain that certain coordinate functions are the images of some other coordinate functions under this action. However, it is not clear that such an action extends to the full function space on Ω and so we need a more complicated version of the above idea. Instead of just working with functions involved into profile points, we need enough functions to represent the function space of a whole σ-algebra. To this end, we extend the above function systems by new functions that are obtained by some natural operations. To keep track of the new functions, we introduce an abstract algebraic formalism involving semigroups. The main difficulty in the proof is to verify that, at the end of this process, we obtain a well-defined operator and it has the desired limiting properties.

First, we will need the next algebraic notion.

Definition 4.1 (Free semigroup with operators)   Let G and L be sets. We denote by $F(G, L)$ the free semigroup with generator set G and operator set L (freely acting on $F(G, L)$). More precisely, we have that $F(G, L)$ is the smallest set of abstract words satisfying the following properties.

(1) $G \subseteq F(G, L)$.
(2) If $w_1, w_2 \in F(G, L)$, then $w_1 w_2 \in F(G, L)$.
(3) If $w \in F(G, L), l \in L$, then $l(w) \in F(G, L)$.

There is a unique length function $m : F(G, L) \to \mathbb{N}$ such that $m(g) = 1$ for $g \in G$, $m(w_1 w_2) = m(w_1) + m(w_2)$ and $m(l(w)) = m(w) + 1$.

An example for a word in $F(G, L)$ is $l_3(l_1(g_1 g_2 l_2(g_2) g_3)) l_3(g_2) g_1$, where $g_1, g_2, g_3 \in G$ and $l_1, l_2, l_3 \in L$. The length of this word is $(((1 + 1 + 2 + 1) + 1) + 1) + 2 + 1 = 10$. Note that if both G and L are countable sets, then so is $F(G, L)$.

Construction of a function system: In this technical part of the proof, we construct a function system $\{v_{i,f} \in L^\infty(\Omega_i)\}_{i \in \mathbb{N}, f \in F}$ for some countable index set F. Later, we will use this function system to construct a probability distribution $\kappa \in \mathcal{P}(\mathbb{R}^{F \times \{0,1\}})$ and an operator $A \in \mathcal{B}_{p,q}(\mathbb{R}^{F \times \{0,1\}}, \kappa)$. We will show that A is an appropriate limit object for the sequence $\{A_i\}_{i=1}^\infty$.

First, we describe the index set F. For every $k \in \mathbb{N}$, let $X'_k \subseteq X_k$ be a countable dense subset in the metric space (X_k, d_H). Let $G := \bigcup_{k=1}^\infty X'_k \times [k]$. The index set F will be the free semigroup generated by G and an appropriate set of nonlinear operators on function spaces. For $y \in \mathbb{Q}$ and $z \in \mathbb{Q}^+$ let $h_{y,z} : \mathbb{R} \to \mathbb{R}$ be the bounded, continuous

function defined by $h_{y,z}(x) = 0$ if $x \notin (y - z, y + z)$ and $h_{y,z}(x) = 1 - |x - y|/z$ if $x \in (y - z, y + z)$. For every $i \in \mathbb{N}$, $l \in L$ and $v \in L^\infty(\Omega_i)$ we define $l(v) := h_{y,z} \circ (vA_i)$, where l is given by the pair $(y, z) \in \mathbb{Q} \times \mathbb{Q}^+$. Note that by definition $\|l(v)\|_\infty \leq 1$. As these functions are naturally identified with $\mathbb{Q} \times \mathbb{Q}^+$, we will use $L = \mathbb{Q} \times \mathbb{Q}^+$. Furthermore, let $F := F(G, L)$ be as in Definition 4.1. We have that F is countable.

Now we describe the functions $\{v_{i,g}\}_{i \in \mathbb{N}, g \in G}$. For every $i, k \in \mathbb{N}$ and $t \in X'_k$ let $\{v_{i,(t,j)}\}_{j=1}^k$ be a system of functions in $L^\infty_{[-1,1]}(\Omega_i)$ such that the joint distribution of

$$(v_{i,(t,1)}, v_{i,(t,2)}, \ldots, v_{i,(t,k)}, v_{i,(t,1)}A_i, v_{i,(t,2)}A_i, \ldots, v_{i,(t,k)}A_i)$$

converges to t as i goes to infinity.

Now we construct the functions $\{v_{i,w}\}_{i \in \mathbb{N}, w \in F}$ recursively to the length of $m(w)$. For words of length 1 the functions are already constructed above. Assume that for some $k \in \mathbb{N}$ we have already constructed all the functions $v_{i,w}$ with $m(w) \leq k$. Let $w \in F$ such that $m(w) = k + 1$. If $w = w_1 w_2$ for some $w_1, w_2 \in F$, then $v_{i,w} := v_{i,w_1} v_{i,w_2}$. If $w = l(w_1)$, then $v_{i,w} := l(v_{i,w_1})$.

Construction of the probability space. Let $\xi_i : \Omega_i \to \mathbb{R}^{F \times \{0,1\}}$ be the function such that for $f \in F$, $e \in \{0, 1\}$, and $\omega_i \in \Omega_i$ the (f, e) coordinate of $\xi_i(\omega_i)$ is equal to $(v_{i,f}A_i^e)(\omega_i)$, where A_i^0 is defined to be the identity operator. Let $\kappa_i \in \mathcal{P}(\mathbb{R}^{F \times \{0,1\}})$ denote the distribution of the random variable ξ_i. (In other words, κ_i is the joint distribution of the functions $\{v_{i,f}\}_{f \in F}$ and $\{v_{i,f}A_i\}_{f \in F}$.) Since $\tau(\kappa_i) \leq c$ holds (recall equation (2.4) for the definition of τ), we have that there is a strictly growing sequence $\{n_i\}_{i=1}^\infty$ of natural numbers such that κ_{n_i} is weakly convergent with limit κ as i goes to infinity. Let $\Omega := \mathbb{R}^{F \times \{0,1\}}$ be the probability space with the Borel σ-algebra $\mathcal{A}$ and probability measure κ. We will consider Ω as topological space, equipped with the product topology. The fact that κ is a probability measure follows from its definition as a weak limit of probability distributions κ_{n_i}.

Construction of the operator: We will define an operator $A \in \mathcal{B}_{p,q}(\Omega)$ with Ω defined above. For $(f, e) \in F \times \{0, 1\}$ let $\pi_{(f,e)} : \mathbb{R}^{F \times \{0,1\}} \to \mathbb{R}$ denote projection function to the coordinate at (f, e). Notice that

$$(4.1) \qquad \pi_{(f,e)} \circ \xi_i = v_{i,f}A_i^e \qquad (i \in \mathbb{N}, (f, e) \in F \times \{0, 1\}).$$

In particular, due to the definition of κ, we also have $\pi_{(f,0)} \in L^\infty_{[-1,1]}(\Omega)$ for $f \in F$. Our goal is to show that there is a unique (p, q)-bounded linear operator A from $L^\infty(\Omega)$ to $L^1(\Omega)$ with $\|A\|_{p \to q} \leq c$ such that $\pi_{(f,0)}A = \pi_{(f,1)}$ holds for every $f \in F$.

Lemma 4.1 *The coordinate functions on $\mathbb{R}^{F \times \{0,1\}}$ have the following properties.*

(1) *If $f_1, f_2 \in F$, then $\pi_{(f_1 f_2, 0)} = \pi_{(f_1, 0)}\pi_{(f_2, 0)}$ holds in $L^\infty(\Omega)$.*
(2) *If $f \in F$ and $l = (y, z)$ holds for some y, z, then $\pi_{(l(f), 0)} = h_{y,z} \circ \pi_{(f,1)}$ holds in $L^\infty(\Omega)$.*
(3) *If $a_1, a_2, \ldots, a_k \in F$, $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{R}$, then*

$$\left\| \sum_{j=1}^k \lambda_j \pi_{(a_j, 1)} \right\|_q \leq c \left\| \sum_{j=1}^k \lambda_j \pi_{(a_j, 0)} \right\|_p.$$

(4) *The linear span of the functions $\{\pi_{(f,0)}\}_{f \in F}$ is dense in the space $L^p(\Omega)$.*

(5) *Assume that $k \in \mathbb{N}$ and $t \in X'_k$. Then $(t, j) \in G \subset F$ holds for $1 \le j \le k$ and*

$$\mathcal{D}(\pi_{((t,1),0)}, \pi_{((t,2),0)}, \ldots, \pi_{((t,k),0)}, \pi_{((t,1),1)}, \pi_{((t,2),1)}, \ldots, \pi_{((t,k),1)}) = t.$$

Remark 4.2 Note that when functions on Ω are treated as functions in $L^r(\Omega)$ for some $r \in [1, \infty]$, then they are considered to be equal if they differ on a κ zero measure set. This kind of weak equality of functions enables various algebraic correspondences between different coordinate functions, which would be impossible in a strict sense. As a toy example, let us consider the uniform measure μ on $\{(x, x) : x \in [0, 1]\}$, which is a Borel measure on $\mathbb{R}^2$. We have that the x-coordinate function $(x, y) \mapsto x$ is equal to the y-coordinate function $(x, y) \mapsto y$ in the space $L^r(\mathbb{R}^2, \mu)$, because they agree on the support of μ.

The proof of Lemma 4.1 will use the following two lemmas.

Lemma 4.3 *Let $r \in [1, \infty)$. For every $v \in L^r(\Omega)$ we have that*

$$\lim_{n \to \infty} \left\| v - \sum_{j=-n^2}^{n^2} (j/n) h_{j/n, 1/n} \circ v \right\|_r = 0.$$

Proof Let $s_n := \sum_{j=-n^2}^{n^2} (j/n) h_{j/n, 1/n}$. The statement is equivalent to $\lim_{n \to \infty} \|v - s_n \circ v\|_r = 0$. An elementary calculation shows that $s_n(x) = x$ if $x \in [-n, n]$ and $|s_n(x)| \le n$ holds for $x \in K_n := \mathbb{R} \setminus [-n, n]$. Now we have that

$$v - s_n \circ v = 1_{K_n} \circ (v - s_n \circ v) = 1_{K_n} \circ v - 1_{K_n} \circ s_n \circ v,$$

and hence

$$\|v - s_n \circ v\|_r \le \|1_{K_n} \circ v\|_r + n\mathbb{P}(|v| \ge n) = \|1_{K_n} \circ v\|_r + n\mathbb{P}(|v|^r \ge n^r),$$

where v is viewed as a random variable on Ω. For $r = 1$, it is well known that $v \in L^1(\Omega)$ implies $\lim_{n \to \infty} n\mathbb{P}(|v| \ge n) = 0$, due to the fact that we have $0 \le n\mathbb{P}(|v| \le n) \le \mathbb{E}(|v|\mathbb{I}(|v| \le n)) \le \mathbb{E}(|v|) < \infty$ in this case. For $r > 1$, by Markov's inequality we have that $\mathbb{P}(|v|^r \ge n^r)$ is at most $\|v\|_r^r / n^r$ and thus $n\mathbb{P}(|v| \ge n)$ converges to 0 as n goes to infinity. Now it suffices to show that $\lim_{n \to \infty} \|1_{K_n} \circ v\|_r = 0$. Let $U_n := \{x : x \in \Omega, |v(x)| > n\}$. We have that $\|1_{K_n} \circ v\|_r^r = \int_{U_n} |v|^r d\kappa$. The facts that v is measurable and κ is a probability measure imply that $\lim_{n \to \infty} \kappa(U_n) = 0$ and $\lim_{n \to \infty} \int_{U_n} |v|^r d\kappa = 0$. ∎

The following lemma is easy to prove, see, e.g., Theorem 22.4 in the lecture notes [13].

Lemma 4.4 *Let $r \in [1, \infty)$. Let $\{v_i \in L^\infty(\Omega)\}_{i \in I}$ be a system of functions for some countable index set I such that for every $a, b \in I$ there is $c \in I$ with $v_a v_b = v_c$. Let $\mathcal{A}_0$ be the σ-algebra generated by the functions $\{v_i\}_{i \in I}$. Suppose that the constant 1 function on Ω can be approximated by a uniformly bounded family of finite linear combinations of $\{v_i\}_{i \in I}$. Then the L^r-closure of the linear span of $\{v_i \in L^\infty(\Omega)\}_{i \in I}$ is $L^r(\Omega, \mathcal{A}_0, \kappa)$.*

Now we return to the proof of Lemma 4.1.

Proof   To prove the first statement of the lemma recall that, by the construction of the function system, we have for every $i \in \mathbb{N}$ and $f_1, f_2 \in F$ that $v_{i,f_1 f_2} = v_{i,f_1} v_{i,f_2}$ holds. Therefore, by equation (4.1) and the continuity of π, we have that each κ_i is supported inside the closed set

$$\left\{ \omega : \omega \in \mathbb{R}^{F \times \{0,1\}}, \ \pi_{(f_1 f_2, 0)}(\omega) = \pi_{(f_1, 0)}(\omega) \pi_{(f_2, 0)}(\omega) \right\}.$$

This implies that κ is also supported inside this set and thus $\pi_{(f_1 f_2, 0)} = \pi_{(f_1, 0)} \pi_{(f_2, 0)}$ holds κ-almost everywhere.

The proof of the second statement is similar to the first one. Again, by the construction of the function system, we have for every $i \in \mathbb{N}$ and $f \in F$, $l = (p, q) \in L$ that $v_{i, l(f)} = l(v_{i,f}) = h_{p,q} \circ (v_{i,f} A_i)$. This means by the definition of κ_i, equation (4.1) and the continuity of π that κ_i is supported on the closed set

$$\left\{ \omega : \omega \in \mathbb{R}^{F \times \{0,1\}}, \ \pi_{(l(f), 0)}(\omega) = h_{p,q}(\pi_{(f,1)}(\omega)) \right\}$$

for every $i \in \mathbb{N}$. Thus, $\pi_{(l(f), 0)} = h_{p,q} \circ \pi_{(f,1)}$ holds κ-almost everywhere.

For the proof of the third statement, let us use that $\|A_i\|_{p \to q} \leq c$ holds for every $i \in \mathbb{N}$ and hence

$$\left\| \sum_{j=1}^{k} \lambda_j v_{i, a_j} A_i \right\|_q \leq c \left\| \sum_{j=1}^{k} \lambda_j v_{i, a_j} \right\|_p.$$

Since the sum on the right-hand side is a function in $L^\infty(\Omega_i)$ whose values are in the compact interval $[-\lambda, \lambda]$ for $\lambda := \sum_{j=1}^{k} |\lambda_j|$, we have that $\sum_{j=1}^{k} \lambda_j \pi_{(a_j, 0)}$ is a bounded, continuous function on the support of κ. Therefore, using $\kappa_{n_i} \xrightarrow{w} \kappa$ and equation (4.1) again (in particular, integrating of the pth power of the absolute values with respect to κ_i), we obtain that

$$\lim_{i \to \infty} \left\| \sum_{j=1}^{k} \lambda_j v_{n_i, a_j} \right\|_p = \left\| \sum_{j=1}^{k} \lambda_j \pi_{(a_j, 0)} \right\|_p.$$

On the other hand, $\left| \sum_{j=1}^{k} \lambda_j \pi_{(a_j, 1)} \right|^q$ is a nonnegative continuous function, thus weak convergence, in this case, implies the following inequality:

$$\left\| \sum_{j=1}^{k} \lambda_j \pi_{(a_j, 1)} \right\|_q \leq \limsup_{i \to \infty} \left\| \sum_{j=1}^{k} \lambda_j v_{n_i, a_j} A_i \right\|_q.$$

These inequalities together yield the third statement.

For the fourth statement, let $\mathcal{H}_r$ denote the L^r-closure of the linear span of the functions $\{\pi_{(f,0)}\}_{f \in F}$ for $r \in [1, \infty)$. First we show that $\pi_{(f,1)} \in \mathcal{H}_q$ holds for every $f \in F$. By the second statement, we have that

$$(4.2) \qquad \sum_{j=-n^2}^{n^2} (j/n) h_{j/n, 1/n} \circ \pi_{(f,1)} = \sum_{j=-n^2}^{n^2} (j/n) \pi_{(l_j(f), 0)},$$

where l_j is given by the pair $(j/n, 1/n)$ for $-n^2 \leq j \leq n^2$. Since the right-hand side of (4.2) is in $\mathcal{H}_q$, we obtain that the left-hand side is also in $\mathcal{H}_q$. On the other hand, we have $\pi_{(f,1)} \in L^q(\Omega)$ due to the third statement. Hence by Lemma 4.3, we obtain that,

as n goes to infinity, the left-hand side of (4.2) converges to $\pi_{(f,1)}$ in $L^q(\Omega)$ and thus $\pi_{(f,1)} \in \mathcal{H}_q$.

Let $\mathcal{A}_0$ be the σ-algebra generated by the functions $\{\pi_{(f,0)}\}_{f \in F}$. Notice that the constant 1 function on Ω can be approximated already in X_1'. We have by the first statement in this lemma and Lemma 4.4 that $\mathcal{H}_r = L^r(\Omega, \mathcal{A}_0, \kappa)$ holds for every $r \in [1, \infty)$. As we have shown, we have for every $f \in F$ that $\pi_{(f,1)} \in \mathcal{H}_q = L^q(\Omega, \mathcal{A}_0, \kappa)$ and thus all coordinate functions on $\mathbb{R}^{F \times \{0,1\}}$ are measurable in $\mathcal{A}_0$. This shows that $\mathcal{H}_r = L^r(\Omega, \mathcal{A}_0, \kappa) = L^r(\Omega, \mathcal{A}, \kappa) = L^r(\Omega)$ holds for every $r \in [1, \infty)$.

The last statement of the lemma follows directly from the definition of the functions $\{\nu_{i,(t,j)}\}_{i \in \mathbb{N}, j \in [k]}$ and the definition of κ. ∎

We are ready to define the operator $A \in \mathcal{B}_{p,q}(\Omega)$. For $f \in F$, let $\pi_{(f,0)}A = \pi_{(f,1)}$. This defines a linear operator on the linear span of $\{\pi_{(f,0)}\}_{f \in F}$, which is bounded due to the third statement of Lemma 4.1. Hence it has a unique continuous linear extension on its L^p-closure. By the fourth statement of the same lemma, we get that there is a unique operator $A \in \mathcal{B}_{p,q}(\Omega)$ with $\|A\|_{p \to q} \le c$ such that $\pi_{(f,0)}A = \pi_{(f,1)}$ holds for every $f \in F$.

Last part of the proof: The last statement of Lemma 4.1 together with the equality $\pi_{((t,j),0)}A = \pi_{((t,j),1)}$ implies that for every $k \in \mathbb{N}$ and $t \in X_k'$ we have $t \in \mathcal{S}_k(A)$. Therefore, for every $k \in \mathbb{N}$ we have that $X_k \subseteq \mathcal{S}_k^*(A)$. Our goal is to show that $X_k = \mathcal{S}_k^*(A)$ for every $k \in \mathbb{N}$ and thus it remains to prove that $\mathcal{S}_k^*(A) \subseteq X_k$.

Let $k \in \mathbb{N}$ and let $v_1, v_2, \ldots, v_k \in L^\infty_{[-1,1]}(\Omega)$. We need to show that $\alpha := \mathcal{D}_A(\{v_j\}_{j=1}^k)$ is in X_k. Let $\varepsilon > 0$ be arbitrary. We have by the fourth statement of Lemma 4.1 that for some large enough natural number m there are elements $f_1, f_2, \ldots, f_m \in F$ and real numbers $\{\lambda_{a,j}\}_{a \in [m], j \in [k]}$ such that for every $j \in [k]$ we have $\|w_j - v_j\|_p \le \varepsilon$, where $w_j := \sum_{a=1}^m \lambda_{a,j} \pi_{(f_a,0)}$ for $j \in [k]$.

Since only vectors with infinity norm at most 1 are considered in the profile, we will use a truncating function. Namely, let $\tilde{h} : \mathbb{R} \to \mathbb{R}$ be the continuous function with $\tilde{h}(x) = x$ for $x \in [-1, 1]$, $\tilde{h}(x) = -1$ for $x \in (-\infty, -1]$ and $\tilde{h}(x) = 1$ for $x \in [1, \infty)$. Since $\|v_j\|_\infty \le 1$ holds for $j \in [k]$, we have that $|w_j(x) - v_j(x)| \ge |\tilde{h} \circ w_j(x) - v_j(x)|$ holds almost everywhere and thus by $\|w_j - v_j\|_p \le \varepsilon$, we obtain $\|\tilde{h} \circ w_j - v_j\|_p \le \varepsilon$ for $j \in [k]$. This also implies by the triangle inequality that

$$(4.3) \qquad \|\tilde{h} \circ w_j - w_j\|_p \le \|\tilde{h} \circ w_j - v_j\|_p + \|v_j - w_j\|_p \le 2\varepsilon$$

holds for $j \in [k]$.

For $i \in \mathbb{N}$ and $j \in [k]$ let $z_{i,j} := \sum_{a=1}^m \lambda_{a,j} \nu_{i,f_a}$. Let $\beta_i := \mathcal{D}_{A_i}(\{z_{i,j}\}_{j=1}^k)$. By the definition of κ, we have that

$$\beta := \lim_{i \to \infty} \beta_{n_i} = \mathcal{D}_A(\{w_j\}_{j=1}^k)$$

holds in d_{LP}. Since $w_j \in L^\infty(\Omega)$, we have $\|v_j A - w_j A\|_1 \le \|v_j A - w_j A\|_q \le c\|v_j - w_j\|_p \le c\varepsilon$. We get from Lemma 13.2 that $d_{\mathrm{LP}}(\alpha, \beta) \le (2k)^{3/4}(c'\varepsilon)^{1/2}$, where $c' := \max(c, 1)$.

Let

$$\beta_i' := \mathcal{D}_{A_i}\left(\{\tilde{h} \circ z_{i,j}\}_{j=1}^k\right).$$

Notice that the functions $\tilde{h} \circ z_{i,j} - z_{i,j}$ are uniformly bounded and their distribution converges weakly to the distribution of $\tilde{h} \circ w_j - w_j$. Hence if i is large enough, then by (4.3) we have that $\|\tilde{h} \circ z_{i,j} - z_{i,j}\|_p \leq 3\varepsilon$ holds for $j \in [k]$ and thus $d_{\mathrm{LP}}(\beta_i', \beta_i) \leq (2k)^{3/4}(3c'\varepsilon)^{1/2}$ by Lemma 13.2.

Let $\{n_i'\}_{i=1}^\infty$ be a subsequence of $\{n_i\}_{i=1}^\infty$ such that $\beta' := \lim_{i\to\infty} \beta_{n_i'}'$ exists. Observe that $\beta' \in X_k$ and $d_{\mathrm{LP}}(\beta', \beta) \leq (2k)^{3/4}(3c'\varepsilon)^{1/2}$. We obtain that

$$d_{\mathrm{LP}}(X_k, \alpha) \leq d_{\mathrm{LP}}(\beta', \alpha) \leq d_{\mathrm{LP}}(\beta', \beta) + d_{\mathrm{LP}}(\beta, \alpha) \leq 3(2k)^{3/4}(c'\varepsilon)^{1/2}.$$

This holds for every $\varepsilon > 0$ and thus $\alpha \in X_k$. □

5 General graph limits

There are various ways of representing graphs by operators and, in particular, by P-operators. Depending on the chosen representation, we get a corresponding limit notion for graphs. In this section, we list four natural operator representations of graphs and investigate the corresponding graph limit notions. Let G be a finite graph on the vertex set $V(G) = [n]$ with edge set $E(G)$.

Adjacency operator convergence: Recall that $\mu_{[n]}$ denotes the uniform distribution on $[n]$. We denote by $A(G) \in \mathcal{B}([n], \mu_{[n]})$ the P-operator defined by

$$(vA(G))(i) := \sum_{(j,i)\in E(G)} v(j)$$

for $i \in [n]$. We have that $d \geq \|A(G)\|_{2\to 2} \geq d^{1/2}$, where d is the maximal degree in G. We can say that a graph sequence $\{G_i\}_{i=1}^\infty$ is convergent if $\{A(G_i)\}_{i=1}^\infty$ is an action convergent sequence of P-operators. We obtain compactness for graphs with uniformly bounded degree. Quite surprisingly (and nontrivially), it turns out that this convergence notion is equivalent to local-global convergence, which is a refinement of Benjamini–Schramm convergence (see [4, 21]). However, for graphs with nonbounded degrees, compactness is not guaranteed. This can be solved by scaling the operators $A(G)$ by some number that depends on G. For example, we have that $\|A(G)/|V(G)|\|_{2\to 2} \leq 1$ holds for every graph G. Again, quite surprisingly, it turns out that convergence of $A(G_i)/|V(G_i)|$ is equivalent to dense graph convergence. (For a definition of dense graph convergence, see [28].) This motivates us to introduce scaling functions that map graphs to positive real numbers. Let $\mathcal{G}$ denote the set of isomorphism classes of finite graphs.

Definition 5.1 Let $f : \mathcal{G} \to \mathbb{R}^+$ be a function. We say that a graph sequence $\{G_i\}_{i=1}^\infty$ is adjacency operator convergent (or just convergent) with scaling f if the sequence $\{A(G_i)/f(G_i)\}_{i=1}^\infty$ is an action convergent sequence of P-operators.

Recall that a *graphop* is a self-adjoint, positivity-preserving P-operator. The word graphop is a mixture of the words, graph, and operator. Note that both graphons and graphings used in graph limit theory are graphops. Theorem 2.9, combined with Proposition 3 and Proposition 3, implies the following.

Theorem 5.1 *Let $\{G_i\}_{i=1}^{\infty}$ be an adjacency operator convergent sequence of graphs with scaling $f : \mathcal{G} \to \mathbb{R}^+$. Assume that there exist $p \in [1, \infty)$, $q \in (1, \infty)$ and $c \in \mathbb{R}^+$ such that $\|G_i/f(G_i)\|_{p \to q} \leq c$ holds for every $i \in \mathbb{N}$. Then there is a graphop A such that $\lim_{i \to \infty} A(G_i)/f(G_i) = A$ holds. We say that A is the adjacency operator limit of $(\{G_i\}_{i=1}^{\infty}, f)$.*

A natural scaling is $f_{p,q}(G) := \|A(G)\|_{p \to q}$ defined for nonempty graphs G ($f_{p,q}$ can be defined as 1 on the empty graph), where $p \in [1, \infty)$, $q \in (1, \infty]$. Let us call it *norm scaling*. With this scaling, every graph sequence has a convergent subsequence, hence we have sequential compactness for arbitrary graph sequences. Norm scaling leads to a general convergence notion that generalizes local-global convergence and recovers dense graph limits up to a constant multiplicative factor in the limit object. The norm scaling is very convenient to use for general graph sequences where no other natural normalization is given.

Random walk convergence of graphs: Let ν_G denote the stationary measure of the random walk on G. It is well known that ν_G is the probability measure on $[n]$ with $\nu_G(i) = d_i/(2|E(G)|)$, where d_i is the degree of the vertex i for $i \in [n]$. We denote by $M(G) \in \mathcal{B}_{2,2}([n], \nu_G)$ the P-operator defined by equation (1.1). The operator $M(G)$ is known as the Markov kernel corresponding to the random walk on G. We have that $M(G)$ is a Markov graphop. Consequently, by Lemma 3.2, we have that $\|M(G)\|_{2 \to 2} = 1$. (If G has no edges, then $M(G)$ is not defined.)

Definition 5.2 *A graph sequence of nonempty graphs $\{G_i\}$ is called random walk convergent if $\{M(G_i)\}_{i=1}^{\infty}$ is a convergent sequence of P-operators.*

The following theorem is a direct consequence of Theorem 3.3.

Theorem 5.2 *Every graph sequence $\{G_i\}_{i=1}^{\infty}$ has a random walk convergent subsequence. If $\{G_i\}_{i=1}^{\infty}$ is random walk convergent, then there is a Markov graphop A such that $\lim_{i \to \infty} M(G_i) = A$. We say that A is the random walk limit of $\{G_i\}_{i=1}^{\infty}$.*

Note that, for regular graphs, random walk convergence is equivalent to adjacency operator convergence with scaling by $\|G\|_{2 \to 2}$. However, if there is a small but nonzero number of very high degree points in G and many low degree points, then adjacency operator convergence may trivialize. Examples for this are the star graphs or the 2-subdivisions of complete graphs. In these cases random walk convergence turns out to be more natural and leads to interesting and nontrivial limit objects (see Section 12).

Extended random walk convergence: A Markov pair is a pair of a Markov graphop $A \in \mathcal{B}(\Omega)$ and a measurable function f on Ω. As we explained in the introduction, a sequence of Markov pairs $\{(A_i, f_i)\}_{i=1}^{\infty}$ is convergent if the extended k-profiles formed by distributions of the form

$$\mathcal{D}_{A_i, f_i}(\{v_j\}_{j=1}^k) := \mathcal{D}(\{v_j\}_{j=1}^k, \{v_j A_i\}_{j=1}^k, f_i)$$

converge in d_H for every $k \in \mathbb{N}$. It can be proved with a slight extension of the proof of Theorem 2.9 (details will be worked out elsewhere) that a convergent sequence of Markov pairs has a limit, which is also a Markov pair. There are two different uses of Markov pairs. The first one is the following. In spite of the fact that the Markov

kernel of a finite graph $M(G)$ determines the sequence of the nonzero degrees (even with multiplicities), this information may be lost in the limit. Even if it is preserved in some way (examples for this are sequences of bounded degree graphs), degrees cannot be read off in the usual way from the limit object which is a Markov graphop. The idea is that we can store the information on the degrees in a normalized version d^* of the degree function d.

Remark 5.3 (Representing graphops by Markov pairs)   In general, every graphop $A \in \mathcal{B}(\Omega, \mathcal{A}, \mu)$ can be naturally represented by a Markov pair in the following way. Let v be the representing measure of A given by Theorem 6.3 and let v' be the marginal measure of v on $(\Omega, \mathcal{A})$. Let $M(A)$ denote the Markov graphop on $(\Omega, \mathcal{A}, v')$ determined by the measure v using again Theorem 6.3. Note that the action of $M(A)$ is given by

$$(vM(A))(x) = (vA)(x)/(1_\Omega A)(x).$$

The representation of A is given by the pair $(M(A), 1_\Omega A)$.

Another interesting use of Markov pairs is that we can use them to represent generalized graphons (related to graphexes [6, 23]) that are symmetric non-negative measurable functions of the form $W : \mathbb{R}^+ \times \mathbb{R}^+ \to [0,1]$ with $\|W\|_1 < \infty$. Let v denote the probability measure on $\mathbb{R}^+ \times \mathbb{R}^+$ defined by $v(S) = \|W\|_1^{-1} \int_S W \, d\lambda^2$. Let v' be the marginal distribution of v on $\mathbb{R}^+$. Let M denote the Markov graphop on $(\mathbb{R}^+, v')$ determined by Theorem 6.3. For $x \in \mathbb{R}^+$ let $f(x) := \int_{\mathbb{R}^+} W(x, y) \, d\lambda$. We can represent the generalized graphon W by the Markov pair (M, f).

Laplace operator convergence: Using the above notations, we denote by $L(G) \in \mathcal{B}_2([n], \mu_{[n]})$ the P-operator defined by

$$(vL(G))(i) := d(i)v(i) - \sum_{(j,i) \in E(G)} v(j).$$

We have that $L(G)$ is a positive self-adjoint operator. Note that, in contrast with $A(G)$ and $M(G)$, the operator $L(G)$ is typically not positivity-preserving. On the other hand, we gain positiveness. Similarly, to the previous definitions, we say that a graph sequence $\{G_i\}_{i=1}^\infty$ is Laplace operator convergent (or just convergent) with scaling f if the sequence $\{L(G_i)/f(G_i)\}_{i=1}^\infty$ has uniformly bounded operator norm and is a convergent sequence of P-operators. Limit objects are positive, self-adjoint P-operators.

Degree weighted operator convergence: Finally, we mention one more interesting P-operator related to G. Similarly, to $M(G)$, we use the stationary distribution of the random walk. Let $F(G) \in \mathcal{B}([n], v_G)$ be defined by

$$(vF(G))(i) := \sum_{(j,i) \in E(G)} v(j)d(j).$$

Again, we have that $F(G)$ is a graphop. Indeed, positivity-preserving property is clear, and self-adjointness can be verified by a simple calculation. Limits of appropriately normalized versions of $F(G_i)$ are graphops.

6 Measure representation of graphops

Let $A \in \mathcal{B}(\Omega, \mathcal{A}, \mu)$ be a graphop. In this section, we construct a measure ν on $\Omega \times \Omega$ that represents the operator A. This means that the operator A can be reconstructed from the measure ν in a natural way. Intuitively, if we think of A as an infinite graph-like object, then ν shows where we can find the edges in $\Omega \times \Omega$. Note that both graphons and graphings are given in terms of such measures rather than in the form of operators. More precisely, graphons are given by a measurable function which is the Radon–Nykodim derivative of a measure on $[0,1] \times [0,1]$. Our goal, in this chapter, is to bring closer the operator language and the existing representations of graph limits.

Assume that $(\Omega, \mathcal{A})$ is a standard Borel space. Let $\mathcal{R}$ denote the set of product sets of the form $S \times T \subseteq \Omega \times \Omega$, where $S, T \in \mathcal{A}$. We have that $\mathcal{R}$ is a so-called semiring. We define the function ν on $\mathcal{R}$ such that $\nu(S \times T) := (1_S, 1_T)_A$ holds for $S, T \in \mathcal{A}$ (recall equation (2.5)).

Lemma 6.1 *The function ν has the following properties.*

(1) *$\nu(S \times T) \geq 0$ for every $S, T \in \mathcal{A}$.*
(2) *If $S \times T$ is the disjoint union of the finitely many product sets $\{S_i \times T_i\}_{i=1}^n$, then $\nu(S \times T) = \sum_{i=1}^n \nu(S_i \times T_i)$.*
(3) *For every $\varepsilon > 0$ there is $\delta = \delta(\varepsilon) > 0$ such that $\nu(S \times T) \leq \varepsilon$ whenever the minimum of $\mu(S)$ and $\mu(T)$ is at most δ.*

Proof By the positivity-preserving property of A and the bilinearity of $(.,.)_A$ we have that ν satisfies the first two properties. To show the last property, observe that, by the self-adjoint property A, we have that $\nu(S \times T) = \nu(T \times S)$ and so the statement is equivalent to showing the existence of $\delta > 0$ such that $\nu(S \times T) \leq \varepsilon$ whenever $\nu(T) \leq \delta$. We have by the first two properties that

$$\nu(S \times T) \leq \nu(S \times T) + \nu((\Omega \backslash S) \times T) = \nu(\Omega \times T) = \int_T f \, d\mu,$$

where $f := 1_\Omega A$. Now, since $f \geq 0$ and $\int f \, d\mu < \infty$, the statement of the lemma follows from the well-known absolute continuity property of integration. ∎

The proof of the following lemma follows a similar scheme as [12, Lemma A.10] or [17, Theorem 454D].

Lemma 6.2 *The function ν is a premeasure on $\mathcal{R}$ and it has a unique extension to a Borel measure on $(\Omega \times \Omega, \mathcal{A} \otimes \mathcal{A})$, denoted also by ν.*

Proof First, we claim that if $\varepsilon > 0$ and $\delta > 0$ satisfy the third property in lemma 6.1 and $\mu(S), \mu(T) \geq 1 - \delta$, then

(6.1) $$\nu(S \times T) \geq \nu(\Omega \times \Omega) - 2\varepsilon.$$

Indeed, by Lemma 6.1, we have

$$\nu(S \times T) = \nu(\Omega \times \Omega) - \nu((\Omega \backslash S) \times T) - \nu(S \times (\Omega \backslash T)) + \nu((\Omega \backslash S) \times (\Omega \backslash T))$$

$$\geq \nu(\Omega \times \Omega) - \nu((\Omega \backslash S) \times \Omega) - \nu(\Omega \times (\Omega \backslash T)) \geq \nu(\Omega \times \Omega) - 2\varepsilon.$$

To prove that v is a premeasure, the only nontrivial part is to show that if $R \in \mathcal{R}$ is the pairwise disjoint union of sets $\{R_i = S_i \times T_i\}_{i=1}^{\infty}$ in $\mathcal{R}$, then $v(R) \le \sum_{i=1}^{\infty} v(R_i)$. Note that it suffices to prove it for $R = \Omega \times \Omega$, since given any other product set R we can obtain $\Omega \times \Omega$ as the disjoint union of R and a finite number of other product sets, in such a way that the claim for $\Omega \times \Omega$ implies the claim for R. Now, since Ω is standard, there exists a topology τ on Ω generating $\mathcal{A}$ such that (1) we can approximate every set in $\mathcal{A}$ by a compact set with arbitrary precision measured in μ; (2) both S_i and T_i are open for every $i \in \mathbb{N}$. Let $\varepsilon > 0$ arbitrary and let $K_1, K_2 \in \mathcal{A}$ be τ-compact sets such that $\mu(K_1), \mu(K_2) \ge 1 - \delta(\varepsilon)$. We have by equation (6.1) that $v(K_1 \times K_2) \ge v(\Omega \times \Omega) - 2\varepsilon$. Since every S_i and T_i is open, we have that $\{S_i \times T_i\}_{i=1}^{\infty}$ is an open cover of the compact set $K_1 \times K_2$, and so there is a finite subcover. Applying the second property of v to this finite subcover, we obtain that

$$v(\Omega \times \Omega) \le v(K_1 \times K_2) + 2\varepsilon \le 2\varepsilon + \sum_{i=1}^{\infty} v(R_i).$$

Since this is true for every $\varepsilon > 0$, we obtain that v is a premeasure. The Carathéodory extension theorem implies that there is a unique extension of v to $\mathcal{A} \otimes \mathcal{A}$. ∎

Theorem 6.3 (Measure representation of graphops)   *If $A \in \mathcal{B}(\Omega, \mathcal{A}, \mu)$ is a graphop, then there is a unique finite measure v on $(\Omega \times \Omega, \mathcal{A} \otimes \mathcal{A})$ with the following properties.*

(1) *v is symmetric, i.e., $v(S \times T) = v(T \times S)$ holds for every $S, T \in \mathcal{A}$.*
(2) *The marginal distribution of v on Ω is absolutely continuous with respect to μ.*
(3) *$(f, g)_A = \int_{(x,y) \in \Omega^2} f(x)g(y) \, dv$ holds for every $f, g \in L^{\infty}(\Omega)$.*

Conversely, if v is a finite measure on $(\Omega \times \Omega, \mathcal{A} \otimes \mathcal{A})$ satisfying the first two properties, then there is a unique graphop A such that the third property is satisfied.

Proof   The existence and uniqueness of v follows from Lemma 6.2. For the converse statement, let $f \in L^{\infty}(\Omega)$ be an arbitrary function. Let F_f denote the $L^{\infty}(\Omega) \to \mathbb{R}$ functional defined by $F_f(g) := \int_{(x,y) \in \Omega^2} f(x)g(y) \, dv$. We have that $|F_f(g)| \le \|f\|_{\infty} \|g\|_{\infty} v(\Omega^2)$, and thus, by duality, there is a unique function $m(f) \in L^1(\Omega)$ such that $F_f(g) = \int m(f)(x)g(x) \, d\mu$. It is clear from the definition that $A : f \mapsto m(f)$ is a self-adjoint, positivity-preserving linear operator with $\|A\|_{\infty \to 1} = v(\Omega^2)$, satisfying the third property. Notice that for this converse statement we assumed that v is finite, which guarantees that we can apply the extension theorem. ∎

Remark 6.4 (Fiber measures)   A natural way of reconstructing A from the representing measure v goes by disintegrating the measure v. By using the disintegration theorem one obtains a family of measures $\{v_x\}_{x \in \Omega}$ on $(\Omega, \mathcal{A})$ (called fiber measures) such that

$$(fA)(x) = \int_{\Omega} f \, dv_x.$$

In general, it is very convenient to describe a graphop in terms of fiber measures. This is illustrated in Figure 4.

Remark 6.5 (Markov graphops as couplings)   Markov graphops are special graphops such that $1_{\Omega} A = 1_{\Omega}$. It follows that the marginal distribution of v on Ω is equal to
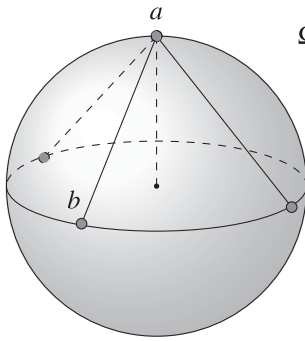
$$\Omega = \{(x, y, z) : x^2 + y^2 + z^2 = 1\}, \ \mu = uniform\ measure$$

$$(a, b) \in E \Leftrightarrow a \perp b \qquad (\Omega, E)\ is\ a\ Borel\ graph$$

$$\nu_a := uniform\ measure\ on\ \{b : b \in \Omega, a \perp b\}$$

$$S : L^2(\Omega, \mu) \to L^2(\Omega, \mu)$$

$$(fS)(a) := \int_{b \perp a} f(b)\, d\nu_a$$

Figure 4: *The spherical graphop.* It is neither a graphon nor a graphing: it is somewhere half way in between.

μ. This means that Markov graphops are completely specified by the data $(\Omega, \mathcal{A}, \nu)$, where ν is a symmetric probability measure on $(\Omega \times \Omega, \mathcal{A} \otimes \mathcal{A})$. Such objects are symmetric self-couplings of probability spaces.

7 Quotient convergence and partitions

In the first part of this section, we relate P-operator convergence to the so-called quotient convergence, which was studied in different forms by different authors [8, 11, 26]. The version that we generalize to P-operators was defined in [26]. In the second part of the chapter, we describe a variant of action convergence that turns out to be equivalent to the original version for uniformly (p, q)-bounded sequences.

Definition 7.1 A function partition of $(\Omega, \mathcal{A}, \mu)$ is a set $\{v_i\}_{i=1}^k$ of $0 - 1$ valued measurable functions on Ω such that $\sum_{i=1}^k v_i = 1_\Omega$. A fractional function partition is a set $\{v_i\}_{i=1}^k$ of functions in $L_{[0,1]}^\infty(\Omega)$ such that $\sum_{i=1}^k v_i = 1_\Omega$. We say that $\{v_i\}_{i=1}^k$ is balanced if $\|v_i\|_1 = 1/k$ holds for every $i \in [k]$.

Definition 7.2 (Quotients of P-operators) Let $k \in \mathbb{N}$ and $A \in \mathcal{B}(\Omega)$. A balanced fractional $k \times k$ quotient of A is a matrix $M \in \mathbb{R}^{k \times k}$ such that there is a balanced fractional function partition $\{v_i\}_{i=1}^k$ of Ω with $M_{i,j} = (v_i, v_j)_A$ for every $i, j \in [k]$. Let $\mathcal{Q}_k(A)$ denote the set of all balanced fractional $k \times k$ quotients of A.

Note that by linearity, the entry sum of any matrix $M \in \mathcal{Q}_k(A)$ is equal to $(1_\Omega, 1_\Omega)_A$ for every $k \in \mathbb{N}$. For two matrices $A, B \in \mathbb{R}^{[k] \times [k]}$ let $d_1(A, B) := \sum_{i,j} |A_{i,j} - B_{i,j}|$ denote the entry-wise l_1 distance. For two subsets $S_1, S_2 \subseteq \mathbb{R}^{[k] \times [k]}$ let $d_{1,H}$ denote the corresponding Hausdorff distance (recall Definition 2.4).

Definition 7.3 (Quotient convergence and metric) A sequence of P-operators $\{A_i\}_{i=1}^\infty$ is quotient convergent if for every k we have that $\mathcal{Q}_k(A_i)$ is convergent in $d_{1,H}$.

The following proposition says that P-operator convergence is stronger than quotient convergence if the sequence has uniformly bounded $\|.\|_{p \to q}$ norm for some $p \in [1, \infty), q \in (1, \infty]$.

Lemma 7.1 *Let us fix $c \geq 1$ and numbers $p \in [1, \infty), q \in (1, \infty]$. For every $k \in \mathbb{N}$ and $\varepsilon > 0$ there exists $\delta = \delta(\varepsilon, k) > 0$ such that whenever two P-operators $A \in \mathcal{B}(\Omega_1), B \in \mathcal{B}(\Omega_2)$ with $\|A\|_{p \to q}, \|B\|_{p \to q} \leq c$ satisfy $d_M(A, B) \leq \delta$, then $d_{1,H}(\mathcal{Q}_k(A), \mathcal{Q}_k(B)) \leq \varepsilon$.*

Proof   Depending on p and q let us choose $p' \in (p, \infty)$ and $q' \in (1, q)$ with $1/p' + 1/q' = 1$. Then we have $\|A\|_{p' \to q'} \leq \|A\|_{p \to q}$. Let $M \in \mathcal{Q}_k(B)$. We need to show that if A and B are sufficiently close in d_M, then there is $M' \in \mathcal{Q}_k(A)$ with $d_1(M, M') \leq \varepsilon$. The other direction (when $M \in \mathcal{Q}_k(A)$) follows by the symmetry of the argument.

Let $v_1, v_2, \ldots, v_k$ be a balanced fractional function partition of Ω_2 such that the corresponding quotient of B is M. We have by the definition of d_M that there are vectors $\{w_i\}_{i=1}^k$ in $L^\infty_{[-1,1]}(\Omega_1)$ such that

$$d_{\mathrm{LP}}(\mathcal{D}_B(\{v_i\}_{i=1}^k)), \mathcal{D}_A(\{w_i\}_{i=1}^k)) \leq 2^{k+1} d_M(A, B).$$

Let $\varepsilon_2 > 0$ be an arbitrary constant, and suppose that $d_M(A, B)$ is so small that the right-hand side of the above inequality is smaller than $\varepsilon_2/2$. The fact that $\{v_i\}_{i=1}^k$ is a balanced fractional function partition can be expressed by the joint empirical distribution of these vectors: it is concentrated on k-tuples of nonnegative numbers with sum 1, and the mean of every marginal is $1/k$. By the inequality above, the joint empirical distribution of $\{w_i\}_{i=1}^k$ is close to this, and it follows that there is a balanced fractional function partition $\{w'_i\}_{i=1}^k$ on Ω_1 such that $\|w_i - w'_i\|_{p'} \leq \varepsilon_2$ holds for every $i \in [k]$. For such a function system, we have for every $i, j \in [k]$ that

$$|(w_i, w_j)_A - (w'_i, w'_j)_A| \leq |(w_i, w_j)_A - (w'_i, w_j)_A| + |(w'_i, w_j)_A - (w'_i, w'_j)_A|$$
$$= |(w_i - w'_i, w_j)_A| + |(w'_i, w_j - w'_j)_A| \leq 2\varepsilon_2 c,$$

where the last inequality is by (2.6) and $\|w_i\|_\infty, \|w'_i\|_\infty \leq 1$. Let $M' \in \mathcal{Q}_k(A)$ be the quotient matrix of A corresponding to $\{w'_i\}_{i=1}^k$ and let $M'\prime$ be defined by $M'\prime_{i,j} := (w_i, w_j)_A$. We have that $d_1(M', M'\prime) \leq 2\varepsilon_2 k^2 c \leq \varepsilon/2$ for appropriate ε_2. It remains to show that $|(v_i, v_j)_A - (w_i, w_j)_A|$ is small for every $i, j \in [k]$ if $d_M(A, B)$ is sufficiently small. This follows from Lemma 13.4.   ∎

The following proposition is a direct consequence of the previous lemma.

Proposition (Action convergence $\Rightarrow$ quotient convergence)   *Let $p \in [1, \infty), q \in (1, \infty]$ and let $\{A_i\}_{i=1}^\infty$ be an action convergent sequence of P-operators with uniformly bounded $\|.\|_{p \to q}$ norms. Then $\{A_i\}_{i=1}^\infty$ is quotient convergent.*

In the rest of this section, we will formulate a version of action convergence. It is clear that $\{v_i\}_{i=1}^k$ is a function partition if and only if there is measurable partition $\mathcal{P} = \{P_1, P_2, \ldots, P_k\}$ of Ω such that $v_i = 1_{P_i}$. Let M_k denote the set of probability measures μ on $\mathbb{R}^{2k}$ such that μ is concentrated on $\{e_i\}_{i=1}^k \times \mathbb{R}^k$, where $e_i \in \mathbb{R}^k$ is the vector with 1 at the ith coordinate and 0 everywhere else. Let $A \in \mathcal{B}(\Omega, \mathcal{A}, \mu)$. We have that $\{v_i\}_{i=1}^k$ is a function partition if and only if $\mathcal{D}_A(\{v_i\}_{i=1}^k) \in M_k$. Let $\mathcal{S}'_k(A) := \mathcal{S}_k(A) \cap M_k$. In other words, $\mathcal{S}'_k(A)$ is the set of all probability measures $\mathcal{D}_A(\{v_i\}_{i=1}^k)$, where $\{v_i\}_{i=1}^\infty$ is a function partition. The next theorem gives a useful equivalent formulation of P-operator convergence for uniformly bounded operators.

Theorem 7.3 *Let $p \in [1, \infty)$, $q \in (1, \infty]$ and let $\{A_i\}_{i=1}^{\infty}$ be a uniformly (p, q)-bounded sequence of P-operators. Then $\{A_i\}_{i=1}^{\infty}$ is convergent if and only if for every k the sequence $\mathcal{S}'_k(A_i)$ is convergent with respect to the Hausdorff distance d_H.*

The proof of this theorem is a direct consequence of the following two lemmas.

Lemma 7.4 *Let $p \in [1, \infty)$, $q \in (1, \infty]$, $c \in \mathbb{R}^+$ be fixed. Then for each $\varepsilon > 0$, $k \in \mathbb{N}$ there is $\delta = \delta(\varepsilon, k) > 0$ such that if A is a P-operator with $\|A\|_{p \to q} \le c$ and $\mu \in \mathcal{S}_k(A)$ satisfies $d_{\mathrm{LP}}(\mu, M_k) \le \delta$, then there is $\mu_2 \in \mathcal{S}'_k(A)$ with $d_{\mathrm{LP}}(\mu, \mu_2) \le \varepsilon$. Furthermore, if for two P-operators A, B with $\|A\|_{p \to q}, \|B\|_{p \to q} \le c$ we have $d_H(\mathcal{S}_k(A), \mathcal{S}_k(B)) \le \delta/2$, then*

$$d_H(\mathcal{S}'_k(A), \mathcal{S}'_k(B)) \le \varepsilon + 2d_H(\mathcal{S}_k(A), \mathcal{S}_k(B)).$$

Proof Assume that $A \in \mathcal{B}(\Omega)$. If μ satisfies the conditions of the first part of the lemma, it can be written as $\mu = \mathcal{D}_A(\{v_i\}_{i=1}^k)$. On the other hand, since $d_{\mathrm{LP}}(\mu, M_k) \le \delta$, we can find a probability measure in M_k with distance at most δ from μ. The first k marginals are concentrated on vectors with entries in $[0, 1]$, hence there are $[0, 1]$-valued functions $w_1, w_2, \ldots, w_k$ in $L^p(\Omega)$ such that $\sum_{i=1}^k w_i = 1_\Omega$ and $\max_{i \in [k]} \|v_i - w_i\|_p \le 2\delta$. It follows that $\max_{i \in [k]} \|v_i A - w_i A\|_q \le 2c\delta$, by the upper bound on the matrix norm. By using Lemma 13.1, it follows that if δ is small enough, then $\mu_2 := \mathcal{D}_A(\{w_i\}_{i=1}^k)$ is in $\mathcal{S}'_k(A)$ and $d_{\mathrm{LP}}(\mu, \mu_2) \le \varepsilon$.

To see the second claim, let $\mu \in \mathcal{S}'_k(A)$. Then there is some $\mu_2 \in \mathcal{S}_k(B)$ with $d_{\mathrm{LP}}(\mu, \mu_2) \le 2d_H(\mathcal{S}_k(A), \mathcal{S}_k(B)) \le \delta$. By definition, we know that $\mu \in \mathcal{S}'_k(A) \subseteq M_k$, hence $d_{\mathrm{LP}}(M_k, \mu_2) \le \delta$. We can apply the first part of the statement for μ_2 and B. We obtain that there is $\mu_3 \in \mathcal{S}'_k(B)$ with $d_{\mathrm{LP}}(\mu_2, \mu_3) \le \varepsilon$. We get that $d_{\mathrm{LP}}(\mu, \mu_3) \le \varepsilon + 2d_H(\mathcal{S}_k(A), \mathcal{S}_k(B))$. ∎

Lemma 7.5 *Let $p \in [1, \infty)$, $q \in (1, \infty]$, $c \in \mathbb{R}^+$ be fixed. For every $\varepsilon > 0$, $k \in \mathbb{N}$ there is $\delta > 0$ and $k' \in \mathbb{N}$ such that if $A \in \mathcal{B}(\Omega_1)$ and $B \in \mathcal{B}(\Omega_2)$ are P-operators with $\|A\|_{p \to q}, \|B\|_{p \to q} \le c$ and $d_H(\mathcal{S}'_{k'}(A), \mathcal{S}'_{k'}(B)) \le \delta$, then $d_H(\mathcal{S}_k(A), \mathcal{S}_k(B)) \le \varepsilon$.*

Proof Let $\mu \in \mathcal{S}_k(A)$ be given by $\mu = \mathcal{D}_A(\{v_i\}_{i=1}^k)$, where $v_i \in L^\infty_{[-1,1]}(\Omega_1)$ holds for $i \in [k]$. For an arbitrary natural number m and $v \in L^p(\Omega)$ let $[v]_m$ denote the $m^{-1}\mathbb{Z}$-discretization of v obtained by composing v with the function $x \mapsto \lceil xm \rceil / m$. It is clear that $\|v - [v]_m\|_p \le m^{-1}$. For every $i \in [k]$ the level sets of $[v_i]_m$ partition Ω_1 into at most $2m$ measurable sets. By taking common refinement of the level sets, we obtain that there is a partition $\{P_i\}_{i=1}^N$ of Ω_1 into $N \le (2m)^k$ measurable sets such that each $[v_i]_m$ is measurable in this partition. This means that there exist real numbers $\{u_{i,j}\}_{i \in [k], j \in [N]}$ between -1 and 1 such that for every $i \in [k]$ we have $[v_i]_m = \sum_{j \in [N]} u_{i,j} 1_{P_j}$.

Let $\varepsilon' > 0$ be some sufficiently small number. If δ is small enough, we have that there is a partition $\{Q_i\}_{i=1}^N$ of Ω_2 such that $\mathcal{D}_A(\{1_{P_i}\}_{i=1}^N)$ and $\mathcal{D}_B(\{1_{Q_i}\}_{i=1}^N)$ are at most ε' far in d_{LP}. Let $w_i := \sum_{j \in [N]} u_{i,j} 1_{Q_j}$. It is clear that if ε' is small enough, then $\kappa := \mathcal{D}_B(\{w_i\}_{i=1}^k)$ is arbitrarily close to $\mathcal{D}_B(\{v_i\}_{i=1}^k)$. We obtain that if m is big enough, and ε' is small enough then $\kappa \in \mathcal{S}_k(B)$ is at most ε far from μ. All the estimates in the proof depend only on c, ε, k and p, q. ∎

8 Dense graph limits and graphons

In this chapter, we explain how the so-called dense graph limit theory fits into our general limit theory. Let us consider the probability space $([0,1], \mathcal{L}, \lambda)$, where λ is the Lebesgue measure on the Lebesgue σ-algebra $\mathcal{L}$. Special P-operators on $L^2([0,1])$ – called graphons—play a crucial role in dense graph limits. A *graphon* is a two-variable measurable function $W : [0,1]^2 \to [0,1]$ with the symmetry property that $W(x,y) = W(y,x)$ holds for every $x, y \in [0,1]$. Graphons act on the Hilbert space $L^2([0,1])$ by

$$(fW)(x) := \int_y f(y) W(y,x) \, d\lambda,$$

where $f \in L^2([0,1])$. It is easy to see that $\|W\|_{2\to2} \le 1$ and thus graphons are P-operators. It is also clear that graphons are positivity-preserving and self-adjoint operators and hence they are also graphops. Let $\mathcal{W}$ denote the space of graphons. For $U, W \in \mathcal{W}$ we say that $U \sim W$ (U is isomorphic to W) if $\delta_\square(U, W) = 0$. Let $\tilde{\mathcal{W}} := \mathcal{W}/\sim$ be the set of equivalence classes. The next theorem from [28] is a fundamental result in graph limit theory.

Theorem 8.1 (Lovász–Szegedy, [28]) *The metric space $(\tilde{\mathcal{W}}, \delta_\square)$ is compact.*

The space $(\tilde{\mathcal{W}}, \delta_\square)$ is basically the graph limit space. Every finite graph $G = (V(G), E(G))$ with $V(G) = [n]$ is represented in $\tilde{\mathcal{W}}$ by the function W_G with $W_G(x,y) := 1$ if $(\lceil xn \rceil, \lceil yn \rceil) \in E(G)$ and $W_G(x,y) := 0$ otherwise. Graph convergence in dense graph limit theory is equivalent to the convergence of the representing functions W_G in $(\tilde{\mathcal{W}}, \delta_\square)$. Our next theorem shows that this limit theory is embedded into our more general limit framework.

Theorem 8.2 *The two pseudometrics $\delta_\square$ and d_M are equivalent on $\mathcal{W}$.*

Proof By Lemma 2.14, it remains to show that for every $\varepsilon > 0$ there is $\delta > 0$ such that if $d_M(U, W) \le \delta$, then $\delta_\square(U, W) \le \varepsilon$. By contradiction, let us assume that there exist $\varepsilon > 0$ and two sequence of graphons $\{U_i\}_{i=1}^\infty$ and $\{W_i\}_{i=1}^\infty$ such that $\delta_\square(U_i, W_i) > \varepsilon$ for every $i \in \mathbb{N}$ and $\lim_{i\to\infty} d_M(U_i, W_i) = 0$. By choosing an appropriate subsequence, we can assume by Theorem 8.1 that $\lim_{i\to\infty} U_i = U$ and $\lim_{i\to\infty} W_i = W$ holds where the convergence is in $\delta_\square$. We obtain that $\delta_\square(U, W) \ge \varepsilon$. On the other hand, by the triangle inequality and Lemma 2.14 we have that

$$d_M(U, W) \le d_M(U, U_i) + d_M(U_i, W_i) + d_M(W_i, W)$$
$$\le d_M(U_i, W_i) + 12\delta_\square(U, U_i)^{1/2} + 12\delta_\square(W, W_i)^{1/2}$$

and thus by taking $\lim_{i\to\infty}$ we get that $d_M(U, W) = 0$. We have by Proposition 7 that $d_{1,H}(\mathcal{Q}_k(U), \mathcal{Q}_k(W)) = 0$ holds for every $k \in \mathbb{N}$. It is well known in graph limit theory (see, e.g., [8]) that such quotient equivalence implies $\delta_\square(U, W) = 0$. ∎

We need the following lemma.

Lemma 8.3 *For every $\varepsilon > 0$ there exists a number n such that if G is a finite graph with $|V(G)| \ge n$, then $d_M(A(G)/|V(G)|, W_G) \le \varepsilon$.*

Proof Fix $k \geq 1$, and recall that the k-profile of $A(G)/|V(G)|$ is based on k functions of the form $h_j : [n] \to [-1,1]$ (where integers represent the vertices of the graph), while the k-profile of W_G is based on functions $v_j : [0,1] \to \mathbb{R}$ with $\|v_j\|_\infty \leq 1$. By the representation given above (after Theorem 8.1), for every such function h_j we can define the function $v_j(x) = h_j(\lceil xn \rceil)$ for $x \in (0,1]$ and $v_j(0) = h_j(1)$. Since this is consistent with the graphon representation of graphs, the empirical distributions

$$\mathcal{D}(h_1, \ldots, h_k, h_1 A(G)/|V(G)|, \ldots, h_k A(G)/|V(G)|)$$

and

$$\mathcal{D}(v_1, \ldots, v_k, v_1 W_G, \ldots, v_k W_G)$$

will be exactly the same. That is, the k-profile of $A(G)/|V(G)|$ is contained by the k-profile of W_G.

Therefore, the main part is to show that for every $\varepsilon > 0$ there exists N such that for $n \geq N$ for every $v_j : [0,1] \to \mathbb{R}$ with $\|v_j\|_\infty \leq 1$ we can find $h_1, h_2, \ldots, h_k : [n] \to [-1,1]$ such that the Lévy–Prokhorov distance of the two empirical distributions is at most ε.

First, we suppose that all possible values of the functions v_j are multiples of an appropriate $\delta > 0$ (to be chosen later, depending only on k and ε. That is, we suppose that $v_j : [0,1] \to T_\delta$, where T_δ is the finite set consisting of the multiples of δ within $[-1,1]$. We will see later that this easily implies the general statement. Now, we use the weak Szemerédi regularity lemma (in the form of Lemma 9.3 and the remark afterward in [27]). This says that given $S \geq 2$ and a graph G on n vertices, there exists a partition $\mathcal{P}$ of the vertices of G to S parts such that (i) the difference between the size of any two classes is at most 1; (ii) we have $d_\square(G, G_\mathcal{P}) \leq 4/\sqrt{\log S}$. Here, $G_\mathcal{P}$ is a weighted graph on the vertex set of G, where the weight of uv is equal to the edge density between the class of u and v. By considering the corresponding graphons, we can say the following. Given $\varepsilon > 0$, let $S = S(\varepsilon)$ be chosen such that $4/\sqrt{\log S} < \varepsilon$. Furthermore, let us partition $[0,1]$ into $\cup_{s=1}^{S} P_s$, where each P_s is the union of the intervals corresponding to the vertices in class s (in the graphon representation of a graph, the ith vertex is represented by the interval $((i-1)/n, i/n)$). Then, the regularity lemma implies that given $\varepsilon > 0$ and W_G, we can find a graphon W'_G such that $\|W_G - W'_G\|_\square \leq \varepsilon$, and W'_G is a step function, which is constant on each set of the form $P_s \times P_t$. Furthermore, by inequality (2.7), this condition implies that $\|W_G - W'_G\|_{\infty \to 1} \leq 4\varepsilon$. Then, by the same argument as in the proof of Lemma 2.12, namely, using Lemma 13.2, we obtain that

$$d_{\mathrm{LP}}\big(\mathcal{D}(v_1, \ldots, v_k, v_1 W_G, \ldots, v_k W_G),$$
$$\mathcal{D}(v_1, \ldots, v_k, v_1 W'_G, \ldots, v_k W'_G) \leq 2\sqrt{\varepsilon} \cdot (2k)^{3/4}.$$

Now, we define a modification of the family of functions $(v_1, \ldots, v_k)$ mapping from $[0,1]$ to T_δ, with the new versions being constants on each interval $((i-1)/n, i/n)$, but with approximately the same joint empirical distribution. For every element $\omega \in T_\delta^k$, we define

$$\mu_{s,\omega} = |\{x \in P_s : (v_1(x), \ldots, v_k(x)) = \omega\}|,$$

where $|\cdot|$ denotes the Lebesgue measure on $[0,1]$. For every $\omega \in T_\delta^k$, we choose $\lfloor \mu_{s,\omega} \cdot n \rfloor$ intervals belonging to P_s, and define $(\bar{v}_1, \ldots, \bar{v}_k)$ to take value ω on these intervals (for different ωs we choose different intervals; since the sum of $\mu_{s,\omega}$ is 1, and we have n intervals, this is possible). On the remaining intervals, we let $(\bar{v}_1, \ldots, \bar{v}_k)$ to be 0.

We will use the following facts about this transformation. First, if the quantities $\mu_{s,\omega}$ are the same for every $s = 1, \ldots, S$ and $\omega \in T_\delta^k$ for $(v_1, \ldots, v_k)$ and some $(u_1, \ldots, u_k)$, then

$$\mathcal{D}(v_1, \ldots, v_k, v_1 W_G', \ldots, v_k W_G') = \mathcal{D}(u_1, \ldots, u_k, u_1 W_G', \ldots, u_k W_G').$$

To see this, recall that W_G' is constant on each $P_i \times P_j$, which means that it is only the average on P_j that counts, and we did not change that. However, the quantities $\mu_{s,\omega}$ are changed when we use the rounding, in order to make the function constant on each interval $((i-1)/n, i/n)$. For $\omega \in T_\delta^k$, we have, at most, one "remainder" interval, and by using the fact that this set has a fixed size, plus, the partition classes have almost the same size, we get that the rounding affects the function only on a set whose measure tends to 0 as $n \to \infty$. Thus, for large enough n, we can find functions $u_1, \ldots, u_k : [0,1] \to T_\delta^k$ such that they have the same quantities $\mu_{s,\omega}$ as $v_1, \ldots, v_k$, and $\|u_j - \bar{v}_j\|_1 \le \varepsilon$. By Lemma 13.2 and the finiteness of the norm of W_G' it follows that

$$d_{\mathrm{LP}}\big(\mathcal{D}(v_1, \ldots, v_k, v_1 W_G', \ldots, v_k W_G'), \mathcal{D}(\bar{v}_1, \ldots, \bar{v}_k, \bar{v}_1 W_G', \ldots, \bar{v}_k W_G')\big) \le \varepsilon$$

for large enough n.

Now, we use the same argument for $\bar{v}_j$ that we used for v_j, which is based on the proof of Lemma 2.12 and Lemma 13.2 and $\|W_G - W_G'\|_\square \le \varepsilon$. This yields

$$d_{\mathrm{LP}}\big(\mathcal{D}(\bar{v}_1, \ldots, \bar{v}_k, \bar{v}_1 W_G', \ldots, \bar{v}_k W_G'),$$
$$\mathcal{D}(\bar{v}_1, \ldots, \bar{v}_k, \bar{v}_1 W_G, \ldots, \bar{v}_k W_G)\big) \le 2\sqrt{\varepsilon} \cdot (2k)^{3/4}.$$

Since $\bar{v}_j$ is constant on each interval $(i/n, (i+1)/n)$, it is straightforward to assign a function $h_j : [n] \to T_\delta$ to it, for which we have

$$\mathcal{D}(\bar{v}_1, \ldots, \bar{v}_k, \bar{v}_1 W_G, \ldots, \bar{v}_k W_G)) = \mathcal{D}\left(h_1, \ldots, h_k, h_1 \frac{A(G)}{|V(G)|}, \ldots, h_k \frac{A(G)}{|V(G)|}\right).$$

To see that this is the right normalization of the adjacency matrix, recall that W_G is defined on $[0,1]^2$, and the length of the intervals on which it is constant 1 is $1/n$.

Putting this together, we conclude that we can find appropriate functions $h_j : [n] \to T_\delta$ such that

$$d_{\mathrm{LP}}\left(\mathcal{D}(v_1, \ldots, v_k, v_1 W_G, \ldots, v_k W_G)),\right.$$
$$\left.\mathcal{D}\left(h_1, \ldots, h_k, h_1 \frac{A(G)}{|V(G)|}, \ldots, h_k \frac{A(G)}{|V(G)|}\right)\right) = O(\varepsilon).$$

Finally, let $w_1, \ldots, w_k : [0,1] \to \mathbb{R}$ with $\|v_j\|_\infty \le 1$ be arbitrary functions (we omit the condition that is goes to T_δ^k). It is clear that we can find $v_1, \ldots, v_k : [0,1] \to T_\delta$

such that $\|w_j - v_j\|_1 \le \delta$. As we have seen before, this implies that

$$d_{\mathrm{LP}}(\mathcal{D}(w_1, \ldots, w_k, w_1 W_G, \ldots, w_k W_G)), \mathcal{D}(v_1, \ldots, v_k, v_1 W_G, \ldots, v_k W_G)) = O(\delta).$$

Therefore, if $\delta = \delta(\varepsilon)$ is a fixed constant smaller than ε, then, together with the first step of the proof, we can conclude that that the Hausdorff distance of the k-profiles of W_G and $A(G)/|V(G)|$ is smaller than ε for large enough n. ∎

Together with Theorem 8.2 it implies the following.

Proposition *If $\{G_i\}_{i=1}^\infty$ is a growing sequence of finite graphs, then the action convergence of $\{A(G_i)/|V(G_i)|\}_{i=1}^\infty$ is equivalent to dense graph convergence.*

9 Benjamini–Schramm and local-global limits

Benjamini–Schramm and local-global limits are used in the study of bounded degree graphs. Let d be a fixed number and let $\mathcal{G}_d$ denote the set of isomorphism classes of graphs with maximal degree at most d. Informally speaking, a graph sequence $\{G_i\}_{i=1}^\infty$ in $\mathcal{G}_d$ is Benjamini–Schramm convergent if for every fixed r the probability distribution of isomorphism classes of neighborhoods of radius r converges when i goes to infinity. It is often useful to refine this convergence notion to the local-global setting. In this framework, we put "colorings" on the vertex sets of G_i in all possible ways and look at all possible colored neighborhood statistics. It is not to confuse with colored Benjamini–Schramm limits, where we put one coloring on each graph G_i. We give the formal definition of local-global limits.

We summarize the notion of local-global convergence based on [21]. A rooted graph is a graph with a distinguished vertex o called *root*. The radius of a rooted graph is the maximal distance from the root over all vertices. A k-coloring of a graph $G = (V, E)$ is a function $f : V \to [k]$. Let $\mathcal{G}_{d,k,r}$ denote the set of isomorphism classes of k-colored rooted graphs of maximal degree at most d and radius at most r. Note that $\mathcal{G}_{d,k,r}$ is a finite set. We denote by $\mathcal{P}(\mathcal{G}_{d,k,r})$ the set of probability distributions on $\mathcal{G}_{d,k,r}$. We have that $\mathcal{P}(\mathcal{G}_{d,k,r})$ together with the total variation distance d_{TV} is a compact metric space. By abusing the notation, we denote by d_H the Hausdorff distance for subsets in $(\mathcal{P}(\mathcal{G}_{d,k,r}), d_{\mathrm{TV}})$.

Let $G = (V, E) \in \mathcal{G}_d$ and let $f : V \to [k]$ be a k-coloring. We denote by $\tau_r(G, f)$ the probability distribution on $\mathcal{G}_{d,k,r}$ obtained by putting the root o on a uniformly chosen random vertex of G and then taking the colored neighborhood of o of radius r. Let $Z_{k,r}(G)$ denote the set of all possible probability distributions $\tau_r(G, f)$, where f runs through all possible k-colorings of G. We have that $Z_{k,r}(G)$ is a subset of $\mathcal{P}(\mathcal{G}_{d,k,r})$.

Definition 9.1 A graph sequence $\{G_i\}_{i=1}^\infty$ is called local-global convergent if for every $r, k \ge 1$ the sequence $\{Z_{k,r}(G_i)\}_{i=1}^\infty$ is convergent in the metric d_H.

It was proved in [21] that limits of local-global convergent graph sequences can be described by certain Borel graphs called graphings. We give the formal definition.

Definition 9.2 (Graphing) Let X be a Polish topological space and let v be a probability measure on the Borel sets in X. A graphing is a graph G on $V(G) = X$

with Borel measurable edge set $E(G) \subset X \times X$ in which $\int_X e(x, y) d\nu(y) \le d$ for all $x \in X$ and

$$(9.1) \qquad \int_A e(x, B) d\nu(x) = \int_B e(x, A) d\nu(x)$$

for all measurable sets $A, B \subseteq X$, where $e(x, S)$ is the number of edges from $x \in X$ to $S \subseteq X$.

A k-coloring of a graphing $G = (X, E)$ is a measurable function $f : X \to [k]$. The probability measure ν allows us to talk about random vertices in G. The colored neighborhood of a random vertex in G is a graph in $\mathcal{G}_{d,k,r}$. The probability distribution $\tau_r(G, f)$, the measure $Z_{k,r}(G)$ and convergence are similarly defined as for finite graphs.

Finite graphs are special graphings, where X is a finite set and ν is the uniform distribution. It will be important that graphings are bounded operators on $L^2(X, \nu)$. The action is given by

$$(\nu G)(x) = \sum_{(y,x) \in E} \nu(y)$$

for $\nu \in L^2(X, \nu)$. We have that $\|G\|_2 \le d$. Note that the integral formula (9.1) is equivalent to the fact that G is a self-adjoint operator. Graphings are also positivity-preserving and hence they are examples of graphops. The next theorem is proved in [36].

Theorem 9.1 *A sequence of graphings $\{G_i\}_{i=1}^{\infty}$ is local-global convergent if and only if $\{Z_{k,1}(G_i)\}_{i=1}^{\infty}$ is convergent in d_H for every fixed $k \ge 1$.*

Our main theorem here says that, restricted to graphings, P-operator convergence is the same as local-global convergence. Consequently, P-operator convergence is a generalization of graphing convergence.

Theorem 9.2 *A sequence of graphings is local-global convergent if and only if it is convergent in the metric d_M.*

Proof We need some preparation. Let $M_{d,k}$ denote the set of vectors $v = (v_1, v_2, \ldots, v_{2k})$ in $\mathbb{N}_0^{[2k]}$ (where $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$) such that $\sum_{i=1}^{k} v_i = 1$ and $\sum_{i=k+1}^{2k} v_i \le d$. There is a natural bijection α between $M_{d,k}$ and $\mathcal{G}_{d,k,1}$ given in the following way. For a vector $v \in M_{d,k}$ let $q(v)$ denote the unique coordinate $i \in [k]$ with $v_i = 1$ and let $s(v) := \sum_{i=k+1}^{2k} v_i$. We denote by $\alpha(v)$ the colored star in which the color of the root (denoted by o) is $q(v)$, the root has $s(v)$ neighbors, and for the neighbors of o the color $i \in [k]$ is used v_{i+k} times. It is clear that the isomorphism type $\alpha(v)$ is determined by this information and each isomorphism type in $\mathcal{G}_{d,k,1}$ is obtained this way. Consequently α is a bijection. We denote by $\hat{\alpha}$ the bijection between the sets of probability measures $\mathcal{P}(M_{d,k})$ and $\mathcal{P}(\mathcal{G}_{d,k,1})$ induced by α using the formula $\hat{\alpha}(\mu)(T) := \mu(\alpha^{-1}(T))$. It is clear that $\hat{\alpha}$ is continuous with respect to the given metrization on $\mathcal{P}(M_{d,k})$ and $\mathcal{P}(\mathcal{G}_{d,k,1})$.

Observe that if G is a graphing (of maximal degree d), then $\mathcal{S}'_k(G) = \mathcal{S}_k(G) \cap \mathcal{P}(M_{d,k})$. Now we show that $\hat{\alpha}(\mathcal{S}'_k(G)) = Z_{k,1}(G)$. To see this, notice that colorings $f : X \to [k]$ are in a one-to-one correspondence with systems of $0 - 1$-valued

functions $\{v_i\}_{i=1}^k$ with $\sum_{i=1}^k v_i = 1_X$, called function partitions. The correspondence is given by $v_i = 1_{f^{-1}(i)}$. It is clear that if $\{v_i\}_{i=1}^\infty$ is a function partition, then

$$\hat{\alpha}(\mathcal{D}_G(\{v_i\}_{i=1}^k)) = \tau_1(G, f).$$

We obtain that $\hat{\alpha}(\mathcal{S}'_k(G)) = \hat{\alpha}(\mathcal{S}_k(G) \cap \mathcal{P}(M_{d,k})) = Z_{k,1}(G)$. Now the continuity of $\hat{\alpha}$ and Theorem 7.3 finish the proof. ∎

10 Generalizations

Action convergence is based on a very general principle. We do not exploit the generality of it in this paper, but as illustration we describe a few useful generalizations.
Complex spaces: The theory developed in this paper can be generalized to operators acting on complex number valued function spaces. Most of the definitions are the same and the proofs of the theorems require some minor changes. Note also that if a P-operator A over $\mathbb{C}$ has the special property that it takes real-valued functions to real valued functions, then its k-profile over $\mathbb{C}$-valued functions can be reconstructed from its $2k$-profile over $\mathbb{R}$ by decomposing functions according to real and complex part.
Simultaneous convergence: We have mentioned in the introduction that it is sometimes useful to introduce simultaneous convergence of pairs (A, f), where $A \in \mathcal{B}(\Omega)$ is a P-operator and f is measurable on Ω. Based on the same principle, one can further generalize this to a simultaneous convergence notion of several P-operators and several functions. An especially interesting case is when matrices and their adjoint matrices are considered simultaneously. In this case, $\mathcal{S}_k(A)$ is defined by

$$\mathcal{D}(v_1, v_2, \ldots, v_k, v_1 A, v_2 A, \ldots, v_k A, v_1 A^*, v_2 A^*, \ldots, v_k A^*).$$

It is not clear whether this leads to a finer convergence notion of matrices or not.
Nonlinear operators: In the definition of the metric d_M, we never use the linearity of the operators. The definition of k-profile and distance $d_M(A, B)$ is meaningful for arbitrary functions $A : L^\infty(\Omega_1) \to L^1(\Omega_1)$ and $B : L^\infty(\Omega_2) \to L^1(\Omega_2)$. (Even multivalued functions can be allowed.) Interesting examples for nonlinear operators are finite matrices composed pointwise with nonlinear functions. For example: $(x, y, z)A :=$ $((x + y)^2, \sin(y + z), z - x)$. Such functions arise in deep learning.

11 Random matrices

In this section, we investigate the convergence of certain dense random matrices with respect to the metric d_M. We consider a sequence of normalized random matrices (H_n) with independent zero-mean $\pm n^{-1/2}$-valued random variables as entries. This is the same as if we choose an element uniformly at random from the set of all $n \times n$ matrices with $\pm n^{-1/2}$ entries. This set will be denoted by $\mathcal{M}_n$. Our goal is to prove the following.

Proposition *For every infinite $S' \subseteq \mathbb{N}$ there exists a P-operator A and an infinite set $S \subseteq S'$ such that the sequence $(H_j)_{j \in S}$ converges to A with respect to d_M with probability 1.*

We start with a statement on the concentration of measure.

Lemma 11.1 *For every $n \in \mathbb{N}$, let $M_n \in \mathcal{M}_n$ be fixed, and H_n be a uniformly chosen element of $\mathcal{M}_n$. Then for every $\eta > 0$, we have*

$$\lim_{n \to \infty} \mathbb{P}\big(\big|d_M(M_n, H_n) - \mathbb{E}(d_M(M_n, H_n))\big| > \eta\big) = 0.$$

Proof For $1 \le j \le n$, let $\mathcal{F}_j$ be the σ-algebra generated by the first j columns of H_n. We apply the well-known concentration inequalities for the martingale $Y_j = \mathbb{E}\big(d_M(M_n, H_n)\big|\mathcal{F}_j\big)$. Notice that if matrices $A, B \in \{-1, 1\}^{n \times n}$ differ only in a single column, then the distance of $\mathcal{D}_A(v_1, \ldots, v_k)$ and $\mathcal{D}_B(v_1, \ldots, v_k)$ is at most $1/n$ in the Lévy–Prokhorov metric (for arbitrary k and vectors v_j), because the two measures coincide everywhere except on an event of probability $1/n$. This implies $d_M(A, B) \le 1/n$. Hence $|Y_j - Y_{j-1}| \le 1/n$ holds for every $j = 1, 2, \ldots, n$. Therefore, by Azuma's inequality, we have that

$$\mathbb{P}\big(\big|d_M(M_n, H_n) - \mathbb{E}(d_M(M_n, H_n))\big| > \eta\big) \le 2 \exp\left(-\frac{\eta^2}{2n \cdot (1/n)^2}\right)$$
$$= 2 \exp(-\eta^2 n / 2). \qquad \blacksquare$$

Let $\mathcal{X}_{p,q,c}$ be the set of matrices M with $\|M\|_{p \to q} \le c$.

Lemma 11.2 *There exists a sequence of matrices $(M_j)_{j \in \mathbb{N}}$ such that the following conditions hold: $(i) M_j \in \mathcal{M}_j \cap \mathcal{X}_{2,2,3}$ for every $j \in \mathbb{N}$; $(ii) d_M(M_j, H_j) \to 0$ in probability as $j \to \infty$, where H_j is a uniformly chosen random element of $\mathcal{M}_j$.*

Proof Given $\varepsilon > 0$, first we find a sequence of matrices around which random matrices are concentrated with error ε. The metric space $(\mathcal{X}_{2,2,3}, d_M)$ is compact by Theorem 2.10, hence it contains a finite $\varepsilon/8$-net. We denote the size of this net by $F(\varepsilon)$. Consider balls of radius $\varepsilon/8$ around the elements of this net. Let $\mathcal{N}_{\varepsilon,n}$ be the set of matrices satisfying the following property: in one of these balls, it is the closest element of $\mathcal{M}_n$ to the center (in case of equality, choose one arbitrarily). Then, $\mathcal{N}_{\varepsilon,n}$ is an $\varepsilon/4$-net in $\mathcal{M}_n \cap \mathcal{X}_{2,2,3}$, and its size is at most $F(\varepsilon)$, as we have chosen at most one element from each ball. It follows that there exists $M'_{\varepsilon,n} \in \mathcal{N}_{\varepsilon,n}$ such that

$$\mathbb{P}\big(d_M(M'_{\varepsilon,n}, H_n) \le \varepsilon/4\big) \ge \frac{1 - \mathbb{P}(\|H_n\|_2 > 3)}{F(\varepsilon)}.$$

Since the operator norm of our random matrix H_n random matrix is concentrated around its expectation 2 (see, e.g., [19]), the probability $\mathbb{P}(\|H_n\|_2 > 3)$ tends to 0 as n goes to infinity. Therefore, for every $\varepsilon > 0$, we have

$$\liminf_{n \to \infty} \mathbb{P}\big(d_M(M'_{\varepsilon,n}, H_n) \le \varepsilon/4\big) > 0.$$

This equation together with Lemma 11.1 for $\eta = \varepsilon/4$ and $(M'_{\varepsilon,n})_{n \in \mathbb{N}}$ implies that

$$\mathbb{E}\big(d_M(M'_{\varepsilon,n}, H_n)\big) \le \frac{\varepsilon}{2}.$$

By combining this with Lemma 11.1 for $\eta = \varepsilon/2$, we conclude that

$$\lim_{n \to \infty} \mathbb{P}\big(d_M(M'_{\varepsilon,n}, H_n) > \varepsilon\big) = 0.$$

The proof can be completed by a standard diagonalization argument. More precisely, we can choose a function $n_0(\varepsilon)$ such that

$$\mathbb{P}\big(d_M(M'_{\varepsilon,n}, H_n) > \varepsilon\big) < \varepsilon \text{ holds for every } n \geq n_0(\varepsilon).$$

Now let $k(n) = \max\{k : n_0(1/k) < n\}$. Then, the sequence $M_j = M'_{1/k(j),j}$ satisfies the conditions of the lemma. ∎

Proof of Proposition 11 Let $(M_j)_{j \in \mathbb{N}}$ be a sequence of matrices satisfying the conditions of Lemma 11.2. By this lemma, we can choose an infinite subset $S \subseteq S'$ such that $(d_M(M_j, H_j))_{j \in S}$ tends to 0 with probability 1 as $j \to \infty$, and $(M_j)_{j \in S}$ converges to a P-operator A with respect to d_M. To guarantee the second condition, we can use Lemma 2.6 and Theorem 2.9, because $M_j \in \mathfrak{X}_{2,2,3}$ for all j. This S will be an appropriate subset of S'. ∎

12 Examples

12.1 Hypercubes and uniform towers

The hypercube graph $Q_n = (V(Q_n), E(Q_n))$ is formed by the vertices and edges of the n-dimensional hypercube. More precisely, $V(Q_n) = \{0,1\}^n$ and two vertices are connected if and only if the representing vectors have Hamming distance one, i.e., they differ at exactly one coordinate. The graph Q_n is n-regular, $|V(Q_n)| = 2^n$ and $|E(Q_n)| = 2^{n-1}n$. This means that the sequence $\{Q_n\}_{n=1}^{\infty}$ is very sparse but not with bounded degrees. Note that Q_n is a Cayley graph of the group Z_2^n where $\{0,1\}$ is identified with the cyclic group Z_2 of order 2 and the generators are the basis vectors $e_i, i \in [n]$ with 1 at the ith coordinate and 0 elsewhere.

Our goal is to show that hypercubes converge to an appropriate Cayley graph of the compact group Z_2^{∞} with a carefully chosen topological basis.

A topological basis is an independent set of vectors in Z_2^{∞} that generates a dense set in Z_2^{∞}. (Note that topological independence is not assumed here.) Quite surprisingly, the usual topological basis $\{e_i\}_{i=1}^{\infty}$ of Z_2^{∞} is not useful for constructing the limit of the hypercubes. The main obstacle is that $\{e_i\}_{i=1}^{\infty}$ is a countable set but there is no natural uniform distribution on an infinite countable set. Instead, we need to find a nice enough topological basis with uncountably many elements and a natural uniform distribution on this basis.

Since Q_n is regular, we have that adjacency operator convergence is equivalent to random walk convergence, so we do not have to choose one of them. The right scaling of the sequence is $A_n := A(Q_n)/n$, where $A(Q_n)$ is the adjacency matrix of Q_n. The operator A_n is a Markov graphop, and if $\{A_n\}_{n=1}^{\infty}$ is convergent, then the limit is also a Markov graphop (recall Theorem 3.3). As we stated above, the purpose of this part of the paper is to show that they indeed converge and to determine the limit object. Some details will be left to the reader regarding the general convergence. We will work with the subsequence $\{Q_{2^n}\}_{n=1}^{\infty}$ that has especially nice properties based on certain uniform mappings between $Q_{2^{n+1}}$ and Q_{2^n}. The general convergence can be obtained from approximate versions of these uniform maps.
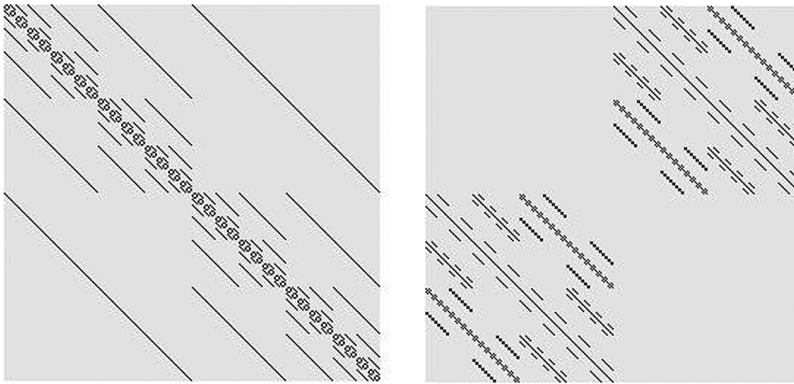
Figure 5: Two representations of the adjacency matrix of the hypercube in dimension 8.

In Figure 5, we show the adjacency matrix of the eight-dimensional hypercube Q_8 using two different orderings of the vertices. Light gray points represent zeros and black points represent ones. The first ordering is based on the binary forms of the numbers $0, 1, 2, \ldots, 255$, which is a rather natural way to order $\{0, 1\}^8$. In the second figure, we compose this ordering with a carefully chosen automorphism of the group Z_2^8. Quite surprisingly, it turns out that the second figure provides a more useful representation when going to the limit. There is a qualitative difference between the two types of representations of Q_8. Intuitively, the first pictures would converge to some "infinite picture," where each vertical (and horizontal) line has countable intersection with the black points. On the other hand, the second figure fits into a sequence such that, after going to the limit, vertical (and horizontal) lines have uncountable intersections with the black points. We will see later that this helps in putting a uniform distribution on the limiting picture.

We will need the following definition. The first part will be a generalization of the graph theoretic notion of covering. Namely, for $b = 1$, we get back the usual definition. For larger b, the map can contract several neighbors of a given vertex into one single vertex, but, as condition (3) shows, only in a balanced way, i.e., the number of contracted neighbors is fixed.

Definition 12.1 (Uniform map and uniform tower)   Let G_1, G_2 be finite graphs. A map $f : V(G_2) \to V(G_1)$ is (a, b)-uniform if

(1) f is a graph homomorphism, i.e., $(f(u), f(w)) \in E(G_1)$ holds for every $(v, w) \in E(G_2)$.
(2) $|f^{-1}(u)| = a$ holds for every $v \in V(G_1)$.
(3) If $u \in V(G_2)$ and w is any neighbor of $f(u)$, then exactly b neighbors of u are mapped to w.

A uniform tower is a sequence $\{G_i, f_i\}_{i=1}^{\infty}$ of finite graphs G_i and maps $f_i : V(G_{i+1}) \to V(G_i)$ such that f_i is (a_i, b_i)-uniform for $i \in \mathbb{N}$.

Recall from Definition 2.7 that $A \prec B$ (A is weakly contained in B) if $\mathcal{S}_k^*(A) \subseteq \mathcal{S}_k^*(B)$ holds for every $k \in \mathbb{N}$, where $\mathcal{S}_k^*(A)$ denotes the closure of $\mathcal{S}_k(A)$ in $(\mathcal{P}(\mathbb{R}^{2k}), d_{\mathrm{LP}})$.

Lemma 12.1   *Let G_1, G_2 be finite graphs and let $f : V(G_2) \to V(G_1)$ be an (a, b)-uniform map for some $a, b \in \mathbb{N}$. Then $A(G_1) \prec A(G_2)/b$.*

Proof   The second property of uniformity implies that if $v_1, v_2, \ldots, v_k \in \mathbb{R}^{V(G_1)}$ for some k, then

$$\mathcal{D}(v_1, v_2, \ldots, v_k) = \mathcal{D}(v_1 \circ f, v_2 \circ f, \ldots, v_k \circ f).$$

Furthermore, the first and third properties imply that if $v \in \mathbb{R}^{V(G_1)}$, then $vA(G_1) \circ f = (v \circ f)A(G_2)/b$. To see this, notice that the left-hand side at a given vertex $u \in V(G_2)$ is the sum of f at the neighbors of $f(u)$. By the third property, exactly b neighbors of u are mapped to any of these vertices. On the other hand, since f is a graph homomorphism, neighbors of u can be mapped only to the neighbors of w. We get that the left-hand side is equal to b times the sum of values of $f(w)$ at the neighbors w of u, and the latter is exactly $(v \circ f)A(G_2)$. This proves the equality. Therefore, we obtain that

$$\mathcal{D}_{A(G_1)}(\{v_i\}_{i=1}^k) = \mathcal{D}_{A(G_2)/b}(\{v_i \circ f\}_{i=1}^k),$$

and hence $\mathcal{S}_k(A(G_1)) \subseteq \mathcal{S}_k(A(G_2)/b)$ holds for every k. ∎

Recall that if $\{X_i\}_{i=1}^\infty$ is a sequence of finite sets with maps $f_i : X_{i+1} \to X_i$, then the inverse limit X is the set of elements in $(x_1, x_2, \ldots) \in \prod_{i=1}^\infty X_i$ such that $f_i(x_{i+1}) = x_i$ holds for every i. Since X is a closed subset of the compact space $\prod_{i=1}^\infty X_i$, we have that X is compact with respect to the subspace topology. The map $\pi_i : X \to X_i$ defined by $\pi_i(x_1, x_2, \ldots) := x_i$ is a continuous map. If each f_i has the property that $|f_i^{-1}(v)| = |f_i^{-1}(w)|$ holds for every $v, w \in X_{i+1}$, then by the Kolmogorov extension theorem there is a unique Borel probability measure μ on X such that for every i the push-forward measure of μ under π_i is uniform on X_i. We call μ the uniform measure on X.

Definition 12.2   Let $\{G_i, f_i\}_{i=1}^\infty$ be a uniform tower such that G_i is finite and d_i-regular for $i \in \mathbb{N}$. Let V be the inverse limit of $\{V(G_i), f_i\}_{i=1}^\infty$. For every $x \in V$ let $N(x) \subseteq V$ denote the inverse limit of the set of neighbors of $\pi_i(x)$ and let v_x denote the uniform measure on $N(x)$ (more precisely, as the previous argument shows, there is a uniform measure on V, as it was constructed as the inverse limit of the finite sets $V(G_i)$, and this induces a uniform measure on $N(x)$). Let A be the P-operator in $\mathcal{B}_{2,2}(V, \mu)$ defined by $(fA)(x) = \int_V f \, dv_x$. We say that A is the inverse limit of the tower $\{G_i, f_i\}_{i=1}^\infty$.

Theorem 12.2 (Convergence of uniform towers)   *Let $\{G_i, f_i\}_{i=1}^\infty$ be a uniform tower such that G_i is d_i-regular for $i \in \mathbb{N}$. Then $\{A(G_i)/d_i\}$ is a convergent sequence of P-operators and the limit object is the inverse limit of $\{G_i, f_i\}_{i=1}^\infty$.*

Proof   Observe that $\|A(G_i)/d_i\|_2 = 1$ and if f_i is (a_i, b_i) uniform, then $d_{i+1} = d_i b_i$. We have by Lemma 12.1 that $A(G_i)/d_i \prec A(G_{i+1})/d_{i+1}$ holds for every $i \in \mathbb{N}$. It follows by compactness that $\mathcal{S}_k(A(G_i)/d_i)$ converges to $\cup_{i=1}^\infty \mathcal{S}_k(A(G_i)/d_i)$ in d_H as i goes to infinity. Let $A \in \mathcal{B}_{2,2}(V, \mu)$ be the inverse limit of the tower $\{G_i, f_i\}_{i=1}^\infty$. By

approximating measurable functions in $L^2(V, \mu)$ by functions of the form $v \circ \pi_i$ we obtain that $\mathcal{S}_k^*(A)$ is the closure of $\cup_{i=1}^\infty S_k(A(G_i)/d_i)$. More precisely, first notice that every function $h : V(G_i) \to \mathbb{R}$ can be represented in the inverse limit in a natural way, by a function which is defined on V but depends only on $\pi_i(v)$. We will call this the lift of h, and denote it by $\tilde{h}$. Going further, we have that $(\tilde{h}A)(v)$ depends only on the values of h at the neighbors of $\pi_i(v)$ in G_i for any $v \in V$. A similar argument works if we consider several functions at the same time, and we get that an element of $S_k(A(G_i)/d_i)$ corresponding to $h_1, \ldots, h_k$ (and $A(G_i)$) is the same as the entry distribution of $(\tilde{h}_1, \ldots, \tilde{h}_k, \tilde{h}_1 A, \ldots, \tilde{h}_k A)$. This implies that $\cup_{i=1}^\infty S_k(A(G_i)/d_i)$ is contained by $\mathcal{S}_k(A)$, and hence the closure of $\cup_{i=1}^\infty S_k(A(G_i)/d_i)$ is contained by $\mathcal{S}_k^*(A)$. For the other direction, let $1 \le j \le k$, and notice that for every $\overline{h}_j \in L^2(V, \mu)$ and $\varepsilon > 0$ we can find a large enough i and $\tilde{h}_j \in L^2(V, \mu)$, which depends only on $\pi_i(v)$ and which satisfies $\|\overline{h}_j - \tilde{h}_j\|_2 \le \varepsilon$. As $\tilde{h}_j$ is the lift of an appropriate $h_j : V(G_i) \to \mathbb{R}$, the same argument as above shows that the element of $\mathcal{S}_k(A)$ corresponding to $(\tilde{h}_1, \ldots, \tilde{h}_k)$ is the same as the element of $\mathcal{S}_k(A(G_i))$ corresponding to $(h_1, \ldots, h_k)$. Then, we can finish the proof by using Lemma 13.2 and the condition of the distance of $\overline{h}_j$ and $\tilde{h}_j$, as it follows that $\mathcal{S}_k^*(A)$ is contained by the closure of $\cup_{i=1}^\infty S_k(A(G_i)/d_i)$.   ∎

Construction of the limiting hypercube: We finally arrived to the construction, which allows us to determine the limit of the sequence $\{Q_{2^n}\}_{n=1}^\infty$. The main observation is that there are $(2^{2^n}, 2)$-uniform maps $f_n : V(Q_{2^{n+1}}) \to V(Q_{2^n})$. Let T_n denote the vertex set of the rooted binary tree of depth n. If $n = \infty$, then T_∞ denotes the infinite rooted binary tree. We have that T_n has 2^n leaves. If v is not a leaf, then we denote by $\alpha_1(v)$ and $\alpha_2(v)$ the two children of v. Recall that Z_2 denotes the group with two elements. For $n \in \mathbb{N} \cup \{\infty\}$ let G_n be the set of functions $f : T_n \to Z_2$ such that $f(v) = f(\alpha_1(v)) + f(\alpha_2(v))$. It is clear that G_n is an elementary abelian 2-group of order 2^{2^n} with respect to pointwise addition. It follows that G_n is a vector space of dimension 2^n over the field with two elements. The group G_∞ is the inverse limit of the groups G_n and it is a compact abelian group with Haar measure μ, as the argument before Definition 12.3 shows. Let B denote the boundary of T_∞. It is well known (see, e.g., [30, Section 1.10]) that B is the Cantor set and every element $b \in B$ is uniquely characterized by an infinite path started at the root of T_∞. By abusing the notation let us identify b with this infinite path. Let g_b denote the element in $Z_2^{T_\infty}$ that takes 1 at the vertices of the path b and 0 otherwise. It is clear that $g_b \in G_\infty$ holds for every $b \in B$. Let $Q := \{g_b : b \in B\}$ and let ν be the probability measure on Q obtained by first choosing b uniformly in the Cantor set B and then taking g_b. For $n \in \mathbb{N} \cup \{\infty\}$ and $m \le n$ we denote by $\pi_{n,m} : G_n \to G_m$ the group homomorphism obtained by restricting a Z_2-labeling of T_n to the subtree T_m. It is easy to see that $\pi_{\infty,n}(Q)$ is a basis in the vector space G_n and thus we can represent Q_{2^n} as the Cayley graph of G_n with generators $\pi_{\infty,n}(Q)$. It is easy to see that the maps

$$\pi_{n+1,n} : V(Q_{2^{n+1}}) \to V(Q_{2^n})$$

are $(2^{2^n}, 2)$ uniform.

It follows from Theorem 12.2 that the limit object of the sequence $\{Q_{2^n}\}_{n=1}^\infty$ is basically the Cayley graph of the compact group $G_\infty \simeq Z_2^\infty$ with generators Q and with uniform

measure on the edges. More precisely, let A denote the P-operator in $\mathcal{B}_{2,2}(G_\infty, \mu)$ defined by

$$(fA)(x) := \int_{z \in Q} f(x + z) \, dv.$$

Then the P-operator A is the limit of the graph sequence $\{Q_{2^n}\}$.

We conclude with some further remarks on the convergence of hypercubes. Similarity of large dimensional hypercubes is interesting even without knowing the limit object. We only establish it for power of two dimensions. Note that it is true in general but it requires an asymptotic version of our covering techniques. This extension of our proof is more technical but the main idea is essentially the same. Our results imply that various extra structures on hypercubes behave similarly in different large dimensions. For example, this includes entry statistics of almost eigenvectors even in the simultaneous setting where many almost eigenvectors are viewed together. (For a more precise definition of this problem see [2].) We emphasize that our results on hypercubes serve more as a theoretical foundation for a large class of questions rather than a complete description of their properties. Any problem that is expressible in terms of the sets $\mathcal{S}_k$ can now be reduced to the study of the limit object, which is a natural algebraic structure. However, this direction is new and needs to be exploited.

12.2 Product graphs

The product of two graphs, G_1 and G_2, is the graph on $V(G_1) \times V(G_2)$ such that $((i, j), (k, l)) \in E(G_1 \times G_2)$ if and only if $(i, k) \in E(G_1)$ and $(j, l) \in E(G_2)$. Graph sequences formed by the powers of a given graph are good test graphs for limit theories. We have that $2|E(G_1 \times G_2)| = 4|E(G_1)||E(G_2)|$ and thus $E(G^i) = 2^{i-1}|E(G)|^i$. It follows that

$$\beta := \lim_{i \to \infty} \log|E(G^i)|/\log|V(G^i)| = \log(2|E(G)|)/\log(|V(G)|).$$

The number $0 \le \beta \le 2$ expresses the exponent of the growth rate of the number of edges in terms of the number of vertices in $\{G^i\}_{i=1}^\infty$. One can view G^i as a fractal-like graph (see Figure 6).

When G is d-regular, we can use Theorem 12.2 to compute the limit object of $\{G^i\}_{i=1}^\infty$. The main observation is that the map $\pi_i : V(G^{i+1}) \to V(G^i)$ given by the projection to the first i coordinates is uniform and thus $\{G_i, \pi_i\}_{i=1}^\infty$ is a uniform tower. It is easy to see that the inverse limit is simply given as the infinite power G^∞ with the uniform distribution on the vertices and on the edges. According to Theorem 12.2, this inverse limit is basically the limit of the normalized P-operator sequence $\{A(G^i)/d^i\}_{i=1}^\infty$. The corresponding graphop $A \in \mathcal{B}(V(G)^\infty)$ is given by

$$(vA)(x) = \mathbb{E}_{(x,y) \in E(G^\infty)} v(y),$$

where the expected value is calculated according to the product measure on the neighbors of x. More precisely, if $x = (x_1, x_2, \dots) \in V(G^\infty)$ is fixed, then the set of neighbors of x is the infinite product $N(x_1) \times N(x_2) \times \cdots$. We define v_x as the product of the uniform measures on $N(x_1) \times N(x_2) \times \cdots$ and the expected value is according to the measure v_x.
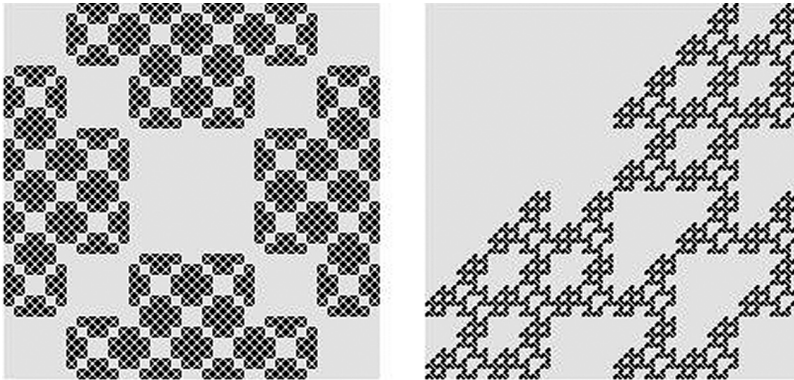
Figure 6: The 3rd power of two different graphs on six vertices.

If G is not regular, then the degrees in G^i are very unevenly distributed. In this case, we can use random walk convergence to get a nontrivial and natural limit object, but we skip the details.

12.3 Star graphs

For every $n \geq 1$, let G_n be the star graph on vertex set $\Omega_n = \{0, 1, \ldots, n-1\}$, namely, in which vertex 0 is connected to every other vertex with a single edge. Since the operator norm of its adjacency matrix is $\sqrt{n-1}$, we should normalize by $\sqrt{n}$ to get a sequence of matrices with bounded operator norm. In this case, the limit will be constant 0, which does not reflect the structure of the graphs. Therefore, instead of the adjacency operator convergence notion, we are interested in the random walk convergence of this sequence, as it was defined in Section 5.

Let v_n be the stationary measure of the random walk on G_n. This puts weight $1/2$ to vertex 0, and $1/(2n-2)$ everywhere else. Then, the Markov operator $M_n := M(G_n)$, which acts on $L^2(\Omega_n)$, is given as follows:

$$(vM_n)(0) = \frac{1}{n-1} \sum_{i=1}^{n-1} v(i); \quad (vM_n)(j) = v(0) \quad (j = 1, \ldots, n-1),$$

where $v \in L^2(\Omega_n)$. Hence the k-profile of M_n consists of the following probability measures on $\mathbb{R}^2$, where $v_1, \ldots, v_k \in L^\infty_{[-1,1]}(\Omega_n)$ are chosen arbitrarily: the measure puts weight $1/2$ to

$$\left(v_1(0), \ldots, v_k(0), \frac{1}{n-1} \sum_{i=1}^{n-1} v_1(i), \ldots, \frac{1}{n-1} \sum_{i=1}^{n-1} v_k(i) \right),$$

and it puts weight $1/(2n-2)$ to

$$\left(v_1(j), \ldots, v_k(j), v_1(0), \ldots, v_k(0) \right)$$

for each $j = 1, 2, \ldots, n-1$.

Now, let $\Omega = [0, 1/2]$ with the following probability measure v: it is the Lebesgue measure on $(0, 1/2]$ together with an atom of weight $1/2$ at 0. We define a P-operator M on $L^2(\Omega)$ by

$$(fM)(0) = 2 \int_0^{1/2} f(y)dy; \qquad (fM)(x) = f(0) \quad (0 < x \le 1/2),$$

where $f \in L^2(\Omega)$. Then the k-profile of the P-operator M is the set of the following probability measures for $f_1, \ldots, f_k \in L^2(\Omega)$: the measure puts weight $1/2$ to

$$\left(f_1(0), \ldots, f_k(0), 2 \int_0^{1/2} f_1(y)dy, \ldots, 2 \int_0^{1/2} f_k(y)dy \right),$$

and puts a uniform distribution (with total weight $1/2$) to

$$\left(f_1(x), \ldots, f_k(x), f_1(0), \ldots, f_k(0) \right).$$

By comparing the profiles of M_n and M, for every $k \ge 1$, we have that $\mathcal{S}_k(M_n) \subset \mathcal{S}_k(M)$. On the other hand, we show that every element of $\mathcal{S}_k(M)$ can be approximated weakly by a sequence whose nth term is chosen from $\mathcal{S}_k(M_n)$. For every $m \ge 1$ we can choose continuous functions $f_1^{(m)}, \ldots, f_k^{(m)} \in L^2(\Omega)$ such that the L^2-distance of f_s and $f_s^{(m)}$ is at most $1/m$ for every $s \in [k]$. Furthermore, if m_n is large enough, then by choosing $v_s(j) = f_s(j/m_n)$, we can find an element of $\mathcal{S}_k(M_{m_n})$ whose Lévy–Prokhorov distance from the probability measure corresponding to $f_1, \ldots, f_k$ in $\mathcal{S}_k(M)$ is arbitrarily small. We conclude that the Hausdorff distance of $\mathcal{S}_k(M_n)$ and $\mathcal{S}_k(M)$ tends to 0 as $n \to \infty$ for each $k \ge 1$, and M is the limit of the sequence of star graphs with respect to random walk convergence.

12.4 Subdivisions of complete graphs

Our second example is the 2-subdivision of the complete graph on n vertices. More precisely, for $n \ge 1$, let

$$\Omega_n = [n] \cup \{w_{ij}, 1 \le i < j \le n\}.$$

When $j < i$, we will use $w_{ji} = w_{ij}$. As for the edges, for every $1 \le i < j \le n$, vertex w_{ij} is connected to i and j. This graph has $n + n(n-1)/2$ vertices and $n(n-1)$ edges.

We denote by M_n the Markov operator of this graph. For every $v : \Omega_n \to \mathbb{R}$ and $1 \le i < j \le n$, we have

$$(vM_n)(i) = \frac{1}{n-1} \sum_{j \ne i} v(w_{ij}); \qquad (vM)(w_{ij}) = \frac{1}{2}\left(v(i) + v(j) \right).$$

The stationary measure puts weight $1/(2n)$ to vertices from $[n]$, and weight $1/(n^2 - n)$ to the other vertices. Hence the k-profile of M_k is given by the set of probability measures putting weight $1/(2n)$ to

$$\left(v_1(i), \ldots, v_k(i), \frac{1}{n-1} \sum_{j \ne i} v_1(w_{ij}), \ldots, \frac{1}{n-1} \sum_{j \ne i} v_k(w_{ij}) \right),$$

and weight $1/(n^2 - n)$ to

$$\left(v_1(w_{ij}), \ldots, v_k(w_{ij}), \frac{1}{2}\big(v_1(i) + v_1(j)\big), \ldots, \frac{1}{2}\big(v_k(i) + v_k(j)\big) \right),$$

where $v_1, \ldots, v_k : \Omega_n \to \mathbb{R}$ are arbitrary functions.

Let $[0,1]^2_\sim$ denote the set of unordered pairs $\{(x,y) : x, y \in [0,1]\}$. In other words, $[0,1]^2_\sim$ is the set $[0,1]^2$ factored by the equivalence $(x,y) \sim (y,x)$. We represent the limit as a P-operator on $\Omega = [0,1] \cup [0,1]^2_\sim$. In this case, a function $f \in L^2(\Omega)$ can be given by a pair (f_1, f_2), where $f_1 \in L^2([0,1])$ and $f_2 \in L^2([0,1]^2_\sim)$. Then we define M as follows:

$$(fM)(x) = \int_0^1 f_2(x, u)\, du; \qquad (fM)(y, z) = \frac{f_1(y) + f_1(z)}{2},$$

where x, y, z are all from the interval $[0,1]$. The k-profile of M consists of probability measures, which are the distributions of the following random variables for some functions $f^{(1)}, \ldots, f^{(k)} \in L^\infty_{[-1,1]}(\Omega)$. With probability $1/2$, we choose x uniformly at random from the interval $[0,1]$ and take

$$\left(f_1^{(1)}(x), \ldots, f_1^{(k)}(x), \int_0^1 f_2^{(1)}(x, u)\, du, \ldots, \int_0^1 f_2^{(k)}(x, u)\, du \right).$$

Otherwise, we choose $(y, z) \in [0,1]^2$ uniformly at random, and take

$$\left(f_2^{(1)}(y, z), \ldots, f_2^{(k)}(y, z), \frac{1}{2}\big(f_1^{(1)}(y) + f_1^{(1)}(z)\big), \ldots, \frac{1}{2}\big(f_1^{(k)}(y) + f_1^{(k)}(z)\big) \right).$$

Similarly, to the previous case, by approximating L^2 functions with continuous ones, it can be proved that M is indeed the limit of M_n, and hence the sequence of 2-subdivisions of complete graphs converges to this P-operator according to random walk convergence.

12.5 Incidence graphs of finite projective planes

Let q be a prime power and let $\mathbb{P}(q)$ denote the projective plane over the finite field with q elements. The plane $\mathbb{P}(q)$ has $q^2 + q + 1$ lines and $q^2 + q + 1$ points. We denote by G_q the bipartite graph, whose vertices are the lines and the points in $\mathbb{P}_q$ and the edges in G_q are incidences in $\mathbb{P}(q)$. This means that a line l is connected with a point p if l contains p. We have that G_q is $(q+1)$-regular, $|V(G_q)| = 2(q^2 + q + 1)$ and $|E(G_q)| = (q^2 + q + 1)(q + 1)$. It follows that the sequence G_q is an intermediate density sequence. The number of edges is roughly the 3/2th power of the number of vertices.

We show that the matrices $A(G_q)/(q+1)$ form a convergent sequence, whose limit is represented by the generalized graphon $W : [0,1]^2 \to [0,2]$ defined by $W(x,y) = 2$ if $(x,y) \in [1/2, 1] \times [0, 1/2] \cup [0, 1/2] \times [1/2, 1]$ and $W(x,y) = 0$ elsewhere.

The proof is based on the fact that the eigenvalues of G_q are known to be $q + 1, -q - 1, \sqrt{q}, -\sqrt{q}$ with multiplicity $1, 1, q^2 + q, q^2 + q$. The two eigenvalues $q + 1$ and $-q - 1$ belong to the constant 1 vector v_1 and the vector v_2, which takes 1 at points and -1 at lines. Let $B_q := (v_1^* v_1 - v_2^* v_2)/2(q^2 + q + 1)$. We have that $\|A(G_q)/(q+1) - B_q\|_{2 \to 2} =$
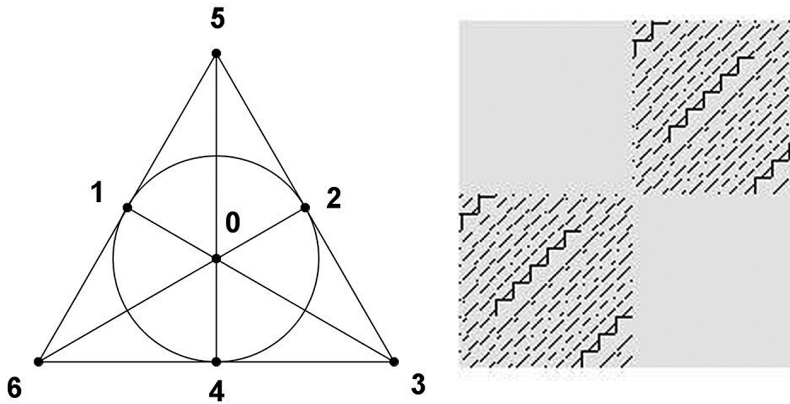
Figure 7: The Fano plane $\mathbb{P}(2)$ and the matrix of the incidence graph of the projective plane over the field $\mathbb{F}_9$.

$q^{-1/2}$. It follows from Lemma 2.13 that

$$d_M(A(G_q)/(q+1), B_q) \leq 3\|A(G_q)/(q+1) - B_q\|_{2\to2}^{1/2} = 3q^{-1/4}$$

and hence the limit $A(G_q)/(q+1)$ is the same as the limit of B_q as the prime power q goes to infinity. On the other hand B_q is twice the normalized adjacency matrix of the complete bipartite graph with equal color classes on $2(q^2 + q + 1)$ points. This proves the claim.

The next question illustrates that this does not end the limiting investigation of G_q. We can look at it at a finer scale by removing the two dominant eigenvectors and normalizing it with a different constant (Figure 7).

Question 12.3 Let $B_q' := (A(G_q)/(q+1) - B_q)q^{1/2}$. We have that $\|B_q'\|_{2\to2} = 1$. Does the sequence of P-operators B_q' converge as the prime power q goes to infinity? If yes, what is the limit object?

Note that by compactness we know that B_q' has convergent subsequences.

13 Appendix (technical lemmas)

Recall from equation (2.1) that $\mathcal{D}(v)$ denotes the empirical distribution of a vector, while (2.4) says that for a probability measure μ on $\mathbb{R}^k$ we defined $\tau(\mu) \in [0, \infty]$ by

$$\max_{1\leq i\leq k} \int_{(x_1, x_2, \ldots, x_k)\in\mathbb{R}^k} |x_i| \, d\mu.$$

Lemma 13.1 Let X, Y be two jointly distributed $\mathbb{R}^k$-valued random variables. Then $d_{\mathrm{LP}}(\mathcal{D}(X), \mathcal{D}(Y)) \leq \tau(X - Y)^{1/2}k^{3/4}$.

Proof First, we claim that for $a > 0$ we have

(13.1) $$\mathbb{P}(|X - Y| \geq a) \leq \tau(X - Y)k^{3/2}/a.$$

Let $\pi_i : \mathbb{R}^k \to \mathbb{R}$ denote the ith coordinate function for $1 \le i \le k$. If the square of every coordinate of $X - Y$ is at most a^2/k, then $|X - Y| \le a$. Hence by the union bound we have that

$$\mathbb{P}(|X - Y| \ge a) \le \sum_{i=1}^{k} \mathbb{P}(|\pi_i(X - Y)| \ge a/k^{1/2}) \le \sum_{i=1}^{k} \mathbb{E}(|\pi_i(X - Y)|)k^{1/2}/a.$$

Since $\tau(X - Y)$ is the maximum of $\mathbb{E}(|\pi_i(X - Y)|)$ over $1 \le i \le k$, inequality (13.1) follows.

Let U be a Borel set in $\mathbb{R}^k$. From (13.1) we have for every $\varepsilon > 0$ that

(13.2) $\mathbb{P}(Y \in U^\varepsilon) \ge \mathbb{P}(X \in U) - c$ and $\mathbb{P}(X \in U^\varepsilon) \ge \mathbb{P}(Y \in U) - c$,

where $c = \tau(X - Y)k^{3/2}\varepsilon^{-1}$. In particular, if $\varepsilon = \tau(X - Y)^{1/2}k^{3/4}$, then $c = \varepsilon$ and (13.2) witnesses that $d_{\mathrm{LP}}(\mathcal{D}(X), \mathcal{D}(Y)) \le \tau(X - Y)^{1/2}k^{3/4}$. ∎

Lemma 13.2   Let $v_1, v_2, \ldots, v_k$ and $w_1, w_2, \ldots, w_k$ be in $L^1(\Omega)$ for some probability space Ω. Let $m := \max_{i \in [k]} \|v_i - w_i\|_1$. Then

$$d_{\mathrm{LP}}(\mathcal{D}(v_1, v_2, \ldots, v_k), \mathcal{D}(w_1, w_2, \ldots, w_k)) \le m^{1/2}k^{3/4}.$$

Proof   We apply Lemma 13.1 to the jointly distributed random variables $X(\omega) := (v_1(\omega), v_2(\omega), \ldots, v_k(\omega))$ and $Y(\omega) := (w_1(\omega), w_2(\omega), \ldots, w_k(\omega))$ defined for $\omega \in \Omega$. Since $\tau(X - Y) = m$, Lemma 13.1 finishes the proof. ∎

For a real number $z \in \mathbb{R}^+$ let $q_z : \mathbb{R} \to \mathbb{R}$ denote the function such that $f_z(x) = 0$ for $|x| \ge 2z$, $f_z(x) = |x + z| - z$ for $x \in [-2z, 0]$ and $f_z(x) = -|x - z| + z$ for $x \in [0, 2z]$.

Lemma 13.3   Let $q \in (1, \infty)$ and let X be a real-valued random variable with $\mathbb{E}(|X|^q) = c < \infty$. Then for $z \in \mathbb{R}^+$ we have that $\mathbb{E}|f_z(X) - X| \le cz^{1-q}$.

Proof   Let $p = q/(q - 1)$. We have that $f_z(x) - x$ is 0 for $x \in [-z, z]$ and $|f_z(x) - x| \le |x|$ for $x \in \mathbb{R} \setminus [-z, z]$. It follows from Hölder's inequality that

$$\mathbb{E}|f_z(X) - X| \le \mathbb{E}(|X|\mathbf{1}_{\mathbb{R} \setminus [-z,z]}(X)) \le \mathbb{E}(|X|^q)^{1/q}\mathbb{E}(\mathbf{1}_{\mathbb{R} \setminus [-z,z]})^{1/p}$$
$$= \mathbb{E}(|X|^q)^{1/q}\mathbb{P}(|X| \ge z)^{1/p} \le c^{1/q}\mathbb{P}(|X| \ge z)^{1/p}.$$

By Markov's inequality, we have that

$$\mathbb{P}(|X| \ge z) = \mathbb{P}(|X|^q \ge z^q) \le \mathbb{E}(|X|^q)/z^q \le c/z^q.$$

This completes the proof.   ∎

Lemma 13.4   Let $q \in (1, \infty)$. Let $\{(X_i, Y_i)\}_{i=1}^{\infty}$ be a sequence of pairs of jointly distributed real-valued random variables such that $X_i \in [-1, 1]$ and $\mathbb{E}(|Y_i|^q) \le c < \infty$ for some $c \in \mathbb{R}^+$. Assume that the distributions of (X_i, Y_i) weakly converge to some probability distribution (X, Y) as i goes to infinity. Then $\mathbb{E}(|Y|^q) \le c$ and

$$\lim_{i \to \infty} \mathbb{E}(X_i Y_i) = \mathbb{E}(XY).$$

Proof   The statement $\mathbb{E}(|Y|^q) \le c$ follows from the compactness of $\{\mu : \mu \in \mathcal{P}(\mathbb{R}), \int_x |x|^q d\mu \le c\}$ in the weak topology. Since f_z is continuous with finite support,

we have that $\lim_{i \to \infty} \mathbb{E}(f_1(X_i) f_z(Y_i)) = \mathbb{E}(f_1(X) f_z(Y))$ holds for every $z \in \mathbb{R}^+$. On the other hand, we have that

$$|\mathbb{E}(f_1(X_i) f_z(Y_i) - X_i Y_i)| = |\mathbb{E}(X_i(f_z(Y_i) - Y_i))| \le \mathbb{E}|f_z(Y_i) - Y_i| \le cz^{(1-q)}$$

by Lemma 13.3, and similarly

$$|\mathbb{E}(f_1(X) f_z(Y) - XY)| \le cz^{(1-q)}.$$

It follows that

$$\left| \lim_{i \to \infty} \mathbb{E}(X_i Y_i) - \mathbb{E}(XY) \right| \le 2cz^{(1-q)}$$

and hence as z goes to infinity, we obtain the statement of the lemma. ∎

Lemma 13.5 *Let μ be a probability measure on $[-c, c]$ for some $c \in \mathbb{R}^+$. Let $p \in [1, \infty)$. Then $\int_{\mathbb{R}} |x|^p \, d\mu \le (2d_{\mathrm{LP}}(\mu, \delta_0))^p + 2d_{\mathrm{LP}}(\mu, \delta_0) c^p$*

Proof Let $d := d_{\mathrm{LP}}(\mu, \delta_0)$. We have that $1 = \delta_0(\{0\}) \le \mu([-2d, 2d]) + 2d$ and so $\mu([-2d, 2d]) \ge 1 - 2d$. It follows that

$$\int_{\mathbb{R}} |x|^p \, d\mu = \int_{[-2d, 2d]} |x|^p \, d\mu + \int_{\mathbb{R} \setminus [-2d, 2d]} |x|^p \, d\mu \le (2d)^p + 2dc^p.$$

∎

Lemma 13.6 *Let $p \in [1, \infty)$ and let $A \in \mathcal{B}(\Omega)$ be a P-operator. Let v_i and w_i be elements in $L^\infty(\Omega)$ with values in $[-1, 1]$ for $i \in [k]$. Then, we have*

$$d_{\mathrm{LP}}(\mathcal{D}_A(\{v_i\}_{i=1}^k), \mathcal{D}_A(\{w_i\}_{i=1}^k)) \le m^{1/2}((2d)^p + 2^{p+1}d)^{1/(2p)}(2k)^{3/4},$$

where $m = \max\{1, \|A\|_{p \to 1}\}$ and $d = \max_{i \in [k]}\{d_{\mathrm{LP}}(\mathcal{D}(v_i - w_i), \delta_0)\}$.

Proof We have by Lemma 13.5 that $\|v_i - w_i\|_p \le ((2d)^p + 2^{p+1}d)^{1/p}$ for every $i \in [k]$. It follows that $\|v_i A - w_i A\|_p \le \|A\|_{p \to 1}((2d)^p + 2^{p+1}d)^{1/p}$ holds for every $i \in [k]$. Then, Lemma 13.2 finishes the proof. ∎

References

[1] Á. Backhausz and B. Szegedy, *On large girth regular graphs and random processes on trees.* Random Struct. Algor. **53**(2018), no. 3, 389–416. http://doi.org/10.1002/rsa.20769

[2] Á. Backhausz and B. Szegedy, *On the almost eigenvectors of random regular graphs.* Ann. Probab. **47**(2019), no. 3, 1677–1725. http://doi.org/10.1214/18-AOP1294

[3] I. Benjamini and N. Curien, *Ergodic theory on stationary random graphs.* Electron. J. Probab. **17**(2012), no. 93, 20 pp. http://doi.org/10.1214/EJP.v17-2401

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

[5] N. Berger, C. Borgs, J. T. Chayes, and A. Saberi, *Asymptotic behavior and distributional limits of preferential attachment graphs.* Ann. Probab. **42**(2014), no. 1, 1–40. http://doi.org/10.1214/12-AOP755

[6] C. Borgs, J. T. Chayes, H. Cohn, and N. Holden, *Sparse exchangeable graphs and their limits via graphon processes.* J. Mach. Learn. Res. **18**(2018), 1–71.

[7] C. Borgs, J. T. Chayes, H. Cohn, and Y. Zhao, *An L^p theory of sparse graph convergence I: limits, sparse random graph models, and power law distributions*. Trans. Am. Math. Soc. 372(2019), no. 5, 3019–3062. http://doi.org/10.1090/tran/7543

[8] C. Borgs, J. T. Chayes, H. Cohn, and Y. Zhao, *An L^p theory of sparse graph convergence II: LD convergence, quotients, and right convergence*. Ann. Probab. 46(2018), no. 1, 337–396. http://doi.org/10.1214/17-AOP1187

[9] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi, *Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing*. Adv. Math. 219(2008), no. 6, 1801–1851. http://doi.org/10.1016/j.aim.2008.07.008

[10] C. Borgs, J. Chayes, and L. Lovász, *Moments of two-variable functions and the uniqueness of graph limits*. Geom. Funct. Anal. 19(2010), no. 6, 1597–1619. http://doi.org/10.1007/s00039-010-0044-0

[11] B. Bollobás and O. Riordan, *Sparse graphs: metrics and random models*. Random Struct. Algorithms 39(2011), 1–38. http://doi.org/10.1002/rsa.20334

[12] P. Candela and B. Szegedy, *Nilspace factors for general uniformity seminorms, cubic exchangeability and limits*. Preprint, 2018. http://arxiv.org/pdf/1803.08758.pdf

[13] B. K. Driver, *Analysis tools with examples*. Lecture Notes, Springer, New York, NY, 2003. http://www.math.ucsd.edu/~bdriver/240A-C-03-04/LectureNotes/anal2p-new.pdf.

[14] G. Elek, *On the limit of large girth graph sequences*. Combinatorica 30(2010), no. 5, 553–563. http://doi.org/10.1007/s00493-010-2559-2

[15] G. Elek and E. Szabó, *Sofic groups and direct finiteness*. J. Algebra 280(2004), no. 2, 426–434. http://doi.org/10.1016/j.jalgebra.2004.06.023

[16] L. Erdős, A. Knowles, H. T. Yau, and J. Yin, *Spectral statistics of Erdős-Rényi graphs I: local semicircle law*. Ann. Probab. 41(2013), no. 3B, 2279–2375. http://doi.org/10.1214/11-AOP734

[17] D. H. Fremlin, *Measure theory*. Vol. 4, Torres Fremlin, Colchester, 2006.

[18] P. E. Frenkel, *Convergence of graphs with intermediate density*. Trans. Am. Math. Soc. 370(2018), no. 5, 3363–3404. http://doi.org/10.1090/tran/7036

[19] S. Geman, *A limit theorem for the norm of random matrices*. Ann. Probab. 8(1980), no. 2, 252–261.

[20] R. A. Gordon, *Real analysis: A first course*. Addison Wesley Higher Mathematics, Reading, MA, 1997.

[21] H. Hatami, L. Lovász, and B. Szegedy, *Limits of locally-globally convergent graph sequences*. Geom. Funct. Anal. 24(2014), no. 1, 269–296. http://doi.org/10.1007/s00039-014-0258-7

[22] M. Hazelwinkel (ed.), *Encyclopaedia of mathematics*. Vol. 4, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1989. http://doi.org/10.1007/978-94-009-6000-8

[23] S. Janson, *Graphons and cut metric on σ-finite measure spaces*. Preprint, 2016. http://arxiv.org/pdf/1608.01833.pdf

[24] G. Kun and A. Thom, *Inapproximability of actions and Kazhdan's property (T)*. Preprint, 2019. https://arxiv.org/abs/1901.03963

[25] D. Kunszenti-Kovács, L. Lovász, and B. Szegedy, *Multigraph limits, unbounded kernels, and Banach space decorated graphs*. Preprint, 2014. arXiv:1406.7846

[26] D. Kunszenti-Kovács, L. Lovász, and B. Szegedy, *Measures on the square as sparse graph limits*. J. Combin. Theory Ser. B 138(2019), 1–40. http://doi.org/10.1016/j.jctb.2019.01.004

[27] L. Lovász, *Large networks and graph limits*. Amer. Math. Soc., 60, American Mathematical Society Colloquium Publications, Providence, RI, 2012.

[28] L. Lovász and B. Szegedy, *Limits of dense graph sequences*. J. Combin. Theory Ser. B 96(2006), no. 6, 933–957. http://doi.org/10.1016/j.jctb.2006.05.002

[29] L. Lovász and B. Szegedy, *Szemerédi's lemma for the analyst*. Geom. Funct. Anal. 17(2007), no. 1, 252–270. http://doi.org/10.1007/s00039-007-0599-6

[30] R. Lyons and Y. Peres, *Probability on trees and networks*. Cambridge Series in Statistical and Probabilistic Mathematics, 42, Cambridge University Press, New York, NY, 2016. http://doi.org/10.1017/9781316672815

[31] C. Male and S. Péché, *Uniform regular weighted graphs with large degree: Wigner's law, asymptotic freeness and graphons limit*. Preprint, 2014. http://arxiv.org/pdf/1410.8126.pdf

[32] J. Nešetřil and P. Ossona de Mendez, *Sparsity, algorithms and combinatorics*. Vol. 28, Springer, Heidelberg, Germany, 2012. http://doi.org/10.1007/978-3-642-27875-4

[33] J. Nešetřil and P. Ossona de Mendez, *Local-global convergence, an analytic and structural approach*. Preprint, 2018. http://arxiv.org/pdf/1805.02051.pdf

[34] A. Rudas, B. Tóth, and B. Valkó, *Random trees and general branching processes*. Random Struct. Algor. 31(2007), no. 2, 186–202. http://doi.org/10.1002/rsa.20137

[35] B. Szegedy, *Sparse graph limits, entropy maximization and transitive graphs*. Preprint, 2015. arXiv:1504.00858

[36] B. Szegedy, *On the colored star metric*. Manuscript in preparation.

[37] Y. Zhu, *A graphon approach to limiting spectral distributions of Wigner-type matrices*. Random Struct. Algor. **56**(2020), no. 1, 251–279. http://doi.org/10.1002/rsa.20894

Department of Probability Theory and Statistics, ELTE Eötvös Loránd University, Budapest, Hungary
and

Alfréd Rényi Institute of Mathematics, Pázmány Péter sétány 1/c, H-1117 Budapest, Hungary
e-mail: agnes.backhausz@ttk.elte.hu

Alfréd Rényi Institute of Mathematics, Reáltanoda utca 13–15, H-1053 Budapest, Hungary
e-mail: szegedy.balazs@renyi.hu