SIAM J. DISCRETE MATH.
Vol. 23, No. 4, pp. 1682–1714

FINDING PLANTED PARTITIONS IN RANDOM GRAPHS WITH GENERAL DEGREE DISTRIBUTIONS*

AMIN COJA-OGHLAN† AND ANDRÉ LANKA‡

Abstract. We consider the problem of recovering a planted partition such as a coloring, a small bisection, or a large cut in an (apart from that) random graph. In the last 30 years many algorithms for this problem have been developed that work provably well on various random graph models resembling the Erdős–Rényi model $G_{n,m}$. In these random graph models edges are distributed uniformly, and thus the degree distribution is very regular. By contrast, the recent theory of large networks shows that real-world networks frequently have a significantly different distribution of the edges and hence also a different degree distribution. Therefore, a variety of new types of random graphs have been introduced to capture these specific properties. One of the most popular models is characterized by a prescribed expected degree sequence. We study a natural variant of this model that features a planted partition. Our main result is that there is a polynomial time algorithm for recovering (a large part of) the planted partition in this model even in the *sparse* case, where the average degree is constant. In contrast to prior work, the input of the algorithm consists *only* of the graph, i.e., no further parameters of the model (such as the expected degree sequence) are revealed to the algorithm.

Key words. random graphs, partitioning, spectral methods, algorithms

AMS subject classification. 05C80

DOI. 10.1137/070699354

1. Introduction.

To solve various types of graph partitioning problems, *spectral heuristics* are in common use. Such heuristics represent the input graph by a suitable matrix and exploit the eigenvectors of that matrix in order to solve the combinatorial problem of interest. Spectral techniques have been used to either cope with "classical" NP-hard graph partitioning problems, such as GRAPH COLORING or MAX CUT, or to solve various types of real-world "clustering problems" where the goal is to recover a "latent" partition of the vertices of a graph. In the latter case there is sometimes no objective function given that the desired partition is supposed to optimize, but the partition has some particular meaning that depends on the application context. For instance, a "cluster" could be a set vertices that span extraordinarily many edges, the idea being that such a dense spot mirrors some kind of special relationship among the vertices involved. Examples of such clustering problems occur in information retrieval [5], scientific simulation [27], or bioinformatics [15]. An important merit of spectral methods is their efficiency (there are very fast algorithms for computing eigenvectors, in particular in the case of sparse graphs/matrices) and their versatility. In the present paper we deal with spectral methods for recovering a latent but "statistically significant" partition in a sparse graph with a highly irregular degree distribution.

Despite their success in applications (e.g., [26, 27]), for most of the known spectral heuristics there are counterexamples known showing that these algorithms perform

*Received by the editors August 2, 2007; accepted for publication (in revised form) July 20, 2009; published electronically November 18, 2009.
http://www.siam.org/journals/sidma/23-4/69935.html

†University of Edinburgh, School of Informatics, 10 Crichton Street, Edinburgh EH8 9AB, UK (acoghlan@inf.ed.ac.uk). This author was supported by grant DFG CO 646.

‡Fakultät für Informatik, Technische Universität Chemnitz, Straße der Nationen 62, 09107 Chemnitz, Germany (lanka@informatik.tu-chemnitz.de).

badly in the "worst case." Thus, understanding the conditions that cause spectral heuristics to succeed (as well as their limitations) is an important research problem. To address this problem, quite a few authors have performed rigorous analyses of spectral techniques on various models of *random graphs*. Examples include Alon and Kahale [3] (Graph Coloring), Boppana [6] (Minimum Bisection), and McSherry [24] (recovering a latent partition).

Since the random graph models studied in the aforementioned papers are closely related to the simple models $G_{n,p}$ and $G_{n,m}$ pioneered by Erdős and Rényi, the resulting graphs have a very simple degree distribution. In fact, the vertex degrees are concentrated about a constant number of values. By contrast, the recent theory of complex networks has shown that in many cases real-world instances of partitioning problems have a considerably more involved degree distribution [1]. Since most spectral heuristics are extremely sensitive to fluctuations of the degree distribution, this means that most of the previous spectral methods simply fail on such inputs. For instance, none of the algorithms from [3, 6, 24] can cope with heavily tailed degree distributions such as those resulting from the ubiquitous "power law."

Therefore, in the present paper we present and analyze a spectral heuristic for partitioning random graphs with a general degree distribution (including, but not limited to "power laws"). In fact, the result applies to *sparse* graphs, i.e., the case that the average degree remains bounded as the number of vertices grows. This case is both of particular algorithmic difficulty and of utmost practical importance, as many real-world networks turn out to be sparse [1].

The present work is an extension of our prior paper [12] on the same subject. The main difference is that in this paper we present an algorithm whose input consists *only* of the graph that the algorithm is to partition. By contrast, the algorithm in [12] requires further inputs (namely, parameters of the random graph model such as the expected degree sequence), which generally will not be available in practice.

1.1. The random graph model and the main result. We consider random graphs with a planted partition and a given expected degree sequence. The model coincides with the one studied in [12] and is very similar to the model investigated in [14]. It is based on the "given expected degrees" model of Chung and Lu [8], modified so as to accommodate a planted partition. The model from [8] can be used to obtain graphs with power law degree distributions, and the same is true for our model. We refer the reader to [8, 14] for a detailed description of how to choose the parameters of the model to obtain a power law distribution with a given exponent.

Let $n > 1$ be an integer, and let $V = V_n = \{1, \dots, n\}$ be a set of nodes. The random graph model has three parameters Φ, w, and $\mathcal{V}$. The first parameter Φ is a symmetric $k \times k$ matrix $(\phi_{ij})_{1\le i,j\le k}$ with nonnegative entries, where $k > 0$ is an integer. Furthermore, $w = (w_u)_{u\in V}$ is an assignment of positive weights to the vertices of the graph. Finally, $\mathcal{V} = (V_1, \dots, V_k)$ is a partition of V into disjoint sets, which we call the *planted partition*. For each $u \in V$, we let $\psi(u)$ denote the index such that $u \in V_{\psi(u)}$. For any two vertices $u, v \in V$, we let

$$p_{uv} = \phi_{\psi(u),\psi(v)} \cdot \frac{w_u \cdot w_v}{\sum_{x\in V} w_x} \quad . \tag{1}$$

We define the random graph $G_n(\Phi, w, \mathcal{V})$ as follows: the vertex set of the graph is V, and for any $u, v \in V$ the edge $\{u, v\}$ is present with probability $\min\{p_{uv}, 1\}$ independently of all others. Here we allow that $u = v$, i.e., the random graph may contain loops. (This has some mild technical advantages. We let a loop contribute one to the vertex-degree.)

We point out that the weight w_u of each vertex u is related to its expected degree. At the end of the second section we precise this dependence and show how one has to choose the weights to obtain a concrete sequence of expected degrees.

The above model comprises a variety of random instances of clustering problems. For example, we can generate random graphs with a planted 3-coloring: let $k = 3$, let V_1, V_2, V_3 be three arbitrary sets (the "color classes"), set $\phi_{ii} = 0$, and let $\phi_{ij} > 0$ for $i \neq j$. Further possibilities are graphs with a small bisection, in which case the V_i are the two sides of the bisection, or graphs with subsets of vertices which are very dense or very sparse.

The algorithmic problem that we deal with is recovering the planted classes V_i (or large parts thereof) efficiently, given just the random graph $G = G_n(\Phi, w, \mathcal{V})$ at the input. Hence, the algorithm does not receive any further parameters of the model (e.g., the matrix Φ). Consequently, the algorithm does not know a priori what "type" of clustering problem it is dealing with, i.e., whether the goal is to find a 3-coloring, a good bisection, or something else.

THEOREM 1. *There exist*

a. *a deterministic polynomial time algorithm* $\mathcal{A}$ *and*

b. *for any* $\alpha, \varepsilon, \delta > 0$, *any integer* $k \geq 2$, *and any* $k \times k$ *matrix* $\Phi = (\phi_{ij})_{1 \leq i,j \leq k}$ *with nonnegative entries numbers* $D = D(\varepsilon, \delta, \Phi) > 0$ *and* $n_0 = n_0(\alpha, \varepsilon, \delta, \Phi)$

such that the following is true. Suppose that $n > n_0$, *that* $w = (w_1, \ldots, w_n)$ *is a tuple of positive reals, and that* $\mathcal{V} = (V_1, \ldots, V_k)$ *is a partition of* $V = \{1, \ldots, n\}$ *such that the following six conditions hold:*

C0. *Let* p_{uv} *be as in* (1). *Then* $p_{uv} \leq 1$.

C1. *The rows of* Φ *are pairwise linearly independent.*

C2. *For all* $u \in V$, *we have* $w_u \leq n^{1-\varepsilon}$.

C3. *Let* $\overline{w} = \sum_{u \in V} w_u / n$. *We have* $w_u \geq \varepsilon \cdot \overline{w}$ *for all* $u \in V$.

C4. $\overline{w} \geq D$.

C5. $|V_i| \geq \delta n$ *for all* $1 \leq i \leq k$.

Then with probability at least $1 - \alpha$, *the algorithm* $\mathcal{A}$ *applied to the random graph* $G = G_n(\Phi, w, \mathcal{V})$ *outputs a partition* $V_1', \ldots, V_k'$ *such that*

$$\sum_{i=1}^{k} |(V_i \setminus V_i') \cup (V_i' \setminus V_i)| = O\left(n / \overline{w}'^{\,0.97}\right), \quad \textit{where}$$

$$\overline{w}' = \frac{1}{n} \sum_{(u,v) \in V} p_{uv} = \sum_{(u,v) \in V} \phi_{\psi(u),\psi(v)} \frac{w_u w_v}{\overline{w} n^2} \qquad \textit{is the expected average degree.}$$

Let us briefly discuss the assumptions of Theorem 1. Conditions C0 and C1 are to ensure that the partitioning problem is well posed. For if Φ has two linearly dependent rows, then the two classes of the partition corresponding to these rows can be merged into one class without changing the probability distribution. More precisely, suppose that the first row is equal to α times the second row. Then by replacing the weights w_v for $v \in V_1$ by αw_v and the classes V_1, V_2 by $V_1 \cup V_2$ and removing the first row of Φ, we would obtain the same distribution on graphs as before but with one less planted class. Condition C2 is to make sure that the (expected) maximum degree is of lower order than n. Condition C3 states that the distribution of the vertex weights must not exhibit an extensive lower tail, and C4 requires that the average weight must be at least some (big enough) number D, which we will choose appropriately in dependence of ε, δ, and Φ. Finally, condition C5 requires that all the classes must contain a nonvanishing fraction of the vertices.

Under these assumptions the theorem states that the planted partition can be recovered with probability close to one, up to a number of $O(n/\overline{w}'^{\,0.97})$ misclassified vertices. This number decreases as $\overline{w}'$ grows, but it is linear in n if $\overline{w}'$ remains bounded as $n \to \infty$. This type of result is best possible, i.e., if $\overline{w}' = O(1)$ as $n \to \infty$, then it is in general impossible to recover the partition $V_1, \ldots, V_k$ perfectly. For instance, with high probability (w.h.p.) the random graph $G_n(\Phi, w, \mathcal{V})$ has $n \cdot \exp(-\Omega(\overline{w}'))$ isolated vertices, which the algorithm cannot possibly partition correctly.

It may be possible to reduce the number of vertices that do not get classified correctly further by combining spectral methods with combinatorial techniques (a nice example of such an approach is the paper [3] on 3-coloring random graphs). Nonetheless, in the present work we do not address this point. Instead, our main contribution is that appropriate spectral methods can be applied to sparse graphs with heavily tailed degree distributions.

To facilitate the further discussion, we say that the random graph $G_n(\Phi, w, \mathcal{V})$ has a property $\mathcal{P}$ w.h.p. if the following is true. For any $\alpha, \varepsilon, \delta > 0$ and for any matrix Φ, there are numbers $D = D(\varepsilon, \delta, \Phi) > 0$, $n_0 = n_0(\alpha, \varepsilon, \delta, \Phi) > 0$ such that for all $n > n_0$, all weight distributions w, and all partitions $\mathcal{V}$ of $V = \{1, \ldots, n\}$ such that C0–C5 are satisfied, the probability that $\mathcal{P}$ occurs in $G_n(\Phi, w, \mathcal{V})$ is at least $1 - \alpha$. This means that the probability of $\mathcal{P}$ tends to one as $n \to \infty$ uniformly w.r.t. w and $\mathcal{V}$, provided that the expected average degree $\bar{w}$ exceeds some number D that depends on $\varepsilon, \delta, \Phi$ only. Hence, the average degree is allowed to remain bounded as $n \to \infty$. Throughout the paper we will use asymptotic notation ($O(\cdot)$, $\Omega(\cdot)$, etc.) to refer to the limit $n \to \infty$, while fixing $\delta, \varepsilon, \Phi$.

1.2. Related work. The general relationship between spectral properties of, say, the adjacency matrix of a graph and clustering properties of the graph itself is well studied. Usually, spectral heuristics (try to) exploit the fact that the desired partition of the vertex set is reflected in the eigenvectors with the largest eigenvalues in absolute value, and that these are separated from the remaining eigenvalues by a significant gap. Phenomena of this type have been exploited in practice extensively. However, most spectral heuristics have a terrible worst-case performance ([28] is a notable exception). Since frequently the computational problems that spectral methods are applied to are NP-hard, this is hardly surprising.

Rigorous positive results on spectral methods have been obtained in the context of random graphs. This leads to provably efficient "average-case" algorithms for clustering problems in situations where no purely combinatorial algorithms are known to work (e.g., [2, 3, 6, 24, 20, 13]). In particular, [3] has lead to further results [16, 17]. The reason for this may be that [3] is based on a rather flexible approach for analyzing spectral properties of random graphs (based on ideas from [18]): spectral properties are inferred directly from the global edge distribution, which in turn is easy to analyze via nonconstructive counting arguments ("first moment method"). We employ an approach of this type kind, too, but we need to enhance the spectral methods of [3, 18] quite significantly because of the more general degree distribution of the graphs that we deal with.

The upper tail of the degree distribution has a dramatic impact on spectral properties of the adjacency matrix. If the graph can be described as sparse (i.e., the number of edges is linear in the number of vertices), then the eigenvalues induced by partitions of the vertices are $\Theta(\bar{d})$, where $\bar{d}$ is the average degree. But if there are vertices of very high degree Δ (say, $\Delta = n^\alpha \gg \bar{d}^2$ for some constant $\alpha > 0$), then these vertices will induce eigenvalues $\pm\sqrt{\Delta} \gg \bar{d}$ in the spectrum of the adjacency matrix. Hence, the dominant eigenvalues will *not* reflect the desired partition, but just the upper tail of the degree distribution. We refer the reader to [25] for an excellent discussion of this phenomenon. Therefore, "classical" spectral methods that rely on the adjacency matrix (or something very similar) are extremely prone to heavily tailed degree distributions and will simply fail in this case.

In the context of sparse Erdős–Rényi type graphs (e.g., $G_{n,p}$, where $p = c/n$ for some constant c) it is also true that the upper tail of the degree distribution affects the spectrum (see [23]). However, the problem is easy to remedy by just deleting the vertices of very high degree (say, more than $2np$) as observed in [3]. Since the number of these vertices is rather small (about $n \cdot \exp(-\Omega(np))$), their removal is not essential, as it does not affect the clustering properties of the graph significantly (at least if the classes of the desired partition are of linear size). By contrast, in the case of a degree distribution with a heavy upper tail this trick is not useful, because significant parts of the graph may just be ignored. This implies that spectral methods that are based on the adjacency matrix are inappropriate for graphs with heavy-tailed degree distributions.

To cope with heavily tailed degree distributions, the Laplacian matrix has been considered; see [7] for a nice exposition. It has also found its way into applications [26]. However, for randomly generated graphs the Laplacian is more difficult to handle theoretically than the adjacency matrix. This is because the entries of the Laplacian are mutually stochastically dependent. Even the Laplacian spectrum of Erdős–Rényi-type graphs is rather difficult to analyze, particularly in the sparse case [10].

Clustering problems on denser random graphs (number of edges $\gg$ number of vertices) can be solved via the Laplacian even in the case of heavily tailed degree distributions. In [14] it is shown that the singular values of a matrix similar to the Laplacian mirror the partition in a random graph model similar to the one we deal with in the present paper, provided that the expected average degree $\overline{w}'$ exceeds $\log^6 n$ *and* that the weight parameters w_u, $u \in V$, are given to the algorithms as additional input. In the dense case the spectral analysis is relatively simple, as it can be reduced to the "trace method" from [19]. This is the approach used in both [14] and [9].

The present paper builds upon our previous work [11, 12]. In [11] we studied the Laplacian eigenvalues of sparse random graphs with a given expected degree sequence. The random graph model is the same as in [9], where dense graphs were studied via different methods. In the present paper we extend the methods from [11] to graphs with a planted partition (see section 4). Furthermore, in [12] we presented an algorithm for recovering a planted partition in the same setting as we consider in the present paper, but with the additional assumption that the *expected* vertex degrees are given to the algorithm as additional input parameters. This assumption is, of course, rather artificial, but it is crucial for the analysis in [11]. Basically, in the present paper we remove this assumption by utilizing the more sophisticated methods for analyzing spectral properties of random graphs that we developed in [11]. We emphasize that also in the paper [14] it was assumed that the algorithm receives further parameters at the input (including the weight parameters w_u for all $u \in V$). Hence, the present paper contributes the first algorithm whose *only* input is the graph that we wish to partition.

In summary, the novel aspects of this paper are

- in comparison to [14] that we can deal with *sparse* graphs where the expected average degree $\overline{w}'$ remains bounded as the number n of vertices grows,
- in comparison to [12, 14] that *only* input is the graph that we need to partition, rather than the graph plus various parameters of the model, and

- in comparison to [2, 3, 6, 24, 20, 13] that we can deal with graphs with heavily tailed degree distributions.

1.3. Organization of the paper. In section 2 we introduce some notation and collect a few results that we will need throughout the paper. Section 3 contains the description of the algorithm. We also reduce the analysis of the algorithm (and thus the proof of Theorem 1) to a result on the spectral properties of a random matrix (Theorem 8 below). The proof of this result is the content of section 4.

2. Preliminaries and notation. We let $\|\cdot\|$ denote the l_2-norm of a vector or a matrix. Here by the l_2-norm of a $\mu \times \nu$ matrix M we mean

$$\|M\| = \max_{x \in \mathbf{R}^\nu : \|x\|=1} \|M \cdot x\|.$$

If M is a real symmetric $\nu \times \nu$ matrix with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_\nu$, then $\|M\| = \max\{\lambda_1, -\lambda_\nu\}$. We denote the transpose of a matrix (or vector) M as M^t. Furthermore, we let $\vec{1}$ denote the vector with all entries equal to one (in any dimension). If $\xi_1, \ldots, \xi_l$ are vectors, then $\langle \xi_1, \ldots, \xi_l \rangle$ denotes the space spanned by $\xi_1, \ldots, \xi_l$. We will occasionally employ the Courant–Fischer characterization of eigenvalues.

FACT 2. *Let $M \in \mathbf{R}^{\nu \times \nu}$ be a symmetric matrix with (real) eigenvalues $\lambda_1 \geq \cdots \geq \lambda_\nu$. Then for all $0 \leq j < n$*

$$\lambda_{j+1} = \min_{\substack{S \\ \dim S = j}} \max_{\substack{x \in S^\perp \\ \|x\|=1}} x^t M x, \lambda_{\nu-j} = \max_{\substack{S \\ \dim S = j}} \min_{\substack{x \in S^\perp \\ \|x\|=1}} x^t M x\,,$$

where S ranges over subspaces of $\mathbf{R}^\nu$ and $\dim S$ is dimension of S.

If $M = (m_{ij})_{1 \leq i \leq \mu, 1 \leq j \leq \nu}$ is a matrix and $X \subseteq \{1, \ldots, \mu\}, Y \subseteq \{1, \ldots, \nu\}$, then $M_{X \times Y}$ denotes the minor of M induced on $X \times Y$; that is,

$$M_{X \times Y} = (m_{ij})_{i \in X, j \in Y}.$$

Furthermore, we let

$$s_M(X, Y) = \sum_{x \in X} \sum_{y \in Y} m_{xy}.$$

To simplify the notation, we usually write $s_M(i, Y)$ instead of $s_M(\{i\}, Y)$.

If $X \subseteq \{1, \ldots, \mu\}$ and $v = (v_1, \ldots, v_\mu)^t \in \mathbf{R}^\mu$ is a vector, then $v_{|X}$ signifies the vector obtained from v by replacing the ith component of v by 0 if $i \notin X$ $(1 \leq i \leq \mu)$. In addition, we let $v_X = (v_i)_{i \in X}$. The difference between v_X and $v_{|X}$ is that $v_X \in \mathbf{R}^{|X|}$, while $v_{|X} \in \mathbf{R}^\mu$.

Let $G = (V, E)$ be a graph and $u \in V$. Then $N(u) = N_G(u) = \{v \in V : \{u, v\} \in E\}$ denotes neighborhood of u in G. Moreover, if $X \subseteq V$, then we let χ_X denote the characteristic vector of X. That is, $\chi_X \in \mathbf{R}^V$, and for $v \in X$ the corresponding entry $\chi_X(v)$ equals one, while for $v \in V \setminus X$ the entry $\chi_X(v)$ is zero.

We need the following Chernoff bounds on the tails of a sum of independent Bernoulli variables [21, Theorems 2.1 and 2.8].

FACT 3. *Let X be the sum of independent* 0–1 *random variables. Then for all $t \geq 0$ we have*

1. $\mathbf{Pr}\left[X \geq \mathbf{E}[X] + t\right] \leq \exp\left(-\frac{t^2}{2 \cdot (\mathbf{E}[X] + t/3)}\right)$,
2. $\mathbf{Pr}\left[X \leq \mathbf{E}[X] - t\right] \leq \exp\left(-\frac{t^2}{2 \cdot \mathbf{E}[X]}\right)$.

We also need a bit of notation concerning the random graph model $G_n(\Phi, w, \mathcal{V})$. For $U_1, U_2 \subseteq V = \{1, \ldots, n\}$, we define the *volume* as

$$\mathrm{Vol}(U_1, U_2) = \sum_{u \in U_1} \sum_{v \in U_2} \phi_{\psi(u),\psi(v)} \cdot \frac{w_u \cdot w_v}{\overline{w} \cdot n}.$$

Since any two vertices u, v of the random graph $G_n(\Phi, w, \mathcal{V})$ are connected with probability $p_{uv} = \phi_{\psi(u),\psi(v)} \cdot \frac{w_u \cdot w_v}{\overline{w} \cdot n}$ independently, we have

$$\mathrm{Vol}(U_1, U_2) = \mathbf{E}\left[s_A(U_1, U_2)\right], \text{ where } A \text{ is the adjacency matrix.}$$

We denote the degree of vertex u by d_u. The expected degree of vertex u is denoted by $w'_u = \mathbf{E}\left[d_u\right]$. Clearly,

$$w'_u = \sum_{v \in V} p_{uv} = \frac{w_u}{\overline{w} \cdot n} \cdot \sum_{v \in V} w_v \cdot \phi_{\psi(u),\psi(v)}.$$

We let

$$\overline{w}' = \sum_{u \in V} w'_u / n$$

signify the arithmetic mean of the expected degrees w'_u. Let us point out that the expected degree w'_u depends on *all* w_v's, *all* sets V_i, and the matrix Φ. Let us summarize a few basic observations.

FACT 4. *Suppose that* $(n, \Phi, w, \mathcal{V})$ *satisfy* C0–C5 *hold.*

1. *Let* u_1 *and* u_2 *be two vertices belonging to the same set of the planted partition. Then* $w_{u_1}/w'_{u_1} = w_{u_2}/w'_{u_2}$.
2. *There is a number* $C = C(\Phi, \varepsilon, \delta)$ *(independent of* n*) such that for all* $u \in V$ *we have* $1/C \le w'_u/w_u \le C$.
3. *The expected average degree of* G *equals* $\overline{w}'$.

Since w_u/w'_u is the same for all $u \in V_i$, we abbreviate

$$W_i = w_u/w'_u = \Theta(1), \qquad \text{and we let} \qquad W = \overline{w}/\overline{w}' = \Theta(1). \tag{2}$$

Instead of thinking of the "vertex weights" w_u as being given, it may sometimes be more natural to consider the *expected degrees* w'_u as given. Then the goal is to derive the vertex weights $(w_u)_{u \in V}$ that lead to the desired expected degree distribution $(w'_u)_{u \in V}$. To obtain such weights w_u, one could choose k positive constants f_i with $1 \le i \le k$ and set $w_u := f_i \cdot w'_u$ for each $u \in V_i$ and each $1 \le i \le k$. The f_i are to compensate the effect of the W_i above. We illustrate the way to find the f_i's by an example.

Suppose we want to plant a small bisection. To this end, we set $\Phi = \left(\begin{smallmatrix} 1 & 0.5 \\ 0.5 & 1 \end{smallmatrix}\right)$, and let $\mathcal{V} = (V_1, V_2)$ be the planted partition. We also assume that we are given the expected degree distribution $(w'_u)_{u \in V}$. To simplify the following calculation, we assume that the sum of the expected degrees on both side of the bisection is the same, that is

$$\sum_{v \in V_1} w'_v = \sum_{v \in V_2} w'_v =: D. \tag{3}$$

Our goal is to find factors f_i and thus weights w_u such that for $u \in V_1$

$$\sum_{v \in V_1} 1 \cdot \frac{w_u \cdot w_v}{\overline{w} \cdot n} + \sum_{v \in V_2} \frac{1}{2} \cdot \frac{w_u \cdot w_v}{\overline{w} \cdot n} = w'_u$$

is satisfied. As for $v \in V_1$ (resp., $v \in V_2$), we have $w_v = f_1 \cdot w'_v$ (resp., $w_v = f_2 \cdot w'_v$); the equality of

$$\sum_{v \in V_1} 1 \cdot \frac{(f_1 \cdot w'_u) \cdot (f_1 \cdot w'_v)}{\overline{w} \cdot n} + \sum_{v \in V_2} \frac{1}{2} \cdot \frac{(f_1 \cdot w'_u) \cdot (f_2 \cdot w'_v)}{\overline{w} \cdot n} = w'_u$$

is needed. This is equal to

$$f_1^2 \cdot \sum_{v \in V_1} w'_v + \frac{f_1 \cdot f_2}{2} \cdot \sum_{v \in V_2} w'_v = \overline{w} \cdot n,$$

and by (3)

$$f_1^2 \cdot D + \frac{f_1 \cdot f_2}{2} \cdot D = \overline{w} \cdot n. \tag{4}$$

For the right-hand side $\overline{w} \cdot n$, we get

$$\overline{w} \cdot n = \sum_{v \in V_1} w_v + \sum_{v \in V_2} w_v = f_1 \cdot \sum_{v \in V_1} w'_v + f_2 \cdot \sum_{v \in V_2} w'_v = (f_1 + f_2) \cdot D.$$

Plugging this into (3) we see that we just need to find positive numbers f_1, f_2 satisfying

$$f_1^2 + \frac{f_1 \cdot f_2}{2} = f_1 + f_2. \tag{5}$$

Repeating the calculation above for some $u \in V_2$ we get analogously

$$\frac{f_1 \cdot f_2}{2} + f_2^2 = f_1 + f_2.$$

This yields $f_1^2 = f_2^2$. Since both are positive, we have $f_1 = f_2$. Using this we get by (5) that $f_1 = f_2 = 4/3$. So, one may obtain the desired sequence of expected degrees w'_u by setting $w_u := 4/3 \cdot w'_u$.

In the above bisection example we were able to find multipliers f_1, f_2 to adapt a given expected degree distribuion. However, in other examples it is not possible to achieve any prescribed sequence of expected degrees. To see this, suppose we plant an independent set V_1. Then $\Phi = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 1 \end{smallmatrix}\right)$. In this case it is impossible to model a sequence where the sum of the expected degrees in V_1 is larger than the sum of the expected degrees in V_2.

3. The algorithm.

3.1. Background: Representing graphs by matrices. In this section we discuss various ways of representing a graph by a matrix. Apart from a few definitions, the section serves purely didactical purposes: it is just meant to facilitate the reader's understanding of the algorithm, which we will describe in the next section. Therefore, we will omit the proofs of a few statements, as none of them will be needed in the proof of Theorem 1.

Recall that if $G = (V, E)$ is a graph, then its *adjacency matrix* $A = A(G) = (a_{vw})_{v,w \in V}$ has entries

$$a_{vw} = \begin{cases} 1 & \text{if } v, w \text{ are adjacent,} \\ 0 & \text{otherwise.} \end{cases}$$

To illustrate the use of this matrix for graph partitioning purposes, let us assume that

(6) the vertex set V has a partition V_1, V_2 into two (disjoint) sets of equal size such that every vertex in V_i has exactly d neighbors in V_{3-i} and exactly $2d$ neighbors in V_i for $i = 1, 2$, where $d \geq 3$ is some integer.

That is, V_1, V_2 is a "good bisection" of G. Let $\xi_i = (\xi_i(v))_{v \in V} = \chi_{V_i}$ be the vector whose entries are one for $v \in V_i$ and 0 for $v \in V_{3-i}$ and $i = 1, 2$. Then the following vectors are eigenvectors of A:

1. $\vec{1} = \xi_1 + \xi_2$ is an eigenvector with eigenvalue $3d$.
2. $\xi_1 - \xi_2$ is an eigenvector with eigenvalue d.

If all other eigenvalues of A are less than d (in absolute value), then it is easy to recover the partition of V_1, V_2 from A. Namely, we just compute any eigenvector ξ with eigenvalue d. Then ξ lies on the line spanned by $\xi_1 - \xi_2$. Hence, ξ is constant on both classes V_1, V_2. Thus, if we let V_1' be the set of vertices $v \in V$ whose corresponding entry ξ_v is positive and $V_2' = V \setminus W_1$, then either $V_1' = V_1$ and $V_2' = V_2$, or $V_2' = V_1$ and $V_1' = V_2$.

Of course, the crucial property is that all but two eigenvalues of A are less than d in absolute value. This property does not hold for all graphs that satisfy (6). However, if G is chosen *uniformly at random* from the set of all graphs that satisfy (6), then all but two eigenvalues of A are in fact $O(\sqrt{d})$ in absolute value w.h.p. This can be shown via techniques from [18]. (Besides, (V_1, V_2) is the optimal bisection w.h.p., provided that d is sufficiently large.)

Let us now consider a slightly different (conceptually simpler) random graph model. Namely, let $V = \{1, \ldots, n\}$ for some even integer n, and let (V_1, V_2) be any partition of V into two sets of size $n/2$. Now, each edge $\{v, v'\}$ is present in the random graph G with probability $4d/n$ if either $v, v' \in V_1$ or $v, v' \in V_2$, and with probability $2d/n$ if $v \in V_1$ and $v' \in V_2$ independently of all other edges. Thus, the *expected* number of neighbors that $v \in V_i$ has in its own class V_i is $2d$, and the *expected* number of neighbors in the opposite class V_{3-i} is d. This random graph model can be expressed as $G = G_n(\Phi, w, \mathcal{V})$ by letting $\Phi_{11} = \Phi_{22} = 4d$, $\Phi_{12} = \Phi_{21} = 2d$, $w_v = 1$ for all v, and $\mathcal{V} = (V_1, V_2)$. We consider d fixed as n grows.

Although the *expected* number of neighbors that each vertex has in V_1, V_2 is as indicated in (6), the actual numbers vary. More precisely, for each $v \in V_i$ the number of neighbors in V_i (respectively, V_{3-i}) is asymptotically Poisson with mean $2d$ (respectively, d). This implies that the maximum degree of the graph is as large as $\Theta(\ln n / \ln \ln n)$ w.h.p., and for each fixed number $0 \leq \Delta = O(1)$ there are $\Omega(n)$ vertices of degree Δ w.h.p.

Let us consider a vertex v of degree $\Delta = 4d^2$. Since $d = O(1)$ as $n \to \infty$, the random graph G is sparse, and therefore the subgraph of G induced on v and its neighbors is a star w.h.p. This implies that the spectrum of the adjacency matrix $A(G)$ contains the eigenvalues of the adjacency matrix of a star $K_{1,\Delta}$, which are $\pm\sqrt{\Delta} = \pm 2d$. This exceeds the eigenvalue d that corresponds to the bisection V_1, V_2. Hence, the bisection cannot be recovered by considering the eigenvector with the second largest eigenvalue. In fact, the spectrum of $A(G)$ contains eigenvalues $\pm\sqrt{\Delta}$ for any fixed $\Delta > 4d^2$ as well. In other words, the upper tail of the degree distribution clutters the spectrum of A wildly, so that it is not straightforward anymore to read the partition off (see [25] for a comprehensive account).

As observed in [3], in the above model this problem is easy to fix. Let G' be the subgraph of G obtained by removing all vertices of degree larger than, say, $4d$ from G. Then the eigenvector η with the second largest eigenvalue of $A(G')$ is "close" in the

ℓ_2-norm to the space spanned by $\vec{1}$ and $\xi_1 - \xi_2$ w.h.p. This implies that η can be used to partition *most* vertices of G' correctly (all but an $O(1/d)$ fraction). Furthermore, the number of vertices of degree bigger than $4d$ is less than $n \cdot \exp(-\Omega(d))$ w.h.p.

The simple trick of just removing vertices of degree bigger than $4d$ works fine in the above model, because the expected degree of any vertex in the graph equals $3d$. If, however, the degree distribution has a significant upper tail, then there may be very many vertices with degree far higher than the average, and therefore removing high degree vertices is not an option. (To obtain this type of graph in the $G_n(\Phi, w, \mathcal{V})$ model, choose Φ and $\mathcal{V}$ as before, but let some weights w_v be larger than one.)

Instead, we will use a different matrix to represent the graph. The basic idea is to normalize the entry corresponding to an edge $\{v, v'\}$ by (some function of) the degrees of v, v'. More precisely, given a graph $G = (V, E)$, we let $M' = M'(G) = (m_{uv})_{u,v \in V}$ be the matrix with entries

$$m'_{uv} = \begin{cases} (d_u d_v)^{-1} & \text{if } u, v \text{ are adjacent,} \\ 0 & \text{otherwise,} \end{cases}$$

where d_v is the degree of v. This is reminiscent of the normalized Laplacian, where the normalization is by the geometric mean $\sqrt{d_u d_v}$ rather than the product $d_u d_v$. Using the latter has certain technical advantages in the present context. In any case, the basic idea is to reduce the weight of high degree vertices, rather than to remove them completely.

As we shall see in Theorem 8 below, the spectrum of M' still mirrors the partition $\mathcal{V}$. More precisely, in the above bisection problem the partition $\mathcal{V}$ induces eigenvalues of order $1/d$. Furthermore, the normalization diminishes the impact of the high vertex degrees on the spectrum. However, there is a new issue: vertices of very *low* degree. The extreme example is an isolated edge $e = \{v, v'\}$. In the random graph $G = G_n(\Phi, w, \mathcal{V})$ described above, there are $\Omega(n)$ isolated edges w.h.p. Each of them yields a 2×2 submatrix of M' whose eigenvalues are ± 1 and thus exceed the eigenvalue of order $1/d$ that corresponds to the planted bisection. To resolve this problem, the algorithm will not work with the matrix M', but with a minor M of this matrix induced on the set of vertices whose degree is at least a certain value d_m. As we shall see, our assumption C3 from section 1.1 ensures that it is feasible to ignore low degree vertices.

As far as the *analysis* of the algorithm is concerned, the matrices M' and M are significantly more difficult to deal with than the adjacency matrix A. This is because if $G = G_n(\Phi, w, \mathcal{V})$, then the entries of A are mutually independent random variables (apart from the obvious dependency resulting from the fact that A is symmetric). By contrast, the entries of M' are mutually dependent, because we divide by the vertex degrees (which are, of course, random variables). Coping with this stochastic dependence will be our main technical challenge.

3.2. Description of the algorithm. For the rest of section 3, we fix $\varepsilon, \delta > 0$ and the $k \times k$ matrix Φ. We let $D = D(\varepsilon, \delta, \Phi)$ and $n_0 = n_0(\varepsilon, \delta, \Phi)$ be large enough numbers and assume that $n > n_0$ and that the weight distribution w and the partition $\mathcal{V} = (V_1, \ldots, V_k)$ are such that $\overline{w} \geq D$. In addition, we assume that conditions C0–C5 hold.

The pseudocode for the graph partitioning algorithm is shown in Figure 1. Steps 1–4 of Algorithm 5 set up the matrix representation from the previous section. As we pointed out there, in order to ensure that the spectrum of the matrix mirrors the desired partition, it is necessary to remove vertices of “atypically low degree.” Instead

ALGORITHM 5. *Input:* A graph $G = (V, E)$ with vertex set $V = \{1, \dots, n\}$. *Output:* A partition of V.

1. Compute $\overline{d} = \sum_{u=1}^{n} d_u/n$ and set $d_{\mathrm{m}} = \overline{d} / \ln \overline{d}$.
2. Set up the matrix $M' = (m'_{uv})_{u,v \in V}$ with entries $m'_{uv} = 1/(d_u \cdot d_v)$ if $\{u, v\} \in E$ and $m'_{uv} = 0$ otherwise.
3. Let $U = \{u \in V : d_u \geq d_{\mathrm{m}}\}$.
4. Obtain M from M' by replacing any entry m'_{uv} with $(u, v) \notin U \times U$ by 0.
5. Compute the eigenvalues $\Lambda_1, \dots, \Lambda_n$ of M and order them such that $|\Lambda_1| \geq \dots \geq |\Lambda_n|$. Let $1 \leq \kappa \leq n$ be such that $|\Lambda_\kappa| \geq 1/{d_{\mathrm{m}}}^{1.1} > |\Lambda_{\kappa+1}|$. Compute a family $s_1, \dots, s_\kappa$ of mutually perpendicular vectors of ℓ_2-norm $\sqrt{n}$ such that s_i is an eigenvector of M with eigenvalue Λ_i for all $1 \leq i \leq \kappa$.
6. Call Algorithm 7 with input $(G, d_{\mathrm{m}}, s_1, \dots, s_\kappa)$ to obtain a partition of V.

FIG. 1. *The graph partitioning algorithm, part* 1.

of actually removing these vertices, step 4 of the algorithm just replaces the entries in the rows and columns corresponding to these vertices by 0.

Since the algorithm does not know the *expected* vertex degrees in the random graph $G_n(\Phi, w, \mathcal{V})$, it has to come up with a good "guess" of a lower bound on the vertex degrees. This guess is $d_{\mathrm{m}} = \bar{d}/\ln \bar{d}$. To see that this makes sense, we need to show that

$$d_{\mathrm{m}} \ll \min_{v \in V} w'_v = \min_{\substack{1 \leq i,j \leq 2 \\ \phi_{ij} > 0}} \min_{u \in V_i} \sum_{v \in V_j} \phi_{ij} \cdot \frac{w_u \cdot w_v}{\overline{w} \cdot n},$$

i.e., we need to derive a lower bound on the minimum expected degree. By assumption C3 we have $w_u \geq \varepsilon \bar{w}$ for all $u \in V$, and C5 ensures that $|V_i| \geq \delta n$. Therefore,

$$\sum_{v \in V_j} \phi_{ij} \cdot \frac{(\varepsilon \cdot \overline{w})^2}{\overline{w} \cdot n} \geq (\delta \cdot n) \cdot \phi_{ij} \cdot \frac{\varepsilon^2 \cdot \overline{w}}{n} \geq \delta \cdot \phi_{ij} \cdot \varepsilon^2 \cdot \overline{w} = \Theta(\overline{w}).$$

Furthermore, we have the following lower bound on $\bar{d}$. Remember that

$$\bar{w}' = \frac{1}{\bar{w} n^2} \sum_{(u,v) \in V \times V} w_u w_v \phi_{\psi(u), \psi(v)}.$$

FACT 6. *If $G = G_n(\Phi, w, \mathcal{V})$, then $\bar{d} \sim \bar{w}'$ w.h.p. Consequently, w.h.p. $\bar{d} = \Theta(\overline{w})$, $d_{\mathrm{m}} > \overline{w}^{2/3}$, $d_{\mathrm{m}} > \overline{w}'^{\,2/3}$, and $d_{\mathrm{m}} \leq (1 + o(1))\overline{w}'/\ln \overline{w}'$.*

Proof. Since $\bar{d}$ is the average degree of G, $n\bar{d}/2$ equals the total number of edges. For any two vertices u, v, the edge $\{u, v\}$ is present in G with probability p_{uv} independently. Therefore, $n\bar{d}/2$ is a sum of independent Bernoulli variables, and hence the Chernoff bound implies that $n\bar{d}/2 \sim \mathrm{E}(n\bar{d}/2)$ w.h.p. Hence, $\bar{d} \sim \mathrm{E}(\bar{d}) = \bar{w}'$ w.h.p. Since the matrix Φ remains fixed as $n \to \infty$, we have $\bar{w}' = \Theta(\bar{w})$, and thus $\bar{d} = \Theta(\bar{w})$ w.h.p. Hence, C3 entails that $d_{\mathrm{m}} = \bar{d}/\ln \bar{d} = \Theta(\bar{w}/\ln \bar{w})$, and $d_{\mathrm{m}} > \overline{w}^{2/3}$ and $d_{\mathrm{m}} > \overline{w}'^{\,2/3}$ w.h.p. ∎

Having set up the matrix M, the algorithm proceeds to compute an orthogonal family $(s_1, \dots, s_\kappa)$ of eigenvectors whose corresponding eigenvalues exceed ${d_{\mathrm{m}}}^{-1.1}$, i.e., are "big." In the analysis of the algorithm we will see that these vectors are closely related to the characteristic vectors of the classes $V_1, \dots, V_k$. Roughly speaking, we will see that the entries of (most) vertices that belong to the same class V_j are essentially identical in all the vectors s_j. On the other hand, for most pairs u, v of vectors that belong to different classes there is at least one j such that the entries of u and v in s_j differ significantly.

ALGORITHM 7. *Input:* The graph G along with vectors $s_1, \ldots, s_\kappa$ and a number d_{m}. *Output:* A partition of V.

1. Let $P := \{V\}$.
2. While there is $V' \in P$ such that there exist $s \in \{s_1, \ldots, s_\kappa\}$ and $l_1 < l_2 < l_3$ such that $f_{V',s}(l_1) = f_{V',s}(l_3) = 1$ and $f_{V',s}(l_2) = f_{V',s}(l_2+1) = 0$
3. set $V'' := \{v \in V' : s(v) < (l_2+1)/\ln d_{\mathrm{m}}\}$ and replace V' in P by V'' and $V' \setminus V''$.
4. Output P.

FIG. 2. *The graph partitioning algorithm, part* 2.

In order to actually partition V we exploit this fact as follows. We start from the trivial partition $P = \{V\}$ and keep refining the partition iteratively as follows. For a set $V' \subseteq V$ and a vector $s \in \mathbf{R}^n$, we define

$$(7)\quad f_{V',s} : \mathbf{Z} \to \{0,1\}, \quad l \mapsto \begin{cases} 1 & \text{if } \left|\left\{v \in V' : \frac{l}{\ln d_{\mathrm{m}}} \le s(v) < \frac{l+1}{\ln d_{\mathrm{m}}}\right\}\right| > \frac{n}{{d_{\mathrm{m}}}^{0.97}}, \\ 0 & \text{otherwise.} \end{cases}$$

We say that an integer l is a *clusterpoint* if $f_{V',s}(l) = 1$. That is, l is a clusterpoint iff there are "a lot" of vertices $v \in V'$ such that their corresponding entry in s lies in the "small" interval $[\frac{l}{\ln d_{\mathrm{m}}}, \frac{l+1}{\ln d_{\mathrm{m}}})$. If l is not a clusterpoint, then we say that l is a *gap*. Now, if the present partition P contains a class V' such that there is a vector $s \in \{s_1, \ldots, s_\kappa\}$ such that V' has two clusterpoints that are separated by at least two subsequent gaps, then V' gets replaced by the set V'' that corresponds to the clusterpoints to the left-hand side of the gap and the set that corresponds to the right-hand side. To be precise, the algorithm proceeds as shown in Figure 2.

Algorithms 5 and 7 have a polynomial running time. For steps 1–4 of Algorithm 5, this is evident. The eigenvalue/eigenvector computation in step 5 can be carried out in polynomial time within any numerical precision. (In fact, it can be shown that in the random graph model $G_n(\Phi, w, \mathcal{V})$ each relevant eigenvector can be computed via $O(\ln n)$ Lanczos iterations.) Furthermore, the main loop of Algorithm 7 gets executed at most n times, because each time one partition class of P gets split into two non-empty parts V'' and $V \setminus V''$, and the vertex set V has n elements. Since all eigenvectors $s(v)$ have ℓ_2-norm $\sqrt{n}$ by construction (cf. step 5 of Algorithm 5), the support of each function $f_{V',s}$ is a subset of $(-\sqrt{n} \cdot \ln n, \sqrt{n} \cdot \ln n)$. Hence, each iteration of steps 2–3 of Algorithm 7 can be executed in polynomial time.

3.3. Proof of Theorem 1. In this section we show that Theorem 1 follows from the following statement about the spectral properties of the matrix M, the proof of which we defer to section 4. Recall the definition of the numbers W_i from (4).

THEOREM 8. *W.h.p. the following two statements hold.*

1. *For all* $1 \le i, j \le k$ *we have*

$$(8)\quad \frac{\vec{1}^t}{\|\vec{1}^t\|} \cdot M_{V_i \times V_j} \cdot \frac{\vec{1}}{\|\vec{1}\|} = \phi_{ij} \cdot W_i \cdot W_j \cdot \frac{\sqrt{|V_i| \cdot |V_j|}}{\overline{w} \cdot n} \cdot \left(1 + O\left({d_{\mathrm{m}}}^{-0.49}\right)\right).$$

2. *For all* $1 \le i, j \le k$ *and any* u, v *with* $\|u\| = \|v\| = 1$ *and* $u \perp \vec{1}$ *or* $v \perp \vec{1}$ *we have*

$$(9)\quad \left|u^t \cdot M_{V_i \times V_j} \cdot v\right| = O\left(1/\overline{w}^{1.49} + 1/{d_{\mathrm{m}}}^{1.5}\right) = O(1/\overline{w}^{1.49}).$$

As ϕ_{ij} and W_i remain constant as $n \to \infty$, we often hide them (as in 1. above) in the $O(\cdot)$.

In section 3.4 we will establish the following lemma.

LEMMA 9. *Let k' be the rank of the matrix Φ. If* (8) *and* (9) *hold, then M has exactly k' eigenvalues whose absolute value is $\Theta(1/\overline{w})$, whereas all the remaining eigenvalues are $O\left({d_{\mathrm{m}}}^{-0.49}/\overline{w}\right)$ in absolute value.*

In combination with Theorem 8, Lemma 9 shows that step 5 of Algorithm 5 correctly identifies (without a priori knowing k') the k' eigenvectors with the dominant eigenvalues. Since $d_{\mathrm{m}} = \overline{d} / \ln \overline{d} = \Theta(\overline{w}/\ln \overline{w})$ w.h.p. by Fact 6 the term $1/{d_{\mathrm{m}}}^{1.1}$ is smaller than the $\Theta(1/\overline{w})$ term and bigger than the $O({d_{\mathrm{m}}}^{-0.49}/\overline{w})$ term.

The following lemma describes the structure of the eigenvectors with the dominant eigenvalues. It shows that these eigenvectors essentially result from the characteristic vectors of the planted partition classes. In addition, the lemma shows that the eigenvectors are essentially constant on the planted partition classes. Furthermore, we will see that for each pair of distinct classes V_j, $V_{j'}$ there is at least one eigenvector that assigns significantly different values to these two sets. This will put us in a position to prove that Algorithm 7 recovers an approximation to the planted partition as claimed in Theorem 1 w.h.p.

LEMMA 10. *Let k' be the rank of the matrix Φ. Let $s_1, s_2, \ldots, s_{k'}$ be mutually orthogonal eigenvectors with the k' largest eigenvalues in absolute value of M such that $\|s_i\| = \sqrt{n}$ for all $1 \le i \le k'$. Let $\chi_1, \ldots, \chi_k$ be the characteristic vectors of $V_1, \ldots, V_k$. If* (8) *and* (9) *hold, then the unique decomposition*

$$s_i = \alpha_{i1} \cdot \chi_1 + \alpha_{i2} \cdot \chi_2 + \cdots + \alpha_{ik} \cdot \chi_k + \gamma_i \cdot u_i,$$

with $u_i \perp \chi_1, \ldots, \chi_k$ and $\|u_i\| = \sqrt{n}$ has the following properties.

1. $|\gamma_i| = O({d_{\mathrm{m}}}^{-0.49})$ *for* $1 \le i \le k'$.
2. *For each $i \in \{1, \ldots, k'\}$ and every $j \in \{1, \ldots, k\}$ there are at most $O(n \cdot \ln^4 d_{\mathrm{m}}/{d_{\mathrm{m}}}^{0.98})$ vertices $v \in V_j$ such that $|\alpha_{ij} - s_i(v)| \ge 1/\ln^2 d_{\mathrm{m}}$.*
3. *For any two distinct indices $1 \le j, j' \le k$ there is $i \in \{1, \ldots, k'\}$ such that*

$$|\alpha_{ij} - \alpha_{ij'}| > 1/\sqrt{\ln d_{\mathrm{m}}}.$$

We defer the proof of Lemma 10 to section 3.5.

Proof of Theorem 1. The values of α_{ij} match the clusterpoints mentioned in section 3.2. Lemma 10 entails that for most $v \in V_j$ the entry $s_i(v)$ deviates from α_{ij} by at most $1/\ln^2 d_{\mathrm{m}}$. As the interval size in the functions $f_{V',s}$ used in Algorithm 7 is $1/\ln d_{\mathrm{m}}$ (see (7)), the typical entries lie in two subsequent intervals, and the entries of at most $O(n \cdot \ln^4 d_{\mathrm{m}}/{d_{\mathrm{m}}}^{0.98})$ vertices $v \in V_j$ do not lie in these intervals. This shows that in total no more than $k \cdot O(n \cdot \ln^4 d_{\mathrm{m}}/{d_{\mathrm{m}}}^{0.98}) < n/{d_{\mathrm{m}}}^{0.97}$ entries of s_i lie outside of such intervals. As all intervals with $\le n/{d_{\mathrm{m}}}^{0.97}$ entries are gaps, each vector s_i has at most $2 \cdot k$ intervals that are clusterpoints.

For two different sets V_j and $V_{j'}$ there is at least one s_i such that the corresponding clusterpoints have a distance of at least $1/\sqrt{\ln d_{\mathrm{m}}}$. Hence, the intervals belonging to these clusterpoints are well separated (there are at least $\sqrt{\ln d_{\mathrm{m}}} - 4$ intervals between them). As we have at most $2k$ nongap intervals, there must be two consecutive gaps. Therefore, Algorithm 7 splits V_j and $V_{j'}$ into different sets of the partition it constructs.

Since we split each set between two gaps, it is impossible to cut some V_j "in the middle." Thus, the partition constructed agrees in large parts the planted partition. The difference between the partition planted and the constructed one is bounded by the number of entries being far away from the clusterpoints, that is, $k \cdot O(n \cdot \ln^4 d_{\mathrm{m}}/{d_{\mathrm{m}}}^{0.98})$. By our choice of $d_{\mathrm{m}} = \overline{d}/\ln \overline{d}$, this is bounded from above by $n/\overline{d}^{\,0.97}$. This completes the proof of Theorem 1. ∎

3.4. Proof of Lemma 9. Let $\chi_1, \dots, \chi_k$ be the characteristic vectors of $V_1, \dots V_k$ (i.e., the uth component of χ_i equals one if $u \in V_i$ and 0 otherwise for all $u \in V$). Consider two unit vectors g and h from the space $\langle \chi_1, \dots, \chi_k \rangle$ spanned by $\chi_1, \dots, \chi_k$. These vectors have decompositions $g = a_1 \cdot \chi_1/\|\chi_1\| + \cdots + a_k \cdot \chi_k/\|\chi_k\|$ and $h = b_1 \cdot \chi_1/\|\chi_1\| + \cdots + b_k \cdot \chi_k/\|\chi_k\|$, where $\sum_{i=1}^k a_i^2 = \sum_{i=1}^k b_i^2 = 1$. Since we are assuming that (8) and (9) hold, we have

$$\begin{aligned}
h^t M g &= \sum_{i,j=1}^{k} b_i \cdot \frac{\chi_i^t}{\|\chi_i\|} \cdot M \cdot a_j \cdot \frac{\chi_j}{\|\chi_j\|} = \sum_{i,j=1}^{k} b_i \cdot a_j \cdot \frac{\vec{1}^t \cdot M_{V_i \times V_j} \cdot \vec{1}}{\sqrt{|V_i| \cdot |V_j|}} \\
&= \sum_{i,j=1}^{k} b_i \cdot a_j \cdot \phi_{ij} \cdot W_i \cdot W_j \cdot \frac{\sqrt{|V_i| \cdot |V_j|}}{\overline{w} \cdot n} \cdot \left(1 + O\left({d_{\mathrm{m}}}^{-0.49}\right)\right) \\
&= \sum_{i,j=1}^{k} \left(b_i \cdot a_j \cdot \phi_{ij} \cdot W_i \cdot W_j \cdot \frac{\sqrt{|V_i| \cdot |V_j|}}{\overline{w} \cdot n} \right) + O\left(\frac{{d_{\mathrm{m}}}^{-0.49}}{\overline{w}} \right) \\
&= \frac{1}{\overline{w}} \cdot \begin{pmatrix} b_1 & \dots & b_k \end{pmatrix} \cdot P \cdot \begin{pmatrix} a_1 \\ \vdots \\ a_k \end{pmatrix} + O\left(\frac{{d_{\mathrm{m}}}^{-0.49}}{\overline{w}} \right),
\end{aligned}$$

where

$$P = \begin{pmatrix} W_1 \cdot \sqrt{\frac{|V_1|}{n}} & & 0 \\ & \ddots & \\ 0 & & W_k \cdot \sqrt{\frac{|V_k|}{n}} \end{pmatrix}$$

$$\times \begin{pmatrix} \phi_{11} & \dots & \phi_{1k} \\ \vdots & \ddots & \vdots \\ \phi_{k1} & \dots & \phi_{kk} \end{pmatrix} \cdot \begin{pmatrix} W_1 \cdot \sqrt{\frac{|V_1|}{n}} & & 0 \\ & \ddots & \\ 0 & & W_k \cdot \sqrt{\frac{|V_k|}{n}} \end{pmatrix}. \tag{10}$$

As $W_i, |V_i| > 0$ for all i, the outer factors in (10) have full rank. Therefore, the rank of P equals the rank of Φ, which is k'. Consequently, P has exactly k' eigenvectors with nonzero eigenvalues $\nu_1, \dots, \nu_{k'}$. As $W_i, |V_i|/n = \Theta(1)$, we conclude that $\nu_1, \dots, \nu_{k'}$ are bounded away from 0 by some constant and that their absolute value is bounded from above by some constant.

Let $\begin{pmatrix} e_1 & \dots & e_k \end{pmatrix}^t$ and $\begin{pmatrix} f_1 & \dots & f_k \end{pmatrix}^t$ be two orthonormal eigenvectors of P with the eigenvalues ν_1 and ν_2. Set

$$g_1 = e_1 \cdot \frac{\chi_1}{\|\chi_1\|} + \cdots + e_k \cdot \frac{\chi_k}{\|\chi_k\|} \qquad \text{and} \qquad g_2 = f_1 \cdot \frac{\chi_1}{\|\chi_1\|} + \cdots + f_k \cdot \frac{\chi_k}{\|\chi_k\|}.$$

Clearly, g_1 and g_2 are orthonormal as well, and the above computation yields

$$\begin{aligned}\left|g_1^t \cdot M \cdot g_1\right| &= \left|\frac{1}{\overline{w}} \cdot \begin{pmatrix} e_1 & \dots & e_k \end{pmatrix} \cdot P \cdot \begin{pmatrix} e_1 \\ \vdots \\ e_k \end{pmatrix} \pm O\left(\frac{{d_{\mathrm{m}}}^{-0.49}}{\overline{w}}\right)\right| \\ &= \left|\frac{1}{\overline{w}} \cdot \nu_1 \pm O\left(\frac{{d_{\mathrm{m}}}^{-0.49}}{\overline{w}}\right)\right| = \Theta\left(\frac{1}{\overline{w}}\right) \qquad \text{and} \\ \left|g_1^t \cdot M \cdot g_2\right| &= \left|\frac{1}{\overline{w}} \cdot \begin{pmatrix} e_1 & \dots & e_k \end{pmatrix} \cdot P \cdot \begin{pmatrix} f_1 \\ \vdots \\ f_k \end{pmatrix} \pm O\left(\frac{{d_{\mathrm{m}}}^{-0.49}}{\overline{w}}\right)\right| \\ &= \left|\frac{1}{\overline{w}} \cdot 0 \pm O\left(\frac{{d_{\mathrm{m}}}^{-0.49}}{\overline{w}}\right)\right| = O\left(\frac{{d_{\mathrm{m}}}^{-0.49}}{\overline{w}}\right).\end{aligned}$$

Proceeding inductively, we obtain k' orthonormal vectors $g_1, \dots, g_{k'}$ such that

$$\left|g_i^t \cdot M \cdot g_j\right| = \begin{cases} \Theta(1/\overline{w}) & \text{for } i = j, \\ O\left({d_{\mathrm{m}}}^{-0.49}/\overline{w}\right) & \text{for } i \neq j\,. \end{cases} \tag{11}$$

Similarly, the kernel of P yields $k - k'$ orthonormal vectors $g_{k'+1}, \dots, g_k$ satisfying

$$\left|g^t \cdot M \cdot g_i\right| = O\left({d_{\mathrm{m}}}^{-0.49}/\overline{w}\right) \tag{12}$$

for any unit vector $g \in \langle g_1, \dots, g_k\rangle$. Note that $\langle g_1, \dots, g_k\rangle = \langle \chi_1, \dots, \chi_k\rangle$.

Let us order $g_1, \dots, g_{k'}$ by the sign of $g_i^t \cdot M \cdot g_i$: let $g_1, \dots, g_l$ be the vectors such that $g_i^t \cdot M \cdot g_i > 0$ and $g_{l+1}, \dots, g_{k'}$ be the vectors with $g_i^t \cdot M \cdot g_i < 0$.

Let x be some unit vector from the space $\langle g_1, \dots, g_l\rangle$. Then there is a decomposition $x = a_1 \cdot g_1 + \dots + a_l \cdot g_l$ with $\sum_{i=1}^l a_i^2 = 1$. Hence,

$$\begin{aligned} x^t M x &= \sum_{i=1}^l a_i^2 \cdot g_i \cdot M \cdot g_i + \sum_{\substack{i,j=1 \\ i \neq j}}^l a_i \cdot a_j \cdot g_i^t \cdot M \cdot g_j \\ &= \Theta(1/\overline{w}) + l^2 \cdot O\left({d_{\mathrm{m}}}^{-0.49}/\overline{w}\right) = \Theta(1/\overline{w}). \end{aligned}$$

Applying the second equality of Fact 2 with $j = n - l$ and $S = \langle g_1, \dots, g_l\rangle^\perp$, we get

$$\lambda_l \geq \min_{\substack{x \in S^\perp \\ \|x\|=1}} x^t A x = \min_{\substack{x \in \langle g_1, \dots, g_l\rangle \\ \|x\|=1}} x^t A x = \Theta(1/\overline{w}).$$

Using the first inequality of Fact 2 with $j = n - (k' - l)$ and $S = \langle g_{l+1}, \dots, g_{k'}\rangle^\perp$, we also obtain $\lambda_{n-k'+l+1} = -\Omega(1/\overline{w})$. Hence, M has (at least) k' eigenvalues whose absolute value is $\Omega(1/\overline{w})$.

We are left to show that the remaining eigenvalues are substantially smaller. To this end consider an arbitrary unit vector x that is perpendicular to $g_{l+1}, \dots, g_{k'}$. We decompose x into $x = a \cdot g + b \cdot h + c \cdot v$, with $g \in \langle g_1, \dots g_l\rangle$, $h \in \langle g_{k'+1}, \dots, g_k\rangle$, and $v \perp g_1, \dots, g_k$ satisfying $\|g\| = \|h\| = \|v\| = 1$, and $a^2 + b^2 + c^2 = 1$. As $g^t \cdot M \cdot g > 0$, we have

$$\begin{aligned} x^t \cdot M \cdot x \;&=\; a^2 \cdot g^t M g + b^2 \cdot h^t M h + c^2 \cdot v^t M v + 2ab \cdot g^t M h \\ &\qquad + 2ac \cdot g^t M v + 2bc \cdot h^t M v \\ &\geq\; b^2 \cdot h^t M h + c^2 \cdot v^t M v + 2ab \cdot g^t M h + 2ac \cdot g^t M v + 2bc \cdot h^t M v \\ &\overset{(12)}{\geq}\; -O\left({d_{\mathrm{m}}}^{-0.49}/\overline{w}\right) + c^2 \cdot v^t M v + 2ac \cdot g^t M v + 2bc \cdot h^t M v. \end{aligned}$$

As $v \perp g_1 \dots, g_k$, we have $v \perp \chi_1, \dots, \chi_k$. By (9) for any unit vector u (and thus in particular for g, h, and v),

$$\left|u^t \cdot M \cdot v\right| \leq \sum_{i,j=1}^{k} \left|u_{V_i}^t \cdot M_{V_i \times V_j} \cdot v_{V_j}\right| = O(1/\overline{w}^{1.49}) = O({d_{\mathrm{m}}}^{-0.49}/\overline{w}).$$

Consequently,

$$x^t \cdot M \cdot x \geq -O\left({d_{\mathrm{m}}}^{-0.49}/\overline{w}\right).$$

Since x was arbitrary from some $(n+l-k')$-dimensional subspace, we conclude (using the second inequality from Fact 2 with $j = k'-l$) that $\lambda_{n+l-k'} \geq -O\left({d_{\mathrm{m}}}^{-0.49}/\overline{w}\right)$. Analogously, using the first inequality from Fact 2, we get $\lambda_{l+1} = O\left({d_{\mathrm{m}}}^{-0.49}/\overline{w}\right)$. Combining these two bounds, we conclude that M has $n - k'$ eigenvalues with absolute value $O\left({d_{\mathrm{m}}}^{-0.49}/\overline{w}\right)$.

To complete the proof, we show that all eigenvalues are $\Theta(1/\overline{w})$ in absolute value. Let x be some arbitrary unit vector. We decompose x into $x = a \cdot g + b \cdot h + c \cdot v$, where $g \in \langle g_1, \dots, g_l\rangle$, $h \in \langle g_{l+1}, \dots, g_k\rangle$, $v \in \langle g_1, \dots, g_k\rangle^\perp$, $\|g\| = \|h\| = \|v\| = 1$, and $a^2 + b^2 + c^2 = 1$. Then similarly as above we get

$$x^t \cdot M \cdot x = a^2 \cdot \Theta(1/\overline{w}) - b^2 \cdot \Theta(1/\overline{w}) + O({d_{\mathrm{m}}}^{-0.49}/\overline{w}).$$

Hence, $|x^t \cdot M \cdot x| = O(1/\overline{w})$, and thus Fact 2 gives the desired bounds on λ_1 and λ_n.

3.5. Proof of Lemma 10. We start with the first assertion. Since s_i is an eigenvector with eigenvalue $\Omega(1/\overline{w})$, we have

$$\left|u_i^t \cdot (M \cdot s_i)\right| = \Theta(1/\overline{w}) \cdot \left|u_i^t \cdot s_i\right| = \Theta(1/\overline{w}) \cdot \left|\gamma_i \cdot u_i^t \cdot u_i\right| = \Theta(n/\overline{w}) \cdot |\gamma_i|.$$

On the other hand, since $u_i \perp \chi_1, \dots, \chi_k$, (9) shows

$$\left|u_i^t \cdot M \cdot s_i\right| = n \cdot \left|\frac{u_i^t}{\|u_i\|} \cdot M \cdot \frac{s_i}{\|s_i\|}\right| = n \cdot O\left(\frac{1}{\overline{w}^{1.49}}\right)$$

Combining both bounds, we conclude that $|\gamma_i| = O(\overline{w}^{\,-0.49}) = O({d_{\mathrm{m}}}^{-0.49})$, thereby proving the first statement.

With respect to the second assertion, let $v \in V_j$ be such that $|\alpha_{ij} - s_i(v)| \geq 1/\ln^2 d_{\mathrm{m}}$. Then clearly $|\gamma_i \cdot u_i(v)| \geq 1/\ln^2 d_{\mathrm{m}}$. As $|\gamma_i| = O({d_{\mathrm{m}}}^{-0.49})$, we have $u_i(v) = \Omega({d_{\mathrm{m}}}^{0.49}/\ln^2 d_{\mathrm{m}})$. Since $u_i^t \cdot u_i = n$, there are at most $n \cdot O(\ln^4 d_{\mathrm{m}}/{d_{\mathrm{m}}}^{0.98})$ such entries in u_i. Hence, we have established the second claim.

To prove the third assertion, assume for contradiction that there are $j, j' \in \{1, \dots, k\}$, with $j \neq j'$ such that for each $i \in \{1, \dots, k'\}$ the inequality $|\alpha_{ij} - \alpha_{ij'}| \leq 1/\sqrt{\ln d_{\mathrm{m}}}$ holds. Consider the vector $v = \chi_j/|V_j| - \chi_{j'}/|V_{j'}|$. This vector is almost perpendicular to *each* s_i, because

$$\left|v^t \cdot s_i\right| = |\alpha_{ij} - \alpha_{ij'}| \leq \frac{1}{\sqrt{\ln d_{\mathrm{m}}}}.$$

Now let

$$v' = v - \sum_{i=1}^{k'} \frac{v^t \cdot s_i}{n} \cdot s_i.$$

Then v' *is* perpendicular to each s_i (and almost parallel to v). Both v and v' have norm $\Theta(1/\sqrt{n})$. Since v' is perpendicular to $s_1, \ldots, s_{k'}$, it lies in the space spanned by eigenvectors with eigenvalues $O({d_{\mathrm{m}}}^{-0.49}/\overline{w})$. Therefore,

$$\|M \cdot v'\| = O\left(\frac{{d_{\mathrm{m}}}^{-0.49}}{\overline{w}}\right) \cdot \|v'\| = O\left(\frac{{d_{\mathrm{m}}}^{-0.49}}{\overline{w} \cdot \sqrt{n}}\right). \tag{13}$$

Furthermore, by the definition of v' we have

$$\begin{aligned}
\|M \cdot v'\| &\geq \|M \cdot v\| - \sum_{i=1}^{k'} \|M \cdot \frac{v^t \cdot s_i}{n} \cdot s_i\| \\
&\geq \|M \cdot v\| - \sum_{i=1}^{k'} \|M\| \cdot \left|\frac{v^t \cdot s_i}{n}\right| \cdot \|s_i\| \\
&\geq \|M \cdot v\| - O\left(\frac{1}{\sqrt{n} \cdot \sqrt{\ln d_{\mathrm{m}}} \cdot \overline{w}}\right). \qquad (14)
\end{aligned}$$

In what follows, we will prove that $\|M \cdot v\| = \Omega(1/(\overline{w} \cdot \sqrt{n}))$. However, due to (14), this contradicts (13). Therefore, we have established that for any j, j' there is i such that the $|\alpha_{ij} - \alpha_{ij'}| > 1/\ln d_{\mathrm{m}}$.

The remaining task is to prove that $\|M \cdot v\| = \Omega(1/(\overline{w} \cdot \sqrt{n}))$. To this end, we set $\eta = M \cdot v$. For each $l \in \{1, \ldots, k\}$ we have

$$\sum_{u \in V_l} \eta(u) = \sum_{u \in V_l} \left(\sum_{u_2 \in V_j} \frac{m_{uu_2}}{|V_j|} - \sum_{u_2 \in V_{j'}} \frac{m_{uu_2}}{|V_{j'}|} \right) = \frac{\vec{1}^t \cdot M_{V_l \times V_j} \cdot \vec{1}}{|V_j|} - \frac{\vec{1}^t \cdot M_{V_l \times V_{j'}} \cdot \vec{1}}{|V_{j'}|}.$$

By (8) this equals

$$\begin{aligned}
\sum_{u \in V_l} \eta(u) &= \left(\frac{\phi_{lj} \cdot W_l \cdot W_j \cdot |V_l|}{\overline{w} \cdot n} - \frac{\phi_{lj'} \cdot W_l \cdot W_{j'} \cdot |V_l|}{\overline{w} \cdot n} \right) \cdot \left(1 + O\left({d_{\mathrm{m}}}^{-0.49}\right)\right) \\
&= \frac{W_l \cdot |V_l|}{\overline{w} \cdot n} \cdot (\phi_{lj} \cdot W_j - \phi_{lj'} \cdot W_{j'}) \cdot \left(1 + O\left({d_{\mathrm{m}}}^{-0.49}\right)\right).
\end{aligned}$$

By Jensen's inequality [22] for convex functions the term $\sum_{u \in V_l} \eta(u)^2$ is minimized iff each $\eta(u)$ equals the arithmetic mean of all $\eta(u)$, $u \in V_l$, which is

$$\frac{\sum_{u \in V_l} \eta(u)}{|V_l|} = \frac{W_l}{\overline{w} \cdot n} \cdot (\phi_{lj} \cdot W_j - \phi_{lj'} \cdot W_{j'}) \cdot \left(1 + O\left({d_{\mathrm{m}}}^{-0.49}\right)\right).$$

Therefore,

$$\begin{aligned}
\sum_{u \in V_l} \eta(u)^2 &\geq |V_l| \cdot \left(\frac{W_l}{\overline{w} \cdot n} \cdot (\phi_{lj} \cdot W_j - \phi_{lj'} \cdot W_{j'}) \cdot \left(1 + O\left({d_{\mathrm{m}}}^{-0.49}\right)\right) \right)^2 \\
&\geq |V_l| \cdot \left(\frac{W_l \cdot (\phi_{lj} \cdot W_j - \phi_{lj'} \cdot W_{j'})}{2 \cdot \overline{w} \cdot n} \right)^2.
\end{aligned}$$

Assumption C1 implies that there is $l \in \{1, \ldots, k\}$ such that $\phi_{lj} \cdot W_j - \phi_{lj'} \cdot W_{j'} \neq 0$. For if $\phi_{lj} \cdot W_j - \phi_{lj'} \cdot W_{j'} = 0$ for all $l \in \{1, \ldots, k\}$, then otherwise columns j and

j' (and by symmetry also rows j and j') of the matrix Φ were linearly dependent. Hence, for this index l we get

$$\sum_{u\in V_l}\eta(u)^2 \geq \left(\frac{W_l\cdot\Theta(1)}{\overline{w}\cdot n}\right)^2\cdot|V_l| = \Theta\left(\frac{1}{\overline{w}^2\cdot n}\right)$$

and

$$\|M\cdot v\| = \|\eta\| = \sqrt{\sum_{u\in V}\eta(u)^2} \geq \sqrt{\sum_{u\in V_l}\eta(u)^2} = \Omega\left(\frac{1}{\overline{w}\cdot\sqrt{n}}\right).$$

4. Proof of Theorem 8. In this section we consider $\varepsilon, \delta > 0$ and the $k\times k$ matrix Φ fixed. We let $D = D(\varepsilon,\delta,\Phi)$ and $n_0 = n_0(\varepsilon,\delta,\Phi)$ be large enough numbers and assume that $n > n_0$ and that the weight distribution w and the partition $\mathcal{V} = (V_1,\ldots,V_k)$ are such that C0–C5 hold.

4.1. Outline of the proof. We will prove that for each pair $(i,j)\in\{1,\ldots,k\}^2$ the submatrix $M_{V_i\times V_j}$ has the two properties stated in Theorem 8 w.h.p. (note that we also need to consider the case $i = j$). Since k is independent of n, the union bound then implies that the two statements hold for all k^2 block simultaneously w.h.p. For the rest of this section we consider i, j fixed.

The matrix $M_{V_i\times V_j}$ seems very difficult to analyze, because its entries are dependent random variables. The dependence of the entries results from the normalization: remember that the entry corresponding to an edge $\{u,v\}$ is $(d_u d_v)^{-1}$, and the degrees of the vertices are mutually dependent random variables (and they are also not independent of the presence of the edge $\{u,v\}$). To deal with this problem, we will relate $M_{V_i\times V_j}$ to a matrix $\mathbf{M}$ whose entries are mutually independent (apart from the fact that $\mathbf{M}$ is symmetric if $i = j$). More precisely, $\mathbf{M} = (\mathbf{m}_{uv})_{u\in V_i, v\in V_j}$ is the $|V_i|\times|V_j|$ matrix with entries

$$\mathbf{m}_{uv} = \begin{cases} 1/(w'_u\cdot w'_v) & \text{if } \{u,v\}\in E,\\ 0 & \text{otherwise,}\end{cases}$$

where w'_u, w'_v are the *expected* degrees of u, v. We analyzed the spectrum of $\mathbf{M}$ in a previous paper [12], and we will build upon the result of this analysis (see Lemma 17 below).

Of course, relating $M_{V_i\times V_j}$ to $\mathbf{M}$ is not immediate, because in $\mathbf{M}$ the entries are normalized by the *expected* degrees, whereas in $M_{V_i\times V_j}$ the normalization is by the *actual* degrees. However, we will show that $M_{V_i\times V_j}$ and $\mathbf{M}$ are sufficiently similar on a certain rectangle $R_{ij}\times C_{ij}\subseteq V_i\times V_j$, where basically $R_{ij}\subseteq V_i$ and $C_{ij}\subseteq V_j$ are the vertices whose actual degrees are sufficiently close to their expectations. The vertices in R_{ij}, C_{ij} are called *good* and those in $V_i\setminus R_{ij}$, $V_j\setminus C_{ij}$ *bad*. We will see that the vast majority of vertices are good and that the matrix $M_{R_{ij}\times C_{ij}}$ can be approximated sufficiently well by $\mathbf{M}_{R_{ij}\times C_{ij}}$. Furthermore, since the three remaining bits $M_{(V_i\setminus R_{ij})\times C_{ij}}$, $M_{R_{ij}\times(V_j\setminus C_{ij})}$, $M_{(V_i\setminus R_{ij})\times(V_j\setminus C_{ij})}$ of $M_{V_i\times V_j}$ are fairly small, we can bound their impact via elementary estimates (mostly based on the Cauchy–Schwarz inequality).

More precisely, $R_{ij}\subseteq V_i$ and $C_{ij}\subseteq V_j$ are the result of the following process. We let Δ denote a sufficiently large constant.

1. Let $F = \{u\in V : \text{for all } l : |s_A(u,V_l) - \text{Vol}(u,V_l)| \leq \text{Vol}(u,V_l)^{0.51}\}$.
2. Set $R'_{ij} := V_i\cap F$ and $C'_{ij} := V_j\cap F$.

3. While there is a vertex $u \in R'_{ij}$ with
$$s_A(u, V_j \setminus C'_{ij}) \geq \mathrm{Vol}(u, V_j) \cdot \Delta/d_{\mathrm{m}}, \qquad \text{let} \quad R'_{ij} := R'_{ij} \setminus \{u\}.$$
4. While there is a vertex $v \in C'_{ij}$ with
$$s_A(V_i \setminus R'_{ij}, v) \geq \mathrm{Vol}(V_i, v) \cdot \Delta/d_{\mathrm{m}}, \qquad \text{let} \quad C'_{ij} := C'_{ij} \setminus \{v\}.$$
5. Repeat step 3 and step 4 until R'_{ij} and C'_{ij} remain unchanged.
6. Let $R_{ij} := R'_{ij}$, $C_{ij} := C'_{ij}$ be the final outcome of the process.

Intuitively, the process does the following. The set F consists of all vertices u such that the actual number $s_A(u, V_l)$ of neighbors of u in *each* class V_l of the planted partition deviates from the expected number $\mathrm{Vol}(u, V_l)$ by at most $\mathrm{Vol}(u, V_l)^{0.51}$. Since $\mathrm{Vol}(u, V_l)^{0.51}$ is slightly bigger than the standard deviation $\mathrm{Vol}(u, V_l)^{0.5}$, we will be able to show that F contains a very large fraction of the vertices. For all vertices in F the degree $d_u = \sum_l s_A(u, V_l)$ will be sufficiently close to its expectation for our purposes. However, it could be that some vertices in F have plenty of neighbors in $V \setminus F$. These vertices are difficult to deal with, and therefore we would like to declare them bad as well. This is the purpose of steps 2–5. In step 2 we initialize R'_{ij} and C'_{ij}; think of this as declaring the vertices in $V_i \setminus F$ and $V_j \setminus F$ bad. Then, in step 3 we keep removing vertices u from R'_{ij} that have many neighbors in $V_j \setminus C'_{ij}$, i.e., many bad neighbors. For C'_{ij} we proceed similarly in step 4, and we keep repeating this process until it stabilizes. Finally, all the remaining vertices are good.

To ease the notation, we abbreviate R_{ij} by $\mathcal{R}$, C_{ij} by $\mathcal{C}$, $V_i \setminus R_{ij}$ by $\overline{\mathcal{R}}$, and $V_j \setminus C_{ij}$ by $\overline{\mathcal{C}}$. The first step of the process ensures that

$$\begin{aligned} & |s_A(u, V) - \mathrm{Vol}(u, V)| \leq 2 \cdot \mathrm{Vol}(u, V)^{0.51} \quad \text{for all } u \in \mathcal{R} \\ \text{and} \quad & |s_A(V, v) - \mathrm{Vol}(V, v)| \leq 2 \cdot \mathrm{Vol}(V, v)^{0.51} \quad \text{for all } v \in \mathcal{C}. \end{aligned} \tag{15}$$

In section 4.4 we will prove the following lemma, which shows that the volumes of $\overline{\mathcal{R}}$ and $\overline{\mathcal{C}}$ are small.

LEMMA 11. *W.h.p. the random graph $G = G_n(\Phi, w, \mathcal{V})$ has the following properties.*
1. $\mathrm{Vol}(\overline{\mathcal{R}}, V_j) \leq n/{d_{\mathrm{m}}}^4$.
2. $\mathrm{Vol}(V_i, \overline{\mathcal{C}}) \leq n/{d_{\mathrm{m}}}^4$.
3. $\mathrm{Vol}(\overline{\mathcal{R}}, \overline{\mathcal{C}}) \leq n/{d_{\mathrm{m}}}^8$.

A consequence of Lemma 11 is that both $\overline{\mathcal{R}}$ and $\overline{\mathcal{C}}$ contain only few vertices. For by the choice of d_{m} (see Fact 6), we have w.h.p. for all $u \in V_i$ and all $v \in V_j$:

$$d_{\mathrm{m}} \leq \mathrm{Vol}(u, V_j) \leq \mathrm{Vol}(u, V) = w'_u \qquad \text{and} \qquad d_{\mathrm{m}} \leq \mathrm{Vol}(V_i, v) \leq w'_v. \tag{16}$$

Summing over all $u \in \overline{\mathcal{R}}$ we get $d_{\mathrm{m}} \cdot |\overline{\mathcal{R}}| \leq \mathrm{Vol}(\overline{\mathcal{R}}, V_j) \leq n/{d_{\mathrm{m}}}^4$; whence $|\overline{\mathcal{R}}| \leq n/{d_{\mathrm{m}}}^5$. As $\delta \cdot n \leq |V_i|$, we get

$$|\overline{\mathcal{R}}| \leq \frac{|V_i|}{\delta \cdot {d_{\mathrm{m}}}^5} \leq \frac{|V_i|}{{d_{\mathrm{m}}}^4} \qquad \text{and} \qquad |\mathcal{R}| = |V_i| - |\overline{\mathcal{R}}| \geq |V_i| \cdot \left(1 - \frac{1}{{d_{\mathrm{m}}}^4}\right), \tag{17}$$

because $d_{\mathrm{m}} > 1/\delta$ (see Fact 6 again). In the same way we get

$$|\overline{\mathcal{C}}| \leq |V_j| / {d_{\mathrm{m}}}^4 \qquad \text{and} \qquad |\mathcal{C}| \geq |V_j| \cdot \left(1 - 1/{d_{\mathrm{m}}}^4\right). \tag{18}$$

Now, we subdivide $M_{V_i \times V_j}$ into four parts $M_{\overline{\mathcal{R}} \times \mathcal{C}}$, $M_{\mathcal{R} \times \overline{\mathcal{C}}}$, $M_{\mathcal{R} \times \mathcal{C}}$, $M_{\overline{\mathcal{R}} \times \overline{\mathcal{C}}}$, which we will analyze separately.

LEMMA 12. *W.h.p. the random graph $G = G_n(\Phi, w, \mathcal{V})$ has the following properties.*

1. $\vec{1}^t \cdot M_{\mathcal{R}\times\mathcal{C}} \cdot \vec{1} = \phi_{ij} \cdot W_i \cdot W_j \cdot ((|\mathcal{R}| \cdot |\mathcal{C}|)/\overline{w} \cdot n) \cdot (1+O(1/{d_{\mathrm{m}}}^{0.49})) = \Theta(\phi_{ij} \cdot n/\overline{w})$.
2. *For any u, v with $\|u\| = \|v\| = 1$ and $u \perp \vec{1}$ or $v \perp \vec{1}$ we have*

$$\left|u^t \cdot M_{\mathcal{R}\times\mathcal{C}} \cdot v\right| = O\left(1/\overline{w}^{1.49}\right).$$

3. $\|M_{\mathcal{R}\times\mathcal{C}}\| = \Theta\left(1/\overline{w}\right)$.

Lemma 12 deals with the "large" block $M_{\mathcal{R}\times\mathcal{C}}$ of M. Its proof is based on relating $M_{\mathcal{R}\times\mathcal{C}}$ to the "easy" matrix $\mathbf{M}_{\mathcal{R}\times\mathcal{C}}$; we defer the details to section 4.2.

The following Lemma 13 takes care of the remaining three blocks. The proof can be found in section 4.3.1.

LEMMA 13. *W.h.p. the random graph $G = G_n(\Phi, w, \mathcal{V})$ has the following properties.*

1. $\|M_{\mathcal{R}\times\overline{\mathcal{C}}}\| = O({d_{\mathrm{m}}}^{-1.5})$.
2. $\|M_{\overline{\mathcal{R}}\times\mathcal{C}}\| = O({d_{\mathrm{m}}}^{-1.5})$.
3. $\|M_{\overline{\mathcal{R}}\times\overline{\mathcal{C}}}\| = O({d_{\mathrm{m}}}^{-1.5})$.

Proof of Theorem 8. With respect to the first item, we have

$$\vec{1}^t \cdot M_{V_i\times V_j} \cdot \vec{1} = \vec{1}^t \cdot M_{\mathcal{R}\times\mathcal{C}} \cdot \vec{1} + \vec{1}^t \cdot M_{\mathcal{R}\times\overline{\mathcal{C}}} \cdot \vec{1} + \vec{1}^t \cdot M_{\overline{\mathcal{R}}\times\mathcal{C}} \cdot \vec{1} + \vec{1}^t \cdot M_{\overline{\mathcal{R}}\times\overline{\mathcal{C}}} \cdot \vec{1}. \tag{19}$$

Item 1 of Lemma 12 gives for the first term

$$\begin{aligned}\vec{1}^t \cdot M_{\mathcal{R}\times\mathcal{C}} \cdot \vec{1} \quad &= \quad \phi_{ij} \cdot W_i \cdot W_j \cdot \frac{|\mathcal{R}| \cdot |\mathcal{C}|}{\overline{w} \cdot n} \cdot \left(1 + O\left(\frac{1}{{d_{\mathrm{m}}}^{0.49}}\right)\right) \\ &\overset{(17),(18)}{=} \phi_{ij} \cdot W_i \cdot W_j \cdot \frac{|V_i| \cdot |V_j|}{\overline{w} \cdot n} \cdot \left(1 + O\left(\frac{1}{{d_{\mathrm{m}}}^{0.49}}\right)\right).\end{aligned}$$

Lemma 13 shows that the second summand in (19) is bounded by

$$\begin{aligned}\left|\vec{1}^t \cdot M_{\mathcal{R}\times\overline{\mathcal{C}}} \cdot \vec{1}\right| &\le \sqrt{|\mathcal{R}| \cdot |\overline{\mathcal{C}}|} \cdot \|M_{\mathcal{R}\times\overline{\mathcal{C}}}\| \overset{(18)}{\le} \sqrt{|V_i| \cdot |V_j| / {d_{\mathrm{m}}}^4} \cdot O({d_{\mathrm{m}}}^{-1.5}) \\ &= \sqrt{|V_i| \cdot |V_j|} \cdot O\left({d_{\mathrm{m}}}^{-2}/\overline{w}\right),\end{aligned}$$

as $d_{\mathrm{m}} > \overline{w}'^{\,2/3}$ by Fact 6. The same bound holds for $\left|\vec{1}^t \cdot M_{\overline{\mathcal{R}}\times\mathcal{C}} \cdot \vec{1}\right|$ as well as $\left|\vec{1}^t \cdot M_{\overline{\mathcal{R}}\times\overline{\mathcal{C}}} \cdot \vec{1}\right|$. Hence, we get

$$\begin{aligned}\frac{\vec{1}^t}{\|\vec{1}^t\|} \cdot M_{V_i\times V_j} \cdot \frac{\vec{1}}{\|\vec{1}\|} &= \phi_{ij} \cdot W_i \cdot W_j \cdot \frac{\sqrt{|V_i| \cdot |V_j|}}{\overline{w} \cdot n} \cdot \left(1 + O\left(\frac{1}{{d_{\mathrm{m}}}^{0.49}}\right)\right) + O\left(\frac{{d_{\mathrm{m}}}^{-2}}{\overline{w}}\right) \\ &= \phi_{ij} \cdot W_i \cdot W_j \cdot \frac{\sqrt{|V_i| \cdot |V_j|}}{\overline{w} \cdot n} \cdot \left(1 + O\left(\frac{1}{{d_{\mathrm{m}}}^{0.49}}\right)\right).\end{aligned}$$

To prove the second item of Theorem 8, we assume that u (from the theorem) is perpendicular to $\vec{1}$, which yields $u^t \cdot (\vec{1}_{|\mathcal{R}} + \vec{1}_{|\overline{\mathcal{R}}}) = 0$. Hence,

$$\left|u^t \cdot \vec{1}_{|\mathcal{R}}\right| = \left|u^t \cdot \vec{1}_{|\overline{\mathcal{R}}}\right| \le \|u\| \cdot \|\vec{1}_{|\overline{\mathcal{R}}}\| \le \sqrt{|\overline{\mathcal{R}}|}. \tag{20}$$

We can decompose u as $u = a \cdot \vec{1}_{|\mathcal{R}} / \|\vec{1}_{|\mathcal{R}}\| + b \cdot u_l$, with $\|u_l\| = 1$ and $u_l \perp \vec{1}_{|\mathcal{R}}$. Clearly $u_{l|\mathcal{R}} \perp \vec{1}_{|\mathcal{R}}$, too. Note that $a^2 + b^2 = 1$. Hence, we can bound $|a|$ as follows:

$$|a| = \left|u^t \cdot \frac{\vec{1}_{|\mathcal{R}}}{\|\vec{1}_{|\mathcal{R}}\|}\right| \overset{(20)}{\le} \sqrt{\frac{|\overline{\mathcal{R}}|}{|\mathcal{R}|}} \overset{(17)}{\le} \sqrt{2/{d_{\mathrm{m}}}^4} < \frac{2}{{d_{\mathrm{m}}}^2}. \tag{21}$$

Let v be an arbitrary unit vector. Then Lemma 12 and Lemma 13 yield

$$\begin{aligned}
\left|u^t \cdot M_{V_i\times V_j} \cdot v\right| &= \left|u^t \cdot M_{V_i\times V_j} \cdot \left(v_{|\mathcal{C}} + v_{|\overline{\mathcal{C}}}\right)\right| \\
&\leq \left|u^t \cdot M_{V_i\times V_j} \cdot v_{|\mathcal{C}}\right| + \|M_{\mathcal{R}\times\overline{\mathcal{C}}}\| + \|M_{\overline{\mathcal{R}}\times\overline{\mathcal{C}}}\| \\
&= \left|\left(a \cdot \frac{\vec{1}^t{}_{|\mathcal{R}}}{\|\vec{1}^t{}_{|\mathcal{R}}\|} + b \cdot u_l^t\right) \cdot M_{V_i\times\mathcal{C}} \cdot v_{\mathcal{C}}\right| + O\left({d_{\mathrm{m}}}^{-1.5}\right) \\
&\leq |a| \cdot \|M_{\mathcal{R}\times\mathcal{C}}\| + \left|\left(b \cdot u_l^t\right) \cdot M_{V_i\times\mathcal{C}} \cdot v_{\mathcal{C}}\right| + O\left({d_{\mathrm{m}}}^{-1.5}\right) \\
&\overset{(21)}{\leq} 2/{d_{\mathrm{m}}}^2 \cdot \|M_{\mathcal{R}\times\mathcal{C}}\| + \left|b \cdot \left(u_{l|\mathcal{R}} + u_{l|\overline{\mathcal{R}}}\right)^t \cdot M_{V_i\times\mathcal{C}} \cdot v_{\mathcal{C}}\right| + O\left({d_{\mathrm{m}}}^{-1.5}\right) \\
&\overset{b\leq 1}{\leq} \left|u_{l\mathcal{R}}^t \cdot M_{\mathcal{R}\times\mathcal{C}} \cdot v_{\mathcal{C}}\right| + \|M_{\overline{\mathcal{R}}\times\mathcal{C}}\| + O\left({d_{\mathrm{m}}}^{-1.5}\right) \\
&\overset{u_{l|\mathcal{R}}\perp\vec{1}_{|\mathcal{R}}}{=} O\left(\frac{1}{\overline{w}^{1.49}}\right) + O\left({d_{\mathrm{m}}}^{-1.5}\right).
\end{aligned}$$

(Remember the difference between $v_{|\mathcal{C}}$ and $v_{\mathcal{C}}$: we have $M_{V_i\times V_j} \cdot v_{|\mathcal{C}} = M_{V_i\times\mathcal{C}} \cdot v_{\mathcal{C}}$.) The case $v \perp \vec{1}$ and u arbitrary can be handled analogously. $\square$

4.2. Proof of Lemma 12: The spectrum of $M_{\mathcal{R}\times\mathcal{C}}$. In this section we analyze the spectrum of the matrix $\mathbf{M}$ with (essentially) independent entries and relate this matrix to $M_{\mathcal{R}\times\mathcal{C}}$. In order to analyze the spectrum of $\mathbf{M}$ we build upon results from [12].

DEFINITION 14. *A random $n \times m$ matrix $X = (x_{uv})$ is a* same-mean-matrix *with mean μ and bound b if the following conditions hold.*

1. *The entries x_{uv} are independent random variables, apart from possibly the trivial dependence induced by symmetry (i.e., $x_{uv} = x_{vu}$ for all pairs (u, v)).*
2. *Each x_{uv} can attain one of exactly two possible values, one of which is* 0.
3. *With probability one we have $x_{uv} \leq b$ for all u, v.*
4. $\mathbf{E}[x_{uv}] = \mu > 0$ *for all u, v.*

The matrix $\mathbf{M}$ is a same-mean-matrix with mean

$$\mu = \mathbf{E}[\mathbf{m}_{uv}] = \frac{1}{w'_u \cdot w'_v} \cdot \phi_{ij} \cdot \frac{w_u \cdot w_v}{\overline{w} \cdot n} \overset{(2)}{=} \frac{\phi_{ij} \cdot W_i \cdot W_j}{\overline{w} \cdot n} = \Theta\left(\frac{1}{\overline{w} \cdot n}\right) \tag{22}$$

and bound

$$b = (1/w'_{\mathrm{m}})^2, \tag{23}$$

where we let $w'_{\mathrm{m}} = \min_{u\in V} w'_u$. By condition C3 on page 1684 and (2) we have $w'_{\mathrm{m}} = \Theta(\overline{w})$.

The following two lemmas are taken from [12]. Lemma 15 is a special case of Lemma 8 in [12] (we set all $a_i := 1$), whereas Lemma 16 is identical to Lemma 9 in that paper.

LEMMA 15. *Let X be a same-mean-matrix with mean μ and bound b. Let $y_1, \ldots, y_l$ be a set of mutually independent entries of X and $Y = \sum_{i=1}^{l} y_i$.*

If $S \leq c \cdot \mathrm{e}^c \cdot l \cdot \mu = c \cdot \mathrm{e}^c \cdot \mathbf{E}[Y]$ for some positive constant c, then

$$\mathbf{Pr}\left[|Y - \mathbf{E}[Y]| \geq S\right] \leq 2 \cdot \exp\left(-S^2/(2 \cdot \mathrm{e}^c \cdot \mathbf{E}[Y] \cdot b)\right).$$

LEMMA 16. *Let X be an $n \times m$ same-mean-matrix with mean μ and bound b and $N = n + m$. Let $R = \{u : \sum_v x_{uv} \leq d \cdot \mu \cdot N\}$ and $C = \{v : \sum_u x_{uv} \leq d \cdot \mu \cdot N\}$ for some arbitrary $d > 1$.*

If $\mu \cdot n \cdot m > b \cdot N$, then with probability $1 - O(1/\sqrt{N})$ we have

$$\sup\{\left|{u_{|R}}^t \cdot X \cdot v_{|C}\right| : \|u_{|R}\| = \|v_{|C}\| = 1 \wedge (u_{|R} \perp \vec{1} \vee v_{|C} \perp \vec{1})\} = O(\sqrt{b \cdot d \cdot \mu \cdot N}).$$

Combining Lemma 15 and Lemma 16, we obtain the following result for $\mathbf{M}$.

LEMMA 17. *With high probability $\mathbf{M}_{\mathcal{R}\times\mathcal{C}}$ has the following three properties.*

1. $\vec{1}^t \cdot \mathbf{M}_{\mathcal{R}\times\mathcal{C}} \cdot \vec{1} = \mu \cdot |\mathcal{R}| \cdot |\mathcal{C}| \cdot (1 + O(1/d_{\mathrm{m}})) = \Theta(n/\overline{w})$.
2. $|u^t \cdot \mathbf{M}_{\mathcal{R}\times\mathcal{C}} \cdot v| = O\left(1/{w'_{\mathrm{m}}}^{1.5}\right)$ *for* $\|u\| = \|v\| = 1$ *and* $u \perp \vec{1}$ *or* $v \perp \vec{1}$.
3. $\|\mathbf{M}_{\mathcal{R}\times\mathcal{C}}\| = \Theta(\mu \cdot \sqrt{|\mathcal{R}| \cdot |\mathcal{C}|}) = \Theta(1/\overline{w})$.

Proof. We start with the first item. Clearly, $\vec{1}^t \cdot \mathbf{M}_{\mathcal{R}\times\mathcal{C}} \cdot \vec{1} = s_{\mathbf{M}}(\mathcal{R}, \mathcal{C})$ is bounded above by $s_{\mathbf{M}}(V_i, V_j)$. Using Lemma 15 with $Y = s_{\mathbf{M}}(V_i, V_j)$, $c = 1$, and $S = \mathbf{E}[Y]/d_{\mathrm{m}}$ we get

$$\begin{aligned}\mathbf{Pr}\left[|Y - \mathbf{E}[Y]| \geq \mathbf{E}[Y]/d_{\mathrm{m}}\right] &\leq 2 \cdot \exp\left(-\mathbf{E}[Y]/(2 \cdot \mathrm{e} \cdot b \cdot {d_{\mathrm{m}}}^2)\right) \\ &\leq 2 \cdot \exp(-\mathbf{E}[Y]/6),\end{aligned}$$

as $b \cdot {d_{\mathrm{m}}}^2 = {d_{\mathrm{m}}}^2/(\min_{u\in V} w'_u)^2 \leq 1$ by (16). As $\mathbf{E}[Y] = |V_i| \cdot |V_j| \cdot \mu = \Theta(n/\overline{w}) = \omega(1)$, we have that w.h.p.

$$s_{\mathbf{M}}(V_i, V_j) = \mu \cdot |V_i| \cdot |V_j| \cdot (1 + O(1/d_{\mathrm{m}})).$$

This gives (together with (17) and (18)) that w.h.p. $s_{\mathbf{M}}(\mathcal{R}, \mathcal{C}) \leq \mu \cdot |\mathcal{R}| \cdot |\mathcal{C}| \cdot (1 + O(1/d_{\mathrm{m}}))$. To obtain a lower bound on $s_{\mathbf{M}}(\mathcal{R}, \mathcal{C})$, we use

$$s_{\mathbf{M}}(\mathcal{R}, \mathcal{C}) \geq s_{\mathbf{M}}(V_i, V_j) - s_{\mathbf{M}}(\overline{\mathcal{R}}, V_j) - s_{\mathbf{M}}(V_i, \overline{\mathcal{C}}).$$

By the construction of $\mathbf{M}$, we have $s_{\mathbf{M}}(\overline{\mathcal{R}}, V_j) \leq b \cdot s_A(\overline{\mathcal{R}}, V_j)$. By Lemma 11, $\mathrm{Vol}(\overline{\mathcal{R}}, V_j) \leq n/{d_{\mathrm{m}}}^4$. In the following calculation we show that w.h.p. all sets $T \subseteq V_i$ (including $\overline{\mathcal{R}}$) with $\mathrm{Vol}(T, V_j) \leq n/{d_{\mathrm{m}}}^4$ satisfy $s_A(T, V_j) < 2n/\overline{w}$. As $|T| \leq \mathrm{Vol}(T, V_j)/d_{\mathrm{m}} \leq n/{d_{\mathrm{m}}}^5$, the number of such sets T is bounded above by

$$\begin{aligned}\binom{n}{n/{d_{\mathrm{m}}}^5} &\leq \left(\mathrm{e} \cdot {d_{\mathrm{m}}}^5\right)^{n/{d_{\mathrm{m}}}^5} \leq \exp\left(\ln(\mathrm{e} \cdot {d_{\mathrm{m}}}^5) \cdot n/{d_{\mathrm{m}}}^5\right) \\ &< \exp\left(n/{d_{\mathrm{m}}}^4\right) \overset{d_{\mathrm{m}} > \overline{w}^{2/3}}{<} \exp\left(n/\overline{w}^2\right).\end{aligned}$$

Fix such a set T. We shall derive the concentration result from Fact 3. To this end, we set $X := s_A(T, V_j)$ and $t := n/\overline{w}$. As $\mathbf{E}[X] = \mathrm{Vol}(T, V_j) \leq n/{d_{\mathrm{m}}}^4 < n/\overline{w}^{8/3}$, we have $\mathbf{E}[X] + t/3 < 2t$. Thus,

$$\mathbf{Pr}\left[s_A(T, V_j) \geq \mathrm{Vol}(T, V_j) + n/\overline{w}\right] \leq \exp(-n/(4\overline{w})).$$

Applying the union bound, we conclude that with probability $> 1 - \exp(-n/(5\overline{w})) = 1 - o(1)$ *all* sets $T \subseteq V_i$ with $\mathrm{Vol}(T, V_j) < n/{d_{\mathrm{m}}}^4$ satisfy

$$s_A(T, V_j) < \mathrm{Vol}(T, V_j) + n/\overline{w} < 2n/\overline{w}.$$

With high probability the same bound holds for $s_A(V_i, T)$, with $T \subseteq V_j$. Hence, w.h.p.

$$\begin{aligned}s_{\mathbf{M}}(\mathcal{R}, \mathcal{C}) &\geq s_{\mathbf{M}}(V_i, V_j) - s_{\mathbf{M}}(\overline{\mathcal{R}}, V_j) - s_{\mathbf{M}}(V_i, \overline{\mathcal{C}}) \geq s_{\mathbf{M}}(V_i, V_j) - b \cdot 4n/\overline{w} \\ &\geq \mu \cdot |V_i| \cdot |V_j| \cdot (1 - O(1/d_{\mathrm{m}})) - O(\mu \cdot n^2/{d_{\mathrm{m}}}^2) \\ &\geq \mu \cdot |V_i| \cdot |V_j| \cdot (1 - O(1/d_{\mathrm{m}})) \geq \mu \cdot |\mathcal{R}| \cdot |\mathcal{C}| \cdot (1 - O(1/d_{\mathrm{m}})).\end{aligned}$$

To prove the second item of Lemma 17, we use Lemma 16. Let $p \in \mathcal{R}$. Then

$$\begin{aligned} s_{\mathbf{M}}(p, V_j) &= \sum_{q \in V_j} \mathbf{m}_{pq} = \sum_{q \in N(p) \cap V_j} \frac{1}{w'_p \cdot w'_q} \leq \frac{|N(p) \cap V_j|}{w'_{\mathrm{m}} \cdot w'_p} = \frac{s_A(p, V_j)}{w'_{\mathrm{m}} \cdot w'_p} \\ &\overset{(15)}{\leq} \frac{2 \cdot \mathrm{Vol}(p, V)}{w'_{\mathrm{m}} \cdot w'_p} = \frac{2}{w'_{\mathrm{m}}} = O\left(\frac{\overline{w}}{w'_{\mathrm{m}}} \cdot \frac{1}{\overline{w}}\right) = O\left(\frac{\overline{w}}{w'_{\mathrm{m}}} \cdot n \cdot \mu\right) \\ &\leq K \cdot \mu \cdot (|V_i| + |V_j|) \end{aligned} \tag{24}$$

for $K = O(\overline{w}/w'_{\mathrm{m}})$. The same bound holds for the column-sum of $v \in \mathcal{C}$. We have $\mu \cdot |\mathcal{R}| \cdot |\mathcal{C}| = \Theta(n/\overline{w})$. As $\mathbf{M}$ has bound $b = 1/w'^2_{\mathrm{m}} = \Theta(1/\overline{w}^2)$, the term $b \cdot (|\mathcal{R}| + |\mathcal{C}|)$ is bounded from above by $O(n/\overline{w}^2)$. Since $\overline{w}$ is large enough, the inequality $\mu \cdot |\mathcal{R}| \cdot |\mathcal{C}| \geq b \cdot (|\mathcal{R}| + |\mathcal{C}|)$ is true. Thus, the assumptions of Lemma 16 are satisfied.

Let u be an arbitrary $|\mathcal{R}|$-dimensional vector. We extend u to an $|V_i|$-dimensional vector u' by setting the additional coordinates of u' to 0. We can do the same for any $|\mathcal{C}|$-dimensional vector v to obtain an $|V_j|$-dimensional vector v'. Clearly, $u^t \cdot \mathbf{M}_{\mathcal{R} \times \mathcal{C}} \cdot v = u'^t \cdot \mathbf{M} \cdot v'$. We apply Lemma 16 with $d = K$ to $\mathbf{M}$. Any nonzero coordinate in u' belongs to $\mathcal{R}$ and by (24) also to R. Similarly, any nonzero coordinate in v' belongs to C. If $\|u\| = \|v\| = 1$ as well as $u \perp \vec{1}$ or $v \perp \vec{1}$, we obtain from Lemma 16

$$\begin{aligned} \left|u^t \cdot \mathbf{M}_{\mathcal{R} \times \mathcal{C}} \cdot v\right| &= \left|u'^t \cdot \mathbf{M} \cdot v'\right| = \left|(u'_{|R})^t \cdot \mathbf{M} \cdot v'_{|C}\right| = O\left(\sqrt{b \cdot K \cdot \mu \cdot (|V_i| + |V_j|)}\right) \\ &= O\left(\sqrt{1/w'^2_{\mathrm{m}} \cdot \overline{w}/w'_{\mathrm{m}} \cdot 1/(\overline{w} \cdot n) \cdot 2 \cdot \delta \cdot n}\right) = O(\sqrt{1/w'^3_{\mathrm{m}}}). \end{aligned}$$

The third item of Lemma 17 is a direct consequence of the previous two. To see this, let x be a unit vector maximizing $\|\mathbf{M}_{\mathcal{R} \times \mathcal{C}} \cdot x\| = \|\mathbf{M}_{\mathcal{R} \times \mathcal{C}}\|$. For $y = \mathbf{M}_{\mathcal{R} \times \mathcal{C}} \cdot x$ we get

$$\|y\|^2 = y^t \cdot y = y^t \cdot \mathbf{M}_{\mathcal{R} \times \mathcal{C}} \cdot x = \|y\| \cdot \frac{y^t}{\|y\|} \cdot \mathbf{M}_{\mathcal{R} \times \mathcal{C}} \cdot x;$$

whence $\|y\| = \frac{y^t}{\|y\|} \cdot \mathbf{M}_{\mathcal{R} \times \mathcal{C}} \cdot x$.

There are unique decompositions $x = a \cdot \frac{\vec{1}}{\|\vec{1}\|} + b \cdot u$ and $\frac{y}{\|y\|} = c \cdot \frac{\vec{1}}{\|\vec{1}\|} + d \cdot u'$, with $u, u' \perp \vec{1}$ and $a^2 + b^2 = c^2 + d^2 = 1$. Thus, we get

$$\begin{aligned} \|\mathbf{M}_{\mathcal{R} \times \mathcal{C}}\| = \|y\| &= ac \cdot \frac{\vec{1}^t}{\|\vec{1}\|} \cdot \mathbf{M}_{\mathcal{R} \times \mathcal{C}} \cdot \frac{\vec{1}}{\|\vec{1}\|} + bc \cdot \frac{\vec{1}^t}{\|\vec{1}\|} \cdot \mathbf{M}_{\mathcal{R} \times \mathcal{C}} \cdot u \\ &\quad + ad \cdot u'^t \cdot \mathbf{M}_{\mathcal{R} \times \mathcal{C}} \cdot \frac{\vec{1}}{\|\vec{1}\|} + bd \cdot u'^t \cdot \mathbf{M}_{\mathcal{R} \times \mathcal{C}} \cdot u \\ &= ac \cdot \Theta\left(\frac{1}{\overline{w}}\right) + (bc + ad + ac) \cdot O\left(\frac{1}{w'_{\mathrm{m}}{}^{1.5}}\right) = O\left(\frac{1}{\overline{w}}\right). \qquad \square \end{aligned}$$

Proof of Lemma 12. Using the notation of (22), we have to show that

$$\vec{1}^t \cdot M_{\mathcal{R} \times \mathcal{C}} \cdot \vec{1} = \vec{1}^t \cdot M_{\mathcal{R} \times \mathcal{C}} \cdot \vec{1} = \mu \cdot |\mathcal{R}| \cdot |\mathcal{C}| \cdot (1 + O(1/d_{\mathrm{m}}{}^{0.49})).$$

Let D_l be the $|\mathcal{R}| \times |\mathcal{R}|$-dimensional diagonal matrix with the entries (w'_u/d_u) for $u \in \mathcal{R}$ on the diagonal and analogous D_r for the vertices in $\mathcal{C}$. Then we have that

$$M_{\mathcal{R} \times \mathcal{C}} = D_l \cdot \mathbf{M}_{\mathcal{R} \times \mathcal{C}} \cdot D_r. \tag{25}$$

In order to show the first item of Lemma 12, we have to bound $\vec{1}^t \cdot D_l \cdot \mathbf{M}_{\mathcal{R}\times\mathcal{C}} \cdot D_r \cdot \vec{1}$. Recalling the process that we used to construct $\mathcal{C}$, we see that (15) entails that $|d_u - w'_u| \le 2 \cdot w_u'^{0.51}$ for all $u \in \mathcal{C}$. Therefore, we obtain for the diagonal entries of D_r and D_l

$$\begin{aligned} & \frac{w'_u}{w'_u + 2 \cdot w_u'^{0.51}} \le \frac{w'_u}{d_u} \le \frac{w'_u}{w'_u - 2 \cdot w_u'^{0.51}} \\ \iff \quad & 1 - \frac{2 \cdot w_u'^{-0.49}}{1 + 2 \cdot w_u'^{-0.49}} \le \frac{w'_u}{d_u} \le 1 + \frac{2 \cdot w_u'^{-0.49}}{1 - 2 \cdot w_u'^{-0.49}} \\ \implies \quad & 1 - 2 \cdot w_u'^{-0.49} \le \frac{w'_u}{d_u} \le 1 + 3 \cdot w_u'^{-0.49}, \end{aligned}$$

because we are assuming that w'_u exceeds some sufficiently large constant.

As D_r (and also D_l) is a diagonal matrix, its norm equals the largest entry in absolute value. Thus, we have $\|D_r\| \le 1 + 3 \cdot w_{\mathrm{m}}'^{-0.49} \le 2$ and $\|D_l\| \le 2$, too. Let $D_r \cdot \vec{1} = a_r \cdot \vec{1} + v_r$, with $v_r \perp \vec{1}$. Then $1 - 2 \cdot w_{\mathrm{m}}'^{-0.49} \le a_r \le 1 + 3 \cdot w_{\mathrm{m}}'^{-0.49}$ and $\|v_r\| \le \sqrt{|\mathcal{C}|} \cdot 5 \cdot w_{\mathrm{m}}'^{-0.49}$. In the same way we get for $\vec{1}^t \cdot D_l = a_l \cdot \vec{1}^t + u_l^t$, with $u_l \perp \vec{1}$ that $1 - 2 \cdot w_{\mathrm{m}}'^{-0.49} \le a_l \le 1 + 3 \cdot w_{\mathrm{m}}'^{-0.49}$ and $\|u_l\| \le \sqrt{|\mathcal{R}|} \cdot 5 \cdot w_{\mathrm{m}}'^{-0.49}$. We bound

$$\begin{aligned} \vec{1}^t \cdot D_l \cdot \mathbf{M}_{\mathcal{R}\times\mathcal{C}} \cdot D_r \cdot \vec{1} = {} & a_l \cdot a_r \cdot \vec{1}^t \cdot \mathbf{M}_{\mathcal{R}\times\mathcal{C}} \cdot \vec{1} + a_l \cdot \vec{1}^t \cdot \mathbf{M}_{\mathcal{R}\times\mathcal{C}} \cdot v_r \\ & + a_r \cdot u_l^t \cdot \mathbf{M}_{\mathcal{R}\times\mathcal{C}} \cdot \vec{1} + u_l^t \cdot \mathbf{M}_{\mathcal{R}\times\mathcal{C}} \cdot v_r \end{aligned} \tag{26}$$

by considering each summand separately. Lemma 17 gives

$$\begin{aligned} a_l \cdot a_r \cdot \vec{1}^t \cdot \mathbf{M}_{\mathcal{R}\times\mathcal{C}} \cdot \vec{1} &= \left(1 + O\left(w_{\mathrm{m}}'^{-0.49}\right)\right)^2 \cdot \mu \cdot |\mathcal{R}| \cdot |\mathcal{C}| \cdot (1 + O(1/d_{\mathrm{m}})) \\ &= \mu \cdot |\mathcal{R}| \cdot |\mathcal{C}| \cdot \left(1 + O\left(w_{\mathrm{m}}'^{-0.49}\right)\right) \end{aligned} \tag{27}$$

and

$$\begin{aligned} \left|a_l \cdot \vec{1}^t \cdot \mathbf{M}_{\mathcal{R}\times\mathcal{C}} \cdot v_r\right| &= a_l \cdot \sqrt{|\mathcal{R}|} \cdot \|v_r\| \cdot \left| \frac{\vec{1}^t}{\|\vec{1}^t\|} \cdot \mathbf{M}_{\mathcal{R}\times\mathcal{C}} \cdot \frac{v_r}{\|v_r\|} \right| \\ &\overset{v_r \perp \vec{1}}{\le} 2 \cdot \sqrt{|\mathcal{R}|} \cdot \left(\sqrt{|\mathcal{C}|} \cdot 5 \cdot w_{\mathrm{m}}'^{-0.49}\right) \cdot O\left(\frac{1}{w_{\mathrm{m}}'^{1.5}}\right) \\ &= O\left(\sqrt{|\mathcal{R}| \cdot |\mathcal{C}|} \cdot w_{\mathrm{m}}'^{-1.99}\right) \\ &= O\left(\mu \cdot |\mathcal{R}| \cdot |\mathcal{C}| \cdot w_{\mathrm{m}}'^{-0.99}\right), \end{aligned} \tag{28}$$

as $\mu = \Theta(1/(\overline{w} \cdot n))$ by (22), $\overline{w} = \Theta(w'_{\mathrm{m}})$, and $|\mathcal{R}|, |\mathcal{C}| = \Theta(n)$. We get the same bound (28) for $|a_r \cdot u_l^t \cdot \mathbf{M}_{\mathcal{R}\times\mathcal{C}} \cdot \vec{1}|$ and $|u_l^t \cdot \mathbf{M}_{\mathcal{R}\times\mathcal{C}} \cdot v_r|$. Thus, (27) is the dominating term in (26), and we get

$$\begin{aligned} \vec{1}^t \cdot M_{\mathcal{R}\times\mathcal{C}} \cdot \vec{1} &\overset{(25)}{=} \vec{1}^t \cdot D_l \cdot \mathbf{M}_{\mathcal{R}\times\mathcal{C}} \cdot D_r \cdot \vec{1} \\ &\overset{(26)}{=} \mu \cdot |\mathcal{R}| \cdot |\mathcal{C}| \cdot \left(1 + O\left(w_{\mathrm{m}}'^{-0.49}\right)\right) \\ &\overset{(22)}{=} \phi_{ij} \cdot W_i \cdot W_j \cdot \frac{|\mathcal{R}| \cdot |\mathcal{C}|}{\overline{w} \cdot n} \cdot (1 + O(w_{\mathrm{m}}'^{-0.49})) \\ &= \phi_{ij} \cdot W_i \cdot W_j \cdot \frac{|\mathcal{R}| \cdot |\mathcal{C}|}{\overline{w} \cdot n} \cdot \left(1 + O\left(\frac{1}{d_{\mathrm{m}}{}^{0.49}}\right)\right). \end{aligned}$$

We come to the second item of Lemma 12. Let $v \perp \vec{1}$ be a unit vector. Then $D_r \cdot v = c \cdot \vec{1}/\|\vec{1}\| + v'$ for some $v' \perp \vec{1}$. As $D_r \cdot \vec{1} = a_r \cdot \vec{1} + v_r$, we get

$$c = \frac{\vec{1}^t \cdot D_r \cdot v}{\sqrt{|\mathcal{C}|}} = \frac{(a_r \cdot \vec{1}^t + v_r^t) \cdot v}{\sqrt{|\mathcal{C}|}} \overset{v \perp \vec{1}}{=} \frac{v_r^t \cdot v}{\sqrt{|\mathcal{C}|}} \leq \frac{\|v_r\| \cdot \|v\|}{\sqrt{|\mathcal{C}|}} \leq 5 \cdot w_{\mathrm{m}}'^{-0.49}$$

and $\|v'\| \leq \|D_r\| \leq 2$. Invoking Lemma 17 and using $w'_{\mathrm{m}} = \Theta(\overline{w})$, we get for any unit vector u

$$\begin{aligned} \left|u^t \cdot D_l \cdot \mathbf{M}_{\mathcal{R}\times\mathcal{C}} \cdot D_r \cdot v\right| &= \left|u^t \cdot D_l \cdot \mathbf{M}_{\mathcal{R}\times\mathcal{C}} \cdot \left(c \cdot \frac{\vec{1}}{\|\vec{1}\|} + v'\right)\right| \\ &\leq \|D_l\| \cdot |c| \cdot \|\mathbf{M}_{\mathcal{R}\times\mathcal{C}}\| + \|D_l\| \cdot \|v'\| \cdot \left|u'^t \cdot \mathbf{M}_{\mathcal{R}\times\mathcal{C}} \cdot \frac{v'}{\|v'\|}\right| \\ &\overset{v' \perp \vec{1}}{\leq} 2 \cdot w_{\mathrm{m}}'^{-0.49} \cdot O\left(\frac{1}{\overline{w}}\right) + 4 \cdot O(w_{\mathrm{m}}'^{-1.5}) = O\left(\frac{1}{\overline{w}^{1.49}}\right). \end{aligned}$$

Applying (25) once more, we obtain

$$\left|u^t \cdot M_{\mathcal{R}\times\mathcal{C}} \cdot v\right| = \left|u^t \cdot D_l \cdot \mathbf{M}_{\mathcal{R}\times\mathcal{C}} \cdot D_r \cdot v\right| = O\left(1/\overline{w}^{1.49}\right).$$

The same bound can be obtained for $u \perp \vec{1}$ and v an arbitrary unit vector. The third item is an immediate consequence of items 1 and 2 of Lemma 12 (cf. the proof of Lemma 17). ∎

4.3. Proof of Lemma 13. We omit the proof for $\|M_{\mathcal{R}\times\overline{\mathcal{C}}}\| = O({d_{\mathrm{m}}}^{-1.5})$, as it is very similar to that for $\|M_{\overline{\mathcal{R}}\times\mathcal{C}}\| = O({d_{\mathrm{m}}}^{-1.5})$.

4.3.1. The spectrum of $M_{\overline{\mathcal{R}}\times\mathcal{C}}$. Let ξ be some $|\mathcal{C}|$-dimensional vector with $\|\xi\| \leq 1$. We show that $\eta = M_{\overline{\mathcal{R}}\times\mathcal{C}} \cdot \xi$ has an l_2-norm bounded above by $O({d_{\mathrm{m}}}^{-1.5})$. We have $\eta_u = 0$ for $u \notin U$ by the construction of M, and for any $u \in U$ we have

$$\eta_u = \sum_{v \in N(u)\cap\mathcal{C}\cap U} \frac{\xi_v}{d_u \cdot d_v};$$

whence

$$\|\eta\|^2 = \sum_{u \in \overline{\mathcal{R}}\cap U} \left(\sum_{v \in N(u)\cap\mathcal{C}\cap U} \frac{\xi_v}{d_u \cdot d_v}\right)^2.$$

Applying the Cauchy–Schwarz inequality, we obtain

$$\begin{aligned} \|\eta\|^2 &\leq \sum_{u \in \overline{\mathcal{R}}\cap U} \left(\sum_{v \in N(u)\cap\mathcal{C}\cap U} \frac{\xi_v^2}{d_u \cdot d_v}\right) \cdot \left(\sum_{v \in N(u)\cap\mathcal{C}\cap U} \frac{1}{d_u \cdot d_v}\right) \\ &\overset{u,v \in U}{\leq} \sum_{u \in \overline{\mathcal{R}}\cap U} \left(\sum_{v \in N(u)\cap\mathcal{C}\cap U} \frac{\xi_v^2}{d_{\mathrm{m}} \cdot d_v}\right) \cdot \frac{1}{d_{\mathrm{m}}} \\ &\leq {d_{\mathrm{m}}}^{-2} \cdot \sum_{u \in \overline{\mathcal{R}}\cap U} \sum_{v \in N(u)\cap\mathcal{C}\cap U} \frac{\xi_v^2}{d_v}. \end{aligned}$$

Since $v \in \mathcal{C}$, v has at most $\mathrm{Vol}(V_i, v) \cdot \Delta/d_\mathrm{m}$ neighbors in $\overline{\mathcal{R}}$. For $v \in \mathcal{C}$ we have $\mathrm{Vol}(V_i, v) < \mathrm{Vol}(V, v) < 2d_v$; otherwise (15) would be false. Thus, each term ξ_v/d_v for $v \in \mathcal{C}$ is counted at most $2d_v \cdot \Delta/d_\mathrm{m}$ times in the sum above. Therefore, we get

$$\|\eta\|^2 \le {d_\mathrm{m}}^{-2} \cdot \sum_{v\in\mathcal{C}} 2d_v \cdot \frac{\Delta}{d_\mathrm{m}} \cdot \frac{\xi_v^2}{d_v} \le 2\Delta \cdot {d_\mathrm{m}}^{-3} \cdot \sum_{v\in\mathcal{C}} \xi_v^2 \le 2\Delta \cdot {d_\mathrm{m}}^{-3}.$$

Consequently, we have $\|M_{\overline{\mathcal{R}}\times\mathcal{C}} \cdot \xi\| = \|\eta\| = O({d_\mathrm{m}}^{-1.5})$ for any ξ with $\|\xi\| \le 1$, which implies the assertion $\|M_{\overline{\mathcal{R}}\times\mathcal{C}}\| = O({d_\mathrm{m}}^{-1.5})$. ∎

4.3.2. The spectrum of $M_{\overline{\mathcal{R}}\times\overline{\mathcal{C}}}$. We postpone the proof of the following lemma to the end of this subsection.

LEMMA 18. *Let $G = G_n(\Phi, w, \mathcal{V})$, and let U be the set constructed by Algorithm 5. Then w.h.p. the following is true for all pairs U_1, U_2 of sets such that $U_1 \subseteq U \cap V_i$, $U_2 \subseteq U \cap V_j$, $\mathrm{Vol}(U_1, V_j) \le n/{d_\mathrm{m}}^4$, and $\mathrm{Vol}(V_i, U_2) \le n/{d_\mathrm{m}}^4$:*

there exist partitions

$$U_1 = U_1^1 \dot\cup U_1^2 \dot\cup \cdots \dot\cup U_1^l \quad \textit{and} \quad U_2 = U_2^1 \dot\cup U_2^2 \dot\cup \cdots \dot\cup U_2^l$$

such that for all $p = 1, \ldots, l$, all $u \in U_1^p$, and all $v \in U_2^p$ simultaneously,

$$\sum_{p'\ge p} s_A(u, U_2^{p'}) \le d_u \cdot \frac{600}{d_\mathrm{m}} \qquad \textit{and} \qquad \sum_{p'\ge p} s_A(U_1^{p'}, v) \le d_v \cdot \frac{600}{d_\mathrm{m}}.$$

By Lemma 11, we can apply Lemma 18 to $U_1 = U \cap \overline{\mathcal{R}}$ and $U_2 = U \cap \overline{\mathcal{C}}$. Let $\overline{\mathcal{R}}^1 \dot\cup \cdots \dot\cup \overline{\mathcal{R}}^l$ $(= U \cap \overline{\mathcal{R}})$ and $\overline{\mathcal{C}}^1 \dot\cup \cdots \dot\cup \overline{\mathcal{C}}^l$ $(= U \cap \overline{\mathcal{C}})$ be the partitions from Lemma 18. To simplify the notation we write $\overline{\mathcal{R}}^{\ge p}$ for $\bigcup_{p'\ge p} \overline{\mathcal{R}}^{p'}$ and analogously $\overline{\mathcal{R}}^{<p}$, $\overline{\mathcal{C}}^{\ge p}$, and $\overline{\mathcal{C}}^{<p}$. Lemma 18 gives for $u \in \overline{\mathcal{R}}^p$ and $v \in \overline{\mathcal{C}}^p$

$$\begin{aligned} &\left|\left\{w \in N(u) \cap \overline{\mathcal{C}}^{\ge p}\right\}\right| = \sum_{p'\ge p} s_A(u, \overline{\mathcal{C}}^{p'}) \le d_u \cdot 600/d_\mathrm{m} \\ \text{and} \quad &\left|\left\{w \in N(v) \cap \overline{\mathcal{R}}^{\ge p}\right\}\right| \le d_v \cdot 600/d_\mathrm{m}. \end{aligned} \tag{29}$$

Let ξ be some $|\overline{\mathcal{C}}|$-dimensional vector with $\|\xi\| \le 1$ and $\eta = M_{\overline{\mathcal{R}}\times\overline{\mathcal{C}}} \cdot \xi$. Let u be an arbitrary vertex from $\overline{\mathcal{R}}$. If $u \notin U$, then $\eta_u = 0$ by the construction of M. Otherwise, $u \in U$, and there exists a p such that $u \in \overline{\mathcal{R}}^p$. Each entry m_{uv} of $M_{\overline{\mathcal{R}}\times\overline{\mathcal{C}}}$ with $v \in \overline{\mathcal{C}} \setminus U$ is set to 0. Such entries do not contribute to η_u. For any nonzero entry m_{uv} we have $v \in \overline{\mathcal{C}}^{<p}$ or $v \in \overline{\mathcal{C}}^{\ge p}$. Thus,

$$\eta_u = \sum_{v\in N(u)\cap\overline{\mathcal{C}}^{<p}} \frac{\xi_v}{d_u \cdot d_v} + \sum_{v\in N(u)\cap\overline{\mathcal{C}}^{\ge p}} \frac{\xi_v}{d_u \cdot d_v}.$$

Further, since $(a+b)^2 \le 2(a^2+b^2)$, we obtain

$$\begin{aligned} \|\eta\|^2 = \sum_{u\in\overline{\mathcal{R}}\cap U} \eta_u^2 \le 2 \cdot \sum_{p=1}^{l} \sum_{u\in\overline{\mathcal{R}}^p} \left(\sum_{v\in N(u)\cap\overline{\mathcal{C}}^{<p}} \frac{\xi_v}{d_u \cdot d_v} \right)^2 \\ + 2 \cdot \sum_{p=1}^{l} \sum_{u\in\overline{\mathcal{R}}^p} \left(\sum_{v\in N(u)\cap\overline{\mathcal{C}}^{\ge p}} \frac{\xi_v}{d_u \cdot d_v} \right)^2. \end{aligned}$$

We can bound the first summand as follows. As $u \in \overline{\mathcal{R}}^p \subseteq U$, we have $d_u \geq d_{\mathrm{m}}$. Using the Cauchy–Schwarz inequality, we get

$$2 \cdot \sum_{p=1}^{l} \sum_{u \in \overline{\mathcal{R}}^p} \left(\sum_{v \in N(u) \cap \overline{\mathcal{C}}^{<p}} \frac{\xi_v}{d_u \cdot d_v} \right)^2$$
$$\leq 2 \cdot \sum_{p=1}^{l} \sum_{u \in \overline{\mathcal{R}}^p} \left(\sum_{v \in N(u) \cap \overline{\mathcal{C}}^{<p}} \frac{\xi_v^2}{d_u \cdot d_v} \right) \cdot \left(\sum_{v \in N(u) \cap \overline{\mathcal{C}}^{<p}} \frac{1}{d_u \cdot d_v} \right)$$
$$\leq 2 d_{\mathrm{m}}^{-2} \cdot \sum_{p=1}^{l} \sum_{u \in \overline{\mathcal{R}}^p} \sum_{v \in N(u) \cap \overline{\mathcal{C}}^{<p}} \frac{\xi_v^2}{d_v}.$$

If $u \in \overline{\mathcal{R}}^p$ and $v \in N(u) \cap \overline{\mathcal{C}}^{<p}$, then there is $p' < p$ such that $v \in \overline{\mathcal{C}}^{p'}$ and $u \in N(v) \cap \overline{\mathcal{R}}^{>p'}$. Therefore, for each v in the sum above, the summand ξ_v^2/d_v occurs at most $\left| N(v) \cap \overline{\mathcal{R}}^{>p'} \right| \leq d_v \cdot 600/d_{\mathrm{m}}$ times; see (29). Hence,

$$2 d_{\mathrm{m}}^{-2} \cdot \sum_{p=1}^{l} \sum_{u \in \overline{\mathcal{R}}^p} \sum_{v \in N(u) \cap \overline{\mathcal{C}}^{<p}} \frac{\xi_v^2}{d_v} \leq 2 d_{\mathrm{m}}^{-2} \cdot \sum_{v \in \overline{\mathcal{C}}} \xi_v^2 \cdot \frac{600}{d_{\mathrm{m}}} \leq 1200 \cdot d_{\mathrm{m}}^{-3}.$$

In a similar way we obtain the same bound for the second summand. Thus, $\|\eta\|^2 = \|M_{\overline{\mathcal{R}} \times \overline{\mathcal{C}}} \cdot \xi\|^2 \leq 2400 \cdot d_{\mathrm{m}}^{-3}$. Since this holds for all ξ, we conclude that $\|M_{\overline{\mathcal{R}} \times \overline{\mathcal{C}}}\| = O(d_{\mathrm{m}}^{-1.5})$.

4.3.3. Proof of Lemma 18. The following lemma is similar to Lemma 8 in [11]. The proof can be done in the same way as in [11].

LEMMA 19. *Let G be a graph generated by our model. Then w.h.p. the following holds for each $1 \leq i, j \leq k$.*

For any two subsets $U \subseteq V_i$ and $U' \subseteq V_j$ such that $u = \max\{|U|, |U'|\} \leq \frac{n}{2}$, one of the following two statements is true:

1. $s_A(U, U') \leq 300 \cdot \mathrm{Vol}(U, U')$.
2. $s_A(U, U') \cdot \ln(s_A(U, U')/\mathrm{Vol}(U, U')) \leq 300 \cdot u \cdot \ln(n/u)$.

We assume that the assertion of Lemma 19 holds. Consider the following procedure with $\Delta \geq 600$ to obtain a partition of U_1 and U_2 as desired:

1. Set $U_1' := U_1, \ U_2' := U_2$.
2. Set $T_1 = T_2 = \emptyset$.
3. While there is some $u \in U_1'$ such that $s_A(u, U_2') \leq d_u \cdot \Delta/d_{\mathrm{m}}$, add u to T_1.
4. While there is some $v \in U_2'$ such that $s_A(U_1', v) \leq d_v \cdot \Delta/d_{\mathrm{m}}$, add v to T_2.
5. Set $U_1' := U_1' \setminus T_1, \ U_2' := U_2' \setminus T_2$.
6. Repeat 2–5 until $|U_1'| = |U_2'| = 0$.

We prove Lemma 18 by showing that the above procedure terminates w.h.p. To obtain a contradiction, we assume the opposite, that at some point at step 6 we have $T_1 = T_2 = \emptyset$, whereas U_1' and U_2' are nonempty. Then for all $u \in U_1'$

$$s_A(u, U_2') > d_u \cdot \Delta/d_{\mathrm{m}} = s_A(u, V) \cdot \Delta/d_{\mathrm{m}}$$

and $s_A(U_1', v) > s_A(V, v) \cdot \Delta/d_{\mathrm{m}}$ for all $v \in U_2'$. Hence,

$$s_A(U_1', U_2') \geq \max\{s_A(U_1', V), s_A(V, U_2')\} \cdot \frac{\Delta}{d_{\mathrm{m}}} \tag{30}$$

$$\overset{U_1', U_2' \subseteq U}{\geq} \Delta \cdot \max\{|U_1'|, |U_2'|\}. \tag{31}$$

As $n/{d_{\mathrm{m}}}^4 \geq \mathrm{Vol}(U_1, V_j) \geq \mathrm{Vol}(U_1', V_j) \geq |U_1'| \cdot d_{\mathrm{m}}$, we have $|U_1'| \leq n/{d_{\mathrm{m}}}^5$ and analogously $|U_2'| \leq n/{d_{\mathrm{m}}}^5$.

Letting $u = \max\{|U_1'|, |U_2'|\}$ and $s = s_A(U_1', U_2')$, we have $u \leq n/{d_{\mathrm{m}}}^5$ and $s \geq \Delta \cdot u$ by (31). We consider several cases to refute the existence of U_1', U_2' as above. In each case we apply Lemma 19 to U_1', U_2'.

1. $\mathrm{Vol}(U_1', U_2') \leq u^{1.5}/\sqrt{n}$.
   (a) Suppose the first condition of Lemma 19 holds. Then

$$300 \cdot u^{1.5}/\sqrt{n} \geq 300 \cdot \mathrm{Vol}(U_1', U_2') \geq s \overset{(31)}{\geq} \Delta \cdot u,$$

   which is false, as $n \geq u$ and $\Delta \geq 300$.
   (b) Suppose the second condition of Lemma 19 holds. Then

$$300 \cdot u \cdot \ln(n/u) \geq s \cdot \ln(s/\mathrm{Vol}(U_1', U_2')).$$

   The right-hand side is monotonically increasing in s for $s > \mathrm{Vol}(U_1', U_2')$. We have $s \geq \Delta \cdot u \geq \Delta \cdot u^{1.5}/n^{0.5} \geq \Delta \cdot \mathrm{Vol}(U_1', U_2')$. Thus, we can replace s by $\Delta \cdot u$ to bound the right-hand side from below. Again, we get a contradiction as $\Delta \geq 600$, as the following calculation shows:

$$\begin{aligned} 300 \cdot u \cdot \ln\left(\frac{n}{u}\right) &\geq \Delta \cdot u \cdot \ln\left(\frac{\Delta \cdot u}{\mathrm{Vol}(U_1', U_2')}\right) \\ &\geq \Delta \cdot u \cdot \ln \frac{\Delta \cdot u}{u^{1.5}/\sqrt{n}} = \frac{\Delta \cdot u}{2} \cdot \ln\left(\Delta^2 \cdot \frac{n}{u}\right). \end{aligned}$$

2. $\mathrm{Vol}(U_1', U_2') > u^{1.5}/\sqrt{n}$ and $s \geq \sqrt{\mathrm{Vol}(U_1', U_2') \cdot n}/d_{\mathrm{m}}$.
   (a) Suppose the first condition of Lemma 19 holds. Then

$$300 \cdot \mathrm{Vol}(U_1', U_2') \geq s \geq \sqrt{\mathrm{Vol}(U_1', U_2') \cdot n}/d_{\mathrm{m}}$$

   yields

$$\mathrm{Vol}(U_1', U_2') \geq n/(300^2 \cdot {d_{\mathrm{m}}}^2),$$

   which is false as $n/{d_{\mathrm{m}}}^4 \geq \mathrm{Vol}(U_1', U_2')$, provided $d_{\mathrm{m}} > 300$.
   (b) Suppose the second condition of Lemma 19 holds. Then

$$\begin{aligned} 300 \cdot u \cdot \ln\left(\frac{n}{u}\right) &\geq s \cdot \ln\left(s/\mathrm{Vol}(U_1', U_2')\right) \\ &\geq \frac{\sqrt{\mathrm{Vol}(U_1', U_2') \cdot n}}{d_{\mathrm{m}}} \cdot \ln \frac{\sqrt{n}}{d_{\mathrm{m}} \cdot \sqrt{\mathrm{Vol}(U_1', U_2')}}. \end{aligned}$$

   The right-hand side of the inequality above is monotonically increasing in $\mathrm{Vol}(U_1', U_2')$ as long as $\mathrm{Vol}(U_1', U_2') \leq n/(\mathrm{e} \cdot d_{\mathrm{m}})^2$, which is true by the assumption of Lemma 18. We replace $\mathrm{Vol}(U_1', U_2')$ by the smaller term $u^{1.5}/\sqrt{n}$ and obtain

$$300 \cdot u \cdot \ln\left(\frac{n}{u}\right) \geq \frac{u^{0.75} \cdot n^{0.25}}{d_{\mathrm{m}}} \cdot \ln\left(\frac{n^{0.75}}{d_{\mathrm{m}} \cdot u^{0.75}}\right).$$

As $u \leq n/{d_{\mathrm{m}}}^5$ the right-hand side is monotonically increasing in n. We decrease this term by using the lower bound $u \cdot {d_{\mathrm{m}}}^5$ on n:

$$\begin{aligned} 300 \cdot u \cdot \ln\left(\frac{n}{u}\right) &\geq \frac{u^{0.75} \cdot (u \cdot {d_{\mathrm{m}}}^5)^{0.25}}{d_{\mathrm{m}}} \cdot \ln\left(\sqrt{\frac{n}{u}} \cdot \frac{(u \cdot {d_{\mathrm{m}}}^5)^{0.25}}{d_{\mathrm{m}} \cdot u^{0.25}}\right) \\ &= u \cdot {d_{\mathrm{m}}}^{0.25} \cdot \ln\left(\sqrt{\frac{n}{u}} \cdot {d_{\mathrm{m}}}^{0.25}\right) \\ &= \frac{{d_{\mathrm{m}}}^{0.25}}{2} \cdot u \cdot \ln\left(\frac{n}{u} \cdot \sqrt{d_{\mathrm{m}}}\right). \end{aligned}$$

Again, we have a contradiction for d_{m} large enough.

3. $\mathrm{Vol}(U_1', U_2') > u^{1.5}/\sqrt{n}$ and $s < \sqrt{\mathrm{Vol}(U_1', U_2') \cdot n}/d_{\mathrm{m}}$.
By the definition of $\mathrm{Vol}(\cdot, \cdot)$, we have

$$\mathrm{Vol}(U_1', U_2') = \frac{\mathrm{Vol}(U_1', V_j) \cdot \mathrm{Vol}(V_i, U_2')}{\mathrm{Vol}(V_i, V_j)}.$$

Assume that $\mathrm{Vol}(U_1', V_j)$ is the larger factor in the enumerator. (If the second one is larger, the contradiction can be derived analogously.) By the assumption $\mathrm{Vol}(U_1', U_2') > u^{1.5}/\sqrt{n}$, we have

$$\mathrm{Vol}(U_1', V_j) \geq \sqrt{\mathrm{Vol}(U_1', U_2') \cdot \mathrm{Vol}(V_i, V_j)} \tag{32}$$

$$> \sqrt{\frac{u^{1.5}}{\sqrt{n}} \cdot d_{\mathrm{m}} \cdot \delta n} > u^{0.75} \cdot n^{0.25} \geq n^{0.25} \cdot |U_1'|^{0.75}. \tag{33}$$

We show below that, by (33) w.h.p., also

$$s_A(U_1', V) \geq s_A(U_1', V_j) \geq \mathrm{Vol}(U_1', V_j)/2 \tag{34}$$

holds. Combining (34) with the assumption

$$s_A(U_1', U_2') = s \leq \sqrt{\mathrm{Vol}(U_1', U_2') \cdot n}/d_{\mathrm{m}}$$

and (30), we obtain

$$\begin{aligned} \frac{\sqrt{\mathrm{Vol}(U_1', U_2') \cdot n}}{d_{\mathrm{m}}} \geq s_A(U_1', U_2') &\overset{(30)}{\geq} s_A(U_1', V) \cdot \frac{\Delta}{d_{\mathrm{m}}} \\ &\overset{(34)}{\geq} \mathrm{Vol}(U_1', V_j) \cdot \frac{\Delta}{2d_{\mathrm{m}}}. \end{aligned}$$

Plugging this estimate into (32), we get

$$\sqrt{\mathrm{Vol}(U_1', U_2') \cdot n} \geq \mathrm{Vol}(U_1', V_j) \cdot \frac{\Delta}{2} \overset{(32)}{>} \sqrt{\mathrm{Vol}(U_1', U_2') \cdot \mathrm{Vol}(V_i, V_j)} \cdot \frac{\Delta}{2};$$

whence $4n/\Delta^2 > \mathrm{Vol}(V_i, V_j)$. This is a contradiction for $\Delta \geq 2$, as we have $\mathrm{Vol}(V_i, V_j) \geq \delta n \cdot d_{\mathrm{m}} > n$.

We are left to show that (33) yields (34). Fix $1 \leq w \leq n/{d_{\mathrm{m}}}^5$ and some set $W \subseteq V_i$, with $|W| = w$ and $\mathrm{Vol}(W, V_j) > n^{0.25} \cdot w^{0.75}$. Item 2 of Fact 3 yields

$$\begin{aligned}\mathbf{Pr}\left[s_A(W, V_j) \leq \mathrm{Vol}(W, V_j)/2\right] &\leq \exp(-\mathrm{Vol}(W, V_j)/8)\\ &\leq \exp\left(-w^{0.75} \cdot n^{0.25}/8\right).\end{aligned}$$

There are at most $\binom{n}{w} \leq (\mathrm{e} \cdot n/w)^w$ sets $W \subseteq V_i$ of size w. Hence, the union bound entails that the probability of the existence of a set W with $|W| = w$ (w is still fixed) is at most

$$\exp\left(-w^{0.75} \cdot n^{0.25}/8 + w \cdot \ln(\mathrm{e} \cdot n/w)\right).$$

For $0 < w \leq n/{d_{\mathrm{m}}}^5$ the exponent is convex, provided d_{m} is large enough. Evaluating at $w = 1$ and $w = n/{d_{\mathrm{m}}}^5$, we see that the exponent is at most $-n^{0.25}/10$. Hence, summing up over all possible values for w, we conclude that the total probability is $o(1)$. Thus, w.h.p. for all W (including $W = U_1'$) we have $s_A(W, V_j) \geq \mathrm{Vol}(W, V_j)/2$.

4.4. Proof of Lemma 11. As a first step, we will show that w.h.p. $\mathrm{Vol}(V \setminus F, V) = \mathrm{Vol}(V, V \setminus F)$ is bounded above by $n/(2{d_{\mathrm{m}}}^4)$. To this end we partition the vertices $u \in V$ according to the numbers $\mathrm{Vol}(u, V_j)$: let

$$I^{t,j} = \{u \in V : 2^t \cdot d_{\mathrm{m}} \leq \mathrm{Vol}(u, V_j) < 2^{t+1} \cdot d_{\mathrm{m}}\}.$$

Since $\mathrm{Vol}(u, V_j)$ is the expected number of neighbors of u in V_j, we have $\mathrm{Vol}(u, V_j) \leq n$, and the choice of d_{m} ensures that $d_{\mathrm{m}} \leq \mathrm{Vol}(u, V_j)$ (see Fact 6). Thus, for each $j \in \{1, 2\}$ the partition $I^{0,j}, I^{1,j}, I^{2,j}, \ldots$ features at most $\log n$ nonempty sets.

Fix j and t. Let $u \in I^{t,j}$, and let X_u be the 0/1 random variable indicating that $u \notin F$ because the number of its neighbors in V_j is not sufficiently concentrated about its expectation (i.e., $X_u = 1$ iff $|s_A(u, V_j) - \mathrm{Vol}(u, V_j)| > \mathrm{Vol}(u, V_j)^{0.51}$). Then by Fact 3 $\mathbf{Pr}[X_u = 1]$ is bounded above by

$$\begin{aligned}\mathbf{Pr}\left[|s_A(u, V_j) - \mathrm{Vol}(u, V_j)| \geq \mathrm{Vol}(u, V_j)^{0.51}\right] &\leq 2 \cdot \exp\left(-\mathrm{Vol}(u, V_j)^{0.02}/4\right)\\ &\leq 2 \cdot \exp\left(-(2^t \cdot d_{\mathrm{m}})^{0.02}/4\right).\end{aligned}$$

Consequently, the expected number of elements $u \in I^{t,j} \setminus F$ is at most $\mathbf{E}\left[\sum_{u \in I^{t,j}} X_u\right]$, which is bounded above by $2 \cdot \exp\left(-(2^t \cdot d_{\mathrm{m}})^{0.02}/4\right) \cdot n$.

As the variance of $\sum_{u \in I^{t,j}} X_u$ is at most linear in $\mathbf{E}\left[\sum_{u \in I^{t,j}} X_u\right]$, Chebyshev's inequality entails that with probability $1 - O(1/n)$

$$\left|I^{t,j} \setminus F\right| = \left|\sum_{u \in I^{t,j}} X_u\right| \leq 4 \cdot \exp\left(-(2^t \cdot d_{\mathrm{m}})^{0.02}/4\right) \cdot n.$$

By the union bound, with probability $1 - O(\log n/n)$ this bound holds for all $0 \leq t < \log n$ and all $j \in \{1, 2\}$ simultaneously.

In case $d_{\mathrm{m}} \geq \log^{51} n$, w.h.p. $\left|I^{t,j} \setminus F\right| = 0$ for all t, j. Hence, $F = V$, both $\overline{\mathcal{R}}$ and $\overline{\mathcal{C}}$ are empty, and Lemma 11 holds trivially.

We proceed with the case $d_{\mathrm{m}} < \log^{51} n$. Each $u \in I^{t,j} \setminus F$ contributes

$$\mathrm{Vol}(u, V) = O(\mathrm{Vol}(u, V_j)) = O(2^{t+1} \cdot d_{\mathrm{m}})$$

to $\mathrm{Vol}(V \setminus F, V)$. Therefore, the total contribution to $\mathrm{Vol}(V \setminus F, V)$ of vertices whose number of neighbors in V_j is not concentrated about its expectation amounts to

$$\begin{aligned}\sum_{t\geq 0} \mathrm{Vol}(I^{t,j} \setminus F, V) &= \sum_{t\geq 0} \sum_{u\in I^{t,j}\setminus F} \mathrm{Vol}(u, V)\\ &\leq \sum_{t\geq 0} \left|I^{t,j} \setminus F\right| \cdot O(2^{t+1} \cdot d_{\mathrm{m}})\\ &\leq \sum_{t\geq 0} 4 \cdot \exp\bigl(-(2^t \cdot d_{\mathrm{m}})^{0.02}/4\bigr) \cdot n \cdot O(2^{t+1} \cdot d_{\mathrm{m}})\\ &\leq \sum_{t\geq 0} n/(2^{t+3} \cdot {d_{\mathrm{m}}}^4) = n/(4 \cdot {d_{\mathrm{m}}}^4).\end{aligned}$$

(The third step holds, provided that d_{m} is large enough. To see this, recall that $e^{-x^{0.02}} \cdot x < 1/x^4$ for sufficiently large x. The same argument is used to obtain the last step from the second last one.) Hence,

$$\mathrm{Vol}(V \setminus F, V) \leq \sum_{j=1}^{2} \sum_{t\geq 0} \mathrm{Vol}(I^{t,j} \setminus F, V) \leq 2 \cdot n/(4 \cdot {d_{\mathrm{m}}}^4) = n/(2 \cdot {d_{\mathrm{m}}}^4).$$

In summary, we have shown that with probability $1 - O(\log n/n)$

$$\mathrm{Vol}(V, V \setminus F) = \mathrm{Vol}(V \setminus F, V) \leq \frac{n}{2 \cdot {d_{\mathrm{m}}}^4}. \tag{35}$$

Assume for contradiction that either of $\mathrm{Vol}(\overline{\mathcal{R}}, V_j)$, $\mathrm{Vol}(V_i, \overline{\mathcal{C}})$ exceeds $n/{d_{\mathrm{m}}}^4$. Then the construction process for $\mathcal{R}$ and $\mathcal{C}$ reaches a point where either $\mathrm{Vol}(V_i \setminus R'_{ij}, V_j)$ or $\mathrm{Vol}(V_i, V_j \setminus C'_{ij})$ exceeds $n/{d_{\mathrm{m}}}^4$. At this point we have either

$$\mathrm{Vol}(V_i \setminus R'_{ij}, V_j) > n/{d_{\mathrm{m}}}^4 \qquad \text{and} \qquad \mathrm{Vol}(V_i, V_j \setminus C'_{ij}) \leq n/{d_{\mathrm{m}}}^4 \tag{36}$$

or

$$\mathrm{Vol}(V_i \setminus R'_{ij}, V_j) \leq n/{d_{\mathrm{m}}}^4 \qquad \text{and} \qquad \mathrm{Vol}(V_i, V_j \setminus C'_{ij}) > n/{d_{\mathrm{m}}}^4.$$

We refute (36) in detail; the second case can be ruled out analogously.

We interrupt our process at the *first* occurrence of (36). Let u be the vertex that causes the interruption. Then we have

$$n/{d_{\mathrm{m}}}^4 < \mathrm{Vol}(V_i \setminus R'_{ij}, V_j) \leq n/{d_{\mathrm{m}}}^4 + w'_u.$$

By C2 all expected degrees are $O(n^{1-\varepsilon})$ for some $\varepsilon > 0$. As $d_{\mathrm{m}} \leq \log^{51} n$, we get

$$n/{d_{\mathrm{m}}}^4 < \mathrm{Vol}(V_i \setminus R'_{ij}, V_j) \leq n/{d_{\mathrm{m}}}^4 + O(n^{1-\varepsilon}) = n/{d_{\mathrm{m}}}^4 \cdot (1 + o(1)). \tag{37}$$

Since $R'_{ij} \subseteq F$, we have $V_i \setminus R'_{ij} = (V_i \setminus F) \cup ((V_i \cap F) \setminus R'_{ij})$; whence

$$\frac{n}{{d_{\mathrm{m}}}^4} \overset{(37)}{<} \mathrm{Vol}(V_i \setminus R'_{ij}, V_j) = \mathrm{Vol}(V_i \setminus F, V_j) + \mathrm{Vol}((V_i \cap F) \setminus R'_{ij}, V_j).$$

As $\mathrm{Vol}(V_i \setminus F, V_j) \le \mathrm{Vol}(V \setminus F, V) \le n/(2 \cdot {d_\mathrm{m}}^4)$ by (35) we have

$$\mathrm{Vol}((V_i \cap F) \setminus R'_{ij}, V_j) > n/(2 \cdot {d_\mathrm{m}}^4). \quad (38)$$

Any vertex $u \in V_i \cap F$ that gets removed from R'_{ij} in step 3 has (at the moment of its deletion) at least $\mathrm{Vol}(u, V_j) \cdot \Delta / d_\mathrm{m}$ neighbors in $V_j \setminus C'_{ij}$. Thus,

$$\begin{aligned} s_A(V_i \setminus R'_{ij}, V_j \setminus C'_{ij}) &\ge s_A((V_i \cap F) \setminus R'_{ij}, V_j \setminus C'_{ij}) \\ &\ge \sum_{u \in (V_i \cap F) \setminus R'_{ij}} \frac{\Delta \cdot \mathrm{Vol}(u, V_j)}{d_\mathrm{m}} \\ &= \frac{\Delta \cdot \mathrm{Vol}((V_i \cap F) \setminus R'_{ij}, V_j)}{d_\mathrm{m}} \overset{(38)}{>} \frac{\Delta \cdot n}{2 \cdot {d_\mathrm{m}}^5}, \end{aligned}$$

whereas

$$\begin{aligned} \mathrm{Vol}(V_i \setminus R'_{ij}, V_j \setminus C'_{ij}) &= \frac{\mathrm{Vol}(V_i \setminus R'_{ij}, V_j) \cdot \mathrm{Vol}(V_i, V_j \setminus C'_{ij})}{\mathrm{Vol}(V_i, V_j)} \\ &\overset{(36),(37)}{\le} \frac{n^2 \cdot (1 + o(1))}{{d_\mathrm{m}}^8 \cdot \mathrm{Vol}(V_i, V_j)} \le \frac{n^2 \cdot (1 + o(1))}{{d_\mathrm{m}}^8 \cdot \delta n \cdot d_\mathrm{m}} \le \frac{n}{{d_\mathrm{m}}^8}. \end{aligned}$$

We apply Lemma 19 with $U = V_i \setminus R'_{ij}$ and $U' = V_j \setminus C'_{ij}$. We have $u = \max\{|U|, |U'|\} \le \mathrm{Vol}(U, V_j)/d_\mathrm{m} \le n/{d_\mathrm{m}}^5 \cdot (1 + o(1)) < n/2$. Rewriting the two inequalities above, we get

$$s_A(U, U') \ge \frac{\Delta \cdot n}{2 \cdot {d_\mathrm{m}}^5}, \qquad \text{whereas} \qquad \mathrm{Vol}(U, U') \le \frac{n}{{d_\mathrm{m}}^8}.$$

Clearly, the first item of Lemma 19 does not hold. Assume that the second one holds. Then we have

$$\frac{\Delta \cdot n}{2 \cdot {d_\mathrm{m}}^5} \cdot \ln \frac{\Delta \cdot {d_\mathrm{m}}^3}{2} < s_A(U, U') \cdot \ln \frac{s_A(U, U')}{\mathrm{Vol}(U, U')} \le 300 \cdot u \cdot \ln \frac{n}{u} \le \frac{600 \cdot n}{{d_\mathrm{m}}^5} \cdot \ln \frac{{d_\mathrm{m}}^5}{2}.$$

However, this is false for $\Delta > 2400$ (remember that we are assuming that Δ exceeds some large enough constant). Thus, w.h.p. (36) is false, and therefore $\mathrm{Vol}(\overline{\mathcal{R}}, V_j) \le n/{d_\mathrm{m}}^4$.

Finally, the third item of Lemma 11 follows from

$$\mathrm{Vol}(\overline{\mathcal{R}}, \overline{\mathcal{C}}) = \frac{\mathrm{Vol}(\overline{\mathcal{R}}, V_j) \cdot \mathrm{Vol}(V_i, \overline{\mathcal{C}})}{\mathrm{Vol}(V_i, V_j)} \le \frac{n^2}{{d_\mathrm{m}}^8 \cdot d_\mathrm{m} \cdot \delta n} \le \frac{n}{{d_\mathrm{m}}^8}.$$

Acknowledgments. We would like to thank two anonymous referees for their comments, which have led to a significantly improved presentation.

REFERENCES

[1] W. Aiello, F. Chung, and L. Lu, *A random graph model for massive graphs*, in Proceedings of the 33rd ACM STOC, 2001, pp. 171–180.

[2] N. Alon, *Spectral Techniques in Graph Algorithms*, Lecture Notes in Comput. Sci. 1380, Springer, New York, 1998, pp. 206–215.

[3] N. Alon and N. Kahale, *A spectral technique for coloring random 3-colorable graphs*, SIAM J. Comput., 26 (1997), pp. 1733–1748.

[4] N. Alon, M. Krivelevich, and V. Vu, *On the concentration of the eigenvalues of random symmetric matrices*, Israel J. Math., 131 (2002), pp. 259–267.
[5] Y. Azar, A. Fiat, A. Karlin, F. McSherry, and J. Saia, *Spectral analysis of data*, in Proceedings of the 33rd ACM STOC, 2001, pp. 619–626.
[6] R. Boppana, *Eigenvalues and graph bisection: An average case analysis*, in Proceedings of the 28th IEEE FOCS, 1987, pp. 280–285.
[7] F. Chung, *Spectral Graph Theory*, American Mathematical Society, Providence, RI, 1997.
[8] F. Chung and L. Lu, *Connected components in random graphs with given expected degree sequences*, Ann. Comb., 6 (2002), pp. 125–145.
[9] F. Chung, L. Lu, and V. Vu, *The spectra of random graphs with given expected degrees*, Internet Math., 1 (2003), pp. 257–275.
[10] A. Coja-Oghlan, *On the Laplacian eigenvalues of* $G_{n,p}$, Combin. Probab. Comput., 16 (2007), pp. 923–946.
[11] A. Coja-Oghlan and A. Lanka, *The Spectral Gap of Random Graphs with Given Expected Degrees*, Lecture Notes in Comput. Sci. 4051, Springer, New York, 2006, pp. 15–26.
[12] A. Coja-Oghlan, A. Goerdt, and A. Lanka, *Spectral Partitioning of Random Graphs with Given Expected Degrees—Detailed Version*, Technical report CSR-09-01, Chemnitz Informatics Reports, Technical University Chemnitz, Chemnitz, Germany.
[13] A. Dasgupta, J. Hopcroft, R. Kannan, and P. Mitra, *Spectral clustering by recursive partitioning*, in Proceedings of the 14th ESA, 2006, pp. 256–267.
[14] A. Dasgupta, J. Hopcroft, and F. McSherry, *Spectral analysis of random graphs with skewed degree distributions*, in Proceedings of the 45th IEEE FOCS, 2004, pp. 602–610.
[15] C. Ding, *Analysis of gene expression profiles: Class discovery and leaf ordering*, in Proceedings of the 6th International Conference on Computational Biology, 2002, pp. 127–136.
[16] U. Feige and E. Ofek, *Spectral techniques applied to sparse random graphs*, Random Structures Algorithms, 27 (2005), pp. 251–275.
[17] A. Flaxman, *A spectral technique for random satisfiable 3CNF formulas*, Random Struct. Algorithms, 32 (2008), pp. 519–534.
[18] J. Friedman, J. Kahn, and E. Szemerédi, *On the second eigenvalue in random regular graphs*, in Proceedings of the 21st ACM STOC, 1989, pp. 587–598.
[19] Z. Füredi and J. Komloś, *The eigenvalues of random symmetric matrices*, Combinatorica, 1 (1981), pp. 233–241.
[20] J. Giesen and D. Mitsche, *Reconstructing many partitions using spectral techniques*, in Proceedings of the 15th FCT, 2005, pp. 433–444.
[21] S. Janson, T. Łuczak, and A. Ruciński, *Random Graphs*, John Wiley and Sons, New York, 2000.
[22] J. Jensen, *Sur les fonctions convexes et les inegalites entre les valeurs moyennes*, Acta Math., 30 (1906), pp. 175–193.
[23] M. Krivelevich and B. Sudakov, *The largest eigenvalue of sparse random graphs*, Combin. Probab. Comput., 12 (2003), pp. 61–72.
[24] F. McSherry, *Spectral partitioning of random graphs*, in Proceedings of the 42nd IEEE FOCS, 2001, pp. 529–537.
[25] M. Mihail and C. Papadimitriou, *On the eigenvalue power law*, in Proceedings of the 6th RANDOM, 2002, pp. 254–262.
[26] A. Pothen, H. Simon, and K. Liou, *Partitioning sparse matrices with eigenvectors of graphs*, SIAM J. Matrix Anal. Appl., 11 (1990), pp. 430–452.
[27] K. Schloegel, G. Karypis, and V. Kumar, *Graph partitioning for high performance scientific simulations*, in CRPC Parallel Computation Handbook, J. Dongarra, I. Foster, G. Fox, K. Kennedy, and W. White, eds., Morgan Kaufmann, San Mateo, CA, 2000.
[28] D. Spielman and S. Teng, *Spectral partitioning works: Planar graphs and finite element meshes*, in Proceedings of the 36th IEEE FOCS, 1996, pp. 96–105.