Partitioning Random Graphs with General Degree Distributions

Amin Coja-Oghlan[1] and André Lanka[2]

[1] Carnegie Mellon University, Department of Mathematical Sciences,
Pittsburgh, PA 15213, USA
amincoja@andrew.cmu.edu
[2] Fakultät für Informatik, Technische Universität Chemnitz
Straße der Nationen 62, 09107 Chemnitz, Germany
lanka@informatik.tu-chemnitz.de

Abstract. We consider the problem of recovering a planted partition (e.g., a small bisection or a large cut) from a random graph. During the last 30 years many algorithms for this problem have been developed that work provably well on models resembling the Erdős-Rényi model $G_{n,m}$. Since in these random graph models edges are distributed very uniformly, the recent theory of large networks provides convincing evidence that real-world networks, albeit looking random in some sense, cannot sensibly be described by these models. Therefore, a variety of new types of random graphs have been introduced. One of the most popular of these new models is characterized by a prescribed expected degree sequence. We study a natural variant of this model that features a planted partition, the main result being that there is a polynomial time algorithm for recovering (a large share of) the planted partition efficiently. In contrast to prior work, the algorithm's input *only* consists of the graph, i.e., no further parameters of the distribution (such as the expected degree sequence) are required.

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 less well defined problems such as recovering a "latent" clustering of the vertices of a graph. Examples of such clustering problems occur in information retrieval [4], scientific simulation [18], or bioinformatics [10]. Furthermore, an important advantage of spectral methods is their efficiency, as there are very fast algorithms for computing eigenvectors, in particular in the case of sparse graphs/matrices.

Despite their success in applications (e.g., [17, 18]), for most of the known spectral heuristics there are counterexamples known showing that these algorithms perform badly in the "worst case". Thus, understanding the conditions that cause spectral heuristics to succeed (as well as their limitations) is an im-

Please use the following format when citing this chapter:

Coja-Oghlan, A. and Lanka, A., 2008, in IFIP International Federation for Information Processing, Volume 273; Fifth IFIP International Conference on Theoretical Computer Science; Giorgio Ausiello, Juhani Karhumäki, Giancarlo Mauri, Luke Ong; (Boston: Springer), pp. 127–141.

portant research problem. To address this problem, quite a few authors have performed rigorous analyses of spectral techniques on suitable models of *random graphs*. Examples include Alon and Kahale [3] (Graph Coloring), Boppana [5] (Minimum Bisection), and McSherry [15] (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 shows that in many cases real-world instances of partitioning problems have a considerably more involved degree distribution [1]. Since most spectral heuristics are very sensitive to fluctuations of the degree distribution, this means that most of the previous spectral methods do not apply to such real-world inputs. Indeed, none of the algorithms from [3, 5, 15] 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 comprises *sparse* graphs, i.e., the case that the average degree remains bounded as the number of vertices grows. This case is of particular practical interest, as many real-world networks turn out to be sparse [1].

The present work is an extension of our prior paper [9] on the same subject. The crucial improvement that we achieve in the present work is that the algorithm *only* requires the graph as an input. By contrast, the algorithm in [9] requires further inputs (namely, parameters of the random graph model such as the expected degree of each vertex), which generally will not be available in practice. Hence, the present work is a step towards spectral methods that apply to graphs with general degree distributions – and in fact to *sparse* graphs.

In Section 2 we describe the random graph model and state the main result. Then, in Section 3 we discuss related work, and Section 4 contains the algorithm and its analysis.

2 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 [9] and resembles the model investigated in Dasgupta, Hopcroft, and McSherry [11]. Moreover, it is based on the "given expected degrees" model of Chung and Lu [7], which we modify in order to incorporate a planted partition.

Let $V = \{1, \ldots, n\}$ be the set of nodes. The first parameter of the model is a symmetric 2×2-matrix $\Phi = (\phi_{ij})$ of full rank with non-negative constants as entries. Furthermore, for each vertex u there is a weight $w_u > 0$; let $\overline{w} =$

$\sum_{u\in V} w_u/n$ be the average weight. In addition, let V_1, V_2 be a partition of V into two subsets; this is going to be *planted partition* that the algorithm is supposed to recover. For each $u \in V$ we let $\psi(u) \in \{1,2\}$ denote the index of the subset u belongs to, that is $u \in V_{\psi(u)}$.

Now, the random graph $G = G(V_1, V_2, \Phi, w_1, \ldots, w_n) = (V, E)$ is obtained by inserting each possible edge $\{u, v\}$ with $u, v \in V$ independently with probability

$$\phi_{\psi(u),\psi(v)} \cdot \frac{w_u \cdot w_v}{\overline{w} \cdot n}. \tag{1}$$

Of course, we insist on the parameters Φ and w_u being chosen such that each of the above terms is bounded above by 1. Let d_u signify the degree of $u \in V$, and let w'_u be the expected degree. Then (1) yields

$$w'_u = \mathbf{E}[d_u] = \frac{w_u}{\overline{w} \cdot n} \cdot \sum_{v\in V} w_v \cdot \phi_{\psi(u),\psi(v)}. \tag{2}$$

We say that the random graph $G = G(V_1, V_2, \Phi, w_1, \ldots, w_n)$ has some property $\mathcal{P}$ *with high probability* ("w.h.p.") if the probability that $\mathcal{P}$ holds tends to 1 as $n \to \infty$, uniformly for any feasible choice of V_1, V_2, Φ and $w_1, \ldots, w_n$.

Let us briefly discuss the meaning of the model's parameters. As (2) shows, the expected degree of $u \in V$ is proportional to w_u. Thus, the purpose of the weights w_u is to model the desired degree sequence (e.g., a power law). Furthermore, the matrix Φ rules the edge density inside the classes V_1, V_2 and the density of the bipartite graph consisting of the V_1-V_2 edges; for by (2) the edge density of V_1 (resp. V_2) is proportional to ϕ_{11} (resp. ϕ_{22}), and the V_1-V_2-edge density is proportional to $\phi_{12} = \phi_{21}$. Thus, the weight w_u influences the degree of u, while the matrix Φ yields what proportion of u's neighbors belong to V_1 or V_2.

For instance, to model a graph with a small bisection, we could set $\phi_{11} = \phi_{22} = 0.51$ and $\phi_{12} = 0.49$. Moreover, we let $V_1, V_2 \subset V$ be two randomly chosen disjoint sets of size $n/2$. Finally, setting $w_u = d \cdot u^{\frac{1}{2}}$, we obtain a graph with a power law degree distribution (with average degree about $2d$) and a "planted bisection" containing about 49% of all edges. Other examples include graphs with planted independent sets, planted dense spots etc.

Theorem 1. *There is a polynomial time algorithm $\mathcal{A}$ such that the following holds. Let $\varepsilon, \delta > 0$ be arbitrarily small but fixed, and let $C = C(\varepsilon, \delta)$ be a sufficiently large constant. Moreover, assume that*

1. $|V_1|, |V_2| \geq \delta n$,
2. for all $u \in V$ the weight w_u satisfies $\varepsilon\overline{w} \leq w_u \leq n^{1-\varepsilon}$, and
3. the average weight satisfies $\overline{w} \geq C$.

Then w.h.p. $\mathcal{A}$ applied to $G = G(V_1, V_2, \Phi, w_1, \ldots, w_n)$ outputs a partition V'_1, V'_2 that differs from the planted partition V_1, V_2 on at most $n \cdot \ln \overline{w} / \overline{w}^{0.98}$ vertices; that is, $\min\{|V_1 \triangle V'_1| + |V_2 \triangle V'_2|, |V_1 \triangle V'_2| + |V_2 \triangle V'_1|\} \leq n \cdot \ln \overline{w} / \overline{w}^{0.98}$.

Note that the number of vertices that $\mathcal{A}$ may not classify correctly decreases as $\overline{w}$ grows. Indeed, if $\overline{w} = O(1)$, i.e., if G is a *sparse* graph with average degree $O(1)$, then it is *impossible* to recover the partition V_1, V_2 perfectly. A simple reason for this is that w.h.p. both V_1 and V_2 will contain a linear number $\Omega(n)$ of isolated vertices. Nevertheless, a large share of the vertices gets partitioned correctly w.h.p. Moreover, we emphasize that the input of the algorithm *only* consists of the graph G; no further parameters of the model are revealed to $\mathcal{A}$.

Although we have stated Thereom 1 only for a planted partition V_1, V_2 with two classes, the techniques generalize to the case of an arbitrarily large but bounded number k of classes. We omit the details to simplify the exposition.

3 Related work

The general relationship between spectral properties of the adjacency matrix of a graph and clustering problems has been investigated thoroughly [2]. Usually this relationship is based on some separation between the few largest eigenvalues in absolute value (which then represent the clusters) and the remaining eigenvalues. Along these lines theoretically rigorous analyses of spectral methods have been conducted, mainly stating that a certain algorithm performs well on a certain random graph model. Indeed, this has lead to provably efficient algorithms for clustering problems in situations where purely combinatorial algorithms do not seem to work; examples include Alon and Kahale [3] (3-coloring), Boppana [5] (graph bisection), and McSherry [15] (recovering a "latent" partition). In particular [3] has inspired further results (e.g., Flaxman's work on 3-SAT [12]).

However, the aforementioned results do not yield spectral algorithms for clustering graphs whose degree distribution features a heavy upper tail, e.g., a power law degree distribution. Nonetheless, these degree distributions occur prominently in large real world networks [1]. In fact, Mihail and Papadimitriou [16] proved that in the case of a power-law the spectrum of the adjacency matrix merely reflects the upper tail of the degree distribution, but provides no clue on global graph properties (such as the presence of dense clusters or a large cut). Furthermore, in the case of a heavily-tailed degree distribution it is not an option to just remove high degree vertices, because significant parts of the graph may just be ignored in this way. Thus, the adjacency matrix is inappropriate to represent graphs with heavy-tailed degree distributions.

To cope with a heavily-tailed degree distribution, the Laplacian matrix has been used in both theoretical (e.g. [6]) and practically oriented work [17]. However, for randomly generated graphs the Laplacian is significantly more difficult to study than the adjacency matrix (because the entries are heavily dependent). Nonetheless, Dasgupta, Hopcroft, and McSherry [11] showed that clustering problems on sufficiently dense random graphs with a general degree distribution (say, average degree $\gg \ln^6(n)$, where n is the number of vertices) can be

solved efficiently using the Laplacian. More precisely, [11] deals with essentially the same model as considered in the present paper (though they additionally deal with the case $k > 2$). However, the assumption that the average degree is $\gg \ln^6 n$ turns out to be crucial in [11] (because the paper employs the "trace method" from Füredi and Komlós [13] for analyzing the Laplacian spectrum). Hence, in comparison to [11] the new aspect of the present work is that our result covers *sparse* graphs (of average degree $O(1)$), which seem most appropriate to model real networks [1]. In fact, the case of sparse graphs is posed as an open problem in [11].

In a prior paper [9] we studied the same random graph model and presented an algorithm for recovering (a large part of) the planted partition efficiently, provided that the *expected* degree distribution $(\mathbf{E}[d_v])_{v\in V}$ is given as a further input parameter to the algorithm. This assumption is crucial in that paper, because the algorithm exploits the spectrum of the matrix $\mathbf{M} = (\mathbf{m}_{uv})_{u,v\in V}$ with entries

$$\mathbf{m}_{uv} = \begin{cases} (\mathbf{E}[d_u]\,\mathbf{E}[d_v])^{-1} & \text{if } u, v \text{ are adjacent,} \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

In fact, in the sparse case (average degree $O(1)$), the vertex degrees d_v are *not* tightly concentrated about their means (as there tails of Poisson type), so that it is impossible to recover/approximate the expected degree distribution $(\mathbf{E}[d_v])_{v\in V}$ sufficiently well in terms of the actual degree distribution $(d_v)_{v\in V}$. Therefore, the assumption that the algorithm is given the expected degree sequence is inevitable in order to set up the matrix (3). Of course, this assumption is rather impractical, because it reduces the applicability of the algorithm to artificially generated instances.

To avoid the assumption that the expected degree sequence is given to the algorithm, we fix (3) by instead considering the matrix $M = (m_{uv})_{u,v\in V}$ with entries

$$m_{uv} = \begin{cases} (d_u d_v)^{-1} & \text{if } u, v \text{ are adjacent,} \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Hence, we replace the expected degrees by the actual vertex degrees of the input graph. In effect, while the entries of (3) are mutually independent (up to the trivial dependence due to symmetry), the entries of (4) are *mutually dependent*. This issue complicates the *analysis* of the algorithm – in particular, the analysis of the spectrum of M – significantly; to cope with these new issues, we build upon methods that we developed recently in [8]. Furthermore, the algorithm needs to proceed more carefully, as the actual vertex degrees may deviate from their means considerably. Thus, in comparison to [9] the contribution of the present work is that we obtain a much more practical algorithm, and present significantly more sophisticated techniques for analyzing its performance on random graphs.

4 The algorithm and its analysis

Throughout this section we keep the notation and the assumptions of Theorem 1.

4.1 Notation and preliminaries

If ξ is a vector, then $\|\xi\|$ denotes its ℓ_2-norm. Moreover, for a $m \times n$ matrix B we let $\|B\| = \max_{\xi \in \mathbf{R}^n,\, \|\xi\|=1} \|B\xi\|$ denote the operator norm. The transpose of B is written as B^t. Furthermore, $\mathbf{1}$ signifies the vector with all entries equal to 1 (in any dimension). If $\xi \in \mathbf{R}^S$ and $U \subseteq S$, then $\xi_{|U} \in \mathbf{R}^S$ signifies the vector obtained by replacing the i'th component of ξ by 0 if $i \notin U$, whereas $\xi_U \in \mathbf{R}^U$ is obtained from ξ by deleting all entries ξ_v with $v \notin U$. In addition, if B is a $m \times n$ matrix and $X \subseteq \{1, \ldots, m\}$, $Y \subseteq \{1, \ldots, n\}$, then $B_{X \times Y}$ denotes the minor of B induced on $X \times Y$. Further, if $M = (m_{uv})$ is a matrix and X (resp. Y) is a set of rows (columns), then we set

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

If u is a vertex of a graph $G = G(V_1, V_2, \Phi, w_1, \ldots, w_n)$, then $N(u) = \{v : \{u, v\} \in E\}$ denotes the neighborhood of u. Moreover, for two sets U_1, U_2 of vertices we define the volume of (U_1, U_2) to be

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

if U_1 and U_2 are disjoint, then $\mathrm{Vol}(U_1, U_2)$ equals the *expected* number of U_1-U_2-edges. In other words, if $A = A(G)$ is the adjacency matrix, then $\mathrm{Vol}(U_1, U_2) = \mathbf{E}[s_A(U_1, U_2)]$.

The following Chernoff bounds will prove useful in several places (cf. [14, Theorems 2.1 and 2.8]).

Fact 2. *Let X be the sum of independent 0–1 random variables. Then*

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

for all $t \geq 0$.

Finally, we collect a few simple observations concerning the random graph model.

Lemma 3. *Suppose that $G = G(V_1, V_2, \Phi, w_1, \ldots, w_n)$ is a random graph.*

1. *Let u_1, 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 exists a constant $C = C(\Phi, \varepsilon, \delta)$ such that $1/C \leq w'_u/w_u \leq C$ for all $u \in V$.*
3. *The expected average degree of G equals $\overline{w}\,' = \sum_{u\in V} w'_u/n = \Theta(\overline{w}\,)$.*

Since by Lemma 3 the quotient w_u/w'_u coincides for all $u \in V_i$, we abbreviate

$$W_i = w_u/w'_u = \Theta(1), \qquad \text{and} \qquad W = \overline{w}\,/\overline{w}\,' = \Theta(1). \tag{5}$$

4.2 The algorithm

The algorithm $\mathcal{A}$ for Theorem 1 reads as follows.

Algorithm 4.
Input: A graph $G = (V, E)$.
Output: A partition V'_1, V'_2 of V.

1. Calculate the average degree $\overline{d} = \sum_{u=1}^{n} d_u/n$ of G and set $d_{\mathrm{m}} = \overline{d}\,/\ln \overline{d}$.
2. Construct the matrix $M = (m_{uv})_{u,v\in V}$ as described in (4).
3. Let $U = \{u \in V : d_u \geq d_{\mathrm{m}}\}$ be the set of all vertices whose degree is "not too small".
4. Obtain M^* from M by replacing any entry m_{uv} with $(u, v) \notin U \times U$ by 0.
5. Let s_1, s_2 be the eigenvectors of M^* with the two largest eigenvalues in absolute value. Scale s_i such that $\|s_i\| = \sqrt{n}$.
6. If at least one of s_1, s_2 enjoys the following property:

   There are $c_1, c_2 \in \mathbb{R}$ with $|c_1 - c_2| > 1/4$ such that more than $n/\sqrt{d_{\mathrm{m}}}$ vertices $v \in U$ satisfy $|s_i(v) - c_1| \leq 1/32$ and more than $n/\sqrt{d_{\mathrm{m}}}$ vertices satisfy $|s_i(v) - c_2| \leq 1/32$, (6)

   then let $s \in \{s_1, s_2\}$ be such an eigenvector. Furthermore, let V'_1 be the vertices whose corresponding entries in s are closer to c_1 than to c_2 and set $V'_2 = V \setminus V'_1$. Otherwise, if neither s_1 nor s_2 enjoys (6), let $V'_1 = V$ and $V'_2 = \emptyset$ (in this case, the algorithm fails).

Before we sketch the analysis of the algorithm, let us briefly discuss the basic ideas that it is based on. In its first step, $\mathcal{A}$ just computes the average degree and the value d_{m}. This value is assumed to be a lower bound on the degree that a vertex should typically have; that is, all vertices with degree $< d_{\mathrm{m}}$ are considered exceptional. Note that this is consistent with assumption 2. of Theorem 1, which entails that $\mathbf{E}[d_u] \geq \delta\varepsilon^2 \cdot \min_{\phi_{ij}>0}(\phi_{ij}) \cdot \overline{d} > d_{\mathrm{m}}$ for all $u \in V$.

Step 2 of the algorithm then sets up the matrix M, whose eigenvectors we are going to use in order to partition G. Note that the entry corresponding

to an edge $\{u, v\}$ is normalized by the product $d_u d_v$ of the vertex degrees; this normalization is crucial as it ensures that the upper tail of the degree distribution does not dominate the spectrum of M (in contrast to the case of the adjacency matrix, cf. Section 3).

While the normalization of the entries of M ensures that the upper tail of the degree distribution does not dominate the spectrum of M, vertices of atypically small degree may induce large eigenvalues (cf. [8]). Therefore, before computing the dominant eigenvectors s_1, s_2 in Step 5, Steps 3 and 4 remove all entries of M that involve low degree vertices. By the Chernoff bound (Fact 2), in this way we just remove a tiny (though linear) fraction of the vertices.

Finally, Step 6 exploits the entries of s_1 and s_2 to compute a partition. The basic insight is that the entries of s_1 and s_2 are essentially constant on the two classes V_1, V_2, and that indeed the entries of s_1 and s_2 differ on each class significantly; this second fact follows from our assumption that the density matrix D has full rank. However, if s_1 and s_2 do not have these properties, then the algorithm will fail to partition the graph correctly and just output a trivial partition.

In order to analyze the algorithm (and thus to prove Theorem 1), we basically need to study the eigenvalues and -vectors of M^*. The main ingredient of the analysis is the following result on the spectrum of the minors $M^*_{V_i \times V_j}$, i.e., the sub-matrices of M^* consisting of the rows V_i and the columns V_j.

Theorem 5. *With high probability the following holds for any two indices $1 \leq i, j \leq 2$.*

1. $$\frac{\mathbf{1}^t}{\|\mathbf{1}^t\|} \cdot {M^*}_{V_i \times V_j} \cdot \frac{\mathbf{1}}{\|\mathbf{1}\|} = \phi_{ij} \cdot W_i \cdot W_j \cdot \frac{\sqrt{|V_i| \cdot |V_j|}}{\overline{w} \cdot n} \cdot \left(1 \pm O\left({d_{\mathrm{m}}}^{-0.49}\right)\right).$$

2. For any u, v with $\|u\| = \|v\| = 1$ and $u \perp \mathbf{1}$ or $v \perp \mathbf{1}$ we have the bound

$$\left|u^t \cdot {M^*}_{V_i \times V_j} \cdot v\right| = O\left(\overline{w}^{-1.49} + {d_{\mathrm{m}}}^{-1.5}\right) = O(1/(\overline{w} \cdot {d_{\mathrm{m}}}^{0.49})).$$

The assumptions of Theorem 1 ensure that the expression on the r.h.s. of 1. is of order $1/\overline{w}$, whereas the expression in 2. is of order $1/(\overline{w} \cdot {d_{\mathrm{m}}}^{0.49})$. Thus, the intuitive meaning of Theorem 5 is that the dominant singular value of $M^*_{V_i \times V_j}$ corresponds approximately to the singular vectors $\mathbf{1}_{V_i}$ and $\mathbf{1}_{V_j}$. By combining the estimates from Theorem 5 for all index pairs $1 \leq i, j \leq 2$, we obtain the following result concerning the eigenvectors of M^*.

Corollary 6. *W.h.p. M^* has exactly two eigenvalues whose absolute value is $\Theta(1/\overline{w})$, whereas all the other eigenvalues are $O\left(1/(\overline{w} \cdot {d_{\mathrm{m}}}^{0.49})\right)$ in absolute value. Moreover, if s_1, s_2 are orthogonal eigenvectors of norm $\sqrt{n}$ with the largest two eigenvalues in absolute value, then there is an index $j \in \{1, 2\}$ such that*

$$s_j = \alpha \mathbf{1}_{|V_1} + \beta \mathbf{1}_{|V_2} + \gamma u, \text{ where } u \perp \mathbf{1}_{|V_1}, \mathbf{1}_{|V_2}, \|u\| = \sqrt{n}$$

and $|\alpha - \beta| > \frac{1}{4}$ and $\gamma = O({d_{\mathrm{m}}}^{-0.49})$.

Corollary 6 implies that w.h.p. step 6 of $\mathcal{A}$ will succeed in finding a vector that satisfies (6). Moreover, a simple calculation based on the above eigenvalue bounds shows that the number of falsely classified vertices (i.e., the symmetric difference of the partitions (V_1', V_2') and (V_1, V_2)) is at most $O(n/{d_\mathrm{m}}^{0.98})$, whence Theorem 1 follows.

The values of α and β correspond to the c_i in the algorithm. If some vertex classified falsely, its entry in s_j is twisted due to its value in $\gamma \cdot u$. Because of the large distance between α and β, such entries are bounded below by some constant. As $|\gamma| = O({d_\mathrm{m}}^{-0.49})$ the value in u has to be $\Omega({d_\mathrm{m}}^{0.49})$. Since $\|u\| = \sqrt{n}$ we have at most $O(n/{d_\mathrm{m}}^{0.98})$ such entries.

4.3 Proof of Corollary 6

At first we show that M^* has the exactly two eigenvalues whose absolute value is $\Theta(1/\overline{w})$, whereas all the other eigenvalues are $O\left(1/(\overline{w} \cdot {d_\mathrm{m}}^{0.49})\right)$ in absolute value. Let g, h be two vectors from the space spanned by $\mathbf{1}_{|V_1}$ and $\mathbf{1}_{|V_2}$. Namely, $g = a_1 \cdot \mathbf{1}_{|V_1}/\|\mathbf{1}_{|V_1}\| + a_2 \cdot \mathbf{1}_{|V_2}/\|\mathbf{1}_{|V_2}\|$ with $a_1^2 + a_2^2 = 1$ and $h = b_1 \cdot \mathbf{1}_{|V_1}/\|\mathbf{1}_{|V_1}\| + b_2 \cdot \mathbf{1}_{|V_2}/\|\mathbf{1}_{|V_2}\|$ with $b_1^2 + b_2^2 = 1$. Note, $\|g\| = \|h\| = 1$. By Theorem 5 we have with probability $1 - o(1)$ that

$$\begin{aligned}
h^t M^* g &= \sum_{i,j=1}^{2} b_i \cdot \frac{\mathbf{1}_{|V_i}}{\|\mathbf{1}_{|V_i}\|} \cdot M^* \cdot a_j \cdot \frac{\mathbf{1}_{|V_j}}{\|\mathbf{1}_{|V_j}\|} = \sum_{i,j=1}^{2} b_i \cdot a_j \cdot \frac{\mathbf{1}^t \cdot {M^*}_{V_i \times V_j} \cdot \mathbf{1}}{\sqrt{|V_i| \cdot |V_j|}} \\
&= \sum_{i,j=1}^{2} b_i \cdot a_j \cdot \phi_{ij} \cdot W_i \cdot W_j \cdot \frac{\sqrt{|V_i| \cdot |V_j|}}{\overline{w} \cdot n} \cdot \left(1 \pm O\left({d_\mathrm{m}}^{-0.49}\right)\right) \\
&= \sum_{i,j=1}^{2} \left(b_i \cdot a_j \cdot \phi_{ij} \cdot W_i \cdot W_j \cdot \frac{\sqrt{|V_i| \cdot |V_j|}}{\overline{w} \cdot n} \right) \pm O\left(1/\left(\overline{w} \cdot {d_\mathrm{m}}^{0.49}\right)\right) \\
&= \frac{1}{\overline{w}} \cdot \begin{pmatrix} b_1 & b_2 \end{pmatrix} \cdot P \cdot \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \pm O\left(1/\left(\overline{w} \cdot {d_\mathrm{m}}^{0.49}\right)\right)
\end{aligned}$$

with

$$P = \begin{pmatrix} W_1 \cdot \sqrt{\frac{|V_1|}{n}} & 0 \\ 0 & W_2 \cdot \sqrt{\frac{|V_2|}{n}} \end{pmatrix} \cdot \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{12} & \phi_{22} \end{pmatrix} \cdot \begin{pmatrix} W_1 \cdot \sqrt{\frac{|V_1|}{n}} & 0 \\ 0 & W_2 \cdot \sqrt{\frac{|V_2|}{n}} \end{pmatrix}.$$

Remember, Φ has full rank as well as both remaining factors of P. We conclude that the matrix P has full rank. The W_i are $\Theta(1)$ as $|V_i|/n$, too. This shows that the spectral properties of P are determined only by Φ, ε and δ and do not rely on $w_1, \ldots, w_n$ or n. P has two eigenvectors with *constant* nonzero eigenvalues. Let $\begin{pmatrix} e_1 & e_2 \end{pmatrix}^t$ and $\begin{pmatrix} f_1 & f_2 \end{pmatrix}^t$ be two orthonormal eigenvectors of P to

the eigenvalues λ_1 and λ_2. Set

$$g_1 = e_1 \cdot \frac{\mathbf{1}_{|V_1}}{\|\mathbf{1}_{|V_1}\|} + e_2 \cdot \frac{\mathbf{1}_{|V_2}}{\|\mathbf{1}_{|V_2}\|} \qquad \text{and} \qquad g_2 = f_1 \cdot \frac{\mathbf{1}_{|V_1}}{\|\mathbf{1}_{|V_1}\|} + f_2 \cdot \frac{\mathbf{1}_{|V_2}}{\|\mathbf{1}_{|V_2}\|}.$$

By the calculation above get

$$\begin{aligned}\left|g_1^t \cdot M^* \cdot g_1\right| &= \left|\frac{1}{\overline{w}} \cdot \begin{pmatrix}e_1 & e_2\end{pmatrix} \cdot P \cdot \begin{pmatrix}e_1\\ e_2\end{pmatrix} \pm O\left(1/\left(\overline{w} \cdot {d_{\mathrm{m}}}^{0.49}\right)\right)\right| \\ &= \left|\frac{1}{\overline{w}} \cdot \lambda_1 \pm O\left(1/\left(\overline{w} \cdot {d_{\mathrm{m}}}^{0.49}\right)\right)\right| = \Theta(1/\overline{w})\end{aligned}$$

whereas

$$\begin{aligned}\left|g_1^t \cdot M^* \cdot g_2\right| &= \left|\frac{1}{\overline{w}} \cdot \begin{pmatrix}e_1 & e_2\end{pmatrix} \cdot P \cdot \begin{pmatrix}f_1\\ f_2\end{pmatrix} \pm O\left(1/\left(\overline{w} \cdot {d_{\mathrm{m}}}^{0.49}\right)\right)\right| \\ &= \left|\frac{1}{\overline{w}} \cdot 0 \pm O\left(1/\left(\overline{w} \cdot {d_{\mathrm{m}}}^{0.49}\right)\right)\right|\end{aligned}$$

Thus for $1 \leq i, j \leq 2$ we have

$$\left|g_i^t \cdot M^* \cdot g_j\right| = \begin{cases}\Theta(1/\overline{w}) & \text{for } i = j\\ O\left(1/\left(\overline{w} \cdot {d_{\mathrm{m}}}^{0.49}\right)\right) & \text{for } i \neq j\end{cases}. \tag{7}$$

For any unit-vector $u \perp g_1, g_2$ (what equals $u \perp \mathbf{1}_{|V_1}, \mathbf{1}_{|V_2}$) we have by Theorem 5 for all unit-vectors v

$$\left|u^t \cdot M^* \cdot v\right| \leq \sum_{i,j=1}^{2} \left|u_{V_i}^t \cdot {M^*}_{V_i \times V_j} \cdot v_{V_j}\right| = O\left(1/\left(\overline{w} \cdot {d_{\mathrm{m}}}^{0.49}\right)\right)$$

and analogously

$$\left|v^t \cdot M^* \cdot u\right| = O\left(1/\left(\overline{w} \cdot {d_{\mathrm{m}}}^{0.49}\right)\right).$$

Both bounds and (7) together with the Courant-Fischer-characterization of eigenvalues yield the first part of the claim.

We are left to show that M^* w.h.p. has an eigenvector s_j as desired. Let e be an eigenvector of M^* with norm $\|e\| = \sqrt{n}$ to the eigenvalue $\Theta(1/\overline{w})$ (in absolute value). We can decompose e such that $e = \alpha \cdot \mathbf{1}_{|V_1} + \beta \cdot \mathbf{1}_{|V_2} + \gamma \cdot u$ for some $u \perp \mathbf{1}_{|V_1}, \mathbf{1}_{|V_2}$ with $\|u\| = \sqrt{n}$. By Theorem 5 we conclude on the one hand

$$\left|e^t \cdot M^* \cdot u\right| = \|e\| \cdot \|u\| \cdot O\left(1/\left(\overline{w} \cdot {d_{\mathrm{m}}}^{0.49}\right)\right) = O\left(n/\left(\overline{w} \cdot {d_{\mathrm{m}}}^{0.49}\right)\right)$$

as $u \perp \mathbf{1}_{|V_1}, \mathbf{1}_{|V_2}$. Because of $e^t \cdot M^* = \Theta(1/\overline{w}) \cdot e^t$ we have on the other hand

$$\left|e^t \cdot M^* \cdot u\right| = \Theta(1/\overline{w}) \cdot \left|e^t \cdot u\right| = \Theta(1/\overline{w}) \cdot |\gamma| \cdot u^t u = \Theta(1/\overline{w}) \cdot |\gamma| \cdot n,$$

so that $|\gamma| = O({d_{\mathrm{m}}}^{-0.49})$. Let s_1, s_2 be as in the lemma and

$$s_j = \alpha_j \cdot \mathbf{1}_{|V_1} + \beta_j \cdot \mathbf{1}_{|V_2} + \gamma_j \cdot u_j$$

the decomposition with $u_j \perp \mathbf{1}_{|V_1}, \mathbf{1}_{|V_2}$ and $\|u_j\| = \sqrt{n}$ as described. Assume for a contradiction that we have $|\alpha_j - \beta_j| \leq 1/4$ for both $j = 1, 2$. As

$$n = s_j^t \cdot s_j = \alpha_j^2 \cdot |V_1| + \beta_j^2 \cdot |V_2| + \gamma_j^2 \cdot n$$

we get

$$\alpha_j^2 + \beta_j^2 \geq \alpha_j^2 \cdot \frac{|V_1|}{n} + \beta_j^2 \cdot \frac{|V_2|}{n} = 1 - \gamma_j^2 \geq 1 - O({d_{\mathrm{m}}}^{-0.98}) \,.$$

Clearly, for both $j = 1, 2$ we have $|\alpha_j| > 1/2$ or $|\beta_j| > 1/2$, yielding that the sign of α_j equals the sign of β_j for both $j = 1, 2$. We get

$$|\alpha_1 \cdot \alpha_2 + \beta_1 \cdot \beta_2| = |\alpha_1 \cdot \alpha_2| + |\beta_1 \cdot \beta_2| \geq \frac{1}{2} \cdot \frac{1}{4} + \frac{1}{4} \cdot \frac{1}{2} = \frac{1}{4}$$

and

$$\begin{aligned} 0 = s_1^t \cdot s_2 &= \left|\alpha_1 \cdot \alpha_2 \cdot |V_1| + \beta_1 \cdot \beta_2 \cdot |V_2| + \gamma_1 \cdot \gamma_2 \cdot u_1^t \cdot u_2\right| \\ &\geq \delta n \cdot |\alpha_1 \cdot \alpha_2 + \beta_1 \cdot \beta_2| - |\gamma_1 \cdot \gamma_2| \cdot n \geq \delta n/4 - O\left(n/{d_{\mathrm{m}}}^{0.98}\right) \,. \end{aligned}$$

This is a contradiction since $\delta > 0$ is constant and d_{m} is large. So at least one s_j has $|\alpha_j - \beta_j| > 1/4$. □

4.4 Proof of Theorem 5: The spectrum of $M^*_{V_i \times V_j}$

The main difficulty in the (rather involved) proof of Theorem 5 is the fact that the entries of M^* are mutually dependent, because we normalize by the actual vertex degrees (cf. Step 2 of the algorithm and (4)). Furthermore, in case of sparse graphs (which is included in Theorem 1), it is possible that all (or most) weights w_u remain bounded as $n \to \infty$. In this case the expected degrees are bounded as well. In effect, the actual degrees of the vertices are *not* concentrated about their expectations, but may deviate by up to $\Omega(\log n / \log\log n)$. Hence, we need to cope with the dependence of the matrix entries as well as with deviations of the vertex degrees from their expectations.

To this end, we mark vertices $u \in V_i$ as "bad" if the number of u's neighbors in V_j is far from its expectation (of course, this is just a part of the *analysis* – the algorithm cannot identify these "bad" vertices). Similarly, we mark vertices from V_j as "bad". Now, it is possible that some "good" vertices inside V_i and/or V_j have many "bad" neighbors. We mark such vertices as "bad", too. Repeating this process, we obtain a subset $R_{ij} \subseteq V_i$ of "good" vertices, which firstly have

about as many neighbors in V_j as expected and secondly have only a few "bad" neighbors in V_j. Analogously we obtain "good" vertices $C_{ij} \subseteq V_j$. Then, we shall analyze the sub-matrix induced on $R_{ij} \times C_{ij}$ separately from the rest.

More precisely, the sets $R_{ij} \subseteq V_i$ and $C_{ij} \subseteq V_j$ are the outcome of the following process. Let c be a sufficiently large constant (the value gets determined later), and let $A = A(G)$ be the adjacency matrix of G.

1. Let $R' = \{u \in V : \forall j' : |s_A(u, V_{j'}) - \mathrm{Vol}(u, V_{j'})| \leq \mathrm{Vol}(u, V_{j'})^{0.51}\}$.
2. Let $C' = \{v \in V : \forall i' : |s_A(V_{i'}, v) - \mathrm{Vol}(V_{i'}, v)| \leq \mathrm{Vol}(V_{i'}, v)^{0.51}\}$.
3. Set $R'_{ij} := R' \cap V_i$ and $C'_{ij} := C' \cap V_j$.
4. While there is some $u \in R'_{ij}$ with
$$s_A(u, V_j \setminus C'_{ij}) \geq \mathrm{Vol}(u, V_j) \cdot c/d_{\mathrm{m}} \qquad \text{then} \quad R'_{ij} := R'_{ij} \setminus \{u\}.$$
5. While there is some $v \in C'_{ij}$ with
$$s_A(V_i \setminus R'_{ij}, v) \geq \mathrm{Vol}(V_i, v) \cdot c/d_{\mathrm{m}} \qquad \text{then} \quad C'_{ij} := C'_{ij} \setminus \{v\}.$$
6. Repeat Steps 4 – 5 until R'_{ij} and C'_{ij} remain unchanged.
7. $R_{ij} := R'_{ij}$. $C_{ij} := C'_{ij}$.

We abbreviate R_{ij} by $\mathcal{R}$ and C_{ij} by $\mathcal{C}$, $V_i \setminus R_{ij}$ by $\overline{\mathcal{R}}$, and $V_j \setminus C_{ij}$ by $\overline{\mathcal{C}}$. Due to the first step of the above process all $u \in \mathcal{R}$ and $v \in \mathcal{C}$ satisfy

$$\begin{aligned} |s_A(u, V) - \mathrm{Vol}(u, V)| &\leq 2 \cdot \mathrm{Vol}(u, V)^{0.51}, \\ |s_A(V, v) - \mathrm{Vol}(V, v)| &\leq 2 \cdot \mathrm{Vol}(V, v)^{0.51}. \end{aligned} \tag{8}$$

Let us briefly discuss the above process. For a vertex $u \in V_1$ the standard deviation of the number $s_A(u, V_j)$ of neighbors of u in V_j from its expectation $\mathrm{Vol}(u, V_j)$ is of order $O(\mathrm{Vol}(u, V_j)^{0.5})$ (because $s_A(u, V_j)$ is a sum of independent 0/1-random variables). Therefore, the Chernoff bound (Fact 2) entails that w.h.p. "most" of the vertices in V_i belong to R'. Moreover, the larger $\mathrm{Vol}(u, V_j)$, the more likely it is that $u \in R'$. Hence, we expect $\mathrm{Vol}(V_i \setminus R', V_j)$ (as well as $\mathrm{Vol}(V_i, V_j \setminus C')$) to be fairly small. Consequently, as a vertex removed from R'_{ij} in Step 4 has relatively many neighbors inside the set $V_j \setminus C'_{ij}$ of small volume, we expect that Step 4 will remove only a small number of vertices. Thus, the final sets $\mathcal{R}$ and $\mathcal{C}$ should constitute the dominant fraction of the volume of G. The following lemma, whose proof is omitted, shows that this is actually the case.

Lemma 7. *W.h.p. we have* $\mathrm{Vol}(\overline{\mathcal{R}}, V_j) \leq n/{d_{\mathrm{m}}}^4$, $\mathrm{Vol}(V_i, \overline{\mathcal{C}}) \leq n/{d_{\mathrm{m}}}^4$, *and* $\mathrm{Vol}(\overline{\mathcal{R}}, \overline{\mathcal{C}}) \leq n/{d_{\mathrm{m}}}^8$.

A consequence of Lemma 7 is that both $\overline{\mathcal{R}}$ and $\overline{\mathcal{C}}$ contain only a few vertices. For by the choice of d_{m} (cf. Step 1 of $\mathcal{A}$) for all $u \in V_i$ and all $v \in V_j$ we have

$$d_{\mathrm{m}} \leq \mathrm{Vol}(u, V_j) \leq \mathrm{Vol}(u, V) = w'_u \qquad \text{and} \qquad d_{\mathrm{m}} \leq \mathrm{Vol}(V_i, v) \leq w'_v\,. \tag{9}$$

Thus, $d_{\mathrm{m}} \cdot \left|\overline{\mathcal{R}}\right| \leq \mathrm{Vol}(\overline{\mathcal{R}}, V_j) \leq n/{d_{\mathrm{m}}}^4$, which yields $\left|\overline{\mathcal{R}}\right| \leq n/{d_{\mathrm{m}}}^5$. As $\delta \cdot n \leq |V_i|$, we get

$$\left|\overline{\mathcal{R}}\right| \leq \frac{|V_i|}{\delta \cdot {d_{\mathrm{m}}}^5} \leq \frac{|V_i|}{{d_{\mathrm{m}}}^4} \qquad \text{and} \qquad |\mathcal{R}| = |V_i| - \left|\overline{\mathcal{R}}\right| \geq |V_i| \cdot \left(1 - \frac{1}{{d_{\mathrm{m}}}^4}\right), \quad (10)$$

(provided that $\overline{w} > 1/\delta^2$ is sufficiently large). Analogously,

$$\left|\overline{\mathcal{C}}\right| \leq |V_i| / {d_{\mathrm{m}}}^4 \qquad \text{and} \qquad |\mathcal{C}| \geq |V_i| \cdot \left(1 - 1/{d_{\mathrm{m}}}^4\right). \quad (11)$$

To proceed, we subdivide ${M^*}_{V_i \times V_j}$ into four parts ${M^*}_{\overline{\mathcal{R}} \times \mathcal{C}}$, ${M^*}_{\mathcal{R} \times \overline{\mathcal{C}}}$, ${M^*}_{\mathcal{R} \times \mathcal{C}}$, and ${M^*}_{\overline{\mathcal{R}} \times \overline{\mathcal{C}}}$, which we shall analyze separately. With respect to ${M^*}_{\mathcal{R} \times \mathcal{C}}$, we have the following.

Lemma 8. *With high probability we have*

1. $\mathbf{1}^t \cdot {M^*}_{\mathcal{R} \times \mathcal{C}} \cdot \mathbf{1} = \phi_{ij} \cdot W_i \cdot W_j \cdot \dfrac{|\mathcal{R}| \cdot |\mathcal{C}|}{\overline{w} \cdot n} \cdot \left(1 \pm O(1/{d_{\mathrm{m}}}^{0.49})\right) = \Theta(n/\overline{w})$,
2. $|u^t \cdot {M^*}_{\mathcal{R} \times \mathcal{C}} \cdot v| = O\left(1/\overline{w}^{1.49}\right)$ *for any* u, v *with* $\|u\| = \|v\| = 1$ *and* $u \perp \mathbf{1}$ *or* $v \perp \mathbf{1}$, *and*
3. $\|{M^*}_{\mathcal{R} \times \mathcal{C}}\| = \Theta\left(1/\overline{w}\right)$.

The proof of Lemma 8 is based on the fact that on $\mathcal{R} \times \mathcal{C}$ the vertex degrees behave at least roughly as expected. Therefore, we can relate the spectrum of ${M^*}_{\mathcal{R} \times \mathcal{C}}$ to the spectrum of $\mathbf{M}_{\mathcal{R} \times \mathcal{C}}$, where $\mathbf{M}$ is the matrix from (3). Since the entries of $\mathbf{M}$ are mutually independent (up to the trivial dependence resulting from symmetry), the analysis of its spectrum is significantly simpler than the analysis of M; in fact, this analysis has been carried out in [9]. Nonetheless, in order to relate $\mathbf{M}_{\mathcal{R} \times \mathcal{C}}$ and ${M^*}_{\mathcal{R} \times \mathcal{C}}$, we need to analyze the degree distribution of G thoroughly, which requires considerable technical work (omitted).

As a next step, we analyze the three "small" blocks ${M^*}_{\overline{\mathcal{R}} \times \mathcal{C}}$, ${M^*}_{\mathcal{R} \times \overline{\mathcal{C}}}$ and ${M^*}_{\overline{\mathcal{R}} \times \overline{\mathcal{C}}}$.

Lemma 9. *With high probability we have that* $\|{M^*}_{\mathcal{R} \times \overline{\mathcal{C}}}\|$, $\|{M^*}_{\overline{\mathcal{R}} \times \mathcal{C}}\|$ *and* $\|{M^*}_{\overline{\mathcal{R}} \times \overline{\mathcal{C}}}\|$ *are* $O({d_{\mathrm{m}}}^{-1.5})$.

The proof of Lemma 9 is based on combinatorial ideas, and, in particular, the fact that the volumes of $\overline{\mathcal{R}}$ and $\overline{\mathcal{C}}$ are relatively small (cf. Lemma 7). Therefore, for instance the subgraph induced on $\overline{\mathcal{R}} \times \overline{\mathcal{C}}$ has a very simple combinatorial structure (it is essentially forest-like), which allows a direct analysis of ${M^*}_{\overline{\mathcal{R}} \times \overline{\mathcal{C}}}$. Details are omitted.

4.4.1 Proof of Theorem 5. With respect to the first statement, we have

$$\mathbf{1}^t \cdot {M^*}_{V_i \times V_j} \cdot \mathbf{1} = \mathbf{1}^t \cdot {M^*}_{\mathcal{R} \times \mathcal{C}} \cdot \mathbf{1} + \mathbf{1}^t \cdot {M^*}_{\mathcal{R} \times \overline{\mathcal{C}}} \cdot \mathbf{1} + \mathbf{1}^t \cdot {M^*}_{\overline{\mathcal{R}} \times \mathcal{C}} \cdot \mathbf{1} + \mathbf{1}^t \cdot {M^*}_{\overline{\mathcal{R}} \times \overline{\mathcal{C}}} \cdot \mathbf{1}. \quad (12)$$

Item 1. of Lemma 8 gives for the first term

$$\begin{aligned}\mathbf{1}^t \cdot {M^*}_{\mathcal{R}\times\mathcal{C}} \cdot \mathbf{1} &= \phi_{ij} \cdot W_i \cdot W_j \cdot \frac{|\mathcal{R}| \cdot |\mathcal{C}|}{\overline{w} \cdot n} \cdot (1 \pm O(1/{d_{\mathrm{m}}}^{0.49})) \\ &\overset{(10),(11)}{=} \phi_{ij} \cdot W_i \cdot W_j \cdot \frac{|V_i| \cdot |V_j|}{\overline{w} \cdot n} \cdot (1 \pm O(1/{d_{\mathrm{m}}}^{0.49}))\,.\end{aligned}$$

Lemma 9 shows that the second summand in (12) is bounded by

$$\begin{aligned}\left|\mathbf{1}^t \cdot {M^*}_{\mathcal{R}\times\overline{\mathcal{C}}} \cdot \mathbf{1}\right| &\leq \sqrt{|\mathcal{R}| \cdot \left|\overline{\mathcal{C}}\right|} \cdot \|{M^*}_{\mathcal{R}\times\overline{\mathcal{C}}}\| \overset{(11)}{\leq} \sqrt{|V_i| \cdot |V_j| \,/ {d_{\mathrm{m}}}^4} \cdot O({d_{\mathrm{m}}}^{-1.5}) \\ &= \sqrt{|V_i| \cdot |V_j|} \cdot O\left({d_{\mathrm{m}}}^{-3.5}\right) = \sqrt{|V_i| \cdot |V_j|} \cdot O(1/(\overline{w} \cdot {d_{\mathrm{m}}}^{0.49}))\,.\end{aligned}$$

The same bound holds for both $\left|\mathbf{1}^t \cdot {M^*}_{\overline{\mathcal{R}}\times\mathcal{C}} \cdot \mathbf{1}\right|$ and $\left|\mathbf{1}^t \cdot {M^*}_{\overline{\mathcal{R}}\times\overline{\mathcal{C}}} \cdot \mathbf{1}\right|$. Dividing each summand for (12) by $\sqrt{|V_i| \cdot |V_j|}$ we get the desired bound on $\dfrac{\mathbf{1}^t}{\|\mathbf{1}^t\|} \cdot {M^*}_{V_i\times V_j} \cdot \dfrac{\mathbf{1}}{\|\mathbf{1}\|}$.

For the second item of Theorem 5 we assume that $u \perp \mathbf{1}$, yielding $u^t \cdot (\mathbf{1}_{|\mathcal{R}} + \mathbf{1}_{|\overline{\mathcal{R}}}) = 0$, so that

$$\left|u^t \cdot \mathbf{1}_{|\mathcal{R}}\right| = \left|u^t \cdot \mathbf{1}_{|\overline{\mathcal{R}}}\right| \leq \|u\| \cdot \|\mathbf{1}_{|\overline{\mathcal{R}}}\| \leq \sqrt{\left|\overline{\mathcal{R}}\right|}\,. \tag{13}$$

We decompose u as $u = a \cdot \mathbf{1}_{|\mathcal{R}} / \|\mathbf{1}_{|\mathcal{R}}\| + b \cdot u_l$ with $\|u_l\| = 1$ and $u_l \perp \mathbf{1}_{|\mathcal{R}}$. Clearly $u_{l|\mathcal{R}} \perp \mathbf{1}_{|\mathcal{R}}$, too, and $a^2 + b^2 = 1$. A straightforward computation yields

$$|a| = \left| u^t \cdot \frac{\mathbf{1}_{|\mathcal{R}}}{\|\mathbf{1}_{|\mathcal{R}}\|} \right| \overset{(13)}{\leq} \frac{\sqrt{\left|\overline{\mathcal{R}}\right|}}{\|\mathbf{1}_{|\mathcal{R}}\|} \overset{(10)}{<} 2/{d_{\mathrm{m}}}^2\,. \tag{14}$$

Let v be some arbitrary unit-vector. Then we can rewrite $\left|u^t \cdot {M^*}_{V_i\times V_j} \cdot v\right|$ as

$$\left|u^t \cdot {M^*}_{V_i\times V_j} \cdot \left(v_{|\mathcal{C}} + v_{|\overline{\mathcal{C}}}\right)\right| \leq \left|u^t \cdot {M^*}_{V_i\times V_j} \cdot v_{|\mathcal{C}}\right| + \|{M^*}_{\mathcal{R}\times\overline{\mathcal{C}}}\| + \|{M^*}_{\overline{\mathcal{R}}\times\overline{\mathcal{C}}}\|\,.$$

The second and the third summand are $O({d_{\mathrm{m}}}^{-1.5})$ by Lemma 9. The first one we bound as follows

$$\begin{aligned}\left|u^t \cdot {M^*}_{V_i\times V_j} \cdot v_{|\mathcal{C}}\right| &= \left|\left(a \cdot \frac{{\mathbf{1}^t}_{|\mathcal{R}}}{\|{\mathbf{1}^t}_{|\mathcal{R}}\|} + b \cdot u_l\right) \cdot {M^*}_{V_i\times\mathcal{C}} \cdot v_{\mathcal{C}}\right| \\ &\leq |a| \cdot \|{M^*}_{\mathcal{R}\times\mathcal{C}}\| + |(b \cdot u_l) \cdot {M^*}_{V_i\times\mathcal{C}} \cdot v_{\mathcal{C}}| \\ &\overset{(14)}{<} 2/{d_{\mathrm{m}}}^2 \cdot O(1/\overline{w}) + \left|b \cdot \left(u_{l|\mathcal{R}} + u_{l|\overline{\mathcal{R}}}\right) \cdot {M^*}_{V_i\times\mathcal{C}} \cdot v_{\mathcal{C}}\right| \\ &\leq O\left({d_{\mathrm{m}}}^{-1.5}\right) + |u_{l\mathcal{R}} \cdot {M^*}_{\mathcal{R}\times\mathcal{C}} \cdot v_{\mathcal{C}}| + \|{M^*}_{\overline{\mathcal{R}}\times\mathcal{C}}\| \\ &= O\left({d_{\mathrm{m}}}^{-1.5}\right) + O\left({\overline{w}}^{-1.49}\right) + O\left({d_{\mathrm{m}}}^{-1.5}\right)\,.\end{aligned}$$

We got the last step because of $u_{l|\mathcal{R}} \perp \mathbf{1}_{|\mathcal{R}}$ and Lemma 8. So, $\left|u^t \cdot {M^*}_{V_i \times V_j} \cdot v\right|$ is $O\left(\overline{w}^{-1.49}\right) + O\left({d_{\mathrm{m}}}^{-1.5}\right)$ as desired. The case $v \perp \mathbf{1}$ and u arbitrary can be handled analogously. □

References

1. Aiello, W., Chung, F., Lu, L.: A random graph model for massive graphs. Proc. 33rd. SToC (2001), 171–180.
2. Alon, N. Spectral techniques in graph algorithms. Proc. LATIN (1998), LNCS 1380, Springer, 206–215.
3. Alon, N., Kahale, N.: A spectral technique for coloring random 3-colorable graphs. SIAM J. Comput. **26** (1997) 1733–1748.
4. Azar, Y., Fiat, A., Karlin, A.R., McSherry, F., Saia, J.: Spectral analysis of data. Proc. 33rd STOC (2001) 619–626
5. Boppana, R.B.: Eigenvalues and graph bisection: An average case analysis. Proc. 28th FoCS (1987), 280–285.
6. Chung, F.K.R.: Spectral Graph Theory. American Mathematical Society (1997).
7. Chung, F.K.R., Lu, L.: Connected components in random graphs with given expected degree sequences. Annals of Combinatorics **6** (2002) 125–145.
8. Coja-Oghlan, A., Lanka, A.: The Spectral Gap of Random Graphs with Given Expected Degrees. Proc. ICALP (2006), LNCS 4051, Springer, 15–26.
9. Coja-Oghlan, A., Goerdt, A., Lanka, A.: Spectral Partitioning of Random Graphs with Given Expected Degrees. Preprint, submitted for publication. Available at http://www.tu-chemnitz.de/informatik/TI/publications/spj10.pdf
10. Ding. C.H.Q.: Analysis of gene expression profiles: class discovery and leaf ordering. Proc. 6th International Conference on Computational biology (2002) 127–136
11. Dasgupta, A., Hopcroft, J.E., McSherry, F.: Spectral Analysis of Random Graphs with Skewed Degree Distributions. Proc. 45th FOCS (2004) 602–610.
12. Flaxman, A.: A spectral technique for random satisfiable 3CNF formulas. Proc. 14th SODA (2003) 357–363.
13. Füredi, Z., Komloś, J.: The eigenvalues of random symmetric matrices. Combinatorica **1** (1981) 233–241.
14. Janson, S., Łuczak, T., Ruciński, A.: Random graphs. John Wiley and Sons 2000.
15. McSherry, F.: Spectral Partitioning of Random Graphs. Proc. 42nd FoCS (2001) 529–537.
16. Mihail, M., Papadimitriou, C.H.: On the Eigenvalue Power Law. Proc. 6th RANDOM (2002) 254–262.
17. Pothen, A., Simon, H.D., Liou, K.-P.: Partitioning sparse matrices with eigenvectors of graphs. SIAM J. Matrix Anal. Appl. **11** (1990) 430–452
18. Schloegel, K., Karypis, G., Kumar, V.: Graph partitioning for high performance scientific simulations. in: Dongarra, J., Foster, I., Fox, G., Kennedy, K., White, A. (eds.): CRPC parallel computation handbook. Morgan Kaufmann (2000)