IOP Publishing | London Mathematical Society Nonlinearity

Nonlinearity **33** (2020) 1404–1423 https://doi.org/10.1088/1361-6544/ab60d2

Generalized quasirandom properties of expanding graph sequences

Marianna Bolla

Institute of Mathematics, Budapest University of Technology and Economics, Műegyetem rkp. 3. H-1111 Budapest, Hungary

E-mail: marib@math.bme.hu

Received 30 November 2018, revised 14 October 2019
Accepted for publication 11 December 2019
Published 13 February 2020

Recommended by Dr Mark F Demers

Abstract

We consider multiclass spectral and discrepancy properties of expanding graph sequences. Unlike the linear nature of spectral characteristics, multiway discrepancy has a nonlinear dependence on the weighted adjacencies of the graph. As we are able to prove equivalences and implications between them and the definition of the generalized quasirandomness of Lovász–Sós (2008), they can be regarded as generalized quasirandom properties akin to the equivalent quasirandom properties of the seminal Chung–Graham–Wilson paper (1989) in the one-class scenario. Since these properties are valid for certain deterministic graph sequences, irrespective of stochastic models, the partial implications also justify for low-dimensional embedding of large-scale graphs and for discrepancy minimizing spectral clustering.

Keywords: generalized quasirandom graphs, multiway discrepancy, normalized modularity spectra, cluster variances
Mathematics Subject Classification numbers: 05C50, 05C60, 05C80, 05C90, 62H30

1. Introduction

The classical linear methods of the multivariate statistical analysis were developed to facilitate the basic idea of factor analysis (static or dynamic) and to make low-dimensional embedding of high-dimensional data. For this purpose, the so-called k-factor model is used, where k is the hidden rank of the covariance structure and it is much less than the number n of the variables. The so-called factor scores also give rise to partition the variables into k clusters. In the fundamental models, k is fixed, whereas $n \to \infty$ together with the sample size N in the static or the time T in the dynamic case. If our data form a graph, then n is the number of vertices,

N is the number of edges, and the usual condition that n and N tend to infinity in a prescribed way (N is superlinear in n) implies that our graph is dense enough. Given a real-life graph on n vertices, that is an instance of an expanding sequence, with the help of graph based matrices, our purpose is to make inference on the number k of the hidden clusters of vertices and to find the clusters themselves.

Graph based matrices together with eigenvalues and eigenvectors have been intensively studied since the 1970s. Hoffman [26] and Fiedler [24] used the eigenvector, corresponding to the smallest positive Laplacian eigenvalue of a connected graph, to find a bipartition of the vertices which approximates the *minimum cut problem*. From the two-clustering point of view ($k = 2$), this eigenvector becomes important when the corresponding eigenvalue is not separated from the trivial zero eigenvalue, but it is separated from the other positive Laplacian eigenvalues. On the contrary, when there is a large spectral gap between the trivial zero and the smallest positive Laplacian eigenvalue (or equivalently, between the trivial 1 and the second largest positive eigenvalue of the transition probability matrix in the random walk view), there is no use of partitioning the vertices, the whole graph forms a highly connected cluster. This $k = 1$ case has frequently been studied since Cheeger [18], establishing a lot of equivalent or near equivalent advisable features of these graphs. There are many results about the relation between this gap and different kinds of expansion constants of the graph (see e.g. [28, 31]), including random walk view of [5, 22, 34]. The vertex subsets of such graphs have a large boundary compared to their volumes characterized by the isoperimetric number, see Mohar [35]. They also show quasirandom properties discussed in Thomason [44, 45], Bollobás [14], and Chung, Graham, Wilson [19, 20]. For these favorable characteristics, they are indispensable in communication networks.

However, less attention has been paid to graphs with a small spectral gap, when several cases can occur: among others, the graph can be a bipartite expander of Alon [2] or its vertices can be divided into two sparsely connected clusters, but the clusters themselves can be good expanders (see [30] and [38]). In case of several clusters of vertices the situation is even more complicated. The pairwise relations between the clusters and the within-cluster relations of the vertices of the same cluster show a great variety. Depending on the number and sign of the so-called *structural eigenvalues* of the *normalized modularity matrix*, to be defined in section 2, we can make inferences on the number of the underlying clusters and the type of connection between them. Furthermore, based on *spectral and singular value decompositions*, low dimensional embedding of the vertices is performed, and classical and modern techniques of the multivariate statistical analysis are used to find the clusters.

The notion of the *multiway discrepancy*, introduced in [12], has a nonlinear nature and cannot be treated with classical linear algebra tools. However, it plays a crucial role in identifying the clusters as it is related to the spectral properties of the graph. In the worst case scenario, when there are no k structural eigenvalues with a moderate k, the Szemerédi regularity lemma guarantees the existence of a universal k (independent of n, it only depends on the discrepancy bound to be attained) such that the vertices can be classified into k (equitable, and a 'small' exceptional) parts such that the between-cluster discrepancies are less than the error bound. This theorem has overwhelming theoretical importance, and also spectral and matrix versions, see Szegedy [42] and Frieze and Kannan [25], whereas our purpose is to give equivalent conditions for the existence of a k-cluster structure with a moderate k. For this purpose, in section 3 we consider k-class (generalized) random and quasirandom graphs, introduced in [32], and in section 3 we establish equivalent properties (including discrepancies and spectra) of them, for a given k. Namely, in theorems 1 and 2 we state implications and equivalences between them. The theorems intensively use notation and facts of already proved theorems that are discussed in section 2. This long preparation is needed to underline the nonlinear

nature of the multiway discrepancy and to be able to precisely formulate the main theorems in section 3. Section 4 is devoted to the proofs, whereas in section 5 we draw consequences of our statements, as for discrepancy based spectral clustering.

2. Preliminaries

First we recall the main result of the Chung–Graham–Wilson [19] paper about quasirandom properties that apply to the one-class quasirandomness, when the graph sequence imitates the properties of the Erdős–Rényi random graph with edge-density $p = \frac{1}{2}$. The authors also anticipate that instead of $\frac{1}{2}$, any fixed $0 < p < 1$ can be considered. Hereby, we enlist only those properties that will be used later for our purposes, and together with the original formulation (with $p = \frac{1}{2}$) we give the analogous form with a general p, while we use the notation of the original [19] paper.

Let (G_n) be a sequence of graphs as $n \to \infty$. The vertex-set of the general term G_n is V_n, and $|V_n| = n$; whereas, the number of edges of G_n is $e(G_n)$. Consider the following properties.

- $P_1(s)$: for all graphs $M(s)$ on s vertices, $N^*_{G_n}(M(s)) = (1 + o(1))n^s \left(\frac{1}{2}\right)^{\binom{s}{2}}$, where $N^*_{G_n}(M(s))$ denotes the number of labelled induced subgraphs of G_n, isomorphic to $M(s)$. With a general p it reads:

$$N^*_{G_n}(M(s)) = (1 + o(1))n^s p^{e(M(s))}(1 - p)^{\binom{s}{2} - e(M(s))},$$

  where $e(M(s))$ is the number of edges in $M(s)$.
- $P_2(t)$: $e(G_n) \geqslant (1 + o(1))\frac{n^2}{4}$ and $N_{G_n}(C_t) \leqslant (1 + o(1))n^t(\frac{1}{2})^t$, where C_t is the cycle with t edges and $N_{G_n}(C_t)$ is the number of its occurrences as a (not necessarily induced) subgraph of G_n, i.e. the number $\hom(C_t, G_n)$ of the $C_t \to G_n$ homomorphisms. (Note that a relation between N and N^* is given in [19].) With a general p it reads:

$$2e(G_n) \geqslant (1 + o(1))\, pn^2, \quad \text{and} \quad \hom(C_t, G_n) \leqslant (1 + o(1))n^t p^t.$$

- P_3: $e(G_n) \geqslant (1 + o(1))\frac{n^2}{4}$, $\lambda_1 = (1 + o(1))\frac{n}{2}$, $\lambda_2 = o(n)$, where λ_1 and λ_2 are the largest and the second largest (in absolute value) eigenvalues of the adjacency matrix of G_n. (Because of the Frobenius theorem, λ_1 is always positive.) With a general p it reads:

$$2e(G_n) \geqslant (1 + o(1))\, pn^2, \quad \lambda_1 = (1 + o(1))pn, \quad \lambda_2 = o(n). \tag{1}$$

- P_4: $\forall S \subseteq V_n$, $e(S) = \frac{1}{4}|S|^2 + o(n^2)$, where $e(S)$ is the number of edges in the subgraph of G_n induced by S, and $o(n^2)$ is uniform in S. With a general p it reads:

$$\forall X \subseteq V_n : \quad e(X, X) = p|X|^2 + o(n^2), \tag{2}$$

  where $e(X, X) = 2e(X)$, and $o(n^2)$ is uniform in X. The notation $e(X, Y)$ will be used later for the number of cut-edges between vertex-subsets X and Y (counting the possible edges in $X \cap Y$ twice).
- P_7: $\sum_{u,v \in V_n} |N_2(u, v) - \frac{n}{4}| = o(n^3)$. With a general p it reads:

$$\sum_{u,v \in V_n} |N_2(u, v) - p^2 n| = o(n^3),$$

  where $N_2(u, v)$ is the number of common neighbors of u, v in G_n.

The main theorem of [19] states that for $s \geqslant 4$ and $t \geqslant 4$ even,

$$P_2(4) \Rightarrow P_2(t) \Rightarrow P_1(s) \Rightarrow P_3 \Rightarrow P_4 \Rightarrow \cdots \Rightarrow P_7 \Rightarrow P_2(4).$$

The authors of [19] called a graph *quasirandom* if it satisfies any, and therefore all, of the above (and some other, here not used) properties.

In the k-class scenario, generalized quasirandom graphs imitate the properties of the generalized random ones. The definition of a generalized random graph sequence is as follows (see, e.g. [1, 32, 41]) .

Definition 1. We are given a model graph H on k vertices with vertex-weights $r_1, \ldots, r_k$ ($r_i > 0$, $\sum_{i=1}^k r_i = 1$) and edge-weights $p_{ij} = p_{ji}$, $1 \leqslant i \leqslant j \leqslant k$ (entries of the $k \times k$ symmetric probability matrix $\mathbf{P}$, where $0 \leqslant p_{ij} \leqslant 1, 1 \leqslant i \leqslant j \leqslant k$, and the diagonal entries correspond to loops at every vertex). G_n is the general term of a generalized random graph sequence on the model graph H if

- it has n vertices;
- to each vertex v a cluster membership $c_v \in \{1, \ldots, k\}$ is assigned according to the probability distribution $r_1, \ldots, r_k$;
- given the memberships, each pair $v \neq u$ of vertices is connected with probability $p_{c_v c_u}$;
- further, all these decisions are made independently.

Note that in definition 1, when $(U_1, \ldots, U_k)$ denotes the membership based clustering of the vertices (they also depend on n, however we will not denote this dependence, unless necessary), the following *strong balancing condition* on the growth of the cluster sizes can be established: with the notation $n_i = |U_i|$, $i = 1, \ldots, k$ ($\sum_{i=1}^k n_i = n$), if $n \to \infty$, then $\frac{n_i}{n} \to r_i$ ($i = 1, \ldots, k$).

Lovász and Sós [32] gave the following definition of a generalized quasirandom graph sequence.

Definition 2. Given a model graph H on k vertices with vertex-weights $r_1, \ldots, r_k$ and edge-weights $p_{ij} = p_{ji}$, $1 \leqslant i \leqslant j \leqslant k$ (entries of $\mathbf{P}$), (G_n) is H-quasirandom if $G_n \to W_H$ as $n \to \infty$ (in terms of the convergence of homomorphism densities, and W_H is the step-function graphon assigned to H).

The above graph-converge means that for every fixed simple graph F, the number of copies of F in G_n is asymptotically the same as the number of copies of F in a generalized random graph on n vertices and the same model graph H. More precisely, by [16], we say that $G_n \to W_H$ if for any simple graph F,

$$\frac{\hom(F, G_n)}{|V(G_n)|^{|V(F)|}} \to \hom(F, H) = \sum_{\psi: V(F) \to V(H)} \prod_{i \in V(F)} r_{\psi(i)} \prod_{ij \in E(F)} p_{\psi(i)\psi(j)}.$$

If $|V(F)| = s$, then we also have an integral formula for the homomorphism density of the simple graph F in the graphon W_H:

$$\hom(F, H) = \hom(F, W_H) = \int_{[0,1]^s} \prod_{\{i,j\} \in E(F)} W_H(x_i, x_j)\, \mathrm{d}x_1 \ldots \mathrm{d}x_s.$$

Note that in [16], the limit object of a convergent graph sequence is also constructed, that is a symmetric, bounded, measurable function $W : [0, 1] \times [0, 1] \to \mathbb{R}$, called *graphon*. This is the continuous extension of the notion of a graph: the interval [0,1] corresponds to the vertices and the values $W(x, y) = W(y, x)$ to the edge-weights (they are usually in the [0,1] interval).

Then the stepfunction graphon W_G is assigned to the vertex- and edge-weighted graph G in the following way: the sides of the unit square are divided into intervals $I_1, \ldots, I_n$ of lengths proportional to the vertex-weights, and over the rectangle $I_i \times I_j$ the stepfunction takes on the value, which is the weight of the $i \sim j$ edge. The so-called *cut distance* between the graphons W and U is

$$\delta_\square(W, U) = \inf_\nu \|W - U^\nu\|_\square \tag{3}$$

where the *cut-norm* of the graphon W is defined by

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

and the infimum in (3) is taken over all measure-preserving bijections $\nu : [0, 1] \to [0, 1]$, while U^ν denotes the transformed U after performing the same measure-preserving bijection ν on both sides of the unit square. An equivalence relation is defined over the set of graphons: two graphons belong to the same class if they can be transformed into each other by a measure-preserving bijection, i.e. their $\delta_\square$ distance is zero. In the sequel, we consider graphons modulo measure preserving maps, and under graphon we understand the whole equivalence class. The authors of [16] also define the $\delta_\square$ distance between a graphon W and a graph G as

$$\delta_\square(W, G) = \delta_\square(W, W_G).$$

Therefore, the $G_n \to W_H$ convergence in definition 2 is also equivalent to the convergence in the $\delta_\square$ metric.

We also remark that in a generalized quasirandom graph sequence of definition 2, the subgraphs are quasirandom, whereas the bipartite subgraps are bipartite quasirandom. For the definitions, we cite the Lovász–Sós [32] paper:

- The graph sequence (G_n) of simple graphs is *quasirandom* with edge density p if for every simple graph F,

$$\frac{\hom(F, G_n)}{v(G_n)^{v(F)}} \to p^{e(F)} \quad \text{as} \quad n \to \infty,$$

  where $\hom(F, G_n)$ counts the number of homomorphisms of F into G_n (adjacency preserving maps of the vertex-set $V(F)$ into the vertex-set $V(G_n)$), whereas $v(\cdot) = |V(\cdot)|$ and $e(\cdot) = |E(\cdot)|$ stands for the number of vertices and edges of the graph in the argument, respectively. Actually, this definition complies with the above $P_1(s)$ property of the Chung–Graham–Wilson [19] paper, since isomorphism densities can be related to homomorphism ones.
- The graph sequence (G_n) of simple bipartite graphs is *bipartite quasirandom* with edge density p if for every simple bipartite graph F,

$$\frac{\hom(F, G_n)}{l(G_n)^{l(F)} r(G_n)^{r(F)}} \to p^{e(F)} \quad \text{as} \quad n \to \infty,$$

  where $l(\cdot)$ and $r(\cdot)$ stands for the number of vertices in the left and right bipartition classes of the bipartite graph in the argument, respectively.

As definition 2 lays the foundation of the generalized quasirandomness, we want to establish equivalent properties resembling those of the Chung–Graham–Wilson [19] paper.

We also introduce some notation and former theorems (called facts) that are needed in our main theorems. Let $G = (V, \boldsymbol{A})$ be an undirected, *edge-weighted graph* on the n-element vertex-set V with the $n \times n$ symmetric weighted adjacency matrix $\boldsymbol{A}$; the entries satisfy $a_{ij} = a_{ji} \geqslant 0$, $a_{ii} = 0$ and they are similarities between the vertex-pairs. If we have a *simple graph* G, then $\boldsymbol{A}$ is the usual 0-1 adjacency matrix.

The *modularity matrix* of G is defined as $\boldsymbol{M} = \boldsymbol{A} - \mathbf{d}\mathbf{d}^T$ (see [37]), where the entries of $\mathbf{d}$ are the *generalized vertex-degrees* $d_i = \sum_{j=1}^n a_{ij}$ $(i = 1, \ldots, n)$, and $\boldsymbol{A}$ is normalized in such a way that $\sum_{i=1}^n \sum_{j=1}^n a_{ij} = 1$; this assumption does not hurt the generality, since neither the forthcoming *normalized modularity matrix* nor the multiway discrepancies to be defined are affected by the scaling of the entries of $\boldsymbol{A}$. The normalized modularity matrix of G (see [8]) is

$$\boldsymbol{M_D} = \boldsymbol{D}^{-1/2}\boldsymbol{M}\boldsymbol{D}^{-1/2} = \boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2} - \sqrt{\mathbf{d}}\sqrt{\mathbf{d}}^T = \boldsymbol{A_D} - \sqrt{\mathbf{d}}\sqrt{\mathbf{d}}^T,$$

where $\boldsymbol{D} = \mathrm{diag}\,(d_1, \ldots, d_n)$ is the diagonal *degree-matrix*, $\boldsymbol{A_D}$ is the normalized adjacency matrix, and the vector $\sqrt{\mathbf{d}} = (\sqrt{d_1}, \ldots, \sqrt{d_n})^T$ has unit norm.

We will assume that G is *connected*, i.e. $\boldsymbol{A}$ is *irreducible*, in which case, the generalized vertex-degrees are all positive. For the relation between the normalized modularity and Laplacian matrices, see [10].

Let $1 < k < n$ be a fixed integer. In the modularity based spectral clustering, we look for the proper k-partition $U_1, \ldots, U_k$ of the vertices such that the within- and between cluster discrepancies are minimized. We refer to U_i's as *clusters*. To motivate the introduction of the exact discrepancy measure observe that the ij entry of $\boldsymbol{M}$ is $a_{ij} - d_i d_j$, which is the difference between the actual connection of the vertices i, j and the connection that is expected under independent attachment of them with probabilities d_i and d_j, respectively. Consequently, the difference between the actual and the expected connectedness of the subsets $X, Y \subset V$ is

$$\sum_{i \in X} \sum_{j \in Y} (a_{ij} - d_i d_j) = a(X, Y) - \mathrm{Vol}\,(X)\mathrm{Vol}\,(Y),$$

where $a(X, Y) = \sum_{i \in X} \sum_{j \in Y} a_{ij}$ is the *weighted cut* between X and Y, and $\mathrm{Vol}\,(X) = \sum_{i \in X} d_i$ is the *volume* of the vertex-subset X. When $\boldsymbol{A}$ is the 0-1 adjacency matrix, $a(X, Y) = e(X, Y)$ is the number of cut-edges between X and Y. Further, let $\rho(X, Y) := \frac{a(X,Y)}{\mathrm{Vol}\,(X)\mathrm{Vol}\,(Y)}$ be the *volume-density* between X and Y. With these, the following definition is given.

Definition 3 (Definition 6 of [12]). The multiway discrepancy of the undirected, edge-weighted graph $G = (V, \boldsymbol{A})$ in the proper k-partition (clustering) $U_1, \ldots, U_k$ of its vertices is

$$\mathrm{md}\,(G; U_1, \ldots, U_k) = \max_{1 \leqslant i \leqslant j \leqslant k} \max_{X \subset U_i,\, Y \subset U_j} \mathrm{md}\,(X, Y; U_i, U_j),$$

where

$$\begin{aligned} \mathrm{md}\,(X, Y; U_i, U_j) &= \frac{|a(X, Y) - \rho(U_i, U_j)\mathrm{Vol}\,(X)\mathrm{Vol}\,(Y)|}{\sqrt{\mathrm{Vol}\,(X)\mathrm{Vol}\,(Y)}} \\ &= |\rho(X, Y) - \rho(U_i, U_j)|\sqrt{\mathrm{Vol}\,(X)\mathrm{Vol}\,(Y)}. \end{aligned}$$

The minimum k-way discrepancy of G is

$$\mathrm{md}_k(G) = \min_{(U_1, \ldots, U_k) \in \mathcal{P}_k} \mathrm{md}\,(G; U_1, \ldots, U_k),$$

where $\mathcal{P}_k$ denotes the set of proper k-partitions of V.

Observe that $\mathrm{md}\,(X,Y;U_i,U_j)$ is unaffected under scaling the edge-weights. Also note that $\mathrm{md}\,(G;U_1,\ldots,U_k)$ is the smallest α such that for every U_i,U_j pair and for every $X\subset U_i$, $Y\subset U_j$,

$$|a(X,Y)-\rho(U_i,U_j)\mathrm{Vol}\,(X)\mathrm{Vol}\,(Y)|\leqslant\alpha\sqrt{\mathrm{Vol}\,(X)\mathrm{Vol}\,(Y)}$$

holds. It resembles the notion of volume-regular cluster pairs of [4] or the ϵ-regular pairs in the Szemerédi regularity lemma [43], albeit with given number of vertex-clusters, which are usually not equitable; further, with volumes, instead of cardinalities.

The forthcoming facts 1 and 2 justify for the following spectral relaxation of the minimum k-way discrepancy problem. Let the eigenvalues of $\boldsymbol{M_D}$, enumerated in decreasing absolute values, be $1\geqslant|\mu_1|\geqslant|\mu_2|\geqslant\cdots\geqslant|\mu_n|=0$. Assume that $|\mu_{k-1}|>|\mu_k|$, and denote by $\mathbf{u}_1,\ldots,\mathbf{u}_{k-1}$ the unit-norm, pairwise orthogonal eigenvectors, corresponding to $\mu_1,\ldots,\mu_{k-1}$. Let $\mathbf{r}_1,\ldots,\mathbf{r}_n\in\mathbb{R}^{k-1}$ be the row vectors of the $n\times(k-1)$ matrix of column vectors $\boldsymbol{D}^{-1/2}\mathbf{u}_1,\ldots,\boldsymbol{D}^{-1/2}\mathbf{u}_{k-1}$; they are called $(k-1)$-dimensional representatives of the vertices.

The *weighted k-variance* of these representatives is defined as

$$\tilde{S}_k^2=\min_{(U_1,\ldots,U_k)\in\mathcal{P}_k}\sum_{i=1}^{k}\sum_{v\in U_i}d_v\|\mathbf{r}_v-\mathbf{c}_i\|^2, \tag{4}$$

where $\mathbf{c}_i=\frac{1}{\mathrm{Vol}\,(U_i)}\sum_{v\in U_i}d_v\mathbf{r}_v$ is the weighted center of the cluster U_i. It is the *weighted k-means algorithm* that provides this minimum. We will also need the plain k-variance of the representatives $\mathbf{r}_1,\ldots,\mathbf{r}_n\in\mathbb{R}^k$ that are row-vectors of the matrix, the columns of which are the unit-norm, pairwise orthogonal eigenvectors corresponding to the k largest (in absolute value) eigenvalues of $\boldsymbol{A}$. This *k-variance* is

$$S_k^2=\min_{(U_1,\ldots,U_k)}\sum_{i=1}^{k}\sum_{v\in U_i}\|\mathbf{r}_v-\mathbf{c}_i\|^2, \tag{5}$$

where $\mathbf{c}_i=\frac{1}{|U_i|}\sum_{v\in U_i}\mathbf{r}_v$ is the center of the cluster U_i. It is the usual *k-means algorithm* that finds this minimum. By an easy analysis of variance argument it follows that the optimum S_k is just the minimum distance between the subspace spanned by the eigenvectors corresponding to the k largest (in absolute value) eigenvalues of $\boldsymbol{A}$ and the one of the step-vectors over the k-partitions of V. Similar holds for $\tilde{S}_k$ with the transformed eigenvectors.

Note that finding a global minimizer for the k-means problem is NP-complete. However, there are efficient polynomial time algorithms for finding an approximate solution whose value is within a constant fraction of the optimal value, under certain conditions. In particular, theorem 4.6 of [39] states that if the data satisfy the k-clusterable criterion ($S_k^2\leqslant\epsilon^2S_{k-1}^2$ with a small enough ϵ), then there is a PTAS (polynomial time approximation scheme) for the k-means problem. Our conditions for the k-variances in theorems 1 and 2 do comply with this requirement.

The spectral relaxation means that we can approximately find discrepancy minimizing clustering via applying the unweighted or weighted k-means algorithm to the k- or $(k-1)$-dimensional vertex representatives. This is supported by the following facts.

Fact 1 (Theorem 7 of [12]). Let $G=(V,\boldsymbol{A})$ be an edge-weighted, undirected graph, $\boldsymbol{A}$ is irreducible. Then for any integer $1\leqslant k<\mathrm{rank}\,(\boldsymbol{A})$,

$$|\mu_k| \leqslant 9\mathrm{md}_k(G)(k+2-9k\ln \mathrm{md}_k(G)) \tag{6}$$

holds, provided $0 < \mathrm{md}_k(G) < 1$, where μ_k is the kth largest eigenvalue (in absolute value) of the normalized modularity matrix $\boldsymbol{M_D}$ of G.

The above fact 1 is a certain converse of the multiclass expander mixing lemma. In the forward direction, in the $k = 1$ case, the following result is considered as the extension of the expander mixing lemma to irregular graphs:

$$\mathrm{md}_1(G) \leqslant \|\boldsymbol{M_D}\| = |\mu_1|,$$

where $\|\boldsymbol{M_D}\|$ is the spectral norm of the normalized modularity matrix of G. This result was proved in [20], where $\mathrm{md}_1(G)$ is denoted by $\mathrm{disc}(G)$. When $k \geqslant 1$, we are able to prove the following generalization of the expander mixing lemma.

Fact 2 (Theorem 3 of [13]). Let G_n be the general term of a connected simple graph sequence, G_n has n vertices. (We do not denote the dependence of the vertex-set V and adjacency matrix $\boldsymbol{A}$ of G_n on n). Assume that there are constants $0 < c < C < 1$ such that except $o(n)$ vertices, the degrees satisfy $cn \leqslant d_v \leqslant Cn$, $v = 1, \ldots, n$. Let the eigenvalues of the normalized modularity matrix $\boldsymbol{M_D}$ of G_n, enumerated in decreasing absolute values, be

$$|\mu_1| \geqslant \cdots \geqslant |\mu_{k-1}| > \varepsilon \geqslant |\mu_k| \geqslant \cdots \geqslant |\mu_n| = 0.$$

The partition $(U_1, \ldots, U_k)$ of V is defined so that it minimizes the weighted k-variance $s^2 = \tilde{S}_k^2$ of the optimal $(k-1)$-dimensional vertex representatives of G_n. Assume that $(U_1, \ldots, U_k)$ satisfies the strong balancing condition. Then

$$\mathrm{md}(G_n; U_1, \ldots, U_k) \leqslant 2\left(\frac{C}{c} + o(1)\right)(\sqrt{2k}s + \varepsilon).$$

Fact 2 implies that $\mathrm{md}_k(G) \leqslant \mathrm{md}(G; U_1, \ldots, U_k) = \mathcal{O}(\sqrt{2k}\tilde{S}_k + |\mu_k|)$. For the $k = 1$ case, $\varepsilon = |\mu_1|$, $\tilde{S}_1 = 0$ (based on the coordinates of the $\boldsymbol{D}^{-1/2}\mathbf{u}_1 = \boldsymbol{D}^{-1/2}\sqrt{\mathbf{d}} = \mathbf{1}$ vector), and so, we get back the original expander mixing lemma up to a constant. In the $k = 2$ bipartite, biregular case we get the statement of [23], see [12] for further explanation. Consequently, a 'small' $|\mu_k|$ and $\tilde{S}_k$ is an indication of k clusters with 'small' within- and between-cluster discrepancies; further, a partition, close to the optimum, can be obtained by spectral tools. Also observe that the degree-condition of fact 2 means that our graph is dense enough.

3. Generalized random and quasirandom properties with equivalence theorems

Now some properties of a generalized random graph sequence are stated so that to get ideas how to formulate generalized quasirandom properties.

Proposition 1. *Let (G_n) be a generalized random graph sequence on the model graph H; G_n has n vertices with vertex-classes $U_1, \ldots, U_k$ of sizes $n_1, \ldots n_k$ (they also depend on n). Let H, and so k be kept fixed, i.e. the $k \times k$ probability matrix $\mathbf{P}$ of rank k and the 'blow-up' ratios $r_1, \ldots, r_k$ are fixed, while $n \to \infty$ under the strong balancing condition. Then the following properties hold almost surely for the convergence of the graph sequence (G_n), for the adjacency matrix $\boldsymbol{A}_n = (a_{ij}^{(n)})$, the normalized modularity matrix $\boldsymbol{M}_{\boldsymbol{D},n}$, the multiway discrepancies, and the within- and between-cluster codegrees of G_n.*

0. $G_n \to W_H$ *as* $n \to \infty$.

1. $\boldsymbol{A}_n$ *has exactly* k *so-called structural eigenvalues that are* $\Theta(n)$*, while the remaining eigenvalues are* $O(\sqrt{n})$ *(in absolute value). Further, the* k*-variance* $S^2_{k,n}$ *(see (5)) of the* k*-dimensional vertex representatives, based on the eigenvectors corresponding to the structural eigenvalues of* $\boldsymbol{A}_n$*, is* $O(\frac{1}{n})$.

2. There exists a positive constant $0 < \delta < 1$ *independent of* n *(it only depends on* k*) such that* $\boldsymbol{M}_{\boldsymbol{D},n}$ *has exactly* $k-1$ *structural eigenvalues of absolute value greater than* δ*, while all the other eigenvalues are* $O(n^{-\tau})$ *for every* $0 < \tau < \frac{1}{2}$*. Further, the weighted* k*-variance* $\tilde{S}^2_{k,n}$ *(see (4)) of the* $(k-1)$*-dimensional vertex representatives, based on the transformed eigenvectors corresponding to the structural eigenvalues of* $\boldsymbol{M}_{\boldsymbol{D},n}$*, is* $O(n^{-2\tau})$*, for every* $0 < \tau < \frac{1}{2}$.

3. There is a constant $0 < \theta < 1$ *(independent of* n*) such that* $\mathrm{md}_1(G_n) > \theta, \ldots,$ $\mathrm{md}_{k-1}(G_n) > \theta$*, and the* k*-way discrepancy* $\mathrm{md}(G_n; U_1, \ldots, U_k)$ *is* $O(n^{-\tau})$*, for every* $0 < \tau < \frac{1}{2}$.

4. For every $1 \leqslant i \leqslant j \leqslant k$ *and* $u \in U_i$*:*

$$N_1(u; U_j) := \sum_{v \in U_j} a^{(n)}_{uv} = p_{ij} n_j + o(n),$$

where $N_1(u; U_j)$ *denotes the number of neighbors of* u *in* U_j.

For every $1 \leqslant i \leqslant j \leqslant k$ *and* $u, v \in U_i$*,* $u \neq v$*:*

$$N_2(u, v; U_j) := \sum_{t \in U_j} a^{(n)}_{ut} a^{(n)}_{vt} = p^2_{ij} n_j + o(n),$$

where $N_2(u, v; U_j)$ *denotes the number of common neighbors of* u, v *in* U_j.

By property 0, a generalized random graph sequence is also generalized quasirandom, though it converges more quickly. Property 0 was proved in [9], while property 1 in [6], and property 2 in [7] in a more general framework of the SVD of rectangular arrays of nonnegative entries, they are summarized in [10]. The basic idea of the proofs is that the adjacency matrix $\boldsymbol{A}_n$ of G_n can be decomposed as a deterministic block-matrix $\boldsymbol{B}_n$ (blown-up of $\mathbf{P}$ with sizes $n_1, \ldots, n_k$) plus a so-called Wigner-noise $\boldsymbol{W}_n$, see [6]. The Wigner-noise has spectral norm $O(\sqrt{n})$ almost surely, while the blown-up matrix has as many non-zero eigenvalues of order n as the rank of $\mathbf{P}$ (in our case, k) with eigenvectors that are stepwise constant over $U_1, \ldots, U_k$.

Note that to prove Properties 1 and 2, even the following *weak balancing condition* suffices: $n \to \infty$ in such a way that $\frac{n_i}{n} \geqslant c$ $(i = 1, \ldots, k)$ with some constant $0 < c \leqslant \frac{1}{k}$, see [6] (proposition 2.1) and [10]. Property 3 is the consequence of property 2, by the the back and forth statements between discrepancy and normalized modularity spectra, see facts 1 and 2.

Property 4 is easy to prove by large deviations. The subgraph of G_n, induced by U_i and denoted by $G_{ii,n}$, is the general term of an Erdős–Rényi type random graph sequence with edge probability p_{ii}, for every $i = 1, \ldots, k$. The bipartite subgraph of G_n, induced by the U_i, U_j pair and denoted by $G_{ij,n}$, is the general term of a bipartite random graph sequence with edge probability p_{ij}, for every $i, j = 1, \ldots, k$; $i \neq j$ pair. Therefore, the subgraphs are almost surely regular, while the bipartite subgraphs are almost surely biregular. Consequently, the vertex-degrees are of order $\Theta(n)$, almost surely. Further, the codegrees are as expected: every two vertices in U_i have approximately the same number of common neighbors in U_j, for every $i, j = 1, \ldots, k$.

Note that the above generalized random graph in another context is discussed as the stochastic block model or planted partition model, see, e.g. [21, 27, 33], though these papers work with a fixed n and do not consider any condition for the growth of the cluster sizes. Indeed, definition 1 provides us with a random graph model without the hidden (planted) clusters revealed. For this purpose, there are algorithms available, e.g. in [21, 33]; however, one wonders whether a large and dense enough real-life graph can be 'close' to a one coming from this model. In the sequel, we will define precisely some properties that are weaker than those of proposition 1, but can characterize a class of graphs, given k. These will be called generalized quasirandom properties. Note that some other papers, e.g. [1, 15, 36] scale the probability matrix with n, and prove the consistency of the clustering under these conditions.

Properties, reminiscent of those of the generalized random graphs, are now formulated for expanding deterministic graph sequences, and we show that there are many equivalences and implications between them, irrespective of stochastic models. Theorem 1 states mainly implications, whereas theorem 2 states equivalences.

Theorem 1. *Let G_n be the general term of a sequence of simple graphs with vertex-set V_n, adjacency matrix $\boldsymbol{A}_n = (a_{ij}^{(n)})$, and normalized modularity matrix $\boldsymbol{M}_{\boldsymbol{D},n}$. Let k be a fixed positive integer, whereas $|V_n| = n \to \infty$. Consider the following properties:*

P0. There exists a vertex- and edge-weighted graph H on k vertices with vertex-weights $r_1, \dots, r_k$ and edge-weights $p_{ij} = p_{ji} \in [0,1]$, $1 \leqslant i \leqslant j \leqslant k$, where the $k \times k$ symmetric probability matrix $\mathbf{P} = (p_{ij})$ has rank k, such that $G_n \to W_H$ as $n \to \infty$. (This is, in fact, the repetition of definition 2.)

PI. $\boldsymbol{A}_n$ has k structural eigenvalues $\lambda_{1,n}, \dots, \lambda_{k,n}$ such that the normalized eigenvalues converge: $\frac{1}{n}\lambda_{i,n} \to q_i$ as $n \to \infty$ $(i = 1, \dots, k)$ with some non-zero reals $q_1, \dots, q_k$, and the remaining eigenvalues are $o(n)$.

The k-variance $S^2_{k,n}$ of the k-dimensional vertex representatives, based on the eigenvectors corresponding to the structural eigenvalues of $\boldsymbol{A}_n$, is $o(1)$. The k-partition $(U_1, \dots, U_k)$ minimizing this k-variance satisfies the strong balancing condition.

PII. G_n has no dominant vertices: there are constants $0 < c < C < 1$ such that the vertex-degrees are between cn and Cn, except of possibly $o(n)$ vertices; further, there exists a constant $0 < \delta < 1$ (independent of n) such that $\boldsymbol{M}_{\boldsymbol{D},n}$ has $k-1$ structural eigenvalues that are greater than δ (in absolute value), while the remaining eigenvalues are $o(1)$.

The weighted k-variance $\tilde{S}^2_{k,n}$ of the $(k-1)$-dimensional vertex representatives, based on the transformed eigenvectors corresponding to the structural eigenvalues of $\boldsymbol{M}_{\boldsymbol{D},n}$, is $o(1)$. The k-partition $(U_1, \dots, U_k)$ minimizing this k-variance satisfies the strong balancing condition.

PIII. There are vertex-classes $U_1, \dots, U_k$ obeying the strong balancing condition, and there is a constant $0 < \theta < 1$ (independent of n) such that $\mathrm{md}_1(G_n) > \theta, \dots, \mathrm{md}_{k-1}(G_n) > \theta$, and $\mathrm{md}(G_n; U_1, \dots, U_k) = o(1)$.

PIV. There are vertex-classes $U_1, \dots, U_k$ of sizes $n_1, \dots, n_k$ obeying the strong balancing condition, and there is a $k \times k$ symmetric probability matrix $\mathbf{P} = (p_{ij})$ of rank k such that, with them, the following holds:

$$\sum_{u,v \in U_i} |N_2(u,v; U_j) - p_{ij}^2 n_j| = o(p_{ij}^2 n_i^2 n_j) = o(n^3), \quad \forall i,j = 1, \dots, k. \tag{7}$$

Then P0 is equivalent to PIV, and they imply PI and PII; further, PII implies PIII. We also consider the following strengthening of property PI:

PI +. $\boldsymbol{A}_n$ *has* k *structural eigenvalues* $\lambda_{1,n},\dots,\lambda_{k,n}$ *such that the normalized eigenvalues converge:* $\frac{1}{n}\lambda_{i,n}\to q_i$ *as* $n\to\infty$ $(i=1,\dots,k)$ *with some non-zero reals* $q_1,\dots,q_k$*, and the remaining eigenvalues are* $o(\sqrt{n})$*.*

The k-variance $S^2_{k,n}$ *of the k-dimensional vertex representatives, based on the eigenvectors corresponding to the structural eigenvalues of* $\boldsymbol{A}_n$*, is* $o(\frac{1}{n})$*. The k-partition* $P_{k,n}=(U_{1n},\dots,U_{kn})$ *of the vertices of* G_n *minimizing this k-variance not only satisfies the strong balancing condition, but we also assume that there is a* $k\times k$ *symmetric probability matrix* $\mathbf{P}=(p_{ij})$ *of rank k such that*

$$d(U_{in},U_{jn}):=\frac{e(U_{in},U_{jn})}{|U_{in}||U_{jn}|}\to p_{ij}\quad(1\leqslant i\leqslant j\leqslant k),\quad n\to\infty. \tag{8}$$

(I.e. the within- and between-cluster edge densities converge to the entries of $\mathbf{P}$*.)*

Then PI + implies P0.

Note that properties P0–PIV are in many ways weaker than those of the generalized random graphs. For example, the generalized quasirandom graphs may converge much more slowly than the generalized random ones that can be decomposed as a deterministic block-matrix plus a Wigner-noise. The Wigner-noise has spectral norm $O(\sqrt{n})$, therefore converges very quickly. In case of a generalized quasirandom graph, the noise has spectral norm $o(n)$ only. Also, in the generalized random case, the codegree condition holds for any pair of the vertices (of the same cluster toward those of a given own or other cluster), whereas in the general quasirandom case, it only holds on average, see (7). However, PI + is even stronger than property 1 of proposition 1: it requires $o(\sqrt{n})$ magnitude of the non-structural eigenvalues, whereas those of a generalized random graph are of order $O(\sqrt{n})$. However, except of $o(n)$ ones, the latter are also of order $o(\sqrt{n})$, in view of concentration results on the eigenvalues, see, e.g. [3]. In summary, the above P0 is stronger than PI, but weaker than PI + . In the $k=1$ case, property P_3 of the Chung–Graham–Wilson [19] paper is equivalent to $P_1(s)$ for every s (our P0). But there, above the separation in the spectrum, a lower bound for the edge-density should be satisfied. In the $k>1$ scenario, the inter- and intra-cluster edge-densities can be bounded only at the expense of each other, so we have to require the convergence of these edge-densities too.

Theorem 2. *Let us define PII + and PIII + as PII and PIII of theorem 1 together with the following additional assumptions, respectively. The k-partition* $(U_1,\dots,U_k)$ *emerging in PII and PIII of theorem 1 not only satisfies the strong balancing condition, but there is a* $k\times k$ *symmetric probability matrix* $\mathbf{P}=(p_{ij})$ *of rank k such that*

$$d(U_i,U_j)=p_{ij}+o(1)\quad(1\leqslant i\leqslant j\leqslant k),\quad n\to\infty \tag{9}$$

(the same as (8)), and for every $1\leqslant i\leqslant j\leqslant k$ *and* $u\in U_i$*,*

$$N_1(u;U_j)=(1+o(1))\,p_{ij}n_j \tag{10}$$

holds. Then properties P0, PII + , PIII + , and PIV are all equivalent.

4. Proofs of theorems 1 and 2

Proof of theorem 1.
Proof of P0 $\Longrightarrow$ PIV. We use the results of [19, 32]. By [32], the vertex set of the generalized quasirandom graph G_n (defined by P0) can be partitioned into classes $U_1, \dots, U_k$ in such a way that $\frac{|U_i|}{n} \to r_i$ $(i = 1, \dots, k)$, that gives the strong balancing; the subgraph $G_{ii,n}$ is the general term of a quasirandom graph sequence with edge-density tending to p_{ii} $(i = 1, \dots, k)$, whereas $G_{ij,n}$ is the general term of a bipartite quasirandom graph sequence with edge-density tending to p_{ij} $(i \neq j)$ as $n \to \infty$. Therefore, for the subgraphs the equivalent statements of [19] of the usual (one-class) quasirandomness are applicable, and similar considerations can be made for the bipartite subgraphs as well, see [14, 29, 45]. We also need two simple lemmas. □

Lemma 1. *If $(G_{ii,n})$ is quasirandom, then*

$$\sum_{u,v\in U_i} N_2(u,v;U_i) \geqslant (1+o(1))\, p_{ii}^2 n_i^3, \quad i = 1, \dots, k.$$

Proof of lemma 1. We drop the index n of the adjacency entries.

$$\sum_{u,v\in U_i} N_2(u,v;U_i) = \sum_{u,v\in U_i}\sum_{t\in U_i} a_{ut}a_{vt} = \sum_{t\in U_i}\sum_{u\in U_i} a_{ut}\sum_{v\in U_i} a_{vt} = \sum_{t\in U_i}[N_1(t;U_i)]^2$$
$$\geqslant \frac{1}{n_i}[\sum_{t\in U_i} N_1(t;U_i)]^2 = \frac{1}{n_i}[2e(U_i)]^2 \geqslant \frac{1}{n_i}[(1+o(1))\, p_{ii}n_i^2]^2 = (1+o(1))\, p_{ii}^2 n_i^3,$$

where $e(U_i)$ is the number of edges within the subgraph $G_{ii,n}$ of G_n, induced by U_i. In the first inequality we used the Cauchy–Schwarz, and in the second one, the first part of the equivalent quasirandom property P_3 of [19], see (6). □

Lemma 2. *If $(G_{ij,n})$ is bipartite quasirandom, then*

$$\sum_{u,v\in U_i} N_2(u,v;U_j) \geqslant (1+o(1))\, p_{ij}^2 n_i^2 n_j, \quad i \neq j.$$

Proof of lemma 2. We drop the index n of the adjacency entries.

$$\sum_{u,v\in U_i} N_2(u,v;U_j) = \sum_{u,v\in U_i}\sum_{t\in U_j} a_{ut}a_{vt} = \sum_{t\in U_j}\sum_{u\in U_i} a_{ut}\sum_{v\in U_i} a_{vt} = \sum_{t\in U_j}[N_1(t;U_i)]^2$$
$$\geqslant \frac{1}{n_j}[\sum_{t\in U_j} N_1(t;U_i)]^2 = \frac{1}{n_j}[e(U_i,U_j)]^2 \geqslant \frac{1}{n_j}[(1+o(1))\, p_{ij}n_i n_j]^2 = (1+o(1))\, p_{ij}^2 n_i^2 n_j,$$

where recall that $e(U_i,U_j)$ is the number of cut-edges between U_i and U_j, i.e. the number of edges in the bipartite subgraph $G_{ij,n}$ of G_n, induced by the U_i, U_j pair. Here, in the first inequality we used the Cauchy–Schwarz, and in the second one, the equivalent quasirandom property of bipartite quasirandom graphs. □

In view of the lemmas, we estimate the square of the left-hand side of (7) by the Cauchy–Schwarz inequality for every $i, j = 1, \dots k$, with n_i^2 terms:

$$\left\{\sum_{u,v\in U_i}|N_2(u,v;U_j)-p_{ij}^2n_j|\right\}^2\leqslant n_i^2\sum_{u,v\in U_i}|N_2(u,v;U_j)-p_{ij}^2n_j|^2$$
$$=n_i^2\left\{\sum_{u,v\in U_i}[N_2(u,v;U_j)]^2-2p_{ij}^2n_j\sum_{u,v\in U_i}N_2(u,v;U_j)+n_i^2(p_{ij}^2n_j)^2\right\}$$
$$\leqslant n_i^2\left\{(1+o(1))\,p_{ij}^4n_i^2n_j^2-2(1+o(1))\,p_{ij}^4n_i^2n_j^2+p_{ij}^4n_i^2n_j^2\right\}$$
$$=n_i^2o(1)\,p_{ij}^4n_i^2n_j^2=o(p_{ij}^4n_i^4n_j^2),$$

where to estimate $\sum_{u,v\in U_i}N_2(u,v;U_j)$ we used lemma 1 in the $i=j$, and lemma 2 in the $i\neq j$ case; further, utilized that $\sum_{u,v\in U_i}[N_2(u,v;U_j)]^2$ is asymptotically $\hom(C_4,G_{ij,n})$, the number of the $C_4\to G_{ij,n}$ homomorphisms, where C_4 is the 4-cycle graph. Indeed, in the $i=j$ case, the equivalent property $P_2(4)$ of [19], guarantees that $\hom(C_4,G_{ii,n})\leqslant(1+o(1))\,p_{ii}^4n_i^4$, the latter is what is expected in an Erdős–Rényi type random graph with edge-density p_{ii}. In the $i\neq j$ case a similar property (see [32]) for bipartite quasirandom graphs implies that the $C_4\to G_{ij,n}$ homomorphism density is

$$\frac{\hom(C_4,G_{ij,n})}{n_i^2n_j^2}=(1+o(1))\,p_{ij}^4.$$

Here $\hom(C_4,G_{ij,n})=\sum_{u,v\in U_i}[N_2(u,v;U_j)]^2$ asymptotically, as by [32] only 4-cycles in the above bipartition have to be considered; these 4-cycles have two vertices from U_i and two from U_j, and any two of the common neighbors of $u,v\in U_i$ in U_j are possible candidates to close a (labelled) 4-cycle with them. □

Proof of PIV $\Longrightarrow$ P0. If the average codegree condition (7) holds for the subgraphs ($i=j$), then by the $P_7\Rightarrow P_1(s)$ ($s=1,2,\dots$) implication of the equivalent quasirandom properties of [19], the subgraphs $G_{ii,n}$ are quasirandom (in terms of the isomorphisms, and so of the homomorphism densities). Likewise, in the $i\neq j$ case, the bipartite subgraphs $G_{ij,n}$ are bipartite quasirandom. Therefore, G_n is built of quasirandom and bipartite quasirandom blocks, so under the strong balancing condition, they together form a generalized quasirandom graph sequence on k classes and model graph H, the vertex-weights of which are $r_1,\dots,r_k$ of the strong balancing condition, and the edge-weights are entries in the condition (7) of the probability matrix $\mathbf{P}$ in PIV. □

Proof of P0 $\Longrightarrow$ PII. We will use the following results proved in former papers.

Fact 3 (Theorem 8 of [11]). Let $G_n=(V_n,\boldsymbol{A}_n)$ be the general term of a convergent sequence of connected edge-weighted graphs whose edge-weights are in [0,1] and the vertex-weights are the generalized degrees. Assume that there are no dominant vertices. Let W denote the limit graphon of the sequence (G_n), and let

$$|\mu_{n,1}|\geqslant|\mu_{n,2}|\geqslant\cdots\geqslant|\mu_{n,n}|=0$$

be the normalized modularity spectrum of G_n (the eigenvalues are indexed by their decreasing absolute values). Further, let $\mu_i(P_{\mathbb{W}})$ be the ith largest (in absolute value) eigenvalue of the integral operator $P_{\mathbb{W}}:L^2(\xi')\to L^2(\xi)$ taking conditional expectation with respect to the joint measure $\mathbb{W}$ embodied by the normalized limit graphon W. Let ξ,ξ' are identically distributed random variables with the marginal distribution of their symmetric joint distribution $\mathbb{W}$, whereas $L^2(\xi)$ denotes the Hilbert space of the measurable functions of ξ with 0 expectation and finite variance. Then for every $i\geqslant1$,

$$\mu_{n,i} \to \mu_i(P_{\mathbb{W}}) \quad \text{as} \quad n \to \infty.$$

Some explanation is in order about the integral operator $P_{\mathbb{W}}$. Let (ξ, ξ') be a pair of identically distributed real-valued random variables defined over the product space $\mathcal{X} \times \mathcal{X}$ having a symmetric joint distribution $\mathbb{W}$ with equal margins $\mathbb{P}$. Suppose that the dependence between ξ and ξ' is regular, i.e. their joint distribution $\mathbb{W}$ is absolutely continuous with respect to the product measure $\mathbb{P} \times \mathbb{P}$, and let w denote its Radon–Nikodym derivative. Note that the Hilbert spaces $H = L^2(\xi)$ and $H' = L^2(\xi')$ defined in fact 3 (with the covariance as inner product) are also isomorphic in the sense that for any $\psi \in H$ there exists a $\psi' \in H'$ and vice versa, such that ψ and ψ' are identically distributed. Then the linear operator $P_{\mathbb{W}} : H' \to H$ taking conditional expectation with respect to the joint distribution is an integral operator; actually, it is a projection restricted to H' and projects onto H. More exactly, to $\psi' \in H'$ the operator $P_{\mathbb{W}}$ assigns $\psi \in H$ such that $\psi = \mathbb{E}_{\mathbb{W}}(\psi' \mid \xi)$, i.e.

$$\psi(x) = \int_{\mathcal{Y}} w(x, y)\psi'(y)\, \mathbb{P}(\mathrm{d}y), \quad x \in \mathcal{X}.$$

If

$$\int_{\mathcal{X}} \int_{\mathcal{X}} w^2(x, y)\mathbb{P}(\mathrm{d}x)\mathbb{P}(\mathrm{d}y) < \infty,$$

then $P_{\mathbb{W}}$ is a Hilbert–Schmidt operator, therefore it is compact and has the spectral decomposition

$$P_{\mathbb{W}} = \sum_{i=1}^{\infty} \mu_i \langle ., \psi_i' \rangle_{H'} \psi_i$$

where for the eigenvalues $|\mu_i| \leqslant 1$ holds and the eigenvalue–eigenfunction equation looks like

$$P_{\mathbb{W}}\psi_i' = \mu_i \psi_i \quad (i = 1, 2, \dots)$$

where ψ_i and ψ_i' are identically distributed, whereas their joint distribution is $\mathbb{W}$. It is easy to see that $P_{\mathbb{W}}$ is self-adjoint and it takes the constantly 1 random variable of H' into the constantly 1 random variable of H; however, the $\psi_0 = 1, \psi_0' = 1$ pair is not regarded as a function pair with eigenvalue $\mu_0 = 1$, since they have no zero expectation. See [10] for more details.

Fact 4 (Theorem 9 of [11]). Assume that there are constants $0 < \varepsilon < \delta \leqslant 1$ such that the normalized modularity spectrum of the above G_n satisfies

$$|\mu_{n,1}| \geqslant \cdots \geqslant |\mu_{n,k-1}| \geqslant \delta > \varepsilon \geqslant |\mu_{n,k}| \geqslant \cdots \geqslant |\mu_{n,n}| = 0.$$

With the notation of fact 3, and assuming that there are no dominant vertices of G_n, the subspace spanned by the vectors $\boldsymbol{D}_n^{-1/2}\mathbf{u}_{n,1}, \dots, \boldsymbol{D}_n^{-1/2}\mathbf{u}_{n,k-1}$, where $\mathbf{u}_{n,1}, \dots, \mathbf{u}_{n,k-1}$ are orthonormal eigenvectors belonging to the $k-1$ largest absolute value eigenvalues of the normalized modularity matrix of G_n, also converges to the corresponding $(k-1)$-dimensional subspace of $P_{\mathbb{W}}$. More exactly, if $\mathbf{P}_{n,k-1}$ denotes the projection onto the subspace spanned by the vectors $\boldsymbol{D}_n^{-1/2}\mathbf{u}_{n,1}, \dots, \boldsymbol{D}_n^{-1/2}\mathbf{u}_{n,k-1}$, and $\mathbf{P}_{k-1}$ denotes the projection onto the analogous eigensubspace of $P_{\mathbb{W}}$, then $\|\mathbf{P}_{n,k-1} - \mathbf{P}_{k-1}\| \to 0$ as $n \to \infty$ (in spectral norm).

So the proof of P0 $\Longrightarrow$ PII is as follows. Since by P0, (G_n) converges to the limit graphon W_H, and the eigenvalues $\mu_i(P_{\mathbb{W}_H})$ of fact 3 consist of $k-1$ non-zero numbers, and the others are zeros, the statement for the convergence of the modularity spectrum follows.

As for the weighted k-variances, we use fact 4, which implies that the subspace spanned by the transformed eigenvectors corresponding to the structural eigenvalues of $\boldsymbol{M}_{\boldsymbol{D},n}$ converges to the subspace of step-vectors; further, the steps are proportional to r_i's, so the strong balancing also follows. As the weighted k-variance depends continuously on the above subspaces, fact 4 implies the convergence of the weighted k-variance as well. Note that here $\|\mathbf{P}_{n,k-1} - \mathbf{P}_{k-1}\|_F \leqslant \sqrt{k-1}\|\mathbf{P}_{n,k-1} - \mathbf{P}_{k-1}\| \to 0$ as $n \to \infty$, where $\|\cdot\|_F$ denotes the Frobenius norm. □

Proof of P0 $\Longrightarrow$ PI. We use theorem 6.7 of [17], where the authors prove that if the sequence (W_{G_n}) of graphons converges to the limit graphon W, then both ends of the spectra of the integral operators, induced by W_{G_n}'s as kernels, converge to the ends of the spectrum of the integral operator induced by W as kernel. We apply this argument to the step-function graphon corresponding to G_n (the eigenvalues of the induced integral operator are the normalized eigenvalues of G_n) and for the limit graphon W_H of (G_n). The same argument as in P0 $\Longrightarrow$ PII can be applied for the convergence of the spectral subspaces, so by the above considerations, the convergence of the k-variances is also obtained. Since the steps of the emerging step-vectors are proportional to r_i's, the strong balancing condition also holds. □

Proof of PII $\Longrightarrow$ PIII. We will use Facts 1 and 2.

Assume that there is a constant $0 < \delta < 1$ such that $\boldsymbol{M}_{\boldsymbol{D},n}$ has $k-1$ eigenvalues that are greater than δ in absolute value, while the remaining eigenvalues are $o(1)$; further, the square-root of weighted k-variance $\tilde{S}^2_{k,n}$ is also $o(1)$. Using that there are no dominant vertices, we apply fact 2. According to this, $\mathrm{md}\,(G_n; U_1, \dots, U_k) = o(1)$. Indirectly, assume that there is no absolute constant $0 < \theta < 1$ such that $\mathrm{md}_1(G_n) > \theta, \dots, \mathrm{md}_{k-1}(G_n) > \theta$. Then there is an $1 \leqslant i \leqslant k-1$ with $\mathrm{md}_i(G_n) \leqslant \varepsilon$ for any $0 < \varepsilon < 1$. Consequently, to any ε there is an N, such that for all $n > N$: $\mathrm{md}_i(G_n) \leqslant \varepsilon$, for this $1 \leqslant i \leqslant k-1$. But to this $\varepsilon > 0$ there is an $\varepsilon' > 0$ such that for all $n > N$: $|\mu_{n,i}| \leqslant \varepsilon'$, for this $1 \leqslant i \leqslant k-1$. We can select ε so small that $\varepsilon' < \delta$. This is possible since the function on the right of equation (6) estimates $|\mu_{n,i}|$ with a (near zero) strictly increasing function of $\mathrm{md}_i(G_n)$. This contradicts to the $|\mu_{n,i}| > \delta$ assumption. □

Proof of PI + $\Longrightarrow$ P0. Under the assumptions of PI + , by theorem 3.1.17 of [10] we are able to find a blown-up matrix $\boldsymbol{B}_n$ of rank k and an error-matrix $\boldsymbol{E}_n$ with $\|\boldsymbol{E}_n\| = o(\sqrt{n})$ such that $\boldsymbol{A}_n = \boldsymbol{B}_n + \boldsymbol{E}_n$ $(n = k, k+1, \dots)$. Say, $\boldsymbol{B}_n$ is the blown-up matrix of the $k \times k$ pattern matrix $\mathbf{P}_n$, the ij entry $p_{ij}^{(n)}$ of which is the common entry of the $U_{in} \times U_{jn}$ block of $\boldsymbol{B}_n$.

Then using the relation between the cut-norm of a graphon and a matrix, further, between the cut-norm and the spectral norm of a matrix, and the transformation of a graph into a step-function graphon, we get that

$$\|W_{\boldsymbol{E}_n}\|_\square \leqslant \frac{1}{n^2}\|\boldsymbol{E}_n\|_\square \leqslant \frac{1}{n^2} n\|\boldsymbol{E}_n\| = \frac{1}{n} o(\sqrt{n}) = o(n^{-1/2}),$$

where $\|\boldsymbol{E}_n\|$ is the spectral-norm, $\|\boldsymbol{E}_n\|_\square$ is the matrix cut-norm of $\boldsymbol{E}_n$ (for the latter one, see [25]), and $W_{\boldsymbol{E}_n}$ denotes the graphon corresponding to the symmetric matrix $\boldsymbol{E}_n$ of uniformly bounded entries. (The sides of the unit square are divided into n equidistant intervals, and the step-function over the small squares of the unit square takes on values corresponding to the matrix entries.) Though, these entries can be negative, the theory of bounded graphons applies to $W_{\boldsymbol{E}_n}$, too.

Using the Steiner equality, we get that the squared Frobenius norm of $\boldsymbol{A}_n - \boldsymbol{B}_n$, restricted to the ij block, is

$$\|(\boldsymbol{A}_n-\boldsymbol{B}_n)_{ij}\|_F^2=\sum_{u\in U_{in}}\sum_{v\in U_{jn}}(a_{uv}^{(n)}-p_{ij}^{(n)})^2$$
$$=\sum_{u\in U_{in}}\sum_{v\in U_{jn}}(a_{uv}^{(n)}-d(U_{in},U_{jn}))^2+|U_{in}||U_{jn}|(d(U_{in},U_{jn})-p_{ij}^{(n)})^2,$$

where the edge-density $d(U_{in},U_{jn})$ of (8) is now viewed as the average of the entries of $\boldsymbol{A}_n$ in the $U_{in}\times U_{jn}$ block. Then by the inequality between the Frobenius and spectral norms,

$$\|(\boldsymbol{A}_n-\boldsymbol{B}_n)_{ij}\|_F^2\leqslant n\|\boldsymbol{A}_n-\boldsymbol{B}_n\|^2=n\|\boldsymbol{E}_n\|^2=n(o(\sqrt{n}))^2.$$

Therefore, for every $1\leqslant i\leqslant j\leqslant k$ pair:

$$(d(U_{in},U_{jn})-p_{ij}^{(n)})^2\leqslant\frac{1}{|U_{in}||U_{jn}|}n(o(\sqrt{n}))^2=\frac{1}{\frac{|U_{in}|}{n}\frac{|U_{jn}|}{n}}n(\frac{o(\sqrt{n})}{n})^2$$
$$=n(o(n^{-1/2}))^2=o(1) \tag{11}$$

as $\frac{|U_{in}|}{n}\to r_i$ when $n\to\infty$ $(i=1,\dots,k)$. Consequently, $p_{ij}^{(n)}\geqslant 0$, provided there are no constantly zero blocks in $\boldsymbol{A}_n$.

Eventually, we prove the $G_n\to W_H$ convergence by verifying that the cut-distance between the corresponding graphons tends to 0. Using the triangle inequality, we get

$$\|W_{G_n}-W_H\|_\square\leqslant\|W_{G_n}-W_{\boldsymbol{B}_n}\|_\square+\|W_{\boldsymbol{B}_n}-W_{G_n/P_{k,n}}\|_\square+\|W_{G_n/P_{k,n}}-W_H\|_\square$$

where $G_n/P_{k,n}$ is the factor graph of G_n with respect to the k-partition $P_{k,n}=(U_{1n},\dots,U_{kn})$. This is an edge- and vertex-weighted graph on k vertices, with vertex-weights $\frac{|U_{in}|}{n}$ and edge-weights $d(U_{in},U_{jn})$, $i,j=1,\dots,k$.

The first term is $\|W_{\boldsymbol{E}_n}\|_\square=o(n^{-1/2})$. To estimate the second term, we use that $\boldsymbol{B}_n$ is the blown-up matrix of $\mathbf{P}_n$ with respect to the k-partition $P_{k,n}$, after conveniently permuting its rows (and columns, accordingly). The graphon $W_{\boldsymbol{B}_n}$ is also stepwise constant over the unit square, where the sides are divided into k parts: the interval I_j has lengths $\frac{|U_{jn}|}{n}$ $(j=1,\dots,k)$, and over $I_i\times I_j$ the step-function takes on the value $p_{ij}^{(n)}$. By its nature, the graphon $W_{G_n/P_{k,n}}$ is stepwise constant with the same subdivision of the unit square, and over $I_i\times I_j$ it takes on the value $d(U_{in},U_{jn}),i,j=1,\dots,k$. But in view of (11), $\|W_{\boldsymbol{B}_n}-W_{G_n/P_{k,n}}\|_\square=\sqrt{n}o(n^{-1/2})=o(1)$. The third term is $o(1)$, because of the assumptions $\frac{|U_{in}|}{n}\to r_i$ $(i=1,\dots,k)$ and $d(U_{in},U_{jn})\to p_{ij}$, $i,j=1,\dots,k$. Therefore, $\|W_{G_n}-W_H\|_\square=o(1)$ and so, $G_n\to H$, which finishes the proof. □

Summarizing: we proved that PIV $\Longleftrightarrow$ P0 $\Longrightarrow$ PI, and PI + $\Longrightarrow$ P0. So PI is weaker and PI + is stronger than P0. We should find something between PI and PI + which is equivalent to P0, what is an open question yet. Further, P0 $\Longrightarrow$ PII $\Longrightarrow$ PIII. We are also able to prove the equivalences of theorem 2 by strengthening PII and PIII.

Proof of theorem 2. We will proceed in the following order: P0 $\Longrightarrow$ PII + $\Longrightarrow$ PIII + $\Longrightarrow$ PIV $\Longrightarrow$ P0, where the last implication has already been proved. For the proof we also need the following lemma.

Lemma 3. *Under P0, the following holds for except $o(n_i)$ vertices $u\in U_i$, and for every $1\leqslant i\leqslant j\leqslant k$:*

$$N_1(u;U_j)=(1+o(1))\,p_{ij}n_j.$$

The statement of the lemma follows from the $P_1(s)$ $(\forall s) \Rightarrow P_0'$ implication of the Chung–Graham–Wilson [19] paper (if $i=j$) and its bipartite analogue (if $i \neq j$). In the multiclass scenario, lemma 3 implies that there are no dominant vertices, i.e. the vertex-degrees are of order $\Theta(n)$, except for at most $o(n)$ vertices. Also, except at most $o(n)$ vertices, the clusters are asymptotically regular, and the cluster pairs are asymptotically biregular. Note that this property, even in the $k=1$ case, is weaker than the equivalent properties of quasirandomness, in particular, than that for the codegrees. □

Proof of P0 $\Longrightarrow$ PII + $\Longrightarrow$ PIII + . These implications hold in view of the proof of theorem 1, as the extra conditions follow from P0, automatically. Indeed, P0 implies the convergence of the edge-densities (see [32], the end of the proof of theorem 2.2) and that of the within- and between-cluster vertex-degrees (see lemma 3), though, latter is weaker than the quasirandomness. □

Proof of PIII + $\Longrightarrow$ PIV. We will prove that P_4, and the analogous statement of [45] holds, whenever PIII + holds. Here we cite the notion of $(p-\beta)$-jumbled and bi-jumbled graphs based on [14, 29, 44, 45]. A graph is $(p-\beta)$-jumbled if for any $X \subset V$

$$\left| e(X) - p\binom{|X|}{2} \right| \leqslant \beta |X|$$

with $0 < p \leqslant 1 < \beta$ and $e(X)$ denotes the number of edges of the underlying graph with both endpoints in X. In the random graph $G_n(p)$, $\beta = O(\sqrt{n})$, which is best possible (see [14]), and for a quasirandom G_n, $\beta = o(n)$, in view of P_4 of [19].

Likewise, a bipartite graph is $(p-\beta)$-bijumbled if for every $X, Y \subset V$

$$|e(X,Y) - p|X||Y|| \leqslant \beta \sqrt{|X||Y|}.$$

It makes sense for bipartite graphs with the partition of vertices (U_1, U_2), where $X \subset U_1$ and $Y \subset U_2$. For bipartite quasirandom graphs, $\beta = o(n)$ again, see [45].

Under the degree-conditions, $\mathrm{Vol}\,(X)$ is approximately n times $|X|$, so the right hand side constant, analogous to β, will be $o(1)$ in our case, when we use volumes instead of cardinalities. Indeed, if $i=j$, then we take into consideration that by (10), for $X \subset U_i$,

$$\mathrm{Vol}\,(X) = |X|(1+o(1)) \sum_{\ell=1}^{k} p_{i\ell} n_\ell .$$

Let $(U_1, \ldots, U_k)$ be the k-partition, guaranteed by PIII + , such that $\mathrm{md}_k(G_n; U_1, \ldots U_k) = o(1)$. Then for $X \subset U_i$,

$$\begin{aligned}
&e(X,X) - p_{ii}|X|^2 = e(X,X) - [d(U_i,U_i) + o(1)]|X|^2 \\
&= e(X,X) - \frac{e(U_i,U_i)}{\frac{\mathrm{Vol}^2(U_i)}{(1+o(1))^2(\sum_{\ell=1}^k p_{i\ell} n_\ell)^2}} \frac{\mathrm{Vol}^2(X)}{(1+o(1))^2(\sum_{\ell=1}^k p_{i\ell} n_\ell)^2} - o(1)|X|^2 \\
&= [e(X,X) - \rho(U_i,U_i)\mathrm{Vol}^2(X)] - o(1)\rho(U_i,U_i)\mathrm{Vol}^2(X) - o(1)|X|^2 \\
&\leqslant \mathrm{md}_k(G_n; U_1, \ldots U_k)\sqrt{\mathrm{Vol}^2(X)} - o(1)e(U_i,U_i)\left(\frac{\mathrm{Vol}\,(X)}{\mathrm{Vol}\,(U_i)}\right)^2 - o(n^2) = o(n^2)
\end{aligned}$$

as $\mathrm{md}_k(G_n; U_1, \ldots U_k) = o(1)$ by PIII+, and we also used (9). Then P_4 of [19] implies P_7 of [19], that is our PIV. If $i \neq j$, then with a similar argument, for $X \subset U_i$, $Y \subset U_j$:

$$
\begin{aligned}
&e(X,Y)-p_{ij}|X||Y| = e(X,Y)-[d(U_i,U_j)+o(1)]|X||Y| \\
&= e(X,Y) - \frac{e(U_i,U_j)}{\frac{\mathrm{Vol}\,(U_i)\mathrm{Vol}\,(U_j)}{(1+o(1))^2(\sum_{\ell=1}^k p_{i\ell}n_\ell)(\sum_{\ell=1}^k p_{j\ell}n_\ell)}}\,\frac{\mathrm{Vol}\,(X)\mathrm{Vol}\,(Y)}{(1+o(1))^2(\sum_{\ell=1}^k p_{i\ell}n_\ell)(\sum_{\ell=1}^k p_{j\ell}n_\ell)} \\
&- o(1)|X||Y| \\
&= [e(X,Y)-\rho(U_i,U_i)\mathrm{Vol}\,(X)\mathrm{Vol}\,(Y)] - o(1)\rho(U_i,U_i)\mathrm{Vol}\,(X)\mathrm{Vol}\,(Y) - o(1)|X||Y| \\
&\leqslant \mathrm{md}_k(G_n;U_1,\dots U_k)\sqrt{\mathrm{Vol}\,(X)\mathrm{Vol}\,(Y)} - o(1)e(U_i,U_j)\left(\frac{\mathrm{Vol}\,(X)}{\mathrm{Vol}\,(U_i)}\right)\left(\frac{\mathrm{Vol}\,(Y)}{\mathrm{Vol}\,(U_j)}\right) \\
&- o(n^2) = o(n^2)
\end{aligned}
$$

as $\mathrm{md}_k(G_n;U_1,\dots U_k)=o(1)$. By theorem 2 of [45], it implies P0, and so, PIV. Further, by PIII+, $\mathrm{md}_1(G_n)>\theta,\dots,\mathrm{md}_{k-1}(G_n)>\theta$, so $\mathrm{md}_i(G_n;U'_1,\dots,U'_i)>\theta$ with any $(U'_1,\dots,U'_i)\in\mathcal{P}_i$ for every $i=1,\dots,k-1$. With the above argument it follows that the requirements of P_4 of [19], see (2), and those of [45] cannot be met with a multiway discrepancy $\mathrm{md}_i(G_n;U'_1,\dots U'_i)$ with $i\in\{1,\dots,k-1\}$ and $(U'_1,\dots U'_i)\in\mathcal{P}_i$. □

5. Conclusions

In course of proving equivalences between generalized quasirandom properties, we have characterized spectra and spectral subspaces of generalized quasirandom graphs; further, used a version of the expander mixing lemma to the k-cluster case and its certain converse. These theorems can give a hint for practitioners about the choice of the number of clusters.

The original expander mixing lemma and its converse (for simple, regular graphs) treat the $k=1$ case only, whereas the Szemerédi regularity lemma applies to the worst case scenario: even if there is not an underlying cluster structure, we can find cluster pairs with small discrepancy with an enormously large k (which does not depend on the number of vertices, it only depends on the discrepancy to be attained). The lemma also has constructive versions, and relations to spectra, see Szegedy [42]. Here we rather treat the intermediate case, and show that a moderate k suffices if our graph has k structural eigenvalues and a hidden k-cluster structure that can be revealed by spectral clustering tools. Under good clustering we generally understand clusters with small within- and between-cluster discrepancies that have a nonlinear dependence on the adjacencies of the graph.

In most of the real-life problems, we want just to find regular partitions that minimize discrepancies both within and between clusters, such that vertices of the same cluster behave similarly towards vertices of the same (own or other) cluster. This new paradigm for structural decomposition (see [40]) relies on minimizing the nonlinear within- and between-cluster discrepancies by means of the normalized modularity matrix. Depending on the number and sign of the so-called structural eigenvalues of this matrix, inferences can be made on the number of the underlying clusters and the type of connection between them. Furthermore, even if not all of the the generalized quasirandom properties hold, the PII $\Rightarrow$ PIII implication of theorem 1 makes it possible to reveal the structure of the actual data with spectral tools. The problem can as well be generalized to rectangular arrays, e.g. for biclustering genes and conditions of microarrays at the same time, when we want to find clusters of similarly functioning genes that equally (not especially weakly or strongly) influence conditions of the same cluster, see [7]. Sometimes, the graph sequence develops in time (e.g. internet or keyword–document matrices), but still we want to find some fixed number of underlying clusters (like user groups or topics) with spectral clustering.

We remark that if $r = \text{rank}\,(\mathbf{P}) < k$, then there are only r structural eigenvalues of the adjacency and normalized modularity matrices of the graph sequence, but the weighted multiway variance is not necessarily dropping down at r. If especially, there are identical rows in $\mathbf{P}$, and the number of distinct rows is ℓ ($r \leqslant \ell \leqslant k$), then it is md_ℓ and $\tilde{S}^2_\ell$ that drops suddenly compared to $\text{md}_{\ell-1}$ and $\tilde{S}^2_{\ell-1}$, respectively. So in practice, when we detect a gap in the spectrum between the $(r-1)$th and rth eigenvalues, it is only an indication of $k > r$ clusters. To find the optimal k, the cluster variances $\tilde{S}^2_r, \tilde{S}^2_{r+1}, \ldots$ should be examined until a 'small enough' $\tilde{S}^2_k$.

Acknowledgments

The author thanks Bojan Mohar for collaboration in proving the P0 $\Leftrightarrow$ PIV equivalence and is also indebted to Vera T Sós and László Lovász for fruitful discussions several years ago. The author also expresses her gratitude to an anonimous reviewer for useful suggestions. The research reported in this paper was supported by the Higher Education Excellence Program of the Ministry of Human Capacities in the frame of Artificial Intelligence research area of Budapest University of Technology (BME FIKP-MI/SC). The research was also supported by the National Research, Development and Innovation Fund (TUDFO/51757/2019-ITM, Thematic Excellence Program).

References

[1] Abbe E and Sandon C 2015 Community detection in general stochastic block models: fundamental limits and efficient algorithms for recovery *IEEE 56th Annual Symp. on FOCS* (arXiv:1503.00609 [MATH.PR])
[2] Alon N 1986 Eigenvalues and expanders *Combinatorica* **6** 83–96
[3] Alon N, Krivelevich M and Vu V H 2002 On the concentration of eigenvalues of random symmetric matrices *Isr. J. Math.* **131** 259–67
[4] Alon N, Coja-Oghlan A, Han H, Kang M, Rödl V and Schacht M 2010 Quasi-randomness and algorithmic regularity for graphs with general degree distributions *SIAM J. Comput.* **39** 2336–62
[5] Azran A and Ghahramani Z 2006 Spectral methods for automatic multiscale data clustering *Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition* ed A Fitzgibbon *et al* (Los Alamitos, CA: IEEE Computer Society) pp 190–7
[6] Bolla M 2005 Recognizing linear structure in noisy matrices *Linear Algebr. Appl.* **402** 228–44
[7] Bolla M, Friedl K and Krámli A 2010 Singular value decomposition of large random matrices (for two-way classification of microarrays) *J. Multivariate Anal.* **101** 434–46
[8] Bolla M 2011 Penalized versions of the Newman–Girvan modularity and their relation to normalized cuts and k-means clustering *Phys. Rev.* E **84** 016108
[9] Bolla M, Kói T and Krámli A 2012 Testability of minimum balanced multiway cut densities *Discrete Appl. Math.* **160** 1019–27
[10] Bolla M 2013 *Spectral Clustering and Biclustering* (New York: Wiley) (https://doi.org/10.1002/9781118650684)
[11] Bolla M 2014 Modularity spectra, eigen-subspaces and structure of weighted graphs *Eur. J. Combin.* **35** 105–16
[12] Bolla M 2016 Relating multiway discrepancy and singular values of nonnegative rectangular matrices *Discrete Appl. Math.* **203** 26–34
[13] Bolla M and Elbanna A 2018 Discrepancy minimizing spectral clustering *Discrete Appl. Math.* **243** 286–9
[14] Bollobás B 2001 *Random Graphs* 2nd edn (Cambridge: Cambridge University Press) (https://doi.org/10.1017/CBO9780511814068)
[15] Bollobás B, Janson S and Riordan O 2007 The phase transition in inhomogeneous random graphs *Random Struct. Algorithms* **31** 3–122

[16] Borgs C, Chayes J T, Lovász L, Sós V T and Vesztergombi K 2008 Convergent graph sequences I: subgraph frequencies, metric properties, and testing *Adv. Math.* **219** 1801–51
[17] Borgs C, Chayes J T, Lovász L, Sós V T and Vesztergombi K 2012 Convergent sequences of dense graphs II: multiway cuts and statistical physics *Ann. Math.* **176** 151–219
[18] Cheeger J 1970 A lower bound for the smallest eigenvalue of the Laplacian *Problems in Analysis* ed R C Gunning (Princeton, NJ:Princeton University Press) pp 195–9 (paper dedicated to Salmon Bochner)
[19] Chung F, Graham R and Wilson R K 1989 Quasi-random graphs *Combinatorica* **9** 345–62
[20] Chung F and Graham R 2008 Quasi-random graphs with given degree sequences *Random Struct. Algorithms* **12** 1–19
[21] Coja-Oghlan A and Lanka A 2009 Finding planted partitions in random graphs with general degree distributions *J. Discret. Math.* **23** 1682–714
[22] Diaconis P and Stroock D 1991 Geometric bounds for eigenvalues of Markov chains *Ann. Appl. Probab.* **1** 36–62
[23] Evra S, Golubev K and Lubotzky A 2015 Mixing properties and the chromatic number of Ramanujan complexes *International Mathematics Research Notices* **2015** 11520–48
[24] Fiedler M 1973 Algebraic connectivity of graphs *Czech. Math. J.* **23** 298–305
[25] Frieze A and Kannan R 1999 Quick approximation to matrices and applications *Combinatorica* **19** 175–220
[26] Hoffman A J 1972 Eigenvalues and partitionings of the edges of a graph *Linear Algebra Appl.* **5** 137–46
[27] Holland P, Laskey K B and Leinhardt S 1983 Stochastic blockmodels: some first steps *Social Netw.* **5** 109–37
[28] Hoory S, Linial N and Widgerson A 2006 Expander graphs and their applications *Bull. Am. Math. Soc.* **43** 439–561
[29] Kohayakawa Y, Rödl V, Schacht M, Sissokho P and Skokan J 2007 Turán's theorem for pseudo-random graphs *J. Comb. Theory* A **114** 631–57
[30] Lee J R, Gharan S O and Trevisan L 2012 Multi-way spectral partitioning and higher-order Cheeger inequalities *Proc. 44th Annual ACM Symp. on the Theory of Computing (New York) pp* 1117–30
[31] Lovász L 1993 Random walks on graphs: a survey *Combinatorics, Paul Erdős is Eighty. János Bolyai Society, Mathematical Studies* vol 2 (Hungary: Keszthely) pp 1–46
[32] Lovász L and Sós V T 2008 Generalized quasirandom graphs *J. Comb. Theor.* B **98** 146–63
[33] McSherry F 2001 Spectral partitioning of random graphs *Proc. 42nd Annual Symp. on Foundations of Computer Science* (Nevada: Las Vegas) pp 529–37
[34] Meila M and Shi J 2001 Learning segmentation by random walks *Proc. 13th Neural Information Processing Systems Conf.* ed T K Leen *et al* (Cambridge, MA: MIT Press) pp 873–9
[35] Mohar B 1988 Isoperimetric inequalities, growth and the spectrum of graphs *Linear Algebra Appl.* **103** 119–31
[36] Mossel E, Neeman J and Sly A 2015 Reconstruction and estimation in the planted partition model *Probab. Theor. Relat. Fields* **162** 431–61
[37] Newman E J M 2010 *Networks, an Introduction* (Oxford: Oxford University Press) (https://doi.org/10.1080.00225X.2012.744247)
[38] Ng A Y, Jordan M I and Weiss Y 2001 On spectral clustering: analysis and an algorithm *Proc. 14th Neural Information Processing Systems Conf.* ed T G Dietterich *et al* (Cambridge, MA: MIT Press) pp 849–56
[39] Ostrovsky R *et al* 2012 The effectiveness of Lloyd-type methods for the *k*-means problem *J. ACM* **59** 28
[40] Pelillo M, Elezi I and Fiorucci M 2017 Revealing structure in large graphs: szemerédi's regularity lemma and its use in pattern recognition *Patter Recognit. Lett.* **87** 4–11
[41] Simonovits M and Sós V T 1991 Szemerédi's partition and quasi-randomness *Random Struct. Algorithms* **2** 1–10
[42] Szegedy B 2011 Limits of kernel operators and the spectral regularity lemma *Eur. J. Combin.* **32** 1156–67
[43] Szemerédi E 1976 Regular partitions of graphs *Coll. Inter. CNRS. No. 260, Problémes Combinatoires et Théorie Graphes* ed J-C Bermond *et al* pp 399–401
[44] Thomason A 1987 Pseudo-random graphs *Ann. Discrete Math.* **33** 307–31
[45] Thomason A 1989 Dense expanders and pseudo-random bipartite graphs *Discret. Math.* **75** 381–6