Electronic Colloquium on Computational Complexity, Report No. 116 (2006)

Graph Partitioning via Adaptive Spectral Techniques

Amin Coja-Oghlan*

Humboldt-Universität zu Berlin, Institut für Informatik,
Unter den Linden 6, 10099 Berlin, Germany
coja@informatik.hu-berlin.de

Abstract. In this paper we study the use of spectral techniques for graph partitioning. Let $G = (V, E)$ be a graph whose vertex set has a "latent" partition $V_1, \ldots, V_k$. Moreover, consider a "density matrix" $\mathcal{E} = (\mathcal{E}_{vw})_{v,w \in V}$ such that for $v \in V_i$ and $w \in V_j$ the entry $\mathcal{E}_{vw}$ is the fraction of all possible V_i-V_j-edges that are actually present in G. We show that on input (G, k) the partition $V_1, \ldots, V_k$ can (almost) be recovered in polynomial time via spectral methods, provided that the following holds: $\mathcal{E}$ approximates the adjacency matrix of G in the operator norm, for vertices $v \in V_i$, $w \in V_j \neq V_i$ the corresponding column vectors $\mathcal{E}_v$, $\mathcal{E}_w$ are separated, and G is sufficiently "regular" w.r.t. the matrix $\mathcal{E}$. This result in particular applies to *sparse* graphs with bounded average degree as $n = \#V \to \infty$, and it yields interesting consequences on partitioning random graphs.
Keywords: graph partitioning, spectral methods, random graphs.

1 Introduction and Results

1.1 Spectral Techniques for Graph Partitioning

To solve various types of graph partitioning problems, *spectral heuristics* are in common use. Such heuristics represent a given graph by a matrix and compute its eigenvalues and -vectors to solve the combinatorial problem in question. Spectral techniques are used either to deal 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. In the present paper we mainly deal with the latter problem, which is of relevance, e.g., in information retrieval [3], scientific simulation [29], or bioinformatics [14].

Despite their success in applications (e.g., [28, 29]), 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 important research problem. To address this problem, quite a few authors have contributed rigorous analyses of spectral techniques on suitable models of *random graphs*. For example, Alon and Kahale [1] analyzed a spectral technique for GRAPH COLORING, Alon, Krivelevich, Sudakov [2] dealt with the MAXIMUM CLIQUE problem, and Boppana [6] and Coja-Oghlan [10] studied random instances of MINIMUM BISECTION. In addition, Flaxman [18] suggested a spectral technique for random 3-SAT.

While the algorithmic techniques of [1, 2, 6, 10, 18] are really tailored for the concrete problems (and random graph models) studied in the respective articles, a remarkable paper of McSherry [27] investigates a more generic spectral partitioning algorithm on a rather general random graph model. McSherry's result comprises the main results from [2, 6], but does not encompass *sparse* random graphs as studied in [1, 10], or graphs in which edges do not occur independently as in [18].

The goal of the present work is to devise a new, generic spectral heuristic that does capture all the previous work [1, 2, 6, 10, 18, 27], and that is indeed applicable to much more general settings. To this end, we shall *not* stick to a specific random graph model (howsoever general this model may be). Instead, we single out as modest conditions as possible that ensure the success of the spectral algorithm. In order to come up with such conditions, let us observe (informally) the most important features that the random graphs in prior work have in common. Let $G = (V, E)$ be a graph whose vertex set has a "latent partition" $V_1, \ldots, V_k$; we think of k being "small" in comparison to $n = \#V$. For $v \in V$ and $1 \leq i \leq k$ we let $e(v, V_i)$ signify the number of v-V_i-edges in G.

* Date: July 4, 2006.

ISSN 1433-8092

Low rank structure. Define a matrix $\mathcal{E} = (\mathcal{E}_{vw})_{v,w\in V}$ of rank $\leq k$ as follows: if $v \in V_i$ and $w \in V_j$, then $\mathcal{E}_{vw}$ is the fraction of all possible V_i-V_j-edges that are actually present in G. Let $A(G)$ be the adjacency matrix of G. Then in [2, 6, 27] the norm $\|A(G) - \mathcal{E}\|$ is "small". By contrast, this is not exactly true in the *sparse* graphs occurring in [1, 10, 18]. Nonetheless, in [1, 10, 18] the graph G' obtained by removing a small number of vertices of "atypically high degree" is such that $A(G')$ is well approximated by the low rank matrix $\mathcal{E}$.

Separation. The low rank matrix $\mathcal{E}$ mirrors the latent partition of the graph in the sense that for vertices $u, v \in H$ that belong to different classes the norm $\|\mathcal{E}_u - \mathcal{E}_v\|$ should be "large enough"; here $\mathcal{E}_w$ denotes the w-column of $\mathcal{E}$.

Approximate Regularity. For all vertices $v \in V_i$ the number $e(v, V_j)$ is "close" to the average number $\#V_i^{-1} \sum_{w\in V_i} e(w, V_j)$.

Core. The graph G has a "large" subgraph H (which is sometimes called the "core" of G) such that for all $v \in H$ the vector $\mathcal{E}_v$ provides a good description of the "densities" $e(v, V_j)/\#V_j$. More precisely, for $w \in V_j$ we define

$$d(v, w) = e(v, V_j)/\#V_j, \text{ and we let } d(v) = (d(v, w))_{w\in V} \in \mathbf{R}^V. \tag{1}$$

Then $\|d(v) - \mathcal{E}_v\|$ should be "small" for all $v \in H$. While in [2, 6, 27] we have $H = G$, in [1, 10, 18] the core H is actually a proper subgraph of G.

The main result of this paper is a spectral algorithm `Partition` that recovers the "latent" partition of a given graph G, provided that (rigorously formulated versions of) the aforementioned conditions are satisfied (cf. Theorem 1). Hence, the result crystallizes "deterministic" conditions that cause spectral methods to succeed, and may thus contribute to a better understanding of such techniques.

Moreover, the fact that we deal with a general graph partitioning problem requires new, generic algorithmic ideas. For instance, the algorithm `Partition` is *adaptive* in the sense that input of the algorithm *only* consists of the graph G and the desired number of vertex classes k. Thus, the algorithm does not require *any* further information about the type of the partition (e.g., no lower bound on the size of the classes or on the separation of vertices in different classes). Furthermore, also the fact that the present work encompasses, e.g., *sparse* graphs (constant average degree) requires new algorithmic solutions.

1.2 The Main Result

To state the four conditions from Section 1.1 rigorously, we need a bit of notation. Throughout the paper, we let $V = \{1, \ldots, n\}$ be a vertex set, and $G = (V, E)$ denotes a graph. Moreover, let $\psi : V \to \{1, \ldots, k\}$, $V_i = \psi^{-1}(v)$, and $n_{\min} = \min_{1\leq i\leq k} \#V_i$. We think of $V_1, \ldots, V_k$ as the "latent" partition of G that we are to recover. Moreover, consider a symmetric $k \times k$ matrix $\boldsymbol{p} = (p_{ij})_{1\leq i,j\leq k}$; the intended meaning is that p_{ij} should equal the "density" of the pair V_i, V_j, i.e., the fraction of all possible V_i-V_j-edges that are actually present in G. Furthermore, let $\mathcal{E} = \mathcal{E}(\psi, \boldsymbol{p}) = (\mathcal{E}_{vw})_{v,w\in V}$ be the $n \times n$ matrix with entries $\mathcal{E}_{vw} = p_{\psi(v)\psi(w)}$; note that $\mathcal{E}$ has rank $\leq k$.

If we think of p_{ij} as the density of the pair V_i, V_j, then we could interpret the entry $\mathcal{E}_{vw}$ as the "probability" that $v, w \in V$ are connected – even though we are *not* assuming that G is a random graph. Moreover, we could consider the term $\sum_{w\in V} \mathcal{E}_{vw}(1 - \mathcal{E}_{vw})$ the "variance" of the number of neighbors of v (because the variance of a Bernoulli experiment with success probability $\mathcal{E}_{vw}$ is $\mathcal{E}_{vw}(1 - \mathcal{E}_{vw})$). Thus,

$$\sigma^* = \max_{v\in V} \sum_{w\in V} \mathcal{E}_{vw}(1 - \mathcal{E}_{vw}) \tag{2}$$

can be interpreted as the "maximum variance" of the vertex degrees of G.

In addition, we need to partition G into a "sparse" and a "dense" part. To this end, let $\Phi = (\Phi_{vw})_{v,w\in V}$ be the matrix with entries

$$\Phi_{vw} = 1 \text{ if } \mathcal{E}_{vw} > \tfrac{1}{2}, \text{ and } \Phi_{vw} = 0 \text{ otherwise.} \tag{3}$$

Then we define

$$G_1 = (V, E_1), \text{ where } E_1 = \{\{v, w\} \in E : \Phi_{vw} = 0\}, \tag{4}$$
$$G_2 = (V, E_2), \text{ where } E_2 = \{\{v, w\} \notin E : \Phi_{vw} = 1,\ v, w \in V,\ v \neq w\}. \tag{5}$$

Thus, G_1 consists of all edges e that are present in G such that Φ_e is 0, while G_2 contains all e with $\Phi_e = 1$ that are *missing* in G. Let $d_{G_1 \cup G_2}(v)$ denote the degree of v in the graph $G_1 \cup G_2 = (V, E_1 \cup E_2)$.

If $M = (m_{vw})_{v,w \in V}$ is a matrix and $X \subset V$, then we let M_X be the matrix obtained from M by replacing all entries m_{vw} with $(v,w) \notin X \times X$ by 0. With this notation, we can state the "low rank structure" condition as follows.

A1. Let A be the adjacency matrix of G, and let $M = \mathcal{E} - A$. There is a number $\sigma^* \leq \lambda \leq \sigma^* \cdot \min\{\sigma^*, n_{\min}/\ln n\}$ such that for any $\Delta > 0$ the set $D(\Delta) = \{v \in V : d_{G_1 \cup G_2}(v) \leq \Delta\}$ satisfies $\|M_{D(\Delta)}\| \leq c_0 k \sqrt{\lambda + \Delta}$, where $c_0 > 0$ is a constant.

Thus, **A1** states that $\mathcal{E}$ "approximates" A within $c_0 k \sqrt{\lambda + \Delta}$ on the subgraph of G obtained by removing all vertices that have degree $> \Delta$ in $G_1 \cup G_2$. The crucial parameter that measures the quality of the approximation is λ, and thus λ will play an important role in the "separation" condition as well. Moreover, we shall see in Section 2 that the occurrence of Δ in the bound in **A1** is actually necessary.

Letting $\mathcal{E}_v = (\mathcal{E}_{vw})_{w \in V}$ signify the v-column of $\mathcal{E}$, we state the "separation" condition as follows.

A2. Let $\rho = c_0^4 \sqrt{k^3 \lambda / n_{\min}}$. Then for all $u, v \in V$ such that $\psi(u) \neq \psi(v)$ we have $\|\mathcal{E}_u - \mathcal{E}_v\| \geq \rho$.

This condition says how much for vertices u, v that belong to different classes the vectors $\mathcal{E}_u, \mathcal{E}_v$ that represent the "expected densities" should differ. Note the dependence of ρ on λ: the tighter $\mathcal{E}$ approximates A, the more "subtle" the differences between $\mathcal{E}_u$ and $\mathcal{E}_v$ can be.

To state the "approximate regularity" condition, for each vertex $v \in V$ and each set $S \subset V$ we let $e(v, S)$ denote the number of edges from v to S in G. Moreover, if we think of $\mathcal{E}_{vw}$ as the "probability" that v, w are connected in G, then we can consider $\mu(v, S) = \sum_{v \neq w \in S} \mathcal{E}_{vw}$ as the "expected" number of v-S-edges.

A3. All $v \in V_j$ obey the bound $\max_{1 \leq i \leq k} |e(v, V_i) - \mu(v, V_i)| \leq 0.1 \left(\frac{1}{k}\sigma^* + \# V_I P_{ij}(1 - p_{ij})\right) + \ln^2 n$.

Hence, **A3** requires that any vertex v should have approximately the "expected" number $\mu(v, V_i)$ of neighbors in each class. The error term on the r.h.s. involves the maximum variance σ^* and in addition the "variance" $\# V_I P_{ij}(1 - p_{ij}) = \sum_{w \in V_i} \mathcal{E}_{vw}(1 - \mathcal{E}_{vw})$ of the number $e(v, V_i)$. Moreover, the additive $\ln^2 n$-term is crucial in the case of sparse graphs (cf. Section 2).

Further, we need a fourth condition that ensures that all classes V_i have at least polylogarithmic size.

A4. $n_{\min} = \min_{1 \leq i \leq k} \# V_i \geq \ln^{30} n$.

As a next step, we shall formulate the "core" condition rigorously; intuitively the "core" is a subgraph H that consists of "well behaved" vertices.

H1. The subgraph H of G satisfies $\# V \setminus H \leq \lambda^{-4} n_{\min}$ and $\sum_{v \in V \setminus H} d_{G_1 \cup G_2}(v)^2 \leq n_{\min}$.
H2. For all $v \in H$ the vector $d(v)$ defined in (1) satisfies $\|\mathcal{E}_v - d(v)\|^2 \leq 0.001 \rho^2$.
H3. All $v \in H$ have degree $\leq 10\sigma^*$ in the graph $G_1 \cup G_2$.
H4. In the graph $G_1 \cup G_2$ each $v \in H$ has at most 100 neighbors in $V \setminus H$.

Thus, **H1** requires that the core H constitutes a "large" share of G, and that the vertices outside of H are not incident with an exorbitant number of edges. Furthermore, by **H2** for all $v \in H$ the vector $d(v)$ should be close to $\mathcal{E}_v$. In addition, **H3** requires that the vertices $v \in H$ do not have a too high degree in $G_1 \cup G_2$, and **H4** means that H should be "well separated" from $V \setminus H$.

Theorem 1. *There are a polynomial time algorithm* `Partition` *and a constant $C > 0$ such that for each $c_0 > C$ and each integer $k \geq 2$ there exists a number n_0 so that the following is true. Suppose that $n \geq n_0$ and $\sigma^* \geq c_0$, that* **A1–A4** *hold, and that H is a subgraph of G that satisfies* **H1–H4**. *Then* `Partition`(G, k) *outputs a partition $(T_1, \ldots, T_k)$ of V such that $T_i \cap H = V_{\tau(i)} \cap H$ for some permutation τ of $\{1, \ldots, k\}$.*

Hence, **A1–A4** and **H1–H4** ensure that `Partition` can recover the planted partition $V_1, \ldots, V_k$ *on the subgraph* H. However, `Partition` cannot recover the *entire* partition $V_1, \ldots, V_k$ in general. In fact, as we shall see in Section 2, recovering the partition $V_1, \ldots, V_k$ perfectly is *impossible* in general; the reason basically is that we only assume that $d(v)$ is close to $\mathcal{E}_v$ for $v \in H$. Thus, loosely speaking Theorem 1 says that if G has a "nice" low rank structure $\mathcal{E}$, then we can recover a large piece of $\mathcal{E}$ in polynomial time.

Furthermore, we emphasize that the input of `Partition` *only* consists of the graph G and the desired number k of classes; *no* other parameters of the partition (e.g., $\mathcal{E}$, ρ, $n_{\min}$) are revealed to the algorithm. Thus, `Partition` is *adaptive* in the sense that the algorithm finds out on its own what "type" of partition it is actually searching for. Indeed, this adaptivity requires new algorithmic ideas, and it is one of the main achievements of this paper; it also seems to be an important feature in applications.

1.3 Random Graphs

We shall apply Theorem 1 to obtain a rather general result on partitioning random graphs. While the scope of Theorem 1 is not limited to the type of random graphs we are considering in this section, the model is interesting because it encompasses the random graphs studied in prior work [1, 2, 6, 10].

Let ψ, $\boldsymbol{p}$, and $\mathcal{E}$ be as in Section 1.2. Then we can define a random graph $G_{n,k}(\psi, \boldsymbol{p})$ as follows: the vertex set of $G_{n,k}(\psi, \boldsymbol{p})$ is $V = \{1, \ldots, n\}$, and any two vertices $v, w \in V$ are connected with probability $\mathcal{E}_{vw}$ independently. We say that $G_{n,k}(\psi, \boldsymbol{p})$ has some propertex $\mathcal{P}$ *with high probability* ("w.h.p.") if the probability that $\mathcal{P}$ holds tends to 1 as $n \to \infty$. As we shall see in Section 2, $G_{n,k}(\psi, \boldsymbol{p})$ comprises various random graph models for specific partitioning problems such as GRAPH COLORING or MAX CUT.

Theorem 2. *Let k be a number independent of n, and suppose that ψ and $\boldsymbol{p}$ satisfy the following.*

R1. *for the quantity σ^* defined in (2) we have $\sigma^* \geq \ln^2(n/n_{\min})$,*
R2. $n_{\min} \geq \ln^{30} n$*, and*
R3. *for all $u, v \in V$ such that $\psi(u) \neq \psi(v)$ the inequality*

$$\|\mathcal{E}_u - \mathcal{E}_v\|^2 \geq \rho^2 = \frac{c_0 k^3 \sigma^*}{n_{\min}} + c_0 \ln\left(\sigma^* + \frac{n}{n_{\min}}\right) \max_{1 \leq i \leq k} \sum_{j=1}^{k} p_{ij}(1 - p_{ij}) \tag{6}$$

holds, where c_0 is a large enough constant.

Then w.h.p. $G = G_{n,k}(\psi, \boldsymbol{p})$ has the properties **A1–A4** *stated in Theorem 1, and G has a subgraph $H = \mathrm{core}(G)$ that satisfies* **H1–H4**. *Furthermore, w.h.p. all components of the graph $(G_1 \cup G_2) - H$ have at most $\ln n$ vertices.*

Letting $d(v, w)$ and $d(v)$ be as in (1), we have $\mathrm{E}(d(v, w)) = \mathcal{E}_{vw}$. The subgraph $\mathrm{core}(G)$ basically consists of those vertices for which $d(v)$ is close to its mean $\mathcal{E}_v$. Thus, $\mathrm{core}(G)$ is actually a "canonically" defined subgraph, and not an artefact produced by the algorithm (cf. Section 8.2 for a precise definition). Moreover, as $(G_1 \cup G_2) - H$ just consists of components of logarithmic size, the graph $G - H$ has a very simple structure.

To explain (6), we note that $\mathcal{E}_{vw} = \mathrm{E}(d(v, w))$, so that $\mathrm{E}(\|\mathcal{E}_v - d_v\|^2)$ quantifies the "variance" of $d(v)$. For $v \in V_i$ we can bound this by $\mathrm{E}(\|\mathcal{E}_v - d(v)\|^2) \leq \sigma^*/n_{\min}$.Furthermore, $\|\mathcal{E}_v - \mathcal{E}_w\|^2$ quantifies how much the planted partition influences $d(v) - d(w)$. Hence, (6) basically says that `Partition` can (almost) recover the planted classes $V_1, \ldots, V_k$ if the influence $\|\mathcal{E}_v - \mathcal{E}_w\|^2$ exceeds the bound $\sigma^*/n_{\min}$ on the "random noise" by a certain amount.

1.4 Related Work

The conditions **A1–A4** in Theorem 1 are reminiscent of the work on quasi-random graphs due to Chung and Graham [9], who investigate the connection between spectral and combinatorial graph properties. Moreover, several authors have investigated the applicability of spectral techniques under various other types of conditions: Bilu and Linial [4] studied stable instances, the work of Frieze and Kannan [19] applies to dense graphs (average degree $\Omega(n)$), Kannan, Vempala, and Vetta [25] considered a bicriteria measure for clustering, and Spielman and Teng [30] investigated planar graphs. In comparison with prior work, the new aspect of the present paper is that our goal is not to optimize some objective function, but to detect and recover a "latent low rank structure" of a given graph. Thus, Theorem 1 is the first result that provides a general *deterministic* formulation of this problem that ensures that the low rank structure can be computed in polynomial time.

The $G_{n,k}(\psi,\boldsymbol{p})$ model was first considered by McSherry [27], who presented a polynomial time algorithm that recovers the planted partition of $G = G_{n,k}(\psi,\boldsymbol{p})$, provided that the following holds. Let $\sigma_{\max}^2 = \max_{1\leq i,j\leq k} p_{ij}(1-p_{ij})$, and let $c_0 > 0$ be a large enough constant; then the assumption reads

$$\|\mathcal{E}_u - \mathcal{E}_v\|^2 \geq c_0 k \cdot \max\left\{\sigma_{\max}^2, \frac{\ln^6 n}{n}\right\} \cdot \left[\frac{n}{n_{\min}} + \ln n\right] \quad \text{if } \psi(u) \neq \psi(v). \tag{7}$$

The two conditions (7) and (6) compare as follows. Due to the $\ln n$-terms occurring in (7), this condition $G_{n,k}(\psi,\boldsymbol{p})$ must have average degree at least $\ln^3 n$ (and $\leq n - \ln^3 n$). By contrast, Theorem 2 also comprises the following three types of graphs.

Sparse graphs. Condition (6) allows that the mean $\mu(v,V_j)$ of the number of v-V_j-edges may be $O(1)$ for all $v \in V$ and $1 \leq j \leq k$. In this case the average degree of $G_{n,k}(\psi,\boldsymbol{p})$ is bounded as $n \to \infty$.

Massive graphs. Similarly, (6) allows that $\mu(v,V_j) = \#V_j - O(1)$ for all v,j. Then $G_{n,k}(\psi,\boldsymbol{p})$ is a massive graph, i.e., the average degree is $n - O(1)$.

Mixtures of both. The most difficult case algorithmically is a "mixture" of the above two cases: for any v and j we *either* have $\mu(v,V_j) = O(1)$ *or* $\mu(v,V_j) = \#V_j - O(1)$. In other words, some of the subgraphs induced on two sets V_i, V_j are sparse, while others are massive.

In fact, the algorithm suggested in [27] fails to produce a partition that is even close to the "planted" one on the three above types of inputs. The reason is essentially that, e.g., sparse random graphs have a considerably more *irregular degree distribution* than random graphs of average degree $\gg \ln n$, and that the tails of the degree distribution affect the spectrum of the adjacency matrix (cf. Section 2).

Furthermore, condition (7) is phrased in terms of $n\sigma_{\max}^2$, which may exceed the expression σ^* from (2) significantly if, e.g., $G_{n,k}(\psi,\boldsymbol{p})$ features a "small" part (say, of size $n^{0.1}$) of density $\frac{1}{2}$. In this case (6) can be a considerably weaker assumption than (7). Nevertheless, (6) does not strictly improve (7), because in (6) there occurs a factor of k^3, while (7) only needs a factor k (recall, however, that $k = O(1)$).

Finally, the algorithm `Partition` presented in this paper is *adaptive* in the sense that it just requires the graph G and the number k at the input. By comparison, the algorithm as it is described in [27] does require further information about the desired partition (e.g., a lower bound on $\|\mathcal{E}_v - \mathcal{E}_w\|$ for v,w in distinct classes, or on $n_{\min}$). In summary, Theorem 2 extends [27] in the following respects.

– The most important point is that `Partition` can cope with the three types of graphs described above (sparse, massive, and mixed).
– The new algorithm requires only the graph G and the number k of classes at the input.
– `Partition` is purely deterministic, while the algorithm in [27] is randomized.

Dasgupta, Hopcroft, Kannan, and Mitra [13] studied the "second eigenvector technique" on $G_{n,k}(\psi,\boldsymbol{p})$; an important point of this work is that it provides a rigorous analysis of this heuristic that contributes to explaining its success in practice. For graphs of moderate density (average degree $\geq$ polylog(n) and $\leq n -$ polylog(n)), the authors obtain a similar result as [27]. Their separation assumption is weaker than both (6) and (7), as they just need to bound $\|\mathcal{E}_u - \mathcal{E}_v\|$ in terms of $\sum_{w\in V} \mathcal{E}_{uw}(1-\mathcal{E}_{uw}) + \mathcal{E}_{vw}(1-\mathcal{E}_{vw})$ rather than in terms of σ^* or $\sigma_{\max}$. However, to achieve this they need some further conditions such as a lower bound of $\Omega(n)$ on $n_{\min}$ (in [27] and in the present work $n_{\min}$ may be as small as polylog(n)).

While in the present paper we are just dealing with the problem of recovering a "latent" partition of a given graph, there are a number of papers dealing with spectral heuristics for "classical" NP-hard problems. For instance, Alon, Krivelevich, and Sudakov [2] studied a "dense" random graph (average degree $\Omega(n)$) with a "planted" clique of size $\Omega(\sqrt{n})$; the main result of [2] can be rederived easily from Theorem 2 as well as from [27]. Further related results that involve partitioning sparse random graphs (constant average degree) include Alon and Kahale [1] (3-coloring), Boppana [6] and Coja-Oghlan [10] (MINIMUM BISECTION), Chen and Frieze [8] (hypergraph 2-coloring), Flaxman [18] (3-SAT), and Goerdt and Lanka [21]

(4-NAE-SAT). These results can only partially be derived using the techniques of [27] (namely, under the additional condition that the average degree must be at least polylogarithmic). Nonetheless, as we shall point out in Section 2, the main results of [1, 6, 8, 10, 18, 21] follow rather easily from Theorems 1 and 2.

A few authors have analyzed spectral techniques on random graphs that cannot be described in terms of the $G_{n,k}(\psi, \boldsymbol{p})$ model. For instance, Dasgupta, Hopcroft, and McSherry [12] suggested a random graph model with a "planted" partition featuring a "skewed" degree distribution. This model is very interesting, because it covers, e.g., random "power law" graphs. Their main result is that the planted partition can be recovered also in this case w.h.p. under a similar assumption as (7). Thus, it is assumed that the average degree is $\geq \mathrm{polylog}(n)$. Applied to the $G_{n,k}(\psi, \boldsymbol{p})$ model, [12] yields a similar result as [27].

Moreover, Dasgupta et al. [13] point out that their algorithm can cope with certain very *regular* sparse random graphs. More precisely, they consider random graphs with a "planted" partition $V_1, \ldots, V_k$, such that for any two vertices $v, w \in V_i$ have (exactly) the same number of randomly chosen neighbors in each class V_j. It is shown in [13] that under a certain separation condition and under the assumption that all classes V_j have size $\Omega(n/k)$ the planted partition can be recovered using the second eigenvector heuristic. However, this model is incomparable to $G_{n,k}(\psi, \boldsymbol{p})$. In fact, due to the very regular degree distribution, the model in [13] behaves actually quite similarly to "dense" $G_{n,k}(\psi, \boldsymbol{p})$ graphs (average degree $\gg \ln n$). We shall see in Section 2 that Theorem 1 also captures the model introduced in [13].

Though some of the currently best results on partitioning random graphs rely on spectral methods, there are quite a few further references on different techniques. Some examples are Bollobás and Scott [5] (randomization), Bui et al. [7] (network flows), Dyer and Frieze [15] (combinatorial methods), Feige and Kilian [16] (semidefinite programming), Jerrum and Sorkin [23] (Metropolis process), and Subramanian and Veni Madhavan [31] (breadth first search).

1.5 Techniques and Outline

Let $G = G_{n,k}(\psi, \boldsymbol{p})$ be a random graph with adjacency matrix A. To recover $V_1, \ldots, V_k$, McSherry [27] employs the following "projection method". Let $\zeta_1, \ldots, \zeta_k$ be the eigenvectors of A with the k largest eigenvalues in absolute value. Let P be a projection of $\mathbf{R}^V$ onto the subspace spanned by $\zeta_1, \ldots, \zeta_k$, and let $\hat{A} = PAP$. Then $\hat{A}$ is called a *rank k approximation* of A. Invoking results on the eigenvalues of random matrices from [20], McSherry shows that $\zeta_1, \ldots, \zeta_k$ mirror the partition $V_1, \ldots, V_k$, and that therefore the Frobenius norm $\|\hat{A} - \mathcal{E}\|_F^2 = \sum_{v \in V} \|\hat{A}_v - \mathcal{E}_v\|^2 \leq kn\sigma_{\max}^2$ is "small" (here $\hat{A}_v$, $\mathcal{E}_v$ denote the v-columns of $\hat{A}, \mathcal{E}$). In effect, $\hat{A}_v$ is "close" to $\mathcal{E}_v$ for "most" vertices v. Thus, due to the separation condition (7) it is possible to recover $V_1, \ldots, V_k$ from $\hat{A}$ (provided that the algorithm is given a lower bound on $\|\mathcal{E}_u - \mathcal{E}_v\|$ for vertices u, v in different classes).

However, this approach breaks down if $G = G_{n,k}(\psi, \boldsymbol{p})$ is a *sparse* graph such that $\#V_i p_{ij} = \Theta(1)$ as $n \to \infty$ for all i, j. In this case the rank k approximation does *not* approximate $\mathcal{E}$ well. The reason is that w.h.p. the degree distribution of $G_{n,k}(\psi, \boldsymbol{p})$ features an *upper tail*; for instance, the maximum degree is $\Omega(\frac{\ln n}{\ln \ln n})$ w.h.p. In fact, vertices of degree $d \gg \sigma^*$ induce eigenvalues that are as large as $\sqrt{d}$ in absolute value, while the assumption (6) just ensures that the eigenvalues corresponding to the partition $V_1, \ldots, V_k$ are about $k\sqrt{\sigma^*}$ in absolute value. In other words, vertices of "atypically high" degree jumble up the spectrum of A, so that the most outstanding eigenvalues do not correspond to the desired partition anymore.

Thus, in the situation of Theorems 1 and 2 we need a more sophisticated approach to obtain a matrix $\hat{A}$ that approximates $\mathcal{E}$ well. Following the work [1] on 3-coloring sparse random graphs, one could try to settle the problem by just removing vertices of degree $\gg \sigma^*$ from G. However, the issue is that the algorithm `Partition` *does not know* σ^* (it is given just G and k); indeed, it is not easy to compute (or approximate) σ^* from G. To cope with this, `Partition` employs a subroutine `Approx` that constructs a "Cauchy sequence" of matrices $\hat{A}_t$ that "converges" to $\mathcal{E}$.

As a next step, `Partition` uses the matrix $\hat{A}$ to compute an initial partition $S_1, \ldots, S_k$ of G. The basic idea is to put $u, v \in V$ into the same S_i iff $\|\hat{A}_u - \hat{A}_v\| \leq 0.1\rho$, say, where ρ is the separation parameter from **A2**. Of course, the problem is that `Partition` does *not* get ρ as an input parameter. Instead, `Partition` employs a procedure `Initial` that computes "centers" $\xi_1, \ldots, \xi_k$ and a partition $S_1, \ldots, S_k$ such that the "squared distance" $\sum_{i=1}^k \sum_{v \in S_i} \|\hat{A}_v - \xi_i\|^2$ is minimized. This partition turns out to be "close" to $V_1, \ldots, V_k$.

Finally, to home in on $V_1, \dots, V_k$, `Partition` calls a local improvement heuristic `Improve`. This heuristic repeats the following operation: to each vertex v we assign a vector $\delta(v)$ that represents the densities $e(v, S_i)/\#S_i$ (reminiscent of (1)). Then, `Improve` shifts each vertex v into that class S_i such that $\|\delta(v) - \xi_i\|$ is minimum. While this procedure is purely combinatorial, its *analysis* relies on spectral arguments and may be of independent interest. A crucial issue here is that `Improve` has to deal with classes $V_1, \dots, V_k$ of (possibly) vastly different sizes, e.g., $\mathrm{polylog}(n)$ vs. $\Theta(n)$.

The paper is organized as follows. In Section 2 we illustrate Theorems 1 and 2 with some examples of concrete graph partitioning problems. Sections 3–7 contain the description of `Partition` and its subroutines and the proof of Theorem 1. Moreover, in Section 8 we apply Theorem 1 to the random graph $G_{n,k}(\psi, \boldsymbol{p})$, thereby proving Theorem 2. Finally, Section 9 contains the proofs of a few technical lemmas.

1.6 Notation and Preliminaries

Throughout the paper we let $V = \{1, \dots, n\}$. If $G = (V, E)$ is a graph, then $A(G)$ denotes its adjacency matrix. Further, for $X, Y \subset V$ we let $e(X, Y) = e_G(X, Y)$ denote the number of X-Y-edges in G, and we set $e(X) = e_G(X) = e_G(X, X)$. Moreover, $d_G(v) = e_G(v, V)$ denotes the degree of v.

We let $\mu(X, Y)$ denote the *expected* number of X-Y-edges in the random graph $G_{n,k}(\psi, \boldsymbol{p})$. Even in Section 3–7, where we do not work with random graphs, it is helpful to use this notation. Further, we set $\mathcal{E}_{V_i} = \mathcal{E}_v$ for any $v \in V_i$. Moreover, we always let Φ denote the matrix (3), and G_1, G_2 denote the graphs defined in (4), (5).

If $M = (m_{vw})_{v,w \in V}$ is a matrix and $v \in V$, then $M_v = (m_{wv})_{w \in V}$ is the v-column of M. We let $\|M\| = \max_{\xi : \|\xi\| = 1} \|M\xi\|$ denote the operator norm and $\|M\|_F = \sqrt{\sum_{v \in V} \|M_v\|^2}$ the Frobenius norm of M. Further, if $X, Y \subset V$, then $M_{X \times Y}$ signifies the matrix obtained from M by replacing all entries m_{xy} with $(x, y) \notin X \times Y$ by 0. For brevity we let $M_X = M_{X \times X}$. If A a matrix of rank $\leq k$, then

$$\|A\|^2 \leq \|A\|_F^2 \leq k\|A\|^2. \tag{8}$$

Furthermore, let A be a symmetric matrix, and let $\zeta_1, \dots, \zeta_k$ denote eigenvectors of A with the k largest eigenvalues in absolute value. Let P be the projection onto the space spanned by $\zeta_1, \dots, \zeta_k$. Then we call $\tilde{A} = PAP$ a *rank k approximation* of A. This definition ensures that if B is any rank k matrix, then

$$\|\tilde{A} - A\| \leq \|B - A\|. \tag{9}$$

2 Applications and Examples

Graph coloring. Alon and Kahale [1] developed a spectral heuristic for coloring 3-colorable graphs generated according to the following model. Let $\psi : V \to \{1, 2, 3\}$ be a random mapping, and let $p_{ij} = p$ if $i \neq j$ and $p_{ii} = 0$ for $i, j = 1, 2, 3$. Then V_1, V_2, V_3 is a planted 3-coloring of $G = G_{n,k}(\psi, \boldsymbol{p})$. In this section we observe that the main result of [1] can be derived from Theorem 2 by adding only few problem specific details (in a similar way one can rederive the results of [6, 10]). We also discuss how the assumptions (7) from [27] and (6) from Theorem 2 relate to each other.

To satisfy (7), we need that $p \geq c'(\ln^3 n)/n$ for a certain constant $c' > 0$. In this case w.h.p. *all* vertices $v \in V_i$ have $(1 + o(1))np/3$ neighbors in the other two classes V_j, $i \neq j$ (by Chernoff bounds), so that G is quite regular. Furthermore, let $\zeta_i \in \mathbf{R}^V$ be the characteristic vector of V_i. Then for $i \neq j$ we have

$$A(G)(\zeta_i - \zeta_j) \sim \frac{np}{3}(\zeta_j - \zeta_i), \tag{10}$$

Moreover, all eigenvectors $\xi \perp \zeta_1, \zeta_2, \zeta_3$ have eigenvalues of order $O(\sqrt{np})$. Hence, the spectrum of $A(G)$ is very "clean" in that the three eigenvectors with the "most outstanding" eigenvalues correspond to V_1, V_2, V_3. In fact, V_1, V_2, V_3 can be read off easily from these three eigenvectors w.h.p.

By comparison, the condition (6) of Theorem 2 only requires that $p \geq c/n$ for a constant $c > 0$, which is exactly the assumption needed in [1]. Let us assume that actually $p = c/n$. Then the numbers $e(v, V_j)$ for $v \in V_i \neq V_j$ are asymptotically Poisson with mean $c/3$. Therefore, w.h.p.

$$\#\{v \in V_i : e(v, V_j) = \gamma\} \sim (c/3)^\gamma \exp(-c/3)n/(3\gamma!) \ . \tag{11}$$

Consequently, it is *impossible* to recover the partition V_1, V_2, V_3 from G perfectly. For by (11) G contains $\Omega(n)$ isolated vertices w.h.p., and of course *no* algorithm can tell which isolated vertex belongs to which V_i. This shows that Theorems 1 and 2 are best possible in the sense that in general we can just hope to recover the correct partition on a subgraph H of G, but not on the entire graph G.

Furthermore, if $p = c/n$, then the spectrum of $A(G)$ does not reflect the planted coloring as nicely as in the "dense" case. For by (11) G contains a large number of stars $K_{1,d}$ with $d \gg c^2$. Thus, the eigenvalues $\pm\sqrt{d} \gg c$ of $A(K_{1,d})$ show up in the spectrum of $A(G)$. In effect, the "relevant" eigenvalues (10) of order c are "hidden" among a lot of eigenvalues $\pm\sqrt{d}$ that result from the upper tail of the degree distribution. Hence, the algorithm from [27] would use eigenvectors merely representing the highest degree vertices, whence it would fail to recover V_1, V_2, V_3. (In fact, it has been observed in [26] that the spectrum undergoes a phase transition as $np \sim \ln n$.)

Nonetheless, by Theorem 2 `Partition` can compute sets S_1, S_2, S_3 such that $S_i \cap H = V_i \cap H$, where $H = \mathrm{core}(G)$. Though S_1, S_2, S_3 do not coincide with V_1, V_2, V_3 perfectly, we can use S_1, S_2, S_3 to 3-color G. To this end, we follow the strategy of Alon and Kahale: by Theorem 2 $G - H$ just consists of components of size $\leq \ln n$. Hence, for each of these components we can compute in polynomial time a 3-coloring that extends the 3-coloring $S_1 \cap H, S_2 \cap H, S_3 \cap H$ of H. Glueing all these 3-colorings together yields the desired 3-coloring of all of G.

Random 3-SAT. Flaxman [18] studied the following model of random 3-SAT. Let $x_1, \ldots, x_n$ be propositional variables, and let $L = \{x_i, \bar{x}_i : 1 \leq i \leq n\}$ be the set of literals. Let $p_i = c_i n^{-3}$. Moreover, pick a random assignment of $x_1, \ldots, x_n$, let T be the set of literals that evaluate to true, and let $F = L \setminus T$. Then, let ϕ be a random 3-SAT formula obtained by including each possible clause over L that contains exactly i literals in T with probability p_i independently.

Flaxman presents an efficient algorithm that computes a satisfying assignment of ϕ, provided (essentially) that c_1, c_2, c_3 exceed a certain large enough constants. The algorithm sets up a graph G with vertex set L in which each clause is represented as a triangle involving the three literals of the clause. Flaxman proves that in G the partition $V_1 = T$, $V_2 = F$ enjoys a separation property (similar to **A2**), and that therefore a partition T', F' of G that coincides with T, F on a large subgraph H of G can be computed via spectral techniques. Then, he uses a brute force algorithm to assign the literals in $G - H$ so that ϕ is satisfied. The same result can be derived easily by employing the algorithm `Partition` from Theorem 1. Observe, however, that the graph G cannot be described in terms of the $G_{n,k}(\psi, \boldsymbol{p})$ model, because edges do not appear independently; thus Theorem 2 does not apply here.

Regular graphs. Bui et al. [7] suggested the following model for MINIMUM BISECTION: suppose that $d' > d$ and that n is even, and let V_1, V_2 be a random partition of V into two pieces of equal size. Then, let G be a graph chosen uniformly at random in which each vertex $v \in V_i$ has exactly d' neighbors in V_i and exactly d neighbors in V_{3-i} $(i = 1, 2)$. Bui et al. show that in this model a minimum bisection (namely V_1, V_2) can be computed in polynomial time (via flow techniques), provided (essentially) that $d' > c$ and $d = o(1)$ for a certain constant $c > 0$.

Using methods from [24], one can show that w.h.p. G has the properties **A1**–**A4**, and that **H1**–**H4** actually hold for $H = G$, provided that $d' \geq d + c(\sqrt{d'} + 1)$ for a certain constant $c > 0$. Thus, Theorem 1 shows that `Partition` yields an optimal bisection w.h.p. This result improves on [7] considerably, since the necessary condition on the parameters is much weaker (but of course the flow techniques suggested in [7] are of independent interest). A similar result was obtained in [13] (via spectral techniques as well).

Once more, G cannot be described in terms of the $G_{n,k}(\psi, \boldsymbol{p})$ model, because the edges do not occur independently. However, even though G can be a sparse graph, due to its very regular degree distribution it is much easier to deal with than a sparse random graph $G_{n,k}(\psi, \boldsymbol{p})$ (e.g., we can set $H = G$ here).

3 The Algorithm `Partition`

Throughout Sections 3–7, we let G be a graph that satisfies **A1**–**A4**. *Moreover, we assume that H is a subgraph of G that has the properties* **H1**–**H4**. *Furthermore, we implicitly assume that n and c_0 are sufficiently* *large. Finally, we use the symbols Φ, σ^*, G_1, and G_2 as defined in Section 1.2.* Note that **A3** implies that

$$d_{G_1 \cup G_2}(v) \leq 7\sigma^* + \ln^3 n \qquad \text{for all } v \in V. \tag{12}$$

In the sequel, we summarize the functioning of `Partition` and its subroutines. We will present and analyze the subroutines `Identify`, `Approx`, and `Improve` in detail in Sections 4–7.

Algorithm 3. `Partition`(G, k)
Input: A graph $G = (V, E)$ and an integer k. *Output:* A partition $T_1, \ldots, T_k$ of G.

1. Run the procedure `Identify`(G, k).
2. If `Identify` fails, then let $\hat{A}$ be a rank k approximation of A.
   otherwise let $\varphi = (\varphi_{vw})_{v,w \in V}$ be the output of `Identify`, and let $\hat{A} =$ `Approx`(G, φ).
3. Let $(S_1, \ldots, S_k, \xi_1, \ldots, \xi_k) =$ `Initial`$(\hat{A}, k)$.
4. Let $(T_1, \ldots, T_k) =$ `Improve`$(G, S_1, \ldots, S_k, \xi_1, \ldots, \xi_k)$.
   Output $(T_1, \ldots, T_k)$.

In Steps 1–2 the goal is to compute a matrix $\hat{A}$ that approximates $\mathcal{E}$ well. If $\sigma^* \gg \ln n$ is not too small, then **A1** ensures that we could just let $\hat{A}$ be any rank k approximation of $A(G)$.

By contrast, if σ^* is small (say, $\sigma^* = O(1)$ as $n \to \infty$) then G consists of "extremely sparse" and/or "extremely dense" parts. Indeed, by **A4** the average degree of $G_1 \cup G_2$ is $\leq 1000\sigma^*$, say. In this case a rank k approximation of $A(G)$ does *not* provide a good approximation of $\mathcal{E}$ (cf. Section 2). Instead, to approximate $\mathcal{E}$, it is instrumental to determine which parts of the graph are sparse and which are dense, i.e., to compute the matrix Φ. This is the aim of the procedure `Identify`, cf. Section 4.

Proposition 4. *`Identify` either outputs the matrix Φ or "fail", and if $\sigma^* \leq \ln^3 n$, then the output is Φ.*

Moreover, in Section 5 we shall establish the following.

Proposition 5. *If $\varphi = \Phi$, then the output $\hat{A}$ of `Approx`(G, φ) is a matrix of rank k such that $\|\hat{A} - \mathcal{E}\| \leq c_0^2 k\sqrt{\lambda}$. Furthermore, if $\sigma^* > \ln^3 n$, then any rank k approximation A' of A satisfies $\|A' - \mathcal{E}\| \leq c_0^2 k\sqrt{\lambda}$.*

Combining Propositions 4 and 5, we conclude that the matrix $\hat{A}$ computed in Step 2 satisfies

$$\|\hat{A} - \mathcal{E}\| \leq c_0^2 k\sqrt{\lambda}, \quad \text{and thus } \|\hat{A} - \mathcal{E}\|_F^2 \overset{(8)}{\leq} c_0^5 k^3 \lambda \overset{\mathbf{A2}}{\leq} c_0^{-3} \rho^2 n_{\min}. \tag{13}$$

Consequently, since $\|\hat{A} - \mathcal{E}\|_F^2 = \sum_{v \in V} \|\hat{A}_v - \mathcal{E}_v\|^2$, for "most" vertices v the distance $\|\hat{A}_v - \mathcal{E}_v\|$ is "small" ($< 0.01\rho$, say). Therefore, `Initial` partitions the vertices $v \in V$ according to the vectors $\hat{A}_v$. More precisely, `Initial` computes k "centers" $\xi_1, \ldots, \xi_k \in \mathbf{R}^V$ and a partition $S_1, \ldots, S_k$ of V such that essentially S_i consists of those vertices v that are close to ξ_i, cf. Section 6.

Proposition 6. *The output of `Initial` enjoys the following properties.*

1. *There is a permutation τ of $\{1, \ldots, k\}$ such that $\|\xi_i - \mathcal{E}_{\tau(i)}\|^2 \leq 0.001\rho^2$ for all $i = 1, \ldots, k$.*
2. $\sum_{i=1}^k \#S_i \triangle V_{\tau(i)} < 0.001 n_{\min}$.
3. $\sum_{a,b=1}^k \#S_a \cap V_b \cdot \|\mathcal{E}_{\tau(a)} - \mathcal{E}_{\tau(b)}\|^2 < 0.001\rho^2 n_{\min}$ *for all $1 \leq j \leq k$.*

While the initial partition $S_1, \ldots, S_k$ is solely determined by the matrix $\hat{A}$, the subroutine `Improve` actually investigates combinatorial properties of G. `Improve` performs iteratively a local improvement of the initial partition $S_1, \ldots, S_k$ that restricted to the subgraph H converges to the planted partition $V_1, \ldots, V_k$.

Proposition 7. *There is a permutation τ such that the output $T_1, \ldots, T_k$ of `Improve` satisfies $T_i \cap H = V_{\tau(i)} \cap H$ for all $i = 1, \ldots, k$.*

A detailed description of `Improve` can be found in Section 7. Finally, since all the procedures run in polynomial time, Theorem 1 is an immediate consequence of Propositions 4–7.

4 Identifying Sparse/Dense Parts

4.1 The Procedure `Identify`

Algorithm 8. $\texttt{Identify}(G, k)$
Input: A graph $G = (V, E)$, the integer k. *Output:* Either a matrix $\varphi = (\varphi_{vw})_{v,w\in V}$ or "fail".

1. Compute a rank k approximation $A^* = (a^*_{vw})_{v,w\in V}$ of $A(G)$.
   Let $B = (b_{vw})_{v,w\in V}$ be the matrix with entries $b_{vw} = 1$ if $a^*_{vw} \geq \frac{1}{2}$ and $b_{vw} = 0$ otherwise.
2. Construct an auxiliary graph $\mathcal{B} = (V, F)$, where $\{v, w\} \in F$ iff $\|B_v - B_w\| > \ln^{24} n$.
   Apply the greedy algorithm for graph coloring to $\mathcal{B}$, and let $T_1, \ldots, T_R$ be the resulting color classes.
3. For all $i, j \in \{1, \ldots, R\}$ and each pair $v \in T_i$, $w \in T_j$ let

$$\varphi_{vw} = \begin{cases} 1 & \text{if } i \neq j \wedge e_G(T_i, T_j) > 0.66\#T_i\#T_j \\ 1 & \text{if } i = j \wedge e_G(T_i) > 0.66\binom{\#T_i}{2}, \\ 0 & \text{otherwise.} \end{cases}$$

4. Let G_1^* be the subgraph of G consisting of all edges $\{v, w\} \in E$ such that $\varphi_{vw} = 0$. Moreover, let G_2^* be the subgraph of $\bar{G}$ consiting of all edges $\{v, w\} \notin E$ satisfying $\varphi_{vw} = 1$.
   If $R \leq k$ and the maximum degree of $G_1^* \cup G_2^*$ is $\leq \ln^4 n$, then return φ. Otherwise output "fail".

The aim of `Identify` is to compute the matrix Φ defined in (3). Let us call two classes V_i, V_j *similar* if for all indices l we have $p_{il} \geq \frac{1}{2} \leftrightarrow p_{jl} \geq \frac{1}{2}$. Moreover, we say that two vertices v, w are similar if they belong to similar classes V_i, V_j.

`Identify` performs a very coarse spectral partitioning of G to identify similar vertices. As a first step, `Identify` computes a low rank approximation A^* of $A(G)$. By **A1** A^* should provide at least a "rough" approximation of $\mathcal{E}$. Then, `Identify` constructs a matrix B by rounding the entries of A^* to 0/1; as the desired output Φ is obtained by rounding the entries of $\mathcal{E}$, B should be "close" to Φ. In fact, if σ^* is "small", then the entries of $\mathcal{E}$ differ from $0/1$ only "a little", so that B should actually be close to $\mathcal{E}$. In Section 4.2 we shall prove the following lemma to estimates $\|B - \mathcal{E}\|$.

Lemma 9. *If $\sigma^* \leq \ln^{10} n$, then $\|B - \mathcal{E}\|_F^2 \leq \log^{23} n$.*

Of course, the difficult part about `Identify` is that we are to compute a matrix φ that coincides *exactly* with Φ. To this end, Step 2 of `Identify` sets up a graph $\mathcal{B}$ in which two vertices v, w are adjacent iff their columns B_v, B_w are far apart. Hence, two vertices should be adjacent in $\mathcal{B}$ iff they are *not* similar, and thus $\mathcal{B}$ should be a complete r-partite graph for some $r \leq k$. Now, the algorithm computes a greedy coloring $T_1, \ldots, T_R$ of $\mathcal{B}$. If $\mathcal{B}$ is indeed complete r-partite, then clearly the greedy algorithm will use $R = r \leq k$ colors. Finally, `Identify` sets up a matrix φ that attains the value 1 on $T_i \times T_j$ if the pair T_i, T_j is "dense", and 0 otherwise. In Section 4.3 we shall prove that this yields the desired output if $\sigma^* \leq \ln^3 n$.

Lemma 10. *If $\sigma^* \leq \ln^{10} n$, then `Identify` either fails or outputs $\varphi = \Phi$. Furthermore, if $\sigma^* \leq \ln^3 n$, then actually `Identify` outputs $\varphi = \Phi$.*

In the light of Lemmas 9 and 10, to complete the proof of Proposition 4 we just need to show that in the case $\sigma^* > \ln^{10} n$ the algorithm outputs "fail" (but does not return a "wrong" matrix $\varphi \neq \Phi$).

Lemma 11. *If $\sigma^* > \ln^{10} n$, then `Identify` outputs "fail".*

The proof of Lemma 11 can be found in Section 4.4. Finally, Proposition 4 is an immediate consequence of Lemmas 9–11.

4.2 Proof of Lemma 9

To prove the lemma, we shall establish the two estimates

$$\|\Phi - \mathcal{E}\|_F^2 \leq \ln^{22} n, \tag{14}$$

$$\|\Phi - B\|_F^2 \leq \ln^{22} n. \tag{15}$$

Then the assertion follows immediately by applying the triangle inequality.

Since Φ, $\mathcal{E}$ both have rank $\leq k$, (8) yields

$$\begin{aligned}\|\Phi-\mathcal{E}\|_F &\leq \sqrt{2k}\|\Phi-\mathcal{E}\| \leq \sqrt{2k}\sum_{i,j=1}^{k}\|\Phi_{V_i\times V_j}-\mathcal{E}_{V_i\times V_j}\| \\ &= \sqrt{2k}\sum_{i,j:p_{ij}\leq\frac{1}{2}} p_{ij}\|\boldsymbol{J}_{V_i\times V_j}\| + \sqrt{2k}\sum_{i,j:p_{ij}>\frac{1}{2}}(1-p_{ij})\|\boldsymbol{J}_{V_i\times V_j}\| \\ &\leq \sqrt{2k}\sum_{i,j=1}^{k} 2p_{ij}(1-p_{ij})\sqrt{\#V_i\#V_j} \leq \sqrt{8k}\sum_{i,j=1}^{k} p_{ij}(1-p_{ij})\#V_j \leq (2k)^{3/2}\sigma^* \leq \ln^{11} n,\end{aligned}$$

because we are assuming that $\sigma^* \leq \ln^{10} n$. Hence, we obtain (14).

Finally, to establish (15) we note that $\max_{v\in V} d_{G_1\cup G_2}(v) \leq 11\ln^{10} n$ by (12). Therefore, **A1** yields $\|A^*-\mathcal{E}\|^2 \leq c_0^2k^2(\sigma^{*\,2}+11\ln^{10} n) \leq 2c_0^2k^2\ln^{20} n$. Hence, as both A^*, $\mathcal{E}$ have rank k, (8) entails

$$\|A^*-\mathcal{E}\|_F^2 \leq 2k\|A^*-\mathcal{E}\|^2 \leq \ln^{21} n. \tag{16}$$

Furthermore, as $\sigma^* \leq \ln^{10} n$, we have $p_{ij}(1-p_{ij}) \leq \ln^{-1} n \leq 0.01$, because $\#V_j \geq n_{\min} \geq \ln^{11} n$ by **A4** and $\sigma^* \geq p_{ij}(1-p_{ij})\#V_j$ for all $1 \leq i,j \leq k$. Hence, if $B_{vw} \neq \Phi_{vw}$, then $|A^*_{vw}-\mathcal{E}_{vw}| \geq \frac{1}{3}$, and thus $\|B-\Phi\|_F^2 \leq 9\|A^*-\mathcal{E}\|_F^2$. Therefore, (15) follows from (16).

4.3 Proof of Lemma 10

Throughout we assume that $\sigma^* \leq \ln^{10} n$. To prove Lemma 10, we need the following observation.

Lemma 12. *Let $1 \leq i,j \leq k$. If $p_{ij} > \frac{1}{2}$, then actually $p_{ij} > 0.9$. Moreover, $p_{ij} \leq \frac{1}{2}$ in fact implies that $p_{ij} < 0.1$.*

Proof. By Lemma 9 we have $\|B-\mathcal{E}\|_F^2 \leq \log^{23} n$. Suppose that $p_{ij} > \frac{1}{2}$. If $p_{ij} < 0.9$, then for all $v \in V_i$ and all $w \in V_j$ we would have $|B_{vw}-\mathcal{E}_{vw}| \geq 0.1$, which yields the contradiction $\|B-\mathcal{E}\|_F^2 \geq 0.01\#V_i\#V_j \geq 0.01n_{\min}^2 > \ln^{50} n$ (cf **A4**). A similar argument establishes the second assertion. □

Corollary 13. *For all $v,w \in V$ we have $\|B_v-B_w\|^2 \leq \ln^{24} n \Leftrightarrow v,w$ are similar.*

Proof. Suppose that $v \in V_i$ and $w \in V_j$ are not similar. Let $1 \leq l \leq k$ be such that $p_{il} > \frac{1}{2}$ and $p_{jl} \leq \frac{1}{2}$. Since $\|B-\mathcal{E}\|_F^2 \leq \ln^{23} n$ by Lemma 9, there are at most $2\ln^{23} n$ vertices $u \in V_l$ such that $B_{vu} = 0$. Similarly, for at most $2\ln^{23} n$ vertices $u \in V_l$ we have $B_{wu} = 1$. Hence, $\|B_v-B_w\|^2 \geq \#V_l - 4\ln^{23} n > \ln^{24} n$ by **A4**.

Conversely, assume that $v,w \in V$ are similar. Let $x = \#\{u \in V : |B_{uv}-\mathcal{E}_{uv}| \geq \frac{1}{3}\}$ and $y = \#\{u \in V : |B_{uw}-\mathcal{E}_{uw}| \geq \frac{1}{3}\}$. Since for all i,j we either have $p_{ij} < 0.1$ or $p_{ij} > 0.9$ by Lemma 12, we obtain $\|B_v-B_w\|^2 \leq x+y \leq 9\|B-\mathcal{E}\|_F^2 \leq \ln^{24} n$, thereby proving the corollary □

Corollary 14. *For all $1 \leq i,j \leq k$ the following holds:*

$$\text{if } p_{ij} > \tfrac{1}{2}, \text{ then } e(V_i,V_j) \geq \tfrac{2}{3}\#V_i\#V_j \ (i \neq j), \text{ resp. } e(V_i,V_j) \geq \tfrac{2}{3}\tbinom{\#V_i}{2}\ (i=j), \tag{17}$$

$$\text{if } p_{ij} \leq \tfrac{1}{2}, \text{ then } e(V_i,V_j) \leq \tfrac{1}{3}\#V_i\#V_j \ (i \neq j), \text{ resp. } e(V_i,V_j) \geq \tfrac{1}{3}\tbinom{\#V_i}{2}\ (i=j). \tag{18}$$

Proof. To prove (17), suppose that $p_{ij} > \frac{1}{2}$. Since then we actually have $p_{ij} > 0.9$ by Lemma 12, in the case $i \neq j$ we get $\mu(V_i,V_j) > 0.9\#V_i\#V_j$, and if $i = j$, then $\mu(V_i) > 0.9\binom{\#V_i}{2}$. Therefore, **A3** implies (17). A similar argument yields (18). □

Proof of Lemma 10. Corollary 13 implies that two vertices $v,w \in V$ are adjacent in the graph $\mathcal{B}$ iff they are not similar. Hence, $\mathcal{B}$ is a complete R-partite graph, whose color classes $T_1,\ldots,T_R$ are exactly the equivalence classes of the similarity relation. Therefore, (17) and (18) entail that φ equals Φ and thus the graphs G_1^*, G_2^* constructed in Step 4 of `Identify` coincide with G_1 and G_2. Consequently, `Identify` either outputs "fail" or $\varphi = \Phi$. Furthermore, if $\sigma^* \leq \ln^3 n$, then the maximum degree of $G_1\cup G_2$ is $\leq \ln^4 n$ by (12), whence `Identify` outputs $\varphi = \Phi$. □

4.4 Proof of Lemma 11

The basic idea of the proof is as follows. If `Identify` does not fail, then $R \le k$ and the maximum degree of $G_1^* \cup G_2^*$ is $\le \ln^4 n$. Thus, G consists of $\le k$ parts $T_1, \ldots, T_R$ such that the graphs induced on the sets T_i and the bipartite graphs consistsing of the T_i-T_j-edges are either *extremely* sparse (maximum degree $\le \ln^4 n$) or *extremely* dense (maximal number of "missing" edges per vertex $\le \ln^4 n$). However, according to the matrix $\mathcal{E}$ of "expectations", G should feature at least one piece of "moderate" density (of average degree between $\ge \frac{1}{2}\ln^{10} n$ in $G_1 \cup G_2$). Hence, if `Identify` does not fail, then $A(G)$ must be far apart from $\mathcal{E}$, which contradicts **A1**.

Let us now carry out this idea in detail. Let $M = \mathcal{E} - A(G)$. Then **A1** and (12) entail that $\|M\| \le c_0^2 k\sqrt{\sigma^* n_{\min}/\ln n}$. By contrast, we shall prove that if `Identify` does not fail, then

$$\|M\| > 10^{-4}k^{-3}\sqrt{n_{\min}\sigma^*} > c_0^2 k\sqrt{\sigma^* n_{\min}/\ln n}, \tag{19}$$

which is a contradiction.

Thus, assume that `Identify` does not fail and hence outputs some matrix φ, which is based on a partition $T_1, \ldots, T_R$ with $R \le k$. Let $1 \le i, j \le k$ be such that $\# V_j p_{ij}(1 - p_{ij}) \ge k^{-1}\sigma^*$ (note that possibly $i = j$). We may assume without loss of generality that $p_{ij} \le \frac{1}{2}$ (if $p_{ij} > \frac{1}{2}$, we just replace G by its complement and $\mathcal{E}$ by $\boldsymbol{J} - \mathcal{E}$). Clearly, for each $v \in V_i$ there is some $1 \le \gamma(v) \le R$ such that $e(v, V_j \cap T_{\gamma(v)}) \ge R^{-1}e(v, V_j)$. Indeed, we will show below that

$$\ln^9 n \le \frac{1}{2kR}\sigma^* \le \# T_{\gamma(v)} \cap V_j \le e(v, V_j \cap T_{\gamma(v)}) + \ln^4 n \quad \text{for all } v \in V_i. \tag{20}$$

Now, let $1 \le \alpha \le R$ be such that $\#\gamma^{-1}(\alpha) \ge R^{-1}\# V_i$. Choose a set $S \subset \gamma^{-1}(\alpha) \subset V_i$ of cardinality

$$s = \# S = \lceil 10^{-4}k^{-3}n_{\min}\rceil \tag{21}$$

arbitrarily, set $T = T_\alpha \cap V_j \setminus S$, and let $t = \# T$. Since $d_{G_1}(w) \le 11\sigma^*$ for all $w \in T_{\gamma(\alpha)}$ by (12), and because $e_{G_1}(\gamma^{-1}(\alpha), T_\alpha \cap V_j) \ge \frac{1}{4kR}\sigma^* \#\gamma^{-1}(\alpha)$ by (20), we conclude that

$$\# T_\alpha \cap V_j \ge \frac{e_{G_1}(\gamma^{-1}(\alpha), T_\alpha \cap V_j)}{\max_{w \in T_\alpha} d_{G_1}(w)} \ge \frac{\#\gamma^{-1}(\alpha)}{44k^2} \ge \frac{\# V_i}{44k^2 R} \ge \frac{n_{\min}}{44k^3} \ge 50s. \tag{22}$$

Therefore,

$$t = \# T \ge \# T_\alpha \cap V_j - s \overset{(22)}{\ge} \frac{1}{2}\# T_\alpha \cap V_j \overset{(20)}{\ge} \frac{\sigma^*}{4k^2}. \tag{23}$$

Further, combining the right inequality from (20) with (22), we conclude that $e(S, T) \ge s(t - \ln^4 n) \ge 0.9st$, while $\mu(S, T) \le p_{ij}st \le \frac{1}{2}st$. Hence,

$$\frac{2}{5}st \le e(S, T) - \mu(S, T) = -\langle M\mathbf{1}_S, \mathbf{1}_T\rangle \le \|M\| \cdot \|\mathbf{1}_S\| \cdot \|\mathbf{1}_T\| = \|M\|\sqrt{st}. \tag{24}$$

Thus, combining (21), (23), and (24), we obtain $\|M\| \ge \frac{2}{5}\sqrt{st} \ge 10^{-4}k^{-3}\sqrt{n_{\min}\sigma^*}$, thereby proving (19).

Finally, to prove (20) we first show that

$$\# T_{\gamma(v)} \cap V_j \ge e(v, V_j \cap T_{\gamma(v)}) \ge \# T_{\gamma(v)} \cap V_j - \ln^4 n \quad \text{for all } v \in V_i. \tag{25}$$

To see this, let $1 \le \beta(v) \le R$ be the index such that $v \in T_{\beta(v)}$. If the entries of φ on the rectangle $T_{\gamma(v)} \times T_{\beta(v)}$ were 0, then $d_{G_1^*}(v) \ge e(v, T_{\gamma(v)}) \ge R^{-1}e(v, V_j)$. As

$$e(v, V_j) \ge \frac{1}{2}\mu(v, V_j) \ge \frac{1}{2}\# V_j p_{ij} \ge \frac{1}{2k}\sigma^* \tag{26}$$

by **A3** and the choice of i, j, we conclude that $d_{G_1^*}(v) \ge \frac{1}{4kR}\sigma^* > \ln^4 n$. But then `Identify` would fail. Consequently, φ attains the value 1 on $T_{\gamma(v)} \times T_{\beta(v)}$. Therefore, $\# T_{\gamma(v)} \cap V_j - e(v, V_j \cap T_{\gamma(v)}) \le d_{G_2^*}(v) \le \ln^4 n$ (because otherwise $d^*_{G_2}(v) > \ln^4 n$, so that `Identify` would fail), whence (25) follows. Furthermore, combining (25) and (26) we obtain (20), because $e(v, V_j \cap T_{\gamma(v)}) \ge R^{-1}e(v, V_j)$ by the definition of $\gamma(v)$.

5 Approximating the Expected Densities

5.1 The Procedure Approx

Algorithm 15. Approx(G, φ)
Input: A graph $G = (V, E)$ and a matrix $\varphi = (\varphi_{vw})_{v,w\in V}$. *Output:* A matrix $\hat{A}$.

1. Let G_1^* be the graph consisting of all edges $\{v, w\} \in E$ such that $\varphi_{vw} = 0$. Further, let G_2^* consist of all edges $\{v, w\} \notin E$ satisfying $\varphi_{vw} = 1$.
   Let $\Delta = n$, set $R_0 = \emptyset$, and let $A_0 = (a_{0,vw})_{v,w\in V} = A(G)$.
2. For $t = 1, \ldots, \log_2 \Delta$ do
3. Let $\Delta_t = 2^{-t}\Delta$ and $R_t = \{v \in V : d_{G_1^*\cup G_2^*}(v) > \Delta_t\}$.
   Let $A_t = (a_{t,vw})_{v,w\in V}$ be the matrix with entries $a_{t,vw} = \varphi_{vw}$ if $(v, w) \in R_t \times V \cup V \times R_t$, and $a_{t,vw} = a_{t-1,vw}$ otherwise.
   If there is an $0 \le s < t$ such that $\|A_s - A_t\| > 4c_0k\sqrt{\Delta_s}$, then abort the for-loop and go to Step 4.
4. Let $\hat{t} = \max\{0, t - 1\}$ and return a rank k approximation of $A_{\hat{t}}$.

The aim of Approx is to compute a low rank matrix $\hat{A}$ that approximates $\mathcal{E}$. To this end, Approx analyses the spectrum of A. On the one hand, if $\lambda \ge \ln^2 n$, then by **A1** the k largest eigenvalues in absolute value of A yield a good enough approximation of the spectrum of $\mathcal{E}$ (cf. Section 3). On the other hand, if λ is "small" – e.g., $\sigma^* \le \lambda = O(1)$ – then the "relevant" eigenvalues of A do not necessarily stand out anymore but may be hidden among "noise" that is due to fluctuations of the vertex degrees (cf. Section 2). Indeed, the "relevant" eigenvalues corresponding to the spectrum of $\mathcal{E}$ are in general about $k\sqrt{\lambda}$, while vertices that have degree $d \gg k^2\lambda$ in the graph $G_1 \cup G_2$ induce eigenvalues $\sqrt{d} \gg k\sqrt{\lambda}$.

Of course, if the parameter σ^* were known to the algorithm, then we could just delete all vertices v such that, say, $d_{G_1\cup G_2}(v) > 10\sigma^*$ to "clean" the spectrum of $A(G)$. We do, however, *not* assume that σ^* is given at the input, but that we are given merely G and k. Furthermore, it is not feasible to just try all possible values of σ^* either. For the algorithm also does not quite know what kind of partition it is looking for, and therefore we could in general not tell from the resulting partition which value of σ^* was correct. For instance, for a wrong value of σ^* the algorithm may easily miss some small planted class V_i but instead split some other big class V_j into two pieces erroneously.

Therefore, Approx pursues the following "adaptive" approach. The algorithm is given the graph G and the matrix $\varphi = \Phi$ (cf. Proposition 4). Thus, the two graphs G_1^*, G_2^* set up in Step 1 coincide with G_1, G_2. Proceeding in $\le \log_2 \Delta$ steps $t = 1, \ldots, \log_2 \Delta$, Step 2 of Approx computes sets R_t of vertices of degree $d_{G_1^*\cup G_2^*}(v) \ge \Delta_t = 2^{-t}\Delta$ and matrices A_t. The A_t's are obtained from $A(G)$ by replacing all entries indexed by $V \times R_t \cup R_t \times V$ by the corresponding entries of φ; the combinatorial meaning is that all edges incident with vertices in R_t get deleted from $G_1 \cup G_2$. Further, to ensure that actually the matrices A_t "converge" to $\mathcal{E}$, Approx aborts the loop as soon as $\|A_s - A_t\|$ gets too large for some $s < t$ (reminiscent of Cauchy's criterion for the convergence of sequences).

Why does this procedure yield a good approximation $\hat{A}$ of $\mathcal{E}$ w.h.p.? Suppose that $\Delta_t > 50\sigma^*$, say. By **H1** and **H3**, the set R_t consists just of $\le n_{\min}/{\sigma^*}^4$ vertices of atypically high degree in $G_1 \cup G_2$. Thus, deleting the vertices R_t removes the eigenvalues caused by the fluctuations of the vertex degrees $> \Delta_t$ while leaving the planted partition essentially intact. Therefore, we can estimate $\|A_t - \mathcal{E}\|$ as follows.

Lemma 16. *Suppose that $\Delta_t \ge 50\lambda$ and that $\varphi = \Phi$. Then $\|A_t - \mathcal{E}\| \le 2c_0k\sqrt{\Delta_t}$.*

Proof. By **A1** we have

$$\|\mathcal{E}_{V\setminus R_t} - A_{V\setminus R_t}\| \le c_0k\sqrt{\Delta_t + \lambda} \le \frac{3}{2}c_0k\sqrt{\Delta_t}. \tag{27}$$

Thus, letting $\mathcal{F} = \mathcal{E}_{R_t} + \mathcal{E}_{R_t\times V\setminus R_t} + \mathcal{E}_{V\setminus R_t\times R_t}$ and $M = \varphi_{R_t} + \varphi_{R_t\times V\setminus R_t} + \varphi_{V\setminus R_t\times R_t}$, we just need to bound the norm of $\mathcal{F} - M = \mathcal{E} - A_t - (\mathcal{E}_{V\setminus R_t} - A_{V\setminus R_t})$. Since $\varphi = \Phi$, we have

$$\|\mathcal{F} - M\|^2 \overset{(8)}{\le} \|\mathcal{F} - M\|_F^2 \le 2\|\mathcal{E}_{R_t\times V} - \Phi_{R_t\times V}\|_F^2$$

$$= 2 \sum_{v \in R_t} \sum_{w \in V} (\Phi_{vw} - p_{\psi(v)\psi(w)})^2 \leq 8 \sum_{v \in R_t} \sum_{w \in V} \left[p_{\psi(v)\psi(w)}(1 - p_{\psi(v)\psi(w)})\right]^2$$
$$= 8 \sum_{a=1}^{k} \sum_{b=1}^{k} \#V_a \cap R_t \cdot \#V_b \cdot [p_{ab}(1-p_{ab})]^2$$
$$\leq 8\sigma^* \sum_{a=1}^{k} \#V_a \cap R_t \sum_{b=1}^{k} p_{ab}(1-p_{ab}) \qquad \text{[because } \sigma^* \geq \#V_b p_{ab}(1-p_{ab})\text{]}$$
$$\leq \frac{8\sigma^{*2}}{n_{\min}} \sum_{a=1}^{k} \#V_a \cap R_t \qquad \text{[because } \sigma^* \geq \textstyle\sum_b n_{\min} p_{ab}(1-p_{ab})\text{]}$$
$$= \frac{8\sigma^{*2}\#R_t}{n_{\min}} \leq \frac{8\sigma^{*2}\#V \setminus H}{n_{\min}} \leq 1 \qquad \text{[by } \mathbf{H1} \text{ and } \mathbf{H3}\text{]}. \tag{28}$$

Combining (27) and (28), we obtain $\|A_t - \mathcal{E}\| \leq \|\mathcal{E}_{V \setminus R_t} - A_{V \setminus R_t}\| + \|\mathcal{F} - M\| \leq 2c_0 k\sqrt{\Delta_t}$. □

Proof of Proposition 5. If `Identify` fails, then $\sigma^* > \ln^3 n$ by Proposition 4, and the matrix $\hat{A}$ defined in Step 2 of `Partition` is a rank k approximation of A. Since $\sigma^* > \ln^3 n$, (12) implies that $d_{G_1 \cup G_2}(v) \leq 20\sigma^*$ for all $v \in V$. Therefore, $\|\mathcal{E} - \hat{A}\| \leq \|\mathcal{E} - A\| + \|\hat{A} - A\| \overset{(9)}{\leq} 2\|\mathcal{E} - A\| \overset{\mathbf{A1}}{\leq} c_0^2 k\sqrt{\lambda}$, as desired.

Let us now assume that `Identify` did not fail, and thus $\varphi = \Phi$. In this case, `Partition` executes `Approx`(G, φ). Let s, t be such that $\Delta_s \geq \Delta_t \geq 50\lambda$. Then by Lemma 16 we have $\|A_s - A_t\| \leq \|A_s - \mathcal{E}\| + \|A_t - \mathcal{E}\| \leq 4c_0 k\sqrt{\Delta_s}$, and thus Step 3 of `Approx` will not abort the loop. Consequently, $\Delta_{\hat{t}} \leq 100\lambda$. Let t^* be maximal such that $\Delta_{t^*} \geq 50\lambda$. Then $\|A_{\hat{t}} - A_{t^*}\| \leq 4c_0 k\sqrt{\Delta_{t^*}}$, because the exit condition in Step 3 of `Approx` was not satisfied for $t = \hat{t}$ and $s = t^*$. Therefore, Lemma 16 entails that

$$\|A_{\hat{t}} - \mathcal{E}\| \leq \|A_{\hat{t}} - A_{t^*}\| + \|A_{t^*} - \mathcal{E}\| \leq 6c_0 k\sqrt{\Delta_{t^*}} \leq 60c_0 k\sqrt{\lambda}. \tag{29}$$

Finally, if $\hat{A}$ is a rank k approximation of $A_{\hat{t}}$, then $\|\hat{A} - A_{\hat{t}}\| \leq \|\mathcal{E} - A_{\hat{t}}\|$ by (9). Therefore, (29) implies $\|\hat{A} - \mathcal{E}\| \leq \|\hat{A} - A_{\hat{t}}\| + \|A_{\hat{t}} - \mathcal{E}\| \leq 2\|A_{\hat{t}} - \mathcal{E}\| \leq 120c_0 k\sqrt{\lambda} \leq c_0^2 k\sqrt{\lambda}$. □

6 Computing an Initial Partition

6.1 The Procedure `Initial`

Algorithm 17. `Initial`$(\hat{A}, k)$
Input: A matrix $\hat{A}$ and the parameter k.
Output: A partition $S_1, \ldots, S_k$ of V and vectors $\xi_1, \ldots, \xi_k \in \mathbf{R}^V$.

1. For $j = 1, \ldots, 2\log n$ do
2. Let $\rho_j = n2^{-j}$ and compute $Q^{(j)}(v) = \{w \in V : \|\hat{A}_w - \hat{A}_v\|^2 \leq 0.01\rho_j^2\}$ for all $v \in V$.
   Then, determine sets $Q_1^{(j)}, \ldots, Q_k^{(j)}$ as follows: for $i = 1, \ldots, k$ do
3. Pick a vertex $v \in V \setminus \bigcup_{l=1}^{i-1} Q_l^{(j)}$ such that $\#Q^{(j)}(v) \setminus \bigcup_{l=1}^{i-1} Q_l^{(j)}$ is maximum.
   Set $Q_i^{(j)} = Q^{(j)}(v) \setminus \bigcup_{l=1}^{i-1} Q_l^{(j)}$ and $\xi_i^{(j)} = \frac{1}{\#Q_i^{(j)}} \sum_{w \in Q_i^{(j)}} \hat{A}_w$.
4. Partition the entire set V as follows.
   - First, let $S_i^{(j)} = Q_i^{(j)}$ for all $1 \leq i \leq k$.
   - Then, add each vertex $v \in V \setminus \bigcup_{l=1}^{k} Q_l^{(j)}$ to a set $S_i^{(j)}$ such that $\|\hat{A}_v - \xi_i^{(j)}\|$ is minimum.

   Set $r_j = \sum_{i=1}^{k} \sum_{v \in S_i^{(j)}} \|\hat{A}_v - \xi_i^{(j)}\|^2$.
5. Let J be such that $r^* = r_J$ is minimum. Return $S_1^{(J)}, \ldots, S_k^{(J)}$ and $\xi_1^{(J)}, \ldots, \xi_k^{(J)}$.

`Initial` is given the approximation $\hat{A}$ of $\mathcal{E}$ and the parameter k, and its goal is to compute a partition of V that is "close" to $V_1, \ldots, V_k$. If the algorithm knew the parameter ρ, then it could partition G as follows. Since the number $z = \#\{v \in V : \|\hat{A}_v - \mathcal{E}_v\|^2 > 0.001\rho^2\}$ satisfies $10^{-3}\rho^2 z \leq \|\hat{A} - \mathcal{E}\|_F^2$, (13) yields $z \le n_{\min}/c_0$. Thus, let $v \in V_i$ be such that $\|\hat{A}_v - \mathcal{E}_v\|^2 \le 10^{-4}\rho^2$, and define $Q(v) = \{w \in V : \|\hat{A}_v - \hat{A}_w\|^2 \le 0.01\rho^2\}$. Since $\mathcal{E}_w = \mathcal{E}_v$ for all $w \in V_i$, we have $\#Q(v) \cap V_i \ge \#V_i - z \ge 0.999\#V_i$. Moreover, $\#Q(v)\setminus V_i \le z$, because of the separation condition **A2**. Thus, $Q(v)$ "almost" coincides with V_i. Hence, we could obtain a good approximation of $V_1, \ldots, V_k$ by just picking iteratively k vertices $v_1, \ldots, v_k$ such that $v_{i+1} \in V \setminus \bigcup_{j=1}^{i-1} Q(v_j)$ and $Q(v_i)$ has maximum cardinality. This is essentially what Steps 2–3 of `Initial` do, and a similar procedure is at the core of McSherry's algorithm [27].

However, since here we do *not* assume that ρ is known to the algorithm, `Initial` has to estimate ρ on its own. To this end, `Initial` applies the above clustering procedure for various "candidate" values $\rho_j = n2^{-j}$, $1 \le j \le 2\log_2 n$. Thus, `Initial` obtains for each j a collection $Q_1^{(j)}, \ldots, Q_k^{(j)}$ of pairwise disjoint subsets of V and vectors $\xi_i^{(j)}$. The idea is that $\xi_i^{(j)}$ should approximate $\mathcal{E}_{V_i}$ well if $Q_i^{(j)}$ is a good approximation of V_i. Hence, if $Q_1^{(j)}, \ldots, Q_k^{(j)}$ is "close" to $V_1, \ldots, V_k$, then $r_j = \sum_{i=1}^k \sum_{v \in S_i^{(j)}} \|\hat{A}_v - \xi_i^{(j)}\|^2 \approx \|\hat{A} - \mathcal{E}\|_F^2$ will be small (cf. (13)). Therefore, the output of `Initial` is just the partition $S_1^{(j)}, \ldots, S_k^{(j)}$ with minmal r_j. In Section 6.2 we shall derive the following bound on this minimum.

Lemma 18. *If $\frac{1}{2}\rho \le \rho_j \le \rho$, then $r_j \le c_0^6 k^3 \lambda$.*

Furthermore, in Section 6.4 we will establish that any partition such that r_j is small yields a good approximation of $V_1, \ldots, V_k$; Proposition 6 is an immediate consequence of Lemmas 18 and 19.

Lemma 19. *Let $S_1, \ldots, S_k$ be a partition and $\xi_1, \ldots, \xi_k$ a sequence of vectors such that $\sum_{i=1}^k \sum_{v \in S_i} \|\xi_i - A_v^*\|^2 \le c_0^6 k^3 \lambda$. Then there is a bijection $\gamma : \{1, \ldots, k\} \to \{1, \ldots, k\}$ such that the following holds.*

1. *$\|\xi_i - \mathcal{E}_{V_{\gamma(i)}}\|^2 \le 0.001\rho^2$ for all $i = 1, \ldots, k$,*
2. *$\sum_{i=1}^k \#S_i \triangle V_{\gamma(i)} < 0.001 n_{\min}$, and*
3. *$\sum_{a,b=1}^k \#S_a \cap V_b \cdot \|\mathcal{E}_{V_{\gamma(a)}} - \mathcal{E}_{V_{\gamma(b)}}\|^2 < 0.001 n_{\min}\rho^2$ for all $1 \le j \le k$.*

6.2 Proof of Lemma 18

Suppose that $\frac{1}{2}\rho \le \rho_j \le \rho$. To ease up the notation, we omit the superscript j; thus, we let $S_i = S_i^{(j)}$, $Q_i = Q_i^{(j)}$ for $1 \le i \le k$, and $Q(v) = Q^{(j)}(v)$ for $v \in V$ (cf. Steps 2–4 of `Initial`). The following lemma, whose proof we postpone to Section 6.3, shows that there is a permutation γ such that ξ_i is "close" to $\mathcal{E}_{V_{\gamma(i)}}$ for all $1 \le i \le k$, and that the sets Q_i are "not too small".

Lemma 20. *Suppose that $\frac{1}{2}\rho \le \rho_j \le \rho$. There is a bijection $\gamma : \{1, \ldots, k\} \to \{1, \ldots, k\}$ such that for each $1 \le i \le k$ we have $\#Q_i \ge \frac{1}{2}\#V_{\gamma(i)}$ and $\|\xi_i - \mathcal{E}_{V_{\gamma(i)}}\|^2 \le 0.1\rho^2$.*

In the sequel, we shall assume without loss of generality that the map γ from Lemma 20 is just the identity, i.e., $\gamma(i) = i$ for all i. Bootstrapping on the estimate $\|\xi_i - \mathcal{E}_{V_i}\|^2 \le 0.1\rho^2$ for $1 \le i \le k$ from Lemma 20, we derive the following stronger estimate.

Corollary 21. *For all $1 \le i \le k$ we have $\|\xi_i - \mathcal{E}_{V_i}\|^2 \le 100\#Q_i^{-1} \sum_{v \in Q_i} \|\hat{A}_v - \mathcal{E}_v\|^2$.*

Proof. By the Cauchy-Schwarz inequality,

$$\|\xi_i - \mathcal{E}_{V_i}\| = \#Q_i^{-1} \left\| \sum_{v \in Q_i} \hat{A}_v - \mathcal{E}_{V_i} \right\| \le \#Q_i^{-1/2} \left[\sum_{v \in Q_i} \|\hat{A}_v - \mathcal{E}_{V_i}\|^2 \right]^{1/2}. \tag{30}$$

Furthermore, as $\|\xi_i - \mathcal{E}_{V_i}\|^2 \le 0.1\rho^2$ by Lemma 20, for all $v \in Q_i \setminus V_i$ we have

$$\|\hat{A}_v - \mathcal{E}_{V_i}\|^2 \le 2(\|\hat{A}_v - \xi_i\|^2 + \|\xi_i - \mathcal{E}_{V_i}\|^2) \le \rho^2/3, \tag{31}$$

because the construction of Q_i in Step 3 of `Initial` ensures that $\|\hat{A}_v - \xi_i\|^2 \le 0.01\rho^2$. Hence, as $\|\mathcal{E}_v - \mathcal{E}_{V_i}\|^2 \ge \rho^2$ by **A2**, (31) implies that $\|\hat{A}_v - \mathcal{E}_v\| \ge 0.1\|\hat{A}_v - \mathcal{E}_{V_i}\|$. Therefore, the assertion follows from (30). □

Corollary 22. *For all $v \in S_i \setminus V_i$ we have $\|\hat{A}_v - \xi_i\| \le 3\|\hat{A}_v - \mathcal{E}_v\|$.*

Proof. Let $i \neq l$ and consider a vertex $v \in S_i \cap V_l$. We shall establish below that

$$\|\hat{A}_v - \xi_i\| \le \|\hat{A}_v - \xi_l\|. \tag{32}$$

Then by Lemma 20 $\|\hat{A}_v - \xi_l\| \le \|\hat{A}_v - \mathcal{E}_v\| + \|\mathcal{E}_v - \xi_l\| \le \|\hat{A}_v - \mathcal{E}_v\| + \rho/3$, and thus $\rho \le \|\mathcal{E}_v - \mathcal{E}_{V_i}\| \le \|\hat{A}_v - \xi_i\| + \|\xi_i - \mathcal{E}_{V_i}\| + \|\hat{A}_v - \mathcal{E}_v\| \le 2\|\hat{A}_v - \mathcal{E}_v\| + \frac{2}{3}\rho$. Consequently, we obtain $\|\hat{A}_v - \mathcal{E}_v\| \ge \frac{1}{6}\rho$, so that the assertion follows from the estimate

$$\|\hat{A}_v - \xi_i\| \overset{(32)}{\le} \|\hat{A}_v - \xi_l\| \le \|\hat{A}_v - \mathcal{E}_v\| + \|\mathcal{E}_v - \xi_l\| \overset{\text{Lemma 20}}{\le} \|\hat{A}_v - \mathcal{E}_v\| + \frac{\rho}{3} \le 3\|\hat{A}_v - \mathcal{E}_v\|.$$

Finally, we prove (32). If $v \in S_i \cap V_l \setminus Q_i$, then the construction of S_i in Step 4 of `Initial` guarantees that $\|\hat{A}_v - \xi_i\| \le \|\hat{A}_v - \xi_l\|$, as claimed. Thus, assume that $v \in Q_i \cap V_l$. Then

$$\begin{aligned} \|\hat{A}_v - \xi_i\| &\le 0.15\rho && \text{[by the definition of } Q_i \text{ in Step 3 of Initial]},\\ \max\{\|\xi_i - \mathcal{E}_{V_i}\|, \|\xi_l - \mathcal{E}_v\|\} &\le \frac{1}{3}\rho && \text{[by Lemma 20]},\\ \|\mathcal{E}_{V_i} - \mathcal{E}_v\| &\ge \rho && \text{[by } \mathbf{A2}]. \end{aligned}$$

Therefore, if $\|\hat{A}_v - \xi_l\| < \|\hat{A}_v - \xi_i\|$, then we would arrive at the contradiction

$$\begin{aligned} \rho &\le \|\mathcal{E}_{V_i} - \mathcal{E}_v\| \le \|\mathcal{E}_{V_i} - \xi_i\| + \|\mathcal{E}_v - \xi_l\| + \|\xi_i - \xi_l\| \\ &\le \frac{2}{3}\rho + \|\hat{A}_v - \xi_i\| + \|\hat{A}_v - \xi_l\| < \frac{2}{3}\rho + 2\|\hat{A}_v - \xi_i\| \le 0.99\rho. \end{aligned}$$

Thus, we conclude that $\|\hat{A}_v - \xi_l\| \ge \|\hat{A}_v - \xi_i\|$, thereby completing the proof. □

Proof of Lemma 18. Since $\#Q_i \ge \frac{1}{2}\#V_i$ by Lemma 20, we have the estimate

$$\begin{aligned} \sum_{i=1}^{k} \sum_{w \in S_i \cap V_i} \|\hat{A}_w - \xi_i\|^2 &\le 2\sum_{i=1}^{k} \sum_{w \in S_i \cap V_i} \left[\|\hat{A}_w - \mathcal{E}_w\|^2 + \|\mathcal{E}_w - \xi_i\|^2\right] \\ &\overset{\text{Cor. 21}}{\le} 2\|\hat{A} - \mathcal{E}\|_F^2 + 200\sum_{i=1}^{k} \frac{\#S_i \cap V_i}{\#Q_i} \sum_{v \in Q_i} \|\hat{A}_v - \mathcal{E}_v\|^2 \le 500\|\hat{A} - \mathcal{E}\|_F^2. \end{aligned} \tag{33}$$

Furthermore, by Corollary 22

$$\sum_{i=1}^{k} \sum_{v \in S_i \setminus V_i} \|\hat{A}_v - \xi_i\|^2 \le 9\sum_{i=1}^{k} \sum_{v \in S_i \setminus V_i} \|\hat{A}_v - \mathcal{E}_v\|^2 \le 9\|\hat{A} - \mathcal{E}\|_F^2. \tag{34}$$

Since $\|\hat{A} - \mathcal{E}\|_F^2 \le c_0^5 k^3 \lambda$ by (13), the bounds (33) and (34) imply the assertion. □

6.3 Proof of Lemma 20

For $1 \le i \le k$ we choose $\gamma(i)$ so that $\#Q_i \cap V_{\gamma(i)}$ is maximum. We shall prove below that for all $1 \le l \le k$ we have

$$\|\xi_l - \mathcal{E}_{V_{\gamma(l)}}\|^2 \le 0.1\rho^2, \tag{35}$$

$$\#Q_l \ge \max\{\#V_i : i \in \{1, \ldots, k\} \setminus \gamma(\{1, \ldots, l-1\})\} - 0.01 n_{\min}, \tag{36}$$

$$\#Q_l \cap V_{\gamma(l)} \ge \#Q_l - 0.01 n_{\min}. \tag{37}$$

These three inequalities imply the assertion. To see that γ is a bijection, let us assume that $\gamma(l) = \gamma(l')$ for two indices $1 \le l < l' \le k$. Indeed, suppose that $l = \min \gamma^{-1}(l)$. Then $\#Q_l \ge \#V_{\gamma(l)} - 0.01 n_{\min}$ by (36), and thus $\#V_{\gamma(l)} \setminus Q_l \le 0.1 n_{\min}$ by (37). Therefore, we obtain the contradiction

$$0.99 n_{\min} \overset{(36)}{\le} \#Q_{l'} \overset{(37)}{\le} 1.1\#Q_{l'} \cap V_{\gamma(l)} \le 1.1\#V_{\gamma(l)} \setminus Q_l \le 0.11 n_{\min}.$$

Finally, as γ is bijective, (36) entails that $\#Q_l \geq 0.9V_{\gamma(l)}$ for all $1 \leq l \leq k$. Hence, due to (37) we obtain $\#Q_l \cap V_l \geq 0.9\#Q_l \geq \frac{1}{2}\#V_{\gamma(l)}$, as desired.

The remaining task is to establish (35)–(37). We proceed by induction on l. Thus, let us assume that (35)–(37) hold for all $l < L$; we are to show that then (35)–(37) are true for $l = L$ as well. As a first step, we establish (36). To this end, consider a class V_i such that $i \notin \gamma(\{1, \ldots, L-1\})$ and let $Z_i = \{v \in V_i : \|\hat{A}_v - \mathcal{E}_v\|^2 \leq 0.001\rho^2\}$. Then $0.001\rho^2(\#V_i - \#Z_i) \leq \sum_{v \in V_i \setminus Z_i} \|\hat{A}_v - \mathcal{E}_v\|^2 \leq \|\hat{A} - \mathcal{E}\|_F^2 \stackrel{(13)}{\leq} c_0^5 k^3 \lambda$, whence the definition $\rho^2 = c_0^8 k^3 \lambda / n_{\min}$ of ρ (cf. **A2**) yields

$$\#Z_i \geq \#V_i - 0.01n_{\min}. \tag{38}$$

Moreover, for all $v \in Z_i$ we have

$$Q(v) = \{w \in V : \|\hat{A}_v - \hat{A}_w\|^2 \leq 0.01\rho^2\} \supset Z_i. \tag{39}$$

In addition, let $w \in Q_l$ for some $l < L$; since our choice of i ensures that $v \in V_i \neq V_{\gamma(l)}$, we have

$$\rho \stackrel{\mathbf{A2}}{\leq} \|\mathcal{E}_{V_{\gamma(l)}} - \mathcal{E}_v\| \leq \|\mathcal{E}_v - \hat{A}_v\| + \|\hat{A}_w - \hat{A}_v\| + \|\xi_l - \hat{A}_w\| + \|\xi_l - \mathcal{E}_{V_{\gamma(l)}}\|. \tag{40}$$

Now, the construction in Step 3 of `Initial` ensures that $\|\hat{A}_w - \xi_l\| \leq 0.1\rho$. Furthermore, $\|\xi_l - \mathcal{E}_{V_{\gamma(l)}}\| \leq \rho/3$ by induction (cf. (35)), and $\|\hat{A}_v - \mathcal{E}_v\| \leq 0.1\rho$, because $v \in Z_i$. Hence, (40) entails that $\|\hat{A}_w - \hat{A}_v\| > 0.1\rho$, so that $w \notin Q(v)$. Consequently, (39) yields

$$Z_i \cap Q_l = \emptyset \text{ for all } l < L. \tag{41}$$

Finally, let v_L signify the vertex chosen by Step 3 of `Initial` to construct Q_L. Then by construction $\#Q_L = \#Q(v_L) \setminus \bigcup_{l=1}^{L-1} Q_l \geq \#Q(v) \setminus \bigcup_{l=1}^{L-1} Q_l$. Therefore,

$$\#Q_L \geq \#Q(v) \setminus \bigcup_{l=1}^{L-1} Q_l \stackrel{(39),(41)}{\geq} \#Z_i \stackrel{(38)}{\geq} \#V_i - 0.01n_{\min}.$$

As this estimate holds for all $i \notin \gamma(\{1, \ldots, L-1\})$, (36) follows.

Thus, we know that Q_L is "big". As a next step, we prove (37), i.e., we show that Q_L "mainly" consists of vertices in $V_{\gamma(L)}$. To this end, let $1 \leq i \leq k$ be such that $\|\mathcal{E}_{V_i} - \hat{A}_{v_L}\|$ is minimum. Let $Y = Q_L \setminus V_i$. Then for all $w \in Y$ we have $\|\mathcal{E}_w - \hat{A}_{v_L}\| \geq \|\mathcal{E}_{V_i} - \hat{A}_v\|$. Further, since by **A2** $\rho \leq \|\mathcal{E}_w - \mathcal{E}_{V_i}\| \leq \|\mathcal{E}_w - \hat{A}_{v_L}\| + \|\mathcal{E}_{V_i} - \hat{A}_{v_L}\| \leq 2\|\mathcal{E}_w - \hat{A}_{v_L}\|$, we conclude that $\|\mathcal{E}_w - \hat{A}_{v_L}\|^2 \geq \frac{1}{4}\rho^2$. On the other hand, as $w \in Q_L$, we have $\|\hat{A}_w - \hat{A}_{v_L}\|^2 \leq 0.01\rho^2$. Therefore, we obtain $\|\hat{A}_w - \mathcal{E}_w\|^2 \geq 0.1\rho^2$ for all $w \in Y$, so that

$$0.1\#Y\rho^2 \leq \sum_{w \in Y} \|\hat{A}_w - \mathcal{E}_w\|^2 \leq \|\hat{A} - \mathcal{E}\|_F^2 \stackrel{(13)}{\leq} c_0^5 k^3 \lambda. \tag{42}$$

As $\rho^2 = c_0^8 k^3 \lambda / n_{\min}$ (cf. **A2**), (42) yields that $\#Y < 0.01n_{\min}$. Consequently, (36) entails that $\#V_i \cap Q_L \geq 0.99\#Q_L$, so that $i = \gamma(L)$. Hence, we obtain $\#Q_L \cap V_{\gamma(L)} = \#Q_L \cap V_i = \#Q_L \setminus Y \geq \#Q_L - 0.01n_{\min}$, thereby establishing (37).

Finally, to show (35), we note that by construction $\|\xi_L - \hat{A}_{v_L}\|^2 \leq 0.01\rho^2$ and $\|\hat{A}_w - \hat{A}_{v_L}\|^2 \leq 0.01\rho^2$ for all $w \in Q_L \cap V_{\gamma(L)}$ (cf. Step 3 of `Initial`). Therefore,

$$\begin{aligned} \#Q_L \cap V_{\gamma(L)} \|\mathcal{E}_{\gamma(L)} - \xi_L\|^2 &\leq 3 \sum_{w \in Q_L \cap V_{\gamma(L)}} \|\xi_L - \hat{A}_{v_L}\|^2 + \|\hat{A}_w - \hat{A}_{v_L}\|^2 + \|\hat{A}_w - \hat{\mathcal{E}}_{\gamma(L)}\|^2 \\ &\leq 0.06\rho^2 \#Q_L \cap V_{\gamma(L)} + 3\|\hat{A} - \mathcal{E}\|_F^2 \stackrel{(13)}{\leq} 0.06\rho^2 \#Q_L \cap V_{\gamma(L)} + 3c_0^5 k^3 \lambda. \end{aligned} \tag{43}$$

Since $\#Q_L \cap V_{\gamma(L)} \geq 0.9n_{\min}$ due to (36) and (37), and because $\rho^2 = c_0^8 k^3 \lambda / n_{\min}$ (cf. **A2**), (43) entails that $\|\mathcal{E}_{\gamma(L)} - \xi_L\|^2 \leq 0.07\rho^2 + \frac{6c_0^5 k^3 \lambda}{n_{\min}} \leq 0.1\rho^2$. Thus, (35) follows.

6.4 Proof of Lemma 19

Set $S_{ab} = S_a \cap V_b$ for $1 \le a, b \le k$. Moreover, for each $1 \le a \le k$ let $1 \le \gamma(a) \le k$ be such that $\|\mathcal{E}_{V_{\gamma(a)}} - \xi_a\|$ is minimum. Then for all $b \ne \gamma(a)$ we have

$$\rho \le \|\mathcal{E}_{V_{\gamma(a)}} - \mathcal{E}_{V_b}\| \le \|\mathcal{E}_{V_{\gamma(a)}} - \xi_a\| + \|\mathcal{E}_{V_b} - \xi_a\| \le 2\|\mathcal{E}_{V_b} - \xi_a\|, \tag{44}$$

so that $\|\mathcal{E}_{V_b} - \xi_a\| \ge \rho/2$. Therefore, by our assumption that $\sum_{i=1}^k \sum_{v\in S_i} \|\xi_i - \hat{A}_v\|^2 \le c_0^6 k^3 \lambda$, we have

$$\frac{\rho^2}{4}\sum_{a=1}^k \sum_{1\le b\le k: b\ne\gamma(a)} \#S_{ab} \le \sum_{a,b=1}^k \#S_{ab}\|\mathcal{E}_{V_b} - \xi_a\|^2 \le 2\sum_{a,b=1}^k \sum_{v\in S_{ab}} \|\mathcal{E}_v - \hat{A}_v\|^2 + \|\hat{A}_v - \xi_a\|^2$$
$$\le\ 2\|\hat{A} - \mathcal{E}\|_F^2 + 2\sum_{a,b=1}^k \sum_{v\in S_{ab}} \|\hat{A}_v - \xi_a\|^2 \stackrel{(13)}{\le} 4c_0^5 k^3\lambda + 2c_0^6 k^3 \lambda \le c_0^7 k^3\lambda. \tag{45}$$

Hence, **A2** entails that

$$\sum_{a=1}^k \#S_a \triangle V_{\gamma(a)} = \sum_{1\le a,b\le k: b\ne\gamma(a)} 2\#S_{ab} \le \frac{8c_0^7 k^3\lambda}{\rho^2} \le 0.001 n_{\min}. \tag{46}$$

Combining (45) and (46), we obtain $\frac{n_{\min}}{2}\|\mathcal{E}_{V_{\gamma(a)}} - \xi_a\|^2 \le \#S_a \cap V_{\gamma(a)}\|\mathcal{E}_{\gamma(a)} - \xi_a\|^2 \le c_0^7 k^3 \lambda$, whence

$$\|\mathcal{E}_{\gamma(a)} - \xi_a\|^2 \le \frac{2c_0^7 k^3 \lambda}{n_{\min}} \stackrel{\mathbf{A2}}{\le} 0.001\rho^2 \quad \text{for all } 1 \le a \le k. \tag{47}$$

Thus, we have established the first two parts of the lemma. In addition, observe that (46) implies that γ is bijective (because the sets $S_1, \ldots, S_k$ are pairwise disjoint and $\#V_a \ge n_{\min}$ for all $1 \le a \le k$). Finally, the third assertion follows from the estimate

$$\sum_{a,b=1}^k \#S_{ab}\|\mathcal{E}_{V_{\gamma(a)}} - \mathcal{E}_{V_{\gamma(b)}}\|^2 \ \le\ 2\sum_{a,b=1}^k \#S_{ab}\left(\|\mathcal{E}_{V_{\gamma(a)}} - \xi_a\|^2 + \|\mathcal{E}_{V_{\gamma(b)}} - \xi_a\|^2\right)$$
$$\stackrel{(44)}{\le} 8\sum_{a,b=1}^k \#S_{ab}\|\mathcal{E}_{V_{\gamma(b)}} - \xi_a\|^2 \stackrel{(45)}{\le} 8c_0^7 k^3 \lambda \stackrel{\mathbf{A2}}{\le} 0.001\rho^2 n_{\min}.$$

7 Local Improvement

7.1 The Procedure `Improve`

Having computed the initial partition $S_1, \ldots, S_k$ with the "centers" $\xi_1, \ldots, \xi_k$, finally `Partition` calls the procedure `Improve` to home in on the planted partition $V_1, \ldots, V_k$ on the subgraph H. In contrast to the previous steps of `Partition`, `Improve` does not rely on spectral methods anymore but just performs a "local" combinatorial procedure.

Algorithm 23. `Improve`$(G, S_1, \ldots, S_k, \xi_1, \ldots, \xi_k)$
Input: The graph $G = (V, E)$, a partition $S_1, \ldots, S_k$ of V, and vectors $\xi_1, \ldots, \xi_k$.
Output: A partition of G.

1. Repeat the following $\lceil \log_2 n \rceil$ times:
2. For all $v \in V$, all $l = 1, \ldots, k$, and all $w \in S_l$ compute the numbers $\delta(v, w) = e(v, S_l)/\#S_l$.
   Let $\delta(v) = (\delta(v, w))_{w\in V} \in \mathbf{R}^V$.
   For all $v \in V$ pick $1 \le \gamma(v) \le k$ such that $\|\delta(v) - \xi_{\gamma(v)}\| = \min_{1\le i\le k} \|\delta(v) - \xi_i\|$ (ties are broken arbitrarily). Then, update $S_i = \gamma^{-1}(i)$ for $i = 1, \ldots, k$.
3. Return the partition $S_1, \ldots, S_k$.

The basic idea behind `Improve` is to compare for each vertex v the actual values $e(v, S_i)$ with the expected values $\mu(v, V_i)$, where the latter are approximated by the entries of ξ_i. More precisely, for each vertex v `Improve` sets up the vector $\delta(v)$ that encodes the densities $e(v, S_i)/\#S_i$. Then, `Improve` updates the partition $S_1, \ldots, S_k$ by putting each vertex v into that class S_j such that $\|\delta(v) - \xi_j\|$ is minimum.

To analyze this procedure, we need a few definitions. For a partition $\mathcal{S} = (\mathcal{S}_1, \ldots, \mathcal{S}_k)$ and a vertex $v \in V$, we define a vector $\delta_{\mathcal{S}}(v) = (\delta_{\mathcal{S}}(v, w))_{w \in V}$ by letting $\delta_{\mathcal{S}}(v, w) = e(v, \mathcal{S}_j)/\#\mathcal{S}_j$ for all $w \in \mathcal{S}_j$ and all $1 \leq j \leq k$; we shall omit the index $\mathcal{S}$ if it is clear from the context. Moreover, we call a partition $\mathcal{R} = (\mathcal{R}_1, \ldots, \mathcal{R}_k)$ an *improvement* of $\mathcal{S}$ if for all $i = 1, \ldots, k$ and all $v \in \mathcal{R}_i$ we have $\|\delta_{\mathcal{S}}(v) - \xi_i\| = \min_{1 \leq j \leq k} \|\delta_{\mathcal{S}}(v) - \xi_j\|$. Thus, each step of `Improve` just computes an improvement $\mathcal{R}$ of the previous partition $\mathcal{S}$.

Furthermore, we say that $\mathcal{S}$ is *feasible* if $\frac{1}{2}\#V_i \leq \#\mathcal{S}_i \leq 2\#V_i$ for all i. In addition, we set $\mathcal{S}_{ij} = \mathcal{S}_i \cap V_j$ and call $\mathcal{S}$ *tight* if $\sum_{i \neq j} \#\mathcal{S}_{ij} \|\mathcal{E}_{V_i} - \mathcal{E}_{V_j}\|^2 \leq 0.001\rho^2 n_{\min}$. Then Proposition 6 entails that the initial partition $\mathcal{S} = (S_1, \ldots, S_k)$ given to `Improve` as an input is both feasible and tight. Therefore, Proposition 7 will follow from the next two lemmas, which we shall prove in Sections 7.2 and 7.3.

Lemma 24. *If $\mathcal{S}$ is feasible and tight, then any improvement $\mathcal{R}$ of $\mathcal{S}$ is tight.*

Lemma 25. *Suppose that $\mathcal{S}$ is feasible and tight and that $\mathcal{R}$ is an improvement of $\mathcal{S}$. Then we have $\sum_{i \neq j} \#\mathcal{R}_{ij} \cap H \leq \frac{1}{10} \sum_{i \neq j} \#\mathcal{S}_{ij} \cap H$.*

Proof of Proposition 7. Let $\mathcal{S} = (\mathcal{S}_1, \ldots, \mathcal{S}_k)$ be a feasible and tight partition such that $\sum_{i=1}^k \#\mathcal{S}_i \triangle V_i \leq 0.001 n_{\min}$,and let $\mathcal{R}$ be an improvement of $\mathcal{S}$. Then by Lemma 24 $\mathcal{R}$ is tight, and by Lemma 25 we have $\sum_{i=1}^k \#\mathcal{R}_i \triangle V_i \leq 0.1 \sum_{i=1}^k \#\mathcal{S}_i \triangle V_i \leq 10^{-4} n_{\min}$, whence $\mathcal{R}$ is feasible. Thus, as the partition $(S_1, \ldots, S_k)$ that `Improve` starts with is feasible and tight by Proposition 6 and in fact satisfies $\sum_{i=1}^k \#S_i \triangle V_i \leq 0.001 n_{\min}$, all the partitions generated by `Improve` remain feasible and tight. Finally, let $\mathcal{T}$ denote the partition returned by `Improve`. Then due to Lemma 25 we have $\sum_{i \neq j} \#\mathcal{T}_{ij} \cap H = 0$, whence $\mathcal{T}_i \cap H = V_i \cap H$ for all $i = 1, \ldots, k$. □

To facilitate the proof of Lemmas 24 and 25, we introduce some notation. Let $A = A(G)$ and

$$M = \mathcal{E}_H - A_H. \tag{48}$$

Then by **A1** and **H3** we have the bound

$$\|M\| \leq c_0^2 k \sqrt{\lambda}. \tag{49}$$

Moreover, for a set $S \subset V$ and a vertex $v \in V$ we let $\mu'(v, S) = \langle \mathcal{E}_v, \mathbf{1}_S \rangle$. Then

$$\|M\mathbf{1}_S\|^2 = \sum_{v \in H} |e(v, S) - \mu'(v, S)|^2 \qquad \text{for all } S \subset V(H). \tag{50}$$

The relation between $\mu(v, S)$ and $\mu'(v, S)$ is that $\mu'(v, S) = \mu(v, S) + p_{\psi(v)\psi(v)}$ if $v \in S$, and $\mu(v, S) = \mu'(v, S)$ if $v \notin S$. If $\mathcal{S} = (\mathcal{S}_1, \ldots, \mathcal{S}_k)$ is a partition of V, then for $v \in V_i$ and $w \in \mathcal{S}_l$ we set

$$\bar{\delta}(v, w) = \frac{\mu'(v, \mathcal{S}_l)}{\#\mathcal{S}_l} \text{ and } \bar{\delta}(v) = (\bar{\delta}(v, w))_{w \in V}.$$

Thus, $\bar{\delta}(v) \in \mathbf{R}^V$ basically is the "expectation" of $\delta(v)$.

7.2 Proof of Lemma 24

For $v \in V$ and $S \subset V$ we let $\mu'_1(v, S) = \sum_{w \in S: \Phi_{vw}=0} \mathcal{E}_{vw}$, $\mu'_2(v, S) = \sum_{w \in S: \Phi_{vw}=1} 1 - \mathcal{E}_{vw}$. We shall prove that if $\mathcal{S}$ is feasible and tight and $\mathcal{R}$ is an improvement of $\mathcal{S}$, then the two inequalities

$$\sum_{a,b=1}^k \#\mathcal{R}_{ab} \|\mathcal{E}_a - \mathcal{E}_b\|^2 \leq 9 \sum_{v \in V} \|\delta(v) - \bar{\delta}(v)\|^2 \leq 0.001\rho^2 n_{\min} \tag{51}$$

hold, so that Lemma 24 follows. Observe that by the definitions of $\delta(v)$, $\bar{\delta}(v)$

$$\sum_{v\in V} \|\delta(v) - \bar{\delta}(v)\|^2 = \sum_{a=1}^{k} \#\mathcal{S}_a^{-1}(e(v,\mathcal{S}_a) - \mu'(v,\mathcal{S}_a))^2. \tag{52}$$

Lemma 26. *Let $\mathcal{S}$ be a feasible partition. Then $\sum_{v\in H} \|\delta(v) - \bar{\delta}(v)\|^2 \le 10^{-4}\rho^2 n_{\min}$.*

To prove Lemma 26, we need the following estimate.

Lemma 27. *For any set $S \subset V$ and any $v \in V$ we have*

$$|e(v,S) - \mu'(v,S)| \le |e_1(v,S) - \mu_1'(v,S) - (e_2(v,S) - \mu_2'(v,S))| + 1.$$

Moreover, if $v \notin S$, then $e(v,S) - \mu'(v,S) = e_1(v,S) - \mu_1'(v,S) - (e_2(v,S) - \mu_2'(v,S))$.

Proof. Let $S_1 = \{w \in S : \Phi_{vw} = 0\}$ and $S_2 = \{w \in S : \Phi_{vw} = 1\}$. Moreover, let $\iota = 1$ if $v \in S_2$ and $\iota = 0$ otherwise. Then by the definition of the graphs G_1, G_2 we have

$$\begin{aligned} e(v,S) - \mu'(v,S) &= e_1(v,S_1) - \mu_1'(v,S_1) + (\#S_2 - \iota - e_2(v,S_2)) - (\#S_2 - \mu_2'(v,S_2)) \\ &= e_1(v,S) - \mu_1'(v,S) - (e_2(v,S) - \mu_2'(v,S)) - \iota, \end{aligned}$$

whence the assertion follows. □

Proof of Lemma 26. Let $A(v) = \sum_{a=1}^{k} \#\mathcal{S}_a^{-1}\left[e(v,\mathcal{S}_a\cap H) - \mu'(v,\mathcal{S}_a\cap H)\right]^2$. Then

$$\sum_{v\in H} A(v) \overset{(50)}{=} \sum_{a=1}^{k} \#\mathcal{S}_a^{-1}\|M\mathbf{1}_{\mathcal{S}_a\cap H}\|^2 \le \|M\|^2 \sum_{a=1}^{k} \frac{\|\mathbf{1}_{\mathcal{S}_a\cap H}\|^2}{\#\mathcal{S}_a} \le k\|M\|^2 \overset{(49)}{\le} c_0^2 k^3 \lambda \overset{\mathbf{A2}}{\le} 10^{-5}\rho^2 n_{\min}. \tag{53}$$

Furthermore, set

$$A'(v) = \sum_{a=1}^{k} \#\mathcal{S}_a^{-1}((e(v,\mathcal{S}_a) - \mu'(v,\mathcal{S}_a))^2 - (e(v,\mathcal{S}_a\cap H) - \mu'(v,\mathcal{S}_a\cap H))^2) =$$

$$\sum_{a=1}^{k} \#\mathcal{S}_a^{-1}\left[e(v,\mathcal{S}_a) - \mu'(v,\mathcal{S}_a) + e(v,\mathcal{S}_a\cap H) - \mu'(v,\mathcal{S}_a\cap H)\right]\left[e(v,\mathcal{S}_a\setminus H) - \mu'(v,\mathcal{S}_a\setminus H)\right]. \tag{54}$$

Then Lemma 27 entails that for all $v \in H$ and all $1 \le a \le k$ we have

$$\begin{aligned} |e(v,\mathcal{S}_a) - \mu'(v,\mathcal{S}_a) + e(v,\mathcal{S}_a\cap H) - \mu'(v,\mathcal{S}_a\cap H)| &\le 2 + 2\sum_{i=1}^{2} e_{G_i}(v,\mathcal{S}_a) + \mu_i'(v,\mathcal{S}_a) \\ &= 2 + 2d_{G_1\cup G_2}(v) + 2(\mu_1'(v,V) + \mu_2'(v,V)) \overset{\mathbf{H3}}{\le} 25\sigma^* \overset{\mathbf{A1}}{\le} 25\lambda, \end{aligned} \tag{55}$$

because $\mu_1'(v,V) + \mu_2'(v,V) \le \mu'(v,V) \le 2\sigma^*$ by the definition of σ^*. Applying Lemma 27 once more, we obtain

$$\begin{aligned} \sum_{v\in H}\sum_{a=1}^{k} |e(v,\mathcal{S}_a\setminus H) - \mu'(v,\mathcal{S}_a\setminus H)| &\le \sum_{v\in H}\sum_{a=1}^{k}\sum_{i=1}^{2} e_{G_i}(v,\mathcal{S}_a\setminus H) + \mu_i'(v,\mathcal{S}_a\setminus H) \\ &= \sum_{i=1}^{2} e_{G_i}(H, V\setminus H) + \sum_{v\in H} \mu_i'(v, V\setminus H) \le \sum_{i=1}^{2} e_{G_i}(H, V\setminus H) + \sum_{w\in V\setminus H} \mu_i'(w,V) \\ &\le e_{G_1\cup G_2}(H, V\setminus H) + \le 2\sigma^*\cdot \#V\setminus H + \sum_{v\in V\setminus H} d_{G_1\cup G_2}(v) \\ &\le 2\sigma^*\cdot \#V\setminus H + \sqrt{\#V\setminus H \sum_{v\in V\setminus H} d_{G_1\cup G_2}(v)} \qquad \text{[by Cauchy-Schwarz]} \\ &\overset{\mathbf{H1}}{\le} (2\sigma^*\lambda^{-4} + \lambda^{-2})n_{\min} \overset{\mathbf{A1}}{\le} 2\lambda^{-2} n_{\min}. \end{aligned} \tag{56}$$

Finally, as $(\mathcal{S}_a)_{1\le a\le k}$ is feasible, we have $\#\mathcal{S}_a \ge \frac{1}{2}n_{\min}$ for all a. Therefore, plugging (55) and (56) into (54), we obtain

$$\sum_{v\in H} A'(v) \le \frac{75 n_{\min}}{\lambda^2 \min_{1\le a\le k}\#\mathcal{S}_a} \le \frac{150}{\lambda^2} \le 1 \tag{57}$$

by our assumption that $\lambda \ge \sigma^* \ge c_0$ for some large enough $c_0 > 0$. Combining (52), (53), and (57), we obtain the desired estimate. □

Lemma 28. *Let $\mathcal{S}$ be a feasible partition. Then $\sum_{v\notin H}\sum_{a=1}^{k}\#\mathcal{S}_a^{-1}(e(v,\mathcal{S}_a)-\mu'(v,\mathcal{S}_a))^2 \le 2$.*

Proof. We decompose $V\setminus H$ into two parts $U_1 = \{v\in V\setminus H : \max_{i=1,2} e_{G_i}(v,V) \le 100\sigma^*\}$, $U_2 = \{v\in V\setminus H : \max_{i=1,2} e_{G_i}(v,V) > 100\sigma^*\}$. Since $\mathcal{S}$ is feasible, we have $\#\mathcal{S}_a \ge \frac{1}{2}n_{\min}$ for all a. Moreover, the definition of σ^* ensures that $\mu'(v,V) \le 2\sigma^*$ for all v. Therefore,

$$\sum_{v\in U_1}\sum_{a=1}^{k}\#\mathcal{S}_a^{-1}(e(v,\mathcal{S}_a)-\mu'(v,\mathcal{S}_a))^2 \le \sum_{v\in U_1}\frac{10^3\sigma^{*\,2}}{n_{\min}} \le \frac{10^3\sigma^{*\,2}\#V\setminus H}{n_{\min}} \overset{\textbf{H1, A1}}{\le} 1. \tag{58}$$

Further, by Lemma 27 and the feasibility of $\mathcal{S}$

$$\sum_{v\in U_2}\sum_{a=1}^{k}\#\mathcal{S}_a^{-1}(e(v,\mathcal{S}_a)-\mu'(v,\mathcal{S}_a))^2 \le \sum_{v\in U_2}\frac{2d_{G_1\cup G_2}(v)^2}{n_{\min}} \overset{\textbf{H1}}{\le} 2. \tag{59}$$

Finally, the assertion follows from (58) and (59). □

Combining Lemmas 26 and 28, we obtain the right inequality in (51). To prove the left one, the following lemma is instrumental.

Lemma 29. *Let $\mathcal{S}$ be any partition. Then for all $1\le i\le k$ and all $v\in V_i$ we have $\|\bar{\delta}(v)-\mathcal{E}_v\|^2 \le 4\sum_{a,b=1}^{k}\#\mathcal{S}_{ab}(p_{ia}-p_{ib})^2$.*

Proof. Let $\tilde{\delta}(v,w) = p_{il}$ for all $w\in\mathcal{S}_l$, and set $\tilde{\delta}(v) = (\tilde{\delta}(v,w))_{w\in V}$. Then

$$\|\tilde{\delta}(v)-\mathcal{E}_v\|^2 = \sum_{a,b=1}^{k}\#\mathcal{S}_{ab}(p_{ia}-p_{ib})^2. \tag{60}$$

Moreover,

$$\begin{aligned}
\|\tilde{\delta}(v)-\bar{\delta}(v)\|^2 &= \sum_{a=1}^{k}\#\mathcal{S}_a\left[\mu'(v,\mathcal{S}_a)\#\mathcal{S}_a^{-1}-p_{ia}\right]^2 = \sum_{a=1}^{k}\#\mathcal{S}_a^{-1}\left[\mu'(v,\mathcal{S}_a)-\#\mathcal{S}_a p_{ia}\right]^2 \\
&= \sum_{a=1}^{k}\#\mathcal{S}_a^{-1}\left[\sum_{b=1}^{k}\mu'(v,\mathcal{S}_{ab})-\#\mathcal{S}_{ab}p_{ia}\right]^2 = \sum_{a=1}^{k}\#\mathcal{S}_a^{-1}\left[\sum_{b=1}^{k}\#\mathcal{S}_{ab}(p_{ib}-p_{ia})\right]^2 \\
&\le \sum_{a,b=1}^{k}\#\mathcal{S}_{ab}(p_{ia}-p_{ib})^2 \qquad \text{[by Cauchy-Schwarz]}.
\end{aligned} \tag{61}$$

Combining (60) and (61) completes the proof. □

Corollary 30. *If $\mathcal{S}$ is tight, then $\sum_{a,b=1}^{k}\#\mathcal{R}_{ab}\|\mathcal{E}_a-\mathcal{E}_b\|^2 \le 9\sum_{v\in V}\|\delta(v)-\bar{\delta}(v)\|^2$.*

Proof. Since $\|\xi_b - \mathcal{E}_b\|^2 \le 0.001\rho^2$ by Proposition 6, Lemma 29 entails that for all $v \in \mathcal{R}_{ab}$

$$\begin{aligned}
\|\delta(v) - \xi_b\| &\le \|\delta(v) - \mathcal{E}_b\| + \|\mathcal{E}_b - \xi_b\| \le \|\delta(v) - \bar{\delta}(v)\| + \|\bar{\delta}(v) - \mathcal{E}_b\| + \rho/30 \\
&\le \|\delta(v) - \bar{\delta}(v)\| + \frac{\rho}{30} + \sqrt{2\sum_{\alpha\neq\beta} \#\mathcal{S}_{\alpha\beta}(p_{b\alpha} - p_{b\beta})^2} \\
&\le \|\delta(v) - \bar{\delta}(v)\| + \frac{\rho}{30} + \sqrt{2\sum_{\alpha\neq\beta} \frac{\#\mathcal{S}_{\alpha\beta}}{\#V_b}\|\mathcal{E}_\alpha - \mathcal{E}_\beta\|^2} \\
&\le \|\delta(v) - \bar{\delta}(v)\| + \frac{\rho}{30} + \sqrt{\frac{2}{n_{\min}}\sum_{\alpha\neq\beta} \#\mathcal{S}_{\alpha\beta}\|\mathcal{E}_\alpha - \mathcal{E}_\beta\|^2} \\
&\le \|\delta(v) - \bar{\delta}(v)\| + \frac{\rho}{20} \qquad \text{[because } \mathcal{S} \text{ is tight].} \qquad (62)
\end{aligned}$$

Furthermore, if $v \in \mathcal{R}_{ab}$, then $v \in V_b$ but $\|\delta(v) - \xi_a\| \le \|\delta(v) - \xi_b\|$. Since $\|\xi_a - \mathcal{E}_a\|^2, \|\xi_b - \mathcal{E}_b\|^2 \le 0.001\rho^2$ by Proposition 6 and $\|\mathcal{E}_a - \mathcal{E}_b\|^2 \ge \rho^2$ by **A2**, we obtain

$$\begin{aligned}
\rho &\le \|\mathcal{E}_a - \mathcal{E}_b\| \le \|\mathcal{E}_a - \xi_a\| + \|\mathcal{E}_b - \xi_b\| + \|\xi_a - \xi_b\| \\
&\le \frac{\rho}{15} + \|\delta(v) - \xi_a\| + \|\delta(v) - \xi_b\| \le \frac{\rho}{15} + 2\|\delta(v) - \xi_b\|,
\end{aligned}$$

and thus $\|\delta(v) - \xi_b\| \ge \frac{2}{5}\|\mathcal{E}_a - \mathcal{E}_b\|$. Hence, (62) yields $\|\delta(v) - \bar{\delta}(v)\| \ge \frac{1}{3}\|\mathcal{E}_a - \mathcal{E}_b\|$. □

As Corollary 30 implies the left inequality in (51), we have completed the proof of Lemma 24.

7.3 Proof of Lemma 25

For all $v \in \mathcal{R}_{ab}$, all $\alpha \in \{1, \ldots, k\}$, and all $w \in \mathcal{S}_\alpha$ we set

$$\begin{aligned}
\Delta(v,w) &= \frac{e(v, \mathcal{S}_\alpha \cap H)}{\#\mathcal{S}_\alpha}, \quad \bar{\Delta}(v,w) = \frac{\mu'(v, \mathcal{S}_\alpha \cap H)}{\#\mathcal{S}_\alpha}, \text{ and we recall that} \\
\delta(v,w) &= \frac{e(v, \mathcal{S}_\alpha)}{\#\mathcal{S}_\alpha}, \quad \bar{\delta}(v,w) = \frac{\mu'(v, \mathcal{S}_\alpha)}{\#\mathcal{S}_\alpha}.
\end{aligned}$$

Moreover, we let $\Delta(v) = (\Delta(v,w))_{w\in H}$, $\bar{\Delta}(v) = (\bar{\Delta}(v,w))_{w\in H}$, and remember that $\delta(v) = (\delta(v,w))_{w\in V}$, $\bar{\delta}(v) = (\bar{\delta}(v,w))_{w\in V}$.

Lemma 31. *Suppose that $\mathcal{S}$ is feasible and tight. Let $\mathcal{R}$ be an improvement of $\mathcal{S}$. Then for all $v \in \mathcal{R}_{ba} \cap H$ we have $\|\Delta(v) - \bar{\Delta}(v)\|^2 \ge 0.1\|\mathcal{E}_{V_a} - \mathcal{E}_{V_b}\|^2$ ($1 \le a, b \le k$).*

Proof. Let $\delta_H(v) = (\delta(v,w))_{w\in H}$ and $\bar{\delta}_H(v) = (\bar{\delta}(v,w))_{w\in H}$, i.e., $\delta_H(v)$, $\bar{\delta}_H(v)$ are the restrictions of $\delta(v)$, $\bar{\delta}(v)$ to H. We claim that for all $v \in \mathcal{R}_{ba} \cap H$ we have

$$\|\delta_H(v) - \bar{\delta}_H(v)\| \ge 0.134\|\mathcal{E}_{V_a} - \mathcal{E}_{V_b}\|. \qquad (63)$$

For as $\mathcal{R}$ is an improvement of $\mathcal{S}$ and because $\|\xi_a - \mathcal{E}_{V_a}\|, \|\xi_b - \mathcal{E}_{V_b}\| \le \rho/33$ by Proposition 6, we have

$$\begin{aligned}
\rho &\le \|\mathcal{E}_{V_a} - \mathcal{E}_{V_b}\| \le \|\delta(v) - \xi_b\| + \|\delta(v) - \xi_a\| + \|\xi_a - \mathcal{E}_{V_a}\| + \|\xi_b - \mathcal{E}_{V_b}\| \le \frac{2}{33}\rho + 2\|\delta(v) - \xi_a\| \\
&\le \frac{2}{33}\rho + 2\|\xi_a - \mathcal{E}_{V_a}\| + 2\|\delta(v) - \mathcal{E}_{V_a}\| \le \frac{4}{33}\rho + 2\|\delta(v) - \mathcal{E}_{V_a}\|.
\end{aligned}$$

Hence, $\|\delta(v) - \mathcal{E}_{V_a}\| \ge \frac{29}{66}\|\mathcal{E}_{V_a} - \mathcal{E}_{V_b}\|$. Furthermore, by Lemma 29 and because $\mathcal{S}$ is tight,

$$\|\bar{\delta}(v) - \mathcal{E}_{V_a}\|^2 \le 4\sum_{\alpha,\beta=1}^{k} \#\mathcal{S}_{\alpha\beta}(p_{a\alpha} - p_{a\beta})^2 \le 4\sum_{\alpha,\beta=1}^{k} \frac{\#S_{\alpha\beta}\|\mathcal{E}_{V_\alpha} - \mathcal{E}_{V_\beta}\|^2}{\#V_a} \le 0.004\rho^2.$$

Thus, $\|\delta(v)-\bar{\delta}(v)\| \geq 0.37\|\mathcal{E}_{V_a}-\mathcal{E}_{V_b}\|$. Therefore, as $\#V\setminus H \leq 10^{-4}n_{\min}$ by **H1** and as $\mathcal{S}$ is feasible, we obtain

$$0.136\|\mathcal{E}_{V_a}-\mathcal{E}_{V_b}\|^2 \leq \|\delta(v)-\bar{\delta}(v)\|^2 = \sum_{\alpha=1}^{k}\#\mathcal{S}_\alpha\left(\frac{e(v,\mathcal{S}_\alpha)-\mu'(v,\mathcal{S}_\alpha)}{\#\mathcal{S}_\alpha}\right)^2$$
$$\leq 1.01\sum_{\alpha=1}^{k}\#\mathcal{S}_\alpha\cap H\left(\frac{e(v,\mathcal{S}_\alpha)-\mu'(v,\mathcal{S}_\alpha)}{\#\mathcal{S}_\alpha}\right)^2 = 1.01\|\delta_H(v)-\bar{\delta}_H(v)\|^2,$$

whence (63) follows.

Now, we shall compare the vectors $\delta_H(v)-\bar{\delta}_H(v)$ and $\Delta(v)-\bar{\Delta}(v)$, so that we can use (63) to bound the norm of the latter vector. Let $\Phi_1(v)=\{w\in V: p_{\psi(v)\psi(w)}\leq 1/2\}$, $\Phi_2(v)=\{w\in V: p_{\psi(v)\psi(w)}>1/2\}$. Then for all $v\in H$ we have

$$\|(\delta_H(v)-\bar{\delta}_H(v))-(\Delta(v)-\bar{\Delta}(v))\|^2 = \sum_{w\in H}(\delta_H(v,w)-\bar{\delta}_H(v,w)-(\Delta(v,w)-\bar{\Delta}(v,w)))^2$$
$$= \sum_{\alpha=1}^{k}\#\mathcal{S}_\alpha^{-1}\left[e(v,\mathcal{S}_\alpha\setminus H)-\mu'(v,\mathcal{S}_\alpha\setminus H)\right]^2$$
$$\leq 2\sum_{\alpha=1}^{k}\#\mathcal{S}_\alpha^{-1}\left[e(v,\Phi_1(v)\cap\mathcal{S}_\alpha\setminus H)-\mu'(v,\Phi_1(v)\cap\mathcal{S}_\alpha\setminus H)\right]^2$$
$$+\#\mathcal{S}_\alpha^{-1}\left[e(v,\Phi_2(v)\cap\mathcal{S}_\alpha\setminus H)-\mu'(v,\Phi_2(v)\cap\mathcal{S}_\alpha\setminus H)\right]^2. \tag{64}$$

Since $v\in H$, by **H4** we have

$$e(v,\Phi_1(v)\cap\mathcal{S}_\alpha\setminus H)\leq 100, \tag{65}$$
$$e(v,\Phi_2(v)\cap\mathcal{S}_\alpha\setminus H)\geq \#\Phi_2(v)\cap\mathcal{S}_\alpha\setminus H-101. \tag{66}$$

Moreover, **H1** entails that

$$\sum_{\alpha=1}^{k}\mu'(v,\Phi_1(v)\cap\mathcal{S}_\alpha\setminus H)\leq \mu'(v,G_1\setminus H)\leq \#V\setminus H\cdot\max_{1\leq\alpha,\beta\leq k:p_{\alpha\beta}\leq\frac{1}{2}}p_{\alpha\beta}\leq\frac{n_{\min}}{\lambda^4}\cdot\frac{2\sigma^*}{n_{\min}}\leq\frac{1}{2},$$

whence $\sum_{\alpha=1}^{k}\mu'(v,\Phi_1(v)\cap\mathcal{S}_\alpha\setminus H)^2\leq\frac{1}{4}$. Consequently, as the fact that $\mathcal{S}$ is feasible implies that $\#\mathcal{S}_\alpha\geq\frac{1}{2}n_{\min}$, we obtain

$$\sum_{\alpha=1}^{k}\frac{\mu'(v,\Phi_1(v)\cap\mathcal{S}_\alpha\setminus H)^2}{\#\mathcal{S}_\alpha}\leq\frac{1}{2n_{\min}}\overset{\textbf{A2}}{\leq}10^{-4}\rho^2. \tag{67}$$

A similar argument shows that

$$\sum_{\alpha=1}^{k}\frac{(\#\Phi_2(v)\cap\mathcal{S}_\alpha\setminus H-\mu'(v,\Phi_2(v)\cap\mathcal{S}_\alpha\setminus H))^2}{\#\mathcal{S}_\alpha}\leq 10^{-4}\rho^2. \tag{68}$$

Plugging (65)–(68) into (64), we get

$$\|(\delta_H(v)-\bar{\delta}_H(v))-(\Delta(v)-\bar{\Delta}(v))\|^2\leq 2\cdot10^{-4}\rho^2+\sum_{\alpha=1}^{k}\frac{10^5}{\#\mathcal{S}_\alpha}\leq 4\cdot10^{-4}\rho^2+\frac{10^6k}{n_{\min}}\overset{\textbf{A2}}{\leq}10^{-3}\rho^2.$$

Therefore, (63) entails that $\|\Delta(v)-\bar{\Delta}(v)\|^2\geq 0.1\|\mathcal{E}_a-\mathcal{E}_b\|^2$ for all $v\in\mathcal{R}_{ba}\cap H$. □

Proof of Lemma 7.3. Let M be the matrix defined in (48). We shall prove below that

$$\sum_{a\neq b}\sum_{v\in\mathcal{R}_{ab}\cap H}\|\Delta(v)-\bar{\Delta}(v)\|^2\leq 0.02\sum_{a\neq b}\rho^2\#\mathcal{R}_{ab}\cap H+\frac{24\|M\|^2}{n_{\min}}\sum_{a\neq b}\#H\cap\mathcal{S}_{ab}. \tag{69}$$

On the other hand, Lemma 31 implies in combination with **A2** that

$$\sum_{a\neq b,\, v\in\mathcal{R}_{ab}\cap H} \|\Delta(v)-\bar{\Delta}(v)\|^2 \geq 0.1\sum_{a\neq b}\#\mathcal{R}_{ab}\cap H\rho^2. \tag{70}$$

Combining (69) and (70), we thus get

$$\sum_{a\neq b}\#\mathcal{R}_{ab}\cap H \leq \frac{1000\|M\|^2}{\rho^2 n_{\min}}\sum_{a\neq b}\#\mathcal{S}_{ab}\cap H \overset{(49),\,\mathbf{A2}}{\leq} \frac{1}{10}\sum_{a\neq b}\#\mathcal{S}_{ab}\cap H,$$

as desired.

Hence, the remaining task is to establish (69). To this end, we note that for all $v\in\mathcal{R}_{ab}\cap H$ such that $a\neq b$ we have

$$\begin{aligned}\|\Delta(v)-\bar{\Delta}(v)\|^2 &\leq 2\sum_{a=1}^{k}\#\mathcal{S}_a^{-1}\left[e(v,H\cap\mathcal{S}_a)-\mu'(v,H\cap\mathcal{S}_a)\right]^2\\ &\leq 6\cdot\sum_{a=1}^{k}\#\mathcal{S}_a^{-1}\left[e(v,H\cap V_a)-\mu'(v,H\cap V_a)\right]^2\\ &\quad+6\cdot\sum_{a=1}^{k}\#\mathcal{S}_a^{-1}\left[e(v,H\cap\mathcal{S}_a\setminus V_a)-\mu'(v,H\cap\mathcal{S}_a\setminus V_a)\right]^2\\ &\quad+6\cdot\sum_{a=1}^{k}\#\mathcal{S}_a^{-1}\left[e(v,H\cap V_a\setminus\mathcal{S}_a)-\mu'(v,H\cap V_a\setminus\mathcal{S}_a)\right]^2\\ &\leq 12[\|d(v)-\mathcal{E}_v\|^2+\sum_{a=1}^{k}\#V_a^{-1}(e(v,H\cap V_a\setminus\mathcal{S}_a)-\mu'(v,H\cap V_a\setminus\mathcal{S}_a))^2\\ &\quad+\sum_{a=1}^{k}\#V_a^{-1}(e(v,H\cap\mathcal{S}_a\setminus V_a)-\mu'(v,H\cap\mathcal{S}_a\setminus V_a))^2],\end{aligned} \tag{71}$$

where the last inequality follows from our assumption that $\mathcal{S}$ is feasible. Since $v\in H$, due to **H2** we have

$$\|d(v)-\mathcal{E}_v\|^2 \leq 0.001\rho^2. \tag{72}$$

Furthermore,

$$\begin{aligned}\sum_{a=1}^{k}\sum_{v\in H}\left[e(v,H\cap V_a\setminus\mathcal{S}_a)-\mu'(v,H\cap V_a\setminus\mathcal{S}_a)\right]^2 &\overset{(50)}{=} \sum_{a=1}^{k}\|M\mathbf{1}_{H\cap V_a\setminus\mathcal{S}_a}\|^2\\ &\leq \|M\|^2\sum_a\|\mathbf{1}_{H\cap V_a\setminus\mathcal{S}_a}\|^2 \;=\; \|M\|^2\sum_a\#H\cap V_a\setminus\mathcal{S}_a,\end{aligned}$$

and analogously $\sum_{a=1}^{k}\sum_{v\in H}\left[e(H\cap\mathcal{S}_a\setminus V_a)-\mu'(v,H\cap\mathcal{S}_a\setminus V_a)\right]^2 \leq \|M\|^2\sum_{a=1}^{k}\#H\cap\mathcal{S}_a\setminus V_a$. Thus,

$$\sum_{a\neq b}\sum_{v\in\mathcal{R}_{ab}\cap H}\left[e(v,V_a\setminus\mathcal{S}_a)-\mu'(v,V_a\setminus\mathcal{S}_a)\right]^2+\left[e(\mathcal{S}_a\setminus V_a)-\mu'(v,\mathcal{S}_a\setminus V_a)\right]^2 \leq 4\|M\|^2\sum_{a\neq b}\#\mathcal{S}_{ab}\cap H. \tag{73}$$

Combining (71), (72), and (73), we obtain (69), thereby completing the proof. □

8 The Random Graph $G_{n,k}(\psi,\boldsymbol{p})$

In this section we prove Theorem 2. We start with some preliminaries on random graphs in Section 8.1. Then, we discuss the construction of the core of $G_{n,k}(\psi,\boldsymbol{p})$ in Section 8.2. Finally, in Section 8.4 we investigate the components of $G_{n,k}(\psi,\boldsymbol{p})-\mathrm{core}(G_{n,k}(\psi,\boldsymbol{p}))$. *Throughout this section, we let ψ, $\boldsymbol{p}$, $\mathcal{E}$, $n_{\min}$, and σ^* be as in Sections 1.2 and 1.3. Furthermore, we always assume that n is sufficiently large.*

8.1 Preliminaries on $G_{n,k}(\psi, p)$

We need to bound the probability that a random variable deviates from its mean significantly. To this end, let ϕ denote the function $\phi : (-1, \infty) \to \mathbf{R}$, $x \mapsto (1+x)\ln(1+x) - x$. A proof of the following Chernoff bound can be found in [22, pages 26–29].

Lemma 32. *Let $X = \sum_{i=1}^{N} X_i$ be a sum of mutually independent Bernoulli random variables with variance $\sigma^2 = \mathrm{Var}(X)$. Then for any $t > 0$ we have*

$$\max\{\mathrm{P}(X \le \mathrm{E}(X) - t), \mathrm{P}(X \ge \mathrm{E}(X) + t)\} \le \exp\left(-\sigma^2 \phi\left(\frac{t}{\sigma^2}\right)\right) \le \exp\left(-\frac{t^2}{2(\sigma^2 + t/3)}\right). \quad (74)$$

The following bound, whose proof can be found in Section 9.1, is a consequence of Azuma's inequality.

Lemma 33. *Let X be a function from graphs to reals that satisfies the following Lipschitz condition.*

Let $G = (V, E)$ be a graph, and let $v, w \in V$. Let G' be the graph obtained from G by removing the edge $\{v, w\}$ if it is present in G, and let G'' be the graph obtained by adding $\{v, w\}$ to G if it is not present. Then $|X(G') - X(G'')| \le 1$. (75)

Then $\mathrm{P}\left[|X(G_{n,k}(\psi, \boldsymbol{p})) - \mathrm{E}(X(G_{n,k}(\psi, \boldsymbol{p})))| > \sqrt{\sigma^* n}\ln^2 n\right] \le n^{-10}$.

In Section 9.2 we shall use Lemma 33 to derive the following estimate on the upper tail of the degree distribution of $G_{n,k}(\psi, \boldsymbol{p})$.

Lemma 34. *Let $U_i = \#\{v \in V : \max_{j=1,2} d_{G_j}(v) \ge 2^{i+1}\sigma^*\}$. Then w.h.p. $\#U_i \le \exp(-2^{i-2}\sigma^*)n$ for all $i = 2, \ldots, \lceil \log_2 n \rceil$.*

Furthermore, in Section 9.3 we shall establish that the graph $G_1 \cup G_2$ does not contain any "atypically dense spots" w.h.p.

Lemma 35. *W.h.p. $G = G_{n,k}(\psi, \boldsymbol{p})$ enjoys the following property.*

For all sets $T \subset V$ such that $\#T \le n\left(\frac{n_{\min}}{n\sigma^*}\right)^2$ we have $e_{G_1 \cup G_2}(T) \le 10\#T$. (76)

Furthermore, with probability $\ge 1 - \exp(-\ln^3 n)$ the following holds.

For all $T \subset V$ such that $\ln^3 n \le \#T \le n\left(\frac{n_{\min}}{n\sigma^*}\right)^2$ we have $e_{G_1 \cup G_2}(T) \le 10\#T$. (77)

Finally, we need the following result on the spectrum of the adjacency matrix of $G_{n,k}(\psi, \boldsymbol{p})$.

Lemma 36. *Let $\Delta > 0$ and $X = \{v \in V : \max_{i=1,2} d_{G_i}(v) \le \Delta\}$. Then $\|A_X - \mathcal{E}_X\| \le ck\sqrt{\sigma^* + \Delta}$.*

In Section 9.4 we indicate how Lemma 36 follows from spectral considerations of Alon and Kahale [1], Feige and Ofek [17], and Füredi and Komloś [20].

8.2 The Core

In this section our objective is to construct a subgraph $\mathrm{core}(G)$ of $G = G_{n,k}(\psi, \boldsymbol{p})$ such that for all vertices $v \in \mathrm{core}(G)$ the numbers $e(v, V_i \cap \mathrm{core}(G))$ do not deviate from the expectations $\mu(v, V_i)$ "too much". To this end, we assign to each $v \in V$ a vector $d(v)$ as in (1), which represents the actual numbers of $e(v, V_i)$-edges. By comparison, $\mathcal{E}_v$ represents the *expected* numbers of v-V_i-edges. The first step of the construction is as follows.

CR1. Initially, remove all vertices v such that $\|d(v) - \mathcal{E}_v\| > 0.01\rho$ from G; that is, set $H = G - \{v \in V : \|d(v) - \mathcal{E}_v\| > 0.01\rho\}$. (Here ρ^2 is the r.h.s. of (6).)

Moreover, recall the decomposition of $G = G_{n,k}(\psi, \boldsymbol{p})$ into the "sparse" part G_1 and the "dense" part G_2 from Section 1.6. Then $\mathrm{E}(d_{G_1 \cup G_2}(v)) \leq 2\sigma^*$ for all $v \in V$. Nevertheless, in the case $\sigma^* = O(1)$ as $n \to \infty$ there may occur vertices such that $d_{G_1 \cup G_2}(v)$ exceeds $2\sigma^*$ significantly. Therefore, as a second step we remove such vertices v.

CR2. Remove all vertices v such that $d_{G_1 \cup G_2}(v) > 10\sigma^*$ from H.

However, in general the result H of **CR1–CR2** will *not* be such that $e(v, V_i \cap H)$ approximates $\mu(v, V_i)$ well for all $v \in H$. The reason is that there may occur vertices $v \in H$ such that "many" neighbors of v got removed. Hence, in the final step of our construction we iteratively remove these vertices v from H.

CR3. While there is a vertex $v \in H$ such that $e_{G_1 \cup G_2}(v, V \setminus H) > 100$, remove v from H.

The outcome of the process **CR1–CR2** is $\mathrm{core}(G) = H$. In Section 8.3 we shall prove that w.h.p. $\mathrm{core}(G)$ constitutes a huge fraction of G.

Proposition 37. *Suppose that (6) holds. Then w.h.p.* $\mathrm{core}(G_{n,k}(\psi, \boldsymbol{p}))$ *contains* $\geq n - n_{\min}\sigma^{*\,-10}$ *vertices. For all* $v \in \mathrm{core}(G)$ *we have* $\|d(v) - \mathcal{E}_v\| \leq 0.01\rho$, $d_{G_1 \cup G_2}(v) \leq 10\sigma^*$, *and* $e_{G_1 \cup G_2}(v, G - H) \leq 100$.

In addition, adapting proof techniques from [1], we shall prove in Section 8.4 that $G - \mathrm{core}(G)$ has the following simple structure w.h.p.

Proposition 38. *If (6) holds, then w.h.p. all components of* $(G_1 \cup G_2) - \mathrm{core}(G)$ *have size* $\leq \ln n$.

Proof of Theorem 2. Assuming that c_0 is a sufficiently large constant and letting $\lambda = \sigma^* > c_0$, we note that Lemma 36 implies that $G_{n,k}(\psi, \boldsymbol{p})$ satisfies **A1** w.h.p. Moreover, our assumption **R3** ensures that **A2** is true. Further, for each vertex $v \in V_j$ and each $1 \leq i \leq k$ the number $e(v, V_i)$ has a binomial distribution with variance $\#V_i P_{ij}(1 - p_{ij}) \leq \sigma^*$; therefore, the Chernoff bound (74) entails that

$$\mathrm{P}\left[|e(v, V_i) - \mu(v, V_i)| > \frac{\sigma^*}{10k} + \ln^2 n\right] \leq 2\exp\left[-\frac{\sigma^{*\,2}k^{-2} + \ln^4 n}{300(\sigma^* + \ln^2 n)}\right] \leq n^{-1}. \tag{78}$$

Thus, we conclude that in both cases **A3** holds w.h.p. Finally, assumption **R2** yields **A4**.

With respect to **H1**, letting $H = \mathrm{core}(G)$ we observe that Proposition 37 entails that $\#V \setminus H \leq n_{\min}\lambda^{-4}$. Furthermore, let $U_i = \{v \in V : 2^{i+1}\sigma^* \leq \max_{j=1,2} d_{G_j}(v) \leq 2^{i+2}\sigma^*\}$. Then Lemma 34 and our assumption that $\sigma^* \geq c_0$ for a large enough number c_0 entail that w.h.p.

$$\begin{aligned}\sum_{v \in V \setminus H} d_{G_1 \cup G_2}(v)^2 &\leq 2^{10}\sigma^{*\,2}\#V \setminus H + \sum_{i \geq 2} 2^{2i+4}\sigma^{*\,2}\#U_i \\ &\leq 2^{10} n_{\min}\sigma^{*\,-2} + \sum_{i \geq 2} 2^{i+2}\sigma^{*\,2}\exp(-2^{i-2}\sigma^*)n \\ &\leq \frac{1}{2}n_{\min} + 8n\exp(-\sigma^*/2) \overset{\textbf{R1}}{\leq} \frac{1}{2}n_{\min} + 8n_{\min}\exp(-\sqrt{\sigma^*}/2) \leq n_{\min},\end{aligned}$$

whence **H1** follows. Moreover, **H2**, **H3**, and **H4** follow directly from Proposition 37. □

8.3 Proof of Proposition 37

To estimate $\#V(\mathrm{core}(G))$, we consider the following modification of the process **CR1–CR3**. Set $\omega = \sigma^* + \frac{n}{n_{\min}}$, and note that $\omega \geq n/n_{\min} \geq k$.

K1. Initially, let K be the subgraph of G obtained by removing all vertices $v \in V$ such that

$$\max_{1 \leq i \leq k} |e(v, V_i) - \mu(v, V_i)| \geq 10^4\left[\sqrt{\#V_i p_{ij}(1 - p_{ij})\ln\omega} + \ln\omega\right].$$

K2. While there is a vertex $v \in K$ such that $e_{G_1 \cup G_2}(v, V \setminus K) > 50$, remove v from K.

To establish Proposition 37, we proceed in two steps. First, we show that $\mathrm{core}(G) \supset K$. Then, we bound $\#V(G-K)$.

Lemma 39. *We have* $\mathrm{core}(G) \supset K$.

Proof. Suppose that $v \in K$. Then

$$\|d(v) - \mathcal{E}_v\|^2 = \sum_{i=1}^{k} \#V_i \left(\frac{e(v,V_i) - \mu(v,V_i)}{\#V_i}\right)^2 = \sum_{i=1}^{k} \#V_i^{-1}(e(v,V_i) - \mu(v,V_i))^2$$

$$\leq 2 \cdot 10^4 \sum_{i=1}^{k} \#V_i^{-1} \left[\#V_i p_{ij}(1-p_{ij}) \ln\omega + \ln^2\omega\right] \qquad \text{[due to } \mathbf{K1}\text{]}$$

$$\leq 2 \cdot 10^4 \left[\sum_{i=1}^{k} p_{ij}(1-p_{ij})\ln\omega + \sum_{i=1}^{k} \frac{\ln^2\omega}{\#V_i}\right] \leq 10^{-4}\rho^2,$$

where the last step follows from (6) and **R1**. Thus, none of the vertices $v \in K$ gets removed by **CR1**. Further, **K1** ensures that $d_{G_1 \cup G_2}(v) \leq 10\sigma^*$ for all $v \in K$, so that K is contained in the subgraph of G obtained in **CR2**. Finally, as **K2** is more restrictive than **CR3**, we conclude that $\mathrm{core}(G) \supset K$. □

Our next aim is to bound $\#V(G-K)$. We first estimate the number of vertices removed by **K1**.

Lemma 40. *W.h.p. there are at most* $n\omega^{-198}$ *vertices* v *such that* $\max_{1 \leq i \leq k} |e(v,V_i) - \mu(v,V_i)| \geq 10^3 \left[\sqrt{\#V_I P_{ij}(1-p_{ij})\ln\omega} + \ln\omega\right]$. *Moreover, if* $\omega \leq n^{1/190}$*, then with probability* $\geq 1 - \exp(-\ln^3 n)$ *there are at most* $n\omega^{-90}$ *such vertices.*

Proof. By the Chernoff bound (74), for each vertex $v \in V_j$ we have

$$P_{ij} = \mathrm{P}\left[|e(v,V_i) - \mu(v,V_i)| \geq 10^3\left(\sqrt{\#V_i p_{ij}(1-p_{ij})\ln\omega} + \ln\omega\right)\right]$$

$$\leq 2\exp\left[-\frac{10^6(\#V_i p_{ij}(1-p_{ij}) + \ln^2\omega)}{2(\#V_i p_{ij}(1-p_{ij}) + 10^3(\sqrt{\#V_i p_{ij}(1-p_{ij})\ln\omega} + \ln\omega)}\right]$$

$$\leq 2\exp\left[-\frac{10^6\#V_i p_{ij}(1-p_{ij})\ln\omega + 10^6\ln^2\omega}{5 \cdot 10^3(\#V_i p_{ij}(1-p_{ij}) + \ln\omega)}\right] \leq 2\omega^{-200}.$$

Hence, letting $Z_{ij} = \#\left\{v \in V_j : |e(v,V_i) - \mu(v,V_i)| \geq 10^3\left(\sqrt{\#V_I P_{ij}(1-p_{ij})\ln\omega} + \ln\omega\right)\right\}$, we have

$$\mathrm{E}(Z_{ij}) \leq 2\#V_j\omega^{-200}. \tag{79}$$

To obtain a bound on Z_{ij} that actually holds w.h.p., we consider two cases.

1st case: $\omega \geq \ln n$**.** Then Markov's inequality entails that w.h.p. $\sum_{i,j=1}^{k} \#Z_{ij} \leq nk\omega^{-199} \leq n\omega^{-198}$.

2nd case: $\omega < \ln n$**.** As adding or removing a single edge $e = \{u,v\}$ affects only the numbers $e(u,V_i)$ and $e(v,V_i)$, the random variable $Z_{ij}/2$ satisfies the Lipschitz condition (75). Further, $\sigma^* \leq \omega \leq \ln n$, and $\#V_j \geq n_{\min} \geq n/\omega > n/\ln n$. Hence, Lemma 33 entails that

$$\mathrm{P}\left[Z_{ij} \geq \#V_j\omega^{-199}\right] \overset{(79)}{\leq} \mathrm{P}\left[Z_{ij} - \mathrm{E}(Z_{ij}) \geq \sqrt{\sigma^* n}\ln^2 n\right] = o(1),$$

and thus $\sum_{i,j=1}^{k} Z_{ij} \leq kn\omega^{-199} \leq n\omega^{-198}$ w.h.p.

Now, assume that $\omega \leq n^{1/190}$. Then the inequalities $\omega \geq \sigma^*$ and $\omega \geq n/n_{\min}$ imply that $\sqrt{n\sigma^*}\ln^2 n \leq \sqrt{n\omega}\ln^2 n \leq n^{96/190}$, while $n\omega^{-92} \geq n^{98/190}$. Therefore, Lemma 33 entails

$$\mathrm{P}\left[Z_{ij} \geq n\omega^{-92}\right] \overset{(79)}{\leq} \mathrm{P}\left[Z_{ij} - \mathrm{E}(Z_{ij}) \geq \sqrt{\sigma^* n}\ln^2 n\right] \leq \exp(-\ln^4 n).$$

Hence, with probability $\geq 1 - \exp(-\ln^3 n)$ the bound $Z_{ij} < n\omega^{-92}$ holds for all $1 \leq i,j \leq k$ simultaneously, and thus $\sum_{i,j=1}^{k} Z_{ij} \leq k^2 n\omega^{-92} \leq n\omega^{-90}$. □

Lemma 40 implies that w.h.p. **K1** removes at most $n\omega^{-198}$ vertices. Finally, we need to bound the number of vertices that get removed during **K2**.

Lemma 41. *W.h.p.* **K2** *removes at most* $n\omega^{-198}$ *vertices.*

Proof. Let S be the set of vertices removed by **K1**. By Lemma 40 we may assume that $s = \#S \leq n\omega^{-198}$. Moreover, let $v_1, \ldots, v_q$ be the vertices removed by **K2** (in this order). Assume that $q \geq s$, and let $T = S \cup \{v_1, \ldots, v_s\}$. We shall prove that T violates (76), so that Lemma 35 entails that actually $q < s$ w.h.p.

To see that T is an "atypically dense" set in $G_1 \cup G_2$ that violates (76), observe that by construction each v_i satisfies $e_{G_1 \cup G_2}(v_i, S \cup \{v_1, \ldots, v_{i-1}\}) \geq 50$. Therefore, $e_{G_1 \cup G_2}(T) \geq 50s \geq 25\#T$, while $\#T = 2s \leq n\omega^{-197}$. □

Combining Lemmas 39–41, we obtain the following corollary, which implies Proposition 37.

Corollary 42. *W.h.p. we have* $\#V(K) \geq n(1 - \omega^{-197})$.

8.4 Proof of Proposition 38

If $\omega = \sigma^* + \frac{n}{n_{\min}} \geq n^{1/190}$, then Lemma 39 and Corollary 42 yield that $\mathrm{core}(G) = G$ w.h.p., and thus there is nothing to prove. Hence, we assume in the sequel that $\omega < n^{1/200}$. We shall prove that in this case w.h.p. the graph $(G_1 \cup G_2) - K$ does not contain a tree on $\ln n$ vertices w.h.p., where K is the outcome of the process **K1**–**K2** defined in Section 8.3. Since $\mathrm{core}(G) \supset K$ by Lemma 39, this implies the assertion.

Thus, let $T = (V_T, E_T)$ be a tree with vertex set $V_T \subset V$ on $t = \#V_T = \lceil \ln n \rceil$ vertices (T is not necessarily a subgraph of G, but just a tree whose vertex set is contained in V). We shall estimate the probability that T is contained in $(G_1 \cup G_2) - K$. To this end, we consider $I_T = \{v \in V_T : d_T(v) \leq 4\}$ and $J_T = V_T \setminus I_T$; as $\#E_T = t - 1$, we have $\#I_T \geq t/2$. Moreover, let K_T be the outcome of the following modification of the process **K1**–**K2** (cf. Section 8.3). Set $\omega = \sigma^* + \frac{n}{n_{\min}}$.

K0'. Let G^* be a graph obtained from G by replacing the edges in E_T by fresh random edges. That is, each edge $e = \{v, w\} \in E_T$ is present in G^* with probability $p_{\psi(v)\psi(w)}$ independently of all others and of the choice of G, and $G^* - E_T = G - E_T$.

K1'. Let K_T be the subgraph of G^* obtained by removing the vertices

$$J_T \cup \left\{ v \in V : \max_{1 \leq i \leq k} |e_G(v, V_i) - \mu_G(v, V_i)| \geq 10^3 \left[\sqrt{\#V_I P_{ij}(1 - p_{ij}) \ln \omega} + \ln \omega \right] \right\}.$$

K2'. While there is a vertex $v \in K_T$ such that $\max_{i=1,2} e_{G_i}(v, V \setminus K_T) > 40$, remove v from K_T.

Lemma 43. *Let K be the result of the process* **K1**–**K2** *(cf. Section 8.3). Then $K_T \subset K$, regardless of the outcome of step* **K0'**.

Proof. Since every vertex $v \in I_T$ is incident with ≤ 4 edges of T, the graph defined in step **K1'** is contained in the graph defined in step **K1**. Consequently, all vertices removed by **K2** also get removed by **K2'**. □

Let us call G *good* if for all trees T as above we have $\#V(G - K_T) \leq n\omega^{-88}$, regardless of the outcome of step **K0'**.

Lemma 44. *We have* $\mathrm{P}\,[G \textit{ is good}] \geq 1 - 2\exp(-\ln^3 n)$.

Proof. Let S be the set of vertices removed by **K1'**, and let $s = \#S$. Since $\omega \leq n^{1/190}$, Lemma 40 entails that with probability $\geq 1 - \exp(-\ln^3 n)$ we have $s \leq \#J_T + n\omega^{-90} \leq n\omega^{-89}$. Furthermore, if **K2'** removes $q \geq n\omega^{-89}$ vertices $v_1, \ldots, v_q$, then consider the set $T = S \cup \{v_1, \ldots, v_{\lceil n\omega^{-89} \rceil}\}$. Then $\ln^3 n \leq n\omega^{-89} \leq \#T \leq s + n\omega^{-89} + 1 \leq n\omega^{-88}$, but $e_{G_1 \cup G_2}(T) \geq 40\#T/2 = 20\#T$ (cf. the proof of Lemma 41). Hence, T violates (77). Consequently, Lemma 35 entails that $q \leq n\omega^{-89}$ with probability $\geq 1 - \exp(-\ln^3 n)$, whence the assertion follows. □

Proof of Proposition 38. Since the construction of K_T is independent of the presence of edges of T in $G_1 \cup G_2$ due to **K0'**, Lemma 43 yields

$$\mathrm{P}\left[T \subset G_1 \cup G_2 \wedge V_T \cap K = \emptyset\right] \leq \mathrm{P}\left[T \subset G_1 \cup G_2\right] \cdot \mathrm{P}\left[I_T \cap K_T = \emptyset\right]. \tag{80}$$

Given their cardinalities, the sets $V_i \cap H_T$ are uniformly distributed random subsets of $V_i \setminus J_T$, as due to **K0'** the distribution of $G^* - J_T$ is invariant under permutations of the vertices within the classes V_i. Therefore, letting $t_i = \#I_T \cap V_i$ and $\nu = \lceil n\omega^{-88}\rceil$, we obtain

$$\begin{aligned}
\mathrm{P}\left[I_T \cap K_T = \emptyset\right] &\leq \mathrm{P}\left[G \text{ is not good}\right] + \prod_{i=1}^{k} \frac{\binom{\#V_i - t_i}{\nu - t_i}}{\binom{\#V_i}{\nu}} \\
&\overset{\text{Lemma 44}}{\leq} \exp(-\ln^3 n) + \prod_{i=1}^{k} \frac{(\#V_i - t_i)_{\nu - t_i}(\nu)_{t_i}}{(\#V_i)_{\nu - t_i}(\#V_i - \nu + t_i)_{t_i}} \\
&\leq \exp(-\ln^3 n) + \prod_{i=1}^{k} \left(\frac{\nu}{\#V_i - \nu}\right)^{t_i} \leq \exp(-\ln^3 n) + \prod_{i=1}^{k} \left(\frac{2\nu}{\#V_i}\right)^{t_i} \\
&\leq \exp(-\ln^3 n) + \omega^{-86\sum_{i=1}^{k} t_i} \leq \exp(-\ln^3 n) + \omega^{-43t} \leq \omega^{-42t}.
\end{aligned} \tag{81}$$

To bound $\mathrm{P}\left[T \subset G_1 \cup G_2\right]$, we note that $\mathrm{P}\left[\{v, w\} \in E(G_1 \cup G_2)\right] \leq 2p_{\psi(v)\psi(w)}(1 - p_{\psi(v)\psi(w)}) \leq 2\sigma^*/n_{\min}$ by the definition of σ^* ($v, w \in V$). Consequently,

$$\mathrm{P}\left[T \subset G_1 \cup G_2\right] \leq \left(\frac{2\sigma^*}{n_{\min}}\right)^{t-1}. \tag{82}$$

Combining (80), (81), and (82), and recalling that $\omega = \sigma^* + \frac{n}{n_{\min}}$, we conclude

$$\mathrm{P}\left[T \subset G_1 \cup G_2 \wedge V_T \cap K = \emptyset\right] \leq \left(\frac{2\sigma^*}{n_{\min}}\right)^{t-1} \omega^{-42t} \leq n^{1-t}\omega^{-39t}. \tag{83}$$

Finally, we are going to apply the union bound to estimate the probability that there *exists* a tree T as above such that $T \subset G_1 \cup G_2$ and $V_T \cap K = \emptyset$. Since by Cayley's formula there are $\binom{n}{t} t^{t-2}$ ways to choose the tree T, (83) entails that

$$\mathrm{P}\left[\exists T : T \subset G_1 \cup G_2 \wedge V_T \cap K = \emptyset\right] \leq \binom{n}{t} t^{t-2} n^{1-t}\omega^{-39t} \leq \exp(t) n^2 \omega^{-39t} \leq n^{-36},$$

because $t \geq \ln n$. Hence, w.h.p. $(G_1 \cup G_2) - K$ contains no tree on $\geq \ln n$ vertices. □

9 Proofs of Auxiliary Lemmas

9.1 Proof of Lemma 33

The proof relies on the following general tail bound, which is a consequence of Azuma's inequality (cf. [22, p. 38] for a proof).

Lemma 45. *Let $\Omega = \prod_{i=1}^{N} \Omega_i$ be a product of probability spaces $\Omega_1, \ldots, \Omega_N$. Let $Y : \Omega \to \mathbf{R}$ be a random variable that satisfies the following condition for all $1 \leq j \leq N$.*

> *If $\omega = (\omega_i)_{1 \leq i \leq N}, \omega' = (\omega_i')_{1 \leq i \leq N} \in \Omega$ differ only in the j'th component (i.e., $\omega_i = \omega_i'$ if $i \neq j$), then $|Y(\omega) - Y(\omega')| \leq \tau$.*

Further, assume that $\mathrm{E}(Y)$ *exists. Then* $\mathrm{P}\left[|Y - \mathrm{E}(Y)| \geq \lambda\right] \leq 2\exp\left(-\lambda^2/(2\tau^2 N)\right)$ *for all* $\lambda > 0$.

To derive Lemma 33 from Lemma 45, we let $\mathcal{P} = \{\{v,w\} : v,w \in V, v \neq w\}$ be the set of all $\binom{n}{2}$ possible edges. Further, for each $e = \{v,w\} \in \mathcal{P}$ we let Ω_e denote a Bernoulli experiment with success probability $p_{\psi(v)\psi(w)}$. Then we have the product decomposition $G_{n,k}(\psi,\boldsymbol{p}) = \prod_{e\in\mathcal{P}} \Omega_e$, because the edges occur independently in $G_{n,k}(\psi,\boldsymbol{p})$. However, we cannot apply Lemma 45 to this decomposition directly, because the number of factors is too large. Therefore, we are going to set up a different product decomposition $G_{n,k}(\psi,\boldsymbol{p}) = \prod_{i=1}^{K} \Omega_i$, where each Ω_i is a product of several Ω_e.

To this end, we partition $\mathcal{P}$ into $K \leq 2\sigma^* n/\ln n$ subsets $\mathcal{P}_1,\ldots,\mathcal{P}_K$ such that $\mathrm{E}(\#E(G_1\cup G_2)\cap\mathcal{P}_i) = \sum_{e\in\mathcal{P}_i} \mathrm{P}\left[e \in G_1 \cup G_2\right] \leq \ln n$ for all $1 \leq i \leq K$; here G_1, G_2 are the graphs defined in (4), (5). Then we have the decomposition

$$G_{n,k}(\psi,\boldsymbol{p}) = \prod_{i=1}^{K} \Omega_i, \quad \text{where } \Omega_i = \prod_{e\in\mathcal{P}_i} \Omega_e. \tag{84}$$

Let us call $\mathcal{P}_i$ *critical* if $\#E(G_1 \cup G_2) \cap \mathcal{P}_i > 100\ln n$. As $\#E(G_1 \cup G_2) \cap \mathcal{P}_i$ is a sum of mutually independent Bernoulli variables, the generalized Chernoff bound (74) entails that $\mathrm{P}\left[\mathcal{P}_i \text{ is critical}\right] \leq n^{-21}$. Therefore, by the union bound

$$\mathrm{P}\left[\exists i : \mathcal{P}_i \text{ is critical}\right] \leq n^{-19}. \tag{85}$$

Now, for $G = G_{n,k}(\psi,\boldsymbol{p})$ we define $\tilde{G} = G - \bigcup_{i:\mathcal{P}_i \text{ is critical}} E(G_1)\cap\mathcal{P}_i + \bigcup_{i:\mathcal{P}_i \text{ is critical}} E(G_2)\cap\mathcal{P}_i$ and set $Y(G) = X(\tilde{G})$. Then (85) yields

$$\mathrm{P}\left[X(G_{n,k}(\psi,\boldsymbol{p})) = Y(G_{n,k}(\psi,\boldsymbol{p}))\right] \geq 1 - n^{-19}. \tag{86}$$

Furthermore, by the Lipschitz condition (75) we have $|X(G) - Y(G)| \leq n^2$ for *all* possible outcomes $G = G_{n,k}(\psi,\boldsymbol{p})$. Therefore, (86) entails that

$$|\mathrm{E}\left(X(G_{n,k}(\psi,\boldsymbol{p}))\right) - \mathrm{E}\left(Y(G_{n,k}(\psi,\boldsymbol{p}))\right)| \leq n^{2-19} \leq 1. \tag{87}$$

Moreover, we claim that for all $1 \leq j \leq K$

if G, G' are such that $G - \mathcal{P}_j = G' - \mathcal{P}_j$, i.e., G, G' differ only on edges corresponding to the factor Ω_j, then $|Y(G) - Y(G')| \leq 200\ln n$ (88)

To prove (88), we let G_1, G_2 and G_1', G_2' be the decompositions of G and G' into the sparse/dense part as defined in (4), (5).

1st case: neither in G nor in G' the set $\mathcal{P}_j$ is critical. Then $\tilde{G}'$ can be obtained from $\tilde{G}$ by either adding or removing the edges in $\mathcal{P}_j \cap (E(G) \triangle E(G'))$. Since $\mathcal{P}_j$ is not critical in both G and G', we have $\#\mathcal{P}_j \cap (E(G) \triangle E(G')) \leq 200\ln n$, so that (88) follows from the Lipschitz condition (75).

2nd case: $\mathcal{P}_j$ is critical in both G and G'. Then $\tilde{G}' = \tilde{G}$, so that $Y(G) = Y(G')$.

3rd case: $\mathcal{P}_j$ is critical in G but not in G'. Then $\tilde{G}'$ is obtained from $\tilde{G}$ by adding or removing the edges in $\mathcal{P}_j \cap E(G')$; since $\#\mathcal{P}_j \cap E(G') \leq 100\ln n$, the Lipschitz condition (75) implies (88).

4th case: $\mathcal{P}_j$ is critical in G' but not in G. Analogous to the 3rd case.

Due to (88), Lemma 45 applied to $Y(G_{n,k}(\psi,\boldsymbol{p}))$ and the decomposition (84) yields

$$\mathrm{P}\left[|Y(G_{n,k}(\psi,\boldsymbol{p})) - \mathrm{E}(Y(G_{n,k}(\psi,\boldsymbol{p})))| > \frac{1}{2}\sqrt{\sigma^* n}\ln^2 n\right] \leq \exp\left[-\frac{\sigma^*\ln^4 n}{160000K\ln^2 n}\right] \leq n^{-11}, \tag{89}$$

provided that n is sufficiently large. Thus, we finally obtain

$$\begin{aligned}
&\mathrm{P}\left[|X(G_{n,k}(\psi,\boldsymbol{p})) - \mathrm{E}(X(G_{n,k}(\psi,\boldsymbol{p})))| \geq \sqrt{\sigma^* n}\ln^2 n\right] \\
\leq\;& \mathrm{P}\left[X(G_{n,k}(\psi,\boldsymbol{p})) \neq Y(G_{n,k}(\psi,\boldsymbol{p}))\right] + \mathrm{P}\left[|Y(G_{n,k}(\psi,\boldsymbol{p})) - \mathrm{E}(X(G_{n,k}(\psi,\boldsymbol{p})))| \geq \sqrt{\sigma^* n}\ln^2 n\right] \\
\overset{(86),(87)}{\leq}\;& n^{-19} + \mathrm{P}\left[|Y(G_{n,k}(\psi,\boldsymbol{p})) - \mathrm{E}(Y(G_{n,k}(\psi,\boldsymbol{p})))| \geq \frac{1}{2}\sqrt{\sigma^* n}\ln^2 n\right] \overset{(89)}{\leq} n^{-19} + n^{-11} \leq n^{-10},
\end{aligned}$$

as desired.

9.2 Proof of Lemma 34

Since for all v and $a = 1, 2$ the degree $d_{G_a}(v)$ of v in G_a is a sum of mutually independent Bernoulli variables with mean $\leq 2\sigma^*$, the Chernoff bound (74) entails that $\mathrm{P}\left[v \in U_i\right] \leq \exp\left[-\frac{1}{3}2^i\sigma^*\right]$. Hence, $\mathrm{E}(\#U_i) \leq \exp\left[-\frac{1}{3}2^i\sigma^*\right] n$. To obtain a bound on $\#U_i$ that actually holds w.h.p., we consider two cases.

1st case: $2^i\sigma^* \geq 24 \ln\ln n$**.** By Markov's inequality, we have

$$\mathrm{P}\left[\#U_i > \exp\left(-2^{i-2}\sigma^*\right) n\right] \leq \frac{\mathrm{E}(\#U_i)}{\exp\left[-2^{i-2}\sigma^*\right] n} \leq \exp\left[-2^i\sigma^*/12\right] \leq \ln^{-2} n. \tag{90}$$

2nd case: $2^i\sigma^* < 24 \ln\ln n$**.** Then $\exp\left[-\frac{1}{3}2^i\sigma^*\right] n \geq n^{1-o(1)}$. Therefore, Lemma 33 yields

$$\mathrm{P}\left[\#U_i > 2\exp\left(-\frac{1}{3}2^i\sigma^*\right) n\right] \leq \mathrm{P}\left[\#U_i - \mathrm{E}(\#U_i) \geq \sqrt{\sigma^* n}\ln^2 n\right] \leq n^{-10}. \tag{91}$$

Finally, combining (90) and (91) and invoking the union bound, we conclude that with probability $\geq 1 - O(\ln^{-1} n)$ we have $\#U_i \leq \exp\left(-2^{i-2}\sigma^*\right) n$ for all $i = 1, \ldots, \lceil \log_2 n \rceil$.

9.3 Proof of Lemma 35

For any two vertices $v, w \in V$ the probability that v, w are connected in $G_1 \cup G_2$ is

$$\mathrm{P}\left[\{v, w\} \in E(G_1 \cup G_2)\right] \leq 2p_{\psi(v)\psi(w)}(1 - p_{\psi(v)\psi(w)}) \leq \frac{2\sigma^*}{n_{\min}}. \tag{92}$$

Let $S \subset V$ be a set of cardinality $s = \#S \leq s_{\max} = n\left(\frac{n_{\min}}{n\sigma^*}\right)^2$. As there are $\binom{\binom{s}{2}}{10s}$ ways to choose a graph with vertex set S that contains $10s$ edges, the union bound entails in combination with (92) that $\mathrm{P}\left[e_{G_1\cup G_2}(S) \geq 10s\right] \leq \binom{\binom{s}{2}}{10s}\left(\frac{2\sigma^*}{n_{\min}}\right)^{10s} \leq \left(\frac{\mathrm{e}s\sigma^*}{10n_{\min}}\right)^{10s}$.Hence, once more due to the union bound we obtain that $P_s = \mathrm{P}\left[\exists S \subset V : \#S \leq s_{\max} \wedge e_{G_1\cup G_2}(S) \geq 10\#S\right] \leq \binom{n}{s}\left(\frac{\mathrm{e}s\sigma^*}{10n_{\min}}\right)^{10s}$.Consequently, we can estimate P_s as follows:

$$\binom{n}{s}P_s \leq \left[\left(\frac{\mathrm{e}n}{s}\right)^2\left(\frac{\mathrm{e}s\sigma^*}{10n_{\min}}\right)^{10}\right]^s \leq \left(\frac{n_{\min}}{n\sigma^*}\right)^{8s} \leq 1. \tag{93}$$

Thus, for any $s_{\min} \geq 1$ we have

$$\mathrm{P}\left[\exists S \subset V : s_{\min} \leq \#S \leq s_{\max} \wedge e_{G_1\cup G_2}(S) \geq 10\#S\right] \leq \sum_{s=s_{\min}}^{s_{\max}} P_s \overset{(93)}{\leq} 2\binom{n}{s_{\min}}^{-1}.$$

Finally, (94) entails that w.h.p. there is no set $S \subset V$ of cardinality $1 \leq \#S \leq s_{\max}$ such that $e_{G_1\cup G_2}(S) \geq 10\#S$, whence the first part of Lemma 35 follows. Furthermore, setting $s_{\min} = \lceil \ln^3 n \rceil$ in (94), we obtain the second assertion.

9.4 Proof of Lemma 36

The proof relies on the following two general lemmas, which are implicit in the work of Alon and Kahale, Feige and Ofek, and Füredi and Komloś [1, 17, 20]; both lemmas are stated and proved explicitly in [11, Chapter 5].

Lemma 46. *There are constants c_1, c_2 such that the following holds. Let $(a_{ij})_{1\leq i<j\leq \nu}$ be a family of mutually independent Bernoulli random variables with mean $0 \leq p \leq 1$. Set $a_{ij} = a_{ji}$ for $1 \leq j < i \leq \nu$, and let $a_{ii} = 0$ for all $1 \leq i \leq \nu$. Moreover, let $A = (a_{ij})_{1\leq i,j\leq \nu}$ and $M = p\boldsymbol{J} - A$. Further, let $d \geq 0$, and set $X = \{i \in \{1, \ldots, \nu\} : \sum_{j=1}^{\nu} a_{ij} \leq d\}$. Then with probability $\geq 1 - O(\nu^{-1})$ we have $\|M_X\| \leq c_2\sqrt{\max\{\nu p, d\}}$.*

Lemma 47. *There are constants c_1, c_2 such that the following holds. Let $(a_{ij})_{1\le i,j\le\nu}$ be a family of mutually independent Bernoulli random variables with mean $0 \le p \le 1$. Moreover, let $A = (a_{ij})_{1\le i,j\le\nu}$ and $M = p\boldsymbol{J} - A$. Further, let $d \ge 0$, and set $X = \{i \in \{1,\ldots,\nu\} : \sum_{j=1}^{\nu} a_{ij} + a_{ji} \le d\}$. Then with probability $\ge 1 - O(\nu^{-1})$ we have $\|M_X\| \le c_2\sqrt{\max\{\nu p, d\}}$.*

Proof of Lemma 36. Let $A = A(G)$ be the adjacency matrix, and set $M^{(i,j)} = p_{ij}\boldsymbol{J}_{V_i\times V_j} - A_{V_i\times V_j}$. Then by Lemmas 46 and 47 (applied to the matrices $A_{V_i\times V_j}$) for all i, j such that $p_{ij} \le \frac{1}{2}$ w.h.p. we have

$$\left\|M_X^{(i,j)}\right\| \le c\sqrt{\max\{\varDelta, \sigma^*\}} \tag{94}$$

for a certain constant $c > 0$. Furthermore, applying Lemmas 46 and 47 to $\boldsymbol{J}_{V_i\times V_j} - A_{V_i\times V_j}$, we conclude that w.h.p. (94) holds for all i, j such that $p_{ij} > \frac{1}{2}$ as well.

To bound $\|M_X\|$, let $\xi, \eta \in \mathbf{R}^V$ be unit vectors. We decompose $\xi = \sum_{i=1}^{k} \xi_i$, where the entries of ξ_i equal the entries of ξ on the coordinates in V_i, and ξ_i is 0 on $V \setminus V_i$. Similarly, we let $\eta = \sum_{i=1}^{k} \eta_i$. Then

$$|\langle M_X\eta, \xi\rangle| = \left|\sum_{i,j=1}^{k} \left\langle M_X^{(i,j)}\eta_j, \xi_i\right\rangle\right| \le \sum_{i,j=1}^{k} \left\|M_X^{(i,j)}\right\| \cdot \|\xi_i\| \cdot \|\eta_j\| \overset{(94)}{\le} \le ck\sqrt{\max\{\varDelta, \sigma^*\}},$$

because $\sum_{i=1}^{k} \|\xi\|^2 = \sum_{i=1}^{k} \|\eta_i\|^2 = 1$. Thus, w.h.p. we have $\|M_X\| = \sup_{\xi,\eta:\|\xi\|=\|\eta\|=1} |\langle M_X\eta, \xi\rangle| \le ck\sqrt{\max\{\varDelta, \sigma^*\}}$, as desired. □

References

1. Alon, N., Kahale, N.: A spectral technique for coloring random 3-colorable graphs. SIAM J. Comput. **26** (1997) 1733–1748
2. Alon, N, Krivelevich, M., Sudakov, B.: Finding a large hidden clique in a random graph. Random Structures and Algorithms **13** (1998) 457–466
3. Azar, Y., Fiat, A., Karlin, A.R., McSherry, F., Saia, J.: Spectral analysis of data. Proc. 33rd STOC (2001) 619–626
4. Bilu, Y., Linial, N.: Are stable instances easy? Preprint (2004)
5. Bollobás, B., Scott, A.D.: Max cut for random graphs with a planted partition. Combinatorics, Probability and Computing **13** (2004) 451–474
6. Boppana, R.: Eigenvalues and graph bisection: an average-case analysis. Proc. 28th FOCS (1987) 280–285
7. Bui, T., Chaudhuri, S., Leighton, T., Sipser, M.: Graph bisection algorithms with good average case behavior. Combinatorica **7** (1987) 171–191
8. Chen, H., Frieze, A.: Coloring bipartite hypergraphs. Proc. 5th IPCO (1996) 345–358
9. Chung, F., Graham, R.: Sparse quasi-random graphs. Combinatorica **22** (2002) 217–244
10. Coja-Oghlan, A.: A spectral heuristic for bisecting random graphs. Proc. 16th SODA (2005) 850–859
11. Coja-Oghlan, A.: Spectral techniques, semidefinite programs, and random graphs. Habilitation thesis, Humboldt University Berlin, 2005.
12. Dasgupta, A., Hopcroft, J.E., McSherry, F.: Spectral analysis of random graphs with skewed degree distributions. Proc. 45th FOCS (2004) 529–537
13. Dasgupta, A., Hopcroft, J.E., Kannan, R., Mitra, P.: Spectral partitioning by recursive partitioning. To appear in Proc. ESA (2006)
14. Ding. C.H.Q.: Analysis of gene expression profiles: class discovery and leaf ordering. Proc. 6th International Conference on Computational biology (2002) 127–136
15. Dyer, M., Frieze, A.: The solution of some NP-hard problems in polynomial expected time. Journal of Algorithms **10** (1989) 451–489
16. Feige, U., Kilian, J.: Heuristics for semirandom graph problems. Journal of Computer and System Sciences **63** (2001) 639–671
17. Feige, U., Ofek, E.: Spectral techniques applied to sparse random graphs. Random Structures and Algorithms **27** (2005) 251–275
18. Flaxman, A.: A spectral technique for random satisfiable 3CNF formulas. Proc. 14th SODA (2003) 357–363
19. Frieze, A., Kannan, R.: Quick approximation to matrices and applications. Combinatorica **19** (1999) 175–220
20. Füredi, Z., Komloś, J.: The eigenvalues of random symmetric matrices. Combinatorica **1** (1981) 233–241
21. Goerdt, A., Lanka, A.: On the hardness and easiness of random 4-SAT formulas. Preprint (2004)

22. Janson, S., Łuczak, T., Ruciński, A.: Random Graphs. Wiley (2000)
23. Jerrum, M., Sorkin, G.: The metropolis algorithm for graph bisection. Discrete Applied Mathematics **82** (1998) 155–175
24. Friedman, J., Kahn, J., Szemerédi, E.: On the second eigenvalue in random regular graphs. Proceedings of the 21st Annual ACM Symposium on Theory of Computing (1989) 587–598
25. Kannan, R., Vempala, S., Vetta, A.: On clusterings: good, bad ans spectral. Journal of the ACM **51** (2004) 497–515
26. Krivelevich, M., Sudakov, B.: The largest eigenvalue of sparse random graphs. Combinatorics, Probability and Computing **12** (2003) 61–72
27. McSherry, F.: Spectral partitioning of random graphs. Proc. 42nd FOCS (2001) 529–537
28. Pothen, A., Simon, H.D., Kang-Pu, L.: Partitioning sparse matrices with eigenvectors of graphs. SIAM J. Matrix Anal. Appl. **11** (1990) 430–452
29. 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)
30. Spielman, D.A., Teng, S.-H.: Spectral partitioning works: planar graphs and finite element meshes. Proc. 36th FOCS (1996) 96–105
31. Subramanian, C.R., Veni Madhavan, C.E.: General partitioning on random graphs. Journal of Algorithms **42** (2002) 153–172

ECCC ISSN 1433-8092

http://eccc.hpi-web.de/