Combinatorics, Probability and Computing (2010) 19, 227–284.
doi:10.1017/S0963548309990514

Graph Partitioning via Adaptive Spectral Techniques

AMIN COJA-OGHLAN†

School of Informatics, University of Edinburgh, Crichton Street, Edinburgh EH8 9AB, UK
(e-mail: acoghlan@inf.ed.ac.uk)

Received 23 July 2007; accepted 10 September 2009; first published online 13 November 2009

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 (very nearly) 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' with respect to the matrix $\mathcal{E}$. This result in particular applies to *sparse* graphs with bounded average degree as $n = \#V \to \infty$, and it has various consequences on partitioning 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 eigenvectors to solve the combinatorial problem in question. Spectral techniques are used to either deal with 'classical' NP-hard graph partitioning problems such as Graph Colouring or Max Cut, or to solve 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 [30], or bio-informatics [14].

Despite their success in applications (*e.g.*, [29], [30]), for most of the known spectral heuristics there are counter-examples 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

† Supported by DFG CO 646. A preliminary version appeared in *ICALP* 2006.

problem, quite a few authors have contributed rigorous analyses of spectral techniques on suitable models of *random graphs*. For example, Alon and Kahale [1] analysed a spectral technique for Graph Colouring, Alon, Krivelevich and 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] has suggested a spectral heuristic for random 3-SAT.

While the algorithmic techniques of [1, 2, 6, 10, 18] are tailored to the concrete problems (and random graph models) studied in the respective articles, in a remarkable paper McSherry [28] 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 `Partition` that does capture all the previous work [1, 2, 6, 10, 18, 28], and that is indeed applicable to much more general settings. We first consider the general random graph model studied by McSherry and develop new algorithmic techniques that apply to the regime of the parameters not covered in [28], including the case of sparse graphs. Then we study the so-obtained spectral heuristic from a *deterministic* point of view, *i.e.*, without referring to any specific random graph model. More precisely, we single out (reasonably modest) conditions on the input graph that ensure the success of the spectral heuristic. To this end, we employ notions related to the concept of *quasi-random graphs* (see Chung and Graham [9] or Krivelevich and Sudakov [27]). The deterministic result thus obtained has a rather broad scope and comprises or improves a number of prior results on spectral methods (see Section 2). Indeed, in order to obtain such a general deterministic result we need to come up with new ideas for *analysing* the spectral heuristic, because the use of 'standard' techniques from the theory of random graphs would impose far too restrictive conditions on the type of input instances.

A further important aspect is that `Partition` is *adaptive* in the sense that its input *only* consists of the graph G and the desired number of vertex classes k. Thus, the algorithm does not need any further *a priori* 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). This aspect requires novel algorithmic ingredients.

1.2. A general random graph model

In this section we state the spectral partitioning result for the random graph model $G_{n,k}(\psi, \mathbf{p})$ from [28]. To define $G_{n,k}(\psi, \mathbf{p})$, we need a bit of notation. Let $V = \{1, \ldots, n\}$ be a vertex set. Moreover, consider a map $\psi : V \to \{1, \ldots, k\}$, set $V_i = \psi^{-1}(v)$, and let $n_{\min} = \min_{1 \leqslant i \leqslant k} \# V_i$. We think of $V_1, \ldots, V_k$ as the 'latent' partition of V that we are to recover. Moreover, consider a symmetric $k \times k$ matrix $\mathbf{p} = (p_{ij})_{1 \leqslant i,j \leqslant k}$, and let $\mathcal{E} = \mathcal{E}(\psi, \mathbf{p}) = (\mathcal{E}_{vw})_{v,w \in V}$ be the $n \times n$ matrix with entries $\mathcal{E}_{vw} = p_{\psi(v)\psi(w)}$.

We define the random graph $G_{n,k}(\psi, \mathbf{p})$ as follows: the vertex set of $G_{n,k}(\psi, \mathbf{p})$ is $V = \{1, \ldots, n\}$, and any two vertices $v, w \in V$ are connected with probability $\mathcal{E}_{vw}$ independently. Thus, $G_{n,k}(\psi, \mathbf{p})$ has a 'planted' partition $V_1, \ldots, V_k$, with the expected edge density between any two sets V_i, V_j being p_{ij}, and the expected density inside each V_i being p_{ii}. As we shall see in Section 2, $G_{n,k}(\psi, \mathbf{p})$ comprises various random graph models for specific partitioning problems such as Graph Colouring or Max Cut.

As we are interested in asymptotic results that hold as n gets 'large', we say that $G_{n,k}(\psi, \mathbf{p})$ has some property $\mathcal{P}$ *with high probability* ('w.h.p.') if the following is true. For any $\varepsilon > 0$ and any integer $k > 1$ there exists $n_0 = n_0(\varepsilon, k)$ such that, for all $n > n_0$, all matrices $\mathbf{p}$, and all maps ψ, the probability that the random graph $G_{n,k}(\psi, \mathbf{p})$ has property $\mathcal{P}$ is at least $1 - \varepsilon$.

To state the algorithmic result, we need to introduce a few parameters of the random graph $G = G_{n,k}(\psi, \mathbf{p})$. Since any two vertices $v, w \in V$ are connected in $G_{n,k}(\psi, \mathbf{p})$ with probability $\mathcal{E}_{vw}$ independently, the variance of the degree of v equals $\sum_{w \in V \setminus \{v\}} \mathcal{E}_{vw}(1 - \mathcal{E}_{vw})$. Thus,

$$\sigma^* = \max_{1 \leqslant i \leqslant k} \sum_{j=1}^{k} \# V_j p_{ij}(1 - p_{ij}) = \max_{v \in V} \sum_{w \in V} \mathcal{E}_{vw}(1 - \mathcal{E}_{vw}) \tag{1.1}$$

is 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

$$\Phi_{vw} = \begin{cases} 1 & \text{if } \mathcal{E}_{vw} > \frac{1}{2}, \\ 0 & \text{otherwise.} \end{cases} \tag{1.2}$$

Then we define

$$G_1 = (V, E_1), \quad \text{where} \quad E_1 = \{\{v, w\} \in E : \Phi_{vw} = 0\}, \tag{1.3}$$

$$G_2 = (V, E_2), \quad \text{where} \quad E_2 = \{\{v, w\} \notin E : \Phi_{vw} = 1,\, v, w \in V,\, v \neq w\}. \tag{1.4}$$

Let $d_{G_1 \cup G_2}(v)$ denote the degree of v in the graph $G_1 \cup G_2 = (V, E_1 \cup E_2)$. Considering $G_1 \cup G_2$ is helpful, because in $G =$ some of the pairs V_i, V_j may be very sparse (*i.e.*, $p_{ij} = o(1)$), while others may be very dense (*i.e.*, $p_{ij} = 1 - o(1)$). Since in $G_1 \cup G_2$ we replace the 'dense' parts of G by their complements, we obtain a graph in which all pairs V_i, V_j are sparse, thus avoiding case distinctions.

Theorem 1.1. *Suppose that the following three conditions hold.*

R1. $\sigma^* \geqslant \ln^2(n/n_{\min})$.

R2. $n_{\min} \geqslant \ln^{30} n$.

R3. *Let* $\mathcal{E}_v = (\mathcal{E}_{wv})_{w \in V}$ *denote the* v*-column of* $\mathcal{E}$. *For all* $u, v \in V$ *such that* $\psi(u) \neq \psi(v)$, *we have*

$$\|\mathcal{E}_u - \mathcal{E}_v\|^2 \geqslant \rho^2 = \frac{c_0 k^3 \sigma^*}{n_{\min}} + c_0 \ln\left(\sigma^* + \frac{n}{n_{\min}}\right) \max_{1 \leqslant i \leqslant k} \sum_{j=1}^{k} p_{ij}(1 - p_{ij}), \tag{1.5}$$

where c_0 *is a sufficiently large constant.*

Then w.h.p. $G = G_{n,k}(\psi, \mathbf{p})$ *has an induced subgraph* H *on* $\geqslant n - \sigma^{*\,-4} n_{\min}$ *vertices such that the following is true.*

https://doi.org/10.1017/S0963548309990514 Published online by Cambridge University Press

- *On input* (G,k), *the polynomial-time algorithm* `Partition` *outputs a partition* $(T_i)_{1\leqslant i\leqslant k}$ *of* G *such that* $H\cap T_i = H\cap V_{\tau(i)}$ *for some permutation* τ *of the indices* $\{1,\ldots,k\}$.
- *All components of the graph* $(G_1\cup G_2)-H$ *obtained from* $G_1\cup G_2$ *by removing the vertices of* H *have at most* $\ln n$ *vertices.*

The first two assumptions **R1** and **R2** of Theorem 1.1 are merely of technical relevance. Condition **R2** just guarantees that none of the 'planted' classes is extremely small. Assumption **R1** is needed to rule out the case that there is an extremely small partition class (of size $o(n)$) while either the random graph or its complement is extremely sparse. Thus, the crucial assumption is **R3**. Since for each $v\in V$ the vector $\mathcal{E}_v=(\mathcal{E}_{wv})_{w\in V}$ represents the probabilities that v is connected to other vertices $w\in V$, $\|\mathcal{E}_u-\mathcal{E}_v\|^2$ measures the 'expected difference' between two vertices u,v with respect to the planted partition. Hence, **R3** quantifies how 'strongly' any two planted classes V_i,V_j must be separated to ensure that `Partition` succeeds. More precisely, the required lower bound ρ^2 increases as a function of σ^* and decreases as a function of $n_{\min}$. That is, the higher the 'random noise level' in $G_{n,k}(\psi,\mathbf{p})$ (*i.e.*, σ^*), and the smaller the size of the classes that we are to recover (*i.e.*, $n_{\min}$), the more the desired partition must 'stand out'.

Theorem 1.1 ensures that under conditions **R1**–**R3** `Partition` w.h.p. yields a partition $T_1,\ldots,T_k$ that coincides with the planted one *on a large subgraph* H (up to a permutation of the indices). Thus, `Partition` will in general not recover the planted partition $V_1,\ldots,V_k$ perfectly. In fact, in general it is plainly *impossible* to compute $V_1,\ldots,V_k$ perfectly w.h.p. (at least for 'small' values of σ^*). To see this, assume that $p_{ij}<\frac{1}{2}$ for all i,j and $\sigma^*=O(1)$. Then in $G=G_{n,k}(\psi,\mathbf{p})$ all vertex degrees have a bounded mean (namely, at most $2\sigma^*$). Therefore, for each vertex v the probability that v is isolated in $G_{n,k}(\psi,\mathbf{p})$ is $\Omega(1)$, whence each class V_i contains $\Omega(\#V_i)$ isolated vertices w.h.p. Since, on input (G,k), it is impossible to tell which isolated vertex stems from which of the planted classes $V_1,\ldots,V_k$, *no* algorithm can recover $V_1,\ldots,V_k$ entirely.

We can characterize the subgraph H on which `Partition` succeeds as follows. Set

$$d(v,w)=e(v,V_j)/\#V_j \quad \text{for } v\in V \text{ and } w\in V_j, \quad \text{and let } d(v)=(d(v,w))_{w\in V}\in\mathbb{R}^V, \tag{1.6}$$

where $e(v,V_j)$ denotes the number of v–V_j-edges in G. Hence, $d(v,w)$ is the 'empirical' edge density between the vertex v and the class of w, so that $\mathrm{E}(d(v,w))=\mathcal{E}_{vw}$ for all w and $\mathrm{E}(d(v))=\mathcal{E}_v$. Now, the induced subgraph H basically consists of those vertices v such that $d(v)$ is 'close' to its expectation $\mathcal{E}_v$. More precisely, set $\lambda=\sigma^*$; then H is the largest induced subgraph of G that enjoys the following four properties.

H1. $\#V\setminus H\leqslant\lambda^{-4}n_{\min}$ and $\sum_{v\in V\setminus H}d_{G_1\cup G_2}(v)^2\leqslant n_{\min}$.
H2. For all $v\in H$ the vector $d(v)$ defined in (1.6) satisfies $\|\mathcal{E}_v-d(v)\|^2\leqslant 0.001\rho^2$.
H3. All $v\in H$ have degree $\leqslant 10\sigma^*$ 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 neighbours in $V\setminus H$.

Thus, **H1** requires that H constitutes a 'large' share of G, and that the vertices outside 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 terms of the parameter ρ. In addition, **H3** requires that the vertices $v\in H$ do not have too high a degree in $G_1\cup G_2$, and **H4** means that H

should be 'well separated' from $V \setminus H$. We call the largest subgraph H of G that satisfies **H1**–**H4** the *core* of G.

If σ^* is sufficiently large (say, $\sigma^* \gg \ln n$), then $H = G$. This is essentially the situation considered in [28]. However, as discussed above, if σ^* is 'small' (say, bounded as $n \to \infty$), then H will be a proper subgraph of G w.h.p.

1.3. The deterministic result

The main result of this paper is that `Partition` actually works under a few *deterministic* assumptions on the input graph G. This result is significantly stronger than Theorem 1.1, because it covers considerably more general types of input graphs than $G_{n,k}(\psi, \mathbf{p})$ (*e.g.*, random graphs in which the edges are dependent or pseudo-random graphs). To state these conditions, let $V = \{1, \ldots, n\}$, let $G = (V, E)$ be a graph, let $k \geqslant 2$ be an integer, let $\psi : V \to \{1, \ldots, k\}$, and let $\mathbf{p} = (p_{ij})_{1 \leqslant i,j \leqslant k}$ be a symmetric $k \times k$ matrix. Then we can define classes $V_i = V_i(\psi) = \psi^{-1}(v)$, set $n_{\min} = n_{\min}(\psi) = \min_{1 \leqslant i \leqslant k} \# V_i$, and define an $n \times n$ matrix $\mathcal{E} = \mathcal{E}(\psi, \mathbf{p}) = (\mathcal{E}_{vw})_{v,w \in V}$ by letting $\mathcal{E}_{vw} = p_{\psi(v)\psi(w)}$. Moreover, we let $\sigma^* = \sigma^*(\psi, \mathbf{p})$, $\Phi = \Phi^*(\psi, \mathbf{p})$, $G_1 = G_1(\psi, \mathbf{p})$, and $G_2 = G_2(\psi, \mathbf{p})$ be as defined in (1.1)–(1.4). Thus, we keep the notation from the previous section, but we no longer assume that G is a random graph. The four deterministic assumptions that we need are the following.

Low-rank structure. The most important assumption is that the matrix $\mathcal{E}$ provides an underlying 'low-rank structure' for the adjacency matrix A of G. 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.

A1. Let A be the adjacency matrix of G, and let $M = \mathcal{E} - A$. There is a number λ such that $\sigma^* \leqslant \lambda \leqslant \sigma^* \cdot \min\{\sigma^*, n_{\min}/\ln n\}$ with the following property. For any $\Delta > 0$ the set $D(\Delta) = \{v \in V : d_{G_1 \cup G_2}(v) \leqslant \Delta\}$ satisfies $\|M_{D(\Delta)}\| \leqslant c_0 k\sqrt{\lambda + \Delta}$, 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 following 'separation' condition as well. In terms of quasi-randomness, the expression λ/σ^{*2} basically corresponds to the 'spectral gap' of the adjacency matrix. We shall see in Section 2 that the occurrence of Δ in the bound in **A1** is actually necessary.

Separation. This condition quantifies 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 (see the **R3** condition in Theorem 1.1).

A2. Let $\rho = c_0^4\sqrt{k^3\lambda/n_{\min}}$ (with λ and c_0 are as in **A1**). Then, for all $u, v \in V$ such that $\psi(u) \neq \psi(v)$, we have $\|\mathcal{E}_u - \mathcal{E}_v\| \geqslant \rho$.

Note the dependence of ρ on λ: the more tightly $\mathcal{E}$ approximates A, the more 'subtle' the differences between $\mathcal{E}_u$ and $\mathcal{E}_v$ can be.

https://doi.org/10.1017/S0963548309990514 Published online by Cambridge University Press

Approximate regularity. The next item is a 'regularity' condition on the degree distribution of G. 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, we let $\mu(v,S) = \sum_{w \in S \setminus \{v\}} \mathcal{E}_{vw}$.

A3. For all $v \in V_j$ we have

$$\max_{1 \leqslant i \leqslant k} |e(v,V_i) - \mu(v,V_i)| \leqslant 0.1 \left(\frac{1}{k}\sigma^* + \#V_i p_{ij}(1-p_{ij}) \right) + \ln^2 n.$$

This condition requires that $e(v,V_i)$ should be close to the number $\mu(v,V_i)$ of edges that we would expect if G were a random graph $G_{n,k}(\psi,\mathbf{p})$. The error term on the right-hand side involves the 'maximum variance' σ^* and 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 (see Section 2).

Lower bound on $n_{\min}$. Finally, we need that all classes V_i have at least polylogarithmic size.

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

The following theorem shows that if G satisfies **A1**–**A4**, and if H is a subgraph of G that satisfies conditions **H1**–**H4** from Section 1.2, then `Partition` can recover the desired partition on H.

Theorem 1.2. *There are a polynomial-time algorithm* `Partition` *and a constant $C > 0$ such that, for each $c_0 > C$ and each integer $k \geqslant 2$, there exists a number n_0 such that the following is true. Suppose that $n \geqslant n_0$, and that $G = (V,E)$ is a graph with vertex set $V = \{1,\ldots,n\}$ so that there exist* $\mathbf{p}$ *and ψ such that:*

- $\sigma^* \geqslant c_0$,
- **A1**–**A4** *hold, and*
- *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 $\{1,\ldots,k\}$.*

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 seems to be an important feature from a practical point of view.

1.4. Related work

Conditions **A1**–**A4** in Theorem 1.2 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 [31] 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 1.2 is the first result that shows that under general *deterministic* assumptions a 'good' low-rank structure can be recovered in polynomial time if there is one.

The $G_{n,k}(\psi, \mathbf{p})$ model was first considered by McSherry [28], who presented a polynomial-time algorithm that recovers the planted partition of $G = G_{n,k}(\psi, \mathbf{p})$, provided that the following holds. Let $\sigma_{\max}^2 = \max_{1 \leqslant i,j \leqslant k} p_{ij}(1 - p_{ij})$, and let $c_0 > 0$ be a sufficiently large constant; then the assumption reads

$$\|\mathcal{E}_u - \mathcal{E}_v\|^2 \geqslant c_0 k \cdot \max\left\{\sigma_{\max}^2, \frac{\ln^6 n}{n}\right\} \cdot \left[\frac{n}{n_{\min}} + \ln n\right] \quad \text{if} \quad \psi(u) \neq \psi(v). \tag{1.7}$$

The two conditions (1.7) and (1.5) compare as follows. Due to the $\ln n$-terms occurring in (1.7), $G_{n,k}(\psi, \mathbf{p})$ must have average degree at least $\ln^3 n$ (and $\leqslant n - \ln^3 n$). In contrast, Theorem 1.1 also comprises the following three types of graphs.

Sparse graphs. Condition (1.5) 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 $1 \leqslant j \leqslant k$. In this case the average degree of $G_{n,k}(\psi, \mathbf{p})$ is bounded as $n \to \infty$.

Massive graphs. Similarly, (1.5) allows that $\mu(v, V_j) = \#V_j - O(1)$ for all v, j. Then $G_{n,k}(\psi, \mathbf{p})$ 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 have *either* $\mu(v, V_j) = O(1)$ *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 [28] 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 (see Section 2).

Furthermore, condition (1.7) is phrased in terms of $n\sigma_{\max}^2$, which may exceed the expression σ^* from (1.1) significantly if, *e.g.*, $G_{n,k}(\psi, \mathbf{p})$ features a 'small' part (say, of size $n^{0.1}$) of density $\frac{1}{2}$. In this case (1.5) can be a considerably weaker assumption than (1.7).

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 [28] 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 $n_{\min}$).

Dasgupta, Hopcroft, Kannan and Mitra [13] studied the 'second eigenvector technique' on $G_{n,k}(\psi, \mathbf{p})$; 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 $\geqslant \text{polylog}(n)$ and $\leqslant n - \text{polylog}(n)$), the authors obtain a result similar to [28] (but with a slightly weaker separation assumption).

Whereas in the present paper we are just dealing with the problem of recovering a 'latent' partition of a given graph, there are several 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 re-derived easily from Theorem 1.1 as well as from [28]. Further related results that involve partitioning sparse random graphs (constant average degree) include Alon and Kahale [1] (3-colouring), Boppana [6] and Coja-Oghlan [10] (Minimum Bisection), Chen and Frieze [8] (hypergraph 2-colouring), Flaxman [18] (3-SAT), and Goerdt and Lanka [21] (4-NAE-SAT). These results can only be partially derived using the techniques of [28] (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 1.1 and 1.2.

A few authors have analysed spectral techniques on random graphs that cannot be described in terms of the $G_{n,k}(\psi, \mathbf{p})$ 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 also be recovered in this case w.h.p. under an assumption similar to (1.7). Thus, it is assumed that the average degree is $\geqslant \text{polylog}(n)$. Applied to the $G_{n,k}(\psi, \mathbf{p})$ model, [12] yields a result similar to [28].

Moreover, Dasgupta, Hopcroft, Kannan and Mitra [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 any two vertices $v, w \in V_i$ have (exactly) the same number of randomly chosen neighbours 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 not comparable to $G_{n,k}(\psi, \mathbf{p})$. In fact, due to the very regular degree distribution, the model in [13] behaves actually quite similarly to 'dense' $G_{n,k}(\psi, \mathbf{p})$ graphs (average degree $\gg \ln n$). We shall see in Section 2 that Theorem 1.2 also captures the model introduced in [13].

Although 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, Chaudhuri, Leighton and Sipser [7] (network flows), Dyer and Frieze [15] (combinatorial methods), Feige and Kilian [16] (semidefinite programming), Jerrum and Sorkin [23] (the Metropolis process), and Subramanian and Veni Madhavan [32] (breadth-first search).

1.5. Techniques and outline

Let $G = G_{n,k}(\psi, \mathbf{p})$ be a random graph with adjacency matrix A. To recover $V_1, \ldots, V_k$, McSherry [28] 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 $\mathbb{R}^V$ onto the subspace spanned by $\zeta_1, \ldots, \zeta_k$, and let $\hat{A} = PAP$. Then $\hat{A}$ is called a *rank k approximation*

https://doi.org/10.1017/S0963548309990514 Published online by Cambridge University Press

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 $\|\hat{A}-\mathcal{E}\|_F^2 = \sum_{v\in V}\|\hat{A}_v-\mathcal{E}_v\|^2 \leqslant kn\sigma_{\max}^2$ 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 (1.7) 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 $G = G_{n,k}(\psi,\mathbf{p})$ is a *sparse* graph such that $\#V_ip_{ij} = \Theta(1)$ 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 $G_{n,k}(\psi,\mathbf{p})$ 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 (1.5) 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 no longer correspond to the desired partition.

Thus, in the situation of Theorems 1.1 and 1.2 we need a more sophisticated approach to obtain a matrix $\hat{A}$ that approximates $\mathcal{E}$ well. Following the work [1] on 3-colouring 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 thus obtained approximation $\hat{A}$ to $\mathcal{E}$ to compute an initial partition $S_1,\ldots,S_k$. The basic idea is to put $u,v\in V$ into the same S_i if and only if $\|\hat{A}_u-\hat{A}_v\| \leqslant 0.1\rho$, 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 'centres' $\xi_1,\ldots,\xi_k$ 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 $V_1,\ldots,V_k$, `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 $e(v,S_i)/\#S_i$ (this is reminiscent of the vector $d(v)$ in (1.6), whose entries represent the densities $e(v,V_i)/\#V_i$). 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. A crucial issue here is that `Improve` has to deal with classes $V_1,\ldots,V_k$ of (possibly) vastly different sizes, *e.g.*, polylog(n) versus $\Theta(n)$.

The paper is organized as follows. In Section 2 we illustrate Theorems 1.1 and 1.2 with some examples of concrete graph partitioning problems. Sections 3–7 contain the description of `Partition` and its subroutines and the proof of Theorem 1.2. Moreover, in Section 8 we apply Theorem 1.2 to the random graph $G_{n,k}(\psi,\mathbf{p})$, thereby proving Theorem 1.1. Finally, Section 9 contains the proofs of a few technical lemmas on the random graph $G_{n,k}(\psi,\mathbf{p})$.

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 $G_{n,k}(\psi, \mathbf{p})$. 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 $v \in V_i$. Note that this is well defined, because all columns $\mathcal{E}_v$ for $v \in V_i$ are identical: they represent the 'expected edge densities' between vertices in V_i and vertices in other classes. Moreover, we always let Φ denote the matrix (1.2), and let G_1, G_2 denote the graphs defined in (1.3), (1.4).

If $\mathcal{M} = (m_{vw})_{v,w \in V}$ is a matrix and $v \in V$, then $\mathcal{M}_v = (m_{wv})_{w \in V}$ is the v-column of $\mathcal{M}$. Moreover, if $X, Y \subset V$, then $\mathcal{M}_{X \times Y}$ signifies the matrix obtained from $\mathcal{M}$ by replacing all entries m_{xy} with $(x, y) \notin X \times Y$ by 0. For brevity we let $\mathcal{M}_X = \mathcal{M}_{X \times X}$. Further, we let $\|\mathcal{M}\| = \max_{\xi : \|\xi\| = 1} \|\mathcal{M}\xi\|$ denote the operator norm and let

$$\|\mathcal{M}\|_F = \sqrt{\sum_{v \in V} \|\mathcal{M}_v\|^2} = \sqrt{\sum_{v,w \in V} m_{vw}^2}$$

denote the Frobenius norm of M. If $\mathcal{M}$ a matrix of rank $\leqslant l$, then

$$\|\mathcal{M}\|^2 \leqslant \|\mathcal{M}\|_F^2 \leqslant l\|\mathcal{M}\|^2. \tag{1.8}$$

Furthermore, suppose that $\mathcal{M}$ is symmetric and let $\zeta_1, \dots, \zeta_l$ denote eigenvectors of $\mathcal{M}$ with the l largest eigenvalues in absolute value. Let P be the projection onto the space spanned by $\zeta_1, \dots, \zeta_l$. Then we call $\tilde{\mathcal{M}} = P\mathcal{M}P$ a *rank l approximation* of $\mathcal{M}$. The definition ensures that

$$\|\tilde{\mathcal{M}} - \mathcal{M}\| \leqslant \|B - \mathcal{M}\| \quad \text{for any matrix } B \text{ of rank} \leqslant l. \tag{1.9}$$

We denote the symmetric difference of two sets A, B by $A \triangle B$. That is,

$$A \triangle B = (A \cup B) \setminus (A \cap B).$$

2. Applications and examples

2.1. Graph colouring

Alon and Kahale [1] developed a spectral heuristic for colouring 3-colourable 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-colouring of $G = G_{n,k}(\psi, \mathbf{p})$. In this section we observe that the main result of [1] can be derived from Theorem 1.1 by adding only a few problem-specific details (in a similar way one can re-derive the results of [6, 10]). We also discuss how the assumptions (1.7) from [28] and (1.5) from Theorem 1.1 relate to each other.

To satisfy (1.7), we need that $p \geqslant c'(\ln^3 n)/n$ for a certain constant $c' > 0$. In this case w.h.p. *all* vertices $v \in V_i$ have $(1 + o(1))np/3$ neighbours in the other two classes V_j,

https://doi.org/10.1017/S0963548309990514 Published online by Cambridge University Press

$i \neq j$ (by Chernoff bounds), so that G is quite regular. Furthermore, let $\zeta_i \in \mathbb{R}^V$ be the characteristic vector of V_i. Then, for $i \neq j$ we have

$$A(G)(\zeta_i - \zeta_j) \sim \frac{np}{3}(\zeta_j - \zeta_i). \tag{2.1}$$

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 (1.5) of Theorem 1.1 only requires that $p \geqslant c/n$ 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.

$$\#\{v \in V_i : e(v, V_j) = \gamma\} \sim (c/3)^\gamma \exp(-c/3)n/(3\gamma!). \tag{2.2}$$

Consequently, it is *impossible* to recover the partition V_1, V_2, V_3 from G perfectly. For by (2.2) 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 1.1 and 1.2 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 colouring as nicely as in the 'dense' case. For by (2.2) 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 (2.1) 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 [28] 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 1.1 `Partition` can compute sets S_1, S_2, S_3 such that $S_i \cap H = V_i \cap H$, where $H = \mathrm{core}(G)$. Although 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-colour G. To this end, we follow the strategy of Alon and Kahale: by Theorem 1.1, $G - H$ just consists of components of size $\leqslant \ln n$. Hence, for each of these components we can compute in polynomial time a 3-colouring that extends the 3-colouring $S_1 \cap H, S_2 \cap H, S_3 \cap H$ of H. Glueing all these 3-colourings together yields the desired 3-colouring of all of G.

It is instructive to compare the 3-colouring algorithm of Alon and Kahale with the algorithm `Partition` from the present paper. The algorithm from [1] essentially proceeds in three phases. In the first ('spectral') phase the algorithm exploits the eigenvectors of a certain matrix A' to obtain a partition that differs from the planted colouring in at most, say, $0.01n$ vertices; here A' is the adjacency matrix of the graph obtained from G by removing all vertices of degree $> 10np$, say. Then, in the second ('combinatorial') phase the algorithm performs a simple local improvement strategy. This improvement strategy colours all vertices of the core H of G correctly. Finally, since all components of $G - H$ have size $O(\log n)$, in the last phase the algorithm can extend the colouring of H to a colouring of G in polynomial time.

The algorithm Partition is based on a similar strategy in that it also combines spectral and combinatorial steps (see Section 1.5). However, since in the present work we deal with a much more general type of problem, both the algorithm and the analysis require new ingredients. With respect to the spectral step, the computation of the matrix $\hat{A}$ requires a rather sophisticated algorithm, instead of just removing high-degree vertices. Furthermore, the combinatorial procedure Improve is significantly more involved, since the desired partition $V_1,\ldots,V_k$ is not necessarily optimal with respect to some simple objective function. Indeed, the analysis of the combinatorial part involves spectral arguments.

2.2. Random 3-SAT

Flaxman [18] studied the following model of random 3-SAT. Let $x_1,\ldots,x_n$ be propositional variables, and let $L=\{x_i,\bar{x}_i : 1\leqslant i\leqslant n\}$ be the set of literals. Let $p_i=c_in^{-2}$. 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) constant. 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 1.2. Observe, however, that the graph G cannot be described in terms of the $G_{n,k}(\psi,\mathbf{p})$ model, because edges do not appear independently; thus Theorem 1.1 does not apply here.

2.3. Regular graphs

Bui, Chaudhuri, Leighton and Sipser [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' neighbours in V_i and exactly d neighbours in V_{3-i} $(i=1,2)$. They 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 $d'\geqslant d+c(\sqrt{d'}+1)$ for a certain constant $c>0$. Thus, Theorem 1.2 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 $G_{n,k}(\psi,\mathbf{p})$ model, because the edges do not occur independently. However, even though G can be a sparse graph, due to its

Algorithm 3.1. `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 $(T_1, \ldots, T_k)$.

Figure 1. Pseudocode for the algorithm `Partition`.

very regular degree distribution it is much easier to deal with than a sparse random graph $G_{n,k}(\psi, \mathbf{p})$ (*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 the constant c_0 from **A1** are sufficiently large. Finally, we use the symbols Φ, σ^*, G_1, and G_2 as defined in Section 1.3. Condition **A3** readily implies the following bound on the maximum degree of $G_1 \cup G_2$:

$$d_{G_1 \cup G_2}(v) \leqslant 7\sigma^* + \ln^3 n \quad \text{for all} \quad v \in V. \tag{3.1}$$

From now on we shall summarize the functioning of `Partition` and its subroutines. We will present and analyse the subroutines `Identify`, `Approx`, and `Improve` in detail in Sections 4–7.

In steps 1–2 the goal is to compute a matrix $\hat{A}$ that approximates $\mathcal{E}$ well. If σ^* is ‘sufficiently large’ (say, $\sigma^* > \ln^3 n$), then actually any rank k-approximation $\hat{A}$ of $A(G)$ is sufficient. But if σ^* is ‘small’ then G consists of ‘extremely sparse’ and/or ‘extremely dense’ parts. In this case the tails of the degree distribution may affect the spectrum of $A(G)$ (see Section 2), and thus a rank k approximation of $A(G)$ may *not* provide a good approximation of $\mathcal{E}$. To obtain an appropriate matrix $\hat{A}$, `Partition` first needs to sort out which parts of the graph are sparse and which are dense. That is, `Partition` needs to compute the matrix Φ, whose entries Φ_{vw} are 0 if the ‘density’ $\mathcal{E}_{vw} \leqslant \frac{1}{2}$, and 1 if $\mathcal{E}_{vw} > \frac{1}{2}$. Computing Φ is the aim of the procedure `Identify`, which we will describe in Section 4. The following proposition summarizes its analysis.

Proposition 3.2. `Identify` *outputs either the matrix* Φ *or ‘fail’, and if* $\sigma^* \leqslant \ln^3 n$*, then the output is* Φ.

Thus, if `Identify` outputs ‘fail’, then `Partition` knows that σ^* is ‘big’. Therefore, in this case it just computes a rank k approximation $\hat{A}$ of $A(G)$. On the other hand, if `Identify` does not fail, `Partition` feeds the resulting matrix φ into the subroutine `Approx`, which we will present in Section 5. The following proposition shows that `Approx` will provide a good approximation $\hat{A}$ to $\mathcal{E}$; here and throughout c_0 signifies the constant from condition **A1**.

Proposition 3.3. *If $\varphi = \Phi$, then the output $\hat{A}$ of* `Approx`(G, φ) *is a matrix of rank k such that $\|\hat{A} - \mathcal{E}\| \leqslant c_0^2 k\sqrt{\lambda}$. Furthermore, if $\sigma^* > \ln^3 n$, then any rank k approximation A' of A satisfies $\|A' - \mathcal{E}\| \leqslant c_0^2 k\sqrt{\lambda}$.*

Combining Propositions 3.2 and 3.3, we conclude that the matrix $\hat{A}$ computed in step 2 satisfies $\|\hat{A} - \mathcal{E}\| \leqslant c_0^2 k\sqrt{\lambda}$. Therefore, our bound (1.8) on the Frobenius norm in terms of the spectral norm entails that $\|\hat{A} - \mathcal{E}\|_F^2 \leqslant 2k\|\hat{A} - \mathcal{E}\| \leqslant c_0^5 k^3 \lambda$, because we are assuming that c_0 is a sufficiently large constant. As, furthermore, $\rho^2 = c_0^8 k^3 \lambda / n_{\min}$ (see **A2**), we obtain the bound

$$\|\hat{A} - \mathcal{E}\|_F^2 \leqslant c_0^5 k^3 \lambda \leqslant c_0^{-3} \rho^2 n_{\min}. \tag{3.2}$$

Having computed the approximation $\hat{A}$ to $\mathcal{E}$, the next step is to obtain an initial partition of the vertices that should be 'close' to the desired partition $V_1, \ldots, V_k$. Computing this partition is the task of the subroutine `Initial` called in step 3. This subroutine basically partitions the vertices by their corresponding column vectors $\hat{A}_v$. More precisely, since the Frobenius norm $\|\hat{A} - \mathcal{E}\|_F^2 = \sum_{v \in V} \|\hat{A}_v - \mathcal{E}_v\|^2$ is 'small', for 'most' vertices v the distance $\|\hat{A}_v - \mathcal{E}_v\|$ is small ($< 0.01\rho$, say). Furthermore, for vertices $v \in V_i$ and $w \in V_j$ in different classes the vectors $\mathcal{E}_v$ and $\mathcal{E}_w$ are well separated by **A2** (namely, $\|\mathcal{E}_v - \mathcal{E}_w\| \geqslant \rho$), while for vertices $u, v \in V_i$ in the same class we have $\mathcal{E}_u = \mathcal{E}_v$. In effect, for 'most' vertices v, w in different classes $\|\hat{A}_v - \hat{A}_w\|$ should be large, while for most u, v in the same class $\|\hat{A}_u - \hat{A}_v\|$ should be small. `Initial` exploits this observation to compute k classes $S_1, \ldots, S_k$ along with k 'centre vectors' $\xi_1, \ldots, \xi_k \in \mathbb{R}^V$ such that for most $v \in S_i$ the norm $\|\hat{A}_v - \xi_i\|$ will be small. We will discuss `Initial` in Section 6 and establish the following.

Proposition 3.4. *There is a permutation τ of $\{1, \ldots, k\}$ such that the output of* `Initial` *has the following properties.*

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

The initial partition $S_1, \ldots, S_k$ is solely determined by the matrix $\hat{A}$, *i.e.*, by spectral properties of G. To home in on the desired output $V_1, \ldots, V_k$, step 4 of `Partition` finally calls a further subroutine `Improve`. This subroutine exploits combinatorial properties of G. More precisely, `Improve` performs an iterative 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 3.5. *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 $i = 1, \ldots, k$.*

Algorithm 4.1. `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 $a^*_{vw} \geqslant \frac{1}{2}$ 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 colouring to $\mathcal{B}$, and let $T_1, \ldots, T_R$ be the resulting colour 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 $\bar{G}$ consisting of all edges $\{v, w\} \notin E$ satisfying $\varphi_{vw} = 1$. If $R \leqslant k$ and the maximum degree of $G^*_1 \cup G^*_2$ is $\leqslant \ln^4 n$, then return φ. Otherwise output 'fail'.

Figure 2. Pseudocode for `Identify`.

A detailed description of `Improve` can be found in Section 7. Since all the procedures run in polynomial time, Theorem 1.2 is an immediate consequence of Propositions 3.2–3.5.

4. Identifying sparse/dense parts

4.1. The procedure `Identify`

The matrix Φ from (1.2) identifies which parts of the input graph G are sparse and which parts are dense: we have $\Phi_{vw} = 1$ if the edge density $\mathcal{E}_{vw} = p_{\psi(v)\psi(w)}$ between the classes of v and w is bigger than $\frac{1}{2}$, and $\Phi_{vw} = 0$ otherwise. Moreover, recall that $\sigma^* = \max_v \sum_w \mathcal{E}_{vw}(1 - \mathcal{E}_{vw})$ denotes the 'maximum variance' of any vertex degree of G (see (1.1)). The objective of `Identify` is to either compute Φ, or detect that $\sigma^* > \ln^3 n$ and output 'fail' (for if $\sigma^* > \ln^3 n$, then `Partition` does not need to know Φ).

Let us call two classes V_i, V_j *similar* if, for all indices l, we have $p_{il} \geqslant \frac{1}{2} \leftrightarrow p_{jl} \geqslant \frac{1}{2}$. In other words, V_i and V_j are similar if, for all $v \in V_i$ and all $w \in V_j$, the corresponding columns Φ_v, Φ_w coincide. Moreover, call two vertices v, w *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 $A(G)$. As we will see, the spectral assumption **A1** entails that A^* provides (at least) a 'rough' approximation of $\mathcal{E}$. Then, `Identify` constructs a matrix B by rounding the entries of A^* to 0/1. Since the desired output Φ is obtained by rounding the entries of $\mathcal{E}$, B should be 'close' to Φ.

Step 2 sets up a graph $\mathcal{B}$ with the same vertex set as G in which two vertices are adjacent if their corresponding column vectors in B are far apart in Euclidean distance.

https://doi.org/10.1017/S0963548309990514 Published online by Cambridge University Press

The intuition is that the columns of similar vertices should be close, unless the variance σ^* of the vertex degrees is very large. Then, `Identify` applies the greedy algorithm for graph colouring to $\mathcal{B}$. Recall that the greedy algorithm processes the vertices in an (arbitrary but) fixed order, assigning to each vertex the least possible colour in $\{1,\ldots,n\}$ that is not yet occupied by a neighbour of that vertex. This colouring is then used in step 3 to set up a matrix φ, which is based on the 'empirical' edge densities between the colour classes.

Finally, step 4 performs a 'consistency check'. First of all, it checks whether the number R of colours used by the greedy algorithm is bounded by k. If not, then φ does not coincide with Φ (because the matrix $\mathcal{E}$ that describes the desired partition has rank at most k). In addition, step 4 sets up two graphs G_1^* and G_2^*. The first graph G_1^* contains all edges $\{v,w\}$ such that $\varphi_{vw}=0$, and in G_2^* two vertices v,w are adjacent if and only if $\{v,w\}\notin E$ and $\varphi_{vw}=1$. This construction mimics the definition of the 'sparse' part G_1 and the 'dense' part G_2 of G as in (1.3). In fact, if $\varphi=\Phi$ (which is what we would like to establish), then $G_1=G_1^*$ and $G_2=G_2^*$. If $\sigma^*\leqslant\ln^3 n$, then condition **A3** implies that the maximum degrees of both G_1 and G_2 are bounded by $\sigma^*+\ln^2 n\leqslant\ln^4 n$, and step 4 checks if this bound holds for G_1^* and G_2^*. If so, `Identify` outputs φ, and otherwise the output is 'fail'.

The analysis of `Identify` proceeds in three steps. Firstly, in Section 4.2 we prove that either B is reasonably close to $\mathcal{E}$ in the Frobenius norm, or σ^* is rather large.

Lemma 4.2. *If $\sigma^*\leqslant\ln^{10} n$, then $\|B-\mathcal{E}\|_F^2\leqslant\log^{23} n$.*

The next step is to rule out that `Identify` outputs a matrix φ that differs from Φ. In Section 4.3 we shall prove the following.

Lemma 4.3. *If $\sigma^*\leqslant\ln^{10} n$, then* `Identify` *either fails or outputs $\varphi=\Phi$. Furthermore, if $\sigma^*\leqslant\ln^3 n$, then actually* `Identify` *outputs $\varphi=\Phi$.*

To complete the proof of Proposition 3.2 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 4.4. *If $\sigma^*>\ln^{10} n$, then* `Identify` *outputs 'fail'.*

The proof of Lemma 4.4 can be found in Section 4.4. Finally, Proposition 3.2 is an immediate consequence of Lemmas 4.2–4.4.

4.2. Proof of Lemma 4.2

We begin with the following simple observation.

Fact 4.5. *Suppose that $\sigma^*\leqslant\ln^{10} n$. Let $1\leqslant i,j\leqslant k$. If $p_{ij}>\frac{1}{2}$, then actually $p_{ij}>0.9$. Moreover, $p_{ij}\leqslant\frac{1}{2}$ in fact implies that $p_{ij}<0.1$.*

Proof. Recall from (1.1) that $\sigma^* = \max_v \sum_w \mathcal{E}_{vw}(1-\mathcal{E}_{vw}) = \max_i \sum_{j=1}^k \#V_j p_{ij}(1-p_{ij})$. Therefore, for all i,j we have

$$\#V_j p_{ij}(1-p_{ij}) \leqslant \sigma^* \leqslant \ln^{10} n. \tag{4.1}$$

Moreover, by assumption **A4** we have $\#V_j \geqslant n_{\min} \geqslant \ln^{11} n$ for all j. Thus, (4.1) shows that $p_{ij}(1-p_{ij}) \leqslant \ln^{-1} n \leqslant 0.01$ for all i,j, provided that n is sufficiently large. □

Lemma 4.6. *Suppose that* $\sigma^* \leqslant \ln^{10} n$. *Then* $\|\Phi - \mathcal{E}\|_F^2 \leqslant \ln^{22} n$.

Proof. Remember that the matrix $\mathcal{E}$ is constant with value p_{ij} on each rectangle $V_i \times V_j$. Moreover, Φ is equal to zero on each rectangle $V_i \times V_j$ such that $p_{ij} \leqslant \frac{1}{2}$, and Φ is equal to one on any rectangle $V_i \times V_j$ such that $p_{ij} > \frac{1}{2}$. Therefore,

$$\begin{aligned}\|\Phi - \mathcal{E}\|_F^2 &= \sum_{i,j=1}^k \|\Phi_{V_i\times V_j} - \mathcal{E}_{V_i\times V_j}\|_F^2 \\ &= \sum_{i,j:p_{ij}\leqslant\frac{1}{2}} p_{ij}^2 \#V_i\#V_j + \sum_{i,j:p_{ij}>\frac{1}{2}} (1-p_{ij})^2\#V_i\#V_j \\ &\leqslant 4\sum_{i,j=1}^k p_{ij}^2(1-p_{ij})^2\#V_i\#V_j. \end{aligned} \tag{4.2}$$

Since $\sigma^* = \max_i \sum_j \#V_j p_{ij}(1-p_{pij})$ (see (1.1)), we have $p_{ij}(1-p_{pij})\#V_j \leqslant \sigma^*$ for any i,j. Consequently, $\|\Phi - \mathcal{E}\|_F^2 \leqslant (2k\sigma^*)^2$. Further, since $\sigma^* \leqslant \ln^{10} n$ and as we are assuming that $n > n_0 = n_0(k)$ for some sufficiently large n_0, we have $2k < \ln n$. Therefore, $\|\Phi - \mathcal{E}\|_F^2 \leqslant (2k\sigma^*)^2 \leqslant \ln^{22} n$. □

Lemma 4.7. *Assume that* $\sigma^* \leqslant \ln^{10} n$. *We have* $\|\Phi - B\|_F^2 \leqslant \ln^{22} n$.

Proof. We recall that $\max_{v\in V} d_{G_1\cup G_2}(v) \leqslant 7\sigma^* + \ln^3 n \leqslant 2\ln^{10} n$ (see (3.1)). Therefore, the spectral condition **A1** yields

$$\|A^* - \mathcal{E}\|^2 \leqslant c_0^2 k^2 \left(\sigma^{*\,2} + \max_{v\in V} d_{G_1\cup G_2}(v)\right) \leqslant 3c_0^2k^2 \ln^{20} n.$$

Hence, as both $A^*, \mathcal{E}$ have rank k, the bound (1.8) on the Frobenius norm in terms of the spectral norm entails

$$\|A^* - \mathcal{E}\|_F^2 \leqslant 2k\|A^* - \mathcal{E}\|^2 \leqslant 6c_0^2k^3\ln^{20} n. \tag{4.3}$$

Since we assume that $n > n_0$ for some sufficiently large number $n_0 = n_0(k)$, (4.3) implies that

$$\|A^* - \mathcal{E}\|_F^2 \leqslant \ln^{21} n. \tag{4.4}$$

Furthermore, B_{vw} is obtained by rounding A^*_{vw} to 0/1, and Φ_{vw} is obtained by rounding $\mathcal{E}_{vw}$ to 0/1. Consequently, Fact 4.5 shows that $B_{vw} \neq \Phi_{vw}$ implies $|A^*_{vw} - \mathcal{E}_{vw}| \geqslant \frac{1}{3}$ for any $v, w \in V$. Therefore, $\|B - \Phi\|_F^2 \leqslant 9\|A^* - \mathcal{E}\|_F^2$, and thus the assertion follows from (4.4). □

https://doi.org/10.1017/S0963548309990514 Published online by Cambridge University Press

Proof of Lemma 4.2. Since we are assuming that $\sigma^* \leqslant \ln^{10} n$, the assertion follows directly from Lemmas 4.6 and 4.7 and the triangle inequality. □

4.3. Proof of Lemma 4.3

Throughout this section we assume that $\sigma^* \leqslant \ln^{10} n$. To prove Lemma 4.3, we need the following observation.

Lemma 4.8. *For all $v, w \in V$ we have $\|B_v - B_w\|^2 \leqslant \ln^{24} n$ if and only if v, w are similar.*

Proof. Suppose that $v \in V_i$ and $w \in V_j$ are not similar. Then there is an index l such that either $p_{il} > \frac{1}{2}$ and $p_{jl} \leqslant \frac{1}{2}$, or $p_{jl} > \frac{1}{2}$ and $p_{il} \leqslant \frac{1}{2}$. Swapping i and j if necessary, we may assume that $p_{il} > \frac{1}{2}$ and $p_{jl} \leqslant \frac{1}{2}$. Since $\|B - \mathcal{E}\|_F^2 \leqslant \ln^{23} n$ by Lemma 4.2, we conclude that

$$\begin{aligned}\frac{1}{4} \cdot \#\{u \in V_l : B_{vu} = 0\} &\leqslant p_{il}^2 \cdot \#\{u \in V_l : B_{vu} = 0\} \\ &\leqslant \|B - \mathcal{E}\|_F^2 \leqslant \ln^{23} n, \quad \text{and similarly} \qquad (4.5)\\ \frac{1}{4} \cdot \#\{u \in V_l : B_{wu} = 1\} &\leqslant (1 - p_{jl})^2 \cdot \#\{u \in V_l : B_{wu} = 1\} \\ &\leqslant \|B - \mathcal{E}\|_F^2 \leqslant \ln^{23} n. \qquad (4.6)\end{aligned}$$

As B is a 0/1 matrix, (4.5) and (4.6) imply that we can bound the distance of the columns B_v, B_w as follows:

$$\begin{aligned}\|B_v - B_w\|^2 &\geqslant \#V_l - \#\{u \in V_l : B_{vu} = B_{wu}\} \qquad (4.7)\\ &\geqslant \#V_l - \#\{u \in V_l : B_{vu} = 0\} - \#\{u \in V_l : B_{wu} = 1\} \geqslant \#V_l - 8\ln^{23} n.\end{aligned}$$

Since $\#V_l \geqslant n_{\min} \geqslant \ln^{30} n$ by condition **A4**, (4.7) implies that

$$\|B_v - B_w\|^2 \geqslant \ln^{30} n - 8\ln^{23} n \geqslant \ln^{24} n.$$

Conversely, assume that $v, w \in V$ are similar. Let $x = \#\{u \in V : |B_{uv} - \mathcal{E}_{uv}| \geqslant \frac{1}{3}\}$ and $y = \#\{u \in V : |B_{uw} - \mathcal{E}_{uw}| \geqslant \frac{1}{3}\}$. Since for all i, j we either have $p_{ij} < 0.1$ or $p_{ij} > 0.9$ by Fact 4.5, we obtain $\|B_v - B_w\|^2 \leqslant x + y \leqslant 9\|B - \mathcal{E}\|_F^2$. This implies the assertion, because $\|B - \mathcal{E}\|_F^2 \leqslant \ln^{23} n$ by Lemma 4.2. □

Lemma 4.9. *For all $1 \leqslant i, j \leqslant k$ the following holds:*

$$\text{if } p_{ij} > \frac{1}{2}, \quad \text{then } e(V_i, V_j) \geqslant \frac{2}{3}\#V_i\#V_j \quad (i \neq j), \quad \text{resp. } e(V_i, V_j) \geqslant \frac{2}{3}\binom{\#V_i}{2} \quad (i = j), \qquad (4.8)$$

$$\text{if } p_{ij} \leqslant \frac{1}{2}, \quad \text{then } e(V_i, V_j) \leqslant \frac{1}{3}\#V_i\#V_j \quad (i \neq j), \quad \text{resp. } e(V_i, V_j) \geqslant \frac{1}{3}\binom{\#V_i}{2} \quad (i = j). \qquad (4.9)$$

Proof. To prove (4.8), suppose that $p_{ij} > \frac{1}{2}$. Then we actually know that $p_{ij} > 0.9$ due to Fact 4.5. Suppose that $i \neq j$. Then the 'expected' number $\mu(V_i, V_j) = \#V_i\#V_j p_{ij}$ of

V_i–V_j-edges satisfies $\mu(V_i, V_j) > 0.9\#V_i\#V_j$. Therefore, assumption **A3** entails that

$$e(V_i, V_j) = \sum_{v \in V_i} e(v, V_j) \geqslant \sum_{v \in V_i} \left[\mu(v, V_i) - 0.1\left(\frac{1}{k}\sigma^* + \#V_j p_{ij}(1-p_{ij})\right) - \ln^2 n\right]$$
$$\geqslant \mu(V_i, V_j) - 0.1k^{-1}\sigma^*\#V_i - 0.1\#V_i\#V_j p_{ij} - \#V_i \ln^2 n.$$

Hence, we obtain

$$\begin{aligned} e(V_i, V_j) &\geqslant 0.9p_{ij}\#V_i\#V_j - \#V_i \cdot \ln^{10} n && (\text{because } \sigma^* \leqslant \ln^{10} n) \\ &\geqslant 0.9p_{ij}\#V_i\#V_j - \#V_i\#V_j/\ln n && (\text{as } \#V_j \geqslant n_{\min} \geqslant \ln^{30} n \text{ by } \mathbf{A4}) \\ &\geqslant 0.8p_{ij}\#V_i\#V_j > \frac{2}{3}\#V_i\#V_j && (\text{because } p_{ij} \geqslant 0.9). \end{aligned}$$

Thus, we have established (4.8) in the case $i \neq j$. If $i = j$, then $\mu(V_i) = \binom{\#V_i}{2}p_{ii} > 0.9\binom{\#V_i}{2}$, and **A3** implies that $e(V_i) \geqslant 0.8p_{ij}\binom{\#V_i}{2} > \frac{2}{3}\binom{\#V_i}{2}$, whence (4.8) follows. A similar argument yields (4.9). □

Proof of Lemma 4.3. Lemma 4.8 implies that two vertices $v, w \in V$ are adjacent in the graph $\mathcal{B}$ if and only if they are not similar. Hence, $\mathcal{B}$ is a complete R-partite graph, whose colour classes $T_1, \ldots, T_R$ are exactly the equivalence classes of the similarity relation. Therefore, (4.8) and (4.9) 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` outputs either 'fail' or $\varphi = \Phi$. Furthermore, if $\sigma^* \leqslant \ln^3 n$, then (3.1) entails that

$$d_{G_1 \cup G_2}(v) \leqslant 7\sigma^* + \ln^3 n \leqslant \ln^4 n \quad \text{for all} \quad v \in V,$$

whence `Identify` outputs $\varphi = \Phi$. □

4.4. Proof of Lemma 4.4

The basic idea of the proof is as follows. If `Identify` does not output 'fail', then the number R of colours used in step 2 is $\leqslant k$. Moreover, the graph $G_1^* \cup G_2^*$ that is obtained by including all edges $\{v, w\}$ of G such that $\varphi_{vw} = 0$ and all $\{v, w\} \notin E(G)$ such that $\varphi_{vw} = 1$ has maximum degree $\leqslant \ln^4 n$ (step 4 checks this condition). Hence, `Identify` has managed to find a partition of G into $R \leqslant k$ classes $T_1, \ldots, T_R$ that led to a matrix φ such that the graph $G_1^* \cup G_2^*$ is very sparse. However, we will show that under our assumption that the maximum 'variance' $\sigma^* = \max_j \sum_{i=1}^k \#V_i p_{ij}(1-p_{ij})$ of the vertex degrees exceeds $\ln^{10} n$, such a partition does not exist. This implies that the assumption that `Identify` does not answer 'fail' was false.

Let us now carry out this idea in detail. We assume in this section that

$$\sigma^* = \max_i \sum_j \#V_j p_{ij}(1-p_{ij}) > \ln^{10} n.$$

Let $1 \leqslant i, j \leqslant k$ be such that $\#V_j p_{ij}(1-p_{ij}) \geqslant k^{-1}\sigma^*$ (note that possibly $i = j$). We may assume without loss of generality that $p_{ij} \leqslant \frac{1}{2}$ (if $p_{ij} > \frac{1}{2}$, we just replace G by its complement and $\mathcal{E}$ by the all-ones matrix minus $\mathcal{E}$). Condition **A3** and the bound

https://doi.org/10.1017/S0963548309990514 Published online by Cambridge University Press

$\sigma^* > \ln^{10} n$ show that, for all $v \in V_i$,

$$|e(v, V_j) - \#V_j p_{ij}| \leqslant \frac{\sigma^*}{10k} + \#V_j p_{ij}(1 - p_{ij})/10 + \ln^2 n < \frac{\sigma^*}{2k}.$$

As $\#V_j p_{ij} \geqslant \sigma^*/k$, this yields

$$e(v, V_j) \geqslant \frac{\sigma^*}{2k} \quad \text{for all} \quad v \in V_i. \tag{4.10}$$

Furthermore, assuming that `Identify` does not fail, we know that its output is a matrix φ that is based on a partition $T_1, \dots, T_R$ with $R \leqslant k$. Recall that each entry φ_{vw} for $v \in T_i$ and $w \in T_j$ is obtained by rounding the 'edge density' $e(T_i, T_j)/\#T_i\#T_j$ if $i \neq j$, resp. $e(T_i)/\binom{\#T_i}{2}$ if $i = j$, to 1 if it is bigger than 0.66 and to 0 otherwise.

To each $v \in V_i$ we assign an index $\gamma(v) \in \{1, \dots, R\}$ such that

$$e(v, V_j \cap T_{\gamma(v)}) = \max_{1 \leqslant l \leqslant R} e(v, V_j \cap T_l).$$

Ties can be broken arbitrarily. Then $e(v, V_j \cap T_{\gamma(v)}) \geqslant R^{-1} e(v, V_j)$.

Fact 4.10. *$e(v, V_j \cap T_{\gamma(v)}) \geqslant \#T_{\gamma(v)} \cap V_j - \ln^4 n$, for all $v \in V_i$.*

Proof. We assign to each $v \in V_i$ the unique index $\beta(v) \in \{1, \dots, R\}$ such that $v \in T_{\beta(v)}$. Assume for contradiction that there exists $v \in V_i$ such that the entries of φ on the rectangle $T_{\gamma(v)} \times T_{\beta(v)}$ are 0. Then all v–$T_{\gamma(v)}$-edges are present in the graph G_1^*. Hence, the degree of v in G_1^* satisfies $d_{G_1^*}(v) \geqslant e(v, T_{\gamma(v)}) \geqslant R^{-1} e(v, V_j)$. In combination with our assumption that $n > n_0$ for some sufficiently large $n_0 = n_0(k)$, (4.10) shows that

$$d_{G_1^*}(v) \geqslant R^{-1} e(v, V_j) \geqslant \frac{\sigma^*}{2kR} > \frac{\ln^{10} n}{2kR} \geqslant \frac{\ln^{10} n}{2k^2} > \ln^4 n.$$

But then step 4 of `Identify` would have failed. This contradiction shows that φ attains the value 1 on the rectangle $T_{\gamma(v)} \times T_{\beta(v)}$ for all $v \in V_i$. Therefore, the degree $d_{G_2^*}(v)$ of $v \in V_i$ in G_2^* provides an upper bound on the number $\#T_{\gamma(v)} \cap V_j - e(v, V_j \cap T_{\gamma(v)})$ of 'missing' edges. Since we are assuming that `Identify` did not fail, we know that $d_{G_2^*}(v) \leqslant \ln^4 n$ (see step 4), and thus

$$\#T_{\gamma(v)} \cap V_j - e(v, V_j \cap T_{\gamma(v)}) \leqslant d_{G_2^*}(v) \leqslant \ln^4 n,$$

as claimed. □

Fact 4.11. *$\ln^9 n \leqslant \frac{\sigma^*}{2kR} \leqslant \#T_{\gamma(v)} \cap V_j \leqslant e(v, V_j \cap T_{\gamma(v)}) + \ln^4 n$, for all $v \in V_i$.*

Proof. To obtain the left-hand inequality, recall that we are assuming that $n > n_0$ for some sufficiently large $n_0 = n_0(k)$, which we can choose so that $\ln n > 2kR$. As $\sigma^* > \ln^{10} n$, we thus get $\sigma^*/(2kR) > \ln^9 n$. The right-hand inequality follows from Fact 4.10. Furthermore, (4.10) entails that $e(v, V_j) \geqslant \sigma^*/(2k)$ for any $v \in V_i$, and the choice of $\gamma(v)$ yields

$$\#V_j \cap T_{\gamma(v)} \geqslant e(v, V_j \cap T_{\gamma(v)}) \geqslant e(v, V_j)/R \geqslant \sigma^*/(2kR).$$

This yields the middle inequality. □

The map $\gamma : V_i \to \{1,\ldots,R\}$ assigns to each $v \in V_i$ an index $\gamma(v)$ such that $e(v, V_j \cap T_{\gamma(v)})$ is maximal, *i.e.*, $e(v, V_j \cap T_{\gamma(v)}) \geqslant e(v, V_j \cap T_l)$ for all l. Let $1 \leqslant \alpha \leqslant R$ be such that $\#\gamma^{-1}(\alpha)$ is maximal. Intuitively, T_α is the class of the partition $T_1,\ldots,T_R$ where most of the vertices in V_i have most of their neighbours. Then the number $\#\gamma^{-1}(\alpha)$ of vertices in V_i that have most of their neighbours in T_α is at least $R^{-1}\#V_i$. Since $R \leqslant k$ and $\#V_i \geqslant n_{\min}$, we have $\#\gamma^{-1}(\alpha) \geqslant n_{\min}/k$. Hence, it is possible to choose a set $S \subset \gamma^{-1}(\alpha) \subset V_i$ of cardinality

$$s = \lceil 10^{-4}k^{-3}n_{\min}\rceil. \tag{4.11}$$

(That is, we choose a set $S \subset \gamma^{-1}(\alpha) \subset V_i$ of this size arbitrarily.) Let $T = T_\alpha \cap V_j \setminus S$, and let $t = \#T$.

Fact 4.12. $e(S,T) \geqslant 0.9st$.

Proof. Double-counting edges in G_1 yields the bound

$$\#T_\alpha \cap V_j \geqslant \frac{e_{G_1}(\gamma^{-1}(\alpha), T_\alpha \cap V_j)}{\max_{w\in T_\alpha} d_{G_1}(w)}, \tag{4.12}$$

which is in terms of the degree $d_{G_1}(w)$ in G_1. Since $\sigma^* > \ln^{10} n$, the bound (3.1) on the maximum degree of $G_1 \cup G_2$ implies that $d_{G_1}(w) \leqslant 7\sigma^* + \ln^3 n \leqslant 8\sigma^*$ for all $w \in T_\alpha$. Furthermore, Fact 4.11 entails that

$$e_{G_1}(\gamma^{-1}(\alpha), T_\alpha \cap V_j) \geqslant \frac{1}{2}\sum_{v\in V_i:\gamma(v)=\alpha} e(v, V_j \cap T_\alpha) \geqslant \frac{\#\gamma^{-1}(\alpha)}{2}\cdot\frac{\sigma^*}{2kR}. \tag{4.13}$$

Since $R \leqslant k$ and $\#\gamma^{-1}(\alpha) \geqslant \#V_i/k$ by the choice of α, (4.12) and (4.13) yield

$$\#T_\alpha \cap V_j \geqslant \frac{e_{G_1}(\gamma^{-1}(\alpha), T_\alpha \cap V_j)}{8\sigma^*} \geqslant \frac{\#V_i}{32k^3} \overset{(4.11)}{\geqslant} 50s. \tag{4.14}$$

Recalling that $T = T_\alpha \cap V_j \setminus S$, we obtain

$$t = \#T \geqslant \#T_\alpha \cap V_j - s \overset{(4.14)}{\geqslant} \frac{1}{2}\#T_\alpha \cap V_j \overset{\text{Fact 4.11}}{\geqslant} \frac{\sigma^*}{4k^2}. \tag{4.15}$$

Since we are assuming that $\sigma^* > \ln^{10} n$ and $n > n_0$ for some $n_0 = n_0(k)$ that is sufficiently large that $\ln n > 4k^2$, (4.15) yields $t - \ln^4 n > 0.9t$. Hence, the right-hand inequality from Fact 4.11 entails $e(S,T) \geqslant s(t - \ln^4 n) \geqslant 0.9st$, as desired. □

To complete the proof of Lemma 4.4, we consider the matrix $M = \mathcal{E} - A(G)$. Its entries are $M_{vw} = \mathcal{E}_{vw} - 1$ if v, w are adjacent, and $M_{vw} = \mathcal{E}_{vw}$ if v, w are not adjacent. Condition **A1** ensures that $\|M\| \leqslant c_0k\sqrt{\lambda + \Delta}$, where $\lambda \leqslant \sigma^* n_{\min}/\ln n$ and Δ is the maximum degree of the graph $G_1 \cup G_2$. Furthermore, (3.1) provides a bound on Δ, namely $\Delta \leqslant 7\sigma^* + \ln^3 n \leqslant 8\sigma^*$, where the last inequality is due to our assumption $\sigma^* > \ln^{10} n$. Hence, assuming that $n > n_0(k)$ is sufficiently large, we obtain

$$\|M\| \leqslant c_0k\sqrt{\sigma^* n_{\min}/\ln n + 8\sigma^*} \leqslant 2c_0k\sqrt{\sigma^* n_{\min}/\ln n} < 10^{-4}k^{-3}\sqrt{\sigma^* n_{\min}}. \tag{4.16}$$

Furthermore, since $S \subset V_i$ and $T \subset V_j$ and $S \cap T = \emptyset$, the definition of M shows that

$$\langle M\mathbf{1}_S, \mathbf{1}_T\rangle = \sum_{(v,w)\in S\times T} M_{vw} = -e(S,T) + \sum_{(v,w)\in S\times T} \mathcal{E}_{vw} = \#S\#T p_{ij} - e(S,T).$$

Since $p_{ij} \leqslant \frac{1}{2}$, Fact 4.12 entails that

$$\begin{aligned}\frac{2}{5}st &\leqslant e(S,T) - \#S\#T p_{ij} = -\langle M\mathbf{1}_S, \mathbf{1}_T\rangle \\ &\leqslant \|M\| \cdot \|\mathbf{1}_S\| \cdot \|\mathbf{1}_T\| = \|M\|\sqrt{st}. \end{aligned} \tag{4.17}$$

Combining (4.11), (4.15), and (4.17), we obtain $\|M\| \geqslant \frac{2}{5}\sqrt{st} \geqslant 10^{-4}k^{-3}\sqrt{n_{\min}\sigma^*}$, in contradiction to (4.16). This shows that our assumption that `Identify` does not answer 'fail' was false.

5. Approximating the expected densities

5.1. The procedure `Approx`

The aim of `Approx` is to compute a low-rank matrix $\hat{A}$ such that $\|\hat{A} - \mathcal{E}\| \leqslant c_0^2 k\sqrt{\lambda}$ (see Proposition 3.3). Here λ denotes the spectral parameter from condition **A1** (and c_0 is the constant from **A1**). Remember that $\sigma^* = \max_j \sum_{i=1}^{k} \#V_i p_{ij}(1-p_{ij})$ is the maximum 'variance' of the vertex degrees, and that $\sigma^* \leqslant \lambda \leqslant \sigma^* \min\{\sigma^*, n_{\min}/\ln n\}$. Thus, the larger σ^* the less accurate an approximation $\hat{A}$ to $\mathcal{E}$ we need to provide. Furthermore, recall the decomposition of the graph G into its 'sparse' part G_1 and the complement G_2 of its 'dense' part (see (1.3)): G_1 contains all edges $\{v,w\} \in E(G)$ such that $\mathcal{E}_{vw} \leqslant \frac{1}{2}$, and G_2 contains all edges $\{v,w\} \notin E(G)$ such that $\mathcal{E}_{vw} > \frac{1}{2}$.

In order to compute the approximation $\hat{A}$ of $\mathcal{E}$, `Approx` analyses the spectrum of the adjacency matrix $A = A(G)$. As we will see, if the maximum variance σ^* is very large (more precisely, $\sigma^* > \ln^3 n$), then it is easy to obtain the desired approximation $\hat{A}$: just computing a rank k approximation of A is sufficient. In contrast, if σ^* is 'small' (say, $\sigma^* < \ln\ln n$), then matters are more involved. In this case the graph $G_1 \cup G_2$ is sparse. In effect, just as in the graph colouring example in Section 2, fluctuations of the vertex degrees affect the spectrum of the adjacency matrix A. Hence, as in Section 2, we could in principle just disregard vertices that have an 'atypically high' degree in $G_1 \cup G_2$. Indeed, condition **A1** suggests that we should ignore all vertices whose degree in $G_1 \cup G_2$ is bigger than 10λ, say. The problem with this is that `Approx` does not know λ. In fact, it is not obvious how to estimate either λ or σ^* given just the graph G and the number k of classes.

Therefore, `Approx` pursues an *adaptive* strategy (see Figure 3). Suppose that $\sigma^* < \ln^3 n$. The input of `Approx` consists of G, k, and the matrix $\varphi = \Phi$ that indicates which parts of G are sparse/dense (see Proposition 3.2). Thus, the two graphs G_1^*, G_2^* set up in step 1 coincide with G_1, G_2. Proceeding in $\log_2 n$ rounds $t = 1, \ldots, \log_2 n$, step 2 of `Approx` computes sets R_t of vertices of 'high' degree $d_{G_1^* \cup G_2^*}(v) \geqslant \Delta_t = 2^{-t}n$ in the graph $G_1^* \cup G_2^*$, and certain matrices A_t. The matrix A_t is obtained from the adjacency matrix $A(G)$ 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 the graph $G_1 \cup G_2$. As discussed above, ideally we would like to return a matrix A_t such

Algorithm 5.1. 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 $\varphi_{vw} = 1$.
   Set $R_0 = \emptyset$, and let $A_0 = (a_{0,vw})_{v,w \in V} = A(G)$.
2. For $t = 1, \ldots, \log_2 n$ do
3. Let $\Delta_t = 2^{-t} n$ 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 $0 \leqslant s < t$ such that $\|A_s - A_t\| > 4c_0 k \sqrt{\Delta_s}$, 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 $A_{\hat{t}}$.

Figure 3. The procedure Approx.

that $\Delta_t \approx \lambda$, say. Since we cannot actually implement this stopping criterion (because λ is unknown), we check instead if the sequence of matrices A_t 'converges' by computing the spectral norm $\|A_s - A_t\|$ for all $s < t$. This condition is somewhat reminiscent of Cauchy's criterion for the convergence of sequences. If $\|A_s - A_t\|$ is 'too large' for some $s < t$, then Δ_t has become too small, and therefore the algorithm returns the previous approximation A_{t-1}.

To analyse Approx, we remember the 'core' subgraph H that satisfies conditions **H1**–**H4** (see Theorem 1.2). Intuitively H consists of the 'well-behaved' vertices of G. Below we will need condition **H1**, which states that $\#V \setminus H \leqslant \lambda^{-4} n_{\min}$, and condition **H3**, according to which all vertices v of H have degree at most $d_{G_1 \cup G_2}(v) \leqslant 10\sigma^*$ in $G_1 \cup G_2$.

Lemma 5.2. *Suppose that* $\Delta_t \geqslant 50\lambda$ *and that* $\varphi = \Phi$. *Then* $\|A_t - \mathcal{E}\| \leqslant 2c_0 k \sqrt{\Delta_t}$.

Proof. The set R_t contains all vertices that have degree greater than Δ_t in $G_1^* \cup G_2^*$. Since we are assuming that $\varphi = \Phi$, we have $G_1^* = G_1$ and $G_2^* = G_2$. Therefore, condition **A1** shows that the minor of A induced on $V \setminus R_t$ satisfies

$$\|\mathcal{E}_{V \setminus R_t} - A_{V \setminus R_t}\| \leqslant c_0 k \sqrt{\Delta_t + \lambda} \leqslant \frac{3}{2} c_0 k \sqrt{\Delta_t}. \tag{5.1}$$

Let $\mathcal{F} = \mathcal{E}_{R_t} + \mathcal{E}_{R_t \times V \setminus R_t} + \mathcal{E}_{V \setminus R_t \times R_t}$ and $M = \varphi_{R_t} + \varphi_{R_t \times V \setminus R_t} + \varphi_{V \setminus R_t \times R_t}$. Since

$$V^2 \setminus (V \setminus R_t)^2 = (R_t \times R_t) \cup (R_t \times V \setminus R_t) \cup (V \setminus R_t \times R_t),$$

the entries $\mathcal{F}_{vw}$ are 0 for $(v, w) \in (V \setminus R_t)^2$, and $\mathcal{F}_{vw} = \mathcal{E}_{vw}$ for $(v, w) \notin (V \setminus R_t)^2$. Similarly, $M_{vw} = 0$ for $(v, w) \in (V \setminus R_t)^2$, and M_{vw} coincides with φ for $(v, w) \notin (V \setminus R_t)^2$. Since on $V^2 \setminus (V \setminus R_t)^2$ the matrix A_t coincides with φ, we have

$$\mathcal{F} - M = \mathcal{E} - A_t - (\mathcal{E}_{V \setminus R_t} - A_{V \setminus R_t}). \tag{5.2}$$

https://doi.org/10.1017/S0963548309990514 Published online by Cambridge University Press

Thus, to complete the proof we need to bound $\|\mathcal{F}-M\|$. Bounding the spectral norm by the Frobenius norm (see (1.8)), we obtain

$$\begin{aligned}\|\mathcal{F}-M\|^2 &\leqslant \|\mathcal{F}-M\|_F^2 \\ &= \|\mathcal{E}_{R_t}-\varphi_{R_t}\|_F^2+\|\mathcal{E}_{R_t\times V\setminus R_t}-\varphi_{R_t\times V\setminus R_t}\|_F^2 \\ &\quad +\|\mathcal{E}_{V\setminus R_t\times R_t}-\varphi_{V\setminus R_t\times R_t}\|_F^2 \\ &\leqslant 2\|\mathcal{E}_{R_t\times V}-\varphi_{R_t\times V}\|_F^2 && \text{(because } \mathcal{E}-\varphi \text{ is symmetric)} \\ &= 2\|\mathcal{E}_{R_t\times V}-\Phi_{R_t\times V}\|_F^2 && \text{(since we assume that } \varphi=\Phi) \\ &= 2\sum_{v\in R_t}\sum_{w\in V}(p_{\psi(v)\psi(w)}-\Phi_{vw})^2 && \text{(as } \mathcal{E}_{vw}=p_{\psi(v)\psi(w)}). \end{aligned} \tag{5.3}$$

The matrix Φ is defined so that $\Phi_{vw}=1$ if $p_{\psi(v)\psi(w)}>\frac{1}{2}$ and $\Phi_{vw}=0$ if $p_{\psi(v)\psi(w)}\leqslant\frac{1}{2}$. Therefore, $(p_{\psi(v)\psi(w)}-\Phi_{vw})^2\leqslant 4(p_{\psi(v)\psi(w)}(1-p_{\psi(v)\psi(w)}))^2$. Hence, (5.3) yields

$$\begin{aligned}\|\mathcal{F}-M\|^2 &\leqslant 8\sum_{v\in R_t}\sum_{w\in V}\left[p_{\psi(v)\psi(w)}(1-p_{\psi(v)\psi(w)})\right]^2 \\ &= 8\sum_{a=1}^{k}\sum_{b=1}^{k}\#V_a\cap R_t\cdot\#V_b\cdot\left[p_{ab}(1-p_{ab})\right]^2 \quad \text{(because } V_a=\psi^{-1}(a)) \end{aligned} \tag{5.4}$$

Since $\sigma^*=\max_b\sum_{a=1}^k\#V_ap_{ab}(1-p_{ab})$, (5.4) entails

$$\begin{aligned}\|\mathcal{F}-M\|^2 &\leqslant 8\sigma^*\sum_{a=1}^{k}\#V_a\cap R_t\sum_{b=1}^{k}p_{ab}(1-p_{ab}) \\ &\leqslant \frac{8\sigma^{*2}}{n_{\min}}\sum_{a=1}^{k}\#V_a\cap R_t && \text{(because } \sigma^*\geqslant\textstyle\sum_b n_{\min}p_{ab}(1-p_{ab})) \\ &= \frac{8\sigma^{*2}\#R_t}{n_{\min}}\leqslant\frac{8\sigma^{*2}\#V\setminus H}{n_{\min}} && \text{(as } R_t\subset V\setminus H \text{ by } \mathbf{H3}). \end{aligned} \tag{5.5}$$

As $\#V\setminus H\leqslant n_{\min}/\lambda^4$ by condition **H1** and $\lambda\geqslant\sigma^*\geqslant c_0$ for some large constant c_0 by assumption **A1**, (5.5) yields $\|\mathcal{F}-M\|^2\leqslant 8\sigma^{*2}\#V\setminus H/n_{\min}\leqslant 1$. Combining this estimate with (5.1) and (5.2), we obtain $\|A_t-\mathcal{E}\|\leqslant\|\mathcal{E}_{V\setminus R_t}-A_{V\setminus R_t}\|+\|\mathcal{F}-M\|\leqslant c_0k\sqrt{\Delta_t}+1\leqslant 2c_0k\sqrt{\Delta_t}$. □

Proof of Proposition 3.3. If `Identify` fails, then $\sigma^*>\ln^3 n$ by Proposition 3.2, and step 2 of `Partition` computes a rank k approximation $\hat{A}$ of A. Since $\sigma^*>\ln^3 n$, the bound (3.1) on the maximum degree of $G_1\cup G_2$ entails that $d_{G_1\cup G_2}(v)\leqslant 8\sigma^*$ for all $v\in V$. Therefore,

$$\|\mathcal{E}-\hat{A}\|\leqslant\|\mathcal{E}-A\|+\|\hat{A}-A\|\overset{(1.9)}{\leqslant}2\|\mathcal{E}-A\|\overset{\mathbf{A1}}{\leqslant}c_0^2k\sqrt{\lambda},$$

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 $\Delta_s\geqslant\Delta_t\geqslant 50\lambda$. Then by Lemma 5.2 and the triangle inequality we have $\|A_s-A_t\|\leqslant\|A_s-\mathcal{E}\|+\|A_t-\mathcal{E}\|\leqslant 4c_0k\sqrt{\Delta_s}$, and thus step 3

https://doi.org/10.1017/S0963548309990514 Published online by Cambridge University Press

of `Approx` will not abort the loop. Consequently, the index $\hat{t}$ chosen in step 4 of `Approx` satisfies $\Delta_{\hat{t}} \leqslant 100\lambda$. Let t^* be maximal such that $\Delta_{t^*} \geqslant 50\lambda$. Then $\|A_{\hat{t}} - A_{t^*}\| \leqslant 4c_0 k\sqrt{\Delta_{t^*}}$, because the exit condition in step 3 of `Approx` was not satisfied for $t = \hat{t}$ and $s = t^*$. Therefore, invoking Lemma 5.2 and the triangle inequality once more, we get

$$\|A_{\hat{t}} - \mathcal{E}\| \leqslant \|A_{\hat{t}} - A_{t^*}\| + \|A_{t^*} - \mathcal{E}\| \leqslant 6c_0 k\sqrt{\Delta_{t^*}} \leqslant 60c_0 k\sqrt{\lambda}. \tag{5.6}$$

Finally, if $\hat{A}$ is a rank k approximation of $A_{\hat{t}}$, then by the key property (1.9) of the rank k approximation we have $\|\hat{A} - A_{\hat{t}}\| \leqslant \|\mathcal{E} - A_{\hat{t}}\|$, because $\mathcal{E}$ has rank at most k. Therefore, (5.6) implies $\|\hat{A} - \mathcal{E}\| \leqslant \|\hat{A} - A_{\hat{t}}\| + \|A_{\hat{t}} - \mathcal{E}\| \leqslant 2\|A_{\hat{t}} - \mathcal{E}\| \leqslant 120c_0 k\sqrt{\lambda} \leqslant c_0^2 k\sqrt{\lambda}$, as claimed. □

6. Computing an initial partition

6.1. The procedure `Initial`

Once `Partition` has obtained the matrix $\hat{A}$, the algorithm calls the subroutine `Initial`. We know from Proposition 3.3 that $\hat{A}$ is a matrix of rank $\leqslant k$ that approximates the 'expected' adjacency matrix $\mathcal{E}$ well. More precisely, in the notation of condition **A1** we have

$$\|\hat{A} - \mathcal{E}\| \leqslant c_0^2 k\sqrt{\lambda}. \tag{6.1}$$

Given the matrix $\hat{A}$ as input, `Initial` now tries to find a partition $(S_1, \ldots, S_k)$ of the vertices of G that is 'close' to the planted partition. In fact, our aim is to show that the output $(S_1, \ldots, S_k)$ satisfies $\sum_{i=1}^{k} \#S_i \triangle V_{\tau(i)} \leqslant 0.001 n_{\min}$, where $\triangle$ denotes the symmetric difference, τ is a permutation of the indices, and $n_{\min}$ is the size of the smallest class of the desired partition.

To construct such a partition, `Initial` basically classifies the vertices $v \in V$ by their corresponding column $\hat{A}_v$ of $\hat{A}$. Two vertices v, w are deemed 'similar' if their columns $\hat{A}_v$, $\hat{A}_w$ are close in ℓ_2. This approach is based on two observations:

(a) If $v \in V_i$ and $w \in V_j$ with $i \neq j$, then $\mathcal{E}_v$ and $\mathcal{E}_w$ are far apart in ℓ_2. More precisely, by condition **A2**

$$\|\mathcal{E}_v - \mathcal{E}_w\|^2 \geqslant \rho^2, \quad \text{where} \quad \rho = c_0^4\sqrt{k^3\lambda/n_{\min}}. \tag{6.2}$$

Furthermore, if $u, v \in V_i$ are in the same class, then $\mathcal{E}_v = \mathcal{E}_u$ by definition.

(b) For 'most' vertices v the vector $\hat{A}_v$ is close to $\mathcal{E}_v$ in ℓ_2.

To establish (b), we combine (6.1) with the bound (1.8) on the Frobenius norm in terms of the operator norm: since both $\hat{A}$ and $\mathcal{E}$ have rank $\leqslant k$, we get

$$\|\hat{A} - \mathcal{E}\|_F^2 \leqslant c_0^4 k^3 \lambda. \tag{6.3}$$

Now, let $z = \#\{v \in V : \|\hat{A}_v - \mathcal{E}_v\|^2 > 0.001\rho^2\}$ be the number of vertices for which $\hat{A}_v$ and $\mathcal{E}_v$ are not close. Then

$$\|\hat{A} - \mathcal{E}\|_F^2 = \sum_{v \in V} \|\hat{A}_v - \mathcal{E}_v\|^2 \geqslant 0.001\rho^2 \cdot z. \tag{6.4}$$

https://doi.org/10.1017/S0963548309990514 Published online by Cambridge University Press

Algorithm 6.1. $\texttt{Initial}(\hat{A}, k)$
Input: A matrix $\hat{A}$ and the parameter k.
Output: A partition $S_1, \ldots, S_k$ of V and vectors $\xi_1, \ldots, \xi_k \in \mathbb{R}^V$.

1. For $j = 1, \ldots, 2\log n$ do
2. Let $\rho_j = n2^{-j}$ and compute $Q^{(j)}(v) = \{w \in V : \|\hat{A}_w - \hat{A}_v\|^2 \leqslant 0.01\rho_j^2\}$ 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 $\xi_i^{(j)} = \sum_{w \in Q_i^{(j)}} \hat{A}_w / \#Q_i^{(j)}$.
4. Partition the entire set V as follows.
   - First, let $S_i^{(j)} = Q_i^{(j)}$ for all $1 \leqslant i \leqslant k$.
   - 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 $\xi_1^{(J)}, \ldots, \xi_k^{(J)}$.

Figure 4. The procedure `Initial`.

If the constant c_0 is chosen to be sufficiently large, then $0.001\rho^2 > c_0^5 k^3 \lambda / n_{\min}$. Hence, (6.1) and (6.4) imply that $z \leqslant n_{\min}/c_0 < n_{\min}/1000$ is small relative to the size $n_{\min}$ of the smallest class of the desired partition.

Thus, for all but $0.001n_{\min}$ vertices we have $\|\hat{A}_v - \mathcal{E}_v\| \leqslant \sqrt{0.001}\rho$. Furthermore, it makes sense to classify such vertices v by their corresponding column $\hat{A}_v$. For assume that $\|\hat{A}_v - \mathcal{E}_v\| \leqslant \sqrt{0.001}\rho$ and $\|\hat{A}_w - \mathcal{E}_w\| \leqslant \sqrt{0.001}\rho$. If $v, w \in V_i$ for some i, then $\mathcal{E}_v = \mathcal{E}_w$, and thus by the triangle inequality

$$\|\hat{A}_v - \hat{A}_w\| \leqslant \|\hat{A}_v - \mathcal{E}_v\| + \|\hat{A}_w - \mathcal{E}_w\| \leqslant 2\sqrt{0.001}\rho. \tag{6.5}$$

In contrast, if $v \in V_i$ and $w \in V_j$, $i \neq j$, then by (6.2)

$$\|\hat{A}_v - \hat{A}_w\| \geqslant \|\hat{\mathcal{E}}_v - \mathcal{E}_v\| - \|\hat{A}_v - \mathcal{E}_v\| - \|\hat{A}_w - \mathcal{E}_w\| \geqslant (1 - 2\sqrt{0.001})\rho. \tag{6.6}$$

If `Initial` knew the parameter ρ, then it could utilize (6.5) and (6.6) to compute a good approximation to the desired partition $V_1, \ldots, V_k$ as follows. For each vertex $v \in V$ we could determine the set $Q(v) = \{w \in V : \|\hat{A}_v - \hat{A}_w\|^2 \leqslant 0.01\rho^2\}$ of vertices w whose vector $\hat{A}_w$ is close to $\hat{A}_v$. If $v \in V_i$ and $\|\hat{A}_v - \mathcal{E}_v\| \leqslant \sqrt{0.001}\rho$, then $Q(v)$ will contain all vertices $w \in V_i$ such that $\|\hat{A}_w - \mathcal{E}_w\| \leqslant \sqrt{0.001}\rho$. As we have seen, there are at least $\#V_i - z \geqslant 0.999\#V_i$ such vertices $w \in V_i$. Hence, $Q(v)$ contains almost all of V_i. Conversely, any vertex $w \in V \setminus V_i$ such that $\|\hat{A}_w - \mathcal{E}_w\| \leqslant \sqrt{0.001}\rho$ does not lie in $Q(v)$ (see (6.6)). Hence, $\#Q(v) \triangle V_i \leqslant z \leqslant 0.001n_{\min}$. Thus, to obtain a partition of V we could first choose a vertex v_1 such that $\#Q(v_1) = \max_{v \in V} \#Q(v)$ is maximum. Then, choose a vertex $v_2 \in V \setminus Q(v_1)$ such that $\#Q(v_2) \setminus Q(v_1) = \max_{v \in V \setminus V_1} \#Q(v) \setminus Q(v_1)$ is maximum, *etc.* This is essentially what steps 2–3 of `Initial` do, and a similar procedure is at the core of McSherry's algorithm [28].

However, we do *not* actually assume that ρ is known to `Initial`. Therefore, the algorithm applies the above clustering procedure for various 'candidate' values $\rho_j = n2^{-j}$. For each j this yields a collection $Q_1^{(j)}, \ldots, Q_k^{(j)}$ of pairwise disjoint subsets of V. For each of them we compute the 'barycentre' $\xi_i^{(j)}$, which is just the arithmetic mean of the vectors $\hat{A}_w$ with $w \in Q_i^{(j)}$. Hence, $\xi_i^{(j)}$ should approximate $\mathcal{E}^{V_i}$ well if $Q_i^{(j)}$ is a good approximation of V_i. In effect, if $Q_1^{(j)}, \ldots, Q_k^{(j)}$ is 'close' to $V_1, \ldots, V_k$, then the error term

$$r_j = \sum_{i=1}^{k} \sum_{v \in S_i^{(j)}} \|\hat{A}_v - \xi_i^{(j)}\|^2$$

should be about as small as $\|\hat{A} - \mathcal{E}\|_F^2$. Therefore, the output of `Initial` is just the partition $S_1^{(j)}, \ldots, S_k^{(j)}$ with minimal r_j. The following lemma shows that, if the above error term is actually small, then the resulting partition is close to $V_1, \ldots, V_k$.

Lemma 6.2. *Let $S_1, \ldots, S_k$ be a partition and let $\xi_1, \ldots, \xi_k$ be a sequence of vectors such that*

$$\sum_{i=1}^{k} \sum_{v \in S_i} \|\xi_i - \hat{A}_v\|^2 \leqslant c_0^6 k^3 \lambda.$$

Then there is a permutation $\gamma : \{1, \ldots, k\} \to \{1, \ldots, k\}$ such that the following holds.

(1) $\|\xi_i - \mathcal{E}^{V_{\gamma(i)}}\|^2 \leqslant 0.001\rho^2$ *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* $1 \leqslant j \leqslant k$.

Proof. Let $S_{ab} = S_a \cap V_b$ be the set of all vertices in class V_b that end up in S_a ($1 \leqslant a, b \leqslant k$). For each $1 \leqslant a \leqslant k$ choose an index $1 \leqslant \gamma(a) \leqslant k$ such that $\|\mathcal{E}^{V_{\gamma(a)}} - \xi_a\|$ is minimum (ties can be broken arbitrarily). Then for all $b \neq \gamma(a)$ we have

$$\begin{aligned} \rho &\leqslant \|\mathcal{E}^{V_{\gamma(a)}} - \mathcal{E}^{V_b}\| && \text{(by condition } \mathbf{A2}) \\ &\leqslant \|\mathcal{E}^{V_{\gamma(a)}} - \xi_a\| + \|\mathcal{E}^{V_b} - \xi_a\| && \text{(by the triangle inequality)} \\ &\leqslant 2\|\mathcal{E}^{V_b} - \xi_a\|, && \text{(by the choice of } \gamma(a)). \end{aligned} \tag{6.7}$$

Hence, $\|\mathcal{E}^{V_b} - \xi_a\| \geqslant \rho/2$. Therefore, the assumption $\sum_{i=1}^{k} \sum_{v \in S_i} \|\xi_i - \hat{A}_v\|^2 \leqslant c_0^6 k^3 \lambda$ implies

$$\begin{aligned} \frac{\rho^2}{4} \sum_{a=1}^{k} \sum_{1 \leqslant b \leqslant k : b \neq \gamma(a)} \#S_{ab} &\leqslant \sum_{a,b=1}^{k} \#S_{ab} \|\mathcal{E}^{V_b} - \xi_a\|^2 \\ &\leqslant 2 \sum_{a,b=1}^{k} \sum_{v \in S_{ab}} \|\mathcal{E}_v - \hat{A}_v\|^2 + \|\hat{A}_v - \xi_a\|^2 \qquad \text{(by the triangle inequality)} \end{aligned}$$

https://doi.org/10.1017/S0963548309990514 Published online by Cambridge University Press

$$= 2\sum_{v\in V}\|\mathcal{E}_v - \hat{A}_v\|^2 + 2\sum_{a,b=1}^{k}\sum_{v\in S_{ab}}\|\hat{A}_v - \xi_a\|^2$$

$$= 2\|\hat{A} - \mathcal{E}\|_F^2 + 2\sum_{a=1}^{k}\sum_{v\in S_a}\|\hat{A}_v - \xi_a\|^2 \leqslant 2\|\hat{A} - \mathcal{E}\|_F^2 + 2c_0^6k^3\lambda. \tag{6.8}$$

Furthermore, (6.1) shows in combination with the bound (1.8) on the Frobenius norm in terms of the spectral norm that $\|\hat{A} - \mathcal{E}\|_F^2 \leqslant c_0^4k^2\lambda$. Hence, (6.8) yields

$$\frac{\rho^2}{4}\sum_{a=1}^{k}\sum_{1\leqslant b\leqslant k:b\neq\gamma(a)}\#S_{ab} \leqslant \sum_{a,b=1}^{k}\#S_{ab}\|\mathcal{E}^{V_b} - \xi_a\|^2 \leqslant 3c_0^6k^3\lambda \leqslant c_0^7k^3\lambda. \tag{6.9}$$

Remembering that $\rho^2 = c_0^8k^3\lambda/n_{\min}$ (see **A2**) for some large constant c_0, we see that (6.9) yields

$$\sum_{a=1}^{k}\#S_a\triangle V_{\gamma(a)} = \sum_{a=1}^{k}\sum_{1\leqslant b\leqslant k:b\neq\gamma(a)}2\#S_{ab} \leqslant \frac{8c_0^7k^3\lambda}{\rho^2} \leqslant 0.001n_{\min}. \tag{6.10}$$

This establishes assertion (2).

In addition, (6.10) shows that γ is a bijection. For assume for contradiction that there are $1 \leqslant a < b \leqslant k$ such that $\gamma(a) = \gamma(b)$. Then (6.10) yields $\#S_a\triangle V_{\gamma(a)} \leqslant 0.001n_{\min}$ and $\#S_b\triangle V_{\gamma(a)} \leqslant 0.001n_{\min}$. But this is impossible, since $S_a \cap S_b = \emptyset$ and $V_{\gamma(a)} \geqslant n_{\min}$.

Furthermore, (6.10) implies that $\#S_a \cap V_{\gamma(a)} \geqslant \#V_{\gamma(a)} - 0.001n_{\min} \geqslant n_{\min}/2$. Thus, by (6.8) for any $1 \leqslant a \leqslant k$ we have

$$\frac{n_{\min}}{2}\|\mathcal{E}^{V_{\gamma(a)}} - \xi_a\|^2 \leqslant \#S_a \cap V_{\gamma(a)}\|\mathcal{E}^{V_{\gamma(a)}} - \xi_a\|^2 \leqslant \sum_{\alpha,\beta=1}^{k}\#S_{\alpha\beta}\|\mathcal{E}^{V_\beta} - \xi_\alpha\|^2 \leqslant c_0^7k^3\lambda.$$

Recalling that $\rho^2 = c_0^8k^3\lambda/n_{\min}$ for a large constant c_0, we thus get

$$\|\mathcal{E}_{\gamma(a)} - \xi_a\|^2 \leqslant \frac{2c_0^7k^3\lambda}{n_{\min}} \leqslant 0.001\rho^2 \quad \text{for all} \quad 1 \leqslant a \leqslant k, \tag{6.11}$$

thereby proving assertion (1).

Finally, to prove assertion (3) we apply the triangle inequality to obtain

$$\sum_{a,b=1}^{k}\#S_{ab}\|\mathcal{E}^{V_{\gamma(a)}} - \mathcal{E}^{V_{\gamma(b)}}\|^2 \leqslant 2\sum_{a,b=1}^{k}\#S_{ab}\big(\|\mathcal{E}^{V_{\gamma(a)}} - \xi_a\|^2 + \|\mathcal{E}^{V_{\gamma(b)}} - \xi_a\|^2\big).$$

Since $\|\mathcal{E}^{V_{\gamma(a)}} - \xi_a\| \leqslant \|\mathcal{E}^{V_{\gamma(b)}} - \xi_a\|$ by the choice of $\gamma(a)$, we obtain

$$\sum_{a,b=1}^{k}\#S_{ab}\|\mathcal{E}^{V_{\gamma(a)}} - \mathcal{E}^{V_{\gamma(b)}}\|^2 \leqslant 4\sum_{a,b=1}^{k}\#S_{ab}\|\mathcal{E}^{V_{\gamma(b)}} - \xi_a\|^2 \overset{(6.9)}{\leqslant} 4c_0^7k^3\lambda.$$

This implies assertion (3), because $\rho^2 = c_0^8k^3\lambda/n_{\min}$ for some large c_0 (see **A2**). □

Finally, in Section 6.2 we shall derive the following bound on the minimum error term r_j.

Lemma 6.3. *Suppose that the index j from step 1 of* `Initial` *is such that $\frac{1}{2}\rho \leqslant \rho_j \leqslant \rho$. Then the term $r_j = \sum_{i=1}^{k} \|\hat{A}_v - \xi_i^{(j)}\|^2$ computed in step 4 satisfies $r_j \leqslant c_0^6 k^3 \lambda$.*

Proof of Proposition 3.4. Lemma 6.3 shows that there is an index j such that $r_j \leqslant c_0^6 k^3 \lambda$. Therefore, the minimum error term r^* computed in step 5 satisfies $r^* \leqslant c_0^6 k^3 \lambda$. Let J be the index chosen in step 5, *i.e.*, $r_J = r^*$. Then Lemma 6.2 shows that the partition $S_1^{(J)}, \ldots, S_k^{(J)}$ satisfies the conditions stated in Proposition 3.4. □

6.2. Proof of Lemma 6.3

Suppose that $\frac{1}{2}\rho \leqslant \rho_j \leqslant \rho$. To ease up the notation, we omit the superscript j; thus, we let $S_i = S_i^{(j)}$, $Q_i = Q_i^{(j)}$ be the sets constructed in iteration j, and we let $Q(v) = Q^{(j)}(v)$ for $v \in V$ (see steps 2–4 of `Initial`). We start by showing that there is a permutation γ such that ξ_i is 'close' to $\mathcal{E}^{V_{\gamma(i)}}$ for all $1 \leqslant i \leqslant k$, and that the sets Q_i are 'not too small'.

Lemma 6.4. *There is a permutation $\gamma : \{1, \ldots, k\} \to \{1, \ldots, k\}$ such that for each $1 \leqslant i \leqslant k$ we have $\#Q_i \geqslant \frac{1}{2}\#V_{\gamma(i)}$ and $\|\xi_i - \mathcal{E}^{V_{\gamma(i)}}\|^2 \leqslant 0.1\rho^2$.*

Proof. For $1 \leqslant i \leqslant k$ choose $\gamma(i)$ so that $\#Q_i \cap V_{\gamma(i)}$ is maximum (ties are broken arbitrarily). We claim that then for all $1 \leqslant l \leqslant k$ the following three inequalities hold:

$$\#Q_l \geqslant \max\{\#V_i : i \in \{1, \ldots, k\} \setminus \gamma(\{1, \ldots, l-1\})\} - 0.01 n_{\min}, \tag{6.12}$$

$$\#Q_l \cap V_{\gamma(l)} \geqslant \#Q_l - 0.01 n_{\min}, \tag{6.13}$$

$$\|\xi_l - \mathcal{E}^{V_{\gamma(l)}}\|^2 \leqslant 0.1\rho^2. \tag{6.14}$$

The proof is by induction on l. Thus, let us assume that (6.12)–(6.14) hold for all $l < L$. We are going to show that (6.12)–(6.14) are then true for $l = L$ as well. As a first step, we establish (6.12). To this end, consider a class V_i such that $i \notin \gamma(\{1, \ldots, L-1\})$ and let $Z_i = \{v \in V_i : \|\hat{A}_v - \mathcal{E}^{V_i}\|^2 \leqslant 0.001\rho^2\}$. Recall our bound (3.2) on $\|\hat{A} - \mathcal{E}\|_F^2$: we have $\|\hat{A} - \mathcal{E}\|_F^2 \leqslant c_0^5 k^3 \lambda$. Therefore,

$$0.001\rho^2(\#V_i - \#Z_i) \leqslant \sum_{v \in V_i \setminus Z_i} \|\hat{A}_v - \mathcal{E}_v\|^2 \leqslant \|\hat{A} - \mathcal{E}\|_F^2 \leqslant c_0^5 k^3 \lambda. \tag{6.15}$$

Since $\rho^2 = c_0^8 k^3 \lambda / n_{\min}$ for some large constant c_0 (see **A2**), (6.15) implies

$$\#Z_i \geqslant \#V_i - 0.01 n_{\min}. \tag{6.16}$$

Moreover, the triangle inequality shows that, for any two vertices $v, w \in Z_i$, the bound $\|\hat{A}_v - \hat{A}_w\|^2 \leqslant 2(\|\hat{A}_v - \mathcal{E}^{V_i}\|^2 + \|\hat{A}_w - \mathcal{E}^{V_i}\|^2) \leqslant 0.004\rho^2$ holds. In effect, for all $v \in Z_i$ we have

$$Q(v) = \{w \in V : \|\hat{A}_v - \hat{A}_w\|^2 \leqslant 0.01\rho^2\} \supset Z_i. \tag{6.17}$$

Now, consider a vertex $v \in Z_i$ and a vertex w such that $w \in Q_l$ for some $l < L$. Since $i \neq \gamma(l)$ by our choice of i, the separation condition **A2** shows that $\|\mathcal{E}^{V_{\gamma(l)}} - \mathcal{E}_v\| \geqslant \rho$. Consequently,

$$\rho \leqslant \|\mathcal{E}^{V_{\gamma(l)}} - \mathcal{E}_v\| \leqslant \|\mathcal{E}_v - \hat{A}_v\| + \|\hat{A}_w - \hat{A}_v\| + \|\xi_l - \hat{A}_w\| + \|\xi_l - \mathcal{E}^{V_{\gamma(l)}}\|. \tag{6.18}$$

https://doi.org/10.1017/S0963548309990514 Published online by Cambridge University Press

Since $w \in Q_l$, the construction in step 3 of `Initial` ensures that $\|\hat{A}_w - \xi_l\| \leqslant 0.1\rho$. Furthermore, $\|\xi_l - \mathcal{E}^{V_{\gamma(l)}}\| \leqslant \rho/3$ by the induction hypothesis (see (6.14)). Moreover, $\|\hat{A}_v - \mathcal{E}_v\| \leqslant 0.1\rho$, because $v \in Z_i$. Hence, (6.18) entails that $\|\hat{A}_w - \hat{A}_v\| > 0.1\rho$, and thus $w \notin Q(v)$. Since this holds for all $w \in Q_l$ with $l < L$, (6.17) yields

$$Z_i \cap Q_l = \emptyset \quad \text{for all} \quad l < L. \tag{6.19}$$

Finally, let v_L signify the vertex chosen by step 3 of `Initial` to construct $Q_L = Q(v_L) \setminus \bigcup_{l<L} Q_l$. The vertex v_L is chosen so that

$$\#Q_L = \#Q(v_L) \setminus \bigcup_{l=1}^{L-1} Q_l \geqslant \#Q(v) \setminus \bigcup_{l=1}^{L-1} Q_l \quad \text{for all} \quad v \in V \setminus \bigcup_{l=1}^{L-1} Q_l.$$

Since $Z_i \cap \bigcup_{l=1}^{L-1} Q_l = \emptyset$ by (6.19), for all $v \in Z_i$ we have

$$\#Q_L \geqslant \#Q(v) \setminus \bigcup_{l=1}^{L-1} Q_l \overset{(6.17)}{\geqslant} \#Z_i \overset{(6.16)}{\geqslant} \#V_i - 0.01 n_{\min}.$$

As this estimate holds for all $i \notin \gamma(\{1, \ldots, L-1\})$, (6.12) follows.

Thus, we know that Q_L is 'big'. As a next step, we prove (6.13), *i.e.*, we show that Q_L 'mainly' consists of vertices of $V_{\gamma(L)}$. Remember that $\gamma(L)$ was chosen so that $\#Q_L \cap V_{\gamma(L)}$ is maximum. In addition, let $1 \leqslant i \leqslant k$ be such that $\|\mathcal{E}^{V_i} - \hat{A}_{v_L}\|$ is minimum. We are going to show that $i = \gamma(L)$. Let $w \in Q_L \setminus V_i$. Then $\|\mathcal{E}_w - \hat{A}_{v_L}\| \geqslant \|\mathcal{E}^{V_i} - \hat{A}_{v_L}\|$ by the choice of i. Further, assumption **A2** ensures that $\|\mathcal{E}_w - \mathcal{E}^{V_i}\| \geqslant \rho$. Hence, by the triangle inequality

$$\rho \leqslant \|\mathcal{E}_w - \mathcal{E}^{V_i}\| \leqslant \|\mathcal{E}_w - \hat{A}_{v_L}\| + \|\mathcal{E}^{V_i} - \hat{A}_{v_L}\| \leqslant 2\|\mathcal{E}_w - \hat{A}_{v_L}\|.$$

Consequently, $\|\mathcal{E}_w - \hat{A}_{v_L}\|^2 \geqslant \frac{1}{4}\rho^2$. On the other hand, as $w \in Q_L \subset Q(v_L)$, we have $\|\hat{A}_w - \hat{A}_{v_L}\|^2 \leqslant 0.01\rho^2$. Therefore, we obtain

$$\|\hat{A}_w - \mathcal{E}_w\| \geqslant \|\mathcal{E}_w - \hat{A}_{v_L}\| - \|\hat{A}_w - \hat{A}_{v_L}\| \geqslant 0.4\rho.$$

Since this is true for all $w \in Q_L \setminus V_i$, we conclude that

$$0.16\rho^2 \cdot \#Q_L \setminus V_i \leqslant \sum_{w \in Q_L \setminus V_i} \|\hat{A}_w - \mathcal{E}_w\|^2 \leqslant \|\hat{A} - \mathcal{E}\|_F^2. \tag{6.20}$$

Furthermore, we know from (3.2) that $\|\hat{A} - \mathcal{E}\|_F^2 \leqslant c_0^5 k^3 \lambda$. Therefore, as $\rho^2 = c_0^8 k^3 \lambda / n_{\min}$ for a large constant $c_0 > 0$ (see **A2**), (6.20) yields

$$\#Q_L \setminus V_i \leqslant \frac{\|\hat{A} - \mathcal{E}\|_F^2}{0.16\rho^2} \leqslant \frac{c_0^5 k^3 \lambda}{0.16\rho^2} < 0.01 n_{\min}. \tag{6.21}$$

Since we have already established (6.12), we know that $\#Q_L \geqslant \#V_i - 0.01 n_{\min}$, and thus (6.21) shows that $\#V_i \cap Q_L \geqslant 0.99 \#Q_L$. This implies $i = \gamma(L)$, because $\gamma(L)$ was chosen to be the index j such that $\#Q_L \cap V_j$ is maximum. In effect, $\#Q_L \cap V_{\gamma(L)} = \#Q_L \cap V_i \geqslant \#Q_L - 0.01 n_{\min}$. This completes the proof of (6.13).

To show (6.14), we note that by construction we have $\|\hat{A}_w - \hat{A}_{v_L}\| \leqslant 0.1\rho$ for all $w \in Q_L$ (see step 3 of `Initial`). As ξ_L is the arithmetic mean of the vectors $\hat{A}_w$ over $w \in Q_L$, this implies $\|\xi_L - \hat{A}_{v_L}\| \leqslant 0.1\rho$. Hence, by the triangle inequality

$$\begin{aligned}
&\#Q_L \cap V_{\gamma(L)} \|\mathcal{E}^{V_{\gamma(L)}} - \xi_L\|^2 \\
&\quad \leqslant 3 \sum_{w \in Q_L \cap V_{\gamma(L)}} \|\xi_L - \hat{A}_{v_L}\|^2 + \|\hat{A}_w - \hat{A}_{v_L}\|^2 + \|\hat{A}_w - \mathcal{E}^{V_{\gamma(L)}}\|^2 \\
&\quad \leqslant 0.06\rho^2 \#Q_L \cap V_{\gamma(L)} + 3 \sum_{w \in Q_L \cap V_{\gamma(L)}} \|\hat{A}_w - \mathcal{E}^{V_{\gamma(L)}}\|^2 \\
&\quad \leqslant 0.06\rho^2 \#Q_L \cap V_{\gamma(L)} + 3\|\hat{A} - \mathcal{E}\|_F^2. \qquad (6.22)
\end{aligned}$$

As the construction of $\hat{A}$ ensures that $\|\hat{A} - \mathcal{E}\|_F^2 \leqslant c_0^5 k^3 \lambda$ (see (3.2)), (6.22) yields

$$\#Q_L \cap V_{\gamma(L)} \|\mathcal{E}^{V_{\gamma(L)}} - \xi_L\|^2 \leqslant 0.06\rho^2 \#Q_L \cap V_{\gamma(L)} + 3c_0^5 k^3 \lambda. \qquad (6.23)$$

Since $\#Q_L \cap V_{\gamma(L)} \geqslant 0.9 n_{\min}$ due to (6.12) and (6.13), and because $\rho^2 = c_0^8 k^3 \lambda / n_{\min}$ for a large constant c_0 (see **A2**), (6.23) entails that $\|\mathcal{E}_{\gamma(L)} - \xi_L\|^2 \leqslant 0.06\rho^2 + 6c_0^5 k^3 \lambda / n_{\min} \leqslant 0.1\rho^2$. Thus, (6.14) follows.

Finally, (6.12)–(6.13) imply the assertion. To see that γ is a bijection, let us assume that $\gamma(l) = \gamma(l')$ for two indices $1 \leqslant l < l' \leqslant k$. Indeed, choose l to be the least index such that $\gamma(l) = \gamma(l')$. Then $\#Q_l \geqslant \#V_{\gamma(l)} - 0.01 n_{\min}$ by (6.12), and thus $\#V_{\gamma(l)} \setminus Q_l \leqslant 0.1 n_{\min}$ by (6.13). Therefore, we obtain the contradiction

$$0.99 n_{\min} \overset{(6.12)}{\leqslant} \#Q_{l'} \overset{(6.13)}{\leqslant} 1.1 \#Q_{l'} \cap V_{\gamma(l)} \leqslant 1.1 \#V_{\gamma(l)} \setminus Q_l \leqslant 0.11 n_{\min}.$$

Finally, as γ is bijective, (6.12) entails that $\#Q_l \geqslant 0.9 V_{\gamma(l)}$ for all $1 \leqslant l \leqslant k$. Hence, due to (6.13) we obtain $\#Q_l \cap V_l \geqslant 0.9 \#Q_l \geqslant \frac{1}{2} \#V_{\gamma(l)}$, as desired. □

From now on we shall assume without loss of generality that the map γ from Lemma 6.4 is just the identity, *i.e.*, $\gamma(i) = i$ for all i. Bootstrapping on the estimate $\|\xi_i - \mathcal{E}^{V_i}\|^2 \leqslant 0.1\rho^2$ for $1 \leqslant i \leqslant k$ from Lemma 6.4, we derive the following stronger estimate.

Corollary 6.5. *For all* $1 \leqslant i \leqslant k$ *we have* $\|\xi_i - \mathcal{E}^{V_i}\|^2 \leqslant 100 \#Q_i^{-1} \sum_{v \in Q_i} \|\hat{A}_v - \mathcal{E}_v\|^2$.

Proof. The vector ξ_i is just the arithmetic mean of the vectors $\hat{A}_v$ over $v \in Q_i$. Therefore, using first the triangle inequality and then Cauchy–Schwarz, we obtain

$$\|\xi_i - \mathcal{E}^{V_i}\| \leqslant \#Q_i^{-1} \left\| \sum_{v \in Q_i} \hat{A}_v - \mathcal{E}^{V_i} \right\| \leqslant \#Q_i^{-1/2} \left[\sum_{v \in Q_i} \|\hat{A}_v - \mathcal{E}^{V_i}\|^2 \right]^{1/2}. \qquad (6.24)$$

Lemma 6.4 shows that $\|\xi_i - \mathcal{E}^{V_i}\|^2 \leqslant 0.1\rho^2$. In addition, all $v \in Q_i$ satisfy $\|\hat{A}_v - \xi_i\| \leqslant 0.2\rho$ due to the construction of Q_i in steps 2–3. Therefore, for all $v \in Q_i \setminus V_i$ we have

$$\|\hat{A}_v - \mathcal{E}^{V_i}\|^2 \leqslant 2(\|\hat{A}_v - \xi_i\|^2 + \|\xi_i - \mathcal{E}^{V_i}\|^2) \leqslant \rho^2/3. \qquad (6.25)$$

Hence, as $\|\mathcal{E}_v - \mathcal{E}^{V_i}\|^2 \geqslant \rho^2$ by the separation condition **A2**, (6.25) implies that $\|\hat{A}_v - \mathcal{E}_v\| \geqslant 0.1\|\hat{A}_v - \mathcal{E}^{V_i}\|$ for all $v \in Q_i$. Therefore, the assertion follows from (6.24). □

https://doi.org/10.1017/S0963548309990514 Published online by Cambridge University Press

As a next step we are going to analyse the set S_i. Recall that steps 2–3 construct the sets Q_i as the set $Q(v_i)$ of vertices w such that $\hat{A}_w$ is 'close' to $\hat{A}_{v_i}$ minus the vertices covered by previous sets $Q_1, \ldots, Q_{i-1}$. Thus, the sets $Q_1, \ldots, Q_k$ do not necessarily contain all vertices. To obtain a partition $S_1, \ldots, S_k$ of all vertices, step 4 assigns each left-over vertex $v \notin Q_1 \cup \cdots \cup Q_k$ to a class S_i such that the distance $\|\hat{A}_v - \xi_i\|$ between $\hat{A}_v$ and the 'centre vector' ξ_i of class Q_i is minimum. The following lemma relates the resulting distance $\|\hat{A}_v - \xi_i\|$ to the distance $\|\hat{A}_v - \mathcal{E}_v\|$ between $\hat{A}_v$ and the 'expected density' vector $\mathcal{E}_v$.

Corollary 6.6. *For all $v \in S_i \setminus V_i$ we have $\|\hat{A}_v - \xi_i\| \leqslant 3\|\hat{A}_v - \mathcal{E}_v\|$.*

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

$$\|\hat{A}_v - \xi_i\| \leqslant \|\hat{A}_v - \xi_l\|. \tag{6.26}$$

If $v \in S_i \cap V_l \setminus Q_i$, then the construction of S_i in step 4 of `Initial` guarantees that $\|\hat{A}_v - \xi_i\| \leqslant \|\hat{A}_v - \xi_l\|$, as claimed. Thus, assume that $v \in Q_i \cap V_l$. Then

$$\begin{aligned} \|\hat{A}_v - \xi_i\| &\leqslant 0.15\rho && \text{(by the definition of } Q_i \text{ in step 3 of } \texttt{Initial}\text{)}, \\ \max\{\|\xi_i - \mathcal{E}^{V_i}\|, \|\xi_l - \mathcal{E}_v\|\} &\leqslant \frac{1}{3}\rho && \text{(by Lemma 6.4)}, \\ \|\mathcal{E}^{V_i} - \mathcal{E}_v\| &\geqslant \rho && \text{(by the separation assumption } \textbf{A2}\text{)}. \end{aligned} \tag{6.27}$$

Assume for contradiction that $\|\hat{A}_v - \xi_l\| < \|\hat{A}_v - \xi_i\|$. Then the three estimates above would yield

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

which is clearly untrue. Thus, we conclude that $\|\hat{A}_v - \xi_l\| \geqslant \|\hat{A}_v - \xi_i\|$, thereby proving (6.26).

To complete the proof, we use the bound $\|\mathcal{E}_v - \xi_l\| \leqslant \rho/3$ once more (see (6.27)). It implies in combination with the triangle inequality and (6.26) that

$$\|\hat{A}_v - \xi_i\| \leqslant \|\hat{A}_v - \mathcal{E}_v\| + \|\mathcal{E}_v - \xi_i\| \leqslant \|\hat{A}_v - \mathcal{E}_v\| + \|\mathcal{E}_v - \xi_l\| \leqslant \|\hat{A}_v - \mathcal{E}_v\| + \rho/3.$$

Hence, as $\|\xi_i - \mathcal{E}^{V_i}\| \leqslant \rho/3$ (again by (6.27)) and $\|\mathcal{E}_v - \mathcal{E}^{V_i}\| = \|\mathcal{E}^{V_i} - \mathcal{E}^{V_l}\| \geqslant \rho$ (by the separation assumption **A2**), we obtain

$$\rho \leqslant \|\mathcal{E}_v - \mathcal{E}^{V_i}\| \leqslant \|\hat{A}_v - \xi_i\| + \|\xi_i - \mathcal{E}^{V_i}\| + \|\hat{A}_v - \mathcal{E}_v\| \leqslant 2\|\hat{A}_v - \mathcal{E}_v\| + \frac{2}{3}\rho.$$

This shows that $\|\hat{A}_v - \mathcal{E}_v\| \geqslant \frac{1}{6}\rho$. Finally, the estimate

$$\|\hat{A}_v - \xi_i\| \overset{(6.26)}{\leqslant} \|\hat{A}_v - \xi_l\| \leqslant \|\hat{A}_v - \mathcal{E}_v\| + \|\mathcal{E}_v - \xi_l\| \overset{(6.27)}{\leqslant} \|\hat{A}_v - \mathcal{E}_v\| + \frac{\rho}{3} \leqslant 3\|\hat{A}_v - \mathcal{E}_v\|$$

implies the assertion. □

https://doi.org/10.1017/S0963548309990514 Published online by Cambridge University Press

Proof of Lemma 6.3. Since $\#Q_i \geqslant \frac{1}{2}\#V_i$ by Lemma 6.4, we have the estimate

$$\sum_{i=1}^{k}\sum_{w\in S_i\cap V_i}\|\hat{A}_w-\xi_i\|^2 \leqslant 2\sum_{i=1}^{k}\sum_{w\in S_i\cap V_i}\left[\|\hat{A}_w-\mathcal{E}_w\|^2+\|\mathcal{E}_w-\xi_i\|^2\right]$$
$$\overset{\text{Cor. 6.5}}{\leqslant} 2\|\hat{A}-\mathcal{E}\|_F^2+200\sum_{i=1}^{k}\frac{\#S_i\cap V_i}{\#Q_i}\sum_{v\in Q_i}\|\hat{A}_v-\mathcal{E}_v\|^2 \leqslant 500\|\hat{A}-\mathcal{E}\|_F^2. \qquad (6.28)$$

Moreover, since by Corollary 6.6 we have $\|\hat{A}_v-\xi_i\| \leqslant 3\|\hat{A}_v-\mathcal{E}_v\|$ for all $v\in S_i\setminus V_i$, we get

$$\sum_{i=1}^{k}\sum_{v\in S_i\setminus V_i}\|\hat{A}_v-\xi_i\|^2 \leqslant 9\sum_{i=1}^{k}\sum_{v\in S_i\setminus V_i}\|\hat{A}_v-\mathcal{E}_v\|^2 \leqslant 9\|\hat{A}-\mathcal{E}\|_F^2. \qquad (6.29)$$

Since $\|\hat{A}-\mathcal{E}\|_F^2 \leqslant c_0^5k^3\lambda$ by (3.2), the bounds (6.28) and (6.29) imply the assertion. □

7. Local improvement

7.1. The procedure `Improve`

When the subroutine `Improve` gets called in step 4 of `Partition`, the first three steps have already computed a partition $S_1,\ldots,S_k$ along with the 'centre vectors' $\xi_1,\ldots,\xi_k\in\mathbb{R}^V$ such that the following three statements are true (see Proposition 3.4; we are assuming without loss of generality that the index permutation τ is just the identity):

$$\|\xi_i-\mathcal{E}^{V_i}\|^2 \leqslant 0.001\rho^2 \quad \text{for} \quad i=1,\ldots,k, \qquad (7.1)$$

$$\sum_{i=1}^{k}\#S_i\triangle V_i < 0.001n_{\min}, \qquad (7.2)$$

$$\sum_{a,b=1}^{k}\#S_a\cap V_b\cdot\|\mathcal{E}^{V_a}-\mathcal{E}^{V_b}\|^2 < 0.001n_{\min}\rho^2. \qquad (7.3)$$

Starting from this partition $S_1,\ldots,S_k$, `Improve` aims to find a partition $T_1,\ldots,T_k$ that coincides with the desired partition $V_1,\ldots,V_k$ on the 'core' subgraph H. Recall that H is an induced subgraph of G that satisfies conditions **H1–H4** detailed in Section 1.2. These conditions basically state that H consists of 'well-behaved' vertices v for which the number $e(v,V_j)$ of neighbours of v in class V_j is approximately equal to the 'expected number' $\mu(v,V_i)=\#V_j\cdot p_{\psi(v)j}$.

Whereas `Partition`, up to this point, has only exploited spectral properties of G, `Improve` is a combinatorial procedure. It is based on comparing the numbers $e(v,S_i)$ of neighbours that vertex v has in the classes $S_1,\ldots,S_k$ with the 'expected' numbers of neighbours $\mu(v,V_i)=\#V_j\cdot p_{\psi(v)j}$. The obvious problem is that `Improve` does not know the latter. But if $\mu(v,V_i)$ were known to the algorithm, a natural approach would be the following. By (7.2) the partition $(S_1,\ldots,S_k)$ is already close to the desired partition $V_1,\ldots,V_k$. Therefore, we would expect that for 'most' vertices v we have $e(v,S_i)\approx e(v,V_i)$. Suppose that $v\in S_i\cap V_j$ and that $v\in H$. Thus, v is in S_i but should get assigned to class

https://doi.org/10.1017/S0963548309990514 Published online by Cambridge University Press

Algorithm 7.1. $\texttt{Improve}(G, S_1, \ldots, S_k, \xi_1, \ldots, \xi_k)$
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 $\delta(v) = (\delta(v, w))_{w \in V} \in \mathbb{R}^V$.
   For all $v \in V$ pick $1 \leqslant \gamma(v) \leqslant k$ such that $\|\delta(v) - \xi_{\gamma(v)}\| = \min_{1 \leqslant i \leqslant k} \|\delta(v) - \xi_i\|$ (ties are broken arbitrarily). Then, update $S_i = \gamma^{-1}(i)$ for $i = 1, \ldots, k$.
3. Return the partition $S_1, \ldots, S_k$.

Figure 5. The subroutine `Improve`.

T_j by `Improve`. Since $v \in H$, we know that $e(v, V_i) \approx \mu(v, V_i) = \#V_i \cdot p_{ji}$. Hence, all we need to do is look for an index $\gamma(v)$ such that $e(v, S_i)$ is 'closest' to $\#V_i \cdot p_{\gamma(v)i}$ for all $i = 1, \ldots, k$. If actually $e(v, S_i) \approx e(v, V_i)$, then due to the separation condition **A2** this will yield $\gamma(v) = j$.

Of course, we need to specify what it means that $e(v, S_i)$ is 'closest' to $\#V_i \cdot p_{\gamma(v)i}$ for all i. Condition **H2** suggests the following. For each vertex v we set up a vector $\delta(v) = (\delta(v, w))_{w \in V}$. The entries are defined as follows: if $w \in S_l$, then $\delta(v, w) = e(v, S_l)/\#S_l$ is the 'empirical' density between v and class S_l. This is somewhat analogous to the vector $d(v) = d(v, w)$ from condition **H2**, whose entries are $d(v, w) = e(v, V_l)/\#V_l$ for $w \in V_l$. Indeed, **H2** states that $\|d(v) - \mathcal{E}_v\|^2 \leqslant 0.001\rho^2$ is 'small' for all $v \in H$. Hence, we expect that for 'most' $v \in H$ $\|\delta(v) - \mathcal{E}_v\|$ is also small, because by (7.2) the partition $S_1, \ldots, S_k$ is already 'close' to $V_1, \ldots, V_k$. Thus, $\gamma(v) \in \{1, \ldots, k\}$ should be chosen so that $\|\delta(v) - \mathcal{E}^{V_{\gamma(v)}}\| = \min_l \|\delta(v) - \mathcal{E}^{V_l}\|$. Then for 'most' $v \in H$ we expect $\gamma(v)$ to be 'correct', *i.e.*, $v \in V_{\gamma(v)}$. Finally, in order to ensure that *all* $v \in H$ get classified correctly, we could repeat this procedure, say, $\log_2 n$ times. That is, we compute γ as before and update the partition $S_1, \ldots, S_k$ by letting $S_j = \gamma^{-1}(j)$. Then recompute γ, and so on.

The remaining gap is that `Improve` should compute $\|\delta(v) - \mathcal{E}^{V_l}\|$ for $l = 1, \ldots, k$ but does not know the vectors $\mathcal{E}^{V_1}, \ldots, \mathcal{E}^{V_k}$. However, `Improve` does know the vectors $\xi_1, \ldots, \xi_k$, which are good approximations of $\mathcal{E}^{V_1}, \ldots, \mathcal{E}^{V_k}$ by (7.1). Hence, instead of computing $\|\delta(v) - \mathcal{E}^{V_l}\|$ `Improve` calculates $\|\delta(v) - \xi_l\|$. This leads to the pseudocode detailed in Figure 5.

In order to establish Proposition 3.5 we need to prove that `Improve` actually homes in on the planted partition. To this end, 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 $1 \leqslant j \leqslant k$. We will 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 $\|\delta_{\mathcal{S}}(v) - \xi_i\| = \min_{1 \leqslant j \leqslant k} \|\delta_{\mathcal{S}}(v) - \xi_j\|$. 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 $\frac{1}{2}\#V_i \leqslant \#\mathcal{S}_i \leqslant 2\#V_i$ for all i. In addition, we set $\mathcal{S}_{ij} = \mathcal{S}_i \cap V_j$. In words, $\mathcal{S}_{ij}$ contains all vertices that are in class $\mathcal{S}_i$ but actually belong in class $\mathcal{S}_j$. Let us call $\mathcal{S}$ *tight* if $\sum_{i \neq j} \#\mathcal{S}_{ij} \|\mathcal{E}^{V_i} - \mathcal{E}^{V_j}\|^2 \leqslant 0.001\rho^2 n_{\min}$. Thus, (7.2) and (7.3)

https://doi.org/10.1017/S0963548309990514 Published online by Cambridge University Press

ensure that the initial partition partition $\mathcal{S} = (S_1, \ldots, S_k)$ given to `Improve` is both feasible and tight. Therefore, Proposition 3.5 will follow from the next two lemmas, which we shall prove in Sections 7.2 and 7.3.

Lemma 7.2. *If $\mathcal{S}$ is feasible and tight, then any improvement $\mathcal{R}$ of $\mathcal{S}$ is tight.*

Lemma 7.3. *Suppose that $\mathcal{S}$ is feasible and tight and that $\mathcal{R}$ is an improvement of $\mathcal{S}$. Then we have $\sum_{i \neq j} \#\mathcal{R}_{ij} \cap H \leqslant \frac{1}{10} \sum_{i \neq j} \#\mathcal{S}_{ij} \cap H$.*

Proof of Proposition 3.5. Let $\mathcal{S} = (\mathcal{S}_1, \ldots, \mathcal{S}_k)$ be a feasible and tight partition such that $\sum_{i=1}^{k} \#\mathcal{S}_i \triangle V_i \leqslant 0.001 n_{\min}$, and let $\mathcal{R}$ be an improvement of $\mathcal{S}$. Then, by Lemma 7.2 $\mathcal{R}$ is tight, and by Lemma 7.3 we have

$$\begin{aligned}\sum_{i=1}^{k} \#\mathcal{R}_i \triangle V_i \leqslant 0.1 \sum_{i=1}^{k} \#\mathcal{S}_i \triangle V_i &\leqslant 0.1 \# V \setminus H + 0.1 \sum_{i \neq j} \#\mathcal{S}_{ij} \cap H \\ &\leqslant 0.1 \# V \setminus H + 10^{-4} n_{\min}.\end{aligned}$$

Since, by condition **H1**, $\# V \setminus H \leqslant n_{\min}/\lambda^4$, where $\lambda > c_0$ for a large constant $c_0 > 0$ (see **A1**), we have $0.1 \# V \setminus H \leqslant 10^{-4} n_{\min}$. Hence, $\mathcal{R}$ is feasible. Thus, as the partition $(S_1, \ldots, S_k)$ with which `Improve` starts is feasible and tight by Proposition 3.4, and in fact satisfies

$$\sum_{i=1}^{k} \# S_i \triangle V_i \leqslant 0.001 n_{\min},$$

all the partitions generated by `Improve` remain feasible and tight.

Let $\mathcal{T} = (\mathcal{T}_1, \ldots, \mathcal{T}_k)$ denote the partition returned by `Improve`. Then, due to Lemma 7.3 we have $\sum_{i \neq j} \#\mathcal{T}_{ij} \cap H \leqslant 10^{-\log_2 n} \cdot n_{\min} < 1$, whence $\mathcal{T}_i \cap H = V_i \cap H$ for all i. □

To facilitate the proof of Lemmas 7.2 and 7.3, we introduce some notation. Let $A = A(G)$ be the adjacency matrix and

$$M = \mathcal{E}_H - A_H. \tag{7.4}$$

Thus, for $v, w \in H$ the entry M_{vw} is $\mathcal{E}_{vw}$ minus one if v, w are adjacent, and just $\mathcal{E}_{vw}$ otherwise. The spectral condition **A1** and condition **H3** on the vertex degrees in H yield the bound

$$\|M\| \leqslant c_0^2 k \sqrt{\lambda}. \tag{7.5}$$

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

$$\|M \mathbf{1}_S\|^2 = \sum_{v \in H} |e(v, S) - \mu'(v, S)|^2 \quad \text{for all} \quad S \subset V(H). \tag{7.6}$$

Recall that $\mu(v, S) = \sum_{w \in S \setminus \{v\}} \mathcal{E}_{vw}$ denotes the 'expected' number of edges between v and S. The relation between $\mu(v, S)$ and $\mu'(v, S)$ is that $\mu'(v, S) = \mu(v, S) + \mathcal{E}_{vv}$ 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 $w \in \mathcal{S}_l$

https://doi.org/10.1017/S0963548309990514 Published online by Cambridge University Press

we set

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

Thus, we can think of $\bar{\delta}(v) \in \mathbb{R}^V$ as the 'expectation' of $\delta(v)$.

7.2. Proof of Lemma 7.2

Throughout this section we fix a partition $\mathcal{S}$ that is both feasible and tight and a partition $\mathcal{R}$ that is an improvement of $\mathcal{S}$. We are going to prove the two inequalities

$$9\sum_{v\in V}\|\delta(v) - \bar{\delta}(v)\|^2 \leqslant 0.001\rho^2 n_{\min}, \tag{7.7}$$

$$\sum_{a,b=1}^{k}\#\mathcal{R}_{ab}\|\mathcal{E}^{V_a} - \mathcal{E}^{V_b}\|^2 \leqslant 9\sum_{v\in V}\|\delta(v) - \bar{\delta}(v)\|^2, \tag{7.8}$$

from which Lemma 7.2 is immediate. Observe that by the definitions of $\delta(v)$, $\bar{\delta}(v)$

$$\begin{aligned}\sum_{v\in V}\|\delta(v) - \bar{\delta}(v)\|^2 &= \sum_{a=1}^{k}\#\mathcal{S}_a\left(\frac{e(v,\mathcal{S}_a) - \mu'(v,\mathcal{S}_a)}{\#\mathcal{S}_a}\right)^2\\ &= \sum_{a=1}^{k}\frac{(e(v,\mathcal{S}_a) - \mu'(v,\mathcal{S}_a))^2}{\#\mathcal{S}_a}.\end{aligned} \tag{7.9}$$

As a first step we establish (7.7). To this end, we need to remember the decomposition of the graph G into its 'sparse' part G_1 and its 'dense' part G_2: recall the matrix $\Phi = (\Phi_{vw})_{v,w\in V}$ whose entries are $\Phi_{vw} = 1$ if $\mathcal{E}_{vw} > \frac{1}{2}$ and $\Phi_{vw} = 0$ if $\mathcal{E}_{vw} \leqslant \frac{1}{2}$. Then G_1 contains all edges $\{v,w\} \in E$ such that $\Phi_{vw} = 0$ and G_2 contains all edges $\{v,w\} \notin E$ such that $\Phi_{vw} = 1$. For $v \in V$ and $S \subset V$ we let

$$\mu_1'(v,S) = \sum_{w\in S:\Phi_{vw}=0}\mathcal{E}_{vw}, \quad \mu_2'(v,S) = \sum_{w\in S:\Phi_{vw}=1} 1 - \mathcal{E}_{vw}.$$

Thus, we can think of $\mu_a'(v,S)$ as the 'expected number' of neighbours that v has in S in the graph G_a $(a = 1,2)$. Moreover, we let $e_a(v,S)$ be the number of v–S-edges in the graph G_a. Finally, recall that $\mu'(v,S) = \sum_{w\in S}\mathcal{E}_{vw}$.

Lemma 7.4. *For any set $S \subset V$ and any $v \in V$ we have*

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

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 $\imath = 1$ if $v \in S_2$ and $\imath = 0$ otherwise. Then, by the definition of the graphs G_1, G_2 we have

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

whence the assertion follows. □

https://doi.org/10.1017/S0963548309990514 Published online by Cambridge University Press

The 'core' subgraph H (which satisfies conditions **H1–H4**) contains vertices for which the numbers $e(v, V_i)$ are 'close' to the expected numbers $\mu(v, V_i) = \sum_{w \in V_i} \mathcal{E}_{vw}$. Moreover, by (7.5) the matrix $M = \mathcal{E}_H - A_H$ has a small spectral norm. Intuitively this means that $A_H \approx \mathcal{E}_H$, *i.e.*, the graph H is quasi-random with respect to the density matrix $\mathcal{E}$. Hence, we expect that for 'most' vertices $v \in H$ and 'sufficiently large' sets $S \subset V$ we have $e(v, S) \approx \mu(v, S)$. In effect, for most $v \in H$ we should have $\delta(v) \approx \bar{\delta}(v)$. The following lemma, which is the key step in the proof of (7.8), provides a precise estimate.

Lemma 7.5. *Let $\mathcal{S}$ be a feasible partition. Then* $\sum_{v \in H} \|\delta(v) - \bar{\delta}(v)\|^2 \leqslant 10^{-4}\rho^2 n_{\min}$.

Proof. 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 by (7.6)

$$\sum_{v \in H} A(v) = \sum_{a=1}^{k} \#\mathcal{S}_a^{-1}\|M\mathbf{1}_{\mathcal{S}_a \cap H}\|^2 \leqslant \|M\|^2 \sum_{a=1}^{k} \frac{\|\mathbf{1}_{\mathcal{S}_a \cap H}\|^2}{\#\mathcal{S}_a}.$$

Since $\|\mathbf{1}_{\mathcal{S}_a \cap H}\|^2 = \#\mathcal{S}_a \cap H \leqslant \#\mathcal{S}_a$, the right-hand side is at most $k\|M\|^2$. Further, (7.5) shows that $k\|M\|^2 \leqslant c_0^2 k^3 \lambda$. Since $\rho^2 = c_0^8 k^3 \lambda / n_{\min}$ for some large constant c_0 (see **A2**), we get $k\|M\|^2 \leqslant 10^{-5}\rho^2 n_{\min}$. Thus,

$$\sum_{v \in H} A(v) = \sum_{v \in H} \sum_{a=1}^{k} \#\mathcal{S}_a^{-1}\left[e(v, \mathcal{S}_a \cap H) - \mu'(v, \mathcal{S}_a \cap H)\right]^2 \leqslant 10^{-5}\rho^2 n_{\min}. \tag{7.10}$$

Furthermore, let

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

Since $e(v, \mathcal{S}_a) = e(v, \mathcal{S}_a \cap H) + e(v, \mathcal{S}_a \setminus H)$ and $\mu'(v, \mathcal{S}_a) = e(v, \mathcal{S}_a \cap H) + \mu'(v, \mathcal{S}_a \setminus H)$, $A'(v)$ is equal to

$$A'(v) = \sum_{a=1}^{k} \#\mathcal{S}_a^{-1}\left[e(v, \mathcal{S}_a) - \mu'(v, \mathcal{S}_a) + e(v, \mathcal{S}_a \cap H) - \mu'(v, \mathcal{S}_a \cap H)\right] \times \left[e(v, \mathcal{S}_a \setminus H) - \mu'(v, \mathcal{S}_a \setminus H)\right]. \tag{7.11}$$

Lemma 7.4 entails that for all $v \in H$ and all $1 \leqslant a \leqslant k$ we have

$$|e(v, \mathcal{S}_a) - \mu'(v, \mathcal{S}_a) + e(v, \mathcal{S}_a \cap H) - \mu'(v, \mathcal{S}_a \cap H)| \leqslant 2 + 2\sum_{i=1}^{2} e_i(v, \mathcal{S}_a) + \mu_i'(v, \mathcal{S}_a).$$

Further, $e_1(v, \mathcal{S}_a) + e_2(v, \mathcal{S}_a)$ is bounded by the degree $d_{G_1 \cup G_2}(v)$ of v in $G_1 \cup G_2$. Since $v \in H$, condition **H3** entails that $d_{G_1 \cup G_2}(v) \leqslant 10\sigma^*$, where $\sigma^* = \max_{v \in V} \sum_{w \in V} \mathcal{E}_{vw}(1 - \mathcal{E}_{vw})$ is the 'maximum variance' of any vertex degree in G. The definition of σ^* also shows that

$$\mu_1'(v, \mathcal{S}_a) + \mu_2'(v, \mathcal{S}_a) \leqslant \mu_1'(v, V) + \mu_2'(v, V) = \sum_{w \in V : \mathcal{E}_{vw} \leqslant \frac{1}{2}} \mathcal{E}_{vw} + \sum_{w \in V : \mathcal{E}_{vw} > \frac{1}{2}} 1 - \mathcal{E}_{vw} \leqslant 2\sigma^*. \tag{7.12}$$

In summary, as $\lambda \geqslant \sigma^*$ by condition **A1**, we obtain that for all $1 \leqslant a \leqslant k$

$$|e(v, \mathcal{S}_a) - \mu'(v, \mathcal{S}_a) + e(v, \mathcal{S}_a \cap H) - \mu'(v, \mathcal{S}_a \cap H)| \leqslant 2 + 24\sigma^* \leqslant 25\lambda. \tag{7.13}$$

https://doi.org/10.1017/S0963548309990514 Published online by Cambridge University Press

Applying Lemma 7.4 once more, we obtain

$$\begin{aligned}\sum_{v\in H}\sum_{a=1}^{k}|e(v,\mathcal{S}_a\setminus H)-\mu'(v,\mathcal{S}_a\setminus H)| &\leqslant \sum_{v\in H}\sum_{a=1}^{k}\sum_{i=1}^{2}e_{G_i}(v,\mathcal{S}_a\setminus H)+\mu_i'(v,\mathcal{S}_a\setminus H)\\ &= \sum_{i=1}^{2}e_{G_i}(H,V\setminus H)+\sum_{v\in H}\mu_i'(v,V\setminus H)\\ &\leqslant \sum_{v\in V\setminus H}d_{G_1\cup G_2}(v)+\mu_1'(v,V)+\mu_1'(v,V)\\ &\leqslant 2\sigma^*\cdot\#V\setminus H+\sum_{v\in V\setminus H}d_{G_1\cup G_2}(v) \qquad \text{(by (7.12))}\\ &\leqslant 2\sigma^*\cdot\#V\setminus H+\sqrt{\#V\setminus H\sum_{v\in V\setminus H}d_{G_1\cup G_2}(v)^2},\end{aligned} \tag{7.14}$$

where the last inequality follows from Cauchy–Schwarz. Now, condition **H1** ensures that $\#V\setminus H\leqslant\lambda^{-4}n_{\min}$, and that $\sum_{v\in V\setminus H}d_{G_1\cup G_2}(v)^2\leqslant n_{\min}$. Hence, (7.14) is at most $(2\sigma^*\lambda^{-4}+\lambda^{-2})n_{\min}$. As $\lambda\geqslant\sigma^*$ by **A1**, we obtain

$$\sum_{v\in H}\sum_{a=1}^{k}|e(v,\mathcal{S}_a\setminus H)-\mu'(v,\mathcal{S}_a\setminus H)|\leqslant(2\sigma^*\lambda^{-4}+\lambda^{-2})n_{\min}\leqslant 2\lambda^{-2}n_{\min}. \tag{7.15}$$

Finally, as $(\mathcal{S}_a)_{1\leqslant a\leqslant k}$ is feasible, we have $\#\mathcal{S}_a\geqslant\frac{1}{2}n_{\min}$ for all a. Therefore, plugging (7.13) and (7.15) into (7.11), we obtain

$$\sum_{v\in H}A'(v)\leqslant\frac{75n_{\min}}{\lambda^2\min_{1\leqslant a\leqslant k}\#\mathcal{S}_a}\leqslant\frac{150}{\lambda^2}. \tag{7.16}$$

Since we are assuming that $\lambda\geqslant\sigma^*\geqslant c_0$ for some sufficiently large $c_0>0$ (see **A1**), (7.16) shows that $\sum_{v\in H}A'(v)\leqslant 1$. Combining (7.9), (7.10), and (7.16), we obtain

$$\sum_{v\in H}\|\delta(v)-\bar{\delta}(v)\|^2\leqslant\sum_{v\in H}A(v)+A'(v)\leqslant 10^{-5}\rho^2n_{\min}+1\leqslant 10^{-4}\rho^2n_{\min},$$

as desired. □

The previous lemma provides an estimate of $\sum_{v\in H}\|\delta(v)-\bar{\delta}(v)\|^2$. The next step is to analyse the remaining vertices, *i.e.*, $\sum_{v\in V\setminus H}\|\delta(v)-\bar{\delta}(v)\|^2$. Since by condition **H1** the subgraph H contains the vast majority of vertices and edges, we expect the latter sum to be quite small.

Lemma 7.6. *Let $\mathcal{S}$ be a feasible partition. Then $\sum_{v\in V\setminus H}\|\delta(v)-\bar{\delta}(v)\|^2\leqslant 2$.*

https://doi.org/10.1017/S0963548309990514 Published online by Cambridge University Press

Proof. For $v \in V$ and $a \in \{1,\ldots,k\}$ let $\imath_{va} = 1$ if $v \in \mathcal{S}_a$ and let $\imath_{va} = 0$ otherwise. Then Lemma 7.4 and Cauchy–Schwarz yield

$$\sum_{v\in V\setminus H}\sum_{a=1}^{k}\#\mathcal{S}_a^{-1}(e(v,\mathcal{S}_a)-\mu'(v,\mathcal{S}_a))^2$$
$$\leqslant \sum_{v\in V\setminus H}\sum_{a=1}^{k}\#\mathcal{S}_a^{-1}\left[\imath_{va}+\sum_{i=1}^{2}e_i(v,\mathcal{S}_a)+\mu_i'(v,\mathcal{S}_a)\right]^2$$
$$\leqslant 5\sum_{v\in V\setminus H}\sum_{a=1}^{k}\#\mathcal{S}_a^{-1}\left[\imath_{va}+\sum_{i=1}^{2}e_i(v,\mathcal{S}_a)^2+\mu_i'(v,\mathcal{S}_a)^2\right]. \tag{7.17}$$

Moreover, we know that $\sum_{i=1}^{2}\sum_{a=1}^{k}\mu_i'(v,\mathcal{S}_a) \leqslant \sum_{i=1}^{2}\mu_i'(s,V) \leqslant 2\sigma^*$ (see (7.12)). In effect,

$$\sum_{i=1}^{2}\sum_{a=1}^{k}\mu_i'(v,\mathcal{S}_a)^2 \leqslant (2\sigma^*)^2.$$

Furthermore, as $\mathcal{S}$ is feasible, we have $\#\mathcal{S}_a \geqslant n_{\min}/2$ for all a. Hence, (7.17) yields

$$\sum_{v\in V\setminus H}\sum_{a=1}^{k}\#\mathcal{S}_a^{-1}(e(v,\mathcal{S}_a)-\mu'(v,\mathcal{S}_a))^2$$
$$\leqslant \frac{\#V\setminus H\cdot(20\sigma^{*2}+5)+\sum_{v\in V\setminus H}d_{G_1\cup G_2}(v)^2}{n_{\min}}. \tag{7.18}$$

As $\#V\setminus H \leqslant \lambda^{-4}n_{\min}$ by **H1** and $\lambda \geqslant \sigma^* \geqslant c_0$ for some large constant $c_0 > 0$ (see **A1**), we have $20\sigma^{*2}\#V\setminus H/n_{\min} \leqslant 20\lambda^{-2} < \frac{1}{2}$. Furthermore, $5\#V\setminus H/n_{\min} \leqslant 5\lambda^{-4} \leqslant \frac{1}{2}$. Hence, (7.18) entails that

$$\sum_{v\in V\setminus H}\sum_{a=1}^{k}\#\mathcal{S}_a^{-1}(e(v,\mathcal{S}_a)-\mu'(v,\mathcal{S}_a))^2 \leqslant 1+\sum_{v\in V\setminus H}d_{G_1\cup G_2}(v)^2/n_{\min}. \tag{7.19}$$

Finally, condition **H1** ensures that $\sum_{v\in V\setminus H}d_{G_1\cup G_2}(v)^2/n_{\min} \leqslant 1$. Hence, (7.19) and (7.9) imply

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

as desired. □

Combining Lemmas 7.5 and 7.6, we obtain $\sum_{v\in V}\|\delta(v)-\bar{\delta}(v)\|^2 \leqslant 10^{-4}\rho^2 n_{\min}+2 < 0.001\rho^2 n_{\min}$, where the last inequality follows from the fact that $\rho^2 n_{\min} = c_0^8 k^3\lambda$ for some large constant c_0 and $\lambda > c_0$ (see **A1**). Thus, we have established (7.7).

To prove (7.8), the following lemma is instrumental. Remember that for each $v \in V$ we have defined a vector $\bar{\delta}(v) = (\bar{\delta}(v,w))_{w\in V}$ by letting $\bar{\delta}(v,w) = \mu'(v,\mathcal{S}_l)/\#\mathcal{S}_l = \sum_{u\in\mathcal{S}_l}\mathcal{E}_{vu}/\#\mathcal{S}_l$ for $w \in \mathcal{S}_l$.

https://doi.org/10.1017/S0963548309990514 Published online by Cambridge University Press

Lemma 7.7. *Let $\mathcal{S}$ be any partition. Then, for all $1 \leqslant i \leqslant k$ and all $v \in V_i$, we have $\|\bar{\delta}(v) - \mathcal{E}_v\|^2 \leqslant 2\sum_{a,b=1}^k \#\mathcal{S}_{ab}(p_{ia} - p_{ib})^2$.*

Proof. Let $\tilde{\delta}(v,w) = p_{ia}$ for all $w \in \mathcal{S}_l$, and set $\tilde{\delta}(v) = (\tilde{\delta}(v,w))_{w\in V}$. Since $\mathcal{E}_{vw} = p_{ib}$ for $w \in V_b$ and $\mathcal{S}_{ab} = \mathcal{S}_a \cap V_b$, we have

$$\|\tilde{\delta}(v) - \mathcal{E}_v\|^2 = \sum_{a,b=1}^k \#\mathcal{S}_{ab}(p_{ia} - p_{ib})^2. \tag{7.20}$$

Moreover, since $\bar{\delta}(v,w) = \mu'(v,\mathcal{S}_a)/\#\mathcal{S}_a$ for $w \in \mathcal{S}_a$, we obtain

$$\begin{aligned}\|\tilde{\delta}(v) - \bar{\delta}(v)\|^2 &= \sum_{a=1}^k \#\mathcal{S}_a\left[\mu'(v,\mathcal{S}_a)\#\mathcal{S}_a^{-1} - p_{ia}\right]^2 = \sum_{a=1}^k \#\mathcal{S}_a^{-1}\left[\mu'(v,\mathcal{S}_a) - \#\mathcal{S}_a p_{ia}\right]^2 \\ &= \sum_{a=1}^k \#\mathcal{S}_a^{-1}\left[\sum_{b=1}^k \mu'(v,\mathcal{S}_{ab}) - \#\mathcal{S}_{ab}p_{ia}\right]^2. \end{aligned} \tag{7.21}$$

Since $\mathcal{S}_{ab} = \mathcal{S}_a \cap V_b$, we have $\mathcal{E}_{vw} = p_{ib}$ for all $w \in \mathcal{S}_{ab}$, and thus $\mu'(v,\mathcal{S}_{ab}) = \sum_{w\in\mathcal{S}_{ab}} \mathcal{E}_{vw} = \#\mathcal{S}_{ab}p_{ib}$. Therefore, (7.21) implies

$$\begin{aligned}\|\tilde{\delta}(v) - \bar{\delta}(v)\|^2 &= \sum_{a=1}^k \#\mathcal{S}_a^{-1}\left[\sum_{b=1}^k \#\mathcal{S}_{ab}(p_{ib} - p_{ia})\right]^2 \\ &\leqslant \sum_{a,b=1}^k \#\mathcal{S}_{ab}(p_{ia} - p_{ib})^2 \qquad \text{(by Cauchy–Schwarz).}\end{aligned} \tag{7.22}$$

Combining (7.20) and (7.22) completes the proof. □

Corollary 7.8. *If $\mathcal{S}$ is tight and $\mathcal{R}$ is an improvement of $\mathcal{S}$, then*

$$\sum_{a,b=1}^k \#\mathcal{R}_{ab}\|\mathcal{E}_a - \mathcal{E}_b\|^2 \leqslant 9\sum_{v\in V}\|\delta(v) - \bar{\delta}(v)\|^2.$$

Proof. Let $v \in \mathcal{R}_{ab}$. Recall that the 'centre vectors' $\xi_1,\ldots,\xi_k$ that `Improve` receives as input parameters satisfy $\|\xi_b - \mathcal{E}_b\|^2 \leqslant 0.001\rho^2$ for $b = 1,\ldots,k$ (see (7.1)). Hence, by the triangle inequality

$$\|\delta(v) - \xi_b\| \leqslant \|\delta(v) - \mathcal{E}_b\| + \|\mathcal{E}_b - \xi_b\| \leqslant \|\delta(v) - \bar{\delta}(v)\| + \|\bar{\delta}(v) - \mathcal{E}_b\| + \sqrt{0.001}\rho.$$

Thus, bounding $\|\bar{\delta}(v) - \mathcal{E}_b\|$ via Lemma 7.7, we obtain

$$\|\delta(v) - \xi_b\| \leqslant \sqrt{0.001}\rho + \|\delta(v) - \bar{\delta}(v)\| + \sqrt{2\sum_{\alpha\neq\beta}\#\mathcal{S}_{\alpha\beta}(p_{b\alpha} - p_{b\beta})^2}. \tag{7.23}$$

For any two indices $\alpha \neq \beta$ and all $w \in V_b$ the w-entry of the vector $\mathcal{E}^{V_\alpha}$ equals $p_{b\alpha}$. Similarly, the w-entry of $\mathcal{E}_w^{V_\beta}$ is equal to $p_{b\beta}$. Therefore, $\|\mathcal{E}^{V_\alpha} - \mathcal{E}^{V_\beta}\|^2 \geqslant \#V_b(p_{b\alpha} - p_{b\beta})^2$.

Plugging this into (7.23), we get

$$\|\delta(v)-\xi_b\| \leqslant \frac{\rho}{30}+\|\delta(v)-\bar{\delta}(v)\|+\sqrt{2\sum_{\alpha\neq\beta}\frac{\#\mathcal{S}_{\alpha\beta}}{\#V_b}\|\mathcal{E}^{V_\alpha}-\mathcal{E}^{V_\beta}\|^2}$$
$$\leqslant \frac{\rho}{30}+\|\delta(v)-\bar{\delta}(v)\|+\sqrt{\frac{2}{n_{\min}}\sum_{\alpha\neq\beta}\#\mathcal{S}_{\alpha\beta}\|\mathcal{E}^{V_\alpha}-\mathcal{E}^{V_\beta}\|^2}, \qquad (7.24)$$

because $\#V_b \geqslant n_{\min}$. Our assumption that $\mathcal{S}$ is tight means that $\sum_{\alpha\neq\beta}\#\mathcal{S}_{\alpha\beta}\|\mathcal{E}^{V_\alpha}-\mathcal{E}^{V_\beta}\|^2 \leqslant 0.001\rho^2 n_{\min}$. Thus, (7.24) yields

$$\|\delta(v)-\xi_b\| \leqslant \|\delta(v)-\bar{\delta}(v)\|+\frac{\rho}{20}. \qquad (7.25)$$

Furthermore, if $v\in\mathcal{R}_{ab}$, then $v\in V_b$ but $\|\delta(v)-\xi_a\|\leqslant\|\delta(v)-\xi_b\|$, because $\mathcal{R}$ is an improvement of $\mathcal{S}$. Since the centre vectors ξ_a,ξ_b satisfy $\|\xi_a-\mathcal{E}_a\|^2, \|\xi_b-\mathcal{E}_b\|^2\leqslant 0.001\rho^2$ (see (7.1)) and $\|\mathcal{E}_a-\mathcal{E}_b\|^2\geqslant\rho^2$ by the separation condition **A2**, we obtain

$$\rho\leqslant\|\mathcal{E}_a-\mathcal{E}_b\|\leqslant\|\mathcal{E}_a-\xi_a\|+\|\mathcal{E}_b-\xi_b\|+\|\xi_a-\xi_b\|$$
$$\leqslant\frac{\rho}{15}+\|\delta(v)-\xi_a\|+\|\delta(v)-\xi_b\|\leqslant\frac{\rho}{15}+2\|\delta(v)-\xi_b\|.$$

Thus, $\|\delta(v)-\xi_b\|\geqslant\frac{2}{5}\|\mathcal{E}_a-\mathcal{E}_b\|\geqslant\frac{2}{5}\rho$. Hence, (7.25) yields $\|\delta(v)-\bar{\delta}(v)\|\geqslant\frac{1}{3}\|\mathcal{E}_a-\mathcal{E}_b\|$. □

As Corollary 7.8 implies (7.8), we have completed the proof of Lemma 7.2.

7.3. Proof of Lemma 7.3

Suppose that $\mathcal{S}$ is a partition that is both feasible and tight; that is, for all $a\in\{1,\dots,k\}$ we have $\frac{1}{2}\#V_a\leqslant\#\mathcal{S}_a\leqslant 2\#V_a$, and letting $\mathcal{S}_{ab}=\mathcal{S}_a\cap V_b$ we have

$$\sum_{a\neq b}\#\mathcal{S}_{ab}\|\mathcal{E}^{V_a}-\mathcal{E}^{V_b}\|^2\leqslant 0.001\rho^2 n_{\min}.$$

Recall that for a vertex v and a set $T\subset V$ we can think of $\mu'(v,T)=\sum_{w\in T}\mathcal{E}_{vw}$ as the 'expected' edge density between v and T. For all $v\in\mathcal{R}_{ab}$, all $\alpha\in\{1,\dots,k\}$, and all $w\in\mathcal{S}_\alpha$, we set

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

Moreover, we let $\Delta(v)=(\Delta(v,w))_{w\in H}$, $\bar{\Delta}(v)=(\bar{\Delta}(v,w))_{w\in H}$, and remember that $\delta(v)=(\delta(v,w))_{w\in V}$, $\bar{\delta}(v)=(\bar{\delta}(v,w))_{w\in V}$. The relationship between $\delta(v)$ and $\Delta(v)$ is that the latter vector disregards the 'exceptional' vertices in $V\setminus H$. We can think of $\bar{\Delta}(v)$ as the 'expectation' of $\Delta(v)$.

Assume that $\mathcal{R}$ is an improvement of $\mathcal{S}$, *i.e.*, for all $v\in\mathcal{R}_a$ we have $\|\delta(v)-\xi_a\|=\min_b\|\delta(v)-\xi_b\|$, where $\xi_1,\dots,\xi_k$ signify the centre vectors that `Improve` received as inputs. Our goal is to show that $\sum_{a\neq b}\#\mathcal{R}_{ab}\cap H\leqslant 0.1\sum_{a\neq b}\#\mathcal{S}_{ab}$. The key step is to prove that for a vertex $v\in\mathcal{R}_{ba}\cap H=\mathcal{R}_b\cap V_a\cap H$ that lies in class b of the improvement $\mathcal{R}$ although it 'belongs' in class $a\neq b$ the vector $\Delta(v)$ is far from its 'expectation' $\bar{\Delta}(v)$.

https://doi.org/10.1017/S0963548309990514 Published online by Cambridge University Press

Lemma 7.9. *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 $\|\Delta(v) - \bar{\Delta}(v)\|^2 \geqslant 0.1\|\mathcal{E}^{V_a} - \mathcal{E}^{V_b}\|^2$ $(a \neq b)$.*

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

$$\|\delta_H(v) - \bar{\delta}_H(v)\| \geqslant 0.134\|\mathcal{E}^{V_a} - \mathcal{E}^{V_b}\|. \tag{7.26}$$

For as $\mathcal{R}$ is an improvement of $\mathcal{S}$ and $v \in \mathcal{R}_b$, we have $\|\delta(v) - \xi_b\| \leqslant \|\delta(v) - \xi_a\|$. Moreover, since $a \neq b$ the separation condition **A2** ensures that $\|\mathcal{E}^{V_a} - \mathcal{E}^{V_b}\| \geqslant \rho$. Therefore, the triangle inequality yields

$$\begin{aligned}\rho \leqslant \|\mathcal{E}^{V_a} - \mathcal{E}^{V_b}\| &\leqslant \|\delta(v) - \xi_b\| + \|\delta(v) - \xi_a\| + \|\xi_a - \mathcal{E}^{V_a}\| + \|\xi_b - \mathcal{E}^{V_b}\| \\ &\leqslant 2\|\delta(v) - \xi_a\| + \|\xi_a - \mathcal{E}^{V_a}\| + \|\xi_b - \mathcal{E}^{V_b}\|. \end{aligned} \tag{7.27}$$

Furthermore, the centre vectors ξ_a, ξ_b are close to $\mathcal{E}^{V_a}, \mathcal{E}^{V_b}$ respectively. More precisely, by (7.1) we have $\|\mathcal{E}^{V_a} - \xi_a\|^2 \leqslant 0.001\rho^2$, $\|\mathcal{E}^{V_b} - \xi_b\|^2 \leqslant 0.001\rho^2$. Hence, (7.27) yields

$$\begin{aligned}\rho \leqslant \|\mathcal{E}^{V_a} - \mathcal{E}^{V_b}\| &\leqslant 2\sqrt{0.001}\rho + 2\|\delta(v) - \xi_a\| \\ &\leqslant \frac{2\rho}{33} + 2\|\xi_a - \mathcal{E}^{V_a}\| + 2\|\delta(v) - \mathcal{E}^{V_a}\| \leqslant \frac{4\rho}{33} + 2\|\delta(v) - \mathcal{E}^{V_a}\|.\end{aligned}$$

Thus,

$$\|\delta(v) - \mathcal{E}^{V_a}\| \geqslant \frac{29}{66}\|\mathcal{E}^{V_a} - \mathcal{E}^{V_b}\| \geqslant \frac{29\rho}{66}. \tag{7.28}$$

Furthermore, Lemma 7.7 shows that

$$\|\bar{\delta}(v) - \mathcal{E}^{V_a}\|^2 \leqslant 4\sum_{\alpha,\beta=1}^{k} \#\mathcal{S}_{\alpha\beta}(p_{a\alpha} - p_{a\beta})^2. \tag{7.29}$$

For all $w \in V_a$ the w-entry of $\mathcal{E}^{V_\alpha}$ is equal to $p_{a\alpha}$, and the w-entry of $\mathcal{E}^{V_\beta}$ equals $p_{a\beta}$. Therefore, $\|\mathcal{E}^{V_\alpha} - \mathcal{E}^{V_\beta}\|^2 \geqslant \#V_a(p_{a\alpha} - p_{a\beta})^2$. Consequently, (7.29) yields, in combination with our assumption that $\mathcal{S}$ is tight,

$$\begin{aligned}\|\bar{\delta}(v) - \mathcal{E}^{V_a}\|^2 &\leqslant 4\sum_{\alpha,\beta=1}^{k} \frac{\#S_{\alpha\beta}\|\mathcal{E}^{V_\alpha} - \mathcal{E}^{V_\beta}\|^2}{\#V_a} \\ &\leqslant \frac{4}{n_{\min}}\sum_{\alpha,\beta=1}^{k} \#S_{\alpha\beta}\|\mathcal{E}^{V_\alpha} - \mathcal{E}^{V_\beta}\|^2 \leqslant 0.004\rho^2.\end{aligned} \tag{7.30}$$

Combining (7.28) and (7.30), we see that

$$\|\delta(v) - \bar{\delta}(v)\| \geqslant 0.37\|\mathcal{E}^{V_a} - \mathcal{E}^{V_b}\|. \tag{7.31}$$

For any vertex $w \in \mathcal{S}_\alpha$ the w-entry of the vector $\delta(v)$ is $e(v, \mathcal{S}_\alpha)/\#\mathcal{S}_\alpha$, and the w-entry of $\bar{\delta}(v)$ is $\mu'(v, \mathcal{S}_\alpha)/\#\mathcal{S}_\alpha$. Therefore, (7.31) entails

$$0.136\|\mathcal{E}^{V_a} - \mathcal{E}^{V_b}\|^2 \leqslant \|\delta(v) - \bar{\delta}(v)\|^2 = \sum_{\alpha=1}^{k} \#\mathcal{S}_\alpha\left(\frac{e(v,\mathcal{S}_\alpha) - \mu'(v,\mathcal{S}_\alpha)}{\#\mathcal{S}_\alpha}\right)^2. \tag{7.32}$$

https://doi.org/10.1017/S0963548309990514 Published online by Cambridge University Press

Further, condition **H1** ensures that $\#V \setminus H \leqslant 10^{-4} n_{\min}$. Moreover, since the partition $\mathcal{S}$ is feasible, we have $\mathcal{S}_\alpha \geqslant \frac{1}{2}\#V_\alpha$ for all α, and $\#V_\alpha \geqslant n_{\min} = \min_\beta \#V_\beta$. Hence, recalling that $\delta_H(v)$ (resp. $\bar{\delta}_H(v)$) is the restriction of δ (resp. $\bar{\delta}$) to H, we obtain from (7.32)

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

whence (7.26) 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 (7.26) to bound the norm of the latter vector. Suppose that $w \in \mathcal{S}_\alpha \cap H$. Then by definition the w-entry of $\delta_H(v)$ is $e(v,\mathcal{S}_\alpha)/\#\mathcal{S}_\alpha$, and the w-entry of $\Delta_H(v)$ is $e(v,\mathcal{S}_\alpha \cap H)/\#\mathcal{S}_\alpha$. Similarly, the w-entry of $\bar{\delta}_H(v)$ equals $\mu'(v,\mathcal{S}_\alpha)/\#\mathcal{S}_\alpha$, and that of $\bar{\Delta}_H(v)$ is $\mu'(v,\mathcal{S}_\alpha \cap H)/\#\mathcal{S}_\alpha$. Plugging these expressions in, we obtain for all $v \in H$

$$\begin{aligned}
&\|(\delta_H(v) - \bar{\delta}_H(v)) - (\Delta(v) - \bar{\Delta}(v))\|^2 \\
&\quad = \sum_{w \in H} (\delta_H(v,w) - \bar{\delta}_H(v,w) - (\Delta(v,w) - \bar{\Delta}(v,w)))^2 \\
&\quad = \sum_{\alpha=1}^{k} \#\mathcal{S}_\alpha \cap H \left(\frac{e(v,\mathcal{S}_\alpha) - e(v,\mathcal{S}_\alpha \cap H) - (\mu'(v,\mathcal{S}_\alpha) - \mu'(v,\mathcal{S}_\alpha \cap H))}{\mathcal{S}_\alpha} \right)^2 \\
&\quad = \sum_{\alpha=1}^{k} \#\mathcal{S}_\alpha \cap H \left(\frac{e(v,\mathcal{S}_\alpha \setminus H) - \mu'(v,\mathcal{S}_\alpha \setminus H)}{\mathcal{S}_\alpha} \right)^2 \\
&\quad \leqslant \sum_{\alpha=1}^{k} \#\mathcal{S}_\alpha^{-1} \left[e(v,\mathcal{S}_\alpha \setminus H) - \mu'(v,\mathcal{S}_\alpha \setminus H) \right]^2.
\end{aligned} \tag{7.33}$$

To continue we need to consider the 'dense' and the 'sparse' bits of the graph separately. Let $\Phi_1(v) = \{w \in V : p_{\psi(v)\psi(w)} \leqslant 1/2\}$ be the set of all vertices w such that the 'edge probability' $p_{\psi(v)\psi(w)}$ is at most $\frac{1}{2}$, and let $\Phi_2(v) = \{w \in V : p_{\psi(v)\psi(w)} > 1/2\}$. Then (7.33) yields

$$\begin{aligned}
\|(\delta_H(v) - \bar{\delta}_H(v)) - (\Delta(v) - \bar{\Delta}(v))\|^2 &\leqslant \sum_{\alpha=1}^{k} \#\mathcal{S}_\alpha^{-1} \left[e(v,\mathcal{S}_\alpha \setminus H) - \mu'(v,\mathcal{S}_\alpha \setminus H) \right]^2 \\
\leqslant 2 \sum_{\alpha=1}^{k} \#\mathcal{S}_\alpha^{-1} &\left[e(v,\Phi_1(v) \cap \mathcal{S}_\alpha \setminus H) - \mu'(v,\Phi_1(v) \cap \mathcal{S}_\alpha \setminus H) \right]^2 \\
&+ \#\mathcal{S}_\alpha^{-1} \left[e(v,\Phi_2(v) \cap \mathcal{S}_\alpha \setminus H) - \mu'(v,\Phi_2(v) \cap \mathcal{S}_\alpha \setminus H) \right]^2.
\end{aligned} \tag{7.34}$$

We recall the decomposition of the graph G into the 'sparse' part G_1 and the complement of the dense part G_2: G_1 contains all edges $e = \{v,w\}$ of G such that $w \in \Phi_1(v)$, and G_2 contains all edges $\{v,w\}$ that are *not* present in G such that $w \in \Phi_2(v)$. Condition **H4** states that in the union $G_1 \cup G_2$ each vertex $v \in H$ has at most 100 neighbours in $V \setminus H$.

Therefore, for all $v \in H$ we have

$$e(v, \Phi_1(v) \cap \mathcal{S}_\alpha \setminus H) \leqslant 100, \tag{7.35}$$
$$e(v, \Phi_2(v) \cap \mathcal{S}_\alpha \setminus H) \geqslant \#\Phi_2(v) \cap \mathcal{S}_\alpha \setminus H - 101. \tag{7.36}$$

Moreover,

$$\begin{aligned}\sum_{\alpha=1}^{k} \mu'(v, \Phi_1(v) \cap \mathcal{S}_\alpha \setminus H) &= \mu'(v, \Phi_1(v) \setminus H) \\ &= \sum_{w \in V \setminus H : \mathcal{E}_{vw} \leqslant \frac{1}{2}} \mathcal{E}_{vw} \leqslant \#V \setminus H \cdot \max_{w \in V : \mathcal{E}_{vw} \leqslant \frac{1}{2}} \mathcal{E}_{vw}. \end{aligned} \tag{7.37}$$

Since $\sigma^* = \max_{u \in V} \sum_{w \in V} \mathcal{E}_{uw}(1 - \mathcal{E}_{uw})$ and the terms $\mathcal{E}_{uw}(1 - \mathcal{E}_{uw})$ are constant on the partition classes $V_1, \ldots, V_k$, we have $\max_{w \in V : \mathcal{E}_{vw} \leqslant \frac{1}{2}} \mathcal{E}_{vw} \leqslant 2\sigma^*/\#V_\alpha$ for some $1 \leqslant \alpha \leqslant k$, and hence $\max_{w \in V : \mathcal{E}_{vw} \leqslant \frac{1}{2}} \mathcal{E}_{vw} \leqslant 2\sigma^*/n_{\min}$. Therefore, (7.37) yields

$$\sum_{\alpha=1}^{k} \mu'(v, \Phi_1(v) \cap \mathcal{S}_\alpha \setminus H) \leqslant \#V \setminus H \cdot \frac{2\sigma^*}{n_{\min}}.$$

As, furthermore, $\#V \setminus H \leqslant n_{\min}/\lambda^4$ by condition **H1** and $\lambda \geqslant \sigma^* \geqslant c_0$ for some large constant c_0, we conclude

$$\sum_{\alpha=1}^{k} \mu'(v, \Phi_1(v) \cap \mathcal{S}_\alpha \setminus H) \leqslant \frac{n_{\min}}{\lambda^4} \cdot \frac{2\sigma^*}{n_{\min}} \leqslant \frac{1}{2}, \tag{7.38}$$

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

$$\sum_{\alpha=1}^{k} \frac{\mu'(v, \Phi_1(v) \cap \mathcal{S}_\alpha \setminus H)^2}{\#\mathcal{S}_\alpha} \leqslant \frac{1}{2n_{\min}} \leqslant 10^{-4}\rho^2 \qquad (\text{as } \rho^2 > 10^4/n_{\min} \text{ by } \mathbf{A2}). \tag{7.39}$$

A similar argument shows that

$$\sum_{\alpha=1}^{k} \frac{(\#\Phi_2(v) \cap \mathcal{S}_\alpha \setminus H - \mu'(v, \Phi_2(v) \cap \mathcal{S}_\alpha \setminus H))^2}{\#\mathcal{S}_\alpha} \leqslant 10^{-4}\rho^2. \tag{7.40}$$

Plugging (7.35), (7.36), (7.39), and (7.40) into (7.34), we get

$$\begin{aligned}\|(\delta_H(v) - \bar{\delta}_H(v)) - (\Delta(v) - \bar{\Delta}(v))\|^2 &\leqslant 2 \cdot 10^{-4}\rho^2 + \sum_{\alpha=1}^{k} \frac{10^5}{\#\mathcal{S}_\alpha} \\ &\leqslant 4 \cdot 10^{-4}\rho^2 + \frac{10^6 k}{n_{\min}}. \end{aligned} \tag{7.41}$$

Since $\rho^2 = c_0^8 k^3 \lambda / n_{\min}$ by condition **A2**, the last expression in (7.41) is less than $10^{-3}\rho^2$. Therefore, (7.26) entails that $\|\Delta(v) - \bar{\Delta}(v)\|^2 \geqslant 0.1\|\mathcal{E}_a - \mathcal{E}_b\|^2$ for all $v \in \mathcal{R}_{ba} \cap H$. □

https://doi.org/10.1017/S0963548309990514 Published online by Cambridge University Press

Lemma 7.10. *Suppose that the partition $\mathcal{S}$ is feasible and tight, and that $\mathcal{R}$ is an improvement of $\mathcal{S}$. Then*

$$\sum_{a\neq b}\sum_{v\in\mathcal{R}_{ab}\cap H}\|\Delta(v)-\bar{\Delta}(v)\|^2 \leqslant 0.02\sum_{a\neq b}\rho^2\#\mathcal{R}_{ab}\cap H + 10^{-5}\rho^2\sum_{a\neq b}\#H\cap\mathcal{S}_{ab}.$$

Proof. Let M be the matrix defined in (7.4). That is, for any two vertices $v, w \in H$ the entry M_{vw} equals $\mathcal{E}_{vw}-1$ if v, w are adjacent, and $M_{vw} = \mathcal{E}_{vw}$ otherwise. Since $\mathcal{R}$ is an improvement of $\mathcal{S}$, for any vertex $v \in \mathcal{R}_a$ the vector $\delta_{\mathcal{S}}(v)$ satisfies $\|\delta_{\mathcal{S}}(v)-\xi_a\| = \min_b\|\delta_{\mathcal{S}}(v)-\xi_b\|$. Also remember that $\mathcal{R}_{ab} = \mathcal{R}_a \cap V_b$. For any $w \in \mathcal{S}_\alpha$ the w-entry of $\Delta(v)$ equals $e(v, \mathcal{S}_\alpha \cap H)/\#\mathcal{S}_\alpha$, and the w-entry of $\bar{\Delta}(v)$ equals $\mu'(v, \mathcal{S}_\alpha\cap H)/\#\mathcal{S}_\alpha$. Therefore, for all $v \in \mathcal{R}_{ab}\cap H$ such that $a \neq b$, we have

$$\begin{aligned}\|\Delta(v)-\bar{\Delta}(v)\|^2 &\leqslant 2\sum_{\alpha=1}^{k}\#\mathcal{S}_\alpha^{-1}\left[e(v,H\cap\mathcal{S}_\alpha)-\mu'(v,H\cap\mathcal{S}_\alpha)\right]^2\\ &\leqslant 6\cdot(S_1(v)+S_2(v)+S_3(v)),\quad\text{where}\end{aligned}\tag{7.42}$$

$$S_1(v) = \sum_{\alpha=1}^{k}\#\mathcal{S}_\alpha^{-1}\left[e(v,H\cap V_\alpha)-\mu'(v,H\cap V_\alpha)\right]^2,$$

$$S_2(v) = \sum_{\alpha=1}^{k}\#\mathcal{S}_\alpha^{-1}\left[e(v,H\cap\mathcal{S}_\alpha\setminus V_\alpha)-\mu'(v,H\cap\mathcal{S}_\alpha\setminus V_\alpha)\right]^2,$$

$$S_3(v) = \sum_{a=1}^{k}\#\mathcal{S}_a^{-1}\left[e(v,H\cap V_a\setminus\mathcal{S}_a)-\mu'(v,H\cap V_a\setminus\mathcal{S}_a)\right]^2.$$

We begin by bounding the last two summands. Since $\mathcal{S}$ is feasible, we have $\#\mathcal{S}_\alpha \geqslant \#V_\alpha \geqslant n_{\min}/2$ for all α. Therefore, the definition of the matrix M entails that (see (7.6))

$$\begin{aligned}\sum_{v\in H}S_3(v) &= \sum_{\alpha=1}^{k}\#\mathcal{S}_\alpha^{-1}\|M\mathbf{1}_{H\cap V_\alpha\setminus\mathcal{S}_\alpha}\|^2 \leqslant \frac{2\|M\|^2}{n_{\min}}\sum_\alpha\|\mathbf{1}_{H\cap V_\alpha\setminus\mathcal{S}_\alpha}\|^2\\ &= \frac{2\|M\|^2}{n_{\min}}\sum_\alpha\#H\cap V_\alpha\setminus\mathcal{S}_\alpha.\end{aligned}\tag{7.43}$$

Analogously,

$$\sum_{v\in H}S_2(v) \leqslant \frac{2\|M\|^2}{n_{\min}}\sum_{\alpha=1}^{k}\#H\cap\mathcal{S}_\alpha\setminus V_\alpha.\tag{7.44}$$

As $\mathcal{S}_{\alpha\beta} = \mathcal{S}_\alpha \cap V_\beta$, we obtain

$$\begin{aligned}\sum_{a\neq b}\sum_{v\in\mathcal{R}_{ab}\cap H}S_2(v)+S_3(v) \leqslant \sum_{v\in H}S_2(v)+S_3(v) &\leqslant \frac{2\|M\|^2}{n_{\min}}\sum_{\alpha=1}^{k}\#H\cap(V_\alpha\triangle\mathcal{S}_\alpha)\\ &\leqslant \frac{2\|M\|^2}{n_{\min}}\sum_{\alpha\neq\beta}\#H\cap\mathcal{S}_{\alpha\beta}.\end{aligned}\tag{7.45}$$

https://doi.org/10.1017/S0963548309990514 Published online by Cambridge University Press

Finally, as $\|M\|^2 \leqslant c_0^2 k\lambda$ by (7.5) and $\rho^2 = c_0^8 k^3\lambda/n_{\min}$ for some large constant $c_0 > 0$ by **A2**, (7.45) yields

$$\sum_{a\neq b}\sum_{v\in\mathcal{R}_{ab}\cap H} S_2(v)+S_3(v) \leqslant \frac{\rho^2}{10^6}\sum_{\alpha\neq\beta}\#H\cap\mathcal{S}_{\alpha\beta}. \tag{7.46}$$

To bound $S_1(v)$ for $v \in \mathcal{R}_{ab}\cap H = \mathcal{R}_a \cap V_b \cap H$ with $a \neq b$, we consider the expression

$$S_4(v) = \sum_{\alpha=1}^{k}\#\mathcal{S}_\alpha^{-1}(e(v,V_\alpha\setminus H)-\mu(v,V_\alpha\setminus H))^2 \leqslant \frac{8(S_{41}(v)^2+S_{42}(v)^2)}{n_{\min}}, \quad \text{where}$$

$$S_{41}(v) = \sum_{\alpha:p_{\alpha b}\leqslant\frac{1}{2}} e(v,V_\alpha\setminus H)+\mu(v,V_\alpha\setminus H),$$

$$S_{42}(v) = \sum_{\alpha:p_{\alpha b}>\frac{1}{2}} (V_\alpha\setminus H - e(v,V_\alpha\setminus H)) + (V_\alpha\setminus H-\mu(v,V_\alpha\setminus H)).$$

In order to estimate $S_{41}(v)$ and $S_{42}(v)$, we remember the decomposition of G into the graphs G_1, G_2: G_1 contains all edges $\{u,w\}$ of G such that $\mathcal{E}_{uw} \leqslant \frac{1}{2}$, and G_2 contains all edges $\{u,w\}$ that are *not* present in G such that $\mathcal{E}_{uw} > \frac{1}{2}$. Then condition **H4** implies that in the union $G_1 \cup G_2$ each vertex $v \in H$ has at most 100 neighbours in $V\setminus H$. Therefore,

$$\sum_{\alpha:p_{\alpha b}\leqslant\frac{1}{2}} e(v,V_\alpha\setminus H)\leqslant 100 \quad\text{and}\quad \sum_{\alpha:p_{\alpha b}>\frac{1}{2}}(V_\alpha\setminus H - e(v,V_\alpha\setminus H))\leqslant 101. \tag{7.47}$$

Since the vector $\mathcal{E}_v$ is constant on the classes $V_1,\dots,V_k$ and $\sigma^* = \max_u\sum_{w\in V}\mathcal{E}_{uw}(1-\mathcal{E}_{uw})$, we have $\max_{w:\mathcal{E}_{vw}\leqslant\frac{1}{2}}\mathcal{E}_{vw} \leqslant 2\sigma^*/n_{\min}$. Consequently,

$$\sum_{\alpha:p_{\alpha b}\leqslant\frac{1}{2}}\mu(v,V_\alpha\setminus H)\leqslant \#V\setminus H\cdot\max_{w:\mathcal{E}_{vw}\leqslant\frac{1}{2}}\mathcal{E}_{vw}\leqslant\#V\setminus H\cdot 2\sigma^*/n_{\min}. \tag{7.48}$$

Analogously, we obtain

$$\sum_{\alpha:p_{\alpha b}>\frac{1}{2}} V_\alpha\setminus H-\mu(v,V_\alpha\setminus H)\leqslant\#V\setminus H\cdot\max_{w:\mathcal{E}_{vw}>\frac{1}{2}}1-\mathcal{E}_{vw}\leqslant\#V\setminus H\cdot 2\sigma^*/n_{\min}. \tag{7.49}$$

Since $\#V\setminus H \leqslant n_{\min}\lambda^{-4}$ by condition **H1**, $\lambda \geqslant \sigma^*$ by condition **A1**, and $\rho^2 = c_0^8k^3\lambda/n_{\min}$ by assumption **A2**, (7.47)–(7.49) yield

$$S_{41}(v)+S_{42}(v)\leqslant 201+\frac{4\sigma^*}{n_{\min}}\cdot\#V\setminus H\leqslant 201+\frac{4\sigma^*}{\lambda^4}\leqslant 202.$$

Thus, we conclude that

$$S_4(v)\leqslant 10^6/n_{\min}\quad\text{for all}\quad v\in\mathcal{R}_{ab}\cap H. \tag{7.50}$$

Finally, to bound $S_1(v)$ consider the vector $d(v) = (d(v,w))_{w\in V}$ with entries $d(v,w) = e(v,V_\alpha)/\#V_\alpha$ for $w \in V_\alpha$. By condition **H2** we have $\|d(v)-\mathcal{E}_v\|^2 \leqslant 0.001\rho^2$ for all $v \in H$. Moreover, as $\mathcal{S}$ is feasible, we have $\#\mathcal{S}_\alpha \geqslant \frac{1}{2}\#V_\alpha \geqslant n_{\min}/2$. Therefore, by Cauchy–Schwarz, for all $v \in H$ we have

$$\begin{aligned}
S_1(v) &= \sum_{\alpha=1}^{k} \#\mathcal{S}_\alpha^{-1}\left[e(v, H \cap V_\alpha) - \mu'(v, H \cap V_\alpha)\right]^2 \\
&\leqslant \sum_{\alpha=1}^{k} \#\mathcal{S}_\alpha^{-1}(e(v, V_\alpha) - \mu(v, V_\alpha))^2 + \#\mathcal{S}_\alpha^{-1}(e(v, V_\alpha \setminus H) - \mu(v, V_\alpha) \setminus H)^2 \\
&\leqslant 2\sum_{\alpha=1}^{k} \#V_\alpha^{-1}(e(v, V_\alpha) - \mu(v, V_\alpha))^2 + S_4(v) \\
&= 2\sum_{\alpha=1}^{k} \#V_\alpha \left(\frac{e(v, V_\alpha) - \mu(v, V_\alpha)}{\#V_\alpha}\right)^2 + S_4(v) \\
&= 2\|d(v) - \mathcal{E}_v\|^2 + S_4(v) \leqslant 0.002\rho^2 + 10^6/n_{\min} \qquad \text{(by (7.50))}.
\end{aligned}$$

Since $\rho^2 = c_0^8 k^3 \lambda / n_{\min}$ for a large constant c_0, we conclude that $S_1(v) \leqslant 0.003\rho^2$ for all $v \in H$. In combination with (7.46) and (7.42), this implies the assertion. □

Proof of Lemma 7.3. Suppose that $\mathcal{S}$ is both feasible and tight, and that $\mathcal{R}$ is an improvement of $\mathcal{S}$. If $v \in \mathcal{R}_{ab} \cap H$ with $a \neq b$, then Lemma 7.9 implies that $\|\Delta(v) - \bar{\Delta}(v)\|^2 \geqslant 0.1\|\mathcal{E}^{V_a} - \mathcal{E}^{V_b}\|^2$. Further, $\|\mathcal{E}^{V_a} - \mathcal{E}^{V_b}\|^2 \geqslant \rho^2$ by condition **A2**. Therefore, we conclude that

$$\sum_{a \neq b} \sum_{v \in \mathcal{R}_{ab} \cap H} \|\Delta(v) - \bar{\Delta}(v)\|^2 \geqslant 0.1 \sum_{a \neq b} \#\mathcal{R}_{ab} \cap H \cdot \rho^2. \tag{7.51}$$

Combining Lemma 7.10 and (7.51), we obtain

$$0.1 \sum_{a \neq b} \#\mathcal{R}_{ab} \cap H \cdot \rho^2 \leqslant 0.02 \sum_{a \neq b} \rho^2 \#\mathcal{R}_{ab} \cap H + \frac{\rho^2}{10^5} \sum_{a \neq b} \#H \cap \mathcal{S}_{ab}.$$

Cancelling ρ^2, we obtain $\sum_{a \neq b} \#\mathcal{R}_{ab} \cap H \leqslant 10^{-3} \sum_{a \neq b} \#H \cap \mathcal{S}_{ab}$, as desired. □

8. The random graph $G_{n,k}(\psi, \mathbf{p})$

In this section we prove Theorem 1.1. We start with some preliminaries on random graphs in Section 8.1. Then, we discuss the construction of the core of $G_{n,k}(\psi, \mathbf{p})$ in Section 8.2. Finally, in Section 8.4 we investigate the components of $G_{n,k}(\psi, \mathbf{p}) - \mathrm{core}(G_{n,k}(\psi, \mathbf{p}))$. Throughout this section, we let ψ, $\mathbf{p}$, $\mathcal{E}$, $n_{\min}$, and let σ^* be as in Sections 1.2 and 1.3. Furthermore, we always assume that n is sufficiently large.

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

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

https://doi.org/10.1017/S0963548309990514 Published online by Cambridge University Press

Lemma 8.1. *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*

$$\max\{\mathrm{P}(X \leqslant \mathrm{E}(X) - t), \mathrm{P}(X \geqslant \mathrm{E}(X) + t)\} \leqslant \exp\left(-\sigma^2\phi\left(\frac{t}{\sigma^2}\right)\right) \leqslant \exp\left(-\frac{t^2}{2(\sigma^2 + t/3)}\right). \tag{8.1}$$

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

Lemma 8.2. *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 $|X(G') - X(G'')| \leqslant 1$. (8.2)

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

In Section 9.2 we shall use Lemma 8.2 to derive the following estimate on the upper tail of the degree distribution of $G_{n,k}(\psi, \mathbf{p})$.

Lemma 8.3. *Let $U_i = \#\{v \in V : \max_{j=1,2} d_{G_j}(v) \geqslant 2^{i+1}\sigma^*\}$. Then w.h.p. $\#U_i \leqslant \exp(-2^{i-2}\sigma^*)n$ 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 8.4. *With high probability $G = G_{n,k}(\psi, \mathbf{p})$ enjoys the following property:*

For all sets $T \subset V$ such that $\#T \leqslant n\left(\frac{n_{\min}}{n\sigma^*}\right)^2$ we have $e_{G_1 \cup G_2}(T) \leqslant 10\#T$. (8.3)

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

For all $T \subset V$ such that $\ln^3 n \leqslant \#T \leqslant n\left(\frac{n_{\min}}{n\sigma^*}\right)^2$ we have $e_{G_1 \cup G_2}(T) \leqslant 10\#T$. (8.4)

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

Lemma 8.5. *Let $\Delta > 0$ and $X = \{v \in V : \max_{i=1,2} d_{G_i}(v) \leqslant \Delta\}$. Then $\|A_X - \mathcal{E}_X\| \leqslant ck\sqrt{\sigma^* + \Delta}$.*

In Section 9.4 we indicate how Lemma 8.5 follows from spectral considerations of Alon and Kahale [1], Feige and Ofek [17], and Füredi and Komloś [20].

https://doi.org/10.1017/S0963548309990514 Published online by Cambridge University Press

8.2. The core

In this section our objective is to construct a subgraph core(G) of $G = G_{n,k}(\psi, \mathbf{p})$ such that for all vertices $v \in \text{core}(G)$ the numbers $e(v, V_i \cap \text{core}(G))$ 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 (1.6), which represents the actual numbers of 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 right-hand side of (1.5).)

Moreover, recall the decomposition of $G = G_{n,k}(\psi, \mathbf{p})$ into the 'sparse' part G_1 and the 'dense' part G_2 from Section 1.6. Then $\mathrm{E}(d_{G_1 \cup G_2}(v)) \leqslant 2\sigma^*$ 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' neighbours of v are 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 core(G) = H. In Section 8.3 we shall prove that w.h.p. core(G) constitutes a huge fraction of G.

Proposition 8.6. *Suppose that* (1.5) *holds. Then w.h.p.* core($G_{n,k}(\psi, \mathbf{p})$) *contains* $\geqslant n - n_{\min}\sigma^{*\,-10}$ *vertices. For all* $v \in \text{core}(G)$ *we have* $\|d(v) - \mathcal{E}_v\| \leqslant 0.01\rho$, $d_{G_1 \cup G_2}(v) \leqslant 10\sigma^*$, *and* $e_{G_1 \cup G_2}(v, G - H) \leqslant 100$.

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

Proposition 8.7. *If* (1.5) *holds, then w.h.p. all components of* $(G_1 \cup G_2) - \text{core}(G)$ *have size* $\leqslant \ln n$.

Proof of Theorem 1.1. Assuming that c_0 is a sufficiently large constant and letting $\lambda = \sigma^* > c_0$, we note that Lemma 8.5 implies that $G_{n,k}(\psi, \mathbf{p})$ satisfies **A1** w.h.p. Moreover, our assumption **R3** ensures that **A2** is true. Further, for each vertex $v \in V_j$ and each $1 \leqslant i \leqslant k$ the number $e(v, V_i)$ has a binomial distribution with variance $\#V_i P_{ij}(1 - p_{ij}) \leqslant \sigma^*$; therefore, the Chernoff bound (8.1) entails that

$$\mathrm{P}\left[|e(v, V_i) - \mu(v, V_i)| > \frac{\sigma^*}{10k} + \ln^2 n\right] \leqslant 2\exp\left[-\frac{\sigma^{*\,2}k^{-2} + \ln^4 n}{300(\sigma^* + \ln^2 n)}\right] \leqslant n^{-1}. \tag{8.5}$$

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

With respect to **H1**, letting $H = \text{core}(G)$ we observe that Proposition 8.6 entails that $\#V \setminus H \leqslant n_{\min}\lambda^{-4}$. Furthermore, let $U_i = \{v \in V : 2^{i+1}\sigma^* \leqslant \max_{j=1,2} d_{G_j}(v) \leqslant 2^{i+2}\sigma^*\}$.

https://doi.org/10.1017/S0963548309990514 Published online by Cambridge University Press

Then Lemma 8.3 and our assumption that $\sigma^* \geqslant c_0$ for a sufficiently large number c_0 entail that w.h.p.

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

whence **H1** follows. Moreover, **H2**, **H3**, and **H4** follow directly from Proposition 8.6. □

8.3. Proof of Proposition 8.6

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 $\omega \geqslant n/n_{\min} \geqslant k$.

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

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

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

Lemma 8.8. *We have* $\mathrm{core}(G) \supset K$.

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

$$\begin{aligned}\|d(v) - \mathcal{E}_v\|^2 &= \sum_{i=1}^k \#V_i\left(\frac{e(v,V_i)-\mu(v,V_i)}{\#V_i}\right)^2 = \sum_{i=1}^k \#V_i^{-1}(e(v,V_i)-\mu(v,V_i))^2 \\ &\leqslant 2\cdot 10^4\sum_{i=1}^k \#V_i^{-1}\left[\#V_i p_{ij}(1-p_{ij})\ln\omega + \ln^2\omega\right] \qquad \text{(due to } \mathbf{K1}\text{)} \\ &\leqslant 2\cdot 10^4\left[\sum_{i=1}^k p_{ij}(1-p_{ij})\ln\omega + \sum_{i=1}^k \frac{\ln^2\omega}{\#V_i}\right] \leqslant 10^{-4}\rho^2,\end{aligned}$$

where the last step follows from (1.5) and **R1**. Thus, none of the vertices $v \in K$ gets removed by **CR1**. Further, **K1** ensures that $d_{G_1\cup G_2}(v) \leqslant 10\sigma^*$ 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**.

https://doi.org/10.1017/S0963548309990514 Published online by Cambridge University Press

Lemma 8.9. *With high probability there are at most $n\omega^{-198}$ vertices v such that*

$$\max_{1\leqslant i\leqslant k} |e(v,V_i)-\mu(v,V_i)| \geqslant 10^3\left[\sqrt{\#V_I P_{ij}(1-p_{ij})\ln\omega}+\ln\omega\right].$$

Moreover, if $\omega\leqslant n^{1/190}$, then with probability $\geqslant 1-\exp(-\ln^3 n)$ there are at most $n\omega^{-90}$ such vertices.

Proof. By the Chernoff bound (8.1), for each vertex $v\in V_j$ we have

$$\begin{aligned}P_{ij} &= \mathrm{P}\left[|e(v,V_i)-\mu(v,V_i)|\geqslant 10^3\left(\sqrt{\#V_i p_{ij}(1-p_{ij})\ln\omega}+\ln\omega\right)\right]\\ &\leqslant 2\exp\left[-\frac{10^6(\#V_i p_{ij}(1-p_{ij})+\ln^2\omega)}{2(\#V_i p_{ij}(1-p_{ij})+10^3(\sqrt{\#V_i p_{ij}(1-p_{ij})\ln\omega}+\ln\omega)}\right]\\ &\leqslant 2\exp\left[-\frac{10^6\#V_i p_{ij}(1-p_{ij})\ln\omega+10^6\ln^2\omega}{5\cdot 10^3(\#V_i p_{ij}(1-p_{ij})+\ln\omega)}\right]\leqslant 2\omega^{-200}.\end{aligned}$$

Hence, letting

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

we have

$$\mathrm{E}(Z_{ij})\leqslant 2\#V_j\omega^{-200}. \tag{8.6}$$

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

Case 1: $\omega\geqslant\ln n$. Then Markov's inequality entails that w.h.p.

$$\sum_{i,j=1}^{k}\#Z_{ij}\leqslant nk\omega^{-199}\leqslant n\omega^{-198}.$$

Case 2: $\omega<\ln n$. 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 (8.2). Further, $\sigma^*\leqslant\omega\leqslant\ln n$, and $\#V_j\geqslant n_{\min}\geqslant n/\omega>n/\ln n$. Hence, Lemma 8.2 entails that

$$\mathrm{P}\left[Z_{ij}\geqslant\#V_j\omega^{-199}\right]\overset{(8.6)}{\leqslant}\mathrm{P}\left[Z_{ij}-\mathrm{E}(Z_{ij})\geqslant\sqrt{\sigma^* n}\ln^2 n\right]=o(1),$$

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

Now, assume that $\omega\leqslant n^{1/190}$. Then the inequalities $\omega\geqslant\sigma^*$ and $\omega\geqslant n/n_{\min}$ imply that $\sqrt{n\sigma^*}\ln^2 n\leqslant\sqrt{n\omega}\ln^2 n\leqslant n^{96/190}$, while $n\omega^{-92}\geqslant n^{98/190}$. Therefore, Lemma 8.2 entails

$$\mathrm{P}\left[Z_{ij}\geqslant n\omega^{-92}\right]\overset{(8.6)}{\leqslant}\mathrm{P}\left[Z_{ij}-\mathrm{E}(Z_{ij})\geqslant\sqrt{\sigma^* n}\ln^2 n\right]\leqslant\exp(-\ln^4 n).$$

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

Lemma 8.9 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**.

https://doi.org/10.1017/S0963548309990514 Published online by Cambridge University Press

Lemma 8.10. *With high probability* **K2** *removes at most* $n\omega^{-198}$ *vertices.*

Proof. Let S be the set of vertices removed by **K1**. By Lemma 8.9 we may assume that $s = \#S \leqslant n\omega^{-198}$. Moreover, let $v_1, \ldots, v_q$ be the vertices removed by **K2** (in this order). Assume that $q \geqslant s$, and let $T = S \cup \{v_1, \ldots, v_s\}$. We shall prove that T violates (8.3), so Lemma 8.4 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 (8.3), observe that by construction each v_i satisfies $e_{G_1 \cup G_2}(v_i, S \cup \{v_1, \ldots, v_{i-1}\}) \geqslant 50$. Therefore, $e_{G_1 \cup G_2}(T) \geqslant 50s \geqslant 25\#T$, while $\#T = 2s \leqslant n\omega^{-197}$. □

Combining Lemmas 8.8–8.10, we obtain the following corollary, which implies Proposition 8.6.

Corollary 8.11. *With high probability we have* $\#V(K) \geqslant n(1 - \omega^{-197})$.

8.4. Proof of Proposition 8.7

If $\omega = \sigma^* + \frac{n}{n_{\min}} \geqslant n^{1/190}$, then Lemma 8.8 and Corollary 8.11 yield that $\mathrm{core}(G) = G$ w.h.p., and thus there is nothing to prove. Hence, we shall assume 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 8.8, 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 $I_T = \{v \in V_T : d_T(v) \leqslant 4\}$ and $J_T = V_T \setminus I_T$; as $\#E_T = t - 1$, we have $\#I_T \geqslant t/2$. Moreover, let K_T be the outcome of the following modification of the process **K1**–**K2** (see 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

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

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 8.12. *Let* K *be the result of the process* **K1**–**K2** (*see 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 $\leqslant 4$ 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′**. □

https://doi.org/10.1017/S0963548309990514 Published online by Cambridge University Press

Let us call G *good* if, for all trees T as above, we have $\#V(G - K_T) \leqslant n\omega^{-88}$, regardless of the outcome of step **K0′**.

Lemma 8.13. *We have* $\mathrm{P}[G \text{ is good}] \geqslant 1 - 2\exp(-\ln^3 n)$.

Proof. Let S be the set of vertices removed by **K1′**, and let $s = \#S$. Since $\omega \leqslant n^{1/190}$, Lemma 8.9 entails that with probability $\geqslant 1 - \exp(-\ln^3 n)$ we have $s \leqslant \#J_T + n\omega^{-90} \leqslant n\omega^{-89}$. Furthermore, if **K2′** removes $q \geqslant n\omega^{-89}$ vertices $v_1, \ldots, v_q$, then consider the set $T = S \cup \{v_1, \ldots, v_{\lceil n\omega^{-89}\rceil}\}$. Then $\ln^3 n \leqslant n\omega^{-89} \leqslant \#T \leqslant s + n\omega^{-89} + 1 \leqslant n\omega^{-88}$, but $e_{G_1 \cup G_2}(T) \geqslant 40\#T/2 = 20\#T$ (see the proof of Lemma 8.10). Hence, T violates (8.4). Consequently, Lemma 8.4 entails that $q \leqslant n\omega^{-89}$ with probability $\geqslant 1 - \exp(-\ln^3 n)$, whence the assertion follows. □

Proof of Proposition 8.7. 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 8.12 yields

$$\mathrm{P}\left[T \subset G_1 \cup G_2 \wedge V_T \cap K = \emptyset\right] \leqslant \mathrm{P}\left[T \subset G_1 \cup G_2\right] \cdot \mathrm{P}\left[I_T \cap K_T = \emptyset\right]. \tag{8.7}$$

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

$$\begin{aligned}
\mathrm{P}\left[I_T \cap K_T = \emptyset\right] &\leqslant \mathrm{P}\left[G \text{ is not good}\right] + \prod_{i=1}^{k} \frac{\binom{\#V_i - t_i}{\nu - t_i}}{\binom{\#V_i}{\nu}} \\
&\overset{\text{Lem. 8.13}}{\leqslant} \exp(-\ln^3 n) + \prod_{i=1}^{k} \frac{(\#V_i - t_i)_{\nu - t_i}(\nu)_{t_i}}{(\#V_i)_{\nu - t_i}(\#V_i - \nu + t_i)_{t_i}} \\
&\leqslant \exp(-\ln^3 n) + \prod_{i=1}^{k} \left(\frac{\nu}{\#V_i - \nu}\right)^{t_i} \\
&\leqslant \exp(-\ln^3 n) + \prod_{i=1}^{k} \left(\frac{2\nu}{\#V_i}\right)^{t_i} \\
&\leqslant \exp(-\ln^3 n) + \omega^{-86\sum_{i=1}^{k} t_i} \\
&\leqslant \exp(-\ln^3 n) + \omega^{-43t} \leqslant \omega^{-42t}.
\end{aligned} \tag{8.8}$$

To bound $\mathrm{P}\left[T \subset G_1 \cup G_2\right]$, we note that

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

by the definition of σ^* ($v, w \in V$). Consequently,

$$\mathrm{P}\left[T \subset G_1 \cup G_2\right] \leqslant \left(\frac{2\sigma^*}{n_{\min}}\right)^{t-1}. \tag{8.9}$$

Combining (8.7), (8.8), and (8.9), and recalling that $\omega = \sigma^* + \frac{n}{n_{\min}}$, we conclude

$$\mathrm{P}\left[T \subset G_1 \cup G_2 \wedge V_T \cap K = \emptyset\right] \leqslant \left(\frac{2\sigma^*}{n_{\min}}\right)^{t-1} \omega^{-42t} \leqslant n^{1-t}\omega^{-39t}. \tag{8.10}$$

https://doi.org/10.1017/S0963548309990514 Published online by Cambridge University Press

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, (8.10) entails that

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

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

9. Proofs of auxiliary lemmas

9.1. Proof of Lemma 8.2

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

Lemma 9.1. *Let $\Omega = \prod_{i=1}^{N} \Omega_i$ be a product of probability spaces $\Omega_1, \ldots, \Omega_N$. Let $Y : \Omega \to \mathbb{R}$ be a random variable that satisfies the following condition for all $1 \leqslant j \leqslant N$.*

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

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

To derive Lemma 8.2 from Lemma 9.1, 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 $G_{n,k}(\psi, \mathbf{p}) = \prod_{e \in \mathcal{P}} \Omega_e$, because the edges occur independently in $G_{n,k}(\psi, \mathbf{p})$. However, we cannot apply Lemma 9.1 to this decomposition directly, because the number of factors is too large. Therefore, we are going to set up a different product decomposition $G_{n,k}(\psi, \mathbf{p}) = \prod_{i=1}^{K} \Omega_i$, where each Ω_i is a product of several Ω_e.

To this end, we partition $\mathcal{P}$ into $K \leqslant 2\sigma^* n / \ln n$ subsets $\mathcal{P}_1, \ldots, \mathcal{P}_K$ such that

$$\mathrm{E}(\#E(G_1 \cup G_2) \cap \mathcal{P}_i) = \sum_{e \in \mathcal{P}_i} \mathrm{P}\left[e \in G_1 \cup G_2\right] \leqslant \ln n \quad \text{for all } \ 1 \leqslant i \leqslant K,$$

where G_1, G_2 are the graphs defined in (1.3), (1.4). Then we have the decomposition

$$G_{n,k}(\psi, \mathbf{p}) = \prod_{i=1}^{K} \Omega_i, \quad \text{where } \ \Omega_i = \prod_{e \in \mathcal{P}_i} \Omega_e. \tag{9.1}$$

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 (8.1) entails that $\mathrm{P}[\mathcal{P}_i \text{ is critical}] \leqslant n^{-21}$. Therefore, by the union bound

$$\mathrm{P}\left[\exists i : \mathcal{P}_i \text{ is critical}\right] \leqslant n^{-19}. \tag{9.2}$$

Now, for $G = G_{n,k}(\psi, \mathbf{p})$ we define

$$\tilde{G} = G - \bigcup_{i : \mathcal{P}_i \text{is critical}} E(G_1) \cap \mathcal{P}_i + \bigcup_{i : \mathcal{P}_i \text{is critical}} E(G_2) \cap \mathcal{P}_i$$

https://doi.org/10.1017/S0963548309990514 Published online by Cambridge University Press

and set $Y(G) = X(\tilde{G})$. Then (9.2) yields

$$\mathrm{P}\big[X(G_{n,k}(\psi,\mathbf{p})) = Y(G_{n,k}(\psi,\mathbf{p}))\big] \geqslant 1 - n^{-19}. \tag{9.3}$$

Furthermore, by the Lipschitz condition (8.2) we have $|X(G) - Y(G)| \leqslant n^2$ for *all* possible outcomes $G = G_{n,k}(\psi,\mathbf{p})$. Therefore, (9.3) entails that

$$|\mathrm{E}\big(X(G_{n,k}(\psi,\mathbf{p}))\big) - \mathrm{E}\big(Y(G_{n,k}(\psi,\mathbf{p}))\big)| \leqslant n^{2-19} \leqslant 1. \tag{9.4}$$

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

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 $|Y(G) - Y(G')| \leqslant 200 \ln n$ (9.5)

To prove (9.5), 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 (1.3), (1.4).

Case 1: the set $\mathcal{P}_j$ is critical in neither G nor G'. In this case $\tilde{G}'$ 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 either G and G', we have $\#\mathcal{P}_j \cap (E(G) \triangle E(G')) \leqslant 200 \ln n$, so that (9.5) follows from the Lipschitz condition (8.2).

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

Case 3: $\mathcal{P}_j$ 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 $\#\mathcal{P}_j \cap E(G') \leqslant 100 \ln n$, the Lipschitz condition (8.2) implies (9.5).

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

Due to (9.5), Lemma 9.1 applied to $Y(G_{n,k}(\psi,\mathbf{p}))$ and the decomposition (9.1) yields

$$\begin{aligned}\mathrm{P}\left[|Y(G_{n,k}(\psi,\mathbf{p})) - \mathrm{E}(Y(G_{n,k}(\psi,\mathbf{p})))| > \frac{1}{2}\sqrt{\sigma^* n}\ln^2 n\right]& \\ \leqslant \exp\left[-\frac{\sigma^* \ln^4 n}{160000K \ln^2 n}\right] \leqslant n^{-11},&\end{aligned} \tag{9.6}$$

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

$$\begin{aligned}&\mathrm{P}\big[|X(G_{n,k}(\psi,\mathbf{p})) - \mathrm{E}(X(G_{n,k}(\psi,\mathbf{p})))| \geqslant \sqrt{\sigma^* n}\ln^2 n\big] \\ &\quad\leqslant \mathrm{P}\big[X(G_{n,k}(\psi,\mathbf{p})) \neq Y(G_{n,k}(\psi,\mathbf{p}))\big] \\ &\qquad + \mathrm{P}\big[|Y(G_{n,k}(\psi,\mathbf{p})) - \mathrm{E}(X(G_{n,k}(\psi,\mathbf{p})))| \geqslant \sqrt{\sigma^* n}\ln^2 n\big] \\ &\quad\overset{(9.3),(9.4)}{\leqslant} n^{-19} + \mathrm{P}\left[|Y(G_{n,k}(\psi,\mathbf{p})) - \mathrm{E}(Y(G_{n,k}(\psi,\mathbf{p})))| \geqslant \frac{1}{2}\sqrt{\sigma^* n}\ln^2 n\right] \\ &\quad\overset{(9.6)}{\leqslant} n^{-19} + n^{-11} \leqslant n^{-10},\end{aligned}$$

as desired.

9.2. Proof of Lemma 8.3

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 $\leqslant 2\sigma^*$, the Chernoff bound (8.1) entails that $\mathrm{P}\big[v \in U_i\big] \leqslant$

https://doi.org/10.1017/S0963548309990514 Published online by Cambridge University Press

$\exp\left[-\frac{1}{3}2^i\sigma^*\right]$. Hence, $\mathrm{E}(\#U_i) \leqslant \exp\left[-\frac{1}{3}2^i\sigma^*\right]n$. To obtain a bound on $\#U_i$ that actually holds w.h.p., we consider two cases.

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

$$\mathrm{P}\left[\#U_i > \exp\left(-2^{i-2}\sigma^*\right)n\right] \leqslant \frac{\mathrm{E}(\#U_i)}{\exp[-2^{i-2}\sigma^*]n} \leqslant \exp\left[-2^i\sigma^*/12\right] \leqslant \ln^{-2} n. \tag{9.7}$$

Case 2: $2^i\sigma^* < 24\ln\ln n$. Then $\exp\left[-\frac{1}{3}2^i\sigma^*\right]n \geqslant n^{1-o(1)}$. Therefore, by Lemma 8.2

$$\begin{aligned}\mathrm{P}\left[\#U_i > 2\exp\left(-\frac{1}{3}2^i\sigma^*\right)n\right] &\leqslant \mathrm{P}\left[\#U_i - \mathrm{E}(\#U_i) \geqslant \sqrt{\sigma^* n}\ln^2 n\right] \\ &\leqslant n^{-10}. \end{aligned}\tag{9.8}$$

Finally, combining (9.7) and (9.8) and invoking the union bound, we conclude that with probability $\geqslant 1 - O(\ln^{-1} n)$ we have $\#U_i \leqslant \exp(-2^{i-2}\sigma^*)n$ for all $i = 1, \ldots, \lceil\log_2 n\rceil$.

9.3. Proof of Lemma 8.4

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

$$\mathrm{P}\left[\{v,w\} \in E(G_1 \cup G_2)\right] \leqslant 2p_{\psi(v)\psi(w)}(1 - p_{\psi(v)\psi(w)}) \leqslant \frac{2\sigma^*}{n_{\min}}. \tag{9.9}$$

Let $S \subset V$ be a set of cardinality $s = \#S \leqslant s_{\max} = n\left(\frac{n_{\min}}{n\sigma^*}\right)^2$. 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 (9.9) that

$$\mathrm{P}\left[e_{G_1\cup G_2}(S) \geqslant 10s\right] \leqslant \binom{\binom{s}{2}}{10s}\left(\frac{2\sigma^*}{n_{\min}}\right)^{10s} \leqslant \left(\frac{es\sigma^*}{10n_{\min}}\right)^{10s}.$$

Hence, once more due to the union bound we obtain that

$$P_s = \mathrm{P}\left[\exists S \subset V : \#S \leqslant s_{\max} \wedge e_{G_1\cup G_2}(S) \geqslant 10\#S\right] \leqslant \binom{n}{s}\left(\frac{es\sigma^*}{10n_{\min}}\right)^{10s}.$$

Consequently, we can estimate P_s as follows:

$$\binom{n}{s}P_s \leqslant \left[\left(\frac{en}{s}\right)^2\left(\frac{es\sigma^*}{10n_{\min}}\right)^{10}\right]^s \leqslant \left(\frac{n_{\min}}{n\sigma^*}\right)^{8s} \leqslant 1. \tag{9.10}$$

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

$$\mathrm{P}\left[\exists S \subset V : s_{\min} \leqslant \#S \leqslant s_{\max} \wedge e_{G_1\cup G_2}(S) \geqslant 10\#S\right] \leqslant \sum_{s=s_{\min}}^{s_{\max}} P_s \overset{(9.10)}{\leqslant} 2\binom{n}{s_{\min}}^{-1}. \tag{9.11}$$

Finally, (9.11) entails that w.h.p. there is no set $S \subset V$ of cardinality $1 \leqslant \#S \leqslant s_{\max}$ such that $e_{G_1\cup G_2}(S) \geqslant 10\#S$, whence the first part of Lemma 8.4 follows. Furthermore, setting $s_{\min} = \lceil\ln^3 n\rceil$ in (9.11), we obtain the second assertion.

9.4. Proof of Lemma 8.5

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].

https://doi.org/10.1017/S0963548309990514 Published online by Cambridge University Press

Lemma 9.2. *There are constants c_1, c_2 such that the following holds. Let $(a_{ij})_{1 \leqslant i<j \leqslant v}$ be a family of mutually independent Bernoulli random variables with mean $0 \leqslant p \leqslant 1$. Set $a_{ij} = a_{ji}$ for $1 \leqslant j < i \leqslant v$, and let $a_{ii} = 0$ for all $1 \leqslant i \leqslant v$. Moreover, let $A = (a_{ij})_{1 \leqslant i,j \leqslant v}$ and $M = p\vec{J} - A$. Further, let $d \geqslant 0$, and set $X = \{i \in \{1, \ldots, v\} : \sum_{j=1}^{v} a_{ij} \leqslant d\}$. Then, with probability $\geqslant 1 - O(v^{-1})$ we have $\|M_X\| \leqslant c_2\sqrt{\max\{vp, d\}}$.*

Lemma 9.3. *There are constants c_1, c_2 such that the following holds. Let $(a_{ij})_{1 \leqslant i,j \leqslant v}$ be a family of mutually independent Bernoulli random variables with mean $0 \leqslant p \leqslant 1$. Moreover, let $A = (a_{ij})_{1 \leqslant i,j \leqslant v}$ and $M = p\vec{J} - A$. Further, let $d \geqslant 0$, and set $X = \{i \in \{1, \ldots, v\} : \sum_{j=1}^{v} a_{ij} + a_{ji} \leqslant d\}$. Then, with probability $\geqslant 1 - O(v^{-1})$ we have $\|M_X\| \leqslant c_2\sqrt{\max\{vp, d\}}$.*

Proof of Lemma 8.5. Let $A = A(G)$ be the adjacency matrix, and set $M^{(i,j)} = p_{ij}\vec{J}_{V_i \times V_j} - A_{V_i \times V_j}$. Then, by Lemmas 9.2 and 9.3 (applied to the matrices $A_{V_i \times V_j}$), for all i, j such that $p_{ij} \leqslant \frac{1}{2}$ w.h.p. we have

$$\|M_X^{(i,j)}\| \leqslant c\sqrt{\max\{\Delta, \sigma^*\}} \tag{9.12}$$

for a certain constant $c > 0$. Furthermore, applying Lemmas 9.2 and 9.3 to $\vec{J}_{V_i \times V_j} - A_{V_i \times V_j}$, we conclude that w.h.p. (9.12) holds for all i, j such that $p_{ij} > \frac{1}{2}$ as well.

To bound $\|M_X\|$, let $\xi, \eta \in \mathbb{R}^V$ 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

$$|\langle M_X\eta, \xi\rangle| = \left|\sum_{i,j=1}^{k} \langle M_X^{(i,j)}\eta_j, \xi_i\rangle\right| \leqslant \sum_{i,j=1}^{k} \|M_X^{(i,j)}\| \cdot \|\xi_i\| \cdot \|\eta_j\| \overset{(9.12)}{\leqslant} ck\sqrt{\max\{\Delta, \sigma^*\}},$$

because $\sum_{i=1}^{k} \|\xi\|^2 = \sum_{i=1}^{k} \|\eta_i\|^2 = 1$. Thus, w.h.p. we have

$$\|M_X\| = \sup_{\xi,\eta : \|\xi\| = \|\eta\| = 1} |\langle M_X\eta, \xi\rangle| \leqslant ck\sqrt{\max\{\Delta, \sigma^*\}},$$

as desired. □

References

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

[2] Alon, N., Krivelevich, M. and Sudakov, B. (1998) Finding a large hidden clique in a random graph. *Random Struct. Alg.* **13** 457–466.

[3] Azar, Y., Fiat, A., Karlin, A. R., McSherry, F. and Saia, J. (2001) Spectral analysis of data. In *Proc. 33rd STOC*, pp. 619–626.

[4] Bilu, Y. and Linial, N. (2004) Are stable instances easy? Preprint.

[5] Bollobás, B. and Scott, A. D. (2004) Max cut for random graphs with a planted partition. *Combin. Probab. Comput.* **13** 451–474.

[6] Boppana, R. (1987) Eigenvalues and graph bisection: An average-case analysis. In *Proc. 28th FOCS*, pp. 280–285.

[7] Bui, T., Chaudhuri, S., Leighton, T. and Sipser, M. (1987) Graph bisection algorithms with good average case behavior. *Combinatorica* **7** 171–191.
[8] Chen, H. and Frieze, A. (1996) Coloring bipartite hypergraphs. In *Proc. 5th IPCO*, pp. 345–358.
[9] Chung, F. and Graham, R. (2002) Sparse quasi-random graphs. *Combinatorica* **22** 217–244.
[10] Coja-Oghlan, A. (2005) A spectral heuristic for bisecting random graphs. In *Proc. 16th SODA*, pp. 850–859.
[11] Coja-Oghlan, A. (2005) Spectral techniques, semidefinite programs, and random graphs. Habilitation thesis, Humboldt University of Berlin.
[12] Dasgupta, A., Hopcroft, J. E. and McSherry, F. (2004) Spectral analysis of random graphs with skewed degree distributions. In *Proc. 45th FOCS*, pp. 529–537.
[13] Dasgupta, A., Hopcroft, J. E., Kannan, R. and Mitra, P. (2006) Spectral partitioning by recursive partitioning. In *Proc. 14th ESA*, pp. 256–267.
[14] Ding, C. H. Q. (2002) Analysis of gene expression profiles: Class discovery and leaf ordering. In *Proc. 6th International Conference on Computational Biology*, pp. 127–136
[15] Dyer, M. and Frieze, A. (1989) The solution of some NP-hard problems in polynomial expected time. *J. Algorithms* **10** 451–489.
[16] Feige, U. and Kilian, J. (2001) Heuristics for semirandom graph problems. *J. Comput. Syst. Sci.* **63** 639–671.
[17] Feige, U. and Ofek, E. (2005) Spectral techniques applied to sparse random graphs. *Random Struct. Alg.* **27** 251–275.
[18] Flaxman, A. (2003) A spectral technique for random satisfiable 3CNF formulas. In *Proc. 14th SODA*, pp. 357–363.
[19] Frieze, A. and Kannan, R. (1999) Quick approximation to matrices and applications. *Combinatorica* **19** 175–220.
[20] Füredi, Z. and Komloś, J. (1981) The eigenvalues of random symmetric matrices. *Combinatorica* **1** 233–241.
[21] Goerdt, A. and Lanka, A. (2004) On the hardness and easiness of random 4-SAT formulas. In *Proc. 15th ISAAC*, pp. 470–483.
[22] Janson, S., Łuczak, T. and Ruciński, A. (2000) *Random Graphs*, Wiley.
[23] Jerrum, M. and Sorkin, G. (1998) The Metropolis algorithm for graph bisection. *Discrete Appl. Math.* **82** 155–175.
[24] Friedman, J., Kahn, J. and Szemerédi, E. (1989) On the second eigenvalue in random regular graphs. In *Proc. 21st STOC*, pp. 587–598.
[25] Kannan, R., Vempala, S. and Vetta, A. (2004) On clusterings: Good, bad and spectral. *J. Assoc. Comput. Mach.* **51** 497–515.
[26] Krivelevich, M. and Sudakov, B. (2003) The largest eigenvalue of sparse random graphs. *Combin. Probab. Comput.* **12** 61–72.
[27] Krivelevich, M. and Sudakov, B. (2006) Pseudo-random graphs. In *More Sets, Graphs and Numbers* (E. Gyori, G. O. H. Katona and L. Lovász, eds), Vol. 15 of *Bolyai Society Mathematical Studies*, pp. 199–262.
[28] McSherry, F. (2001) Spectral partitioning of random graphs. In *Proc. 42nd FOCS*, pp. 529–537.
[29] Pothen, A., Simon, H. D. and Kang-Pu, L. (1990) Partitioning sparse matrices with eigenvectors of graphs. *SIAM J. Matrix Anal. Appl.* **11** 430–452.
[30] Schloegel, K., Karypis, G. and Kumar, V. (2000) Graph partitioning for high performance scientific simulations. In *CRPC Parallel Computation Handbook* (J. Dongarra, I. Foster, G. Fox, K. Kennedy and A. White, eds), Morgan Kaufmann.
[31] Spielman, D. A. and Teng, S.-H. (1996) Spectral partitioning works: Planar graphs and finite element meshes. In *Proc. 36th FOCS*, pp. 96–105.
[32] Subramanian, C. R. and Veni Madhavan, C. E. (2002) General partitioning on random graphs. *J. Algorithms* **42** 153–172.