A SPECTRAL METHOD FOR COMMUNITY DETECTION IN MODERATELY SPARSE DEGREE-CORRECTED STOCHASTIC BLOCK MODELS

LENNART GULIKERS,* ** *Microsoft Research - INRIA Joint Centre and École Normale Supérieure*

MARC LELARGE,*** *INRIA and École Normale Supérieure*

LAURENT MASSOULIÉ,* **** *Microsoft Research - INRIA Joint Centre*

Abstract

We consider community detection in degree-corrected stochastic block models. We propose a spectral clustering algorithm based on a suitably normalized adjacency matrix. We show that this algorithm consistently recovers the block membership of all but a vanishing fraction of nodes, in the regime where the lowest degree is of order $\log(n)$ or higher. Recovery succeeds even for very heterogeneous degree distributions. The algorithm does not rely on parameters as input. In particular, it does not need to know the number of communities.

Keywords: Social and information network; random graph; degree-corrected stochastic block model; spectral algorithm; machine learning

2010 Mathematics Subject Classification: Primary 91D30

Secondary 05C80; 60B20; 05C80; 68W40; 91C20

1. Introduction

Social and information networks are omnipresent in our daily lives and have been the interest of much recent research activity [25]. Studies have been focusing on local properties of network systems as well as their large-scale properties. Among those large-scale phenomena, community structure has received a lot of attention. A wide variety of networks are found to have communities or blocks: groups of vertices with many links between themselves and substantially fewer to the rest of the network, or vice versa. One of the fundamental problems in network inference considers the detection of such communities; see [11] and [24] and the references therein for an overview.

In this paper we consider an instance of a certain probabilistic model that might be a fit on the observed data. One of the best known such models is the stochastic block model (SBM) [13]. In its simplest form, each of n vertices belongs to precisely one of K communities. Edges are independently drawn between different nodes with probabilities depending only on the block memberships of the involved vertices. This model is able to generate a diverse collection of random graphs, while remaining analytically tractable.

In practice however, the SBM fails to accurately describe the observed data: due to the stochastic inidentifiability of nodes in the same community, it does not allow for degree heterogeneity within blocks. The degree-correlated SBM (DC-SBM) was proposed in [16] to overcome this issue. The DC-SBM allows, additional to a block structure, the fitting of arbitrary degree sequences such that the expected degree of a vertex is independent of its community. In this paper we deal with community detection in this model.

Several methods for community detection can be found in the literature. They include, but are not limited to, modularity maximization [26], belief propagation [9], and spectral clustering. For the latter, see, for instance, [30, Section 4].

Spectral algorithms employ eigenvectors of matrices representing network data to return nonlocal information of the network. The most commonly used matrices are the adjacency matrix and the (un)normalized graph laplacian [30]. In [32], the authors studied the spectra of the adjacency matrix for networks possessing arbitrary degree distributions while simultaneously exhibiting a community structure. They demonstrate that those spectra consist, in general, of two components: a part containing the bulk of eigenvalues and a separated part with outliers whose number is, in general, equal to the number of blocks present.

The contribution of our paper is as follows. We demonstrate with a clean analysis that community detection in a moderately sparse DC-SBM is feasible under rather general conditions on the degree sequence.

More specifically, we consider the matrix $\widehat{H}$ with entry (u, v) given by $\widehat{H}_{uv} = (1/\widehat{D}_u \widehat{D}_v) A_{uv}$ if $A_{uv} = 1$ and $\widehat{H}_{uv} = 0$ otherwise (here A is the adjacency matrix of the graph and $\widehat{D}_u$ is the observed degree of vertex u), which we shall call the *normalized adjacency matrix*.

We show that this matrix concentrates around a deterministic matrix P of rank $L \leq K$ when the minimum expected degree is as small as $\log(n)$. To establish this concentration result, we use Lemma 11 below, which could be of independent interest, as a simple alternative to the commonly used Davis–Kahan theorem.

Due to the underlying community structure, the matrix that has the first L eigenvectors of P as its columns has the favourable property that it has only K different rows. Hence, due to this fact and the concentration of $\widehat{H}$ around P, the rows of the corresponding eigenvector matrix of $\widehat{H}$ considered as points in an L-dimensional Euclidean space, must cluster around K centres. This property indicates that $\widehat{H}$ is the correct matrix to analyse when dealing with the DC-SBM. Indeed, associating each vertex with its corresponding row, we show in this paper that we retrieve the correct community of all but a vanishing fraction of nodes.

Further, we point out a natural connection between $\widehat{H}$ and a random walk on the observed graph.

The paper is organized as follows. First, we formally introduce the DC-SBM together with the necessary notation. Next, we state our main result for community detection in this model, followed by a discussion in Section 4: a discussion of relevant literature, performance on real data, the conditions in the main theorem, and a connection between $\widehat{H}$ and random walks. Section 5 outlines the approach we take to prove the main theorem, which is accompanied by a statement of all auxiliary lemmas. Section 6 contains algebraic preliminaries. All proofs are deferred to Section 7. Finally, in Section 8 we give a suggestion for future research.

2. Model and notation

The DC-SBM denoted by $\mathcal{G}(B, K, \{\sigma_u\}_{u=1}^n, \{D_u\}_{u=1}^n)$ is a generalization of the Erdős–Rényi classical model of random graphs. We introduce a random graph on $V := \{1, \ldots, n\}$.

We partition the set of vertices into K communities of $\alpha_k n$ members each: each vertex u is given a label $\sigma_u \in S := \{1, \ldots, K\}$. A weight D_u is given to each vertex u to encode its expected degree. Without loss of generality, we assume that $D_1 \leq D_2 \leq \cdots \leq D_n$. All weights and labels will depend on n, but this is suppressed in the notation here. For each pair (u, v), we include the edge (u, v) with probability

$$\mathbb{P}(u \sim v) = \begin{cases} \dfrac{D_u D_v}{n \overline{D}} B_{\sigma_u \sigma_v} & \text{if } u \neq v, \\ 0 & \text{if } u = v, \end{cases} \tag{1}$$

where $B \in (\mathbb{R}^+)^{K \times K}$ is a symmetric matrix, independent of n and $\overline{D} = 1/n \sum_{l=1}^n D_l$, the average weight. Matrix B may be chosen *completely* independent of the weights $\{D_u\}_{u=1}^n$: all information about the community structure is then captured by B alone.

We make some further assumptions on the parameters of the model. For (1) to define a probability, we assume that

$$\frac{D_u D_v}{n \overline{D}} B_{\sigma_u \sigma_v} \leq 1 \quad \text{for all } u, v. \tag{2}$$

The vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)$ is assumed to be constant. Hence, the clusters are well balanced, as the size of each community grows linearly with n. Further, the average weight in a cluster,

$$\overline{D}_i = \frac{1}{\alpha_i n} \sum_{u=1}^n D_u \mathbf{1}_{\{\sigma_u = i\}},$$

is assumed to be asymptotically a fraction of the average weight $\overline{D}$. That is, we assume that there exist nonzero constants $d_1, \ldots, d_K$ such that

$$\lim_{n \to \infty} \frac{\overline{D}_i}{\overline{D}} = d_i \quad \text{for all } i. \tag{3}$$

Under this assumption, the following limit exists for all i:

$$\overline{M}_i = \lim_{n \to \infty} \frac{M_i}{\sum_{l=1}^n D_l} = \sum_{k=1}^K B_{ik} \alpha_k d_k, \tag{4}$$

where $M_i = \sum_l D_l B_{i\sigma_l}$.

We shall see in Section 4.4 that we need the following condition for the communities to be identifiable: we assume that, for all i, l, there exists j such that

$$\frac{B_{ij}}{\overline{M}_i} \neq \frac{B_{lj}}{\overline{M}_l}. \tag{5}$$

In the analysis that follows, we shall consider the random graph in a moderately sparse regime, i.e. we assume that either

$$\lim_{n \to \infty} \frac{D_1}{\log(n)} = \infty, \tag{6}$$

or, for some constant $c < \frac{1}{2}$,

$$D_1 \geq C_{B, \boldsymbol{M}} \log(n) \quad \text{and} \quad \lim_{n \to \infty} \frac{D_n^2}{n^c} \to 0, \tag{7}$$

where $C_{B,M}$ is some constant depending on B, $M = (\overline{M}_1, \ldots, \overline{M}_K)$ and the convergence rate in (4). Further, we assume the following condition on the weights:

$$\frac{D_1^2}{\overline{D}} = \Omega(\log(n)). \tag{8}$$

Note that under those assumptions, D_u represents the expected degree of vertex u up to a multiplicative factor that depends only on the community σ_u. Indeed, if $\widehat{D}_u$ denotes the observed degree of vertex u then

$$\mathbb{E}[\widehat{D}_u] = \frac{D_u}{n\overline{D}} \sum_{l \neq u} D_l B_{\sigma_u \sigma_l} = \frac{D_u}{n\overline{D}} (M_{\sigma_u} - D_u B_{\sigma_u \sigma_u}) = D_u \overline{M}_{\sigma_u} (1 + \varepsilon_n), \tag{9}$$

where $\varepsilon_n \leq 2/\overline{M}_{\sigma_u} D_u = o_n(1)$.

As an example, we let $\{\sigma_u\}_{u=1}^n$ be any sequence such that $n/2$ of its elements are 1 and the other $n/2$ elements are 2. Then there are two equally sized communities: $K = 2$ and $\alpha_1 = \alpha_2 = \frac{1}{2}$. Let $\{D_u\}_{u=1}^n$ be any nondecreasing sequence with $D_1 > 0$. Set

$$B = \begin{pmatrix} a & b \\ b & a \end{pmatrix} \quad \text{for some constants } a \text{ and } b.$$

Then

$$\mathbb{P}(u \sim v) = \frac{D_u D_v}{n\overline{D}} \begin{cases} a & \text{if } \sigma_u = \sigma_v, \\ b & \text{otherwise}. \end{cases}$$

This is exactly the extended-planted partition (EPP) model considered in [4].

3. Main results

Our aim is to retrieve the underlying community structure from a single observation of the random graph. We do this by analysing the spectral properties of $\widehat{H} \in \mathbb{R}^{n \times n}$ defined, for $u, v \in V$, by

$$\widehat{H}_{uv} = \begin{cases} \dfrac{1}{\widehat{D}_u \widehat{D}_v} A_{uv} & \text{if } A_{uv} = 1, \\ 0 & \text{otherwise}, \end{cases}$$

where A is the adjacency matrix of the observed graph. We shall demonstrate that this matrix is close (in a sense to be specified below) to the matrix P defined for (u, v) as

$$P_{uv} = \frac{1}{n\overline{D}} \frac{B_{\sigma_u \sigma_v}}{\overline{M}_{\sigma_u} \overline{M}_{\sigma_v}}.$$

Denote the rank of P by L. Due to the community structure, $L \leq K$ (see below for details).

In the regime where (6) holds, let f be any function tending to 0, such that

$$f(n) \gg \frac{1}{\widehat{D}_1} + \frac{1}{\sqrt{\log(n)}} + \sqrt{\frac{\log(n)}{\widehat{D}_1}}.$$

For the regime where (7) holds, let f tend to 0 in such a way that

$$f(n) \gg \frac{1}{\widehat{D}_1} + \frac{1}{\sqrt{\log(n)}} + \frac{1}{\log^{1/3}(n)}.$$

Further, let $\tau(n) = 1/f(n)^{1/3}$.

In Algorithm 1 we use $\widehat{H}$ to reconstruct the communities.

Algorithm 1. (i) Calculate the average degree in the graph, call it $\widehat{D}_{\text{average}}$. Let $\widehat{L}$ be the number of eigenvalues of $\widehat{H}$ that are in absolute value larger than $f(n)/\widehat{D}_{\text{average}}$.

(ii) Compute the first $\widehat{L}$ orthonormal eigenvectors of $\widehat{H}$ ordered according to their absolute eigenvalues. Denote these eigenvectors and their corresponding eigenvalues by $\widehat{x}_1, \ldots, \widehat{x}_{\widehat{L}}$ and $\widehat{\lambda}_1, \ldots, \widehat{\lambda}_{\widehat{L}}$, respectively.

(iii) Associate to each node $u \in V$ the vector

$$\widehat{z}_u = (\widehat{x}_1(u), \ldots, \widehat{x}_{\widehat{L}}(u)). \tag{10}$$

Cluster the vectors $(\widehat{z}_u)_{u=1}^n$ as follows. Pick $\tau(n)$ pairs of vertices and label them

$$(u(1), u'(1)), \ldots, (u(\tau(n)), u'(\tau(n))).$$

Calculate $\delta(t) = \sqrt{n}\|\widehat{z}_{u(t)} - \widehat{z}_{u'(t)}\|$, and $\varepsilon = \min_{t:\delta(t) > f^{2/3}(n)} \delta(t)$. Find a vertex m so that $\{u': \sqrt{n}\|\widehat{z}_m - \widehat{z}_{u'}\| \leq \varepsilon/8\}$ has cardinality larger than $f^{1/3}(n)n$. Form a community consisting of all nodes in $\{u': \sqrt{n}\|\widehat{z}_m - \widehat{z}_{u'}\| \leq \varepsilon/4\}$. Remove those nodes and iterate this procedure.

Theorem 1. *Consider a DC-SBM $\mathcal{G}(B, \boldsymbol{\alpha}, \{\sigma_u\}_{u=1}^n, \{D_u\}_{u=1}^n)$. Assume that assumptions (2), (3), (5), (8), and either (6) or (7) hold. Then Algorithm 1 retrieves the community of all but a vanishing fraction of nodes.*

The first step of the algorithm estimates L. Indeed, by definition there are only L nonzero eigenvalues of P. Those are all of order $1/\overline{D}$ and the corresponding first eigenvalues of $\widehat{H}$ are of the same order. The remaining eigenvalues of $\widehat{H}$ are negligible with respect to $f(n)/\overline{D}$.

Under the assumptions in Theorem 1, all but a negligible number of rows of the matrix having the first L eigenvectors of $\widehat{H}$ as it columns, cluster for large n to within negligible distance of block-specific representatives that are separated by some nonvanishing gap (call the corresponding vertices typical). This is exploited in the third step. There, with high probability, all picked vertices are typical. Thus, for a pair t, $\delta(t)$ vanishes in front of $f^{2/3}(n)$ if the vertices in the pair belong to the same community. Hence, by calculating the distance between the other vertices, we obtain ε as an estimator for the gap mentioned above. At most $f(n)^{2/3}n$ vertices are not typical. Hence, the chosen ball around m with radius $\varepsilon/8$ contains a negligible number of nontypical vertices, the remaining vertices should necessarily be in the same community. By enlarging the radius of the ball around m, we include all vertices of a single community. See the proof of Theorem 1 below for more details.

Remark 1. Note that the only input to the algorithm is the regime (i.e. either $D_1(n) = \Theta(\log(n))$ or $D_1(n) \gg \log(n)$). This information is used to pick the correct form of the function f. Alternatively, we could adapt the algorithm so that it requires $L = \text{rank}(B)$ and $\alpha_{\min}$ instead of the regime: step (i) can then be skipped, in step (ii) we replace $\widehat{L}$ by L and in step (iii) we chose a vertex m that contains in its $\varepsilon/8$ neighbourhood at least $\alpha_{\min}n/2$ vertices.

4. Discussion

Before we prove the main theorem, we make some observations and remarks.

4.1. Adjacency matrix

In [19] and [21], the authors used the adjacency matrix A of a graph to recover the underlying community structure. They considered the matrix having the first K eigenvectors of A as its columns and showed that, under appropriate conditions, its rows cluster in K different

directions. However, results in [6] and [22] suggest that the algorithms in [19] and [21] fail when the expected degree sequence is too irregular. Intuitively, if the prescribed degree sequence follows a power law then so does the spectrum of the adjacency matrix. Further, as we shall demonstrate below, the first K eigenvectors correspond only to the K top-degree nodes, and should therefore not be expected to capture more global features of a graph, such as its underlying block structure. The following theorem makes this observation more rigorous.

Theorem 2. *Consider a DC-SBM* $\mathscr{G}(B, K, \{\sigma_u\}_{u=1}^n, \{D_u\}_{u=1}^n)$ *such that*

$$D_u = \begin{cases} D_1 & \text{if } 1 \le u < n - k, \\ D_1 n^\gamma (u + 1 - (n - k)) & \text{if } u \ge n - k, \end{cases}$$

where $k = n^\beta$ *and the constants* β *and* γ *obey*

$$D_1^2(n) n^{2\gamma + 4\beta - 1} \to 0 \tag{11}$$

and

$$\gamma > 4\beta. \tag{12}$$

Further, assume that

$$\sigma_u = \begin{cases} 2 & \text{if } u \le n/2, \\ 1 & \text{if } u > n/2. \end{cases}$$

Under these conditions, the first k *eigenvectors become, for large n, indistinguishable from the eigenvectors of a graph that is the disjoint union of k stars having degrees* $D_n + o(1), \ldots, D_{n-k} + o(1)$.

For instance, $D_1(n) = n^{1/20}$, $\beta = \frac{1}{20}$ and $\gamma = \frac{1}{5}$, meets the assumptions in Theorem 2. Further, it verifies the conditions in the main theorem (Theorem 1): Algorithm 1 will successfully return the community membership of all but a vanishing fraction of nodes.

We remark that the above theorem is inspired by the main result in [22]. There, random graphs without community structure were considered and the power-law behaviour of the corresponding spectrum was obtained. To say something about the eigenvectors, we additionally introduce a gap between the top k-degreed nodes and the remaining $n - k$ nodes. This allows us to use Lemma 11; see the proof of Theorem 2 below.

4.2. Spectral clustering on ratios-of-eigenvectors

Interestingly, the first eigenvectors of A do contain information about the underlying community structure, but in a hidden way. Indeed, the spectral clustering on ratios-of-eigenvectors (SCORE) method proposed in [15] shows that, under some conditions, using the coordinate-wise ratios of the leading eigenvectors leads to consistent clustering.

Note that we obtain the same random graph model as in [15] by putting $\theta(u) := D_u / \sqrt{n \overline{D} \alpha}$ and $P(i, j) = \alpha B_{ij}$, where $\alpha^{-1} = \max_{i,j} B_{ij}$. We further note that the conditions are more stringent: (2.7) demands that P (or B) is nonsingular which is unnecessary here; see Remark 2 below.

4.3. Laplacian

As we pointed out, the adjacency matrix does not capture accurately global properties of a graph. The normalized Laplacian is a more suitable candidate. It is defined by $L = I - D^{-1/2} A D^{-1/2}$, where I is the identity matrix, A is the adjacency graph, and D the diagonal

matrix with the row sums of A on its diagonal (i.e. the degrees). The object of study in [6] was the Laplacian spectra of random graphs with a given degree sequence $(d_1, \ldots, d_n)$, where edges are independently present between each pair of vertices (u, v) with probability $d_u d_v / \sum_{l=1}^{n} d_l$. In the regime $d_1^2 \gg \overline{D}$, with $\overline{D} = 1/n \sum_{l=1}^{n} d_l$, the eigenvalues satisfy the semicircle law with respect to the circle of radius $2/\sqrt{\overline{D}}$ centred at 1.

Denote the eigenvalues of the normalized laplacian by $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n \leq 2$. It is a well-known fact that all eigenvalues are located in the interval $[0, 2]$ and that the algebraic multiplicity of 0 is equal to the number of components in the graph. The authors of [6] further studied the spectral gap $\lambda = \min\{\lambda_2, 2 - \lambda_n\}$, which reflects global properties of the random graph. According to [6], when $d_1 \gg \log^2(n)$,

$$\lambda \geq 1 - \frac{1 + o(1)}{4/\sqrt{\overline{w}}} - \frac{\log^2(n)}{d_1},$$

thus in this dense regime, all nonzero eigenvalues are close to 1 and, thus, the spectrum of the Laplacian contains no outliers, in contrast with the adjacency matrix. This bound is improved in [5], to

$$\lambda \geq 1 - 2\sqrt{\frac{6 \log(2n)}{d_1}} \quad \text{for } d_1 \gg \log(n).$$

The SBM is a special case of the latent space model [12]. In this model an unknown vector z_u is associated to each node u (in a social network, this vector would represent the unknown social position of person u) and an edge between u and v is present with probability depending only on z_u and z_v. If A is the adjacency matrix of the graph, D the diagonal matrix containing the degrees, and $L = D^{-1/2} A D^{-1/2}$, then the population version of these matrices are defined as $\mathcal{A} = \mathbb{E}[A \mid z_1, \ldots, z_n]$, $\mathcal{D} = \text{diag}(\sum_{v=1}^{n} \mathcal{A}_{1v}, \ldots, \sum_{v=1}^{n} \mathcal{A}_{nv})$, and $\mathcal{L} = \mathcal{D}^{-1/2} \mathcal{A} \mathcal{D}^{-1/2}$. In [28], convergence of the empirical eigenvectors of L to the population eigenvectors of $\mathcal{L}$ was shown. This follows from their novel result establishing the convergence of L^2 to $\mathcal{L}^2$ in Frobenius norm. This forms the basis of an algorithm that uses the first k eigenvectors to recover the hidden communities in the SBM (thus, without degree corrections). The algorithm is shown to succeed if those first k eigenvalues are sufficiently separated from the rest of the eigenvalues and if the minimum expected degree exceeds $\sqrt{2}n/\sqrt{\log n}$, which is more restrictive than the lower bound of $\log n$.

In [8], the matrix $\mathbb{E}[D]^{-1/2} A \mathbb{E}[D]^{-1/2}$ (reminiscent of the normalized Laplacian) was used to retrieve the underlying community structure in the DC-SBM. Note that this method requires the expected degrees to be known. It succeeds if the minimum degree is of order $\log^6 n$.

To deal with low-degree nodes, the authors in [4] used the degree-corrected random walk laplacian: $I - (D + \tau I)^{-1} A$, where $\tau > 0$ is a constant, to find clusters in the EPP model, where the expected minimum degree is $\Omega(\log n)$. In the EPP model, B is a matrix where an element is equal to p if it is on the diagonal, and q otherwise; it is, thus, a special case of the DC-SBM. The algorithm based on the random walk laplacian requires τ as input and the optimal value of τ depends in a complex way on the degree distribution of the graph. The main theorem in [4] comes with lengthy conditions that are not easy to compare with other results. This theorem, restricted to the setting where all the d_u are equal to d, assumes q to be a constant, which is more restrictive than our assumptions. It is unclear whether the results for the EPP model can be neatly generalized using the same operator to the DC-SBM, given the complexity of the present conditions.

Although the Laplacian captures global properties of a graph much better than the adjacency matrix, its spectrum is still influenced by the underlying degree structure. Indeed, consider a DC-SBM with 3000 vertices divided in $K = 3$ equally sized communities, with

$$B = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 0 & 2 \\ 3 & 2 & 5 \end{pmatrix},$$

degree sequence

$$D_u = \begin{cases} u^{1/3} & \text{if } u = 1, \ldots, 1000, \\ (u - 1000)^{1/3} & \text{if } u = 1001, \ldots, 2000, \\ (u - 2000)^{1/3} & \text{if } u = 2001, \ldots, 3000, \end{cases}$$

and community membership

$$\sigma_u = \begin{cases} 1 & \text{if } u = 1, \ldots, 1000, \\ 2 & \text{if } u = 1001, \ldots, 2000, \\ 3 & \text{if } u = 2001, \ldots, 3000. \end{cases}$$

In Figure 1, we plot the eigenvectors corresponding to the first and second largest absolute eigenvalues of $I - \mathbb{E}[D]^{-1/2}\mathbb{E}[A]\mathbb{E}[D]^{-1/2}$, where A is the adjacency matrix and D is the diagonal matrix containing the row sums of A. The Laplacian concentrates around $I - \mathbb{E}[D]^{-1/2}\mathbb{E}[A]\mathbb{E}[D]^{-1/2}$ if the minimum degree is large enough (see Section 8). The community structure is clearly perturbed by the degree sequence. In general, an additional step is needed to determine the community membership of all nodes when using the Laplacian.

Compare this figure to Figure 2, containing the first two eigenvectors of

$$\mathbb{E}[D]^{-1}\mathbb{E}[A]\mathbb{E}[D]^{-1}.$$

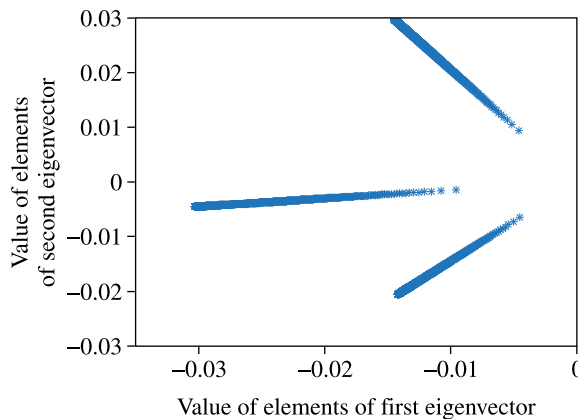The vertices are seen to be clearly divided into three communities.

FIGURE 1: Plot of the eigenvectors corresponding to the first and second largest absolute eigenvalue of $I - \mathbb{E}[D]^{-1/2}\mathbb{E}[A]\mathbb{E}[D]^{-1/2}$, where A is the adjacency matrix of a random graph drawn according to the DC-SBM defined at the end of Section 4.2, and D is the diagonal matrix containing the row sums of A. For those eigenvectors, say $(x_1, \ldots, x_n)'$ and $(y_1, \ldots, y_n)'$, we draw a dot (x_u, y_u) for each element u.

Now consider another two-community DC-SBM on n vertices with

$$B = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix},$$

degree sequence

$$D_u = \begin{cases} \log^2(n) & \text{if } u \leq n/2, \\ 100 \log^2(n) & \text{if } u > n/2. \end{cases}$$

and community membership

$$\sigma_u = \begin{cases} 1 & \text{if } u \leq n/2, \\ 2 & \text{if } u > n/2. \end{cases}$$

Then, according to Lemma 11, the eigenvectors of H become eventually indistinguishable from the eigenvectors of the $n \times n$ matrix with zero diagonal and all other elements equal to $1/n\overline{D}$. Clearly, the communities cannot be recovered from the latter matrix.

The off-diagonal elements of $\mathbb{E}[D]^{-1/2}\mathbb{E}[A]\mathbb{E}[D]^{-1/2}$ are given by

$$\frac{1}{n\overline{D}}\sqrt{D_u}\sqrt{D_v} = \frac{2}{101n}Z_{\sigma_u\sigma_v} \quad \text{with } Z = \begin{pmatrix} 1 & 10 \\ 10 & 100 \end{pmatrix}.$$

Now, Z has eigenvector

$$\begin{pmatrix} 1 \\ 10 \end{pmatrix}$$

corresponding to eigenvalue 101. The other eigenvalue is 0. Then that the minimal gap between different eigenvalues of $\mathbb{E}[D]^{-1/2}\mathbb{E}[A]\mathbb{E}[D]^{-1/2}$ is $2 - \mathcal{O}(1/n)$. According to [5], $\rho(D^{-1/2}AD^{-1/2} - \mathbb{E}[D]^{-1/2}\mathbb{E}[A]\mathbb{E}[D]^{-1/2}) = o(1)$ with high probability, where $\rho(X)$ denotes the spectral radius of a matrix X. Consequently, Lemma 11 entails that for large n, clustering according to the eigenvector of $D^{-1/2}AD^{-1/2}$, corresponding to its largest eigenvalue, reveals the community membership of all but a vanishing fraction of nodes.
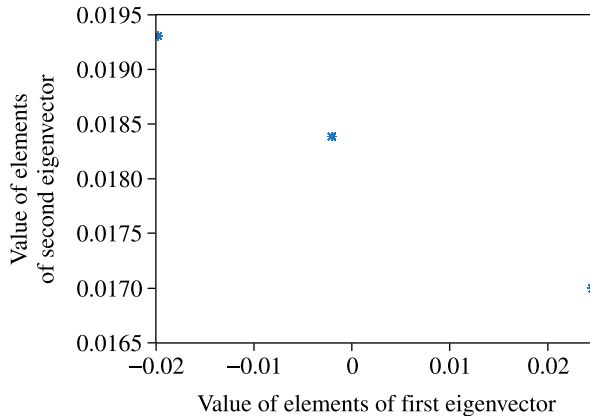

FIGURE 2: Plot of the eigenvectors corresponding to the first and second largest absolute eigenvalue of $\mathbb{E}[D]^{-1}\mathbb{E}[A]\mathbb{E}[D]^{-1}$, where A is the adjacency matrix of a random graph drawn according to the DC-SBM defined at the end of Section 4.2, and D is the diagonal matrix containing the row sums of A. For those eigenvectors, say $(x_1, \ldots, x_n)'$ and $(y_1, \ldots, y_n)'$, we plot a point (x_u, y_u) for each element u. Note that many elements are represented by the same point, clearly reflecting the community structure.

Those two examples hint that whether the Laplacian L or the degree-normalized adjacency matrix H should be used depends on the correlation between the degrees and the communities, and the 'signal strength' of B. The first example shows that if the degrees are uncorrelated, L seems to add some extra noise, whereas H 'filters' the degrees and reflects immediately the underlying communities. In the second example, B gives no information about the communities, but the vertices can be clustered according to their degrees. Matrix H ignores this degree structure and, thus, fails to detect the communities. Laplacian L on its turn still reflects the degree sequence and, therefore, the communities.

4.4. Regularized spectral clustering

Qin and Rohe [27] dealt with the shortcomings of the Laplacian by inflating the degrees: given a number $\tau > 0$, the regularized graph Laplacian [4], [27] is defined as

$$L_\tau = D_\tau^{-1/2} A D_\tau^{-1/2}, \tag{13}$$

where $D_\tau = D + \tau I$.

The regularized spectral clustering algorithm in [27] starts with computing the matrix $X = [X_1, X_2, \ldots, X_K]$, where $X_1, X_2, \ldots, X_K$ are the eigenvectors corresponding to the K largest eigenvalues. A matrix X^* is then formed by projecting each row of X on the unit sphere. Considering each row of X^* as a point in $\mathbb{R}^K$, and applying k-means with K centres on these points gives an almost exact clustering, provided some conditions on $\delta + \tau$ (δ is the smallest expected degree) and the smallest strictly positive eigenvalue of L_τ hold. In particular, [27, Theorem 4.2(a)] demands that $\delta + \tau \gg \log(n)$. Since simulation results suggest that τ should be taken as the average degree, it is unclear if this method outperforms the algorithm proposed in [27].

We note that [27] is the first work that relates the leverage scores (the Euclidean norm of the rows of X) to the quality of the outputted clustering.

4.5. When does the degree-normalized adjacency matrix fail?

Consider a DC-SBM with $K \geq 2$ communities, such that for two different communities $i \neq j$, for all l, $B_{il}/\overline{M}_i = B_{jl}/\overline{M}_j$. Then it can be verified that, for large n, in a dense enough regime, the eigenvectors of H corresponding to nonzero eigenvalues do not distinguish between communities i and j.

Further, the method breaks down in too sparse a regime. For instance, two low-degreed vertices connected by an edge cause the top eigenvectors to concentrate around them. We observed this when applying $\widehat{H}$ on the sparse political blogs network [2], see Section 4.7.

4.6. Degree-normalized adjacency matrix

The same matrix H was used in [7] to perform community detection on the DC-SBM in the sparse regime (the minimum degree is bounded from below by a constant). The main restriction in their setting is that the minimum degree must be of the same order as the average degree, more precisely, there exists $\varepsilon > 0$ such that $D_i > \varepsilon \overline{D}$ for all i. Hence, too much irregularity in the degree sequence is not captured. In this sense our work complements their results.

Spectral clustering was performed in [7] on a minor of $\widehat{H}$ where the rows and columns of vertices with a degree smaller than $D_{\text{average}}/\log(n)$ (where D_{average} is the observed average degree in the graph) are set to 0, which is not the same as leaving out completely the nodes with too low a degree. Due to the assumption that all expected degrees are of the same order, most observed degree will exceed the lower bound $D_{\text{average}}/\log(n)$.

There are alternative ways to deal with low-degree nodes; see, for instance, Section 8 on future research.

4.7. Performance on real networks: karate club, dolphins, and political blogs

We have tested our method on three real networks, namely, Zachary's karate club [31], the dolphin social network [20], and the political blogs dataset [2]. The error rate for Zachary's karate club is 2/34 and for the dolphin social network 0/62.

The error rate for the political blogs dataset is 230/1221 when thresholding the Frobenius eigenvector. We restricted ourselves to the giant component of 1221 nodes, as is common in most other works (the original data contained 1490 blogs). Our clustering is worse than that obtained by SCORE (where the error rate is 58/1221), but similar to the nonbacktracking matrix (where around 15 percent of the nodes are misclassified [17]).

We observed that the leading eigenvectors are concentrated on a few nodes, due to the presence of certain problematic structures (such as two low-degreed vertices connected by an edge). However, the value of the Frobenius eigenvector on the remaining vertices is still correlated with their community membership as can be observed in Figures 3 and 4.

Figure 3 is a histogram of the Frobenius eigenvector restricted to the roughly 600 nodes that have corresponding value in the interval $[0, 10^{-9}]$. The nodes seem to concentrate around two centres according to their community. However, this phenomenon is only weakly visible (note that our theory does not apply for *sparse* graphs).

In Figure 4 we have sorted the 1221 indices of the Frobenius eigenvector according to an increasing corresponding value: the community structure then becomes clear.

We further observed that thresholding the eighth eigenvector leads to only 160 misclassified vertices. Interestingly, if we inflate the degrees by replacing $H = A_{uv}/\widehat{D}_u\widehat{D}_v$ by $H_{\text{inflated}} = A_{uv}/\max\{\widehat{D}_u\widehat{D}_v, 200\}$, we obtain an error rate of 74/1221 by thresholding its second eigenvector. This suggests that initial misclassifications are indeed due to low-degree nodes (the average degree is 27, but there are also many leaves present).

4.8. Interpretation of the conditions

Note that, since $\mathbb{E}\widehat{D}_u$ is related to D_u according to (9), $\widehat{H}$ normalizes the tendency of communities to connect by the average degree of their nodes and loses, therefore, giving some information about the graph.
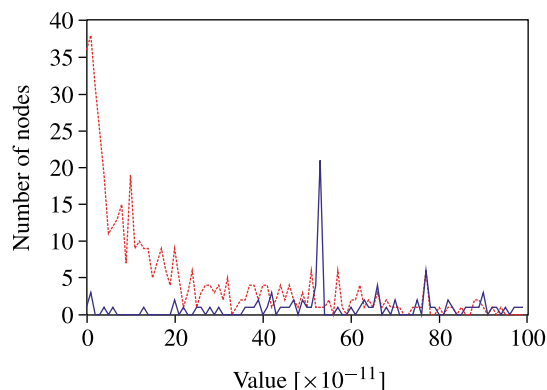

FIGURE 3: Histogram of the Frobenius eigenvector restricted to the roughly 600 nodes that have corresponding value in the interval $[0, 10^{-9}]$. Two communities are shown.
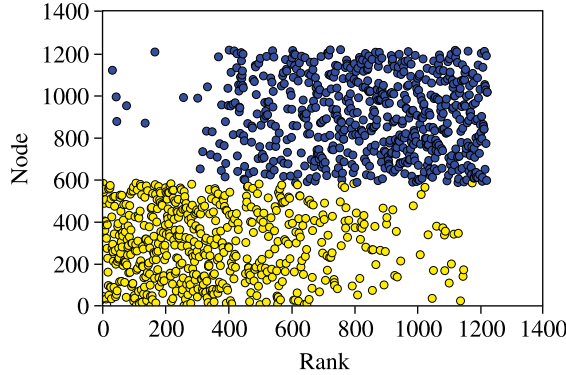
FIGURE 4: Ranking of the 1221 indices of the Frobenius eigenvector according to an increasing corresponding value. Rank 1 is the node with smallest value in the eigenvector and rank 1221 the node with largest value. Shaded circles indicate community membership.

Observation 1. If, for some $i, j, l \in S$,

$$\frac{B_{ij}}{M_i} = \frac{B_{lj}}{M_l}$$

then

$$\frac{\mathbb{E}[\text{\#edges between community } i \text{ and } j]}{\mathbb{E}[\text{total degree of vertices in community } i]} = \frac{\mathbb{E}[\text{\#edges between community } l \text{ and } j]}{\mathbb{E}[\text{total degree of vertices in community } l]}.$$

Remark 2. The identifiability condition is violated if there are distinct i and l and there exists some constant $c > 0$ such that

$$B_{ij} = cB_{lj} \quad \text{for all } j.$$

Indeed, in that case, $M_i = cM_l$ and, thus,

$$\frac{B_{ij}}{M_i} = \frac{cB_{lj}}{cM_l} = \frac{B_{lj}}{M_l}.$$

However, unlike the setting considered in [15] and [19], it is not necessary for B to be full rank. Indeed, consider

$$B = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 0 & 2 \\ 3 & 2 & 5 \end{pmatrix}$$

which has rank 2. Let $\alpha_1 = \alpha_2 = \alpha_3 = \frac{1}{3}$ and $\sum_{\sigma_u=i} D_u = i\alpha_i n \log^2(n)$ for all $i = 1, 2, 3$. Then it is easily verified that the identifiability condition is met.

Note that $\mathcal{G}(B, K, \{\sigma_u\}_{u=1}^n, \{D_u\}_{u=1}^n)$ and $\mathcal{G}(B^*, K, \{\sigma_u\}_{u=1}^n, \{D_u^*\}_{u=1}^n)$ generate the same ensemble of random graphs whenever

$$\frac{D_u B_{\sigma_u \sigma_v} D_v}{\overline{D}} = \frac{D_u^* B_{\sigma_u \sigma_v}^* D_v^*}{\overline{D^*}}.$$

Hence, the underlying block-matrix B cannot be estimated from a single observation of the graph. Rather, we may estimate

$$n\overline{D} \approx \sum_u \widehat{D}_u,$$

and, denoting the assigned community membership (after applying our reconstruction algorithm) of l by τ_l,

$$\left(\sum_u \widehat{D}_u \right) \frac{\sum_{u:\, \tau_u=i} \sum_{v:\, \tau_v=j} \widehat{H}_{uv}}{(\sum_{u:\, \tau_u=i} 1)(\sum_{v:\, \tau_v=j} 1)} \approx \frac{B_{ij}}{\overline{M}_i \overline{M}_j}.$$

Hence, for a DC-SBM $\mathcal{G}(B, K, \{\sigma_u\}_{u=1}^n, \{D_u\}_{u=1}^n)$, the matrix

$$\left(\frac{B_{ij}}{\overline{M}_i \overline{M}_j} \right)_{i,j=1}^K$$

is identifiable, not B. It is due to this degeneracy of the DC-SBM and the structure of $\widehat{H}$ that condition (5) in Theorem 1 is the best possible.

Lemma 1. *Consider a DC-SBM $\mathcal{G}(B, K, \{\sigma_u\}_{u=1}^n, \{D_u\}_{u=1}^n)$. Fix i and l then the following are equivalent:*

  (i) *for all j, we have*
$$\frac{B_{ij}}{\overline{M}_i} = \frac{B_{lj}}{\overline{M}_l};$$

  (ii) *there exist a DC-SBM $\mathcal{G}(B^*, K, \{\sigma_u\}_{u=1}^n, \{D_u^*\}_{u=1}^n)$ with the same community structure $\{\sigma_u\}_u$, such that, for all j,*
$$B_{ij}^* = B_{lj}^*$$

  and, for all u, v,
$$\frac{D_u B_{\sigma_u \sigma_v} D_v}{D} = \frac{D_u^* B_{\sigma_u \sigma_v}^* D_v^*}{D^*}.$$

4.9. Random walk point of view

The matrix $\widehat{H}$ is related to a random walk on an instance of the random graph. Indeed,

$$\widehat{H}_{uv} = \begin{cases} \dfrac{1}{\widehat{D}_u} \dfrac{1}{\widehat{D}_v} A_{uv} = \dfrac{A_{uv}}{\widehat{D}_u} \dfrac{A_{vu}}{\widehat{D}_v} & \text{if } A_{uv} = 1, \\ 0 & \text{if } A_{uv} = 0, \end{cases}$$

since $A_{uv} = A_{vu}$ is either 1 (in case edge uv is present) or 0 (when u and v are not connected). Now, $\widehat{D}_u = \sum_{l=1}^n A_{lu}$, as it is the observed degree, which we denoted here in increasing order: $\widehat{D}_1 \le \widehat{D}_2 \le \cdots \le \widehat{D}_n$. Thus, $A_{uv}/\widehat{D}_u$ is exactly the probability that a random walk (in an undirected graph without weights) jumps from vertex u to v, given that it is currently at vertex u. Denoting the latter probability by $\mathbb{P}_u(u \to v)$, we see that

$$\widehat{H}_{uv} = \mathbb{P}_u(u \to v)\mathbb{P}_v(v \to u) = \mathbb{P}_u(u \to v \to u)$$

due to the Markov property of the random walk. In other words, $\widehat{H}_{uv}$ is the probability that a random walk currently at vertex u will consecutively traverse edge uv and back.

Extending this observation to powers of $\widehat{H}$ leads to

$$(\widehat{H}^k)_{uv} = \sum_{l_1=1,\ldots,l_{k-1}=1}^n \mathbb{P}_u(u \to l_1 \to \cdots \to l_{k-1} \to v)\mathbb{P}_v(v \to l_{k-1} \to \cdots \to l_1 \to u),$$

the probability that a random walk, after traversing a path of length k starting at u and ending at v, subsequently traverses that path in the exact opposite direction.

Further, note that

$$(\widehat{D}_1, \dots, \widehat{D}_n)\widehat{H} = (\mathbf{1}_{\{\widehat{D}_1 \neq 0\}}, \dots, \mathbf{1}_{\{\widehat{D}_n \neq 0\}}),$$

hence if $\boldsymbol{v}$ is an eigenvector of $\widehat{H}$ with eigenvalue λ then

$$\sum_{u=1}^{n} \mathbf{1}_{\{\widehat{D}_u \neq 0\}} v_u = \lambda \sum_{u=1}^{n} \widehat{D}_u v_u.$$

Since it can be easily verified that $\widehat{H}$ is primitive on connected components, the Perron–Frobenius theorem implies that the eigenvector $\boldsymbol{v}_{\max}$ corresponding to the largest eigenvalue $\lambda_{\max}$ (which is positive) has only positive elements. Hence,

$$\lambda_{\max} = \frac{\sum_{u=1}^{n} \mathbf{1}_{\{\widehat{D}_u \neq 0\}} v_u}{\sum_{u=1}^{n} \widehat{D}_u v_u} \geq \frac{\sum_{u=1}^{n} \mathbf{1}_{\{\widehat{D}_u \neq 0\}} v_u}{\widehat{D}_n \sum_{u=1}^{n} \mathbf{1}_{\{\widehat{D}_u \neq 0\}} v_u} = \frac{1}{\widehat{D}_n}.$$

We may derive an upper bound by noting that the spectral radius is bounded from above by the maximal absolute row sum

$$\lambda_{\max} \leq \max_{u=1}^{n} \left(\sum_{v=1}^{n} \mathbb{P}_u(u \to v \to u) \right).$$

5. Outline of proof of main theorem

In this section we consider the setting of Theorem 1. All lemmas here, except Lemma 5, assume that either (6) or (7) hold. Lemma 5 assumes condition (6): the minimum degree should grow faster than $\log(n)$. Lemma 6 assumes (7): the minimum degree is of order $\log(n)$.

Our first objective is to show that $\widehat{H}$ is close to some matrix P, in the sense that their difference $W := \widehat{H} - P$ has negligible spectral radius relatively to that of P. Here, an entry (u, v) of P is defined as

$$P_{uv} = \frac{1}{n\overline{D}} \frac{B_{\sigma_u \sigma_v}}{\overline{M}_{\sigma_u} \overline{M}_{\sigma_v}}.$$

We relate P in turn to Z defined by

$$Z_{ij} = \frac{\alpha_j B_{ij}}{\overline{M}_i \overline{M}_j}, \qquad i, j \in S.$$

Indeed, we show that if $y = (y(1), \dots, y(K))^{\top}$ is an eigenvector of Z with eigenvalue λ then $(y(\sigma_1), \dots, y(\sigma_n))^{\top}$ fulfils that role for P with eigenvalue $(1/\overline{D})\lambda$. As a consequence, the eigenvectors of P associated to nonzero eigenvalues are constant on blocks.

Finally, we consider the matrix that has the first L eigenvectors of P as its columns. We show that the rows of this matrix cluster to within vanishing distance of block-specific representatives. We start by inspection of the difference

$$W = \widehat{H} - P = (\widehat{H} - H) + (H - \mathbb{E}[H]) + (\mathbb{E}[H] - P), \tag{14}$$

where H is defined as

$$H_{uv} = \begin{cases} \dfrac{1}{\mathbb{E}\widehat{D}_u \mathbb{E}\widehat{D}_v} A_{uv} & \text{if } A_{uv} = 1, \\ 0 & \text{otherwise}. \end{cases}$$

Define
$$\Delta(P) = \min\{|\lambda - \mu| \colon \lambda \neq \mu, \lambda, \mu \text{ eigenvalue of } P\},$$
i.e. the smallest gap between consecutive eigenvalues. A crucial role will be played by Lemma 11 below, which says that to any eigenvector $\widehat{x}$ of $\widehat{H}$ there exists an eigenvector x of P such that $\|x - \widehat{x}\| \to 0$ as $n \to \infty$, whenever

$$\frac{\rho(W)}{\Delta(P)} \to 0 \quad \text{as } n \to \infty,$$

where we recall that $\rho(X)$ denotes the spectral radius of a matrix X. Hence, we need to calculate $\Delta(P)$.

Lemma 2. *The smallest gap between subsequent eigenvalues of P is given by*

$$\Delta(P) = \Omega\left(\frac{1}{\overline{\overline{D}}}\right).$$

All terms in the right-hand side of (14) have negligible spectral radius with respect to $\Delta(P)$.

Lemma 3. *The matrix $\mathbb{E}[H]$ is close to P in the following sense:*

$$\rho(\mathbb{E}[H] - P) = \mathcal{O}\left(\frac{1}{D_1}\right)\frac{1}{\overline{\overline{D}}} = o_n(1)\frac{1}{\overline{\overline{D}}}.$$

Lemma 4. *The matrix H concentrates with high probability around its expectation, as*

$$\rho(H - \mathbb{E}[H]) = \mathcal{O}\left(\frac{1}{\sqrt{\log(n)}}\right)\frac{1}{\overline{\overline{D}}} = o_n(1)\frac{1}{\overline{\overline{D}}}.$$

Lemma 5. *Consider the DC-SBM in the dense regime, where (6) holds. Then, for the spectral radius of the difference $\widehat{H} - H$, it holds with high probability that*

$$\rho(\widehat{H} - H) = \mathcal{O}\left(\sqrt{\frac{\log(n)}{D_1(n)}}\right)\frac{1}{\overline{\overline{D}}(n)} = o_n(1)\frac{1}{\overline{\overline{D}}}.$$

Lemma 6. *Consider the DC-SBM in the regime where (7) holds. Then, for the spectral radius of the difference $\widehat{H} - H$, it holds with high probability that*

$$\rho(\widehat{H} - H) = \mathcal{O}\left(\frac{1}{\log^{1/3}(n)}\right)\frac{1}{\overline{\overline{D}}(n)} = o_n(1)\frac{1}{\overline{\overline{D}}}.$$

We use these lemmas in conjunction with Lemma 11 below to prove the following lemma.

Lemma 7. *To each normed eigenvector $\widehat{x}$ of $\widehat{H}$ corresponds a normed eigenvector x of P such that*

$$\widehat{x} \cdot x = 1 - \mathcal{O}\left(\left(\frac{\rho(W)}{\Delta(P)}\right)^2\right) = 1 - o_n(1),$$

where

$$\rho(W) \leq \rho(\widehat{H} - H) + \rho(H - \mathbb{E}[H]) + \rho(\mathbb{E}[H] - P).$$

Having proved this lemma, we show that Algorithm 1 indeed correctly reconstructs the community of all but a vanishing fraction of vertices.

Recall the definition of $\widehat{H}$ and observe that $\widehat{H}$ is symmetric. Consequently, there exist n eigenvectors of $\widehat{H}$ that form an *orthonormal* basis, thus, we are indeed able to find L *orthonormal* eigenvectors of $\widehat{H}$ corresponding to its first eigenvectors.

Next we show that the $(\widehat{z}_u)_{u \in V}$, defined in (10), tend to block-representatives.

Lemma 8. *There exist K vectors $\{t_k\}_{k \in S}$, i.e. block-representatives, such that*

$$\|\sqrt{n}\widehat{z}_u - t_{\sigma_u}\| = \mathcal{O}\left(\left(\frac{\rho(W)}{\Delta(P)}\right)^{2/3}\right) = o_n(1)$$

for all but $\mathcal{O}(n(\rho(W)/\Delta(P))^{2/3})$ nodes.

The remaining and crucial step is to demonstrate that those block-representatives are indeed distinct.

Lemma 9. *Assume that, for all i, j, there exists i' such that*

$$\frac{B_{ii'}}{\overline{M}_i \overline{M}_{i'}} \neq \frac{B_{ji'}}{\overline{M}_j \overline{M}_{i'}} \tag{15}$$

then $|t_k - t_l| = \Omega(1)$ for all $k \neq l$.

Proof of Theorem 1. After proving the above lemmas, it remains to show that $\widehat{L}$ in step (i) of Algorithm 1 with high probability is equal to L. Further, we should verify that the procedure in step (iii) forms the correct clusters. For the first step, note the following. In the regime where (6) holds,

$$\rho(W) = \mathcal{O}\left(\frac{1}{D_1} + \frac{1}{\sqrt{\log(n)}} + \sqrt{\frac{\log(n)}{D_1}}\right)\frac{1}{\overline{D}},$$

and in the other regime, where (7) holds,

$$\rho(W) = \mathcal{O}\left(\frac{1}{D_1} + \frac{1}{\sqrt{\log(n)}} + \frac{1}{\log^{1/3}(n)}\right)\frac{1}{\overline{D}}.$$

Compare this to f as in Algorithm 1: depending on the regime, the term in parentheses goes to 0 upon division by $f(n)$. To see this, note that due to Bernstein's inequality (16) and (25), $\widehat{D}_u \in (1/2\overline{M}_{\sigma_u}, 3/2\overline{M}_{\sigma_u})D_u$ for $u = 1$ and $u = n$ with high probability. Hence, $\widehat{D}_1$ ($\widehat{D}_n$) is of the same order of magnitude as D_1 (respectively D_n). Now, due to Lemma 11 below, the first L eigenvalues of $\widehat{H}$ are of order $1/\overline{D} - \mathcal{O}(\rho(W)) \gg f(n)/\overline{D}$. The remaining eigenvalues are of order $\mathcal{O}(\rho(W)) \ll f(n)/\overline{D}$. Further, D_{average} may be written as twice the sum of $\Omega(n^2)$ independent Bernoulli random variables. It is thus with high probability a constant away from $\overline{D}$. Hence, $\widehat{L} = L$ with high probability.

In step (iii), the probability that all picked pairs contain only typical vertices (i.e. whose corresponding rows cluster around K centres) is larger than $(1 - f^{2/3}(n))^{2\tau(n)}$ which tends to 1, since $f^{2/3}(n)\tau(n) \to 0$ as $n \to \infty$. Thus, with high probability, for a pair t, $\delta(t)$ vanishes in front of $f^{2/3}(n)$ if the vertices in the pair belong to the same community. Then $\delta(t)$ is of order $\Omega(1)$ otherwise. Hence, ε, as defined in step (iii) of Algorithm 1, is of order $\Omega(1)$; it thus estimates the separation distance in Lemma 9.

Further, at most $f(n)^{2/3}n$ vertices are not typical. Hence, the chosen ball around m with radius $\varepsilon/8$ contains at least $(f(n)^{1/3} - f(n)^{2/3})n \gg f(n)^{2/3}n$ typical vertices. Those must

necessarily belong to the same community. Since all typical vertices belonging to the same community are at most a distance $f(n)^{2/3}$ apart, all of them are located in the ball of radius $\varepsilon/4$ around m.

We see that the algorithm puts, with high probability, all but a vanishing fraction of nodes in K clusters. □

6. Algebraic preliminaries

We shall make use of the following fact about the spectral radius.

Lemma 10. *If* $|X| \leq Y$ *holds entry-wise for two real symmetric matrices X and Y then* $\rho(X) \leq \rho(Y)$.

Proof. Due to the Rayleigh–Ritz theorem, we have

$$\rho(X) = \max_{\|z\|=1} \|Xz\|.$$

Hence,

$$\rho(X) = \max_{\|z\|=1} \|Xz\| \leq \max_{\|z\|=1} \| Y|z| \| = \max_{\|z\|=1} \|Yz\| = \rho(Y). \qquad \square$$

The following lemma could be of independent interest as a simple alternative to the commonly used David–Kahan theorem.

Lemma 11. *Let A and δA be two $n \times n$ symmetric matrices. Let $\lambda_1 \geq \cdots \geq \lambda_n$ be the eigenvalues of $A + \delta A$ and $\mu_1 \geq \cdots \geq \mu_n$ be the eigenvalues of A. Let $\Delta = \min\{|\mu_i - \mu_j|: \mu_i \neq \mu_j, \mu_i, \mu_j$ eigenvalue of $A\}$. Assume that $\rho(\delta A) < \Delta/2$. Let $\boldsymbol{v}_i$ be a normed eigenvector of $A + \delta A$ corresponding to eigenvalue λ_i, for any $i = 1, \ldots, n$. Then*

  (i) $|\lambda_i - \mu_i| \leq \rho(\delta A)$;

 (ii) *the dimension of the eigenspace E_i of $A + \delta A$ corresponding to the eigenvalue λ_i is no larger than the dimension of the eigenspace of A corresponding to the eigenvalue μ_i;*

 (iii) *there exists a normed eigenvector $\widehat{\boldsymbol{v}}_i$ of A corresponding to eigenvalue μ_i such that*

$$\boldsymbol{v}_i \cdot \widehat{\boldsymbol{v}}_i \geq \sqrt{1 - \left(\frac{\rho(\delta A)}{\Delta/2}\right)^2}.$$

Proof. (i) is due to Weyl's inequality (see, for instance, [14]).

(ii) Let d be the dimension of E_i and write $\lambda_i = \lambda_{i+1} = \cdots = \lambda_{i+d-1}$. Since $|\lambda_i - \mu_i| \leq \rho(\delta A)$, we have $|\mu_i - \mu_{i+1}| \leq 2\rho(\delta A) < \Delta$. Thus, $\mu_i = \mu_{i+1}$, and similarly for the other eigenvalues.

(iii) We start with some notation: let m be the number of distinct eigenvalues of A; denote those distinct numbers as $\gamma_1 > \cdots > \gamma_m$. Define $S_i = \{u \in \{1, \ldots, n\}: \mu_u = \gamma_i\}$, the set of indices of eigenvalues that are all equal to γ_i. For $u \in \{1, \ldots, n\}$, define $\tau_u \in \{1, \ldots, m\}$ as the unique index such that $u \in S_{\tau_u}$. Write

$$\boldsymbol{v}_i = \sum_j \alpha_j \boldsymbol{w}_j,$$

where $\{\boldsymbol{w}_j\}_j$ are orthonormal eigenvectors of A with associated eigenvalues $\{\mu_j\}_j$. Then

$$(A + \delta A)\boldsymbol{v}_i = \sum_j \alpha_j \mu_j \boldsymbol{w}_j + (\delta A)\boldsymbol{v}_i.$$

Hence,

$$(\delta A)\boldsymbol{v}_i = \sum_{j \notin S_{\tau_i}} \alpha_j (\lambda_i - \mu_j) \boldsymbol{w}_j + \sum_{j \in S_{\tau_i}} \alpha_j (\lambda_i - \mu_j) \boldsymbol{w}_i.$$

Taking norms on both sides,

$$(\rho(\delta A))^2 \geq \sum_{j \notin S_{\tau_i}} \alpha_j^2 (\lambda_i - \mu_j)^2 \geq \sum_{j \notin S_{\tau_i}} \alpha_j^2 \left(\Delta - \frac{\Delta}{2} \right)^2 = \left(1 - \sum_{j \in S_{\tau_i}} \alpha_j^2 \right) \left(\frac{\Delta}{2} \right)^2,$$

since, by definition $|\mu_i - \mu_j| \geq \Delta$ if $\tau_i \neq \tau_j$ and our observation $|\lambda_i - \mu_i| \leq \rho(\delta A) < \Delta/2$. Set

$$\widehat{\boldsymbol{v}}_i = \frac{1}{\sqrt{\sum_{j \in S_{\tau_i}} \alpha_j^2}} \sum_{j \in S_{\tau_i}} \alpha_j \boldsymbol{w}_j$$

then

$$\boldsymbol{v}_i \cdot \widehat{\boldsymbol{v}}_i = \sqrt{\sum_{j \in S_{\tau_i}} \alpha_j^2} \geq \sqrt{1 - \left(\frac{\rho(\delta A)}{\Delta/2} \right)^2}. \qquad \square$$

Lemma 12. *Consider a square $n \times n$ symmetric zero-diagonal random matrix A such that its elements $A_{uv} = A_{vu}$ are independent Bernoulli random variables with parameters*

$$\mathbb{E}[A_{uv}] = a_{uv} \frac{\widehat{\omega}(n)}{n},$$

where the a_{uv} are constants independent of n and $\widehat{\omega}(n) = \Omega(\log(n))$. Then, with probability larger than $1 - \mathcal{O}(1/n^2)$, the spectral radius of $A - \mathbb{E}[A]$ satisfies

$$\rho(A - \mathbb{E}[A]) \leq \mathcal{O}(\sqrt{\widehat{\omega}(n)}).$$

Proof. This is precisely Lemma 2 in [29], where the authors quantified the term *with high probability*. This was achieved by choosing $c_1 > 3$ in its proof. Note that the latter proof builds further on results by Feige and Ofek [10]. $\square$

Lemma 13. (Bernstein's inequality.) *Let $X_1, \ldots, X_n$ be zero-mean independent random variables all bounded from above by 1. Set $\sigma^2 = (1/n) \sum_{u=1}^{n} \mathrm{var}(X_u)$. Then*

$$\mathbb{P}\left(\frac{1}{n} \sum_{u=1}^{n} X_u > \varepsilon \right) \leq \exp\left(-\frac{n\varepsilon^2}{2(\sigma^2 + \varepsilon/3)} \right). \qquad (16)$$

Proof. See [3]. $\square$

Note that Bernstein's lemma can easily be extended to the case of noncentred random variables.

7. Proofs

In the proofs below, we shall often write

$$D_u = \phi_u \omega(n), \qquad (17)$$

where $1 = \phi_1 \leq \phi_2 \leq \cdots \leq \phi_n$, and

$$\omega(n) = D_1. \qquad (18)$$

Further, we introduce

$$g(n) = \sum_{l=1}^{n} \phi_l, \tag{19}$$

$$\overline{\phi}(n) = \frac{g(n)}{n}, \tag{20}$$

Proof of Lemma 2. Write

$$P_{uv} = \frac{1}{n\overline{D}} \frac{B_{\sigma_u \sigma_v}}{\overline{M}_{\sigma_u} \overline{M}_{\sigma_v}} = \frac{1}{n\overline{D}} \frac{Z_{\sigma_u \sigma_v}}{\alpha_{\sigma_v}}.$$

Let $\boldsymbol{y} = (y(1), \ldots, y(K))^{\top}$ be an eigenvector of Z with eigenvalue λ, we show that $\boldsymbol{w} = (y(\sigma_1), \ldots, y(\sigma_n))^{\top}$ is an eigenvector of P with eigenvalue $(1/\overline{D})\lambda$. Indeed,

$$
\begin{aligned}
P\boldsymbol{w} &= \begin{pmatrix} \sum_{l=1}^{n} P_{1l} \cdot y(\sigma_l) \\ \vdots \\ \sum_{l=1}^{n} P_{nl} \cdot y(\sigma_l) \end{pmatrix} \\
&= \frac{1}{n\overline{D}} \begin{pmatrix} \sum_{l=1}^{n} Z_{\sigma_1 \sigma_l}/\alpha_{\sigma_l} \cdot y(\sigma_l) \\ \vdots \\ \sum_{l=1}^{n} Z_{\sigma_n \sigma_l}/\alpha_{\sigma_l} \cdot y(\sigma_l) \end{pmatrix} \\
&= \frac{1}{n\overline{D}} \begin{pmatrix} \sum_{k=1}^{K} n\alpha_k Z_{\sigma_1 k}/\alpha_k \cdot y(k) \\ \vdots \\ \sum_{k=1}^{K} n\alpha_k Z_{\sigma_n k}/\alpha_k \cdot y(k) \end{pmatrix} \\
&= \frac{1}{\overline{D}} \begin{pmatrix} \lambda y(\sigma_1) \\ \vdots \\ \lambda y(\sigma_n) \end{pmatrix} \\
&= \frac{1}{\overline{D}} \lambda \boldsymbol{w}.
\end{aligned}
$$

Thus, $(1/\overline{D})\lambda$ is an eigenvalue of P.

For the other direction, note that if $\sigma_u = \sigma_v$ then row u and row v in P are identical. Hence, if $\boldsymbol{w} = (w(1), \ldots, w(n))^{\top}$ is an eigenvector of P corresponding to a nonzero eigenvalue then $w(u) = w(v)$. Let $\boldsymbol{w} = (w(\sigma_1), \ldots, w(\sigma_n))^{\top}$ be an eigenvector of P with eigenvalue $\lambda \neq 0$. By carrying out a similar calculation as above, we see that $(w(1), \ldots, w(K))^{\top}$ is an eigenvector of Z with eigenvalue $\overline{D}\lambda$.

The statement follows from this one-to-one correspondence between the eigenvectors of both matrices corresponding to nonzero eigenvalues. □

Proof of Lemma 3. Note that

$$\mathbb{E}[H] - P = \mathbb{E}[H] - (P - \operatorname{diag}(P_{11}, \ldots, P_{nn})) + \operatorname{diag}(P_{11}, \ldots, P_{nn}).$$

Now,

$$\rho(\operatorname{diag}(P_{11}, \ldots, P_{nn})) = \mathcal{O}\left(\frac{1}{n\overline{D}}\right)$$

as $\mathrm{diag}(P_{11}, \ldots, P_{nn})$ contains only K different elements, each of order $1/n\overline{D}$. Further, for $u \neq v$,

$$\mathbb{E}[H_{uv}] = \frac{D_u}{\mathbb{E}[\widehat{D}_u]} \frac{D_v}{\mathbb{E}[\widehat{D}_v]} B_{\sigma_u \sigma_v} \frac{1}{n\overline{D}} = (1 + \delta(n)) \frac{B_{\sigma_u \sigma_v}}{\overline{M}_{\sigma_u} \overline{M}_{\sigma_v}} \frac{1}{n\overline{D}} = P_{uv} + \delta(n) P_{uv},$$

where $\delta(n) = \mathcal{O}(\varepsilon_n)$ with, due to (9),

$$\varepsilon_n \leq \max_i \frac{2}{\overline{M}_i} \frac{1}{\omega(n)}$$

tending to 0 uniformly for all u, v. Hence, due to Lemma 10,

$$\rho(\mathbb{E}[H] - (P - \mathrm{diag}(P_{11}, \ldots, P_{nn})) + \mathrm{diag}(P_{11}, \ldots, P_{nn})) = \mathcal{O}\left(\frac{1}{D_1}\right) \frac{1}{\overline{D}}. \qquad \square$$

Proof of Lemma 4. We start by introducing the constants $C_{ij} = B_{ij}/\overline{M}_i \overline{M}_j$ and $\alpha = \max_{ij}(1/\overline{M}_i \overline{M}_j)$. Set, for $u < v$,

$$X_{uv} = X_{vu} = \alpha^{-1} \omega^2(n)(H_{uv} - \mathbb{E}[H_{uv}]),$$

where $\omega(n)$ is defined in (18). That is,

$$X_{uv} = \frac{1 + o(1)}{\alpha} \left(\frac{1}{\overline{M}_{\sigma_u} \overline{M}_{\sigma_v}} \frac{1}{\phi_u \phi_v} \mathrm{Ber}\left(\frac{D_u D_v}{n\overline{D}} B_{\sigma_u \sigma_v} \right) - C_{\sigma_u \sigma_v} \frac{\omega(n)}{\overline{\phi}} \frac{1}{n} \right),$$

with ϕ_u and $\overline{\phi}$ defined in (17) and (20), respectively. Due to our choice of α and the assumption that $\phi_u \geq 1$ for all u,

$$X_{uv} \in (1 + o(1))[-p_{uv}, 1 - p_{uv}],$$

where

$$p_{uv} = \frac{C_{\sigma_u \sigma_v}}{\alpha} \frac{\omega(n)}{\overline{\phi}} \frac{1}{n}.$$

Let $\widehat{X}_{uv} = X_{uv}/(1 + o(1))$ such that $\widehat{X}_{uv} \in [-p_{uv}, 1 - p_{uv}]$. We shall compare the symmetric zero-diagonal matrix $\widehat{X}$ to the deviation from its expectation of another symmetric zero-diagonal matrix, where elements uv are given by $\mathrm{Ber}(p_{uv})$ for $u \neq v$. Since, by assumption (8),

$$\frac{\omega(n)}{\overline{\phi}(n)} = \frac{D_1^2(n)}{\overline{D}(n)} = \Omega(\log(n)), \tag{21}$$

Lemma 12 applies. Following an argument given in [29], we construct a function Y_{uv} such that Y_{uv} has values only in $\{-p_{uv}, 1 - p_{uv}\}$ and $\mathbb{E}[Y_{uv} \mid \widehat{X}_{uv}] = \widehat{X}_{uv}$. First, let $\{U_{uv}\}_{u<v}$ be independent uniformly distributed random variables. Fix $u < v$. Define, for $x \in [-p_{uv}, 1 - p_{uv}]$ and $w \in [0, 1]$,

$$F_{uv}(x, w) = 1 - p_{uv} - 1_{x \leq w - p_{uv}} \quad \text{and} \quad Y_{uv} = Y_{vu} = F_{uv}(\widehat{X}_{uv}, U_{uv}).$$

Then

$$\mathbb{P}(F_{uv}(\widehat{X}_{uv}, U_{uv}) = 1 - p_{uv} \mid \widehat{X}_{uv}) = \widehat{X}_{uv} + p_{uv}$$

and

$$\mathbb{P}(F_{uv}(\widehat{X}_{uv}, U_{uv}) = -p_{uv} \mid \widehat{X}_{uv}) = 1 - p_{uv} - \widehat{X}_{uv};$$

thus,

$$\mathbb{E}[Y_{uv} \mid \widehat{X}_{uv}] = \widehat{X}_{uv} \quad \text{and} \quad \mathbb{P}(Y_{uv} = 1 - p_{uv}) = \mathbb{E}[\widehat{X}_{uv}] + p_{uv} = p_{uv}.$$

Hence, indeed, $Y_{uv} = \text{Ber}(p_{uv}) - p_{uv}$.

Let Y be the symmetric zero-diagonal matrix with each element uv given by Y_{uv} for $u \neq v$. Then, according to Lemma 12,

$$\mathbb{P}\left(\rho(Y) \leq \mathcal{O}\left(\sqrt{\frac{\omega(n)}{\overline{\phi}}}\right)\right) \geq 1 - \mathcal{O}\left(\frac{1}{n^2}\right). \tag{22}$$

We shall use this observation in the following comparison:

$$\rho(\widehat{X}) = \rho(\mathbb{E}[Y \mid \widehat{X}]) \leq \mathbb{E}[\rho(Y) \mid \widehat{X}],$$

by Jensen's inequality. Set $S = \mathbb{E}[\rho(Y) \mid \widehat{X}]$, we shall show that it is also upper bounded by $\mathcal{O}(\sqrt{\omega(n)/\overline{\phi}})$.

First, note that $|Y|$ is element-wise dominated by the all-one matrix, hence $\rho(Y) \leq n$. Second, the sigma-algebra generated by S is contained in the sigma-algebra generated by $\widehat{X}$. Hence,

$$\mathbb{E}[\rho(Y) \mid S] = \mathbb{E}[\mathbb{E}[\rho(Y) \mid \widehat{X}] \mid S] = \mathbb{E}[S \mid S] = S.$$

Further, both Y and $\widehat{X}$ take only finitely many different values, and thus $\rho(Y)$ and S take values in a finite space. It therefore makes sense to consider, for $t > 0$, the function

$$\beta(\cdot) = \mathbb{P}(\rho(Y) > t \mid S = \cdot).$$

We have

$$S = \mathbb{E}[\rho(Y) \mid S] \leq \beta(S)n + (1 - \beta(S))t,$$

i.e.

$$\beta(S) \geq \frac{S - t}{n - t}.$$

Denote $\gamma = \mathbb{P}(S > t + 1)$ then

$$\mathbb{P}(\rho(Y) > t) = \mathbb{E}[\beta(S)] \geq \mathbb{E}[\beta(S)\mathbf{1}_{\{S > t+1\}}] \geq \frac{\gamma}{n - t}.$$

As a consequence, for $t = \mathcal{O}(\sqrt{\omega(n)/\overline{\phi}})$, by (22) we have

$$\mathbb{P}(S > t + 1) = \gamma \leq (n - t)\mathbb{P}(\rho(Y) > t) = (n - t)\mathcal{O}\left(\frac{1}{n^2}\right) = \mathcal{O}\left(\frac{1}{n}\right).$$

Therefore,

$$\rho(H - \mathbb{E}[H]) = \frac{\alpha}{\omega^2(n)}\rho(X) \leq (1 + o(1))\frac{\alpha}{\omega^2(n)}\rho(\widehat{X}) \leq \mathcal{O}\left(\sqrt{\frac{1}{\overline{\phi}\omega^3(n)}}\right),$$

where the first inequality stems from the fact that the order $1 + \varepsilon_n$ term in (9) holds uniformly over all vertices. Finally, due to (21),

$$\rho(H - \mathbb{E}[H]) = \mathcal{O}\left(\sqrt{\frac{\overline{\phi}}{\omega(n)}}\frac{1}{\overline{\phi}\omega(n)}\right) = \mathcal{O}\left(\frac{1}{\sqrt{\log(n)}}\right)\frac{1}{D}. \qquad \square$$

Proof of Lemma 5. To prove this theorem we claim that in the present setting, it suffices to show that with high probability,

$$(\widehat{H} - H)_{uv} = \varepsilon_{uv} H_{uv},$$

where, for some constant $\widehat{C}$ and all large enough n,

$$|\varepsilon_{uv}| \le \widehat{C}\varepsilon(n) \tag{23}$$

with

$$\varepsilon(n) := \sqrt{\frac{6}{\min_i \overline{M}_i} \frac{2\log(n)}{\omega(n)}} = \mathcal{O}\left(\sqrt{\frac{\log(n)}{D_1(n)}}\right) \to 0 \tag{24}$$

by assumption. Consequently, after an appeal to Lemma 10,

$$\rho(\widehat{H} - H) \le \rho(|\widehat{H} - H|) \le \widehat{C}\varepsilon(n)\rho(H).$$

Since, $H = \mathbb{E}[H] + H - \mathbb{E}[H]$, it follows from Lemmas 3 and 4 that

$$\rho(H) = \mathcal{O}\left(\frac{1}{D}\right).$$

Consider the difference

$$
\frac{1}{\widehat{D}_u}\frac{1}{\widehat{D}_v} - \frac{1}{\mathbb{E}\widehat{D}_u}\frac{1}{\mathbb{E}\widehat{D}_v}
$$
$$
= \frac{1}{\mathbb{E}\widehat{D}_u}\frac{1}{\mathbb{E}\widehat{D}_v}\frac{1}{1+(\widehat{D}_u - \mathbb{E}\widehat{D}_u)/\mathbb{E}\widehat{D}_u}\frac{1}{1+(\widehat{D}_v - \mathbb{E}\widehat{D}_v)/\mathbb{E}\widehat{D}_u} - \frac{1}{\mathbb{E}\widehat{D}_u}\frac{1}{\mathbb{E}\widehat{D}_v}
$$
$$
= \frac{1}{\mathbb{E}\widehat{D}_u \mathbb{E}\widehat{D}_v}\varepsilon_{uv};
$$

thus,

$$
\varepsilon_{uv} = \frac{\mathbb{E}\widehat{D}_u - \widehat{D}_u}{\mathbb{E}\widehat{D}_u} + \frac{\mathbb{E}\widehat{D}_v - \widehat{D}_v}{\mathbb{E}\widehat{D}_v} + \mathcal{O}\left(\left(\frac{\mathbb{E}\widehat{D}_u - \widehat{D}_u}{\mathbb{E}\widehat{D}_u}\right)^2\right) + \mathcal{O}\left(\left(\frac{\mathbb{E}\widehat{D}_v - \widehat{D}_v}{\mathbb{E}\widehat{D}_v}\right)^2\right).
$$

We quantify $(\mathbb{E}\widehat{D}_u - \widehat{D}_u)/\mathbb{E}\widehat{D}_u$. Since $\widehat{D}_u$ is a sum of Bernoulli random variables with mean

$$\mathbb{E}[\widehat{D}_u] = D_u \overline{M}_{\sigma_u}(1 - o(1)),$$

where the $o(1)$ term follows from (9), we have for $\varepsilon(n)$ as in (24), the Bernstein inequality (see (16)),

$$
\mathbb{P}\left(\left|\frac{\mathbb{E}\widehat{D}_u - \widehat{D}_u}{\mathbb{E}\widehat{D}_u}\right| > \varepsilon(n)\right) \le 2\exp\left(-\frac{\varepsilon^2(n)}{2+\varepsilon(n)/3}\mathbb{E}[\widehat{D}_u]\right)
$$
$$
= 2\exp\left(-\frac{\varepsilon^2(n)}{2+\varepsilon(n)/3}D_u\overline{M}_{\sigma_u}(1-o(1))\right)
$$
$$
\le 2\exp\left(-\frac{\varepsilon^2(n)}{3}\omega(n)\frac{\overline{M}_{\sigma_u}}{2}\right)
$$
$$
\le \frac{2}{n^2}. \tag{25}
$$

Invoking this we establish the union bound

$$\mathbb{P}\left(\left|\frac{\mathbb{E}\widehat{D}_1 - \widehat{D}_1}{\mathbb{E}\widehat{D}_1}\right| \le \varepsilon(n), \ldots, \left|\frac{\mathbb{E}\widehat{D}_n - \widehat{D}_n}{\mathbb{E}\widehat{D}_n}\right| \le \varepsilon(n)\right) \ge 1 - \sum_{u=1}^{n}\mathbb{P}\left(\left|\frac{\mathbb{E}\widehat{D}_u - \widehat{D}_u}{\mathbb{E}\widehat{D}_u}\right| > \varepsilon(n)\right)$$

$$\ge 1 - \frac{2}{n} \to 1 \quad \text{as } n \to \infty.$$

Hence,

$$E = \left\{\left|\frac{\mathbb{E}\widehat{D}_u - \widehat{D}_u}{\mathbb{E}\widehat{D}_u}\right| \le \varepsilon(n) \text{ for all } u \in V\right\}$$

holds with high probability. Thus, we establish (23): $|\varepsilon_{uv}| \le 2\varepsilon(n) + \mathcal{O}(\varepsilon^2(n)) \le \widehat{C}\varepsilon(n)$ with $\widehat{C}$ a large enough constant.

Henceforth, we condition on E. Then, for $u \ne v$,

$$\widehat{H}_{uv} - H_{uv} = \left(\frac{1}{\widehat{D}_u}\frac{1}{\widehat{D}_v} - \frac{1}{\mathbb{E}\widehat{D}_u\mathbb{E}\widehat{D}_v}\right)A_{uv} = \varepsilon_{uv}\frac{1}{\mathbb{E}\widehat{D}_u\mathbb{E}\widehat{D}_v}A_{uv} = \varepsilon_{uv}H_{uv}. \qquad \square$$

Proof of Lemma 6. We define

$$\varepsilon(n) = \frac{1}{\log^{1/3}(n)} \tag{26}$$

and we shall call a vertex u good if $|\mathbb{E}\widehat{D}_u - \widehat{D}_u| \le \varepsilon(n)\mathbb{E}\widehat{D}_u$. We use this definition to split

$$\left|\frac{1}{\widehat{D}_u}\frac{1}{\widehat{D}_v} - \frac{1}{\mathbb{E}\widehat{D}_u}\frac{1}{\mathbb{E}\widehat{D}_v}\right|A_{uv} = M_{uv} + M_{uv}^{\mathrm{c}} + M_{uv}^{\mathrm{r}} - M_{uv}^{\mathrm{cr}}, \tag{27}$$

where

$$M_{uv} = \left|\frac{1}{\widehat{D}_u}\frac{1}{\widehat{D}_v} - \frac{1}{\mathbb{E}\widehat{D}_u}\frac{1}{\mathbb{E}\widehat{D}_v}\right|A_{uv}\mathbf{1}_{\{u \text{ and } v \text{ good}\}},$$

$$M_{uv}^{\mathrm{c}} = \left|\frac{1}{\widehat{D}_u}\frac{1}{\widehat{D}_v} - \frac{1}{\mathbb{E}\widehat{D}_u}\frac{1}{\mathbb{E}\widehat{D}_v}\right|A_{uv}\mathbf{1}_{\{v \text{ bad}\}},$$

$$M_{uv}^{\mathrm{r}} = \left|\frac{1}{\widehat{D}_u}\frac{1}{\widehat{D}_v} - \frac{1}{\mathbb{E}\widehat{D}_u}\frac{1}{\mathbb{E}\widehat{D}_v}\right|A_{uv}\mathbf{1}_{\{u \text{ bad}\}},$$

$$M_{uv}^{\mathrm{cr}} = \left|\frac{1}{\widehat{D}_u}\frac{1}{\widehat{D}_v} - \frac{1}{\mathbb{E}\widehat{D}_u}\frac{1}{\mathbb{E}\widehat{D}_v}\right|A_{uv}\mathbf{1}_{\{u \text{ and } v \text{ bad}\}}.$$

We shall show that all terms in (27) have a negligible spectral radius compared to $\Delta(P)$. First note that the difference

$$\frac{1}{\widehat{D}_u}\frac{1}{\widehat{D}_v} - \frac{1}{\mathbb{E}\widehat{D}_u}\frac{1}{\mathbb{E}\widehat{D}_v}$$

may be written as

$$\frac{1}{\mathbb{E}\widehat{D}_u\mathbb{E}\widehat{D}_v}\varepsilon_{uv},$$

where

$$\varepsilon_{uv} = \frac{\mathbb{E}\widehat{D}_u - \widehat{D}_u}{\mathbb{E}\widehat{D}_u} + \frac{\mathbb{E}\widehat{D}_v - \widehat{D}_v}{\mathbb{E}\widehat{D}_v} + \mathcal{O}\left(\left(\frac{\mathbb{E}\widehat{D}_u - \widehat{D}_u}{\mathbb{E}\widehat{D}_u}\right)^2\right) + \mathcal{O}\left(\left(\frac{\mathbb{E}\widehat{D}_v - \widehat{D}_v}{\mathbb{E}\widehat{D}_v}\right)^2\right).$$

Now, similarly as in the proof of Lemma 5, there exists a constant $\widehat{C}$ such that $\varepsilon_{uv} \le \widehat{C}\varepsilon(n)$ if both u and v are good. Consequently, $\rho(M) \le \widehat{C}\varepsilon(n)\rho(\widehat{H})$.

Next we analyse the other terms in (27). We start with M^c. The idea is that, although now

$$\left| \frac{1}{\widehat{D}_u} \frac{1}{\widehat{D}_v} - \frac{1}{\mathbb{E}\widehat{D}_u} \frac{1}{\mathbb{E}\widehat{D}_v} \right| = \mathcal{O}\left(\frac{1}{\mathbb{E}\widehat{D}_u} \frac{1}{\mathbb{E}\widehat{D}_v} \right),$$

the total number of nonzero elements in a column of M^c is very small, so that its spectral radius indeed vanishes upon division by $\Delta(P)$. We note that

$$(A_{uv} \mathbf{1}_{\{v \text{ bad}\}})_{u,v} = (A_{uv} \mathbf{1}_{\{u \text{ bad}\}})_{u,v}^{\top},$$

so that a similar statement holds for the maximal row sum of M^r. Obviously, $M^{cr} \leq M^c$ as do their spectral radii.

As a consequence of these observations, it thus suffices to prove our claim for M^c. To do so, we proceed in three steps. First, we show that

$$\mathbb{P}(\mathcal{E}_1) = \mathbb{P}\left(\left\{ \text{for all } u : \left| \frac{\mathbb{E}\widehat{D}_u - \widehat{D}_u}{\mathbb{E}\widehat{D}_u} \right| \leq \frac{1}{2} \right\} \right) \geq 1 - \frac{2}{n^2}. \tag{28}$$

From which it follows, after a short computation, that, with probability larger than $1 - 2/n^2$, for all u, v,

$$\left| \frac{1}{\widehat{D}_u} \frac{1}{\widehat{D}_v} - \frac{1}{\mathbb{E}\widehat{D}_u} \frac{1}{\mathbb{E}\widehat{D}_v} \right| \leq 3 \frac{1}{\mathbb{E}\widehat{D}_u} \frac{1}{\mathbb{E}\widehat{D}_v}.$$

Keeping this in mind, it thus suffices to demonstrate that $(A_{uv} \mathbf{1}_{\{v \text{ bad}\}})_{uv}$ has a spectral radius much smaller than the spectral radius of A. The column sum in the former is equal to the number of bad neighbours in a vertex. That is, the spectral radius is bounded by $\max_u X_u$, where, for $u \in V$,

$$X_u = \sum_{v \in \mathcal{N}(u)} Z_v \tag{29}$$

with $Z_v = \mathbf{1}_{\{v \text{ is bad}\}}$. Caution is needed here as the indicator functions in (29) are not independent.

In the second step we shall show that with high probability the number of edges between vertices in the neighbourhood of u is negligible compared to the expected degree of vertex u. That is,

$$\mathbb{P}(\mathcal{E}_2(u)) = \mathbb{P}\left(\left\{ \sum_{x,y \in \mathcal{N}(u)} A_{xy} \leq \frac{1}{2} \varepsilon(n)\omega(n) \right\} \right) \geq 1 - \frac{2}{n^2}, \tag{30}$$

where $\omega(n)$ is defined in (18). Hence, except for possibly $\frac{1}{4}\varepsilon(n)\omega(n)$ of them, the variables in (29) form an independent set (conditional on not having any neighbours among $\mathcal{N}(u)$).

The last step consists in showing that this leads to

$$\mathbb{P}(X_u > \varepsilon(n)\mathcal{O}(\mathbb{E}\widehat{D}_u) \mid \mathcal{E}_1, \mathcal{E}_2) = o\left(\frac{1}{n} \right). \tag{31}$$

The assertion follows now straightforwardly. With high probability, we have

$$\sum_v M_{uv}^c \leq 3 \frac{1}{\mathbb{E}\widehat{D}_u} \max_v \frac{1}{\mathbb{E}\widehat{D}_v} X_u$$

$$\leq 3 \frac{1}{\mathbb{E}\widehat{D}_u} \max_v \frac{1}{\mathbb{E}\widehat{D}_v} \varepsilon(n)\mathcal{O}(\mathbb{E}\widehat{D}_u)$$

$$\leq \mathcal{O}\left(\frac{\varepsilon(n)}{\min_v \mathbb{E}\widehat{D}_v}\right)$$
$$= \mathcal{O}\left(\frac{\varepsilon(n)}{\omega(n)}\right).$$

Now $\overline{D} = \mathcal{O}(\omega(n))$ since $D_1^2(n)/\overline{D}(n) = \Omega(\log(n))$. Consequently, due to the choice of $\varepsilon(n)$ in (26),

$$\rho(M^c) = \mathcal{O}\left(\frac{\varepsilon(n)}{\omega(n)}\right) = \mathcal{O}\left(\frac{1}{\log^{1/3}(n)}\right)\frac{1}{\overline{D}(n)} = o_n(1)\frac{1}{\overline{D}(n)}.$$

The first step, i.e. demonstrating (28), is easily carried out: fix $u \in V$ and use Bernstein's inequality (16) to verify the bound

$$\mathbb{P}\left(\left|\frac{\mathbb{E}\widehat{D}_u - \widehat{D}_u}{\mathbb{E}\widehat{D}_u}\right| > \frac{1}{2}\right) \leq 2\exp\left(-\frac{3}{26}\mathbb{E}[\widehat{D}_u]\right).$$

Now, for large enough n, $\mathbb{E}\widehat{D}_u \geq \overline{M}C_{B,M}\log n$, and by assumption, $C_{B,M}$ from (7) is so large that $\frac{3}{26}\mathbb{E}[\widehat{D}_u] > 2\log(n)$. Hence,

$$\mathbb{P}\left(\frac{1}{2\mathbb{E}\widehat{D}_u} \leq \widehat{D}_u \leq \frac{3}{2\mathbb{E}\widehat{D}_u}\right) \geq 1 - \frac{2}{n^2}.$$

We proceed with the second step, i.e. (30). Set $\overline{M} = \max_i \overline{M}_i$, $\overline{B} = \max_{i,j} B_{ij}$. Set $C = \max\{1/2\overline{M}, 5\overline{M}^2, \overline{B}\}$. Consider, conditional on $\widehat{D}_u \leq 2\mathbb{E}\widehat{D}_u$,

$$\sum_{x,y \in \mathcal{N}(u)} A_{xy} = \sum_{x,y \in \mathcal{N}(u)} \text{Ber}\left(B_{\sigma_x \sigma_y}\frac{\phi_x \phi_y \omega(n)}{g(n)}\right)$$
$$\leq \text{bin}\left(4(\mathbb{E}\widehat{D}_u)^2, \overline{B}\frac{\phi_x \phi_y \omega(n)}{g(n)}\right)$$
$$\leq \text{bin}\left(5\overline{M}^2\phi_u^2\omega^2(n), \overline{B}\frac{\phi_x \phi_y \omega(n)}{g(n)}\right)$$
$$\leq \text{bin}\left(5\overline{M}^2\phi_n^2\omega^2(n), \overline{B}\frac{\phi_n^2 \omega(n)}{g(n)}\right)$$
$$\leq \text{bin}\left(C\phi_n^2\omega^2(n), C\frac{\phi_n^2 \omega(n)}{g(n)}\right),$$

where ϕ_u and $g(n)$ are defined in (17) and (19), respectively. We now show that

$$\mathbb{P}\left(\text{bin}\left(C\phi_n^2\omega^2(n), C\frac{\phi_n^2\omega(n)}{g(n)}\right) \geq \frac{1}{2}\varepsilon(n)\omega(n)\right) = o\left(\frac{1}{n}\right).$$

First, note that

$$\mathbb{P}\left(\text{bin}\left(C\phi_n^2\omega^2(n), C\frac{\phi_n^2\omega(n)}{g(n)}\right) \geq \frac{1}{2}\varepsilon(n)\omega(n)\right) \leq \binom{C\phi_n^2\omega^2(n)}{\frac{1}{2}\varepsilon(n)\omega(n)}\left(C\frac{\phi_n^2\omega(n)}{g(n)}\right)^{\varepsilon(n)\omega(n)/2}. \tag{32}$$

Using the fact that

$$\binom{n}{k} \leq \left(\frac{ne}{k}\right)^k,$$

we have

$$\binom{C\phi_n^2\omega^2(n)}{\frac{1}{2}\varepsilon(n)\omega(n)} \leq \left(2Ce\frac{\phi_n^2\omega(n)}{\varepsilon(n)}\right)^{\varepsilon(n)\omega(n)/2}$$

$$= \exp\left(\frac{1}{2}\varepsilon(n)\omega(n)\log\left(2Ce\frac{\phi_n^2\omega(n)}{\varepsilon(n)}\right)\right)$$

$$\leq \exp\left(\frac{c}{2}\varepsilon(n)\omega(n)\log(g(n)) + \frac{1}{2}\varepsilon(n)\omega(n)\log(2Ce)\right),$$

where $c < \frac{1}{2}$ from (7) is such that $\phi_n^2\omega^2(n)/\log^{2/3}(n)n^c \to 0$ (and thus $\phi_n^2\omega(n)/g^c(n) = o_n(1)/\log^{1/3}(n) \leq \varepsilon(n)$, since $g(n) = \Theta(n)$ in the particular setting of this lemma). Write

$$\left(C\frac{\phi_n^2\omega(n)}{g(n)}\right)^{\varepsilon(n)\omega(n)/2} = \exp\left(-\frac{1}{2}\varepsilon(n)\omega(n)\log\left((g(n))^{1-c}\frac{(g(n))^c}{C\phi_n^2\omega(n)}\right)\right)$$

$$\leq \exp\left(-\frac{1}{2}\varepsilon(n)\omega(n)\log((g(n))^{1-c})\right)$$

$$= \exp\left(-\frac{1-c}{2}\varepsilon(n)\omega(n)\log(g(n))\right)$$

if n large enough. Combining these estimates, we see that (32) may be bounded from above by

$$\exp\left(-\frac{1-2c}{2}\varepsilon(n)\omega(n)\log(g(n)) + \frac{1}{2}\varepsilon(n)\omega(n)\log(2Ce)\right)$$

$$\leq \left(-\frac{1-2c}{4}\varepsilon(n)\omega(n)\log(g(n))\right)$$

since $g(n) \geq n \gg 2Ce$. Finally, since $((1-2c)/4)\varepsilon(n)\omega(n) \geq 2$, for large n,

$$\mathbb{P}(\mathcal{E}_2) = 1 - \mathbb{P}\left(\sum_{x,y \in \mathcal{N}(u)} A_{xy} \geq \frac{1}{2}\varepsilon(n)\omega(n)\right) \geq 1 - e^{-\log(g^2(n))} \geq 1 - \frac{1}{n^2},$$

i.e. (30).

We proceed with the last step, i.e. establishing (31). Write

$$X_u = \sum_{v \in \mathcal{N}(u):\, \mathcal{N}(v) \cap \mathcal{N}(u) \neq \varnothing} Z_v + \sum_{v \in \mathcal{N}(u):\, \mathcal{N}(v) \cap \mathcal{N}(u) = \varnothing} Z_v.$$

We already know from (30) that the first sum is smaller than $\frac{1}{2}\varepsilon(n)\omega(n)$, with high probability. The variables in the second sum, $\{Z_v\}_{v \in \mathcal{N}(u):\, \mathcal{N}(v) \cap \mathcal{N}(u) = \varnothing}$, are independent. For such a vertex $v \in \mathcal{N}(u)$ that has no neighbour with u in common, we have $\widehat{D}_v = d'_v + 1$, where

$$d'_v = \sum_{x \notin \mathcal{N}(u),\, x \neq u} \mathrm{Ber}\left(B_{\sigma_v\sigma_x}\frac{D_vD_x}{n\overline{D}}\right),$$

the degree of v outside $\mathcal{N}(u) \cup \{u\}$. We show that v is a good vertex with high probability by proving that d'_v concentrates on its mean which on its turn is close to $\mathbb{E}[\widehat{D}_v]$. First, define

$$\mathbb{E}_*[\cdot] := \mathbb{E}[\cdot \mid \mathcal{N}(u), \mathcal{E}_2, \widehat{D}_u \leq 2\mathbb{E}[\widehat{D}_u]]$$

then

$$\mathbb{E}_*[d'_v] = \sum_{x \notin \mathcal{N}(u),\, x \neq u} B_{\sigma_v \sigma_x} \frac{D_v D_x}{n \overline{D}}$$

$$= \sum_{x \neq v} B_{\sigma_v \sigma_x} \frac{D_v D_x}{n \overline{D}} - \sum_{x \in \mathcal{N}(u) \cup \{u\},\, x \neq v} B_{\sigma_v \sigma_x} \frac{D_v D_x}{n \overline{D}}$$

$$\geq \mathbb{E}[\widehat{D}_v] - \mathcal{O}\left(\frac{\phi_n^2 \omega(n)}{g(n)} \right) \mathbb{E}[\widehat{D}_u]$$

$$= \mathbb{E}[\widehat{D}_v] - \mathcal{O}\left(\frac{\phi_n^3 \omega^2(n)}{g(n)} \right)$$

$$= \mathbb{E}[\widehat{D}_v] - o_n(1).$$

Second, we use Bernstein's inequality (16) to prove that d'_v concentrates around $\mathbb{E}[\widehat{D}_v]$ up to a factor $\varepsilon(n)$ as in (26), i.e.

$$\mathbb{P}(d'_v \geq (1 + \varepsilon(n))\mathbb{E}\widehat{D}_v \mid \mathcal{N}(u), \mathcal{E}_2, \widehat{D}_u \leq 2\mathbb{E}[\widehat{D}_u])$$

$$\leq \exp\left(-\frac{(\varepsilon(n)\mathbb{E}_* d'_v + (1 + \varepsilon(n))o_n(1))^2}{2(\mathbb{E}_* d'_v + 1/3(\varepsilon(n)\mathbb{E}_* d'_v + (1 + \varepsilon(n))o_n(1)))} \right)$$

$$\leq \exp\left(-\frac{(\varepsilon(n)\mathbb{E}_* d'_v)^2 (1 + o_n(1)/\varepsilon(n)\mathbb{E}_* d'_v)}{4\mathbb{E}_* d'_v} \right)$$

$$\leq \exp(-C\varepsilon^2(n)\log(n)),$$

where we *redefined* $C = \frac{1}{8}$. Similarly,

$$\mathbb{P}(d'_v \leq (1 - \varepsilon(n))\mathbb{E}\widehat{D}_v \mid \mathcal{N}(u), \mathcal{E}_2, \widehat{D}_u \leq 2\mathbb{E}[\widehat{D}_u]) \leq \exp(-C\varepsilon^2(n)\log(n)).$$

Hence, each vertex $v \in \mathcal{N}(u)$ that has no neighbour with u in common is thus a good vertex with probability $2\exp(-C\varepsilon^2(n)\log(n))$. Consequently, conditional on $\mathcal{N}(u), \mathcal{E}_2, \widehat{D}_u \leq 2\mathbb{E}[\widehat{D}_u]$,

$$\sum_{v \in \mathcal{N}(u):\, \mathcal{N}(v) \cap \mathcal{N}(u) = \varnothing} Z_v \leq \operatorname{bin}(2\mathbb{E}\widehat{D}_u, 2\exp(-C\varepsilon^2(n)\log n)).$$

We have

$$\mathbb{P}\big(\operatorname{bin}(2\mathbb{E}\widehat{D}_u, 2\exp(-C\varepsilon^2(n)\log n)) \geq \tfrac{1}{2}\varepsilon(n)\mathbb{E}\widehat{D}_u\big)$$

$$\leq \binom{2\mathbb{E}\widehat{D}_u}{\frac{1}{2}\varepsilon(n)\mathbb{E}\widehat{D}_u} (2\exp(-C\varepsilon^2(n)\log n))^{\varepsilon(n)\mathbb{E}\widehat{D}_u/2}$$

$$\leq \left(\frac{4\mathrm{e}}{\varepsilon(n)} \right)^{\varepsilon(n)\mathbb{E}\widehat{D}_u/2} (2\exp(-C\varepsilon^2(n)\log n))^{\varepsilon(n)\mathbb{E}\widehat{D}_u/2}$$

$$= \exp\left(\frac{1}{2}\varepsilon(n)\mathbb{E}\widehat{D}_u \left(\log \frac{8e}{\varepsilon(n)} - C\varepsilon^2(n)\log n \right) \right)$$

$$= o\left(\frac{1}{n} \right),$$

since $\varepsilon(n) = 1/\log^{1/3}(n)$. Hence,

$$\mathbb{P}\left(X_u > \frac{1}{2}\varepsilon(n)(\omega(n) + \mathbb{E}\widehat{D}_u) \,\bigg|\, \mathcal{E}_1, \mathcal{E}_2 \right) = o\left(\frac{1}{n} \right).$$

The last step (i.e. (31)) is completed by noting that $\omega(n) = \mathcal{O}(\mathbb{E}\widehat{D}_u)$.   □

Proof of Lemma 7. All matrices in

$$W = (\widehat{H} - H) + (H - \mathbb{E}[H]) + (\mathbb{E}[H] - P)$$

are real and symmetric, hence, combining Lemmas 2–6,

$$\rho(W) \le \rho(\widehat{H} - H) + \rho(H - \mathbb{E}[H]) + \rho(\mathbb{E}[H] - P) = o_n(1)\frac{1}{\overline{D}(n)}.$$

Employing Lemma 11 yields that to each eigenvector $\widehat{x}$ of $\widehat{H} = P + W$ corresponds an eigenvector x of P such that

$$\widehat{x} \cdot x \ge \sqrt{1 - (\frac{\rho(W)}{\Delta(P)})^2} = 1 - \mathcal{O}\left(\left(\frac{\rho(W)}{\Delta(P)}\right)^2\right) = 1 - o_n(1),$$

since $\Delta(P) = \Omega(1/\overline{D}(n))$. $\qquad\square$

Proof of Lemma 8. Invoking Lemma 7, to each $\widehat{x}_i$ (with eigenvalue $\widehat{\lambda}_i$) there exists a normed eigenvector x_i (with eigenvalue λ_i) of P such that

$$\widehat{x}_i \cdot x_i = 1 - f_i(n)$$

with $f_i(n) = o_n(1)$. We claim that all λ_i are larger than 0 (note that we refer here to a set of $\widehat{L}$ eigenvalues). This can be seen as follows. From Lemma 2 we know that the first L eigenvalues of P are of order $1/\overline{D}$ and all other eigenvalues are 0. By Lemma 11,

$$|\lambda_i - \widehat{\lambda}_i| \le \rho(W) \ll \frac{1}{\overline{D}},$$

hence the first L eigenvalues of $\widehat{H}$ are also of order $\Omega(1/\overline{D}) - \mathcal{O}(\rho(W)) = \Omega(1/\overline{D})$, and the other $n - L$ are of order $\mathcal{O}(\rho(W))$. Now, the $\widehat{L}$ eigenvalues of $\widehat{H}$ that are picked in step (i) of Algorithm 1 are precisely those whose absolute eigenvalue exceeds $f(n)/\overline{D}_{\text{average}} = \Omega(f(n)/\overline{D}) \gg \rho(W)$, by the construction of f in Section 3. Hence, those eigenvalues must necessarily be of order $\Omega(1/\overline{D})$ (i.e. they are indeed nonzero) and $L = \widehat{L}$ with high probability.

Since x_i corresponds to a nonzero eigenvalue, it follows from the proof of Lemma 2 that x_i is constant on each block, i.e. $x_i(u) = x_i(v)$ if $\sigma_u = \sigma_v$. Let $x_i^{(k)}$ be the value of x_i on block $k \in S$. Set

$$t_k = \sqrt{n}(x_1^{(k)}, \ldots, x_L^{(k)}).$$

Then

$$\frac{1}{n}|\{u \in V : \|\sqrt{n}\widehat{z}_u - t_{\sigma_u}\|^2 \ge T^2\}| \le \frac{1}{nT^2}\sum_{m=1}^{n}\|\sqrt{n}\widehat{z}_u - t_{\sigma_u}\|^2$$

$$= \frac{1}{T^2\sum_{u=1}^{n}\|(\widehat{x}_1^{(u)}, \ldots, \widehat{x}_L^{(u)}) - (x_1^{(\sigma_u)}, \ldots, x_L^{(\sigma_u)})\|^2}$$

$$= \frac{1}{T^2\sum_{k=1}^{L}\|\widehat{x}_k - x_k\|^2}$$

$$= \frac{1}{T^2\sum_{k=1}^{L}f_k(n)},$$

and to complete the proof, let $T = (\sum_{k=1}^{L}f_k(n))^{1/3} = \mathcal{O}((\rho(W)/\Delta(P))^{2/3}) = o_n(1)$. $\qquad\square$

Proof of Lemma 9. Below we shall make a spectral decomposition in terms of L orthonormal eigenvectors of Z that span the union of all eigenspaces corresponding to nonzero eigenvalues. Recall from the proof of Lemma 2 how we can obtain the eigenvectors of Z from the eigenvectors of P.

Recall that, by construction, $\{\widehat{x}_i\}_{i=1}^L$ are orthonormal eigenvectors of $\widehat{H}$ corresponding to nonzero eigenvalues spanning an L-dimensional space. Recall further from the proof of Lemma 8 that the corresponding eigenvectors $\{x_i\}_{i=1}^L$ of P are associated with nonzero eigenvalues. Lemma 11(ii) entails that the space spanned by those $\{x_i\}_{i=1}^L$ has also dimension L. And Lemma 7 implies that $\{x_i\}_{i=1}^L$ become an orthonormal set for n tending to ∞ (because they become more and more aligned with the orthonormal set $\{\widehat{x}_i\}_{i=1}^L$).

Let, as in the proof of Lemma 8, $x_i^{(k)}$ be the value of x_i on block $k \in S$. Note that $\sum_k n\alpha_k(x_i^{(k)})^2 = 1$ for $i \in \{1, \dots, L\}$. Setting $y_i = \sqrt{n}(x_i^{(1)}, \dots, x_i^{(K)})^\top$, we see that each y_i is a *normalized* eigenvector of Z in the sense that $\sum_k \alpha_k(y_i(k))^2 = 1$. Now, assume for a contradiction that $|t_k - t_l| \to 0$ as $n \to \infty$, i.e.

$$\sum_{i=1}^L |\sqrt{n}x_i^{(k)} - \sqrt{n}x_i^{(l)}|^2 = \sum_{i=1}^L |y_i(l) - y_i(k)|^2 \to 0.$$

We conclude that there exist orthonormal eigenvectors of Z, $\{\overline{y}_1, \dots, \overline{y}_L\}$ (with eigenvalues $\{\lambda_i\}_{i=1}^L$ after a possible relabelling of indices), that span the range of Z, such that

$$\overline{y}_u(k) = \overline{y}_u(l) \quad \text{for all } u.$$

The other $K - L$ eigenvectors have 0 as an eigenvalue.

To proceed, consider the matrix

$$N = \left(\sqrt{\alpha_u} \frac{B_{uv}}{\overline{M}_u \overline{M}_v} \sqrt{\alpha_v}\right)_{u,v}.$$

If $(x(1), \dots, (x(K))^\top$ is an eigenvector of Z then $(\sqrt{\alpha_1}x(1), \dots, \sqrt{\alpha_K}x(K))^\top$ is an eigenvector of N, as is easily verified. Hence, N has $\{(\sqrt{\alpha_1}\overline{y}_i(1), \dots, \sqrt{\alpha_K}\overline{y}_i(K))^\top\}_{i=1}^L$ as eigenvectors corresponding to nonzero eigenvalues, and $K - L$ eigenvectors with 0 as an eigenvalue (which do not contribute to the spectral decomposition of N). Hence,

$$N = \left(\sum_{l=1}^L \sqrt{\alpha_u}\overline{y}_l(u)\lambda_l\sqrt{\alpha_v}\overline{y}_l(v)\right)_{u,v}.$$

Thus, for all u,

$$\frac{B_{ku}}{\overline{M}_k \overline{M}_u} = \sum_m \overline{y}_m(k)\lambda_m\overline{y}_m(u) = \sum_m \overline{y}_m(l)\lambda_m\overline{y}_m(u) = \frac{B_{lu}}{\overline{M}_l \overline{M}_u},$$

violating assumption (15).   □

7.1. Comparison to spectral analysis on the adjacency matrix

Proof of Theorem 2. This proof leans strongly on ideas borrowed from [22], where graphs *without* a community structure were considered. Parts of their proof carry through for the DC-SBM considered here. Note that $\lim_{n\to\infty} g(n)/n = 1$.

By definition, we require, without lose of generality, $D_1 \leq D_2 \leq \cdots \leq D_n$. However, we obtain the same graph (with now a decreasing degree sequence) by a rearrangement of indices, if we set

$$\phi_u = \begin{cases} \dfrac{\phi_1}{u} & \text{if } u \leq 1 \leq k = n^\beta, \\ 1 & \text{if } u > n^\beta, \end{cases}$$

where $\phi_1 = n^{\gamma+\beta}$, and $D_u = \phi_u \omega(n)$ (with ω as in (18)). Then

$$\sigma_u = \begin{cases} 1 & \text{if } u \leq n/2 \\ 2 & \text{if } u > n/2. \end{cases}$$

Denote a sample of the random graph by G. We decompose G into the following graphs (exactly as in [22]):

- G_1, which is a union of vertex disjoint stars $S_1, \ldots, S_k$. Star S_u has as its centre node u and as leaves those vertices from among $\{k+1, \ldots, n\}$ adjacent to u, but not adjacent to $\{1, \ldots, u-1\}$;

- G_1' is the graph consisting of all edges of G with one endpoint in $\{1, \ldots, k\}$ and the other endpoint in $\{k+1, \ldots, n\}$, except for those edges in G_1;

- G_2 is the subgraph of G, which is induced by $\{1, \ldots, k\}$;

- G_3 is the subgraph of G, which is induced by $\{k+1, \ldots, n\}$.

Further, let F_u be the subset of vertices in $\{k+1, \ldots, n\}$ that are adjacent to $\{1, \ldots, u-1\}$ and let C be a constant, independent of n, whose value might change along the course of the proof.

We claim that $\widehat{d_u}$, the degree of vertex u in G_1, concentrates around its mean. Indeed, consider

$$\widehat{d_u} = \sum_{l=k+1}^{n} \text{Ber}\left(\frac{D_u D_l}{g(n)\omega(n)} B_{\sigma_u \sigma_l} \right) - \sum_{l \in F_u} \text{Ber}\left(\frac{D_u D_l}{g(n)\omega(n)} B_{\sigma_u \sigma_l} \right),$$

where g is defined in (19). Then

$$d_u = \mathbb{E}[\widehat{d_u}] \geq \frac{\omega(n)\phi_u}{g(n)} \left(\sum_{l=k+1}^{n} B_{\sigma_u \sigma_l} - C\mathbb{E}[|F_u|] \right),$$

which we bound from below by estimating $\mathbb{E}[|F_u|]$ for $u \leq k = n^\beta$: for large enough n,

$$\mathbb{E}[|F_u|] = \sum_{l=k+1}^{n} \sum_{v=1}^{u-1} \frac{D_l D_v}{\omega(n) g(n)} B_{\sigma_u \sigma_l}$$

$$\leq C \frac{\omega(n)\phi_1}{g(n)} \sum_{l=k+1}^{n} \sum_{v=1}^{u-1} \frac{1}{v}$$

$$\leq C\omega(n)\phi_1 \frac{n - n^\beta}{g(n)} n^\beta$$

$$\leq C\omega(n)n^{\gamma+2\beta},$$

after recalling the special choice for the degree sequence.

Consequently, we have

$$\frac{n}{g(n)}\frac{B_{11}+B_{12}}{2}D_u \geq d_u \geq \frac{B_{11}+B_{12}}{2}D_u\left(\frac{n}{g(n)}-C\frac{\omega(n)n^{\gamma+2\beta}}{g(n)}\right).$$

Invoking large deviation theory on $\widehat{d}_u$ (which is a sum of Bernoulli random variables), we deduce that

$$\mathbb{P}(|\widehat{d}_u-d_u|>\sqrt{c'd_u\log n}) \leq \frac{2}{n^{c'/4}} \quad \text{for } c'>0 \text{ a constant.} \tag{33}$$

We take $c'=8$ to establish (33) uniformly over all vertices. We next investigate $\Delta(G_1)$, the smallest gap between different eigenvalues of G_1. This graph is the union of vertex disjoint stars with degree $\widehat{d}_u$ so that its spectrum is given by

$$\left\{\pm\sqrt{\widehat{d}_1-1},\ldots,\pm\sqrt{\widehat{d}_k-1}\right\}.$$

We claim that

$$\Delta(G_1) \geq C\sqrt{\omega(n)}n^{(\gamma-3\beta)/2} \to \infty \tag{34}$$

with high probability. Indeed, define

$$x_u^{\pm}=d_u\pm\sqrt{c'd_u\log n}$$

and note that with high probability $\widehat{d}_u \geq x_u^-$ and $\widehat{d}_{u+1} \leq x_{u+1}^+$. To investigate the difference $x_u^- - x_{u+1}^+$, we first bound d_u-d_{u+1} from below, i.e.

$$\begin{aligned}
d_u-d_{u+1} &\geq \frac{B_{11}+B_{12}}{2}\omega(n)\phi_1\left(\frac{n/g(n)}{u(u+1)}-C\frac{\omega(n)n^{\gamma+2\beta}}{n}\frac{1}{u}\right)\\
&= \frac{B_{11}+B_{12}}{2}\omega(n)\phi_1\frac{1}{u}\left(\frac{n/g(n)}{u+1}-C\frac{\omega(n)n^{\gamma+2\beta}}{n}\right)\\
&\geq \frac{B_{11}+B_{12}}{2}\omega(n)\phi_1\frac{1}{u}\left(\frac{n/g(n)}{n^\beta+1}-C\frac{\omega(n)n^{\gamma+2\beta}}{n}\right)\\
&\geq \frac{B_{11}+B_{12}}{4}\omega(n)\phi_1\frac{1}{n^\beta}\frac{1}{n^\beta}\\
&= \frac{B_{11}+B_{12}}{4}\frac{\omega(n)n^{\gamma+\beta}}{n^{2\beta}}\\
&= \frac{B_{11}+B_{12}}{4}\omega(n)n^{\gamma-\beta}.
\end{aligned}$$

Next we show that the $\sqrt{d_u\log n}$ terms are negligible, i.e.

$$\begin{aligned}
\sqrt{d_u\log n} &\leq \sqrt{\frac{((B_{11}+B_{12})/2)n}{g(n)D_u\log n}}\\
&\leq C\sqrt{\omega(n)\log(n)n^{\gamma+\beta}}\\
&\leq C\omega(n)n^{(\gamma+\beta)/2}\\
&\ll \omega(n)n^{\gamma-\beta}
\end{aligned}$$

due to (12). Hence,

$$x_u^- - x_{u+1}^+ \geq C\omega(n)n^{\gamma-\beta}.$$

As a consequence,

$$
\begin{aligned}
\Delta(G_1) &\geq \min_{u\in\{1,\dots,k\}} \left(\sqrt{\widehat{d}_u - 1} - \sqrt{\widehat{d}_{u+1} - 1} \right) \\
&\geq \min_{u\in\{1,\dots,k\}} \left(\sqrt{x_u^- - 1} - \sqrt{x_{u+1}^+ - 1} \right) \\
&= \min_{u\in\{1,\dots,k\}} \left(\frac{x_u^- - x_{u+1}^+}{\sqrt{x_u^- - 1} + \sqrt{x_{u+1}^+ - 1}} \right) \\
&\geq C \frac{\omega(n)n^{\gamma-\beta}}{\sqrt{\omega(n)}n^{(\gamma+\beta)/2}} \\
&= C\sqrt{\omega(n)}n^{(\gamma-3\beta)/2},
\end{aligned}
$$

i.e. (34).

We continue with an inspection of G_1', i.e. we focus on $\widehat{m}_u = \widehat{D}_u|_{\widehat{G}_1'}$, the degree of vertex u in G_1', and show that

$$\widehat{m}_u \leq 2c' \log n, \tag{35}$$

with high probability (here, m_u is the expectation of $\widehat{m}_u$). We shall use this in combination with the fact that the spectral radius of a graph is bounded by its largest degree. Write

$$\widehat{m}_u = \sum_{l\in F_u} \text{Ber}\left(\frac{\phi_u\omega(n)}{g(n)} B_{1\sigma_l} \right).$$

This expression allows us to deduce an upper bound for m_u,

$$m_u = \mathbb{E}[\widehat{m}_u] \leq C\mathbb{E}[F_u]\frac{\phi_u\omega(n)}{g(n)} \leq C\omega(n)n^{\gamma+2\beta}n^{\gamma+\beta}\frac{1}{u}\frac{\omega(n)}{g(n)} \leq C\omega^2(n)\frac{n^{2\gamma+3\beta}}{n},$$

which tends to 0 due to (11). Standard bounds for Bernoulli random variables yield

$$\mathbb{P}(|m_u - \widehat{m}_u| \leq c'\log n) \leq 2\exp\left(-\frac{(c'\log n)^2}{2(m_u + c'\log(n)/3)} \right) \leq 2\exp\left(-\frac{1}{4}c'\log n \right) = \frac{2}{n^{c'/4}}.$$

We conclude that, with probability at least $1 - 2/n^2$,

$$\widehat{m}_u \leq m_u + c'\log n \leq 2c'\log n,$$

i.e. (35) holds. An identical estimate holds when $u > k$.

We next bound the number of edges in G_2, denoted by $E(G_2)$. The square root of $E(G_2)$ is an upper bound for the spectral radius of G_2. Thus,

$$\mathbb{E}[|E(G_2)|] = C\sum_{u=1}^{k}\sum_{v=1}^{k} \frac{\phi_u\phi_v\omega(n)}{g(n)} \leq C\frac{n^{\gamma+\beta}n^{\gamma+\beta}\omega(n)n^\beta n^\beta}{g(n)} \leq C\frac{n^{2\gamma+4\beta}}{n},$$

vanishing for large n. Again, upon invoking standard large deviation theory, we have, with probability at least $1 - 2/n^2$,

$$\mathbb{E}[|E(G_2)|] \leq 2c'\log n. \tag{36}$$

Consider the degree of a vertex $u > k$ in G_3,

$$\mathbb{E}[\widehat{D}_u|_{G_3}] = \sum_{v=k}^{n} \frac{\phi_u \phi_v \omega(n)}{g(n)} B_{\sigma_v \sigma_u} \leq C \frac{\omega(n)}{g(n)} n \leq C\omega(n).$$

Hence,

$$\mathbb{P}(\widehat{D}_u|_{G_3} > C\omega(n) + \sqrt{c' \log(n) C\omega(n)}) \leq \frac{2}{n^{c'/4}}. \tag{37}$$

Combining these observations leads to our assertion that the first k eigenvectors of A become undistinguishable from those of the k stars when n tends to ∞. Indeed, split A according to the described graph composition, i.e.

$$A = A|_{G_1} + A|_{G_1'} + A|_{G_2} + A|_{G_3}$$

and note that the spectral radii of $A|_{G_1'}$, $A|_{G_2}$, and $A|_{G_3}$ vanish in the presence of $\Delta(G_1)$. This follows because (as mentioned above) for any graph its spectral radius is bounded by the minimum of its largest degree and the square root of its number of edges. Hence, due to (35)–(37),

$$\rho(A|_{G_1'}) \leq 2c' \log n, \qquad \rho(A|_{G_2}) \leq \sqrt{2c' \log n},$$

and

$$\rho(A|_{G_3}) \leq C\omega(n) + \sqrt{c' \log(n) C\omega(n)},$$

with high probability. All these three bounds vanish indeed upon division by $\Delta(G_1) \geq C\omega(n)n^{(\gamma-3\beta)/2}$. An application of Lemma 11 completes the proof. $\qquad\square$

7.2. Interpretation of the conditions

Proof of Remark 1. Assume that $B_{ij}/M_i = B_{lj}/M_l$ then

$$\frac{B_{ij}}{M_i M_j} = \frac{B_{lj}}{M_l M_j}.$$

Now, set $\overline{\phi}_i = (1/\alpha_i n) \sum_{\sigma_u=i} \phi_u$ then

$$\frac{B_{ij}}{M_i M_j} = \frac{\alpha_i \overline{\phi}_i B_{ij} \alpha_j \overline{\phi}_j}{\alpha_i \overline{\phi}_i M_i M_j \alpha_j \overline{\phi}_j}.$$

We give a probabilistic interpretation to the terms appearing in the denominator, i.e.

$$n\alpha_i \overline{\phi}_i M_i = n\alpha_i \overline{\phi}_i \sum_{k=1}^{K} \sum_{u:\sigma_u=k} \phi_u B_{i\sigma_u}$$

$$= n\alpha_i \overline{\phi}_i \sum_{k=1}^{K} n\alpha_k \overline{\phi}_k B_{ik}$$

$$= \sum_{k=1}^{K} (n\alpha_i)(n\alpha_k) \overline{\phi}_i B_{ik} \overline{\phi}_k$$

$$= \sum_{k=1}^{K} \sum_{u:\sigma_u=i} \phi_u \sum_{v:\sigma_v=k} \phi_v B_{ik}$$

$$= \frac{n}{\omega(n)} \sum_{k=1}^{K} \sum_{u:\sigma_u=i} \sum_{v:\sigma_v=k} \mathbb{P}(l \leftrightarrow m)$$

$$= \frac{n}{\omega(n)} \sum_{u:\sigma_u=i} \sum_{m=1}^{n} \mathbb{P}(l \leftrightarrow m)$$

$$= \frac{n}{\omega(n)} \{\text{expected total degree of vertices in community } i\}.$$

An inspection of the numerator reveals that

$$n\alpha_i \overline{\phi}_i B_{ij} \overline{\phi}_j \alpha_j n = \sum_{u:\sigma_u=i} \phi_u \sum_{v:\sigma_v=j} \phi_v B_{ij}$$

$$= \frac{n}{\omega(n)} \sum_{u:\sigma_u=i} \sum_{v:\sigma_v=j} \mathbb{P}(u \leftrightarrow v)$$

$$= \frac{n}{\omega(n)} \{\text{expected \#edges between community } i \text{ and } j\}. \qquad \square$$

Proof of Lemma 1. Assume first that for some i and l, we have, for all j, $\widehat{B}_{ij} = \widehat{B}_{lj}$ and, for all u, v,

$$\frac{\phi_u B_{\sigma_u \sigma_v} \phi_v}{g(n)} = \frac{\widehat{\phi}_u \widehat{B}_{\sigma_u \sigma_v} \widehat{\phi}_v}{\widehat{g}(n)}$$

with ϕ_u defined in (17) and g in (19) ($\widehat{\phi}_u$ and $\widehat{g}$ are defined analogously). Fix j. Let α, β, and γ be *any* indices such that $\sigma_\alpha = i$, $\sigma_\beta = j$, and $\sigma_\gamma = l$. Then

$$\frac{\widehat{\phi}_\alpha \widehat{B}_{ij} \widehat{\phi}_\beta}{\widehat{g}(n)} = \frac{\phi_\alpha B_{ij} \phi_\beta}{g(n)} \quad \Longrightarrow \quad B_{ij} = \frac{\widehat{\phi}_\alpha}{\phi_\alpha} \frac{\widehat{\phi}_\beta}{\phi_\beta} \frac{g(n)}{\widehat{g}(n)} \widehat{B}_{ij}$$

and

$$\frac{\widehat{\phi}_\gamma \widehat{B}_{lj} \widehat{\phi}_\beta}{\widehat{g}(n)} = \frac{\phi_\gamma B_{lj} \phi_\beta}{g(n)} \quad \Longrightarrow \quad B_{lj} = \frac{\widehat{\phi}_\gamma}{\phi_\gamma} \frac{\widehat{\phi}_\beta}{\phi_\beta} \frac{g(n)}{\widehat{g}(n)} \widehat{B}_{lj},$$

implying that (since $\widehat{B}_{ij} = \widehat{B}_{lj}$)

$$B_{ij} = \frac{\widehat{\phi}_\alpha}{\phi_\alpha} \frac{\widehat{\phi}_\beta}{\phi_\beta} \frac{g(n)}{\widehat{g}(n)} \widehat{B}_{lj} = \frac{\widehat{\phi}_\alpha}{\phi_\alpha} \frac{\phi_\gamma}{\widehat{\phi}_\gamma} B_{lj}.$$

Since j was arbitrary, there exist c such that, for all j, $B_{ij} = cB_{lj}$, hereby violating the identifiability condition, as pointed out in Remark 2, i.e.

$$\frac{B_{ij}}{M_i} = \frac{B_{lj}}{M_l} \quad \text{for all } j.$$

Now assume that (a) holds, i.e.

$$\frac{B_{ij}}{M_i} = \frac{B_{lj}}{M_l} \quad \text{for all } j.$$

Define, for $k, l \in S$ and $u \in V$,

$$\widehat{B}_{kl} = \frac{1}{M_k} \frac{1}{M_l} B_{kl} \quad \text{and} \quad \widehat{\phi}_u = f(n) \phi_u M_{\sigma_u},$$

where $f(n) = \sum_v \phi_v M_{\sigma_v} / \sum_w \phi_w$. Then

$$\widehat{B}_{ij} = \frac{1}{M_i}\frac{1}{M_j}B_{ij} = \frac{1}{M_i}\frac{1}{M_j}\frac{M_i}{M_l}B_{lj} = \frac{1}{M_j}\frac{1}{M_l}B_{lj} = \widehat{B}_{lj}$$

and (as above, we define $\widehat{g}$ analogously to g)

$$
\begin{aligned}
\frac{\widehat{\phi}_u \widehat{B}_{\sigma_u \sigma_v} \widehat{\phi}_v}{\widehat{g}(n)} &= \frac{1}{\sum_w f(n)\phi_w M_{\sigma_w}}\phi_u M_{\sigma_u} f(n)\frac{B_{\sigma_u\sigma_v}}{M_{\sigma_u}M_{\sigma_v}}f(n)M_{\sigma_v}\phi_v \\
&= \frac{\phi_u B_{\sigma_u\sigma_v}\phi_v}{\sum_w \phi_w} \\
&= \frac{\phi_u B_{\sigma_u\sigma_v}\phi_v}{g(n)}.
\end{aligned}
$$
□

8. Future research

8.1. Exact recovery

The obtained clustering here is almost exact: only a vanishing fraction of nodes is mis-classified. It is plausible that an exact clustering could be obtained from this clustering, by using it as input to the 'clean-up' algorithm presented in [1, Section 7.2] or, alternatively, [23, Algorithm 2].

8.2. Nonconstant *B*

In this paper we assumed B to be a constant matrix. The current analysis could be extended to a setting where B is allowed to change with n. We need, however, the existence of a constant $\delta > 0$ such that for all n, $\rho(Z) \geq \delta$ for $\widehat{H}$ to concentrate. For identifiability we need the existence of some $\varepsilon > 0$ such that for all i, j, and n, $\max_{i'} |B_{ii'}/\overline{M}_i \overline{M}_{i'} \neq B_{ji'}/\overline{M}_j \overline{M}_{i'}| \geq \varepsilon$.

8.3. Sparser graphs

The main issue with both the normalized adjacency matrix and the Laplacian is proving when those matrices concentrate around a deterministic matrix. For the Laplacian, if the degrees are of order $\Omega(\log(n))$, matrices concentrate according to [5]. But, if the minimum degree is of order $o(\log(n))$, the graph is seen to have some isolated vertices. Those contribute to multiple zeros in the spectrum; hence, the matrix does not concentrate. There are multiple ways to overcome this issue, for instance removing the low-degree vertices or raising all the degrees. The latter strategy was proposed in [18] for the inhomogeneous Erdős–Rényi random graph (where edges are independently present with probabilities $(p_{uv})_{u,v=1}^{n}$) and also in [4] and [27] (see Section 4.4) for the DC-SBM. According to [18], for $\tau \sim d$ with $d = n \max_{uv} p_{uv}$, with high probability,

$$\rho(L_\tau - (\mathbb{E}[D_\tau]^{-1/2}\mathbb{E}[A]\mathbb{E}[D_\tau]^{-1/2})) = \mathcal{O}\left(\frac{1}{\sqrt{d}}\right),$$

where L_τ is defined in (13).

Based on these observations, it might be fruitful to use $\widehat{H}$ on a graph where the degrees have been artificially inflated.

References

[1] ABBE, E., BANDEIRA, A. S. AND HALL, G. (2016). Exact recovery in the stochastic block model. *IEEE Trans. Inf. Theory* **62,** 471–487.
[2] ADAMIC, L. A. AND GLANCE, N. (2005). The political blogosphere and the 2004 U.S. election: divided they blog. In *Proc. LinkKDD'05*, ACM, New York, pp. 36–43.

[3] BERNSTEIN, S. (1946). *The Theory of Probabilities*. Gastehizdat, Moscow.

[4] CHAUDHURI, K., CHUNG, F. AND TSIATAS, A. (2012). Spectral clustering of graphs with general degrees in the extended planted partition model. In *Proc. 25th Annual Conf. Learning Theory*, pp. 35.1–35.23.

[5] CHUNG, F. AND RADCLIFFE, M. (2011). On the spectra of general random graphs. *Electron. J. Combin.* **18,** 215.

[6] CHUNG, F., LU, L. AND VU, V. (2004). The spectra of random graphs with given expected degrees. *Internet Math.* **1,** 257–275.

[7] COJA-OGHLAN, A. AND LANKA, A. (2010). Finding planted partitions in random graphs with general degree distributions. *SIAM J. Discrete Math.* **23,** 1682–1714.

[8] DASGUPTA, A., HOPCROFT, J. E. AND MCSHERRY, F. (2004). Spectral analysis of random graphs with skewed degree distributions. In *Proc. 45th Ann. IEEE Symp. Foundations Comput. Sci.*, IEEE, New York, pp. 602–610.

[9] DECELLE, A., KRZAKALA, F., MOORE, C. AND ZDEBOROVÁ, L. (2011). Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys. Rev. E* **84,** 066106.

[10] FEIGE, U. AND OFEK, E. (2005). Spectral techniques applied to sparse random graphs. *Random Structures Algorithms* **27,** 251–275.

[11] GIRVAN, M. AND NEWMAN, M. E. J. (2002). Community structure in social and biological networks. *Proc. Nat. Acad. Sci. USA* **99,** 7821–7826.

[12] HOFF, P. D., RAFTERY, A. E. AND HANDCOCK, M. S. (2002). Latent space approaches to social network analysis. *J. Amer. Statist. Assoc.* **97,** 1090–1098.

[13] HOLLAND, P. W., LASKEY, K. B. AND LEINHARDT, S. (1983). Stochastic blockmodels: first steps. *Social Networks* **5,** 109–137.

[14] HORN, R. A. AND JOHNSON, C. R. (1985). *Matrix Analysis*. Cambridge University Press.

[15] JIN, J. (2015). Fast community detection by SCORE. *Ann. Statist.* **43,** 57–89.

[16] KARRER, B. AND NEWMAN, M. E. J. (2011). Stochastic blockmodels and community structure in networks. *Phys. Rev. E* **83,** 016107.

[17] KRZAKALA, F. *et al.* (2013). Spectral redemption in clustering sparse networks. *Proc. Nat. Acad. Sci. USA* **110,** 20935–20940.

[18] LE, C. M. AND VERSHYNIN, R. (2015). Concentration and regularization of random graphs. Preprint. Available at https://arxiv.org/abs/1506.00669.

[19] LEI, J. AND RINALDO, A. (2015). Consistency of spectral clustering in stochastic block models. *Ann. Statist.* **43,** 215–237.

[20] LUSSEAU, D. *et al.* (2003). The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations. *Behavioral Ecol. Sociobiol.* **54,** 396–405.

[21] MCSHERRY, F. (2001). Spectral partitioning of random graphs. In *Proc. 42nd IEEE Symp. Foundations Comput. Sci.*, IEEE, Los Alamitos, CA, pp. 529–537.

[22] MIHAIL, M. AND PAPADIMITRIOU, C. (2002). On the eigenvalue power law. In *Randomization and Approximation Techniques in Computer Science* (Lecture Notes Comput. Sci. **2483**), Springer, Berlin, pp. 254–262.

[23] MOSSEL, E., NEEMAN, J. AND SLY, A. (2015). Consistency thresholds for the planted bisection model. In *STOC'15—Proc. 2015 ACM Symp. Theory Comput.*, ACM, New York, pp. 69–75.

[24] NEWMAN, M. E. J. (2004). Detecting community structure in networks. *Europ. Phys. J. B* **38,** 321–330.

[25] NEWMAN, M. E. J. (2010). *Networks: An Introduction*. Oxford University Press.

[26] NEWMAN, M. E. J. AND GIRVAN, M. (2004). Finding and evaluating community structure in networks. *Phys. Rev. E* **69,** 026113.

[27] QIN, T. AND ROHE, K. (2013). Regularized spectral clustering under the degree-corrected stochastic blockmodel. In *Advances in Neural Information Processing Systems 26*, Curran Associates, Red Hook, NY, pp. 3120–3128.

[28] ROHE, K., CHATTERJEE, S. AND YU, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist.* **39,** 1878–1915.

[29] TOMOZEI, D.-C. AND MASSOULIÉ, L. (2014). Distributed user profiling via spectral methods. *Stoch. Systems* **4,** 1–43.

[30] VON LUXBURG, U. (2007). A tutorial on spectral clustering. *Statist. Comput.* **17,** 395–416.

[31] ZACHARY, W. W. (1977). An information flow model for conflict and fission in small groups. *J. Anthropological Res.* **33,** 452–473.

[32] ZHANG, X., NADAKUDITI, R. AND NEWMAN, M. E. J. (2014). Spectra of random graphs with community structure and arbitrary degrees. *Phys. Rev. E* **89,** 042816.