Journal of Machine Learning Research 18 (2018) 1-49 Submitted 5/17; Revised 2/18; Published 4/18

Improved Spectral Community Detection in Large Heterogeneous Networks

Hafiz Tiomoko Ali TIO.HAFIZ@GMAIL.COM
Romain Couillet ROMAIN.COUILLET@CENTRALESUPELEC.FR
CentraleSupélec
Université Paris Saclay
Laboratoire des Signaux et Systèmes
3 rue Joliot Curie, 91192 Gif-Sur-Yvette

Editor: Ulrike von Luxburg

Abstract

In this article, we propose and study the performance of spectral community detection for a family of "α-normalized" adjacency matrices $\mathbf{A}$, of the type $\mathbf{D}^{-\alpha}\mathbf{A}\mathbf{D}^{-\alpha}$ with $\mathbf{D}$ the degree matrix, in heterogeneous dense graph models. We show that the previously used normalization methods based on $\mathbf{A}$ or $\mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-1}$ are in general suboptimal in terms of correct recovery rates and, relying on advanced random matrix methods, we prove instead the existence of an optimal value $\alpha_{\rm opt}$ of the parameter α in our generic model; we further provide an online estimation of $\alpha_{\rm opt}$ only based on the node degrees in the graph. Numerical simulations show that the proposed method outperforms state-of-the-art spectral approaches on moderately dense to dense heterogeneous graphs.

Keywords: community detection, random networks, heterogeneous graphs, random matrix theory, spectral clustering

1. Introduction and Motivations

The advent of the big data era is creating an unprecedented need for automating large network analysis. Community detection is among the most important tasks in automated network mining (Fortunato, 2010). Given a network graph, detecting communities consists in retrieving hidden clusters of nodes based on some similarity metric (the edges are dense inside communities and sparse across communities). While quite simple to define, community detection is usually not an easy task and many methods arising from different fields have been proposed to carry it out. The most important of them are statistical inference, modularity maximization and graph partitioning methods. Statistical inference methods consist in fitting the observed network to a structured network model and infer its parameters (among which the assignment of the nodes to the communities) (Hastings, 2006; Newman and Leicht, 2007). Modularity maximization algorithms rely instead on the modularity metric which quantifies the subdivision of networks into communities (Fortunato, 2010).[1] However, retrieving the modularity maximizing graph partition is generally an NP-hard problem and many polynomial-time approximation methods have been proposed: greedy methods (Newman, 2004), simulated annealing (Guimera et al., 2004), extremal optimization (Duch and Arenas,

1. Precisely, the modularity is defined as the difference between the total number of edges inside the communities for a given partition and the total number of edges if the partition were created randomly in the graph.

2005) and spectral methods (Newman, 2006b). Spectral algorithms consist in retrieving the communities from the eigenvectors associated with the dominant eigenvalues of some matrix representation of the graph structure (adjacency matrix, modularity matrix, Laplacian matrix). By relaxing the modularity optimization problem from binary values of the community memberships to continuous scores, it is shown that approximate modularity maximization and even statistical inference methods can be performed via a low dimensional clustering of the entries of the dominant eigenvectors of the representation matrix (Ng et al., 2002; Newman, 2016) in polynomial time. Precisely, the steps of spectral methods for community detection are described in the following algorithm.

Algorithm 1: Spectral algorithm

1: Compute the, say, ℓ eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_\ell \in \mathbb{R}^n$ corresponding to the dominant (largest or smallest) eigenvalues of one of the matrix representations of the network (adjacency, modularity, Laplacian) of size $n \times n$.

2: Stack the vectors $\mathbf{u}_i$'s columnwise in a matrix $\mathbf{W} = [\mathbf{u}_i, \ldots, \mathbf{u}_\ell] \in \mathbb{R}^{n \times \ell}$.

3: Let $\mathbf{r}_1, \ldots, \mathbf{r}_n \in \mathbb{R}^\ell$ be the rows of $\mathbf{W}$. Cluster $\mathbf{r}_i \in \mathbb{R}^\ell$, $1 \leq i \leq n$ in one of K groups using any low-dimensional classification algorithm (e.g., k-means (Hartigan and Wong, 1979) or Expectation Maximization (EM) (Ng et al., 2012)). The label assigned to $\mathbf{r}_i$ then corresponds to the label of node i.

Most of the works proposing statistical analysis of the performance of community detection (for dense as well as sparse networks) consider the basic Stochastic Block Model (SBM) as a model for networks decomposable into communities. Denoting $\mathcal{G}$ a K-class graph of n vertices with communities $\mathcal{C}_1, \ldots, \mathcal{C}_K$ with g_i the group assignment of node i, the SBM assumes an adjacency matrix $\mathbf{A} \in \{0,1\}^{n \times n}$ with A_{ij} independent Bernoulli random variables with parameter $P_{g_i g_j}$ where P_{ab} represents the probability that any node of class $\mathcal{C}_a$ is connected to any node of class $\mathcal{C}_b$. The main limitation of this model is that it is only suited to homogeneous graphs where all nodes have the same average degree in each community (besides, class sizes are often taken equal). As suggested in the practical case of the popular Political Blogs graph (Adamic and Glance, 2005) (see Figure 4), a more realistic model, the Degree-Corrected SBM (DCSBM), was proposed in (Coja-Oghlan and Lanka, 2009; Karrer and Newman, 2011) to account for degree heterogeneity inside communities. For the same graph $\mathcal{G}$ defined above, by letting q_i, $1 \leq i \leq n$, be some intrinsic weights which affect the probability for node i to connect to any other network node, the adjacency matrix $\mathbf{A} \in \{0,1\}^{n \times n}$ of the graph generated by the DCSBM is such that A_{ij} are independent Bernoulli random variables with parameter $q_i q_j C_{g_i g_j}$, where $C_{g_i g_j}$ is a class-wise correction factor.

Community detection in DCSBMs has recently been studied, providing "consistent"[2] algorithms ranging from modularity/likelihood based approaches to spectral clustering methods. Sufficient conditions under which likelihood based approaches (Karrer and Newman, 2011) and modularity optimization methods (Newman, 2006b) are weakly and strongly consistent, have been provided in (Zhao et al., 2012). The so-called CMM (Convexified Modularity Maximization) algorithm was proposed in (Chen et al., 2015) to cope with the computa-

2. Consistency is mainly defined in two forms. Informally, a community detection algorithm is **weakly** consistent whenever the fraction of misclassified nodes vanishes asymptotically with high probability and a community detection algorithm is **strongly** consistent whenever the labels estimated by the algorithm match exactly the true labelling asymptotically with high probability.

tional expensiveness of modularity/likelihood methods (Karrer and Newman, 2011; Newman, 2006b) by solving a convex programming relaxation of the modularity optimization. Asymptotic minimax risks for misclassification loss under the DCSBM have been established in (Gao et al., 2016). There a consistent algorithm achieving the minimax optimal rates was derived, which is similar to spectral methods but proceeds without the explicit computation of eigenvectors and is hence computationally less expensive. As far as spectral clustering methods are concerned, (Lyzinski et al., 2014) and (Lei et al., 2015) show consistency of the classical spectral clustering procedure for community detection applied to the adjacency matrix of moderately sparse DCSBM (for not too irregular degree distributions) where the expected degree is as small as $\log n$. Later, its has been shown (Coja-Oghlan and Lanka, 2009; Qin and Rohe, 2013; Jin et al., 2015; Gulikers et al., 2015) that when the degrees are highly heterogeneous, the classical spectral methods fail to detect the genuine communities. To illustrate those limitations of spectral methods under the DCSBM, the two graphs of Figure 1 provide 2D representations of dominant eigenvector 1 versus eigenvector 2 for the standard modularity matrix and the Bethe Hessian matrix[3], when three quarters of the nodes connect with low weight $q_{(1)}$ and one quarter of the nodes with high weight $q_{(2)}$. For both methods, it is clear that k-means or EM alike would erroneously induce the detection of extra communities and even a confusion of genuine communities in the Bethe Hessian approach. Those extra communities are produced by some biases created by the intrinsic weights q_i's; intuitively, nodes sharing the same intrinsic connection weights tend to create their own sub-cluster inside each community, thereby forming additionnal sub-communities inside the genuine communities. To overcome this issue, a number of regularized spectral clustering techniques have been proposed to normalize either the adjacency matrix or the leading eigenvectors by the degrees. In (Coja-Oghlan and Lanka, 2009; Gulikers et al., 2015), the authors have proposed to cluster the nodes based on the eigenvectors of a normalized adjacency matrix $\mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-1}$ with $\mathbf{D}$ the diagonal matrix containing the observed degrees on the main diagonal. The SCORE algorithm devised in (Jin et al., 2015) consists instead in using the leading eigenvectors of the adjacency matrix (pre-normalized by the dominant eigenvector which, as shown subsequently, is equivalent to normalizing by the inverse degree matrix $\mathbf{D}^{-1}$) and (Qin and Rohe, 2013) proposed to use the eigenvectors of the Laplacian matrix $\mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}$.

As previously stated, the aforementioned works have shown that under some regularity (or regularization) conditions, an almost perfect or perfect reconstruction of the nodes labels can be achieved asymptotically. Our motivation in this article is to go beyond mere consistency results by understanding the performances of the different regularized spectral clustering algorithms for large but finite network sizes n. We place ourselves in a regime where communities are too close to induce perfect reconstructions. In order to encompass most aforementioned methods, we study here a generalized regularization of the adjacency

3. The Bethe Hessian (BH) spectral method (Saade et al., 2014) is based on the union of the eigenvectors associated to the negative eigenvalues of $H(r_c)$ and $H(-r_c)$ respectively where $H(r) = (r^2-1)\mathbf{I}_n - r\mathbf{A} + \mathbf{D}$ for $r_c = \frac{\sum_i d_i^2}{\sum_i d_i} - 1$ with d_i the degree of node i ($\mathbf{D}$ and d_i are defined subsequently).

matrix[4] given, for any $\alpha \in \mathbb{R}$, by

$$\mathbf{L}_\alpha = (2m)^\alpha \frac{1}{\sqrt{n}} \mathbf{D}^{-\alpha} \left[\mathbf{A} - \frac{\mathbf{d}\mathbf{d}^\mathsf{T}}{2m} \right] \mathbf{D}^{-\alpha}$$

where $\mathbf{d}$ is the vector of degrees ($d_i = \sum_{j=1}^n A_{ij}$), $\mathbf{D}$ is the diagonal matrix of degrees (containing $\mathbf{d}$ on the main diagonal) and $m = \frac{1}{2}\mathbf{d}^\mathsf{T}\mathbf{1}_n$ is the number of edges in the network. In particular, $\mathbf{L}_0$ is the modularity matrix (Newman, 2006b; Jin et al., 2015), $\mathbf{L}_{\frac{1}{2}}$ is a modularity equivalent to the normalized Laplacian matrix (Qin and Rohe, 2013; Chung, 1997) and $\mathbf{L}_1$ is the form used in (Coja-Oghlan and Lanka, 2009; Gulikers et al., 2015; Tiomoko Ali and Couillet, 2016).

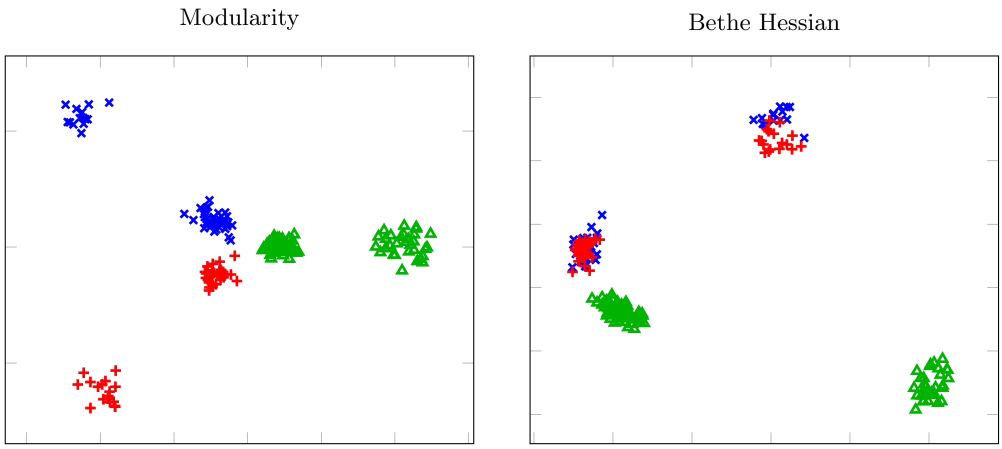


Figure 1: Two dominant eigenvectors (x-y axes) for $n = 2000$, $K = 3$ classes $\mathcal{C}_1$, $\mathcal{C}_2$ and $\mathcal{C}_3$ of sizes $|\mathcal{C}_1| = |\mathcal{C}_2| = \frac{n}{4}$, $|\mathcal{C}_3| = \frac{n}{2}$, $\frac{3}{4}$ of the nodes having $q_i = 0.1$ and $\frac{1}{4}$ of the nodes having $q_i = 0.5$, matrix of weights $\mathbf{C} = \mathbf{1}_3\mathbf{1}_3^\mathsf{T} + \frac{100}{\sqrt{n}}\mathbf{I}_3$. Colors and shapes correspond to ground truth classes.

Besides, while we believe (up to more involved mathematical treatment) that our results essentially hold true in moderately sparse graphs (of average degree of order $\Omega(\log n)$), we focus here on a *dense* DCSBM model where $q_i = \Omega(1)$. In this regime, when the correction factors $C_{g_ig_j}$ differ by a rate greater than $\mathcal{O}(n^{-\frac{1}{2}})$, weak consistency is shown for all regularized spectral algorithms. Instead, we induce a regime where clusters remain at an asymptotically "constant" distance. This is ensured by letting $C_{g_ig_j} = \Omega(1)$ individually but with the $C_{g_ig_j}$'s differing by $\mathcal{O}(n^{-\frac{1}{2}})$. Under this regime, we are able to fully study the dominant eigenvalues and associated eigenvectors (used for classification) of $\mathbf{L}_\alpha$ for large dimensional dense graphs following the DCSBM, thus allowing one to assess the performances for very large but finite-size graphs.

In a nutshell, our main findings are as follows.

4. The leading term $\frac{\mathbf{d}\mathbf{d}^\mathsf{T}}{2m}$ (not providing any information about the communities) is shown in simulations to have asymptotically no impact on the clustering performance. It is discarded here mostly for mathematical simplicity. Note in passing that $\mathbf{A} - \frac{\mathbf{d}\mathbf{d}^\mathsf{T}}{2m}$ corresponds to the so-called modularity matrix (Newman, 2006a), therefore $\mathbf{L}_\alpha$ may be seen as a "α-normalized" modularity matrix.

- We prove the existence of and obtain an expression for an optimal value α_{opt} of α for which the community detectability threshold[5] is maximally achievable.This value needs not be either 0 or 1 and its proper choice is of utmost importance in highly heterogeneous graphs.
- We provide a consistent estimator $\hat{\alpha}_{\text{opt}}$ of α_{opt} based on $\mathbf{d}$ alone.
- We show that to achieve consistent clustering in the DCSBM model, the dominant eigenvectors used for clustering should be pre-multiplied by $\mathbf{D}^{\alpha-1}$ prior to the low dimensional classification (step 3 of Algorithm 1), thereby recovering the SCORE algorithm (Jin et al., 2015) for $\alpha = 0$ and the algorithm in (Gulikers et al., 2015) for $\alpha = 1$, as special cases.
- Our proposed method is summarized under the form of Algorithm 2. As the estimation of α_{opt} is essentially linear with n, Algorithm 2 does not impair the computational cost of the underlying spectral method. A Python implementation of the algorithm is available in (Tiomoko Ali and Couillet, 2017).
- A deeper study of the regularized eigenvectors allows us to improve the initial setting of the EM algorithm (in the step 3 of the spectral algorithm described above) in comparison with a random setting.
- Numerical simulations (throughout the article) show that our methods outperform state-of-the-art spectral methods both on synthetic graphs and on real world networks.

All proofs are deferred to a separate section while sketches are provided for the main results of the article.

Notations: Vectors (matrices) are denoted by lowercase (uppercase) boldface letters. $\{\mathbf{v}_a\}_{a=1}^n$ is the column vector $\mathbf{v}$ with (scalar or vector) entries v_a and $\{\mathbf{V}_{ab}\}_{a,b=1}^n$ is the matrix $\mathbf{V}$ with (scalar or matrix) entries V_{ab}. For a vector $\mathbf{v}$, the operator $\mathcal{D}(\mathbf{v}) = \mathcal{D}\left(\{v_a\}_{a=1}^n\right)$ is the diagonal matrix having the scalars v_a down its diagonal and for a matrix $\mathbf{V}$, $\mathcal{D}(\mathbf{V})$ is the vector containing the diagonal entries of $\mathbf{V}$. The vector $\mathbf{1}_n \in \mathbb{R}^n$ stands for the column vector filled with ones. The Dirac measure at x is δ_x. The vector $\mathbf{j}_a$ is the canonical vector of class $\mathcal{C}_a$ defined by $(\mathbf{j}_a)_i = \delta_{i\in\mathcal{C}_a}$ and $\mathbf{J} = [\mathbf{j}_1, \ldots, \mathbf{j}_K] \in \{0,1\}^{n\times K}$. The set $\mathbb{C}^+$ is $\{z \in \mathbb{C},\ \Im[z] > 0\}$. We denote $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ to indicate that $\mathbf{x}$ is a Gaussian distributed random vector with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$.

2. Preliminaries

This section describes the network model under study, which is based on the DCSBM defined in the previous section, and provides preliminary technical results.

Consider an n-node random graph with K classes $\mathcal{C}_1, \ldots, \mathcal{C}_K$ of sizes $|\mathcal{C}_k| = n_k$. Each node is characterized by an intrinsic connexion weight q_i which affects the probability that this node gets attached to another node in the graph. A null model would consider that the existence of an edge between i and j has probability $q_i q_j$. In order to take into account the

5. The community detectability threshold is the point beyond which there exists a clustering algorithm which can do better than a random guess.

membership of the nodes to some group, we define $\mathbf{C} \in \mathbb{R}^{K\times K}$ as a matrix of class weights C_{ab}, independent of the q_i's, affecting the connection probability between nodes in $\mathcal{C}_a$ and nodes in $\mathcal{C}_b$. Following (Karrer and Newman, 2011), the adjacency matrix $\mathbf{A}$ of the graph generated from a DCSBM model has independent entries (up to symmetry) which are Bernoulli random variables with parameter $P_{ij} = q_i q_j C_{g_i g_j} \in (0,1)$ where g_i is the group assignment of node i. We set $A_{ii} = 0$ for all i. For convenience of exposition and without loss of generality, we assume that node indices are sorted by clusters, i.e nodes 1 to n_1 constitute $\mathcal{C}_1$, nodes n_1+1 to n_1+n_2 form $\mathcal{C}_2$, and so on.

The matrix under study is given by

$$\mathbf{L}_\alpha = (2m)^\alpha \frac{1}{\sqrt{n}} \mathbf{D}^{-\alpha} \left[\mathbf{A} - \frac{\mathbf{d}\mathbf{d}^\mathsf{T}}{2m}\right] \mathbf{D}^{-\alpha} \tag{1}$$

where $\mathbf{d} = \mathbf{A}\mathbf{1}_n$, $\mathbf{D} = \mathcal{D}(\mathbf{d})$ and $m = \frac{1}{2}\mathbf{d}^\mathsf{T}\mathbf{1}_n$.

We are mainly interested in a dense network regime where clustering is not asymptotically trivial. This regime is ensured by the following growth rate conditions.

Assumption 1 *As $n \to \infty$, K remains fixed and, for all $i,j \in \{1,\dots,n\}$*

1. $C_{g_i g_j} = 1 + \frac{M_{g_i g_j}}{\sqrt{n}}$, *where* $M_{g_i g_j} = \Omega(1)$; *we shall denote* $\mathbf{M} = \{M_{ab}\}_{a,b=1}^K$.
2. q_i *are i.i.d. random variables with measure* μ *having compact support in* $(0,1)$.
3. $\frac{n_i}{n} \to c_i > 0$ *and we will denote* $\mathbf{c} = \{c_k\}_{k=1}^K$.

The goal of the article is to study deeply the eigenstructure of $\mathbf{L}_\alpha$ in order to understand the different mechanisms into play when performing spectral clustering on $\mathbf{L}_\alpha$. As can be observed, $\mathbf{L}_\alpha$ has non independent entries as $\mathbf{D}$ (and $\mathbf{d}$) depend on $\mathbf{A}$, and it thus does not follow a standard random matrix model. Our strategy is to approximate $\mathbf{L}_\alpha$ by a more tractable random matrix $\tilde{\mathbf{L}}_\alpha$ which asymptotically preserves the eigenvalue distribution and isolated eigenvectors of $\mathbf{L}_\alpha$. We obtain the corresponding approximate of $\mathbf{L}_\alpha$ as follows.

Theorem 1 *Let Assumption 1 hold and let $\mathbf{L}_\alpha$ be given by (1). Then, for $\mathbf{D}_q \triangleq \mathcal{D}(\mathbf{q})$, as $n \to \infty$, $\|\mathbf{L}_\alpha - \tilde{\mathbf{L}}_\alpha\| \to 0$ in operator norm, almost surely, where*

$$\tilde{\mathbf{L}}_\alpha = \frac{1}{\sqrt{n}} \mathbf{D}_q^{-\alpha} \mathbf{X} \mathbf{D}_q^{-\alpha} + \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\mathsf{T},$$
$$\mathbf{U} = \begin{bmatrix} \frac{\mathbf{D}_q^{1-\alpha}\mathbf{J}}{\sqrt{n}} & \frac{\mathbf{D}_q^{-\alpha}\mathbf{X}\mathbf{1}_n}{\mathbf{q}^\mathsf{T}\mathbf{1}_n} \end{bmatrix},$$
$$\mathbf{\Lambda} = \begin{bmatrix} \left(\mathbf{I}_K - \mathbf{1}_K\mathbf{c}^T\right)\mathbf{M}\left(\mathbf{I}_K - \mathbf{c}\mathbf{1}_K^T\right) & -\mathbf{1}_K \\ -\mathbf{1}_K^T & 0 \end{bmatrix},$$

with $\mathbf{X} = \{X_{ij}\}_{i,j=1}^n$ symmetric with independent entries (up to the symmetry), X_{ij} having zero mean and variance $q_i q_j(1-q_i q_j)$, and $\mathbf{J} = [\mathbf{j}_1,\dots,\mathbf{j}_K] \in \{0,1\}^{n\times K}$ with $(\mathbf{j}_a)_i = \delta_{\{g_i=a\}}$.

Sketch of Proof *The proof relies on the fact that we may write $A_{ij} = q_i q_j + q_i q_j \frac{M_{g_i g_j}}{\sqrt{n}} + X_{ij}$ where X_{ij} is a zero mean random variable with variance $q_i q_j(1-q_i q_j) + \Theta(n^{-\frac{1}{2}})$, since A_{ij} is a Bernoulli random variable with parameter $q_i q_j(1 + \frac{M_{g_i g_j}}{\sqrt{n}})$. From there, the terms: $\mathbf{d} =$*

$\mathbf{A}\mathbf{1}_n$, $\mathbf{d}^\mathsf{T}\mathbf{1}_n$, $\mathbf{d}\mathbf{d}^\mathsf{T}$ and $\mathbf{D} = \mathcal{D}(\mathbf{d})$ composing $\mathbf{L}_\alpha$ can be evaluated. Notably, $\mathbf{D}$ and $\mathbf{d}^\mathsf{T}\mathbf{1}_n$ can be decomposed as the sum of dominant terms (with higher spectral norms with respect to n) and trailing terms (vanishing spectral norms with respect to n), so that we can write a Taylor expansion of $\mathbf{D}^{-\alpha}$ and $(\mathbf{d}^\mathsf{T}\mathbf{1}_n)^\alpha$ for $\alpha \in \mathbb{R}$. By computing $\mathbf{L}_\alpha$ using the asymptotic approximations of $\mathbf{D}^{-\alpha}$, $(\mathbf{d}^\mathsf{T}\mathbf{1}_n)^\alpha$, $\mathbf{A}$, $\mathbf{d}\mathbf{d}^\mathsf{T}$, we obtain $\tilde{\mathbf{L}}_\alpha$. The complete proof is provided in Section 6.1.

This result immediately implies the following Corollary.

Corollary 2 *Under Assumption 1, let $\lambda_i(\mathbf{L}_\alpha)$ (resp., $\lambda_i(\tilde{\mathbf{L}}_\alpha)$) be the eigenvalues of $\mathbf{L}_\alpha$ (resp., $\tilde{\mathbf{L}}_\alpha$) with associated eigenvectors $\mathbf{u}_i(\mathbf{L}_\alpha)$ (resp., $\mathbf{u}_i(\tilde{\mathbf{L}}_\alpha)$). We have*

$$\max_{1\leq i\leq n} \left|\lambda_i(\mathbf{L}_\alpha) - \lambda_i(\tilde{\mathbf{L}}_\alpha)\right| \xrightarrow{\text{a.s.}} 0$$

and, if $\liminf_n \min_{j\neq i} |\lambda_i(\mathbf{L}_\alpha) - \lambda_j(\tilde{\mathbf{L}}_\alpha)| > 0$,

$$\left\|\mathbf{u}_i(\mathbf{L}_\alpha) - \mathbf{u}_i(\tilde{\mathbf{L}}_\alpha)\right\| \xrightarrow{\text{a.s.}} 0.$$

Thus, for large enough n, the spectral analysis of $\mathbf{L}_\alpha$ can be performed through that of $\tilde{\mathbf{L}}_\alpha$.

The matrix $\tilde{\mathbf{L}}_\alpha$ is essentially a classical random matrix model and the study of its eigenvalues and dominant eigenvectors can be performed using standard random matrix theory (RMT) approaches (Benaych-Georges and Nadakuditi, 2012; Hachem et al., 2013).

3. Main Results

3.1 Spike model and dominant eigenvector regularization

The matrix $\tilde{\mathbf{L}}_\alpha$ is an additive spiked random matrix (Baik et al., 2005) as it is the sum of the standard full rank symmetric random matrix $n^{-\frac{1}{2}}\mathbf{D}_q^{-\alpha}\mathbf{X}\mathbf{D}_q^{-\alpha}$ having independent zero mean entries and a low rank matrix $\mathbf{U}\mathbf{\Lambda}\mathbf{U}^\mathsf{T}$. As shown in Figure 2, the spectrum (eigenvalue distribution) of spiked random matrices is generally composed of (one or several) bulks of concentrated eigenvalues and, when a phase transition is met, of additionnal eigenvalues which isolate from the aforementioned bulks. The eigenvectors corresponding to the isolated eigenvalues of the spiked random matrix become more correlated to the eigenvectors of the low rank matrix when the corresponding eigenvalues are far away from the rest of the eigenvalues.

From Theorem 1, the low rank matrix $\mathbf{U}\mathbf{\Lambda}\mathbf{U}^\mathsf{T}$ contains the matrix $\mathbf{D}_q^{1-\alpha}\mathbf{J}$; so, when the phase transition is met, the eigenvectors of $\tilde{\mathbf{L}}_\alpha$ will be correlated to some extent to $\mathbf{D}_q^{1-\alpha}\mathbf{J}$ as long as the corresponding informative eigenvalues are isolated from the bulk of eigenvalues. This is well illustrated in Figure 2 where the eigenvectors associated to non-isolated eigenvalues are noisy, i.e., classes can be barely distinguished from those eigenvectors. On the other hand, the eigenvectors associated to isolated eigenvalues consist of noisy plateaus characterizing the classes and thus a consistent classification can be expected using those eigenvectors. However, for a better clustering, one expects instead the vectors used for classification to be correlated to the canonical vectors $\mathbf{j}_a$, $1 \leq a \leq K$, instead of $\mathbf{D}_q^{1-\alpha}\mathbf{j}_a$.

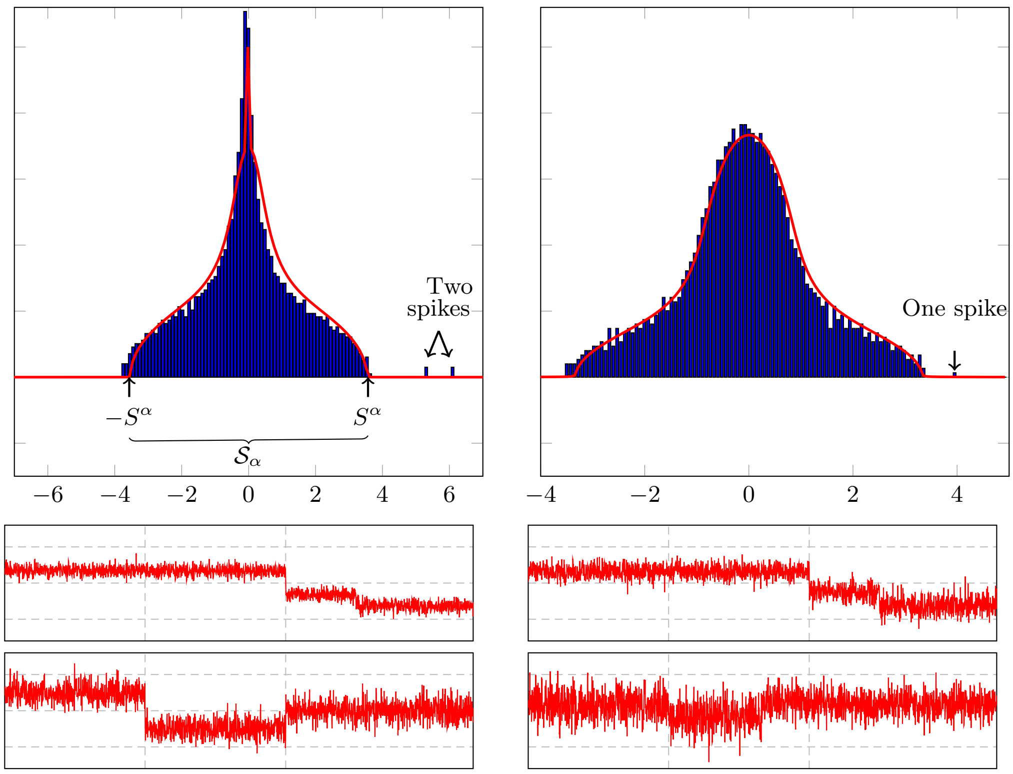


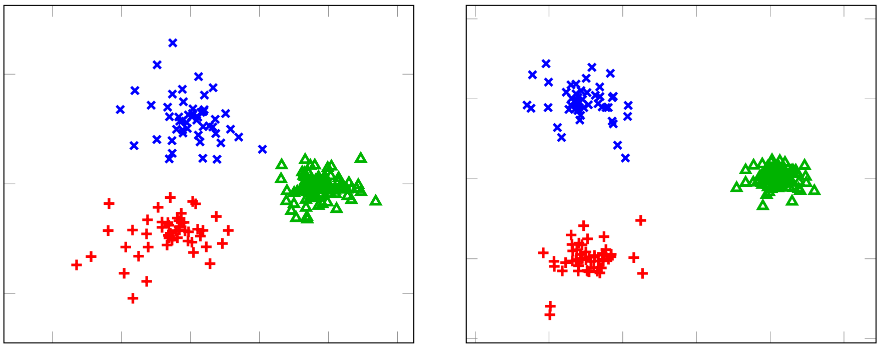
Figure 2: Two graphs generated upon the DCSBM with $K = 3$, $n = 2000$, $c_1 = 0.3, c_2 = 0.3, c_3 = 0.4$, $\mu = \frac{1}{2}\delta_{q_{(1)}} + \frac{1}{2}\delta_{q_{(2)}}$, $q_{(1)} = 0.4$, $q_{(2)} = 0.9$ and two different affinity matrices $\mathbf{M}$. **(Left)** $M_{ii} = 12$, $M_{ij} = -4, i \neq j$, **(Right)**: $M_{ii} = -3$, $M_{ij} = -10, i \neq j$, **(Top)**: Eigenvalue distribution of $\mathbf{L}_\alpha$, $\alpha = 0$. **(Bottom)**: First and second leading eigenvectors of $\mathbf{L}_\alpha$, $\alpha = 0$.

As a consequence, we claim that, letting $\mathbf{u}_1, \ldots, \mathbf{u}_l$ *be the eigenvectors associated to the* ℓ *isolated eigenvalues of* $\mathbf{L}_\alpha$*, the vectors* $\mathbf{v}_i = \mathbf{D}^{\alpha-1}\mathbf{u}_i$ *for* $1 \leq i \leq \ell$ *should be the ones used for the classification instead of the* $\mathbf{u}_i$*'s.*[6]

This important observation helps correcting the biases (creation of artificial classes) introduced by the degree heterogeneity observed earlier in Figure 1. As shown in Figure 3, which assumes the same setting as Figure 1, when the aforementioned eigenvector regularization is performed prior to EM or k-means classification, the genuine communities are correctly recovered.

As mentionned earlier, the eigenvectors corresponding to eigenvalues in the bulk are asymptotically of no use for clustering. It is thus important to characterize the phase transition point beyond which eigenvalues isolate from the bulk and determine which α best ensures this transition. To this end, we will first determine the support $\mathcal{S}^\alpha$ of the limiting spectral

6. As far as the eigenvectors are concerned, we may freely replace $\mathbf{D}_q$ (unknown in practice) by $\mathbf{D}$ (which can be computed from the observed graph) since, from Lemma 10 in the subsequent section 3.4, the vector of degrees $\mathbf{d}$ is, up to a scale factor β, a consistent estimator of the vector of intrinsic weights $\mathbf{q}$ and thus $\|\beta\mathbf{D} - \mathbf{D}_q\| \to 0$ almost surely.



($\alpha = 1$) ($\alpha = \hat{\alpha}_{\text{opt}}$)

Figure 3: Two dominant eigenvectors of $\mathbf{L}_\alpha$ pre-multiplied by $\mathbf{D}^{\alpha-1}$(x-y axes) for $n = 2000$, $K = 3$, $\mu = \frac{3}{4}\delta_{q_{(1)}} + \frac{1}{4}\delta_{q_{(2)}}$, $q_{(1)} = 0.1$, $q_{(2)} = 0.5$, $c_1 = c_2 = \frac{1}{4}$, $c_3 = \frac{1}{2}$, $\mathbf{M} = 100\mathbf{I}_3$ with $\hat{\alpha}_{\text{opt}}$ defined in Section 3.4. Same setting as Figure 1.

distribution of $\mathbf{L}_\alpha$. Then, following popular spiked model tools, we will find conditions for the existence of isolated eigenvalues. This is the objective of the next sections.

3.2 Limiting support

In this section, we characterize the limiting eigenvalue distribution of $\mathbf{L}_\alpha$ where most eigenvalues concentrate. This in turn shall allow to determine the transition point beyond which informative eigenvalues isolate from the main bulk of eigenvalues and consistent clustering can thus be achieved by using the corresponding eigenvectors associated to those eigenvalues. The limiting eigenvalue distribution of $\mathbf{L}_\alpha$ is given in the following result.

Theorem 3 (Limiting spectrum) *Let $\pi_n^\alpha = \frac{1}{n}\sum_{i=1}^n \delta_{\lambda_i(\mathbf{L}_\alpha)}$ be the empirical spectral distribution (e.s.d.) of $\mathbf{L}_\alpha$. Then, as $n \to \infty$, $\pi_n^\alpha \to \bar{\pi}^\alpha$ almost surely where $\bar{\pi}^\alpha$ is a probability measure with compact symmetric support $\mathcal{S}^\alpha = [-S_\alpha, S_\alpha]$ defined, for $z \in \mathbb{C}^+ \setminus \mathcal{S}^\alpha$, by its Stieltjes transform*

$$m^\alpha(z) \equiv \int (t-z)^{-1}\,\mathrm{d}\bar{\pi}^\alpha(t) = \int \frac{1}{-z - f^\alpha(z)q^{1-2\alpha} + g^\alpha(z)q^{2-2\alpha}}\mu(dq)$$

where $(f^\alpha(z), g^\alpha(z)) \in (\mathbb{C}^+)^2$ (resp., $(\mathbb{R}^-)^2$) is the unique solution for $z \in \mathbb{C}^+$ (resp., $\mathbb{R}^+$), of

$$f^\alpha(z) = \int \frac{q^{1-2\alpha}\mu(dq)}{-z - f^\alpha(z)q^{1-2\alpha} + g^\alpha(z)q^{2-2\alpha}}$$
$$g^\alpha(z) = \int \frac{q^{2-2\alpha}\mu(dq)}{-z - f^\alpha(z)q^{1-2\alpha} + g^\alpha(z)q^{2-2\alpha}}. \tag{2}$$

Sketch of Proof *Since $\tilde{\mathbf{L}}_\alpha = \frac{1}{\sqrt{n}}\mathbf{D}_q^{-\alpha}\mathbf{X}\mathbf{D}_q^{-\alpha} + \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\mathsf{T}$ is a spiked random matrix, the e.s.d. π_n^α of $\mathbf{L}_\alpha$ converges weakly to the the e.s.d. $\tilde{\pi}_n^\alpha$ of $\frac{1}{\sqrt{n}}\mathbf{D}_q^{-\alpha}\mathbf{X}\mathbf{D}_q^{-\alpha}$ (by Weyl interlacing lemma) since $\mathbf{U}\mathbf{\Lambda}\mathbf{U}^\mathsf{T}$ is a low rank matrix. We thus find an asymptotic limit $\bar{\pi}^\alpha$ for $\tilde{\pi}_n^\alpha$ so that $\pi_n^\alpha \to \bar{\pi}^\alpha$ almost surely. To do so, we show that the Stieltjes transform of $\tilde{\pi}_n^\alpha$ converges to $m^\alpha(z)$ for $z \in \mathbb{C}^+$, which is the Stieltjes transform of the probability measure $\bar{\pi}^\alpha$ so that* *the convergence also holds for the probability measures (the e.s.d.). The Stieltjes transform of the e.s.d.* $\tilde{\pi}_n^\alpha$ *is* $n^{-1}\operatorname{tr}(\frac{1}{\sqrt{n}}\mathbf{D}_q^{-\alpha}\mathbf{X}\mathbf{D}_q^{-\alpha} - z\mathbf{I}_n)^{-1}$ *(where* $(\frac{1}{\sqrt{n}}\mathbf{D}_q^{-\alpha}\mathbf{X}\mathbf{D}_q^{-\alpha} - z\mathbf{I}_n)^{-1}$ *is the so-called resolvent of the random matrix* $\frac{1}{\sqrt{n}}\mathbf{D}_q^{-\alpha}\mathbf{X}\mathbf{D}_q^{-\alpha}$*), the deterministic limit of which gives* $m^\alpha(z)$*, computed using classical random matrix theory (RMT) tools (Pastur et al., 2011). The calculus details are provided in Section 6.2.*

Remark 4 (Stochastic Block Model) *Particularizing Theorem 3 to the Stochastic Block Model (SBM) (where* $q_i = q_0$ *for all* i*), the limiting probability measure* $\bar{\pi}^\alpha$ *is the popular semi-circle distribution with density* $\bar{\pi}^\alpha(dt) = \frac{2}{\pi(S^\alpha)^2}\sqrt{\max\{(S^\alpha)^2 - t^2, 0\}}dt$ *with* $S^\alpha = 2q_0^{1-2\alpha}\sqrt{1-q_0^2}$*. The associated Stieljes transform* $m^\alpha(z)$ *is explicit with in particular*

$$q_0^{1-2\alpha}m^\alpha(zq_0^{1-2\alpha}) = q_0^{\frac{1}{2}-\alpha}f^\alpha(zq_0^{\frac{1}{2}-\alpha}) = q_0^{-1}g^\alpha(zq_0^{1-2\alpha}) = -\frac{z}{2(1-q_0^2)} - \sqrt{\left(\frac{z}{2(1-q_0^2)}\right)^2 - \frac{1}{1-q_0^2}}.$$

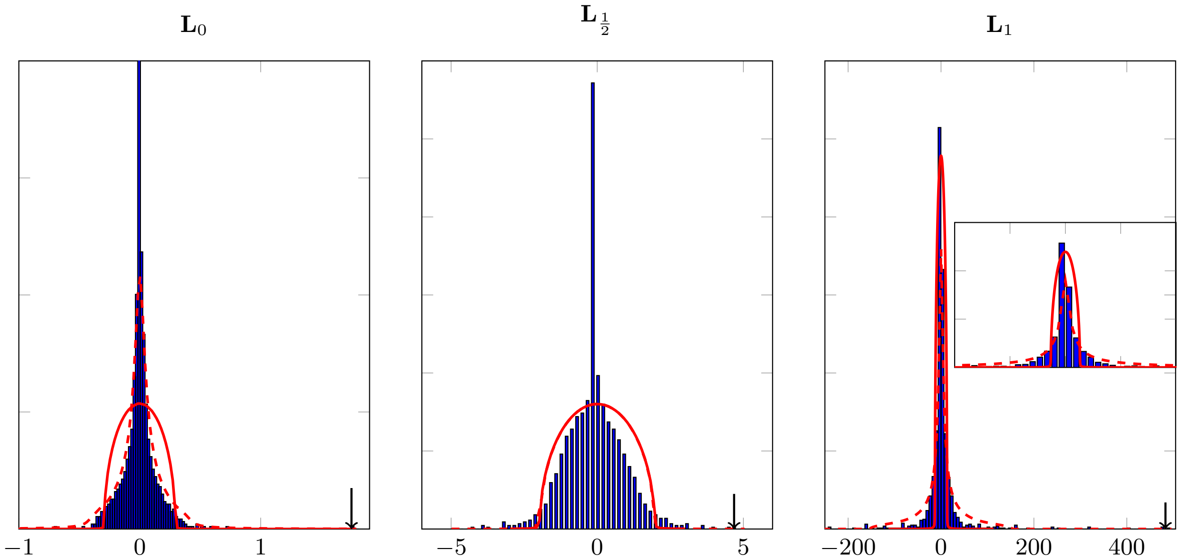


Figure 4: Political blogs (Adamic and Glance, 2005) network. Empirical versus Theoretical law of the eigenvalues of $\mathbf{L}_{\hat{\alpha}_{\text{opt}}}$ when fitting this network with the DCSBM (dashed) and the SBM (solid). Here $\hat{\alpha}_{\text{opt}} = 0$. The arrow shows the position of the largest eigenvalue.

The top of Figure 2, already discussed above, shows the density of the limiting $\bar{\pi}^\alpha$, for $\alpha = 0$, superimposed over the histogram of π_n^α. Figure 4 similarly displays the histogram π^α of the empirical eigenvalues of $\mathbf{L}_\alpha$ corresponding to the real network of Political blogs (Adamic and Glance, 2005) versus the theoretical limiting distribution $\bar{\pi}^\alpha$ obtained by fitting the network to the DCSBM (from Theorem 3, with μ the actual degree distribution of the graph) and the theoretical limiting distribution obtained by fitting the network to the SBM instead (in solid lines).[7] We note importantly that the DCSBM is a good fit for the political blogs

7. The SBM assumes here $q_i = q_0$ for all i.

network except possibly for $L_{\frac{1}{2}}$ while the SBM does not fit the network in any case. This suggests that the DCSBM is a more appropriate model when studying real world networks.

3.3 Phase transition

We also observe in Figure 4 (and more obviously in the synthetic case of Figure 2) that different choices of α lead to different behaviors in the position of the dominant eigenvalues. We shall determine here when separation of one or several eigenvalues from the bulk occurs. To this end, we follow popular spiked model techniques (Benaych-Georges and Nadakuditi, 2012; Hachem et al., 2013) for phase transition characterization. This entails the following result.

Theorem 5 (Phase transition) *Let Assumption 1 hold and let $\lambda(\bar{\mathbf{M}})$ be a non zero eigenvalue with multiplicity η of $\bar{\mathbf{M}} \equiv \left(\mathcal{D}(\mathbf{c}) - \mathbf{c}\mathbf{c}^{\mathsf{T}}\right)\mathbf{M}$. Then, for $\alpha \in \mathbb{R}$, there exists corresponding isolated eigenvalues $\lambda_i(\mathbf{L}_\alpha), \ldots, \lambda_{i+\eta-1}(\mathbf{L}_\alpha) \in \mathbb{R}\backslash\mathcal{S}^\alpha$ of $\mathbf{L}_\alpha$ all converging to $\rho \in \mathbb{R} \setminus \mathcal{S}^\alpha$, as $n \to \infty$, almost surely, if and only if* [8]

$$\left|\lambda(\bar{\mathbf{M}})\right| > \tau^\alpha \triangleq -\lim_{x \downarrow \mathcal{S}^\alpha} \frac{1}{g^\alpha(x)},$$

with $g^\alpha(x)$ defined in Theorem 3. In this case, ρ is defined by

$$\rho = (g^\alpha)^{-1}\left(-\frac{1}{\lambda(\bar{\mathbf{M}})}\right).$$

Sketch of Proof *From Theorem 3, the e.s.d. of $\mathbf{L}_\alpha$ converges weakly to the e.s.d. of $\frac{1}{\sqrt{n}}\mathbf{D}_q^{-\alpha}\mathbf{X}\mathbf{D}_q^{-\alpha}$ with support $\mathcal{S}^\alpha$ (defined in Theorem 3) but since $\frac{1}{\sqrt{n}}\mathbf{D}_q^{-\alpha}\mathbf{X}\mathbf{D}_q^{-\alpha}$ and $\mathbf{L}_\alpha$ only differ by a finite rank matrix $\mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\mathsf{T}}$, some eigenvalues of $\mathbf{L}_\alpha$ may isolate from the support $\mathcal{S}^\alpha$. To find those isolated eigenvalues, we solve for $\rho \notin \mathcal{S}^\alpha$, $\det(\mathbf{L}_\alpha - \rho\mathbf{I}_n) = 0$. This leads to find the ρ's for which $0 = \det(\mathbf{I}_{K+1} + \mathbf{U}^{\mathsf{T}}\mathbf{Q}_\rho^\alpha\mathbf{U}\mathbf{\Lambda})$ where $\mathbf{Q}_\rho^\alpha = (\frac{1}{\sqrt{n}}\mathbf{D}_q^{-\alpha}\mathbf{X}\mathbf{D}_q^{-\alpha} - \rho\mathbf{I}_n)^{-1}$ is the resolvent of $\frac{1}{\sqrt{n}}\mathbf{D}_q^{-\alpha}\mathbf{X}\mathbf{D}_q^{-\alpha}$. By using standard RMT calculus (Benaych-Georges and Nadakuditi, 2012), we obtain a deterministic approximation of $\mathbf{I}_{K+1} + \mathbf{U}^{\mathsf{T}}\mathbf{Q}_z^\alpha\mathbf{U}\mathbf{\Lambda}$ which leads to the phase transition condition in Theorem 5.*

Remark 6 ($\boldsymbol{\tau^\alpha}$ in SBM setting) *From Remark 4, in the SBM setting, τ^α no longer depends on α and is given by $\tau^\alpha = \frac{\sqrt{1-q_0^2}}{q_0}$.*

Remark 7 (Number of isolated eigenvalues) *From Theorem 5, there is a one-to-one mapping between the limiting isolated eigenvalues ρ of $\mathbf{L}_\alpha$ and non zero eigenvalues of $\bar{\mathbf{M}} = \left(\mathcal{D}(\mathbf{c}) - \mathbf{c}\mathbf{c}^{\mathsf{T}}\right)\mathbf{M}$. As $\mathbf{1}_K^{\mathsf{T}}\bar{\mathbf{M}} = 0$, $\bar{\mathbf{M}}$ has a maximum of $K-1$ non zero eigenvalues which means that at most $K-1$ eigenvalues of $\mathbf{L}_\alpha$ can be found at macroscopic distance from $\mathcal{S}^\alpha$. Thus, at most $K-1$ eigenvectors of $\mathbf{L}_\alpha$ can be used in the first step of the spectral algorithm described in the introduction.*

Remark 8 (The complete spectrum of $\mathbf{L}_{\boldsymbol{\alpha}}$) *Strictly speaking, the aforementioned statements are somewhat inaccurate. An exhaustive analysis of $\mathbf{L}_\alpha$ indeed reveals that, under some*

8. The limit $\lim_{x\downarrow\mathcal{S}^\alpha} g^\alpha(x)$ is well defined in $(-\infty, 0]$ as $x \mapsto g^\alpha(x)$ can be shown to be a continuous growing negative function on the right side of $\mathcal{S}^\alpha$.

conditions on μ, and irrespective of the clustering matrix $\mathbf{M}$, extra isolated eigenvalues can be found in the spectrum of $\mathbf{L}_\alpha$, the eigenvectors of which do not contain any structural information about the classes. This rather unfamiliar scenario has also been evidenced in the context of spectral kernel clustering in (Couillet et al., 2016). Since this hypothetical eigenvalue and eigenvector pair is of no value for the interest of clustering, it shall no longer be discussed in the following. Besides, most settings of practical interest do not present this singular behavior. A thorough discussion of this peculiarity is provided in Section 6.5.

The value τ^α defined in Theorem 5 is a community detectability threshold which in the dense regime for the SBM case was shown to split the community detectability into two regions: a region where no algorithm can succed better than a random guess in classifying the nodes and a region where a non trivial detection is possible (Decelle et al., 2011; Nadakuditi and Newman, 2012). When the separability condition of Theorem 5 is ensured, the alignment between the properly normalized eigenvectors of $\mathbf{L}_\alpha$ and linear combinations of the class vectors $\mathbf{j}_a$'s (defined in Theorem 1) is away from zero, thus ensuring a non trivial classification performance. The larger $\lambda(\bar{\mathbf{M}})$, the closer are the vectors used for classification to the class vectors $\mathbf{j}_a$'s.

Theorem 9 *Under Assumption 1, let $\lambda(\bar{\mathbf{M}})$ and $\lambda(\mathbf{L}_\alpha)$ be an eigenvalue pair as defined in Theorem 5. We further assume $\lambda(\bar{\mathbf{M}})$ of unit multiplicity and denote $\mathbf{u}$ the eigenvector associated to the eigenvalue $\lambda(\mathbf{L}_\alpha)$. Then, letting $\bar{\mathbf{v}} = \frac{\mathbf{D}^{\alpha-1}\mathbf{u}}{\|\mathbf{D}^{\alpha-1}\mathbf{u}\|}$ and $\mathbf{\Pi} = \sum_{a=1}^{K} \frac{\mathbf{j}_a\mathbf{j}_a^\mathsf{T}}{n_a}$, for all $\epsilon > 0$, there exists $\gamma_-, \gamma_+ > 0$ such that, for all n large, almost surely,*

$$0 < \left|\lambda(\bar{\mathbf{M}})\right| - \tau^\alpha < \gamma_- \Rightarrow \bar{\mathbf{v}}^\mathsf{T}\mathbf{\Pi}\bar{\mathbf{v}} < \epsilon$$
$$\left|\lambda(\bar{\mathbf{M}})\right| - \tau^\alpha > \gamma_+ \Rightarrow \bar{\mathbf{v}}^\mathsf{T}\mathbf{\Pi}\bar{\mathbf{v}} > 1 - \epsilon.$$

This result is a direct corrolary of Theorem 14 which is introduced later in Section 3.5.

Figure 5 illustrates Theorem 9, which confirms that, below the phase transition threshold τ^α, there is asymptotically no correlation between the vectors $\bar{\mathbf{v}}$ and the class canonical vectors $\mathbf{j}_a$'s and thus no consistent clustering can be achieved in this regime. The theoretical curve is obtained by using the deterministic asymptotic approximation of $\bar{\mathbf{v}}^\mathsf{T}\mathbf{\Pi}\bar{\mathbf{v}}$ which is explicitely given in Section 6.4.

3.4 Optimal α

In this section, we determine the values of α for which the community detectability threshold is maximally achieved. This, in turn, is expected to allow for the optimal extraction of information about the classes from the extreme eigenvectors although this is not easily proved (and in our opinion, most likely not always true).

From Theorem 5, since $\bar{\mathbf{M}}$ does not depend on α, the smaller τ^α the more likely the detectability condition $\left|\lambda(\bar{\mathbf{M}})\right| > \tau^\alpha$ is met. We then seek α for which τ^α is minimal. For any compact set $\mathcal{A} \subset \mathbb{R}$, we may thus define

$$\alpha_{\rm opt} \triangleq \operatorname{argmin}_{\alpha\in\mathcal{A}} \{\tau^\alpha\}$$

which we shall assume is unique (if $q_i = q_0$ is constant, τ^α is constant across α; this case is thus excluded). The estimation of $\alpha_{\rm opt}$ however requires the knowledge of $g^\alpha(x)$ for each $\alpha \in \mathcal{A}$. The estimation of $g^\alpha(x)$ can be done numerically by solving the fixed point equation

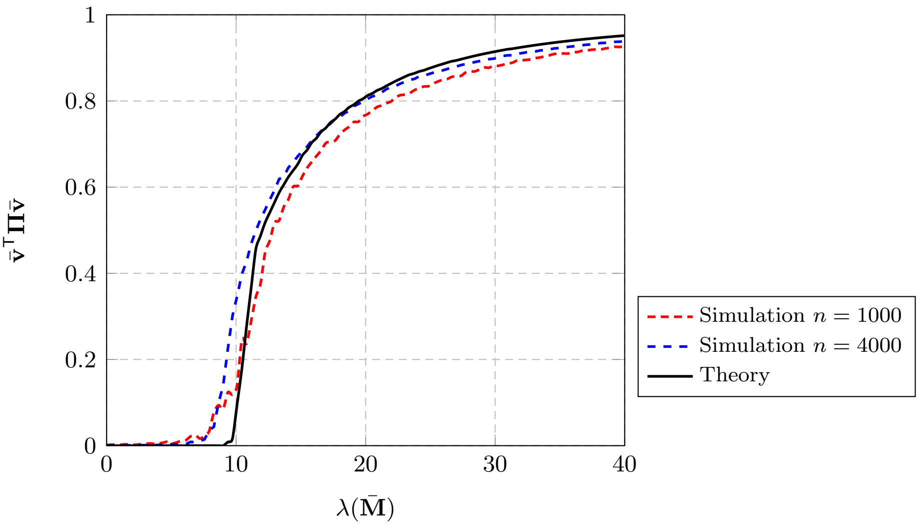


Figure 5: Simulated versus empirical $\bar{\mathbf{v}}^\mathsf{T}\mathbf{\Pi}\bar{\mathbf{v}}$ for $K = 3$, $\mu = \frac{3}{4}\delta_{q_{(1)}} + \frac{1}{4}\delta_{q_{(2)}}$, $q_{(1)} = 0.1$, $q_{(2)} = 0.2$, $c_1 = c_2 = \frac{1}{4}$, $c_3 = \frac{1}{2}$, $\mathbf{M} = \Delta\mathbf{I}_3$ with Δ ranging from 0 to 100.

defined in Theorem 3 provided μ is known. As a direct consequence of Assumption 1-(1), μ can in fact be estimated from the empirical graph degrees irrespective of the class matrix $\mathbf{C}$, according to the following result.

Lemma 10 *Let* $\hat{q}_i = \frac{d_i}{\sqrt{\mathbf{d}^\mathsf{T}\mathbf{1}_n}}$. *Then, under Assumption 1,*

$$\max_{1\leq i\leq n} |q_i - \hat{q}_i| \to 0 \tag{3}$$

almost surely.

We thus have all the ingredients to estimate α_{opt}.[9]

Proposition 11 *Define* $\hat{\mu} \triangleq \frac{1}{n}\sum_{i=1}^n \delta_{\hat{q}_i}$ *with* $\hat{q}_i = \frac{d_i}{\sqrt{\mathbf{d}^\mathsf{T}\mathbf{1}_n}}$ *and* $\hat{\mathcal{S}}^\alpha$, $\hat{f}^\alpha(z)$, $\hat{g}^\alpha(z)$, *as in Theorem 3 but for* μ *replaced by* $\hat{\mu}$. *Then, as* $n \to \infty$,

$$\hat{\alpha}_{\text{opt}} \to \alpha_{\text{opt}}$$

almost surely, where $\hat{\alpha}_{\text{opt}} \triangleq \operatorname{argmin}_{\alpha\in\mathcal{A}}\{\hat{\tau}^\alpha\}$ *with*

$$\hat{\tau}_\alpha \equiv -\frac{1}{\lim_{x\downarrow\hat{\mathcal{S}}^\alpha} \hat{g}^\alpha(x)}.$$

Remark 12 (Numerical evaluation of S^α) *Estimating* $\hat{\tau}_\alpha$ *requires to determine* $\hat{\mathcal{S}}^\alpha$. *To this end, we use the fact that* $\hat{g}^\alpha(x)$ *is only defined for* $x \notin \hat{\mathcal{S}}^\alpha$. *We thus evaluate* $\hat{\mathcal{S}}^\alpha$ *by an iterative dichotomic search in intervals of the type* $[l, r]$ *for which* $\hat{g}^\alpha(l)$ *is undefined (and thus the algorithm in Equation 2 does not converge) and* $\hat{g}^\alpha(r)$ *is defined (the algorithm converges), starting from e.g.,* $l = 0$ *and* r *quite large.*

9. Note here that imposing $\mathcal{A}$ to be a compact set ensures the uniform validity of Theorem 3.

Remark 13 (Relevance of the choice of α) *Following Remarks 4 and 6, note that the choice of α is only relevant to heterogeneous graphs, as in the SBM case, the phase transition threshold τ^α is constant irrespective of α. This suggests that the more heterogeneous the graph the more important an appropriate setting of α.*

The aforementionned importance of choosing $\alpha = \hat{\alpha}_{\rm opt}$ along with the need to pre-multiply the dominant eigenvectors of $\mathbf{L}_\alpha$ by $\mathbf{D}^{\alpha-1}$ before classification, as discussed after exposing Theorem 1, naturally bring us to an improved version of Algorithm 1 provided below. The

---

Algorithm 2: Improved spectral algorithm

1: Evaluate $\alpha = \hat{\alpha}_{\rm opt} = \operatorname{argmin}_{\alpha\in\mathcal{A}} \lim_{x\downarrow \hat{S}^\alpha} \hat{g}^\alpha(x)$ as per Proposition 11.
2: Retrieve the ℓ eigenvectors corresponding to the ℓ largest eigenvalues of $\mathbf{L}_\alpha = (2m)^\alpha \frac{1}{\sqrt{n}}\mathbf{D}^{-\alpha}\left[\mathbf{A} - \frac{\mathbf{d}\mathbf{d}^{\sf T}}{2m}\right]\mathbf{D}^{-\alpha}$. Denote $\mathbf{u}_1^\alpha, \ldots, \mathbf{u}_\ell^\alpha$ those eigenvectors.
3: Letting $\mathbf{v}_i^\alpha = \mathbf{D}^{\alpha-1}\mathbf{u}_i^\alpha$ and $\bar{\mathbf{v}}_i^\alpha = \frac{\mathbf{v}_i^\alpha}{\|\mathbf{v}_i^\alpha\|}$, stack the vectors $\bar{\mathbf{v}}_i^\alpha$'s columnwise in a matrix $\mathbf{W} = [\bar{\mathbf{v}}_1^\alpha, \ldots, \bar{\mathbf{v}}_\ell^\alpha] \in \mathbb{R}^{n\times\ell}$.
4: Let $\mathbf{r}_1, \ldots, \mathbf{r}_n \in \mathbb{R}^\ell$ be the rows of $\mathbf{W}$. Cluster $\mathbf{r}_i \in \mathbb{R}^\ell$, $1 \le i \le n$ in one of the K groups using any low-dimensional classification algorithm (e.g., k-means or EM). The label assigned to $\mathbf{r}_i$ then corresponds to the label of node i.

---

performances of Algorithm 2 mainly depend on the content of the eigenvectors $\bar{\mathbf{v}}_i^\alpha$'s. These regularized eigenvectors happen to be shaped like noisy "plateaus" (step functions), each plateau characterizing a class. The objective of the next section is to provide deterministic limits of the parameters of those noisy plateaus from which the asymptotic performances of Algorithm 2 unfold.

3.5 Eigenvectors and improvement of Expectation Maximization (EM) algorithm

In this section, we provide a precise characterization of the asymptotic class means and class covariances of the dominant eigenvectors entries (used for clustering) which in turn allows to improve the classical EM algorithm used in the last step of spectral clustering procedures. The eigenvectors of $\mathbf{L}_\alpha$ have the property of remaining "stable" in the large dimensional limit, thereby allowing for a precise characterization of their content. This behavior (classical in the spike model analysis of random matrices) however only holds for eigenvectors associated to strictly isolated eigenvalues (in the sense that the latter remain at macroscopic distance of all other eigenvalues). In the remainder, we thus assume that the normalized eigenvector $\bar{\mathbf{v}}_i^\alpha$ under study is associated with such a strictly isolated eigenvalue.

As one can see in Figure 3, the different clusters of points (rows of $\mathbf{W}$ in Algorithm 2) have different dispersions (variances) in the DCSBM model under consideration. The most appropriate algorithm to use in step 4 of Algorithm 2 is the expectation maximization (EM) method. EM considers each point $\mathbf{r}_i \in \mathbb{R}^\ell$ arising from $[\bar{\mathbf{v}}_1^\alpha, \ldots, \bar{\mathbf{v}}_\ell^\alpha]$ as a mixture of K Gaussian random vectors with means $\boldsymbol{\nu}_{EM}^a$ and covariances $\boldsymbol{\Sigma}_{EM}^a \in \mathbb{R}^{\ell\times\ell}$, $a \in \{1, \ldots, K\}$. Starting from initial means and covariances, they are sequentially updated until convergence. To identify $\boldsymbol{\nu}_{EM}^a$, $\boldsymbol{\Sigma}_{EM}^a$ and thus understand the performance of Algorithm 2, we may write $\bar{\mathbf{v}}_i^\alpha$ [10] as the

10. Recall that the graph nodes were assumed labeled by class, and thus the entries of $\bar{\mathbf{v}}_i^\alpha$ are similarly sorted by class.

"noisy plateaus" vector

$$\bar{\mathbf{v}}_i^{\alpha} = \sum_{a=1}^{K} \nu_i^a \frac{\mathbf{j}_a}{\sqrt{n_a}} + \sqrt{\sigma_{ii}^a}\mathbf{w}_i^a \tag{4}$$

where $\mathbf{w}_i^a \in \mathbb{R}^n$ is a random vector orthogonal to $\mathbf{j}_a$, of norm $\sqrt{n_a}$ and supported on the indices of $\mathcal{C}_a$ and

$$\nu_i^a = \frac{1}{\sqrt{n_a}} \left(\bar{\mathbf{v}}_i^{\alpha}\right)^{\mathsf{T}} \mathbf{j}_a = \frac{1}{\sqrt{n_a}} \frac{(\mathbf{u}_i^{\alpha})^{\mathsf{T}} \mathbf{D}^{\alpha-1} \mathbf{j}_a}{\sqrt{(\mathbf{u}_i^{\alpha})^{\mathsf{T}} \mathbf{D}^{2(\alpha-1)} \mathbf{u}_i^{\alpha}}} \tag{5}$$

$$\sigma_{ij}^a = \frac{(\mathbf{u}_i^{\alpha})^{\mathsf{T}} \mathbf{D}^{\alpha-1} \mathcal{D}_a \mathbf{D}^{\alpha-1} \mathbf{u}_j^{\alpha}}{\sqrt{(\mathbf{u}_i^{\alpha})^{\mathsf{T}} \mathbf{D}^{2(\alpha-1)} \mathbf{u}_i} \sqrt{(\mathbf{u}_j^{\alpha})^{\mathsf{T}} \mathbf{D}^{2(\alpha-1)} \mathbf{u}_j^{\alpha}}} - \nu_i^a \nu_j^a \tag{6}$$

with $\mathcal{D}_a = \mathcal{D}(\mathbf{j}_a)$. The vector $\boldsymbol{\nu}^a = (\nu_i^a)_{i=1}^{\ell} \in \mathbb{R}^{\ell}$ and the matrix $\boldsymbol{\Sigma}^a = (\sigma_{ij}^a)_{i,j=1}^{\ell} \in \mathbb{R}^{\ell \times \ell}$ represent respectively the empirical means and empirical covariances of the points $\mathbf{r}_i$ (defined in Algorithm 2) belonging to class $\mathcal{C}_a$. Thus, provided that EM converges to the correct solution, $(\boldsymbol{\nu}_{EM}^a)_i$ and $(\boldsymbol{\Sigma}_{EM}^a)_{ij}$ shall converge asymptotically to the limiting values of $\nu_i^a \in \mathbb{R}$ and σ_{ij}^a respectively. Clearly, for small values of $\boldsymbol{\Sigma}^a$ compared to $\boldsymbol{\nu}^a$, clustering the vectors $\bar{\mathbf{v}}_i^{\alpha}$ shall lead to good performances.

We find the asymptotic limits of the class means ν_i^a and the class covariances σ_{ij}^a. The explicit expressions of those limits are provided in the proof section (Theorems 21 and 22) for readability reasons.

Theorem 14 *For ν_i^a, σ_{ij}^a defined in (26), (27) respectively, there exist deterministic limits $\nu_i^{a,\infty}$ and $\sigma_{ij}^{a,\infty}$ (explicitely defined in Theorems 21 and 22 in Section 6.4) such that, as $n \to \infty$, almost surely*

$$\left|(\nu_i^a)^2 - (\nu_i^{a,\infty})^2\right| \to 0$$
$$\left|\sigma_{ij}^a - \sigma_{ij}^{a,\infty}\right| \to 0.$$

Sketch of Proof *Technically, the standard tools used in spiked random matrix analysis do not allow for an immediate assessment of the quantities ν_i^a and σ_{ij}^a. As a workaround, we follow the approach used in (Couillet et al., 2016) which relies on the possibility to estimate bilinear forms of the type $\mathbf{a}^{\mathsf{T}}\mathbf{u}_i^{\alpha}(\mathbf{u}_i^{\alpha})^{\mathsf{T}}\mathbf{b}$ for given vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ and unit multiplicity eigenvectors $\mathbf{u}_i^{\alpha}$ of $\mathbf{L}_{\alpha}$ since we have from Cauchy formula, as $n \to \infty$ almost surely, (since $\lambda_i(\mathbf{L}_{\alpha}) \to \rho$)*

$$\mathbf{a}^{\mathsf{T}}\mathbf{u}_i^{\alpha}(\mathbf{u}_i^{\alpha})^{\mathsf{T}}\mathbf{b} = -\frac{1}{2\pi i}\oint_{\Gamma_\rho} \mathbf{a}^{\mathsf{T}} \left(\mathbf{L}_{\alpha} - z\mathbf{I}_n\right)^{-1} \mathbf{b} \mathrm{d}z$$

and for a given matrix $\mathbf{D}$

$$(\mathbf{u}_i^{\alpha})^{\mathsf{T}}\mathbf{D}\mathbf{u}_i^{\alpha} = \operatorname{tr} \mathbf{u}_i^{\alpha}(\mathbf{u}_i^{\alpha})^{\mathsf{T}}\mathbf{D} = -\frac{1}{2\pi i}\oint_{\Gamma_\rho} \operatorname{tr} \left(\mathbf{L}_{\alpha} - z\mathbf{I}_n\right)^{-1} \mathbf{D} \mathrm{d}z$$

where Γ_ρ is a positively oriented contour circling around the limiting eigenvalue ρ of $\lambda_i(\mathbf{L}_{\alpha})$ associated to the eigenvector $\mathbf{u}_i^{\alpha}$ of $\mathbf{L}_{\alpha}$. The calculus details are provided in Section 6.4.

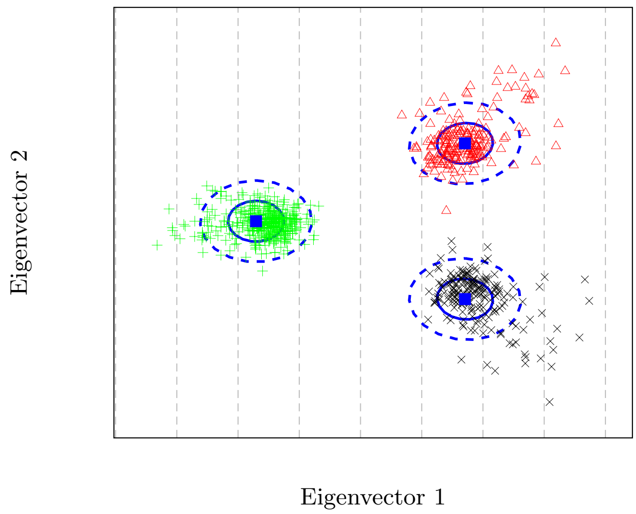


Figure 6: $n = 800$, $K = 3$ classes $\mathcal{C}_1$, $\mathcal{C}_2$ and $\mathcal{C}_3$ of sizes $|\mathcal{C}_1| = |\mathcal{C}_2| = \frac{n}{4}$, $|\mathcal{C}_3| = \frac{n}{2}$, $\frac{3}{4}$ of the nodes having $q_i = 0.3$ and the others having $q_i = 0.8$, matrix of weights $\mathbf{C} = \mathbf{1}_3\mathbf{1}_3^\mathsf{T} + \frac{30}{\sqrt{n}}\mathbf{I}_3$. Two dimensional representation of the dominant eigenvectors 1 and 2 of $\mathbf{L}_\alpha$. In blue, theoretical means and one- and two- standard deviations.

Using the asymptotic results in Theorem 14, we display in Figures 6 and 7 the theoretical means and standard deviations versus ground truths for each class-wise block of the eigenvectors entries. The good fit between the ground truths and the theoretical findings of the class means and class covariances, calls for the improvement of the random initialization of the EM procedure in the last step of spectral clustering.

The performances of EM highly depend on the chosen starting parameters; a first natural choice is to set them randomly, which as we shall see leads to poor performances especially in cases where the clusters are not easily separable. Since the theoretical limiting means $\boldsymbol{\nu}^{a,\infty}$ and covariances $\boldsymbol{\Sigma}^{a,\infty}$ are respectively the limiting values of $\boldsymbol{\nu}_{EM}^a$ and covariances $\boldsymbol{\Sigma}_{EM}^a$ provided EM converges to the correct solution, we may set as initial parameters of EM our findings $\boldsymbol{\nu}^{a,\infty}$ (Theorem 21) and $\boldsymbol{\Sigma}^{a,\infty}$ (Theorem 22) for $a \in \{1,\ldots,K\}$ provided those can be estimated. In most scenarios, the many unknowns prevent such an estimation. Nonetheless, from Corollary 24 (Section 6.4), provided the class proportions (or the sizes of each class) are (more or less) known, we can consistently estimate $\boldsymbol{\nu}^\infty$ and $\boldsymbol{\Sigma}^\infty$ in a 2-class scenario. As we shall see, this new setting of initial parameters is much better than other initializations approaches.

To show the effect of our setting of initial parameters of EM based on the findings $\boldsymbol{\nu}^\infty$ and $\boldsymbol{\Sigma}^\infty$, Figure 8 compares the empirical performances of our new spectral algorithm based on the regularized eigenvector of $\mathbf{L}_{0.5}$ for different initial settings of the EM parameters *i)* random setting (Random EM) *ii)* our theoretical setting (by assuming that the class proportions are known) and *iii)* the ground truth setting (oracle EM where we set the initial points to the empirically evaluated means and covariances of each cluster based on ground truth). Below the transition point, no consistent clustering can be achieved for large n using the eigenvectors associated to highest eigenvalues since the clusters are not separable and our theoretical limiting means and covariances are not defined since there is no isolated eigenvalues

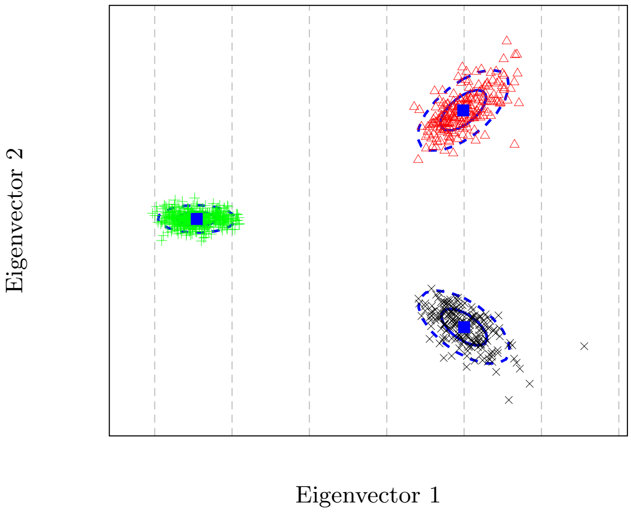


Figure 7: $n = 800$, $K = 3$ classes $\mathcal{C}_1$, $\mathcal{C}_2$ and $\mathcal{C}_3$ of sizes $|\mathcal{C}_1| = |\mathcal{C}_2| = \frac{n}{4}$, $|\mathcal{C}_3| = \frac{n}{2}$, q_i's uniformly distributed over $[0.1, 0.9]$, matrix of weights $\mathbf{C} = \mathbf{1}_3\mathbf{1}_3^\mathsf{T} + \frac{100}{\sqrt{n}}\mathbf{I}_3$. Two dimensional representation of the dominant eigenvectors 1 and 2 of $\mathbf{L}_\alpha$. In blue, theoretical means and one- and two- standard deviations

in that case. We have thus initialized EM at random in this non interesting regime (as for Random EM). The EM algorithm may in that regime set all the nodes to the same cluster, which will then result to a classification rate close to the proportion of the nodes in the cluster of largest size. In the interesting regime (after the transition point), we see that the performances (in terms of correct classification rate) of the algorithm using our theoretical setting of EM closely match the performances of an ideal setting with ground truth (oracle EM). The performances of the algorithm using a random initialization (Random EM) are completely degraded especially around critical cases (small values of Δ). Random EM becomes reliable only for very large values of Δ where clustering is somewhat trivial.

4. Numerical simulations

We restrict ourselves to $\alpha \in \mathcal{A} = [0, 1]$ for the numerical simulations. To illustrate the importance of the choice of α_{opt}, Figure 9 presents the theoretical (asymptotic) ratio between the limiting largest eigenvalue ρ of $\mathbf{L}_\alpha$ and the right edge S^α of the limiting support $\mathcal{S}^\alpha$ with respect to the amplitude of the eigenvalues of $\bar{\mathbf{M}}$. Although α_{opt} only ensures in theory to have the best isolation of the eigenvalues only in "worst cases scenarios"(i.e., when $\lambda(\bar{\mathbf{M}})$ is slighty larger than $\tau^{\alpha_{\text{opt}}}$), Figure 9 shows that taking $\alpha = \alpha_{\text{opt}}$ provides the largest gap $\frac{\rho}{S^\alpha}$ for all values of $\lambda(\bar{\mathbf{M}})$. This suggests (again, without any theoretical support) best performances with $\alpha = \alpha_{\text{opt}}$ in all cases (for any value of $\mathbf{M}$).

In the sequel, to compare the different algorithms, we will use the performance evaluation measure known as *overlap to ground truth communities*, defined in (Krzakala et al., 2013) as

$$\text{Overlap} \equiv \frac{\frac{1}{n}\sum_{i=1}^{n} \delta_{g_i \hat{g}_i} - \frac{1}{K}}{1 - \frac{1}{K}},$$

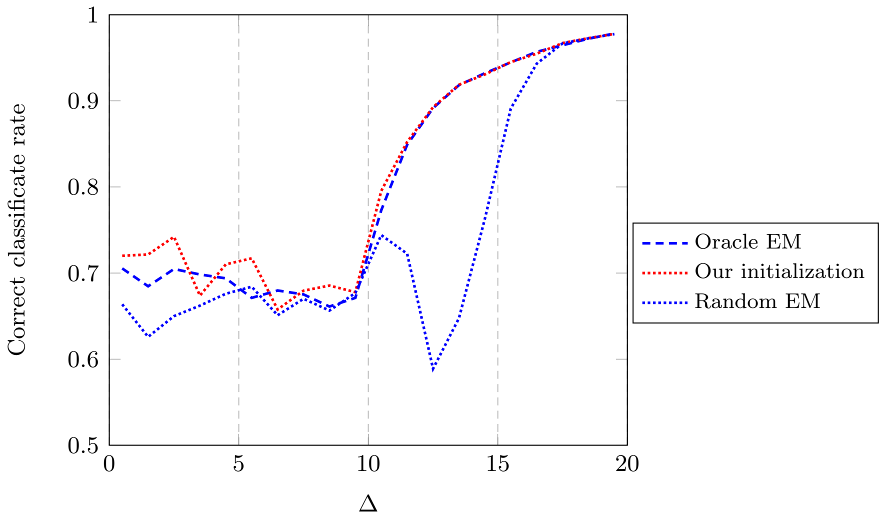


Figure 8: Probability of correct recovery for $\alpha = 0.5$, $n = 4000$, $K = 2$, $c_1 = 0.8$, $c_2 = 0.2$, $\mu = \frac{3}{4}\delta_{q_{(1)}} + \frac{1}{4}\delta_{q_{(2)}}$ with $q_{(1)} = 0.2$ and $q_{(2)} = 0.8$, $\mathbf{M} = \Delta\mathbf{I}_2$, for $\Delta \in [0, 20]$.

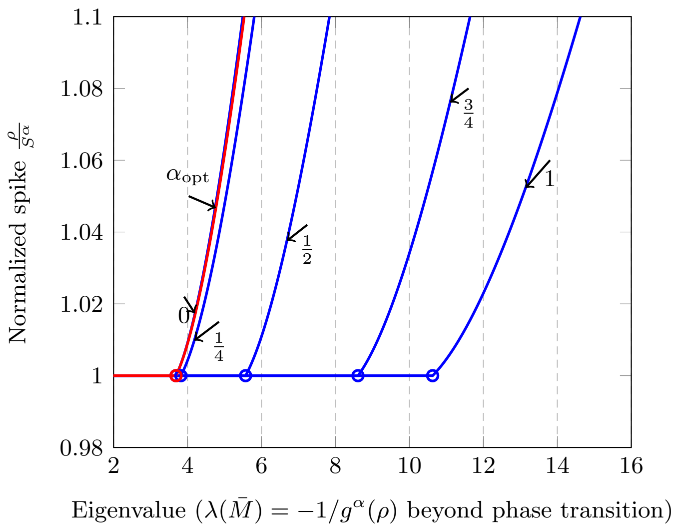


Figure 9: Ratio between the limiting largest eigenvalue ρ of $\mathbf{L}_\alpha$ and the right edge of the support $\mathcal{S}^\alpha$, as a function of the largest eigenvalue $\lambda(\bar{\mathbf{M}})$ of $\bar{\mathbf{M}}$, $\mathbf{M} = \Delta\mathbf{I}_3$, $c_i = \frac{1}{3}$, for $\Delta \in [10, 150]$, $\mu = \frac{3}{4}\delta_{q_{(1)}} + \frac{1}{4}\delta_{q_{(2)}}$ with $q_{(1)} = 0.1$ and $q_{(2)} = 0.5$, for $\alpha \in \{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1, \alpha_{\rm opt}\}$ (indicated on the curves of the graph). Here, $\alpha_{\rm opt} = 0.07$. Circles indicate phase transition.

where g_i and $\hat{g}_i$ are the true and estimated labels of node i, respectively. Figure 10 subsequently shows the overlap performance under the setting of Figure 9 for a simulated graph of $n = 3000$ nodes. Note that the empirically observed phase transitions closely match the theoretical ones (drawn in circles and the same as in Figure 9). We then consider in Fig-

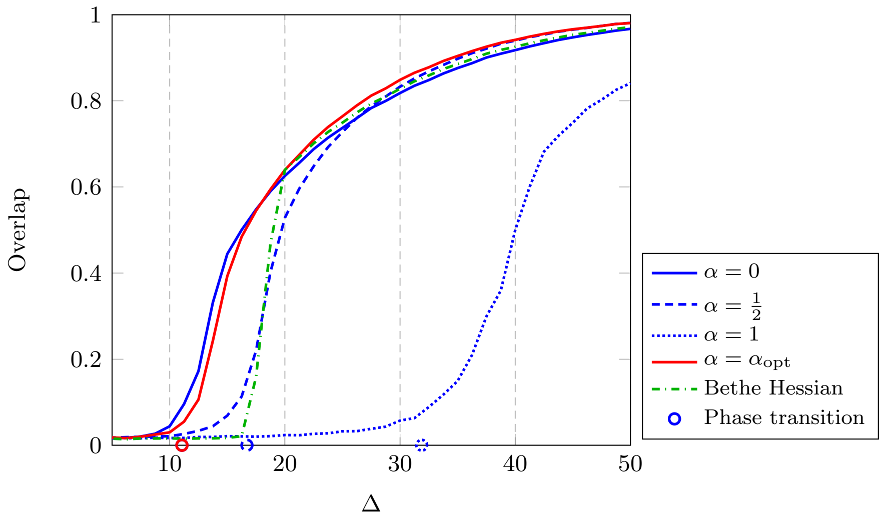


Figure 10: Overlap performance for $n = 3000$, $K = 3$, $c_i = \frac{1}{3}$, $\mu = \frac{3}{4}\delta_{q_{(1)}} + \frac{1}{4}\delta_{q_{(2)}}$ with $q_{(1)} = 0.1$ and $q_{(2)} = 0.5$, $\mathbf{M} = \Delta\mathbf{I}_3$, for $\Delta \in [5, 50]$. Here $\alpha_{\text{opt}} = 0.07$.

ure 11 a DCSBM graph where $\mathbf{M}$ is fixed and three quarters of the nodes connect with a fixed intrinsic low weight $q_{(1)} = 0.1$ and we vary the intrinsic weights $q_{(2)}$ of the remaining quarter of the nodes from low to high weights. We observe a sudden drop of the BH overlap for large $q_{(2)} - q_{(1)}$. This phenomenon is consistent with the fact, observed earlier in Figure 1, that BH creates artificial communities out of nodes with the same q_i parameter. This is a practical demonstration of the need for a proper eigenvector normalization to avoid degree biases. This observation has recently led (Newman, 2013) to consider a regularization for the non-backtracking operator on which the BH method is based, which still awaits for proper analysis.

In Figure 12, we consider a more realistic synthetic graph where the q_i's assume a power law of support $[0.05, 0.3]$ which simulates a sparse graph characteristic of real world networks. Although this is not the regime we study in this article, our method for $\alpha = \hat{\alpha}_{\text{opt}}$ still competes with the BH method which was developped for sparse homogeneous graphs. However, it is seen that the theoretical phase transitions do not closely match the empirical ones especially for the case $\alpha = 1$. This mismatch is likely due to the fact that our theoretical results in this article require $P_{ij} = \Omega(1)$ which is not always the case in this scenario.

We finally confront the performances (in terms of overlap and modularity [11]) of the different spectral algorithms on the Political blogs graph (Adamic and Glance, 2005) in Table 1. We should note that while $\alpha_{\text{opt}} = 0$ in this case, it achieves the best performance both in terms of the overlap to the ground truth and of the modularity. [12] Likely, the reason why

11. The modularity Q for a given graph partition with class labels g_i's is defined as : $Q = \frac{1}{2m}\sum_{i,j=1}^{n}\left(A_{ij} - \frac{d_i d_j}{2m}\right)\delta_{g_i=g_j}$ where $\mathbf{d} = \mathbf{A}\mathbf{1}_n$ is the degree vector and $m = \frac{1}{2}\mathbf{d}^{\mathsf{T}}\mathbf{1}_n$ is the total number of edges.
12. We should note here that the scores for the BH are different from the ones found in the article (Saade et al., 2014) since here we are running k-means algorithm in the last step of the spectral algorithm while the authors of (Saade et al., 2014) have instead used a sign classification of the eigenvector components for networks with two communities.

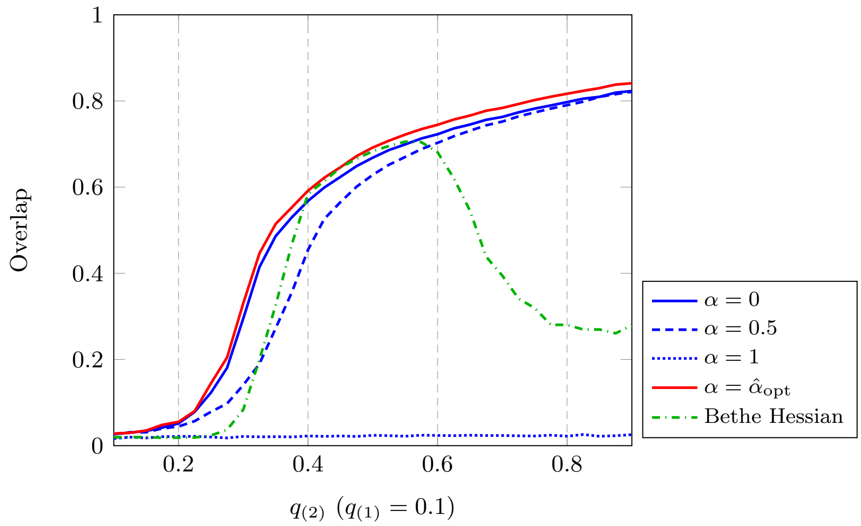


Figure 11: Overlap for $n = 3000$, $K = 3$, $\mu = \frac{3}{4}\delta_{q_{(1)}} + \frac{1}{4}\delta_{q_{(2)}}$ with $q_{(1)} = 0.1$ and $q_{(2)} \in [0.1, 0.9]$, $\mathbf{M}$ defined by $M_{ii} = 10$, $M_{ij} = -10, i \neq j$, $c_i = \frac{1}{3}$.

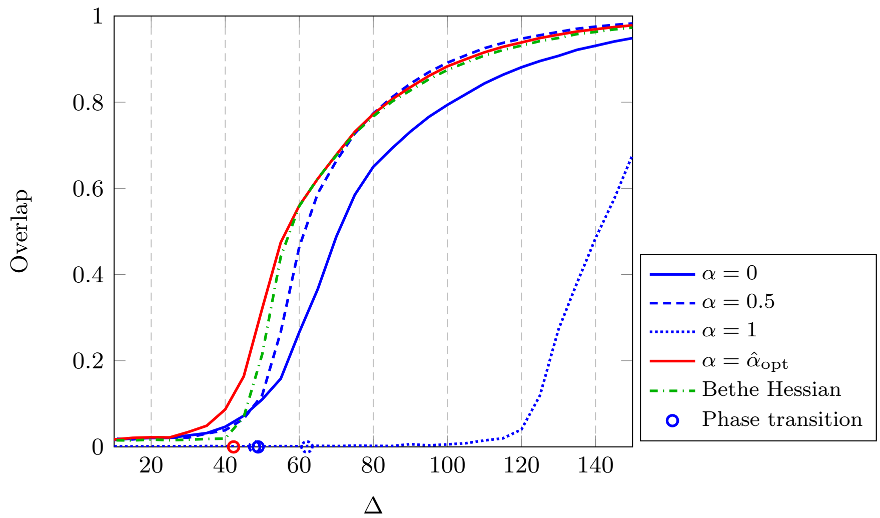


Figure 12: Overlap for $n = 3000$, $K = 3$, $c_i = \frac{1}{3}$, μ a power law with exponent 3 and support $[0.05, 0.3]$, $\mathbf{M} = \Delta \mathbf{I}_3$, for $\Delta \in [10, 150]$. Here $\hat{\alpha}_{\text{opt}} = 0.28$.

$\alpha = 0$ is optimal on the Political blogs data set can be seen in Figure 4, where $\mathbf{L}_0$ is the similarity matrix for which the isolated eigenvalue is the farthest from the bulk of the other eigenvalues and thus the associated eigenvector is more aligned to the classes compared to the eigenvectors of $\mathbf{L}_{\frac{1}{2}}$ and $\mathbf{L}_1$.

Algo	Overlap	Modularity
$\hat{\alpha}_{\text{opt}}$ ($\simeq 0$)	**0.897**	**0.4246**
$\alpha = 0.5$	0.035	$\simeq 0$
$\alpha = 1$	0.040	$\simeq 0$
BH	0.304	0.2723

Table 1: Overlap performance and Modularity after applying the different spectral algorithms on the Political blogs graph (Adamic and Glance, 2005).

5. Concluding Remarks

The thorough study of $\mathbf{L}_\alpha$ performed in this article allows us to go further than the observation of (Gulikers et al., 2015) and (Jin et al., 2015) which state that it is important to use the eigenvectors of $\mathbf{L}_1$ or a normalization of the eigenvectors of $\mathbf{L}_0$ rather than the eigenvectors of $\mathbf{L}_0$ themselves for community detection when the network has an heterogeneous degree distribution (to avoid misclassifications induced by degree biases). Our main finding in particular is to show that there exists an optimal α, denoted here α_{opt}, for which taking the eigenvectors of $\mathbf{L}_{\alpha_{\text{opt}}}$ pre-multiplied by $\mathbf{D}^{\alpha_{\text{opt}}-1}$ ensures best performances (or to be more precise best asymptotic cluster detectability).

The results and methods in this article are all based on the strong assumption that the average node degree is of order $\mathcal{O}(n)$ and that the class-wise correction factors $C_{g_ig_j}$ differ by $\mathcal{O}(n^{-\frac{1}{2}})$ since $\forall i, j \in \{1, \ldots, n\}$, $C_{g_ig_j} = 1 + \frac{M_{g_ig_j}}{\sqrt{n}}$. Previous works (Lyzinski et al., 2014; Lei et al., 2015; Gulikers et al., 2015) suggest that the present analysis, which only considers "first order spectral statistics", should naturally extend to moderately sparse graphs (of as little as $\Omega(\log n)$ average degree). Under the sparse DCSBM graph assumption, strikingly different tools are required, opening up a challenging area of improved algorithm research. Similarly, if the $C_{g_ig_j}$'s differ at a rate $n^{-\frac{1}{2}} \ll r_n \ll 1$, mere refinements of our analysis ensure asymptotic weak consistency for all values of α based on the present tools. In passing, this shows that identifiability considerations are equivalent to those delineated for any α, as in (Gulikers et al., 2015) for $\alpha = 1$. Formally, the case where $r_n = \Omega(1)$ breaks Lemma 10 and therefore the validity of our present analysis but this scenario is also by far covered by previous works.

6. Proofs

Preliminaries

The random matrix under study $\mathbf{L}_\alpha$ is not a classically studied matrix in random matrix theory. We will thus first find in Section 6.1 an approximate tractable random matrix $\tilde{\mathbf{L}}_\alpha$ which asymptotically preserves the eigenvalue distribution and the extreme eigenvectors of $\mathbf{L}_\alpha$. In Section 6.2, we study the empirical distribution of the eigenvalues of $\mathbf{L}_\alpha$ and in Section 6.3, we characterize the exact localizations of those eigenvalues. Finally, a thorough study of the eigenvectors associated to the aforementioned eigenvalues is investigated in Sections 6.4 and 6.5.

We follow here the proof technique of (Couillet et al., 2016). In the sequel, we will make some approximations of random variables in the asymptotic regime where $n \to \infty$. For the sake of random variables comparisons, we give the following stochastic definitions. For $x \equiv x_n$ a random variable and $u_n \geq 0$, we write $x = \mathcal{O}(u_n)$ if for any $\eta > 0$ and $D > 0$, $n^D \mathbb{P}(x \geq n^\eta u_n) \to 0$ as $n \to \infty$. For $\mathbf{v}$ a vector or a diagonal matrix with random entries, $\mathbf{v} = \mathcal{O}(u_n)$ means that the maximal entry of $\mathbf{v}$ in absolute value is $\mathcal{O}(u_n)$ in the sense defined previously. When $\mathbf{M}$ is a square matrix, $\mathbf{M} = \mathcal{O}(u_n)$ means that the operator norm of $\mathbf{M}$ is $\mathcal{O}(u_n)$. For $\mathbf{x}$ a vector or a matrix with random entries, $\mathbf{x} = o(u_n)$ means that there is $\kappa > 0$ such that $\mathbf{x} = \mathcal{O}(n^{-\kappa} u_n)$.

Most of the proofs here are classical in random matrix theory (see e.g., (Baik and Silverstein, 2006)) but require certain controls inherent to our model. The goal of the article not being an exhaustive development of the proofs techniques, we will admit a number of technical results already studied in the literature. However, we will exhaustively develop the calculus to obtain our final results which are not trivial.

6.1 Random equivalent for $\mathbf{L}_\alpha$

The matrix $\mathbf{L}_\alpha = (\mathbf{d}^\mathsf{T}\mathbf{1}_n)^\alpha \frac{1}{\sqrt{n}} \mathbf{D}^{-\alpha} \left[\mathbf{A} - \frac{\mathbf{d}\mathbf{d}^\mathsf{T}}{\mathbf{d}^\mathsf{T}\mathbf{1}_n} \right] \mathbf{D}^{-\alpha}$ has non independent entries and is not a classical random matrix model. The idea is thus to approximate $\mathbf{L}_\alpha$ by a more tractable random matrix model $\tilde{\mathbf{L}}_\alpha$ in such a way that they share asymptotically the same set of outlying eigenvalues/eigenvectors which are of interest in our clustering scenario. We recall that the entries A_{ij} of the adjacency matrix is defined from the DCSBM model as independent Bernoulli random variables with parameter $q_i q_j \left(1 + \frac{M_{g_i g_j}}{\sqrt{n}}\right)$; one may thus write

$$A_{ij} = q_i q_j + q_i q_j \frac{M_{g_i g_j}}{\sqrt{n}} + X_{ij}$$

where X_{ij}, $1 \leq i, j \leq n$, are independent (up to the symmetry) zero mean random variables of variance $q_i q_j (1 - q_i q_j) + \mathcal{O}(n^{-\frac{1}{2}})$, since A_{ij} has mean $q_i q_j + q_i q_j \frac{M_{g_i g_j}}{\sqrt{n}}$ and variance $q_i q_j (1 - q_i q_j) + \mathcal{O}(n^{-\frac{1}{2}})$. We can then write the normalized adjacency matrix as follows

$$\frac{1}{\sqrt{n}} \mathbf{A} = \frac{1}{\sqrt{n}} \mathbf{q}\mathbf{q}^\mathsf{T} + \frac{1}{n} \left\{ \mathbf{q}_{(a)} \mathbf{q}_{(b)}^\mathsf{T} M_{ab} \right\}_{a,b=1}^K + \frac{1}{\sqrt{n}} \mathbf{X} \tag{7}$$

$$= \underbrace{\frac{\mathbf{q}\mathbf{q}^\mathsf{T}}{\sqrt{n}}}_{\mathbf{A}_{d,\sqrt{n}}} + \underbrace{\frac{1}{n} \mathbf{D}_q \mathbf{J}\mathbf{M}\mathbf{J}^\mathsf{T} \mathbf{D}_q}_{\mathbf{A}_{d,1}} + \underbrace{\frac{\mathbf{X}}{\sqrt{n}}}_{\mathbf{A}_{r,1}}, \tag{8}$$

where[13] $\mathbf{q}_{(i)} = [q_{n_1+\ldots+n_{i-1}+1}, \ldots, q_{n_1+\ldots+n_i}]^\mathsf{T} \in \mathbb{R}^{n_i}$ $(n_0 = 0)$, $\mathbf{X} = \{X_{ij}\}_{i,j=1}^n$ and $\mathbf{D}_q = \mathcal{D}(\mathbf{q})$. The idea of the proof is to write all the terms of $\mathbf{L}_\alpha$ based on Equation (8), since all those terms depend on $\mathbf{A}$. To this end, we will evaluate successively $\mathbf{d} = \mathbf{A}\mathbf{1}_n$, $\mathbf{D} = \mathcal{D}(\mathbf{d})$, $\mathbf{d}\mathbf{d}^\mathsf{T}$ and $2m = \mathbf{d}^\mathsf{T}\mathbf{1}_n$. It will appear that $\mathbf{D}$ and $\mathbf{d}^\mathsf{T}\mathbf{1}_n$ are composed of dominant terms (with higher operator norm) and vanishing terms (with smaller operator norm); we may then proceed to writing a Taylor expansion of $\mathbf{D}^{-\alpha}$ and $(2m)^\alpha = (\mathbf{d}^\mathsf{T}\mathbf{1}_n)^\alpha$ for any α around their dominant terms to finally retrieve a Taylor expansion of $\mathbf{L}_\alpha$.

13. We recall that subscript 'd, n^k' stands for deterministic term whose operator norm is of order n^k and 'r, n^k' for random term with operator norm of order n^k.

Let us start by developing the degree vector $\mathbf{d} = \mathbf{A}\mathbf{1}_n$. We have

$$\mathbf{d} = \mathbf{q}\mathbf{q}^\mathsf{T}\mathbf{1}_n + \frac{1}{\sqrt{n}}\mathbf{D}_q\mathbf{J}\mathbf{M}\mathbf{J}^\mathsf{T}\mathbf{D}_q\mathbf{1}_n + \mathbf{X}\mathbf{1}_n = \mathbf{q}^\mathsf{T}\mathbf{1}_n\Big(\underbrace{\mathbf{q}}_{\mathcal{O}(n^{\frac{1}{2}})} + \underbrace{\frac{1}{\sqrt{n}}\frac{\mathbf{D}_q\mathbf{J}\mathbf{M}\mathbf{J}^\mathsf{T}\mathbf{D}_q\mathbf{1}_n}{\mathbf{q}^\mathsf{T}\mathbf{1}_n}}_{\mathcal{O}(n^{-\frac{1}{2}})} + \underbrace{\frac{\mathbf{X}\mathbf{1}_n}{\mathbf{q}^\mathsf{T}\mathbf{1}_n}}_{\mathcal{O}(n^{-\frac{1}{2}})}\Big). \tag{9}$$

Let us then write the expansions of $\mathbf{d}^\mathsf{T}\mathbf{1}_n$, $(\mathbf{d}^\mathsf{T}\mathbf{1}_n)^\alpha$, $\mathbf{d}\mathbf{d}^\mathsf{T}$ and $\frac{\mathbf{d}\mathbf{d}^\mathsf{T}}{(\mathbf{d}^\mathsf{T}\mathbf{1}_n)}$ respectively. From (9), we obtain

$$\mathbf{d}^\mathsf{T}\mathbf{1}_n = (\mathbf{q}^\mathsf{T}\mathbf{1}_n)^2\Big[1 + \underbrace{\frac{1}{\sqrt{n}}\frac{\mathbf{1}_n^\mathsf{T}\mathbf{D}_q\mathbf{J}\mathbf{M}\mathbf{J}^\mathsf{T}\mathbf{D}_q\mathbf{1}_n}{(\mathbf{q}^\mathsf{T}\mathbf{1}_n)^2}}_{\mathcal{O}(n^{-\frac{1}{2}})} + \underbrace{\frac{\mathbf{1}_n^\mathsf{T}\mathbf{X}\mathbf{1}_n}{(\mathbf{q}^\mathsf{T}\mathbf{1}_n)^2}}_{\mathcal{O}(n^{-\frac{1}{2}})}\Big]. \tag{10}$$

Thus for any α, proceeding to a 1^{st} order Taylor expansion, we may write

$$(\mathbf{d}^\mathsf{T}\mathbf{1}_n)^\alpha = (\mathbf{q}^\mathsf{T}\mathbf{1}_n)^{2\alpha}\Big[1 + \frac{\alpha}{\sqrt{n}}\underbrace{\frac{\mathbf{1}_n^\mathsf{T}\mathbf{D}_q\mathbf{J}\mathbf{M}\mathbf{J}^\mathsf{T}\mathbf{D}_q\mathbf{1}_n}{(\mathbf{q}^\mathsf{T}\mathbf{1}_n)^2}}_{\mathcal{O}(n^{-\frac{1}{2}})} + \alpha\underbrace{\frac{\mathbf{1}_n^\mathsf{T}\mathbf{X}\mathbf{1}_n}{(\mathbf{q}^\mathsf{T}\mathbf{1}_n)^2}}_{\mathcal{O}(n^{-\frac{1}{2}})} + o(n^{-\frac{1}{2}})\Big]. \tag{11}$$

Besides, from (9) we have

$$\begin{aligned}\mathbf{d}\mathbf{d}^\mathsf{T} = (\mathbf{q}^\mathsf{T}\mathbf{1}_n)^2\Big[&\underbrace{\mathbf{q}\mathbf{q}^\mathsf{T}}_{\mathcal{O}(n)} + \underbrace{\frac{1}{\sqrt{n}}\frac{\mathbf{q}\mathbf{1}_n^\mathsf{T}\mathbf{D}_q\mathbf{J}\mathbf{M}\mathbf{J}^\mathsf{T}\mathbf{D}_q}{\mathbf{q}^\mathsf{T}\mathbf{1}_n}}_{\mathcal{O}(\sqrt{n})} + \underbrace{\frac{1}{\sqrt{n}}\frac{\mathbf{D}_q\mathbf{J}\mathbf{M}\mathbf{J}^\mathsf{T}\mathbf{D}_q\mathbf{1}_n\mathbf{q}^\mathsf{T}}{\mathbf{q}^\mathsf{T}\mathbf{1}_n}}_{\mathcal{O}(\sqrt{n})} + \underbrace{\frac{\mathbf{q}\mathbf{1}_n^\mathsf{T}\mathbf{X}}{\mathbf{q}^\mathsf{T}\mathbf{1}_n}}_{\mathcal{O}(\sqrt{n})} + \underbrace{\frac{\mathbf{X}\mathbf{1}_n\mathbf{q}^\mathsf{T}}{\mathbf{q}^\mathsf{T}\mathbf{1}_n}}_{\mathcal{O}(\sqrt{n})}\\
&+ \underbrace{\frac{1}{n}\frac{\mathbf{D}_q\mathbf{J}\mathbf{M}\mathbf{J}^\mathsf{T}\mathbf{D}_q\mathbf{1}_n\mathbf{1}_n^\mathsf{T}\mathbf{D}_q\mathbf{J}\mathbf{M}\mathbf{J}^\mathsf{T}\mathbf{D}_q}{(\mathbf{q}^\mathsf{T}\mathbf{1}_n)^2}}_{\mathcal{O}(1)} + \underbrace{\frac{1}{\sqrt{n}}\frac{\mathbf{D}_q\mathbf{J}\mathbf{M}\mathbf{J}^\mathsf{T}\mathbf{D}_q\mathbf{1}_n\mathbf{1}_n^\mathsf{T}\mathbf{X}}{(\mathbf{q}^\mathsf{T}\mathbf{1}_n)^2}}_{\mathcal{O}(1)} + \underbrace{\frac{1}{\sqrt{n}}\frac{\mathbf{X}\mathbf{1}_n\mathbf{1}_n^\mathsf{T}\mathbf{D}_q\mathbf{J}\mathbf{M}\mathbf{J}^\mathsf{T}\mathbf{D}_q}{(\mathbf{q}^\mathsf{T}\mathbf{1}_n)^2}}_{\mathcal{O}(1)}\\
&+ \underbrace{\frac{\mathbf{X}\mathbf{1}_n\mathbf{1}_n^\mathsf{T}\mathbf{X}}{(\mathbf{q}^\mathsf{T}\mathbf{1}_n)^2}}_{\mathcal{O}(1)} + o(1)\Big].\end{aligned} \tag{12}$$

Keeping in mind that we shall only need terms with non vanishing operator norms asymptotically, we will require $\frac{1}{\sqrt{n}}\left[\mathbf{A} - \frac{\mathbf{d}\mathbf{d}^\mathsf{T}}{\mathbf{d}^\mathsf{T}\mathbf{1}_n}\right]$ to have terms with spectral norms of order at least $\mathcal{O}(1)$. We get from multiplying (12) and (11) (with $\alpha = -1$)

$$\begin{aligned}\frac{1}{\sqrt{n}}\frac{\mathbf{d}\mathbf{d}^\mathsf{T}}{\mathbf{d}^\mathsf{T}\mathbf{1}_n} = \frac{\mathbf{q}\mathbf{q}^\mathsf{T}}{\sqrt{n}} &+ \frac{1}{n}\frac{\mathbf{q}\mathbf{1}_n^\mathsf{T}\mathbf{D}_q\mathbf{J}\mathbf{M}\mathbf{J}^\mathsf{T}\mathbf{D}_q}{\mathbf{q}^\mathsf{T}\mathbf{1}_n} + \frac{1}{n}\frac{\mathbf{D}_q\mathbf{J}\mathbf{M}\mathbf{J}^\mathsf{T}\mathbf{D}_q\mathbf{1}_n\mathbf{q}^\mathsf{T}}{\mathbf{q}^\mathsf{T}\mathbf{1}_n} + \frac{1}{\sqrt{n}}\frac{\mathbf{q}\mathbf{1}_n^\mathsf{T}\mathbf{X}}{\mathbf{q}^\mathsf{T}\mathbf{1}_n} + \frac{1}{\sqrt{n}}\frac{\mathbf{X}\mathbf{1}_n\mathbf{q}^\mathsf{T}}{\mathbf{q}^\mathsf{T}\mathbf{1}_n}\\
&- \frac{1}{n}\frac{\mathbf{1}_n^\mathsf{T}\mathbf{D}_q\mathbf{J}\mathbf{M}\mathbf{J}^\mathsf{T}\mathbf{D}_q\mathbf{1}_n}{(\mathbf{q}^\mathsf{T}\mathbf{1}_n)^2}\mathbf{q}\mathbf{q}^\mathsf{T} - \frac{1}{\sqrt{n}}\frac{\mathbf{1}_n^\mathsf{T}\mathbf{X}\mathbf{1}_n}{(\mathbf{q}^\mathsf{T}\mathbf{1}_n)^2}\mathbf{q}\mathbf{q}^\mathsf{T} + \mathcal{O}(n^{-\frac{1}{2}}).\end{aligned} \tag{13}$$

By subtracting (13) from (8), we obtain

$$\begin{aligned}\frac{1}{\sqrt{n}}\left(\mathbf{A} - \frac{\mathbf{d}\mathbf{d}^\mathsf{T}}{\mathbf{d}^\mathsf{T}\mathbf{1}_n}\right) = &\frac{1}{n}\mathbf{D}_q\mathbf{J}\mathbf{M}\mathbf{J}^\mathsf{T}\mathbf{D}_q - \frac{1}{n}\frac{\mathbf{q}\mathbf{1}_n^\mathsf{T}\mathbf{D}_q\mathbf{J}\mathbf{M}\mathbf{J}^\mathsf{T}\mathbf{D}_q}{\mathbf{q}^\mathsf{T}\mathbf{1}_n} - \frac{1}{n}\frac{\mathbf{D}_q\mathbf{J}\mathbf{M}\mathbf{J}^\mathsf{T}\mathbf{D}_q\mathbf{1}_n\mathbf{q}^\mathsf{T}}{\mathbf{q}^\mathsf{T}\mathbf{1}_n}\\
&+ \frac{1}{n}\frac{\mathbf{1}_n^\mathsf{T}\mathbf{D}_q\mathbf{J}\mathbf{M}\mathbf{J}^\mathsf{T}\mathbf{D}_q\mathbf{1}_n}{(\mathbf{q}^\mathsf{T}\mathbf{1}_n)^2}\mathbf{q}\mathbf{q}^\mathsf{T} + \frac{\mathbf{X}}{\sqrt{n}} - \frac{1}{\sqrt{n}}\frac{\mathbf{q}\mathbf{1}_n^\mathsf{T}\mathbf{X}}{\mathbf{q}^\mathsf{T}\mathbf{1}_n} - \frac{1}{\sqrt{n}}\frac{\mathbf{X}\mathbf{1}_n\mathbf{q}^\mathsf{T}}{\mathbf{q}^\mathsf{T}\mathbf{1}_n}\\
&+ \frac{1}{\sqrt{n}}\frac{\mathbf{1}_n^\mathsf{T}\mathbf{X}\mathbf{1}_n}{(\mathbf{q}^\mathsf{T}\mathbf{1}_n)^2}\mathbf{q}\mathbf{q}^\mathsf{T} + \mathcal{O}(n^{-\frac{1}{2}}).\end{aligned} \tag{14}$$

It then remains to evaluate $\mathbf{D}^{-\alpha}$. From (9), we may write $\mathbf{D} = \mathcal{D}(\mathbf{d})$ as

$$\mathbf{D} = \mathbf{q}^\mathsf{T}\mathbf{1}_n\Big(\underbrace{\mathbf{D}_q}_{\mathcal{O}(1)} + \underbrace{\mathcal{D}\left(\frac{1}{\sqrt{n}}\frac{\mathbf{D}_q\mathbf{J}\mathbf{M}\mathbf{J}^\mathsf{T}\mathbf{D}_q\mathbf{1}_n}{\mathbf{q}^\mathsf{T}\mathbf{1}_n}\right)}_{\mathcal{O}(n^{-\frac{1}{2}})} + \underbrace{\mathcal{D}\left(\frac{\mathbf{X}\mathbf{1}_n}{\mathbf{q}^\mathsf{T}\mathbf{1}_n}\right)}_{\mathcal{O}(n^{-\frac{1}{2}})}\Big).$$

The right hand side of $\mathbf{D}$ (in brackets) having a leading term in $\mathcal{O}(1)$ and residual terms in $\mathcal{O}(n^{-\frac{1}{2}})$, the Taylor expansion of the $(-\alpha)$-power of $\mathbf{D}$ is then retrieved

$$\mathbf{D}^{-\alpha} = \left(\mathbf{q}^\mathsf{T}\mathbf{1}_n\right)^{-\alpha}\Big(\underbrace{\mathbf{D}_q^{-\alpha}}_{\mathcal{O}(1)} - \alpha\mathbf{D}_q^{-(\alpha+1)}\underbrace{\mathcal{D}\left(\frac{1}{\sqrt{n}}\frac{\mathbf{D}_q\mathbf{J}\mathbf{M}\mathbf{J}^\mathsf{T}\mathbf{D}_q\mathbf{1}_n}{\mathbf{q}^\mathsf{T}\mathbf{1}_n}\right)}_{\mathcal{O}(n^{-\frac{1}{2}})} - \alpha\mathbf{D}_q^{-(\alpha+1)}\underbrace{\mathcal{D}\left(\frac{\mathbf{X}\mathbf{1}_n}{\mathbf{q}^\mathsf{T}\mathbf{1}_n}\right)}_{\mathcal{O}(n^{-\frac{1}{2}})} + \mathcal{O}(n^{-1})\Big). \tag{15}$$

By combining the expressions (11), (14) and (15), we obtain a Taylor approximation of $\mathbf{L}_\alpha$ as follows

$$\begin{aligned}\mathbf{L}_\alpha &= \mathbf{D}_q^{-\alpha}\frac{\mathbf{X}}{\sqrt{n}}\mathbf{D}_q^{-\alpha} + \frac{1}{n}\mathbf{D}_q^{1-\alpha}\mathbf{J}\mathbf{M}\mathbf{J}^\mathsf{T}\mathbf{D}_q^{1-\alpha} - \frac{1}{n}\frac{\mathbf{D}_q^{1-\alpha}\mathbf{1}_n\mathbf{1}_n^\mathsf{T}\mathbf{D}_q\mathbf{J}\mathbf{M}\mathbf{J}^\mathsf{T}\mathbf{D}_q^{1-\alpha}}{\mathbf{q}^\mathsf{T}\mathbf{1}_n}\\
&- \frac{1}{n}\frac{\mathbf{D}_q^{1-\alpha}\mathbf{J}\mathbf{M}\mathbf{J}^\mathsf{T}\mathbf{D}_q\mathbf{1}_n\mathbf{1}_n^\mathsf{T}\mathbf{D}_q^{1-\alpha}}{\mathbf{q}^\mathsf{T}\mathbf{1}_n} + \frac{1}{n}\frac{\mathbf{1}_n^\mathsf{T}\mathbf{D}_q\mathbf{J}\mathbf{M}\mathbf{J}^\mathsf{T}\mathbf{D}_q\mathbf{1}_n}{(\mathbf{q}^\mathsf{T}\mathbf{1}_n)^2}\mathbf{D}_q^{1-\alpha}\mathbf{1}_n\mathbf{1}_n^\mathsf{T}\mathbf{D}_q^{1-\alpha} - \frac{1}{\sqrt{n}}\frac{\mathbf{D}_q^{1-\alpha}\mathbf{1}_n\mathbf{1}_n^\mathsf{T}\mathbf{X}\mathbf{D}_q^{-\alpha}}{\mathbf{q}^\mathsf{T}\mathbf{1}_n}\\
&- \frac{1}{\sqrt{n}}\frac{\mathbf{D}_q^{-\alpha}\mathbf{X}\mathbf{1}_n\mathbf{1}_n^\mathsf{T}\mathbf{D}_q^{1-\alpha}}{\mathbf{q}^\mathsf{T}\mathbf{1}_n} + \frac{1}{\sqrt{n}}\frac{\mathbf{1}_n^\mathsf{T}\mathbf{X}\mathbf{1}_n}{(\mathbf{q}^\mathsf{T}\mathbf{1}_n)^2}\mathbf{D}_q^{1-\alpha}\mathbf{1}_n\mathbf{1}_n^\mathsf{T}\mathbf{D}_q^{1-\alpha} + \mathcal{O}(n^{-\frac{1}{2}}).\end{aligned}$$

The three following arguments allow to complete the proof

- $\mathbf{1}_n = \mathbf{J}\mathbf{1}_K$ and $\mathbf{D}_q\mathbf{1}_n = \mathbf{q}$.
- We may write $(\frac{1}{n}\mathbf{J}^\mathsf{T}\mathbf{q})_i = \frac{n_i}{n}\left(\frac{1}{n_i}\sum_{a\in\mathcal{C}_i} q_a\right)$. For classes of large sizes n_i, from the law of large numbers, $\left(\frac{1}{n_i}\sum_{a\in\mathcal{C}_i} q_a\right) \xrightarrow{\text{a.s.}} m_\mu$ and so, $\frac{1}{n}\mathbf{J}^\mathsf{T}\mathbf{q} \xrightarrow{\text{a.s.}} m_\mu\mathbf{c}$ where we recall that $m_\mu = \int t\mu(\mathrm{d}t)$.
- As $\mathbf{X}$ is a symmetric random matrix having independent entries of zero mean and finite variance, from the law of large numbers, we have $\frac{1}{n}\frac{\mathbf{1}_n^\mathsf{T}\mathbf{X}\mathbf{1}_n}{\sqrt{n}} \xrightarrow{\text{a.s.}} 0$.

Using those three arguments, $\mathbf{L}_\alpha$ may be further rewritten

$$\begin{aligned}\mathbf{L}_\alpha &= \mathbf{D}_q^{-\alpha}\frac{\mathbf{X}}{\sqrt{n}}\mathbf{D}_q^{-\alpha} + \frac{1}{n}\mathbf{D}_q^{1-\alpha}\mathbf{J}\mathbf{M}\mathbf{J}^\mathsf{T}\mathbf{D}_q^{1-\alpha} - \frac{1}{n}\mathbf{D}_q^{1-\alpha}\mathbf{J}\mathbf{1}_K\mathbf{c}^\mathsf{T}\mathbf{M}\mathbf{J}^\mathsf{T}\mathbf{D}_q^{1-\alpha}\\
&- \frac{1}{n}\mathbf{D}_q^{1-\alpha}\mathbf{J}\mathbf{M}\mathbf{c}\mathbf{1}_K^\mathsf{T}\mathbf{J}^\mathsf{T}\mathbf{D}_q^{1-\alpha} + \frac{1}{n}\mathbf{D}_q^{1-\alpha}\mathbf{J}\mathbf{1}_K\mathbf{c}^\mathsf{T}\mathbf{M}\mathbf{c}\mathbf{1}_K^\mathsf{T}\mathbf{J}^\mathsf{T}\mathbf{D}_q^{1-\alpha}\\
&- \frac{1}{\sqrt{n}\mathbf{q}^\mathsf{T}\mathbf{1}_n}\mathbf{D}_q^{1-\alpha}\mathbf{J}\mathbf{1}_K\mathbf{1}_n^\mathsf{T}\mathbf{X}\mathbf{D}_q^{-\alpha} - \frac{1}{\sqrt{n}\mathbf{q}^\mathsf{T}\mathbf{1}_n}\mathbf{D}_q^{-\alpha}\mathbf{X}\mathbf{1}_n\mathbf{1}_K^\mathsf{T}\mathbf{J}^\mathsf{T}\mathbf{D}_q^{1-\alpha} + \mathcal{O}(n^{-\frac{1}{2}}).\end{aligned} \tag{16}$$

By rearranging the terms of (16), we obtain the expected result

$$\begin{aligned}\mathbf{L}_\alpha &= \mathbf{D}_q^{-\alpha}\frac{\mathbf{X}}{\sqrt{n}}\mathbf{D}_q^{-\alpha}\\
&+ \begin{bmatrix}\frac{\mathbf{D}_q^{1-\alpha}\mathbf{J}}{\sqrt{n}} & \frac{\mathbf{D}_q^{-\alpha}\mathbf{X}\mathbf{1}_n}{\mathbf{q}^\mathsf{T}\mathbf{1}_n}\end{bmatrix}\begin{bmatrix}\left(\mathbf{I}_K - \mathbf{1}_K\mathbf{c}^T\right)\mathbf{M}\left(\mathbf{I}_K - \mathbf{c}\mathbf{1}_K^T\right) & -\mathbf{1}_K\\ -\mathbf{1}_K^T & 0\end{bmatrix}\begin{bmatrix}\frac{\mathbf{J}^\mathsf{T}\mathbf{D}_q^{1-\alpha}}{\sqrt{n}}\\ \frac{\mathbf{1}_n^\mathsf{T}\mathbf{X}\mathbf{D}_q^{-\alpha}}{\mathbf{q}^\mathsf{T}\mathbf{1}_n}\end{bmatrix} + \mathcal{O}(n^{-\frac{1}{2}}).\end{aligned}$$

This proves Theorem 1.

6.2 Limiting spectral distribution of $\mathbf{L}_\alpha$

It follows from Theorem 1 that $\tilde{\mathbf{L}}_\alpha = \mathbf{D}_q^{-\alpha}\frac{\mathbf{X}}{\sqrt{n}}\mathbf{D}_q^{-\alpha} + \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\mathsf{T}$ is equivalent to an additive spiked random matrix (Chapon et al., 2012) where

$$\mathbf{U} = \begin{bmatrix} \frac{\mathbf{D}_q^{1-\alpha}\mathbf{J}}{\sqrt{n}} & \frac{\mathbf{D}_q^{-\alpha}\mathbf{X}\mathbf{1}_n}{\mathbf{q}^\mathsf{T}\mathbf{1}_n} \end{bmatrix},$$
$$\mathbf{\Lambda} = \begin{bmatrix} \left(\mathbf{I}_K - \mathbf{1}_K\mathbf{c}^T\right)\mathbf{M}\left(\mathbf{I}_K - \mathbf{c}\mathbf{1}_K^T\right) & -\mathbf{1}_K \\ -\mathbf{1}_K^T & 0 \end{bmatrix},$$

with the difference that the deterministic part $\mathbf{U}\mathbf{\Lambda}\mathbf{U}^\mathsf{T}$ is not independent of the random part $\mathbf{D}_q^{-\alpha}\frac{\mathbf{X}}{\sqrt{n}}\mathbf{D}_q^{-\alpha}$ (an issue that we solve here) and $\mathbf{U}$ is not composed of orthonormal vectors. Let us then study $\bar{\mathbf{X}} = \mathbf{D}_q^{-\alpha}\frac{\mathbf{X}}{\sqrt{n}}\mathbf{D}_q^{-\alpha}$ (having entries $\bar{X}_{ij}$ with zero mean and variance σ_{ij}^2/n with $\sigma_{ij}^2 = q_i^{1-2\alpha}q_j^{1-2\alpha}(1 - q_iq_j) + \mathcal{O}(n^{-\frac{1}{2}})$) and show that its empirical spectral distribution (e.s.d.) $\tilde{\pi}^\alpha$ converges weakly to $\bar{\pi}^\alpha$ with Stieljes transform $e_{00}^\alpha(z) = \int (t-z)^{-1}\,\mathrm{d}\bar{\pi}^\alpha(t)$ for $z \in \mathbb{C}^+$. This will imply (By Weyl interlacing formula) that the empirical spectral measure $\pi^\alpha \equiv \frac{1}{n}\sum_{i=1}^n \delta_{\lambda_i(\tilde{\mathbf{L}}_\alpha)}$ (with $\lambda_i(\tilde{\mathbf{L}}_\alpha)$ eigenvalues of $\tilde{\mathbf{L}}_\alpha$) will also converge to $\bar{\pi}^\alpha$.

The matrix $\bar{\mathbf{X}}$ is a classical random matrix model in RMT already studied in similar cases (Pastur et al., 2011). It is well known for those random matrix models (having entries with given means, variances and bounded first order moments) that the law of the $\bar{X}_{ij}$'s does not change the results on the limiting law of the e.s.d. $\tilde{\pi}^\alpha$: this property is kwown as *universality* (e.g., (Silverstein and Bai, 1995)). For technical reasons, we can thus assume that the $\bar{X}_{ij}$'s are Gaussian random variables with the same means and variances in order to use standard Gaussian calculus, introduced in (Pastur et al., 2011). The objective of the proof is to find the deterministic limit $e_{00}^\alpha(z)$ for the random quantity $\frac{1}{n}\operatorname{tr}\left(\bar{\mathbf{X}} - z\mathbf{I}_n\right)^{-1}$ which is the Stieljes transform of the e.s.d. $\tilde{\pi}^\alpha$. Deterministic equivalents for the Stieljes transform of empirical spectral measures associated with centered and symmetric random matrix models with a variance profile have already been studied in for example (Ajanki et al., 2015; Hachem et al., 2007). We give in Appendix C an exhaustive development of the Gaussian calculus to obtain $e_{00}^\alpha(z)$. The final result is as follows.

Lemma 15 (A first deterministic equivalent) *Let* $\mathbf{Q} = (\bar{\mathbf{X}} - z\mathbf{I}_n)^{-1}$. *Then, for all* $z \in \mathbb{C}^+$,

$$\mathbf{Q} \leftrightarrow \bar{\mathbf{Q}} = \left(-z\mathbf{I}_n - \mathcal{D}\left(e_i(z)\right)_{i=1}^n\right)^{-1} \tag{17}$$

where $e_i(z)$ *the unique solution of* $e_i(z) = \frac{1}{n}\operatorname{tr}\mathcal{D}\left(\sigma_{ij}^2\right)_{j=1}^n\left(-z\mathbf{I}_n - \mathcal{D}\left(e_j(z)\right)_{j=1}^n\right)^{-1}$ *and the notation* $\mathbf{A} \leftrightarrow \mathbf{B}$ *stands for* $\frac{1}{n}\operatorname{tr}\mathbf{C}\mathbf{A} - \frac{1}{n}\operatorname{tr}\mathbf{C}\mathbf{B} \to 0$ *and* $\mathbf{d}_1^\mathsf{T}(\mathbf{A}-\mathbf{B})\mathbf{d}_2 \to 0$ *almost surely, for all deterministic Hermitian matrix* $\mathbf{C}$ *and deterministic vectors* $\mathbf{d}_i$ *of bounded norms (spectral norm for matrices and Euclidian norm for vectors).*

From Lemma 15, we get directly $\frac{1}{n}\operatorname{tr}\mathbf{Q} - e_{00}^\alpha(z) \xrightarrow{\text{a.s.}} 0$ with $e_{00}^\alpha(z) = \frac{1}{n}\sum_{i=1}^n \frac{1}{-z-e_i(z)}$. Observe now that

$$\begin{aligned} e_i(z) &= \frac{1}{n}\sum_{j=1}^n \frac{q_i^{1-2\alpha}q_j^{1-2\alpha} - q_i^{2-2\alpha}q_j^{2-2\alpha}}{-z-e_j(z)} \\ &= q_i^{1-2\alpha}e_{11}^\alpha(z) - q_i^{2-2\alpha}e_{21}^\alpha(z) \end{aligned} \tag{18}$$

where

$$e_{11}^{\alpha}(z) = \frac{1}{n}\sum_{j=1}^{n}\frac{q_j^{1-2\alpha}}{-z - q_j^{1-2\alpha}e_{11}^{\alpha}(z) + q_j^{2-2\alpha}e_{21}^{\alpha}(z)}$$
$$e_{21}^{\alpha}(z) = \frac{1}{n}\sum_{j=1}^{n}\frac{q_j^{2-2\alpha}}{-z - q_j^{1-2\alpha}e_{11}^{\alpha}(z) + q_j^{2-2\alpha}e_{21}^{\alpha}(z)} \tag{19}$$

from which we get

$$e_{00}^{\alpha}(z) = \int \frac{1}{-z - e_{11}^{\alpha}(z)q^{1-2\alpha} + e_{21}^{\alpha}(z)q^{2-2\alpha}}\mu(dq).$$

where for $z \in \mathbb{C}^+$ and $a, b \in \mathbb{Z}$ we define

$$e_{ab}^{\alpha}(z) = \int \frac{q^{a-2b\alpha}\mu(dq)}{-z - e_{11}^{\alpha}(z)q^{1-2\alpha} + e_{21}^{\alpha}(z)q^{2-2\alpha}}. \tag{20}$$

with $\mu(\mathrm{d}q) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\delta_{q_i}$. From this, we have that $e_{00}^{\alpha}(z)$ does not depend on n, so that $\frac{1}{n}\operatorname{tr}\mathbf{Q} \xrightarrow{\text{a.s.}} E_0^{\alpha}(z)$, $\tilde{\pi}^{\alpha} \to \bar{\pi}^{\alpha}$, and thus $\pi^{\alpha} \to \bar{\pi}^{\alpha}$ since $\tilde{\mathbf{L}}_{\alpha}$ and $\bar{\mathbf{X}}$ only differ by a finite rank matrix. This proves Theorem 3.

In the main core of the article, we have defined $e_{00}^{\alpha}(z) \triangleq m^{\alpha}(z)$, $e_{11}^{\alpha}(z) \triangleq f^{\alpha}(z)$ and $e_{21}^{\alpha}(z) \triangleq g^{\alpha}(z)$ for readability reasons. For future use, we define for $z, \tilde{z} \in \mathbb{C}\setminus\mathcal{S}^{\alpha}$

$$e_{ab;2}^{\alpha}(z,\tilde{z}) = \int \frac{q^{a-2b\alpha}\mu(dq)}{(-z - E_1^{\alpha}(z)q^{1-2\alpha} + E_2^{\alpha}(z)q^{2-2\alpha})(-\tilde{z} - E_1^{\alpha}(\tilde{z})q^{1-2\alpha} + E_2^{\alpha}(\tilde{z})q^{2-2\alpha})} \tag{21}$$

and

$$e_{ab;3}^{\alpha}(z,\tilde{z}) = \int \frac{q^{a-2b\alpha}\mu(dq)}{(-z - E_1^{\alpha}(z)q^{1-2\alpha} + E_2^{\alpha}(z)q^{2-2\alpha})^2(-\tilde{z} - E_1^{\alpha}(\tilde{z})q^{1-2\alpha} + E_2^{\alpha}(\tilde{z})q^{2-2\alpha})}. \tag{22}$$

Convergence of the e_i's

Similar results to Lemma 15 have been derived for example in (Hachem et al., 2007) and the fixed point algorithm (17) which consists of iterating the e_i's is shown to converge. Since the calculation of the e_{ab}'s is an intermediary step of (17) from (18), the fixed point algorithm (19) also converges. From the analyticity of the Stieljes transform outside its support, Lemma 15 extends naturally to $\mathbb{C}\setminus\mathcal{S}^{\alpha}$. This proves Theorem 3.

Remark 16 *Similarly to (Hachem et al., 2007), when none of the* $(\mathbf{D}_q^{-\alpha})_{ii}$*'s is isolated, the random matrix* $\bar{\mathbf{X}}$ *does not produce isolated eigenvalues outside the support* $\mathcal{S}^{\alpha}$ *of* $\bar{\pi}^{\alpha}$*. Here, for large* n*, this property is verified since from Assumption 1, the* q_i*'s are i.i.d. arising from a law with compact support (the probability that a* $(\mathbf{D}_q^{-\alpha})_{ii}$ *gets isolated tends to* 0 *asymptotically). This gives Proposition 17 which we will not prove here; similar proofs are provided for example in (Bai and Silverstein, 1998).*

Proposition 17 (No eigenvalues outside the support) *Following the statement of Theorem 3, let S_-^α and S_+^α be respectively the left and right edges of $\mathcal{S}^\alpha$. Then, for any $\epsilon > 0$, by letting $\mathcal{S}_\epsilon^\alpha = [S_-^\alpha - \epsilon; S_+^\alpha + \epsilon]$, for all large n almost surely,*

$$\left\{\lambda_i\left(\mathbf{D}_q^{-\alpha}\frac{\mathbf{X}}{\sqrt{n}}\mathbf{D}_q^{-\alpha}\right), 1 \le i \le n\right\} \cap (\mathbb{R} \setminus \mathcal{S}_\epsilon^\alpha) = \emptyset.$$

Remark 18 *The support $\mathcal{S}^\alpha$ is symmetric i.e., $\bar{\pi}^\alpha([a,b]) = \bar{\pi}^\alpha([-b,-a])$. We have in particular $S_-^\alpha = -S_+^\alpha = -S^\alpha$ where we denote $S_+^\alpha \triangleq \sup \mathcal{S}^\alpha$ and $S_-^\alpha \triangleq \inf \mathcal{S}^\alpha$.*

6.3 Isolated eigenvalues of $\mathbf{L}_\alpha$ and phase transition.

In the previous section, we have shown that the e.s.d. of $\mathbf{L}_\alpha$ converges weakly to the limiting law of the eigenvalues of $\bar{\mathbf{X}}$ since they only differ by a finite rank matrix. We shall have in addition isolated eigenvalues of $\mathbf{L}_\alpha$ induced by the aforementionned low rank matrix. We are interested here in the localization of eigenvalues of $\mathbf{L}_\alpha$ isolated from the support $\mathcal{S}^\alpha$ of the limiting law of its e.s.d. According to Proposition 17, there is almost surely no eigenvalue of $\bar{\mathbf{X}}$ at non-vanishing distance from $\mathcal{S}^\alpha$ asymptotically as $n \to \infty$ and hence the plausible isolated eigenvalues of $\mathbf{L}_\alpha$ are only due to the matrix $\mathbf{U}\mathbf{\Lambda}\mathbf{U}^\mathsf{T}$. We follow classical random matrix approaches used for the study of the spectrum of spiked random matrices (Benaych-Georges and Nadakuditi, 2012; Chapon et al., 2012). From Theorem 1, the eigenvalues ρ of $\mathbf{L}_\alpha$ falling at non-vanishing distance from the limiting support $\mathcal{S}^\alpha$ solve for large n, $0 = \det(\mathbf{L}_\alpha - \rho\mathbf{I}_n)$ almost surely for $\rho \notin \mathcal{S}^\alpha$. Since $\|\mathbf{L}_\alpha - \tilde{\mathbf{L}}_\alpha\| \xrightarrow{\text{a.s.}} 0$, $\rho_i(\mathbf{L}_\alpha) - \rho_i(\tilde{\mathbf{L}}_\alpha) \xrightarrow{\text{a.s.}} 0$ for all eigenvalues $\rho_i(\mathbf{L}_\alpha)$. We may then just solve $0 = \det(\mathbf{D}_q^{-\alpha}\frac{\mathbf{X}}{\sqrt{n}}\mathbf{D}_q^{-\alpha} + \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\mathsf{T} - \rho\mathbf{I}_n)$. Now, as from Proposition 17, the random matrix $\bar{\mathbf{X}}$ does not have eigenvalues at non-vanishing distance from $\mathcal{S}^\alpha$ asymptotically, for $\rho \notin \mathcal{S}^\alpha$, we can thus factor and cancel out $\det(\bar{\mathbf{X}} - \rho\mathbf{I}_n)$ from the previous determinant equation, so that we are left to solve

$$0 = \det(\mathbf{I}_n + \mathbf{Q}_\rho^\alpha\mathbf{U}\mathbf{\Lambda}\mathbf{U}^\mathsf{T}) = \det(\mathbf{I}_{K+1} + \mathbf{U}^\mathsf{T}\mathbf{Q}_\rho^\alpha\mathbf{U}\mathbf{\Lambda})$$

where $\mathbf{Q}_\rho^\alpha = (\bar{\mathbf{X}} - \rho\mathbf{I}_n)^{-1}$. As we will show next, the matrix $\mathbf{I}_{K+1} + \mathbf{U}^\mathsf{T}\mathbf{Q}_\rho^\alpha\mathbf{U}\mathbf{\Lambda}$ converges to a deterministic matrix, almost surely for large n. By the argument principle (similar to e.g., (Chapon et al., 2012)), the roots of $\mathbf{I}_{K+1} + \mathbf{U}^\mathsf{T}\mathbf{Q}_\rho^\alpha\mathbf{U}\mathbf{\Lambda}$ are asymptotically those of the limiting matrix, with same multiplicity and it suffices to study the latter.

We then proceed to retrieving a limit for $\mathbf{I}_{K+1} + \mathbf{U}^\mathsf{T}\mathbf{Q}_\rho^\alpha\mathbf{U}\mathbf{\Lambda}$. From Theorem 1, we have

$$\mathbf{U}^\mathsf{T}\mathbf{Q}_\rho^\alpha\mathbf{U} = \begin{pmatrix} \frac{1}{n}\mathbf{J}^\mathsf{T}\mathbf{D}_q^{1-\alpha}\mathbf{Q}_\rho^\alpha\mathbf{D}_q^{1-\alpha}\mathbf{J} & \frac{1}{\sqrt{n}(\mathbf{q}^\mathsf{T}\mathbf{1}_n)}\mathbf{J}^\mathsf{T}\mathbf{D}_q^{1-\alpha}\mathbf{Q}_\rho^\alpha\mathbf{D}_q^{-\alpha}\mathbf{X}\mathbf{1}_n \\ \frac{1}{\sqrt{n}(\mathbf{q}^\mathsf{T}\mathbf{1}_n)}\mathbf{1}_n^\mathsf{T}\mathbf{X}\mathbf{D}_q^{-\alpha}\mathbf{Q}_\rho^\alpha\mathbf{D}_q^{1-\alpha}\mathbf{J} & \frac{1}{(\mathbf{q}^\mathsf{T}\mathbf{1}_n)^2}\mathbf{1}_n^\mathsf{T}\mathbf{X}\mathbf{D}_q^{-\alpha}\mathbf{Q}_\rho^\alpha\mathbf{D}_q^{-\alpha}\mathbf{X}\mathbf{1}_n \end{pmatrix}.$$

The entries $(1,2)$, $(2,1)$ and $(2,2)$ of $\mathbf{U}^\mathsf{T}\mathbf{Q}_\rho^\alpha\mathbf{U}$ are random as they contain the random matrix $\mathbf{X}$ but tend to be deterministic in the limit. In fact, using the resolvent identity, we have that $\mathbf{Q}_\rho^\alpha\mathbf{D}_q^{-\alpha}\frac{\mathbf{X}}{\sqrt{n}}\mathbf{D}_q^{-\alpha} = \mathbf{I}_n + \rho\mathbf{Q}_\rho^\alpha$, the entry $(1,2)$ becomes $\frac{1}{(\mathbf{q}^\mathsf{T}\mathbf{1}_n)}\mathbf{J}^\mathsf{T}\mathbf{D}_q\mathbf{1}_n + \rho\frac{1}{\sqrt{n}(\mathbf{q}^\mathsf{T}\mathbf{1}_n)}\mathbf{J}^\mathsf{T}\mathbf{D}_q^{1-\alpha}\mathbf{Q}_\rho^\alpha\mathbf{D}_q^\alpha\mathbf{1}_n$ and the entry $(2,2)$ is equal to $\frac{n}{(\mathbf{q}^\mathsf{T}\mathbf{1}_n)^2}\Big(\mathbf{1}_n^\mathsf{T}\mathbf{X}\mathbf{1}_n + \rho\mathbf{1}_n^\mathsf{T}\mathbf{D}_q^{2\alpha}\mathbf{1}_n + \rho^2\mathbf{1}_n^\mathsf{T}\mathbf{D}_q^\alpha\mathbf{Q}_\rho^\alpha\mathbf{D}_q^\alpha\mathbf{1}_n\Big)$. Now, we can freely use Lemma 15 to evaluate the limits of the entries of $\mathbf{U}^\mathsf{T}\mathbf{Q}_\rho^\alpha\mathbf{U}$ since all the terms are of the form $\mathbf{a}^\mathsf{T}\mathbf{Q}_\rho^\alpha\mathbf{b}$ with $\mathbf{a}$ and $\mathbf{b}$ deterministic vectors. From Lemma 15, the entries $(1,1)$, $(1,2)$ and $(2,2)$ converge almost surely respectively to $\frac{1}{n}\mathbf{J}^\mathsf{T}\mathbf{D}_q^{1-\alpha}\bar{\mathbf{Q}}_\rho^\alpha\mathbf{D}_q^{1-\alpha}\mathbf{J}$, $\frac{1}{(\mathbf{q}^\mathsf{T}\mathbf{1}_n)}\mathbf{J}^\mathsf{T}\mathbf{D}_q\mathbf{1}_n + \rho\frac{1}{(\mathbf{q}^\mathsf{T}\mathbf{1}_n)}\mathbf{J}^\mathsf{T}\mathbf{D}_q^{1-\alpha}\bar{\mathbf{Q}}_\rho^\alpha\mathbf{D}_q^\alpha\mathbf{1}_n$ and $\frac{n}{(\mathbf{q}^\mathsf{T}\mathbf{1}_n)^2}\Big(\mathbf{1}_n^\mathsf{T}\mathbf{X}\mathbf{1}_n + \rho\mathbf{1}_n^\mathsf{T}\mathbf{D}_q^{2\alpha}\mathbf{1}_n + \rho^2\mathbf{1}_n^\mathsf{T}\mathbf{D}_q^\alpha\bar{\mathbf{Q}}_\rho^\alpha\mathbf{D}_q^\alpha\mathbf{1}_n\Big)$ for large n.

Now, using the fact that for any bounded continuous function f, from the law of large numbers,

$$\frac{1}{n}\sum_{j\in\mathcal{C}_i} f(q_j) = \frac{n_i}{n}\frac{1}{n_i}\sum_{j\in\mathcal{C}_i} f(q_j) \xrightarrow{\text{a.s.}} c_i \int f(q)\mu(dq). \tag{23}$$

After some algebra, we obtain $\frac{1}{n}\mathbf{J}^\mathsf{T}\mathbf{D}_q^{1-\alpha}\bar{\mathbf{Q}}_\rho^\alpha\mathbf{D}_q^{1-\alpha}\mathbf{J} \xrightarrow{\text{a.s.}} e_{21}^\alpha(\rho)\mathcal{D}(\mathbf{c})$ where the e_{ij}'s are given in Theorem 3. Similarly for the terms $(1,2)$ and $(2,2)$, we obtain respectively

$$\frac{1}{(\mathbf{q}^\mathsf{T}\mathbf{1}_n)}\mathbf{J}^\mathsf{T}\mathbf{D}_q\mathbf{1}_n + \rho\frac{1}{(\mathbf{q}^\mathsf{T}\mathbf{1}_n)}\mathbf{J}^\mathsf{T}\mathbf{D}_q^{1-\alpha}\bar{\mathbf{Q}}_\rho^\alpha\mathbf{D}_q^\alpha\mathbf{1}_n \xrightarrow{\text{a.s.}} \left(1+\frac{\rho}{m_\mu}e_{10}^\alpha(\rho)\right)\mathbf{c}$$

and

$$\frac{n}{(\mathbf{q}^\mathsf{T}\mathbf{1}_n)^2}\left(\mathbf{1}_n^\mathsf{T}\mathbf{X}\mathbf{1}_n + \rho\mathbf{1}_n^\mathsf{T}\mathbf{D}_q^{2\alpha}\mathbf{1}_n + \rho^2\mathbf{1}_n^\mathsf{T}\mathbf{D}_q^\alpha\bar{\mathbf{Q}}_\rho^\alpha\mathbf{D}_q^\alpha\mathbf{1}_n\right) \xrightarrow{\text{a.s.}} \frac{1}{m_\mu^2}\left(\rho v_\mu + \rho^2 e_{0;-1}^\alpha(\rho)\right)$$

with $v_\mu = \int q^{2\alpha}\mu(\mathrm{d}q)$ and where we have also used the fact that $\frac{1}{n}\mathbf{1}_n^\mathsf{T}\frac{\mathbf{X}}{\sqrt{n}}\mathbf{1}_n \xrightarrow{\text{a.s.}} 0$ again from the law of large numbers.

The limit of $\mathbf{I}_{K+1} + \mathbf{U}^\mathsf{T}\mathbf{Q}_\rho^\alpha\mathbf{U}\mathbf{\Lambda}$ is then obtained as

$$\mathbf{I}_{K+1} + \mathbf{U}^\mathsf{T}\mathbf{Q}_\rho^\alpha\mathbf{U}\mathbf{\Lambda} \xrightarrow{\text{a.s.}}$$
$$\begin{pmatrix} \mathbf{I}_K + e_{21}^\alpha(\rho)(\mathcal{D}(\mathbf{c}) - \mathbf{c}\mathbf{c}^\mathsf{T})\mathbf{M}(\mathbf{I}_K - \mathbf{c}\mathbf{1}_K^\mathsf{T}) - \left(1+\frac{\rho}{m_\mu}e_{10}^\alpha(\rho)\right)\mathbf{c}\mathbf{1}_K^\mathsf{T} & -e_{21}^\alpha(\rho)\mathbf{c} \\ \frac{\rho}{m_\mu^2}\left(v_\mu + \rho e_{0;-1}^\alpha(\rho)\right)\mathbf{1}_K^\mathsf{T} & -\rho\frac{e_{10}^\alpha(\rho)}{m_\mu} \end{pmatrix}.$$

Using the Schur complement formula for the determinant of block matrices, we have that the determinant of the RHS matrix is zero whenever

$$-\rho\frac{e_{10}^\alpha(\rho)}{m_\mu}\det\Big[\mathbf{I}_K + e_{21}^\alpha(\rho)(\mathcal{D}(\mathbf{c}) - \mathbf{c}\mathbf{c}^\mathsf{T})\mathbf{M}(\mathbf{I}_K - \mathbf{c}\mathbf{1}_K^\mathsf{T}) - \left(1+\frac{\rho}{m_\mu}e_{10}^\alpha(\rho)\right)\mathbf{c}\mathbf{1}_K^\mathsf{T} + \frac{\left(v_\mu + \rho e_{0;-1}^\alpha(\rho)\right)e_{21}^\alpha(\rho)}{m_\mu e_{10}^\alpha(\rho)}\mathbf{c}\mathbf{1}_K^\mathsf{T}\Big] = 0$$

or equivalently $\det(\underline{\mathbf{G}}_\rho^\alpha) = 0$ where

$$\underline{\mathbf{G}}_\rho^\alpha = \mathbf{I}_K + e_{21}^\alpha(\rho)(\mathcal{D}(\mathbf{c}) - \mathbf{c}\mathbf{c}^\mathsf{T})\mathbf{M}(\mathbf{I}_K - \mathbf{c}\mathbf{1}_K^\mathsf{T}) + \theta^\alpha(\rho)\mathbf{c}\mathbf{1}_K^\mathsf{T}$$
$$\theta^\alpha(\rho) = -1 - \frac{\rho}{m_\mu}e_{10}^\alpha(\rho) + \frac{\left(v_\mu + \rho e_{0;-1}^\alpha(\rho)\right)e_{21}^\alpha(\rho)}{m_\mu e_{10}^\alpha(\rho)}.$$

The isolated eigenvalues ρ of $\mathbf{L}_\alpha$, which are the ρ for which $\det(\mathbf{I}_{K+1} + \mathbf{U}^\mathsf{T}\mathbf{Q}_\rho^\alpha\mathbf{U}\mathbf{\Lambda}) = 0$, are then asymptotically the ρ such that $\det(\underline{\mathbf{G}}_\rho^\alpha) = 0$.

Remark 19 (Two types of isolated eigenvalues) *From the previous paragraph,* $1+\theta^\alpha(\rho)$ *is an eigenvalue of* $\underline{\mathbf{G}}_\rho^\alpha$ *with associated left eigenvector* $\mathbf{1}_K$ *and right eigenvector* $\mathbf{c}$ *since* $\mathbf{1}_K^T\underline{\mathbf{G}}_\rho^\alpha = (1+\theta^\alpha(\rho))\mathbf{1}_K^T$ *and* $\underline{\mathbf{G}}_\rho^\alpha\mathbf{c} = (1+\theta^\alpha(\rho))\mathbf{c}$.

Letting ρ *be such that* $\det(\underline{\mathbf{G}}_\rho^\alpha) = 0$, *we can discriminate two cases*

- $1+\theta^\alpha(\rho) = 0$: *isolated eigenvalues are found for those* $\rho \in \mathbb{R}\setminus\mathcal{S}^\alpha$ *such that* $1+\theta^\alpha(\rho) = 0$. *We shall denote by* $\tilde{\rho}$ *such eigenvalues when they exist.*

- $1+\theta^\alpha(\rho)\neq 0$: *the left and right eigenvectors associated to the zero eigenvalues of* $\underline{\mathbf{G}}_\rho^\alpha$ *are respectively orthogonal to the right and left eigenvectors associated to the non-zero eigenvalues. So, by letting* $\mathbf{V}_l$, $\mathbf{V}_r$ *be matrices containing in columns the respectively left and right eigenvectors of* $\underline{\mathbf{G}}_\rho^\alpha$ *associated with the zero eigenvalues, we have* $\mathbf{V}_l^\mathsf{T}\mathbf{c}=\mathbf{0}$ *and* $\mathbf{1}_K^\mathsf{T}\mathbf{V}_r=\mathbf{0}$ *since* $1+\theta^\alpha(\rho)\neq 0$*. It is thus immediate that* $(\mathbf{V}_l,\mathbf{V}_r)$ *is also a pair of eigenvectors (with multiplicity) of* $\mathbf{I}_K+e_{21}^\alpha(\rho)\left(\mathcal{D}(\mathbf{c})-\mathbf{c}\mathbf{c}^\mathsf{T}\right)\mathbf{M}\left(\mathbf{I}_K-\mathbf{c}\mathbf{1}_K^T\right)$ *associated to the zero eigenvalues.*

As we show in Section 6.5, for $1+\theta^\alpha(\tilde{\rho})=0$, the eigenvectors associated to the aforementioned isolated eigenvalues $\tilde{\rho}$ will not contain information about the classes. This case is thus of no interest for clustering. It is nevertheless important from a practical viewpoint to note that, even in the absence of communities, spurious isolated eigenvalues may be found that may deceive the experimenter in suggesting the presence of node clusters. From now on, we will only consider the isolated eigenvalues ρ for which $1+\theta^\alpha(\rho)\neq 0$.

We now have all the ingredients to determine the conditions under which we may have eigenvalues of $\mathbf{L}_\alpha$ which isolate from $\mathcal{S}^\alpha$. Let l be a non zero eigenvalue of $\mathbf{G}_\rho^\alpha=(\mathcal{D}(\mathbf{c})-\mathbf{c}\mathbf{c}^\mathsf{T})\mathbf{M}(\mathbf{I}_K-\mathbf{c}\mathbf{1}_K^T)$. Since $\det((\mathcal{D}(\mathbf{c})-\mathbf{c}\mathbf{c}^\mathsf{T})\mathbf{M}(\mathbf{I}_K-\mathbf{c}\mathbf{1}_K^\mathsf{T}))=\det((\mathbf{I}_K-\mathbf{c}\mathbf{1}_K^\mathsf{T})(\mathcal{D}(\mathbf{c})-\mathbf{c}\mathbf{c}^\mathsf{T})\mathbf{M})=\det((\mathcal{D}(\mathbf{c})-\mathbf{c}\mathbf{c}^\mathsf{T})\mathbf{M})$, l is also a non zero eigenvalue of $\bar{\mathbf{M}}=(\mathcal{D}(\mathbf{c})-\mathbf{c}\mathbf{c}^\mathsf{T})\mathbf{M}$. For each isolated eigenvalue ρ of $\mathbf{L}_\alpha$ we have a one-to-one mapping with a non zero eigenvalue l of $\bar{\mathbf{M}}$ such that $l=-\frac{1}{E_2^\alpha(\rho)}$. Hence, to show the existence of isolated eigenvalues of $\mathbf{L}_\alpha$, we need to solve for $\rho\in\mathbb{R}\setminus\mathcal{S}^\alpha$, $l=-\frac{1}{E_2^\alpha(\rho)}$ for each non zero eigenvalue l of $\bar{\mathbf{M}}$. Precisely, let us write $\mathcal{S}^\alpha=\bigcup_{m=1}^M[S_{m,-}^\alpha,S_{m,+}^\alpha]$ with $S_{1,-}^\alpha\leq S_{1,+}^\alpha<S_{2,-}^\alpha\leq\ldots<S_{M,+}^\alpha$ and define $S_{0,+}=-\infty$ and $S_{M+1,-}=+\infty$. Then, recalling that the Stieltjes transform of a real supported measure is necessarily increasing on $\mathbb{R}$, there exist isolated eigenvalues of $\mathbf{L}^\alpha$ in $(S_{m,+}^\alpha,S_{m+1,-}^\alpha)$, $m\in\{0,\ldots,M\}$, for all large n almost surely, if and only if there exists eigenvalues ℓ of $\bar{\mathbf{M}}$ such that

$$\lim_{x\downarrow S_{m,+}^\alpha}E_2^\alpha(x)<-\ell^{-1}<\lim_{x\uparrow S_{m+1,-}^\alpha}E_2^\alpha(x). \tag{24}$$

In particular, when $\mathcal{S}^\alpha=[S_-^\alpha,S_+^\alpha]$ is composed of a single connected component (as when $\mathcal{S}^\alpha$ is the support of the semi-circle law as well as most cases met in practice), then isolated eigenvalues of $\mathbf{L}^\alpha$ may only be found beyond S_+^α if $\ell>\lim_{x\downarrow S_+^\alpha}-\frac{1}{E_2^\alpha(x)}$ ($l>0$) or below S_-^α if $\ell<\lim_{x\uparrow S_-^\alpha}-\frac{1}{E_2^\alpha(x)}$ ($l<0$), for some non-zero eigenvalue ℓ of $\bar{\mathbf{M}}$. From the asymptotic spectrum of $\mathbf{L}_\alpha$, $S_-^\alpha=-S_+^\alpha$ as one can show that for any $z\in\mathbb{R}\setminus\mathcal{S}^\alpha$, $E_2^\alpha(-z)=-E_2^\alpha(z)$ so that both previous conditions reduce to $\mid\ell\mid>\lim_{x\downarrow S_+^\alpha}-\frac{1}{E_2^\alpha(x)}$. This proves Theorem 5.

The next section is advocated to the study of the eigenvectors associated to isolated eigenvalues of $\mathbf{L}_\alpha$.

6.4 Informative eigenvectors

In this section, in order to fully characterize the performances of Algorithm 2, we study in depth the normalized eigenvectors $\bar{\mathbf{v}}_i^\alpha$ used for the classification in the algorithm (step 3 of Algorithm 2). We consider here the eigenvectors corresponding to the eigenvalues for which $1+\theta^\alpha(\rho)\neq 0$ (when $1+\theta^\alpha(\rho)=0$, the corresponding eigenvectors do not contain any structural information about the classes; this case is treated in Section 6.5). For technical reasons, we restrict ourselves here to those eigenpairs $(\lambda_i, \bar{\mathbf{v}}_i^\alpha)$'s for which there exists no $\lambda_j \neq \lambda_i$ such that, if $\lambda_i \to \rho$, $\lambda_j \to \rho$.

We recall that we may write $\bar{\mathbf{v}}_i^\alpha$ [14] as the "noisy plateaus" vector

$$\bar{\mathbf{v}}_i^\alpha = \sum_{a=1}^{K} \nu_i^a \frac{\mathbf{j}_a}{\sqrt{n_a}} + \sqrt{\sigma_{ii}^a}\mathbf{w}_i^a \tag{25}$$

where $\mathbf{w}_i^a \in \mathbb{R}^n$ is a random vector orthogonal to $\mathbf{j}_a$, of norm $\sqrt{n_a}$ and supported on the indices of $\mathcal{C}_a$ and

$$\nu_i^a = \frac{1}{\sqrt{n_a}} (\bar{\mathbf{v}}_i^\alpha)^\mathsf{T} \mathbf{j}_a = \frac{1}{\sqrt{n_a}} \frac{(\mathbf{u}_i^\alpha)^\mathsf{T}\mathbf{D}^{\alpha-1}\mathbf{j}_a}{\sqrt{(\mathbf{u}_i^\alpha)^\mathsf{T}\mathbf{D}^{2(\alpha-1)}\mathbf{u}_i^\alpha}} \tag{26}$$

$$\sigma_{ij}^a = \frac{(\mathbf{u}_i^\alpha)^\mathsf{T}\mathbf{D}^{\alpha-1}\mathcal{D}_a\mathbf{D}^{\alpha-1}\mathbf{u}_j^\alpha}{\sqrt{(\mathbf{u}_i^\alpha)^\mathsf{T}\mathbf{D}^{2(\alpha-1)}\mathbf{u}_i}\sqrt{(\mathbf{u}_j^\alpha)^\mathsf{T}\mathbf{D}^{2(\alpha-1)}\mathbf{u}_j^\alpha}} - \nu_i^a\nu_j^a \tag{27}$$

with $\mathcal{D}_a = \mathcal{D}(\mathbf{j}_a)$.

- We estimate the ν_i^a's by obtaining an estimator of the $K \times K$ matrix

$$\frac{1}{n}\frac{\mathbf{J}^\mathsf{T}\mathbf{D}^{\alpha-1}\mathbf{u}_i^\alpha(\mathbf{u}_i^\alpha)^\mathsf{T}\mathbf{D}^{\alpha-1}\mathbf{J}}{(\mathbf{u}_i^\alpha)^\mathsf{T}\mathbf{D}^{2(\alpha-1)}\mathbf{u}_i^\alpha},$$

  the diagonal entries of which allow to estimate $|\nu_i^a|$ while the off-diagonal entries are used to decide on the signs of the ν_i^a's (up to a convention in the sign of $\mathbf{u}_i^\alpha$).

- Similarly, we first estimate the more involved object

$$\frac{1}{n}\frac{\mathbf{J}^\mathsf{T}\mathbf{D}^{\alpha-1}\mathbf{u}_i^\alpha(\mathbf{u}_i^\alpha)^\mathsf{T}\mathbf{D}^{\alpha-1}\mathcal{D}_a\mathbf{D}^{\alpha-1}\mathbf{u}_j^\alpha(\mathbf{u}_j^\alpha)^\mathsf{T}\mathbf{D}^{\alpha-1}\mathbf{J}}{\left((\mathbf{u}_i^\alpha)^\mathsf{T}\mathbf{D}^{2(\alpha-1)}\mathbf{u}_i^\alpha\right)\left((\mathbf{u}_j^\alpha)^\mathsf{T}\mathbf{D}^{2(\alpha-1)}\mathbf{u}_j^\alpha\right)}$$

  from which $\frac{(\mathbf{u}_i^\alpha)^\mathsf{T}\mathbf{D}^{\alpha-1}\mathcal{D}_a\mathbf{D}^{\alpha-1}\mathbf{u}_j^\alpha}{\sqrt{(\mathbf{u}_i^\alpha)^\mathsf{T}\mathbf{D}^{2(\alpha-1)}\mathbf{u}_i^\alpha}\sqrt{(\mathbf{u}_j^\alpha)^\mathsf{T}\mathbf{D}^{2(\alpha-1)}\mathbf{u}_j^\alpha}}$ is retrieved by dividing any entry e, f of the former quantity by non-vanishing quantities $\nu_i^e\nu_i^f$. For the eigenvectors $\mathbf{u}_i^\alpha$ used for clustering, there is always at least one index f such that ν_i^f is non zero (otherwise, this eigenvector is of no use for clustering).

6.4.1 Evaluation of the class means ν_i^a's

The estimation of the ν_i^a's requires the evaluation of $\frac{1}{n}\frac{\mathbf{J}^\mathsf{T}\mathbf{D}^{\alpha-1}\mathbf{u}_i^\alpha(\mathbf{u}_i^\alpha)^\mathsf{T}\mathbf{D}^{\alpha-1}\mathbf{J}}{(\mathbf{u}_i^\alpha)^\mathsf{T}\mathbf{D}^{2(\alpha-1)}\mathbf{u}_i^\alpha}$ for $\mathbf{u}_i^\alpha$ eigenvector associated to a limiting isolated eigenvalue ρ with unit multiplicity of $\mathbf{L}_\alpha$. By residue calculus, we have that

$$\frac{1}{n}\mathbf{J}^\mathsf{T}\mathbf{D}^{\alpha-1}\mathbf{u}_i^\alpha(\mathbf{u}_i^\alpha)^\mathsf{T}\mathbf{D}^{\alpha-1}\mathbf{J} = -\frac{1}{2\pi i}\oint_{\Gamma_\rho}\frac{1}{n}\mathbf{J}^\mathsf{T}\mathbf{D}^{\alpha-1}\left(\mathbf{L}_\alpha - z\mathbf{I}_n\right)^{-1}\mathbf{D}^{\alpha-1}\mathbf{J}\mathrm{d}z \tag{28}$$

14. Recall that the graph nodes were assumed labeled by class, and thus the entries of $\bar{\mathbf{v}}_i^\alpha$ are similarly sorted by class.

for large n almost surely, where Γ_ρ is a complex (positively oriented) contour circling around the limiting eigenvalue ρ only. As from Theorem 1, $\mathbf{L}_\alpha = \mathbf{D}_q^{-\alpha}\frac{\mathbf{X}}{\sqrt{n}}\mathbf{D}_q^{-\alpha} + \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\mathsf{T} + o(1)$, we apply the Woodburry identity to the inverse in the previous integrand and we get

$$\begin{aligned}\frac{1}{n}\mathbf{J}^\mathsf{T}\mathbf{D}^{\alpha-1}\left(\mathbf{L}_\alpha - z\mathbf{I}_n\right)^{-1}\mathbf{D}^{\alpha-1}\mathbf{J} &= \frac{1}{n}\mathbf{J}^\mathsf{T}\mathbf{D}^{\alpha-1}\mathbf{Q}_z^\alpha\mathbf{D}^{\alpha-1}\mathbf{J}\\ &+ \frac{1}{n}\mathbf{J}^\mathsf{T}\mathbf{D}^{\alpha-1}\mathbf{Q}_z^\alpha\mathbf{U}\mathbf{\Lambda}\left(\mathbf{I}_{K+1} + \mathbf{U}^\mathsf{T}\mathbf{Q}_z^\alpha\mathbf{U}\mathbf{\Lambda}\right)^{-1}\mathbf{U}^\mathsf{T}\mathbf{Q}_z^\alpha\mathbf{D}^{\alpha-1}\mathbf{J} + o(1).\end{aligned}$$

The first right-hand side has asymptotically no residue when we integrate over the contour Γ_ρ (as per Proposition 17 there is no eigenvalues of $\bar{\mathbf{X}}$ in Γ_ρ for all large n almost surely). We are then left with the second right-most term. Using the block structure used in Section 6.3, we may write

$$\begin{aligned}&\left(\mathbf{I}_{K+1} + \mathbf{U}^\mathsf{T}\mathbf{Q}_z^\alpha\mathbf{U}\mathbf{\Lambda}\right)^{-1} \xrightarrow{\text{a.s.}}\\ &\begin{pmatrix}\mathbf{I}_K + e_{21}^\alpha(z)(\mathcal{D}(\mathbf{c}) - \mathbf{c}\mathbf{c}^\mathsf{T})\mathbf{M}(\mathbf{I}_K - \mathbf{c}\mathbf{1}_K^\mathsf{T}) - \left(1 + \frac{z}{m_\mu}e_{10}^\alpha(z)\right)\mathbf{c}\mathbf{1}_K^\mathsf{T} & -e_{21}^\alpha(z)\mathbf{c}\\ \frac{z}{m_\mu^2}\left(v_\mu + ze_{0;-1}^\alpha(z)\right)\mathbf{1}_K^\mathsf{T} & -z\frac{e_{10}^\alpha(z)}{m_\mu}\end{pmatrix}^{-1}.\end{aligned}$$

Let us write $\gamma(z) = \frac{z}{m_\mu^2}\left(v_\mu + ze_{0;-1}^\alpha(z)\right)$. We can now use a block inversion formula to write

$$\left(\mathbf{I}_{K+1} + \mathbf{U}^T\mathbf{Q}_z^\alpha\mathbf{U}\mathbf{\Lambda}\right)^{-1} \xrightarrow{\text{a.s.}} \begin{pmatrix}(\underline{\mathbf{G}}_z^\alpha)^{-1} & -\frac{e_{10}^\alpha(z)\left[\underline{\mathbf{G}}_z^\alpha - \frac{\gamma(z)m_\mu e_{21}^\alpha(z)}{ze_{10}^\alpha(z)}\mathbf{c}\mathbf{1}_K^\mathsf{T}\right]^{-1}\mathbf{c}}{-\frac{ze_{21}^\alpha(z)}{m_\mu} + \gamma(z)e_{10}^\alpha(z)\mathbf{1}_K^\mathsf{T}\left[\underline{\mathbf{G}}_z^\alpha - \frac{\gamma(z)m_\mu e_{21}^\alpha(z)}{ze_{10}^\alpha(z)}\mathbf{c}\mathbf{1}_K^\mathsf{T}\right]^{-1}\mathbf{c}}\\ \frac{\gamma(z)m_\mu}{ze_{21}^\alpha(z)}\mathbf{1}_K^\mathsf{T}(\underline{\mathbf{G}}_z^\alpha)^{-1} & \frac{1}{-\frac{ze_{21}^\alpha(z)}{m_\mu} + \gamma(z)e_{10}^\alpha(z)\mathbf{1}_K^\mathsf{T}\left[\underline{\mathbf{G}}_z^\alpha - \frac{\gamma(z)m_\mu e_{21}^\alpha(z)}{ze_{10}^\alpha(z)}\mathbf{c}\mathbf{1}_K^\mathsf{T}\right]^{-1}\mathbf{c}}\end{pmatrix} \tag{29}$$

with $\underline{\mathbf{G}}_z^\alpha = \mathbf{I}_K + e_{21}^\alpha(z)\left(\mathcal{D}(\mathbf{c}) - \mathbf{c}\mathbf{c}^\mathsf{T}\right)\mathbf{M}\left(\mathbf{I}_K - \mathbf{c}\mathbf{1}_K^T\right) + \theta^\alpha(z)\mathbf{c}\mathbf{1}_K^T$. The entries of the previous matrix seem to be cumbersome but as we will see, the residue calculus will greatly simplify. In fact, we have that $\mathbf{1}_K^T\underline{\mathbf{G}}_z^\alpha = (1 + \theta^\alpha(z))\mathbf{1}_K^T$ so that $\mathbf{1}_K^T(\underline{\mathbf{G}}_z^\alpha)^{-1} = \frac{1}{1+\theta^\alpha(z)}\mathbf{1}_K^T$ which is well defined since we are considering the case $1 + \theta^\alpha(z) \neq 0$. Similarly, we have that

$$\left[\underline{\mathbf{G}}_z^\alpha - \frac{\gamma(z)m_\mu e_{10}^\alpha(z)}{ze_{21}^\alpha(z)}\mathbf{c}\mathbf{1}_K^\mathsf{T}\right]\mathbf{c} = \left(-z\frac{e_{10}^\alpha(z)}{m_\mu}\right)\mathbf{c}$$

meaning that $\left[\underline{\mathbf{G}}_z^\alpha - \frac{\gamma(z)m_\mu e_{21}^\alpha(z)}{ze_{10}^\alpha(z)}\mathbf{c}\mathbf{1}_K^\mathsf{T}\right]^{-1}\mathbf{c} = -\frac{m_\mu}{ze_{10}^\alpha(z)}\mathbf{c}$. So finally, the terms $(1,2)$, $(2,1)$ and $(2,2)$ of $\left(\mathbf{I}_{K+1} + \mathbf{U}^T\mathbf{Q}_z^\alpha\mathbf{U}\mathbf{\Lambda}\right)^{-1}$ do no longer depend on $(\underline{\mathbf{G}}_z^\alpha)^{-1}$ and thus do not have poles in the contour Γ_ρ. We can then write

$$\left(\mathbf{I}_{K+1} + \mathbf{U}^T\mathbf{Q}_z^\alpha\mathbf{U}\mathbf{\Lambda}\right)^{-1} = \begin{pmatrix}(\underline{\mathbf{G}}_z^\alpha)^{-1} & 0\\ 0 & 0\end{pmatrix} + \mathbf{R}_1(z)$$

with $\mathbf{R}_1(z)$ having no residue in the contour Γ_ρ. Thus, to perform the contour integration of the integrand in (28) around Γ_ρ, we just need to evaluate the top-left entries of $\mathbf{J}^\mathsf{T}\mathbf{D}^{\alpha-1}\mathbf{Q}_z^\alpha\mathbf{U}\mathbf{\Lambda}$ and $\mathbf{U}^\mathsf{T}\mathbf{Q}_z^\alpha\mathbf{D}^{\alpha-1}\mathbf{J}$. Those are easily retrieved from the calculus in Section 6.3.

We have in particular $(\frac{1}{\sqrt{n}}\mathbf{J}^\mathsf{T}\mathbf{D}^{\alpha-1}\mathbf{Q}_z^\alpha\mathbf{U}\boldsymbol{\Lambda})_{11} \xrightarrow{\text{a.s.}} e_{00}^\alpha(z)(\mathcal{D}(\mathbf{c})-\mathbf{c}\mathbf{c}^\mathsf{T})\mathbf{M}(\mathbf{I}_K-\mathbf{c}\mathbf{1}_K^\mathsf{T}) - \beta^\alpha(z)\mathbf{c}\mathbf{1}_K^\mathsf{T}$ where $\beta^\alpha(z)=\frac{1}{m_\mu}\left[\int t^{2\alpha-1}\mu(\mathrm{d}t)+e_{-1;-1}^\alpha(z)\right]$ and similarly $(\mathbf{U}^\mathsf{T}\mathbf{Q}_z^\alpha\mathbf{D}^{\alpha-1}\mathbf{J})_{11} \xrightarrow{\text{a.s.}} e_{00}^\alpha(z)\mathcal{D}(\mathbf{c})$, so that finally

$$\frac{1}{n}\mathbf{J}^\mathsf{T}\mathbf{D}^{\alpha-1}\mathbf{u}_i^\alpha(\mathbf{u}_i^\alpha)^\mathsf{T}\mathbf{D}^{\alpha-1}\mathbf{J} \xrightarrow{\text{a.s.}}$$
$$-\frac{1}{2\pi i}\oint_{\Gamma_\rho}\left[\left(e_{00}^\alpha(z)(\mathcal{D}(\mathbf{c})-\mathbf{c}\mathbf{c}^\mathsf{T})\mathbf{M}(\mathbf{I}_K-\mathbf{c}\mathbf{1}_K^\mathsf{T})-\beta^\alpha(z)\mathbf{c}\mathbf{1}_K^\mathsf{T}\right)(\underline{\mathbf{G}}_z^\alpha)^{-1}\times e_{00}^\alpha(z)\mathcal{D}(\mathbf{c})+\mathbf{R}_2(z)\mathrm{d}z\right]$$

where $\mathbf{R}_2(z)$ is a matrix having no residue in the considered contour. Now, we are ready to compute the integral. From the Cauchy integral formula,

$$\frac{1}{n}\mathbf{J}^\mathsf{T}\mathbf{D}^{\alpha-1}\mathbf{u}_i^\alpha(\mathbf{u}_i^\alpha)^\mathsf{T}\mathbf{D}^{\alpha-1}\mathbf{J} \xrightarrow{\text{a.s.}}$$
$$\lim_{z\to\rho}(z-\rho)\left[e_{00}^\alpha(z)(\mathcal{D}(\mathbf{c})-\mathbf{c}\mathbf{c}^\mathsf{T})\mathbf{M}(\mathbf{I}_K-\mathbf{c}\mathbf{1}_K^\mathsf{T})-\beta^\alpha(z)\mathbf{c}\mathbf{1}_K^\mathsf{T}\right](\underline{\mathbf{G}}_z^\alpha)^{-1}\times e_{00}^\alpha(z)\mathcal{D}(\mathbf{c}).$$

By writing $\underline{\mathbf{G}}_z^\alpha=\rho_z\mathbf{v}_{r,z}\mathbf{v}_{l,z}^T+\tilde{\mathbf{V}}_{r,z}\tilde{\boldsymbol{\Sigma}}_z\tilde{\mathbf{v}}_{l,z}^T$ where $\mathbf{v}_{r,z}$ and $\mathbf{v}_{l,z}$ are respectively right and left eigenvectors associated with the vanishing eigenvalue ρ_z of $\underline{\mathbf{G}}_z^\alpha$ when $z\to\rho$; $\tilde{\mathbf{V}}_{r,z}\in\mathbb{R}^{n\times\eta_\rho}$ and $\tilde{\mathbf{V}}_{l,z}\mathbb{R}^{n\times\eta_\rho}$ are respectively sets of right and left eigenspaces associated with non vanishing eigenvalues, we then have

$$\lim_{z\to\rho}(z-\rho)(\underline{\mathbf{G}}_z^\alpha)^{-1}\overset{(1)}{=}\lim_{z\to\rho}(z-\rho)\frac{\mathbf{v}_{r,z}\mathbf{v}_{l,z}^T}{\rho_z'}$$

where we have used the l'Hopital rule and the fact that the non vanishing eigenvalue part of $\underline{\mathbf{G}}_z^\alpha$ will produce zero in the limit $z\to\rho$. Using $\rho_z=\mathbf{v}_{l,z}^\mathsf{T}\underline{\mathbf{G}}_z^\alpha\mathbf{v}_{r,z}$, we obtain

$$\frac{1}{n}\mathbf{J}^\mathsf{T}\mathbf{D}^{\alpha-1}\mathbf{u}_i^\alpha(\mathbf{u}_i^\alpha)^\mathsf{T}\mathbf{D}^{\alpha-1}\mathbf{J} \xrightarrow{\text{a.s.}}$$
$$\left[e_{00}^\alpha(\rho)(\mathcal{D}(\mathbf{c})-\mathbf{c}\mathbf{c}^\mathsf{T})\mathbf{M}(\mathbf{I}_K-\mathbf{c}\mathbf{1}_K^\mathsf{T})-\beta^\alpha(\rho)\mathbf{c}\mathbf{1}_K^\mathsf{T}\right]\frac{\mathbf{v}_{r,\rho}\mathbf{v}_{l,\rho}^\mathsf{T}}{\left(\mathbf{v}_{l,z}^\mathsf{T}\underline{\mathbf{G}}_z^\alpha\mathbf{v}_{r,z}\right)'_{z=\rho}}\times e_{00}^\alpha(\rho)\mathcal{D}(\mathbf{c}).$$

Since $(\mathbf{v}_{l,\rho})^\mathsf{T}\underline{\mathbf{G}}_\rho^\alpha=\underline{\mathbf{G}}_\rho^\alpha\mathbf{v}_{r,\rho}=0$,

$$\begin{aligned}\left((\mathbf{v}_{l,z})^T\underline{\mathbf{G}}_z^\alpha\mathbf{v}_{r,z}\right)'_{z=\rho}&=((\mathbf{v}_{l,z})^T)'_{z=\rho}\underline{\mathbf{G}}_\rho^\alpha\mathbf{v}_{r,\rho}+(\mathbf{v}_{l,\rho})^T\left(\underline{\mathbf{G}}_z^\alpha\right)'_{z=\rho}\mathbf{v}_{r,\rho}+(\mathbf{v}_{l,\rho})^T\underline{\mathbf{G}}_\rho^\alpha(\mathbf{v}_{r,z})'_{z=\rho}\\&=(\mathbf{v}_{l,\rho})^T\left(\underline{\mathbf{G}}_z^\alpha\right)'_{z=\rho}\mathbf{v}_{r,\rho}\\&=\left(e_{21}^\alpha(\rho)\right)'(\mathbf{v}_{l,\rho})^T\left(\mathcal{D}(\mathbf{c})-\mathbf{c}\mathbf{c}^\mathsf{T}\right)\mathbf{M}\left(\mathbf{I}_K-\mathbf{c}\mathbf{1}_K^T\right)\mathbf{v}_{r,\rho}\end{aligned}$$

where the subscript $'$ denotes the first derivative with respect to z. Using the fact that $\mathbf{v}_{r,\rho}$ is orthogonal to $\mathbf{1}_K^T$, and $(\mathbf{v}_{r,\rho},\mathbf{v}_{l,\rho})$ is also a pair of eigenvectors of $\left(\mathcal{D}(\mathbf{c})-\mathbf{c}\mathbf{c}^\mathsf{T}\right)\mathbf{M}\left(\mathbf{I}_K-\mathbf{c}\mathbf{1}_K^T\right)$ associated with eigenvalue $-\frac{1}{e_{21}^\alpha(\rho)}$, we get

$$\frac{1}{n}\mathbf{J}^\mathsf{T}\mathbf{D}^{\alpha-1}\mathbf{u}_i^\alpha(\mathbf{u}_i\alpha)^\mathsf{T}\mathbf{D}^{\alpha-1}\mathbf{J}\xrightarrow{\text{a.s.}}\frac{(e_{00}^\alpha(\rho))^2}{e_{21}^\alpha(\rho)'}\frac{\mathbf{v}_{r,\rho}(\mathbf{v}_{l,\rho})^T}{\mathbf{v}_{l,\rho}^T\mathbf{v}_{r,\rho}}\mathcal{D}(\mathbf{c}). \tag{30}$$

By introducing $\mathbf{v}_\rho = \mathcal{D}(\mathbf{c})^{\frac{1}{2}}\mathbf{v}_{l,\rho} = \mathcal{D}(\mathbf{c})^{-\frac{1}{2}}\mathbf{v}_{r,\rho}$ eigenvector of the symmetric matrix $\mathcal{D}(\mathbf{c})^{\frac{1}{2}}\left(\mathbf{I}_K - \mathbf{1}_K\mathbf{c}^\mathsf{T}\right)\mathbf{M}\left(\mathbf{I}_K - \mathbf{c}\mathbf{1}_K^T\right)\mathcal{D}(\mathbf{c})^{\frac{1}{2}}$, we obtain the final result

$$\frac{1}{n}\mathbf{J}^\mathsf{T}\mathbf{D}^{\alpha-1}\mathbf{u}_i^\alpha(\mathbf{u}_i^\alpha)^\mathsf{T}\mathbf{D}^{\alpha-1}\mathbf{J} \xrightarrow{\text{a.s.}} \frac{(e_{00}^\alpha(\rho))^2}{e_{21}^\alpha(\rho)'}\mathcal{D}\left(\mathbf{c}\right)^{1/2}\mathbf{v}_\rho(\mathbf{v}_\rho)^T\mathcal{D}\left(\mathbf{c}\right)^{1/2}. \tag{31}$$

Next, we need to estimate the denominator term $(\mathbf{u}_i^\alpha)^\mathsf{T}\mathbf{D}^{2(\alpha-1)}\mathbf{u}_i^\alpha$ of $\frac{1}{n}\frac{\mathbf{J}^\mathsf{T}\mathbf{D}^{\alpha-1}\mathbf{u}_i^\alpha(\mathbf{u}_i^\alpha)^\mathsf{T}\mathbf{D}^{\alpha-1}\mathbf{J}}{(\mathbf{u}_i^\alpha)^\mathsf{T}\mathbf{D}^{2(\alpha-1)}\mathbf{u}_i^\alpha}$ of ν_i^a. For $\mathbf{u}_i^\alpha$ an eigenvector of $\mathbf{L}_\alpha$ associated to an isolated eigenvalue converging to ρ asymptotically, we have

$$\begin{aligned}(\mathbf{u}_i^\alpha)^\mathsf{T}\mathbf{D}^{2(\alpha-1)}\mathbf{u}_i^\alpha &= \operatorname{tr}(\mathbf{u}_i^\alpha(\mathbf{u}_i^\alpha)^\mathsf{T}\mathbf{D}^{2(\alpha-1)}) \\ &= \operatorname{tr}\left(-\frac{1}{2\pi i}\oint_{\Gamma_\rho}\left(\mathbf{L}_\alpha - z\mathbf{I}_n\right)\mathbf{D}^{2(\alpha-1)}\mathrm{d}z\right).\end{aligned}$$

As in the previous section, by applying Woodburry idendity, this is equivalent to evaluating

$$\operatorname{tr}\left(-\frac{1}{2\pi i}\oint_{\Gamma_\rho}\left[\mathbf{U}^\mathsf{T}\mathbf{Q}_z^\alpha\mathbf{D}^{2(\alpha-1)}\mathbf{Q}_z^\alpha\mathbf{U}\mathbf{\Lambda}\begin{pmatrix}(\mathbf{G}_z^\alpha)^{-1} & 0\\ 0 & 0\end{pmatrix} + \mathbf{R}_3(z)\right]\mathrm{d}z\right),$$

where $\mathbf{R}_3(z)$ is a matrix having no residue in the considered contour.

Again here, we just need the top left entry of $\mathbf{U}^\mathsf{T}\mathbf{Q}_z^\alpha\mathbf{D}^{2(\alpha-1)}\mathbf{Q}_z^\alpha\mathbf{U}\mathbf{\Lambda}$ which is given from Theorem 1 by

$$(\mathbf{U}^\mathsf{T}\mathbf{Q}_z^\alpha\mathbf{D}^{2(\alpha-1)}\mathbf{Q}_z^\alpha\mathbf{U}\mathbf{\Lambda})_{11} = \underbrace{\frac{1}{n}\mathbf{J}^\mathsf{T}\mathbf{D}^{1-\alpha}\mathbf{Q}_z^\alpha\mathbf{D}^{2(\alpha-1)}\mathbf{Q}_z^\alpha\mathbf{D}^{1-\alpha}\mathbf{J}(\mathbf{I}_K - \mathbf{1}_K\mathbf{c}^\mathsf{T})\mathbf{M}(\mathbf{I}_K - \mathbf{c}\mathbf{1}_K^\mathsf{T})}_{(\mathbf{I})} \tag{32}$$

$$-\underbrace{\frac{1}{\sqrt{n}(\mathbf{q}^\mathsf{T}\mathbf{1}_n)}\mathbf{D}^{1-\alpha}\mathbf{Q}_z^\alpha\mathbf{D}^{2(\alpha-1)}\mathbf{Q}_z^\alpha\mathbf{D}^{-\alpha}\mathbf{X}\mathbf{1}_n\mathbf{1}_K^\mathsf{T}}_{(\mathbf{II})}. \tag{33}$$

We can get rid of the term (**II**) since after residue calculus, we will get (similar to Equation 30) $\mathbf{1}_K^\mathsf{T}\mathbf{v}_{r,\rho} = 0$ which cancels out the whole term. Let us now concentrate on the term (**I**). At this point, we need to introduce the following result which, for any deterministic vectors of bounded Euclidean norm $\mathbf{a}$, $\mathbf{b}$ and any deterministic diagonal matrix $\mathbf{\Xi}$, approximates the random quantity $\mathbf{a}^\mathsf{T}\mathbf{Q}_{z_1}^\alpha\mathbf{\Xi}\mathbf{Q}_{z_2}^\alpha\mathbf{b}$ by a deterministic equivalent.

Lemma 20 (Second deterministic equivalents) *For all $z \in \mathbb{C}\backslash\mathcal{S}^\alpha$, we have the following deterministic equivalent*

$$\mathbf{Q}_{z_1}^\alpha\mathbf{\Xi}\mathbf{Q}_{z_2}^\alpha \leftrightarrow \bar{\mathbf{Q}}_{z_1}^\alpha\mathbf{\Xi}\bar{\mathbf{Q}}_{z_2}^\alpha + \bar{\mathbf{Q}}_{z_1}^\alpha\mathcal{D}\left[\left(\mathbf{I}_n - \mathbf{\Upsilon}_{z_1,z_2}\right)^{-1}\mathbf{\Upsilon}_{z_1,z_2}\operatorname{diag}\left(\mathbf{\Xi}\right)\right]\bar{\mathbf{Q}}_{z_2}^\alpha$$

where $\mathbf{\Xi}$ is any diagonal matrix, $\bar{\mathbf{Q}}_z^\alpha$ is given in Lemma 15 and

$$\Upsilon_{z_1,z_2}(i,j) = \frac{1}{n}\frac{q_i^{1-2\alpha}q_j^{1-2\alpha}\left(1 - q_iq_j\right)}{\left(-z_1 - e_{11}^\alpha(z_1)q_i^{1-2\alpha} + e_{21}^\alpha(z_1)q_i^{2-2\alpha}\right)\left(-z_2 - e_{11}^\alpha(z_2)q_j^{1-2\alpha} + e_{21}^\alpha(z_2)q_j^{2-2\alpha}\right)}.$$

The equivalence relation $\leftrightarrow$ is as defined in Lemma 15.

Thanks to Lemma 20 (proof provided in Appendix D), a deterministic approximation of the term (**I**) in Equation (32) can be obtained. We get in particular

$$\frac{1}{n}\mathbf{J}^{\mathsf{T}}\mathbf{D}^{1-\alpha}\mathbf{Q}_z^{\alpha}\mathbf{D}^{2(\alpha-1)}\mathbf{Q}_z^{\alpha}\mathbf{D}^{1-\alpha}\mathbf{J} = \frac{1}{n}\mathbf{J}^{\mathsf{T}}\mathbf{D}^{1-\alpha}\bar{\mathbf{Q}}_z^{\alpha}\mathbf{D}^{2(\alpha-1)}\bar{\mathbf{Q}}_z^{\alpha}\mathbf{D}^{1-\alpha}\mathbf{J} \tag{34}$$

$$+\frac{1}{n}\mathbf{J}^{\mathsf{T}}\mathbf{D}^{1-\alpha}\bar{\mathbf{Q}}_z^{\alpha}\mathcal{D}\left[(\mathbf{I}_n-\boldsymbol{\Upsilon}_{z,z})^{-1}\boldsymbol{\Upsilon}_{z,z}\mathbf{d}^{\alpha}\right]\bar{\mathbf{Q}}_z^{\alpha}\mathbf{D}^{1-\alpha}\mathbf{J} \tag{35}$$

where $\mathbf{d}^{\alpha} = \{q_i^{2(\alpha-1)}\}_{i=1}^n$ and $\boldsymbol{\Upsilon}_{z_1,z_2}$ was defined in Lemma 20. Using similar argument as in Equation (23), we can easily show that the first right hand side term of (34) converges almost surely to $e_{00;2}^{\alpha}\mathcal{D}(\mathbf{c})$. It then remains to estimate the second right-most term of (34). $\boldsymbol{\Upsilon}_{z_1,z_2}$ (as defined in Lemma 20) may be written as the sum of two rank-one matrices

$$\boldsymbol{\Upsilon}_{z_1,z_2} = \frac{1}{n}\left(\mathbf{a}_{z_1}\mathbf{a}_{z_2}^{\mathsf{T}} - \mathbf{b}_{z_1}\mathbf{b}_{z_2}^{\mathsf{T}}\right)$$

where $\mathbf{a}_z = \left\{\frac{q_j^{1-2\alpha}}{-z-q_j^{1-2\alpha}e_{11}^{\alpha}(z)+q_j^{2-2\alpha}e_{21}^{\alpha}(z)}\right\}_{j=1}^n$ and $\mathbf{b}_z = \left\{\frac{q_j^{2-2\alpha}}{-z-q_j^{1-2\alpha}e_{11}^{\alpha}(z)+q_j^{2-2\alpha}e_{21}^{\alpha}(z)}\right\}_{j=1}^n$.

The matrix $\boldsymbol{\Upsilon}_{z,z}$ can thus be further written $\boldsymbol{\Upsilon}_{z,z} = \frac{1}{n}\begin{pmatrix}\mathbf{a}_z & \mathbf{b}_z\end{pmatrix}\mathbf{I}_2\begin{pmatrix}\mathbf{a}_z^{\mathsf{T}}/n \\ -\mathbf{b}_z^{\mathsf{T}}/n\end{pmatrix}$. Using matrix inversion lemmas, we have

$$(\mathbf{I}_n-\boldsymbol{\Upsilon}_{z,z})^{-1}\boldsymbol{\Upsilon}_{z,z}\mathbf{d}^{\alpha} = \begin{pmatrix}\mathbf{a}_z & \mathbf{b}_z\end{pmatrix}\begin{pmatrix}1-\frac{\mathbf{a}_z^{\mathsf{T}}\mathbf{a}_z}{n} & -\frac{\mathbf{a}_z^{\mathsf{T}}\mathbf{b}_z}{n} \\ \frac{\mathbf{b}_z^{\mathsf{T}}\mathbf{a}_z}{n} & 1+\frac{\mathbf{b}_z^{\mathsf{T}}\mathbf{b}_z}{n}\end{pmatrix}^{-1}\begin{pmatrix}\frac{\mathbf{a}_z^{\mathsf{T}}\mathbf{d}^{\alpha}}{n} \\ -\frac{\mathbf{b}_z^{\mathsf{T}}\mathbf{d}^{\alpha}}{n}\end{pmatrix}.$$

Using again the argument in Equation (23) , we can easily show that $\frac{\mathbf{a}_z^{\mathsf{T}}\mathbf{a}_z}{n}$, $\frac{\mathbf{a}_z^{\mathsf{T}}\mathbf{b}_z}{n}$, $\frac{\mathbf{b}_z^{\mathsf{T}}\mathbf{b}_z}{n}$, $\frac{\mathbf{a}_z^{\mathsf{T}}\mathbf{d}^{\alpha}}{n}$ and $\frac{\mathbf{b}_z^{\mathsf{T}}\mathbf{d}^{\alpha}}{n}$ converge for large n almost surely respectively to $e_{22;2}^{\alpha}(z)$, $e_{32;2}^{\alpha}(z)$, $e_{42;2}^{\alpha}(z)$, $e_{-1;0}^{\alpha}(z)$ and $e_{00}^{\alpha}(z)$ with $e_{ij;2}^{\alpha}$ defined in Equation (21). This given, we can show that

$$\frac{1}{n}\mathbf{J}^{\mathsf{T}}\mathbf{D}^{1-\alpha}\bar{\mathbf{Q}}_z^{\alpha}\mathcal{D}\left[(\mathbf{I}_n-\boldsymbol{\Upsilon}_{z,z})^{-1}\boldsymbol{\Upsilon}_{z,z}\mathbf{d}^{\alpha}\right]\bar{\mathbf{Q}}_z^{\alpha}\mathbf{D}^{1-\alpha}\mathbf{J} \xrightarrow{\text{a.s.}} \chi^{\alpha}(z)\mathcal{D}(\mathbf{c})$$

with

$$\chi^{\alpha}(z) = \frac{\left[\left(1+e_{42;2}^{\alpha}(z)\right)e_{-1,0}^{\alpha}(z)-e_{32;2}^{\alpha}(z)e_{00}^{\alpha}(z)\right]e_{32;3}^{\alpha}(z)-\left[e_{22;2}^{\alpha}(z)e_{-1,0}^{\alpha}(z)+\left(1-e_{22;2}^{\alpha}(z)\right)e_{00}^{\alpha}(z)\right]e_{42;3}^{\alpha}(z)}{\left(1+e_{42;2}^{\alpha}(z)\right)\left(1-e_{22;2}^{\alpha}(z)\right)+\left[e_{32;2}^{\alpha}(z)\right]^2}.$$

We thus have

$$(\mathbf{u}_i^{\alpha})^{\mathsf{T}}\mathbf{D}^{2(\alpha-1)}\mathbf{u}_i^{\alpha} \xrightarrow{\text{a.s.}} \operatorname{tr}\left(\lim_{z\to\rho}\left(e_{00;2}^{\alpha}(z)+\chi^{\alpha}(z)\right)(\mathcal{D}(\mathbf{c})-\mathbf{c}\mathbf{c}^{\mathsf{T}})\mathbf{M}(\mathbf{I}_K-\mathbf{c}\mathbf{1}_K^{\mathsf{T}})(\underline{\mathbf{G}}_z^{\alpha})^{-1}\right).$$

By applying l'Hopital rule to evaluate this limit as in the previous section, we obtain

$$(\mathbf{u}_i^{\alpha})^{\mathsf{T}}\mathbf{D}^{2(\alpha-1)}\mathbf{u}_i^{\alpha} \xrightarrow{\text{a.s.}} \frac{e_{00;2}(\rho)+\chi^{\alpha}(\rho)}{\left(e_{21}^{\alpha}(\rho)\right)'}.$$

Finally,

$$\frac{1}{n}\frac{\mathbf{J}^{\mathsf{T}}\mathbf{D}^{\alpha-1}\mathbf{u}_i^{\alpha}(\mathbf{u}_i^{\alpha})^{\mathsf{T}}\mathbf{D}^{\alpha-1}\mathbf{J}}{(\mathbf{u}_i^{\alpha})^{\mathsf{T}}\mathbf{D}^{2(\alpha-1)}\mathbf{u}_i^{\alpha}} \xrightarrow{\text{a.s.}} \frac{(e_{00}^{\alpha}(\rho))^2}{e_{00;2}(\rho)+\chi^{\alpha}(\rho)}\mathcal{D}(\mathbf{c})^{1/2}\mathbf{v}_{\rho}(\mathbf{v}_{\rho})^{T}\mathcal{D}(\mathbf{c})^{1/2}. \tag{36}$$

We recall that one goal of this section is to estimate $\nu_i^a = \frac{1}{\sqrt{n_a}} \frac{\mathbf{u}_i^\mathsf{T} \mathbf{D}^{\alpha-1} \mathbf{j}_a}{\sqrt{\mathbf{u}_i^\mathsf{T} \mathbf{D}^{2(\alpha-1)} \mathbf{u}_i}}$, the square of which is $\frac{n}{n_a} \left[\frac{1}{n} \frac{\mathcal{D}(\mathbf{c})^{-\frac{1}{2}} \mathbf{J}^\mathsf{T} \mathbf{D}^{\alpha-1} \mathbf{u}_i^\alpha (\mathbf{u}_i^\alpha)^\mathsf{T} \mathbf{D}^{\alpha-1} \mathbf{J} \mathcal{D}(\mathbf{c})^{-\frac{1}{2}}}{(\mathbf{u}_i^\alpha)^\mathsf{T} \mathbf{D}^{2(\alpha-1)} \mathbf{u}_i^\alpha} \right]_{aa}$. From Equation (36), the former quantity is easily retrieved and we have

$$|\nu_i^a|^2 = \frac{(e_{00}^\alpha(\rho_i))^2}{e_{00;2}^\alpha(\rho_i) + \chi^\alpha(\rho_i)} \left| v_a^i \right|^2 . \tag{37}$$

This proves the following Theorem giving the limit of the empirical class means ν_i^a's.

Theorem 21 (Means) *For each eigenpair* $(\lambda(\bar{\mathbf{M}}), \mathbf{v})$ *of* $\mathcal{D}(\mathbf{c})^{\frac{1}{2}} \left(\mathbf{I}_K - \mathbf{1}_K \mathbf{c}^\mathsf{T} \right) \mathbf{M} \left(\mathbf{I}_K - \mathbf{c} \mathbf{1}_K^T \right) \mathcal{D}(\mathbf{c})^{\frac{1}{2}}$ *of unit multiplicity, mapped to eigenpair* $(\rho, \mathbf{u}_i^\alpha)$ *of* $\mathbf{L}_\alpha$ *as defined in Corollary 5, under the conditions of Assumption 1 and for* ν_i^a *defined in* (26)*, we have almost surely as* $n \to \infty$, $\left| (\nu_i^a)^2 - (\nu_i^{a,\infty})^2 \right| \to 0$ *where*

$$(\nu_i^{a,\infty})^2 \equiv \frac{\left[e_{00}^\alpha(\rho) \right]^2}{e_{00;2}^\alpha(\rho, \rho) + \chi^\alpha(\rho)} (v_a)^2$$

with $\chi^\alpha(\rho) = \frac{\left[\left(1 + e_{42;2}^\alpha(\rho)\right) e_{-1,0}^\alpha(\rho) - e_{32;2}^\alpha(\rho) e_{00}^\alpha(\rho) \right] e_{32;3}^\alpha(\rho) - \left[e_{22;2}^\alpha(\rho) e_{-1,0}^\alpha(\rho) + \left(1 - e_{22;2}^\alpha(\rho)\right) e_{00}^\alpha(\rho) \right] e_{42;3}^\alpha(\rho)}{\left(1 + e_{42;2}^\alpha(\rho)\right)\left(1 - e_{22;2}^\alpha(\rho)\right) + \left[e_{32;2}^\alpha(\rho) \right]^2}$ *and* v_a *is the component* a *of* $\mathbf{v}$.

Using the definition of ν_a^i in (26) and of $\bar{\mathbf{v}}$, $\mathbf{\Pi}$ in Theorem 9, Theorem 9 unfolds easily since $\bar{\mathbf{v}}^\mathsf{T} \mathbf{\Pi} \bar{\mathbf{v}} = \sum_{a=1}^K (\nu_a^i)^2 = \frac{\left[e_{00}^\alpha(\rho) \right]^2}{e_{00;2}^\alpha(\rho, \rho) + \chi^\alpha(\rho)} (v_a)^2$.

6.4.2 Evaluation of the class covariances σ_{ij}^a's

We have shown at the beginning of this section that to estimate the σ_{ij}^a's, we need to evaluate the more involved object

$$\frac{1}{n} \frac{\mathbf{J}^\mathsf{T} \mathbf{D}^{\alpha-1} \mathbf{u}_i^\alpha (\mathbf{u}_i^\alpha)^\mathsf{T} \mathbf{D}^{\alpha-1} \mathcal{D}_a \mathbf{D}^{\alpha-1} \mathbf{u}_j^\alpha (\mathbf{u}_j^\alpha)^\mathsf{T} \mathbf{D}^{\alpha-1} \mathbf{J}}{\left((\mathbf{u}_i^\alpha)^\mathsf{T} \mathbf{D}^{2(\alpha-1)} \mathbf{u}_i^\alpha \right) \left((\mathbf{u}_j^\alpha)^\mathsf{T} \mathbf{D}^{2(\alpha-1)} \mathbf{u}_j^\alpha \right)}.$$

Similarly to what was done previously for the estimation of $\frac{1}{n} \frac{\mathbf{J}^\mathsf{T} \mathbf{D}^{\alpha-1} \mathbf{u}_i^\alpha (\mathbf{u}_i^\alpha)^\mathsf{T} \mathbf{D}^{\alpha-1} \mathbf{J}}{(\mathbf{u}_i^\alpha)^\mathsf{T} \mathbf{D}^{2(\alpha-1)} \mathbf{u}_i^\alpha}$, we need here to evaluate

$$\left(\frac{1}{2\pi i} \right)^2 \oint_{\Gamma_{\rho 1}} \oint_{\Gamma_{\rho 2}} \frac{1}{n} \mathbf{J}^\mathsf{T} \mathbf{D}^{\alpha-1} \left(\mathbf{L}_\alpha - z_1 \mathbf{I}_n \right)^{-1} \mathbf{D}^{\alpha-1} \mathbf{D}_a \mathbf{D}^{\alpha-1} \left(\mathbf{L}_\alpha - z_2 \mathbf{I}_n \right)^{-1} \mathbf{D}^{\alpha-1} \mathbf{J} \mathrm{d}z_1 \mathrm{d}z_2$$

where Γ_{ρ_1} and Γ_{ρ_2} are two positively oriented contours circling around some limiting isolated eigenvalues ρ_1 and ρ_2 respectively. We will use the same technique as in the proof of Theorem 21 to evaluate this integrand. Namely, by applying the Woodburry identity to each of the inverse in the integrand, we get

$$\begin{aligned} \left(\frac{1}{2\pi i} \right)^2 \oint_{\Gamma_{\rho_1}} \oint_{\Gamma_{\rho_2}} \frac{1}{n} \mathbf{J}^T \mathbf{D}^{\alpha-1} \mathbf{Q}_{z_1}^\alpha \mathbf{U} \mathbf{\Lambda} \left(\mathbf{I}_{K+1} + \mathbf{U}^T \mathbf{Q}_{z_1}^\alpha \mathbf{U} \mathbf{\Lambda} \right)^{-1} \mathbf{U}^T \mathbf{Q}_{z_1}^\alpha \mathbf{D}^{\alpha-1} \mathbf{D}_a \mathbf{D}^{\alpha-1} \mathbf{Q}_{z_2}^\alpha \mathbf{U} \\ \times \mathbf{\Lambda} \left(\mathbf{I}_{K+1} + \mathbf{U}^T \mathbf{Q}_{z_2}^\alpha \mathbf{U} \mathbf{\Lambda} \right)^{-1} \mathbf{U}^T \mathbf{Q}_{z_2}^\alpha \mathbf{D}^{\alpha-1} \mathbf{J} \mathrm{d}z_1 \mathrm{d}z_2 \end{aligned}$$

where we have used the fact that the cross-terms $\frac{1}{n}\mathbf{J}^T\mathbf{D}^{\alpha-1}\mathbf{Q}^\alpha_{z_i}\mathbf{D}^{\alpha-1}\mathbf{J}$, $i=1,2$ will vanish asymptotically as the latter do not have poles in the considered contours. By using the identity $\mathbf{\Lambda}\left(\mathbf{I}_{K+1}+\mathbf{U}^T\mathbf{Q}^\alpha_{z_1}\mathbf{U}\mathbf{\Lambda}\right)^{-1}\mathbf{U}^T=\left(\mathbf{I}_{K+1}+\mathbf{\Lambda}\mathbf{U}^T\mathbf{Q}^\alpha_{z_1}\mathbf{U}\right)^{-1}\mathbf{\Lambda}\mathbf{U}^T$, the previous integral writes

$$\left(\frac{1}{2\pi i}\right)^2\oint_{\Gamma_{\rho_1}}\oint_{\Gamma_{\rho_2}}\frac{1}{n}\mathbf{J}^T\mathbf{D}^{\alpha-1}\mathbf{Q}^\alpha_{z_1}\mathbf{U}\mathbf{\Lambda}\left(\mathbf{I}_{K+1}+\mathbf{U}^T\mathbf{Q}^\alpha_{z_1}\mathbf{U}\mathbf{\Lambda}\right)^{-1}\mathbf{U}^T\mathbf{Q}^\alpha_{z_1}\mathbf{D}^{\alpha-1}\mathbf{D}_a\mathbf{D}^{\alpha-1}\mathbf{Q}^\alpha_{z_2}\mathbf{U}$$
$$\times\left(\mathbf{I}_{K+1}+\mathbf{\Lambda}\mathbf{U}^T\mathbf{Q}^\alpha_{z_2}\mathbf{U}\right)^{-1}\mathbf{\Lambda}\mathbf{U}^T\mathbf{Q}^\alpha_{z_2}\mathbf{D}^{\alpha-1}\mathbf{J}\mathrm{d}z_1\mathrm{d}z_2$$

Most of those quantities have been evaluated in the evaluation of the ν_i^a's. We thus obtain

$$\left(\frac{1}{2\pi i}\right)^2\oint_{\Gamma_{\rho_1}}\oint_{\Gamma_{\rho_2}}\left[\frac{1}{n}\mathbf{J}^T\mathbf{D}^{\alpha-1}\mathbf{Q}^\alpha_{z_1}\mathbf{U}\mathbf{\Lambda}\begin{pmatrix}(\underline{\mathbf{G}}^\alpha_{z_1})^{-1} & 0\\ 0 & 0\end{pmatrix}\mathbf{U}^T\mathbf{Q}^\alpha_{z_1}\mathbf{D}^{\alpha-1}\mathbf{D}_a\mathbf{D}^{\alpha-1}\mathbf{Q}^\alpha_{z_2}\mathbf{U}\right.$$
$$\left.\times\begin{pmatrix}\left((\underline{\mathbf{G}}^\alpha_{z_2})^{-1}\right)^T & 0\\ 0 & 0\end{pmatrix}\mathbf{\Lambda}\mathbf{U}^T\mathbf{Q}^\alpha_{z_2}\mathbf{D}^{\alpha-1}\mathbf{J}+\mathbf{R}_4(z_1,z_2)\right]\mathrm{d}z_1\mathrm{d}z_2$$

where $\mathbf{R}_4(z_1,z_2)$ has no poles in the considered contours. It is then sufficient to evaluate the top left entry of each of the matrices $\mathbf{J}^T\mathbf{D}^{\alpha-1}\mathbf{Q}^\alpha_{z_1}\mathbf{U}\mathbf{\Lambda}$, $\mathbf{U}^T\mathbf{Q}^\alpha_{z_1}\mathbf{D}^{\alpha-1}\mathbf{D}_a\mathbf{D}^{\alpha-1}\mathbf{Q}^\alpha_{z_2}\mathbf{U}$ and $\mathbf{\Lambda}\mathbf{U}^T\mathbf{Q}^\alpha_{z_2}\mathbf{D}^{\alpha-1}\mathbf{J}$ to compute the whole integrand. The first and the third of the latter matrices have been evaluated in the proof of Theorem 21. We are then left with the top left entry of $\mathbf{U}^T\mathbf{Q}^\alpha_{z_1}\mathbf{D}^{\alpha-1}\mathbf{D}_a\mathbf{D}^{\alpha-1}\mathbf{Q}^\alpha_{z_2}\mathbf{U}$ which is $\frac{1}{n}\mathbf{J}^T\mathbf{D}^{\alpha-1}\mathbf{Q}^\alpha_{z_1}\mathbf{D}^{\alpha-1}\mathbf{D}_a\mathbf{D}^{\alpha-1}\mathbf{Q}^\alpha_{z_2}\mathbf{D}^{\alpha-1}\mathbf{J}$ from Theorem 1. The former quantity has already been evaluated in the previous section but for (z,z) replaced by (z_1,z_2) and the diagonal matrix between $\mathbf{Q}^\alpha_{z_1}$ and $\mathbf{Q}^\alpha_{z_2}$ being here $\mathbf{D}^{\alpha-1}\mathbf{D}_a\mathbf{D}^{\alpha-1}$ instead of $\mathbf{D}^{2(\alpha-1)}$. We thus have

$$\left(\mathbf{U}^T\mathbf{Q}^\alpha_{z_1}\mathbf{D}^{\alpha-1}\mathbf{D}_a\mathbf{D}^{\alpha-1}\mathbf{Q}^\alpha_{z_2}\mathbf{U}\right)_{11}\xrightarrow{\text{a.s.}}c_a\left(e^\alpha_{00;2}(z_1,z_2)\mathcal{D}\left(\boldsymbol{\delta}_{i=a}\right)_{i=1}^K+\chi^\alpha(z_1,z_2)\mathcal{D}(\mathbf{c})\right).$$

Finally, we are left to evaluate

$$\left(\frac{1}{2\pi i}\right)^2\oint_{\Gamma_{\rho_1}}\oint_{\Gamma_{\rho_2}}\frac{1}{n}\left[e^\alpha_{00}(z_1)(\mathcal{D}(\mathbf{c})-\mathbf{c}\mathbf{c}^\mathsf{T})\mathbf{M}(\mathbf{I}_K-\mathbf{c}\mathbf{1}_K^\mathsf{T})-\beta^\alpha(z_1)\mathbf{c}\mathbf{1}_K^\mathsf{T}\right](\underline{\mathbf{G}}^\alpha_{z_1})^{-1}$$
$$\times c_a\left(e^\alpha_{00;2}(z_1,z_2)\mathcal{D}\left(\boldsymbol{\delta}_{i=a}\right)_{i=1}^K+\chi^\alpha(z_1,z_2)\mathcal{D}(\mathbf{c})\right)$$
$$\times\left((\underline{\mathbf{G}}^\alpha_{z_2})^{-1}\right)^\mathsf{T}\left[e^\alpha_{00}(z_2)(\mathbf{I}_K-\mathbf{1}_K\mathbf{c}^\mathsf{T})\mathbf{M}(\mathcal{D}(\mathbf{c})-\mathbf{c}\mathbf{c}^\mathsf{T})-\beta^\alpha(z_2)\mathbf{1}_K\mathbf{c}^\mathsf{T}\right]\mathrm{d}z_1\mathrm{d}z_2.$$

We can then perform a residue calculus similar to what was done in the proof of Theorem 21. Additionnaly, we use the fact that the eigenvectors $\mathbf{v}_{\rho_1}$ and $\mathbf{v}_{\rho_2}$ corresponding to distinct eigenvalues ρ_1 and ρ_2 of the symmetric matrix $\mathcal{D}(\mathbf{c})^{\frac{1}{2}}\left(\mathbf{I}_K-\mathbf{1}_K\mathbf{c}^\mathsf{T}\right)\mathbf{M}\left(\mathbf{I}_K-\mathbf{c}\mathbf{1}_K^T\right)\mathcal{D}(\mathbf{c})^{\frac{1}{2}}$ are orthogonals. All calculus done, we get

$$\left(\frac{1}{n}\frac{\mathbf{J}^\mathsf{T}\mathbf{D}^{\alpha-1}\mathbf{u}_i^\alpha(\mathbf{u}_i^\alpha)^\mathsf{T}\mathbf{D}^{\alpha-1}\mathbf{D}_a\mathbf{D}^{\alpha-1}\mathbf{u}_j^\alpha(\mathbf{u}_j^\alpha)^\mathsf{T}\mathbf{D}^{\alpha-1}\mathbf{J}}{\left((\mathbf{u}_i^\alpha)^\mathsf{T}\mathbf{D}^{2(\alpha-1)}\mathbf{u}_i^\alpha\right)\left((\mathbf{u}_j^\alpha)^\mathsf{T}\mathbf{D}^{2(\alpha-1)}\mathbf{u}_j^\alpha\right)}\right)_{ef}\xrightarrow{\text{a.s.}}$$
$$\frac{e^\alpha_{00}(\rho_i)e^\alpha_{00}(\rho_j)}{\left(e^\alpha_{00;2}(\rho_i,\rho_i)+\chi^\alpha(\rho_i)\right)\left(e^\alpha_{00;2}(\rho_j,\rho_j)+\chi^\alpha(\rho_j)\right)}$$
$$\times\left[e^\alpha_{00;2}(\rho_i,\rho_j)\sqrt{c_ec_f}v_e^iv_f^jv_a^iv_a^j+\delta_{\rho_i=\rho_j}c_a\chi^\alpha(\rho_i)\sqrt{c_ec_f}v_e^iv_f^i\right].\tag{38}$$

We are thus now ready to evaluate the σ_{ij}^a's. By definition,

$$\sigma_{ij}^a = \left[\frac{(\mathbf{u}_i^\alpha)^\mathsf{T} \mathbf{D}^{\alpha-1} \mathbf{D}_a \mathbf{D}^{\alpha-1} \mathbf{u}_j^\alpha}{\sqrt{(\mathbf{u}_i^\alpha)^\mathsf{T} \mathbf{D}^{2(\alpha-1)} \mathbf{u}_i^\alpha} \sqrt{(\mathbf{u}_j^\alpha)^\mathsf{T} \mathbf{D}^{2(\alpha-1)} \mathbf{u}_j^\alpha}} - \frac{1}{n_a} \frac{(\mathbf{u}_i^\alpha)^\mathsf{T} \mathbf{D}^{\alpha-1} \mathbf{j}_a}{\sqrt{(\mathbf{u}_i^\alpha)^\mathsf{T} \mathbf{D}^{2(\alpha-1)} \mathbf{u}_i^\alpha}} \frac{(\mathbf{u}_j^\alpha)^\mathsf{T} \mathbf{D}^{\alpha-1} \mathbf{j}_a}{\sqrt{(\mathbf{u}_j^\alpha)^\mathsf{T} \mathbf{D}^{2(\alpha-1)} \mathbf{u}_j^\alpha}} \right]. \tag{39}$$

The first right hand side term is estimated by dividing $\left(\frac{1}{n} \frac{\mathbf{J}^\mathsf{T} \mathbf{D}^{\alpha-1} \mathbf{u}_i^\alpha (\mathbf{u}_i^\alpha)^\mathsf{T} \mathbf{D}^{\alpha-1} \mathbf{D}_a \mathbf{D}^{\alpha-1} \mathbf{u}_j^\alpha (\mathbf{u}_j^\alpha)^\mathsf{T} \mathbf{D}^{\alpha-1} \mathbf{J}}{\left((\mathbf{u}_i^\alpha)^\mathsf{T} \mathbf{D}^{2(\alpha-1)} \mathbf{u}_i^\alpha\right)\left((\mathbf{u}_j^\alpha)^\mathsf{T} \mathbf{D}^{2(\alpha-1)} \mathbf{u}_j^\alpha\right)} \right)_{ef}$ (Equation 38) by $\frac{1}{\sqrt{n}} \frac{(\mathbf{u}_i^\alpha)^\mathsf{T} \mathbf{D}^{\alpha-1} \mathbf{j}_e}{\sqrt{(\mathbf{u}_i^\alpha)^\mathsf{T} \mathbf{D}^{2(\alpha-1)} \mathbf{u}_i^\alpha}} \neq 0$ and $\frac{1}{\sqrt{n}} \frac{(\mathbf{u}_j^\alpha)^\mathsf{T} \mathbf{D}^{\alpha-1} \mathbf{j}_f}{\sqrt{(\mathbf{u}_j^\alpha)^\mathsf{T} \mathbf{D}^{2(\alpha-1)} \mathbf{u}_j^\alpha}} \neq 0$ for any couple of indexes (e, f) such that the aforementioned quantities are non zeros. Indeed from the definition of ν_i^a and Equation (37), we get

$$\frac{1}{\sqrt{n}} \frac{(\mathbf{u}_i^\alpha)^\mathsf{T} \mathbf{D}^{\alpha-1} \mathbf{j}_e}{\sqrt{(\mathbf{u}_i^\alpha)^\mathsf{T} \mathbf{D}^{2(\alpha-1)} \mathbf{u}_i^\alpha}} \xrightarrow{\text{a.s.}} \sqrt{c_e} \frac{e_{00}^\alpha(\rho)}{\sqrt{e_{00;2}^\alpha(\rho, \rho) + \chi^\alpha(\rho)}} \left| v_e^i \right|. \tag{40}$$

The covariances σ_{ij}^a's are then found by combining the previous estimates (38) and (40) as per the Definition (39) of the σ_{ij}^a's. This proves the following theorem giving the limit of the empirical class covariances σ_{ij}^a's.

Theorem 22 (Covariances) *For two unit multiplicity eigenpairs $(\lambda_1(\bar{\mathbf{M}}), \mathbf{v}^1)$ and $(\lambda_2(\bar{\mathbf{M}}), \mathbf{v}^2)$ of $\mathcal{D}(\mathbf{c})^{\frac{1}{2}} \left(\mathbf{I}_K - \mathbf{1}_K \mathbf{c}^\mathsf{T}\right) \mathbf{M} \left(\mathbf{I}_K - \mathbf{c} \mathbf{1}_K^T\right) \mathcal{D}(\mathbf{c})^{\frac{1}{2}}$ mapped respectively to $(\rho_1, \mathbf{u}_i^\alpha)$ and $(\rho_2, \mathbf{u}_j^\alpha)$ eigenpairs of $\mathbf{L}_\alpha$ and for σ_{ij}^a defined in (27), we have almost surely as $n \to \infty$, $\left|\sigma_{ij}^a - \sigma_{ij}^{a,\infty}\right| \to 0$ where*

$$\sigma_{ij}^{a,\infty} \equiv \frac{\left[\left(e_{00;2}^\alpha(\rho_1, \rho_2) - e_{00}^\alpha(\rho_1) e_{00}^\alpha(\rho_2)\right) v_a^{\rho_1} v_a^{\rho_2} + \delta_{\rho_1}^{\rho_2} c_a \chi^\alpha(\rho_1)\right]}{\sqrt{e_{00;2}^\alpha(\rho_1) + \chi^\alpha(\rho_1)} \sqrt{e_{00;2}^\alpha(\rho_2) + \chi^\alpha(\rho_2)}}$$

where $\chi^\alpha(\rho)$ is defined in Theorem 21.

From Theorems 21 and 22, $\nu_i^{a,\infty}$ and $\sigma_{ij}^{a,\infty}$ depend on the e_{ij}'s (defined in Theorem 3), the normalized eigenvectors $\mathbf{v}$ of $\mathcal{D}(\mathbf{c})^{\frac{1}{2}} \left(\mathbf{I}_K - \mathbf{1}_K \mathbf{c}^\mathsf{T}\right) \mathbf{M} \left(\mathbf{I}_K - \mathbf{c} \mathbf{1}_K^T\right) \mathcal{D}(\mathbf{c})^{\frac{1}{2}}$ and the proportions c_a's of classes. Thanks to Lemma 10, the e_{ij}'s can consistently be estimated similarly to what was described in Proposition 11. Namely, the q_i's can be estimated using $\hat{q}_i = \frac{d_i}{\sqrt{\mathbf{d}^\mathsf{T} \mathbf{1}_n}}$ and replaced in Equations (20), (21), (22) to obtain consistent estimates for the e_{ij}'s. However, the eigenvectors $\mathbf{v}$ and the class proportions are not directly accessible in practice. Nevertheless, in the particular case of $K = 2$ classes, we know exactly $\mathbf{v}$.

Remark 23 ($K = 2$ classes) *Here, only one isolated eigenvector is used for the classification. Since $\mathbf{v}_r$ (right eigenvector of $\bar{\mathbf{M}}$) is orthogonal to $\mathbf{1}_2$, $\mathbf{v}_r$ is necessarily the vector $[1, -1]^\mathsf{T}$. Hence, the normalized eigenvector $\mathbf{v} = \frac{\mathcal{D}(\mathbf{c})^{-\frac{1}{2}} \mathbf{v}_r}{\|\mathcal{D}(\mathbf{c})^{-\frac{1}{2}} \mathbf{v}_r\|}$ is $\frac{1}{\sqrt{1/c_1 + 1/c_2}} \left[\frac{1}{\sqrt{c_1}}, -\frac{1}{\sqrt{c_2}}\right]^\mathsf{T}$.*

We thus obtain from Theorems 21 and 22 along with Remark 23,

Corollary 24 (Means and covariances for $K = 2$ classes) *For $a = 1, 2$*

$$(\nu^{a,\infty})^2 = \frac{[e_{00}^{\alpha}(\rho)]^2}{\left(e_{00;2}^{\alpha}(\rho,\rho) + \chi^{\alpha}(\rho)\right)\left(1 + \frac{c_a}{1-c_a}\right)}$$

$$(\sigma^{a,\infty})^2 = \frac{\left[\frac{\left(e_{00;2}^{\alpha}(\rho,\rho) - e_{00}^{\alpha}(\rho)^2\right)}{\left(1+\frac{c_a}{1-c_a}\right)} + c_a \chi^{\alpha}(\rho)\right]}{e_{00;2}^{\alpha}(\rho,\rho) + \chi^{\alpha}(\rho)}$$

for ρ the unique isolated eigenvalue of $\mathbf{L}_\alpha$ (if it exists).

6.5 Non informative eigenvectors

The objective of this section is to show that the eigenvectors $\tilde{\mathbf{u}}^\alpha$ of $\mathbf{L}_\alpha$ associated to the limiting eigenvalue $\tilde{\rho}$ for which $1 + \theta^\alpha(\tilde{\rho}) = 0$ (Remark 19) are not useful for the classification.

Let us write as in Section 6.4

$$\tilde{\mathbf{u}}^\alpha = \sum_{a=1}^{K} \tilde{\nu}^a \frac{\mathbf{j}_a}{\sqrt{n_a}} + \sqrt{\tilde{\sigma}_{ii}^\alpha}\mathbf{w}^a \tag{41}$$

where $\mathbf{w}^a \in \mathbb{R}^n$ is a random vector orthogonal to $\mathbf{j}_a$ of norm $\sqrt{n_a}$, supported on the indices of $\mathcal{C}_a$ with identically distributed entries. We shall show that $\tilde{\nu}^a$ is independent of class $\mathcal{C}_a$ and thus, any correct classification cannot be done using $\tilde{\mathbf{u}}^\alpha$. From (41), $\tilde{\nu}^a = \frac{(\tilde{\mathbf{u}}^\alpha)^\mathsf{T}\mathbf{j}_a}{\sqrt{n_a}}$ which can be retrieved from the diagonal elements of $\frac{1}{n}\mathbf{J}^\mathsf{T}\tilde{\mathbf{u}}^\alpha(\tilde{\mathbf{u}}^\alpha)^\mathsf{T}\mathbf{J}$. We will evaluate this object by using the same technique as in Section 6.4. By the residue formula, we have

$$\frac{1}{n}\mathbf{J}^\mathsf{T}\tilde{\mathbf{u}}^\alpha(\tilde{\mathbf{u}}^\alpha)^\mathsf{T}\mathbf{J} = -\frac{1}{2\pi i}\oint_{\Gamma_{\tilde{\rho}}} \frac{1}{n}\mathbf{J}^\mathsf{T}\left(\mathbf{L}_\alpha - z\mathbf{I}_n\right)^{-1}\mathbf{J}\mathrm{d}z \tag{42}$$

$$= -\frac{1}{2\pi i}\oint_{\Gamma_{\tilde{\rho}}} \frac{1}{n}\mathbf{J}^\mathsf{T}\mathbf{Q}_z^\alpha\mathbf{J}\mathrm{d}z + \frac{1}{2\pi i}\oint_{\Gamma_{\tilde{\rho}}} \frac{1}{n}\mathbf{J}^\mathsf{T}\mathbf{Q}_z^\alpha\mathbf{U}\mathbf{\Lambda}\left(\mathbf{I}_{K+1} + \mathbf{U}^\mathsf{T}\mathbf{Q}_z^\alpha\mathbf{U}\mathbf{\Lambda}\right)^{-1}\mathbf{U}^\mathsf{T}\mathbf{Q}_z^\alpha\mathbf{J} \tag{43}$$

for large n almost surely, where $\Gamma_{\tilde{\rho}}$ is a complex (positively oriented) contour circling around the limiting eigenvalue $\tilde{\rho}$ only. The first integral $-\frac{1}{2\pi i}\oint_{\Gamma_{\tilde{\rho}}}\frac{1}{n}\mathbf{J}^\mathsf{T}\mathbf{Q}_z^\alpha\mathbf{J}\mathrm{d}z$ is asymptotically zero since, from Proposition 17, the integrand has no poles in the contour $\Gamma_{\tilde{\rho}}$. We thus obtain similarly as in Section 6.4

$$\frac{1}{n}\mathbf{J}^\mathsf{T}\tilde{\mathbf{u}}^\alpha(\tilde{\mathbf{u}}^\alpha)^\mathsf{T}\mathbf{J} = \frac{1}{n}\mathbf{J}^\mathsf{T}\mathbf{Q}_{\tilde{\rho}}^\alpha\mathbf{U}\mathbf{\Lambda}\left[\lim_{z\to\tilde{\rho}}(z-\tilde{\rho})(\mathbf{I}_{K+1} + \mathbf{U}^\mathsf{T}\mathbf{Q}_z^\alpha\mathbf{U}\mathbf{\Lambda})^{-1}\right]\mathbf{U}^\mathsf{T}\mathbf{Q}_{\tilde{\rho}}^\alpha\mathbf{J}. \tag{44}$$

From (29), the entries $(1,2)$ and $(2,2)$ of $(\mathbf{I}_{K+1} + \mathbf{U}^\mathsf{T}\mathbf{Q}_z^\alpha\mathbf{U}\mathbf{\Lambda})^{-1}$ do not contain $(\underline{\mathbf{G}}_z^\alpha)^{-1}$ since $\left[\underline{\mathbf{G}}_z^\alpha - \frac{\gamma(z)m_\mu e_{21}^\alpha(z)}{ze_{10}^\alpha(z)}\mathbf{c}\mathbf{1}_K^\mathsf{T}\right]^{-1}\mathbf{c} = -\frac{m_\mu}{ze_{10}^\alpha(z)}\mathbf{c}$ and thus, the above limit will give zero for those entries. We thus get

$$\frac{1}{n}\mathbf{J}^\mathsf{T}\tilde{\mathbf{u}}^\alpha(\tilde{\mathbf{u}}^\alpha)^\mathsf{T}\mathbf{J} = \frac{1}{n}\mathbf{J}^\mathsf{T}\mathbf{Q}_{\tilde{\rho}}^\alpha\mathbf{U}\mathbf{\Lambda}\left[\lim_{z\to\tilde{\rho}}(z-\tilde{\rho})\begin{pmatrix}(\underline{\mathbf{G}}_z^\alpha)^{-1} & 0\\ \frac{\gamma(z)m_\mu}{ze_{21}^\alpha(z)}\mathbf{1}_K^\mathsf{T}(\underline{\mathbf{G}}_z^\alpha)^{-1} & 0\end{pmatrix}\right]\mathbf{U}^\mathsf{T}\mathbf{Q}_{\tilde{\rho}}^\alpha\mathbf{J}. \tag{45}$$

We recall that in the case under study ($1 + \theta^\alpha(\tilde{\rho}) = 0$), $\mathbf{1}_K$ and $\mathbf{c}$ are respectively left and right eigenvectors of $\underline{\mathbf{G}}_z^\alpha$ associated to the vanishing eigenvalue. We can thus write

$\underline{\mathbf{G}}_z^\alpha = \rho_z \mathbf{c}\mathbf{1}_K^\mathsf{T} + \tilde{\mathbf{V}}_{r,z}\tilde{\mathbf{\Sigma}}_z\tilde{\mathbf{V}}_{l,z}^T$ where ρ_z is the vanishing eigenvalue when $z \to \tilde{\rho}$ and $\tilde{\mathbf{V}}_{r,z}$ and $\tilde{\mathbf{V}}_{l,z}$ are respectively sets of right and left eigenspaces associated with non vanishing eigenvalues. Hence, we have

$$\lim_{z\to\tilde{\rho}}(z-\tilde{\rho})(\underline{\mathbf{G}}_z^\alpha)^{-1} \overset{(1)}{=} \lim_{z\to\tilde{\rho}} \frac{\mathbf{c}\mathbf{1}_K^\mathsf{T}}{\rho_z'} \overset{(2)}{=} \lim_{z\to\tilde{\rho}} \frac{\mathbf{c}\mathbf{1}_K^\mathsf{T}}{\mathbf{1}_K^\mathsf{T}(\underline{\mathbf{G}}_z^\alpha)'\mathbf{c}} \overset{(3)}{=} \lim_{z\to\tilde{\rho}} \frac{\mathbf{c}\mathbf{1}_K^\mathsf{T}}{(\theta^\alpha(z))'} \overset{(4)}{=} \frac{\mathbf{c}\mathbf{1}_K^\mathsf{T}}{(\theta^\alpha(\tilde{\rho}))'} \tag{46}$$

where in (1) we have used the l'Hopital rule, in (2) we used the fact that ρ_z can be written $\rho_z = \mathbf{1}_K^\mathsf{T}\underline{\mathbf{G}}_z^\alpha\mathbf{c}$ and in (3) we have used $(\underline{\mathbf{G}}_{\tilde{\rho}}^\alpha)' = (e_{21}^\alpha(\tilde{\rho}))' \left(\mathcal{D}(\mathbf{c}) - \mathbf{c}\mathbf{c}^\mathsf{T}\right)\mathbf{M}\left(\mathbf{I}_K - \mathbf{c}\mathbf{1}_K^T\right) + (\theta^\alpha(\tilde{\rho}))'\mathbf{c}\mathbf{1}_K^T$ and $\mathbf{1}_K^\mathsf{T}\mathbf{c} = 1$. We then have

$$\frac{1}{n}\mathbf{J}^\mathsf{T}\tilde{\mathbf{u}}^\alpha(\tilde{\mathbf{u}}^\alpha)^\mathsf{T}\mathbf{J} = \frac{1}{n}(\mathbf{J}^\mathsf{T}\mathbf{Q}_{\tilde{\rho}}^\alpha\mathbf{U}\mathbf{\Lambda})_{11}\frac{\mathbf{c}\mathbf{1}_K^\mathsf{T}}{(\theta^\alpha(\tilde{\rho}))'}(\mathbf{U}^\mathsf{T}\mathbf{Q}_{\tilde{\rho}}^\alpha\mathbf{J})_{11} \tag{47}$$

$$+ \frac{1}{n}(\mathbf{J}^\mathsf{T}\mathbf{Q}_{\tilde{\rho}}^\alpha\mathbf{U}\mathbf{\Lambda})_{12}\frac{\gamma(\tilde{\rho})m_\mu\mathbf{1}_K^\mathsf{T}}{\tilde{\rho}e_{21}^\alpha(\tilde{\rho})(\theta^\alpha(\tilde{\rho}))'}(\mathbf{U}^\mathsf{T}\mathbf{Q}_{\tilde{\rho}}^\alpha\mathbf{J})_{11}. \tag{48}$$

All calculus done similarly as in Section 6.4, we get

$$\begin{aligned}\frac{1}{n}\mathbf{J}^\mathsf{T}\tilde{\mathbf{u}}^\alpha(\tilde{\mathbf{u}}^\alpha)^\mathsf{T}\mathbf{J} \xrightarrow{\text{a.s.}} \frac{e_{1,\frac{1}{2}}^\alpha(\tilde{\rho})}{(\theta^\alpha(\tilde{\rho}))'}&\Big[e_{1,\frac{1}{2}}^\alpha(\tilde{\rho})\left(\mathcal{D}(\mathbf{c}) - \mathbf{c}\mathbf{c}^\mathsf{T}\right)\mathbf{M}\left(\mathbf{I}_K - \mathbf{c}\mathbf{1}_K^T\right)\\
&-\frac{1}{m_\mu}\left(\int t^\alpha\mu(\mathrm{d}t) + e_{0,\frac{1}{2}}^\alpha(\tilde{\rho})\right)\mathbf{c}\mathbf{1}_K^\mathsf{T}\Big]\mathbf{c}\mathbf{c}^\mathsf{T}\\
&-(e_{1,\frac{1}{2}}^\alpha(\tilde{\rho}))^2\frac{\gamma(\tilde{\rho})m_\mu}{\tilde{\rho}e_{21}^\alpha(\tilde{\rho})(\theta^\alpha(\tilde{\rho}))'}\mathbf{c}\mathbf{c}^\mathsf{T}.\end{aligned}$$

Finally,

$$\frac{1}{n}\mathbf{J}^\mathsf{T}\tilde{\mathbf{u}}^\alpha(\tilde{\mathbf{u}}^\alpha)^\mathsf{T}\mathbf{J} \xrightarrow{\text{a.s.}} -\frac{e_{1,\frac{1}{2}}^\alpha(\tilde{\rho})}{m_\mu(\theta^\alpha(\tilde{\rho}))'}\left[\int t^\alpha\mu(\mathrm{d}t) + e_{0,\frac{1}{2}}^\alpha(\tilde{\rho}) + \frac{e_{1,\frac{1}{2}}^\alpha(\tilde{\rho})\gamma(\tilde{\rho})m_\mu^2}{\tilde{\rho}e_{21}^\alpha(\tilde{\rho})}\right]\mathbf{c}\mathbf{c}^\mathsf{T}. \tag{49}$$

By recalling that $\tilde{\nu}^a = \frac{(\tilde{\mathbf{u}}^\alpha)^\mathsf{T}\mathbf{j}_a}{\sqrt{n_a}} = \sqrt{\frac{1}{n_a}\left[\frac{1}{n}\mathcal{D}(\mathbf{c})^{-\frac{1}{2}}\mathbf{J}^\mathsf{T}\tilde{\mathbf{u}}^\alpha(\tilde{\mathbf{u}}^\alpha)^\mathsf{T}\mathbf{J}\mathcal{D}(\mathbf{c})^{-\frac{1}{2}}\right]_{aa}}$, from (49) we deduce that

$$\tilde{\nu}^a \xrightarrow{\text{a.s.}} -\frac{e_{1,\frac{1}{2}}^\alpha(\tilde{\rho})}{m_\mu(\theta^\alpha(\tilde{\rho}))'}\left[\int t^\alpha\mu(\mathrm{d}t) + e_{0,\frac{1}{2}}^\alpha(\tilde{\rho}) + \frac{e_{1,\frac{1}{2}}^\alpha(\tilde{\rho})\gamma(\tilde{\rho})m_\mu^2}{\tilde{\rho}e_{21}^\alpha(\tilde{\rho})}\right]$$

which is independent of the class information (class proportions or inter-class affinities). This concludes the proof.

Acknowledgements

This work is supported by the ANR RMT4GRAPH Project (ANR-14-CE28-0006).

Appendix A. Stein Lemma and Nash Poincare inequality

Lemma 25 *Let x be a standard real Gaussian random variable and $f : \mathbb{R} \to \mathbb{R}$ be a $\mathbb{C}^1$ function with first derivative $f'(x)$ having at most polynomial growth. Then,*

$$\mathbb{E}[xf(x)] = \mathbb{E}[f'(x)].$$

Lemma 26 *Let x be a standard real Gaussian random variable and $f : \mathbb{R} \to \mathbb{R}$ be a $\mathbb{C}^1$ function with first derivative $f^{'}(x)$. Then, we have*

$$\mathrm{Var}[f(x)] \leq \mathbb{E}[|f^{'}(x)|^2].$$

The proofs of those lemma can be found in (Pastur et al., 2011).

Appendix B. Consistent estimates of the averages connectivity weights q_i's

Lemma 27 *Under Assumption 1,*

$$\max_{1\leq i\leq n} |q_i - \hat{q}_i| \to 0 \tag{50}$$

almost surely, where $\hat{q}_i = \frac{d_i}{\sqrt{\mathbf{d}^\mathsf{T}\mathbf{1}_n}}$.

We need to prove that $\sum_{n=1}^{\infty} \mathbb{P}\left(\max_{1\leq i\leq n} |q_i - \hat{q}_i| > \eta\right) < \infty$ for any $\eta > 0$ so that we can conclude from the first Borel Cantelli lemma (Theorem 4.3 in (Billingsley, 1995)) that $\mathbb{P}\left(\limsup_n \max_{1\leq i\leq n} |q_i - \hat{q}_i| > \eta\right) = 0$ from which Lemma 27 unfolds. We have that

$$\begin{aligned}\mathbb{P}\left(\max_{1\leq i\leq n} |q_i - \hat{q}_i| > \eta\right) &\leq \sum_{i=1}^{n} \mathbb{P}(|q_i - \hat{q}_i| > \eta) \\ &\leq \sum_{i=1}^{n} \mathbb{P}(\hat{q}_i - q_i > \eta) + \mathbb{P}(q_i - \hat{q}_i > \eta). \end{aligned}\tag{51}$$

Let us treat for instance the term $\mathbb{P}(\hat{q}_i - q_i > \eta)$ in the following. Since $A_{ij} = q_i q_j + q_i q_j \frac{M_{g_i g_j}}{\sqrt{n}} + X_{ij}$ with X_{ij} a zero mean random variable, we have
$\frac{1}{n}\sum_{j=1}^{n} \mathbb{E}A_{ij} \to q_i m_\mu$ and $\frac{1}{n^2}\sum_{i,j} \mathbb{E}A_{ij} \to m_\mu^2$ in the limit $n \to \infty$. For $\hat{q}_i = \frac{\sum_{j=1}^{n} A_{ij}}{\sqrt{\sum_{i,j} A_{ij}}}$, we can write

$$\begin{aligned}\hat{q}_i - q_i = &\underbrace{\frac{\frac{1}{n}\sum_{j=1}^{n}(A_{ij} - \mathbb{E}A_{ij})}{\sqrt{\frac{1}{n^2}\sum_{i,j} \mathbb{E}A_{ij}}}}_{A} + \underbrace{\frac{\frac{1}{n}\sum_{j=1}^{n} A_{ij}}{\sqrt{\frac{1}{n^2}\sum_{i,j} A_{ij}}} - \frac{\frac{1}{n}\sum_{j=1}^{n} A_{ij}}{\sqrt{\frac{1}{n^2}\sum_{i,j} \mathbb{E}A_{ij}}}}_{B} \\ &+ \underbrace{\frac{\frac{1}{n}\sum_{j=1}^{n} \mathbb{E}A_{ij}}{\sqrt{\frac{1}{n^2}\sum_{i,j} \mathbb{E}A_{ij}}} - q_i}_{C}\end{aligned}$$

Since A, B and C tend to zero in the limit $n \to \infty$, we will next use the fact that $\mathbb{P}(\hat{q}_i - q_i > \eta) \leq \mathbb{P}(A > \eta/3) + \mathbb{P}(B > \eta/3) + \mathbb{P}(C > \eta/3)$ and show that all those individual probabilities vanish asymptotically. Since the term C is deterministic and tends to zero in the limit $n \to \infty$, we have $\mathbb{P}(C > \eta/3) = 0$ for all large n. Let us then control $\mathbb{P}(A > \eta/3)$ and $\mathbb{P}(B > \eta/3)$. We have

$$\begin{aligned}\mathbb{P}(A > \eta/3) &= \mathbb{P}\left(\frac{1}{n}\sum_{j=1}^{n}(A_{ij} - \mathbb{E}A_{ij}) > \frac{\eta m_\mu}{3} + o(1)\right) \\ &\leq \exp\left[-\frac{n\eta^2 m_\mu^2}{18\left(\sigma^2 + \eta m_\mu/9\right)} + o(1)\right]\end{aligned}\tag{52}$$

with $\sigma^2 = \limsup_n \max_{1\leq i\leq n} q_i(\sum_j q_j) - q_i^2(\sum_j q_j^2)$ and where in the last inequality of (52), we have used Bernstein's inequality (Theorem 3 in (Boucheron et al., 2013)) since the A_{ij}'s are independent Bernoulli random variables with variance $\sigma_{ij}^2 = q_iq_j(1-q_iq_j) + \mathcal{O}(n^{-\frac{1}{2}})$. For the term B we have

$$\begin{aligned}
\mathbb{P}(B > \eta/3) &= \mathbb{P}\left(\frac{1}{n}\sum_{j=1}^n A_{ij}\frac{\sqrt{\frac{1}{n^2}\sum_{i,j}\mathbb{E}A_{ij}} - \sqrt{\frac{1}{n^2}\sum_{i,j}A_{ij}}}{\sqrt{\frac{1}{n^2}\sum_{i,j}A_{ij}}} > \frac{\eta m_\mu}{3} + o(1)\right)\\
&\overset{(1)}{\leq} \mathbb{P}\left(\left|\frac{\sqrt{\frac{1}{n^2}\sum_{i,j}\mathbb{E}A_{ij}} - \sqrt{\frac{1}{n^2}\sum_{i,j}A_{ij}}}{\sqrt{\frac{1}{n^2}\sum_{i,j}A_{ij}}}\right| > \frac{\eta m_\mu}{3} + o(1)\right)\\
&\overset{(2)}{\leq} \mathbb{P}\left(\left|\sqrt{\frac{1}{n^2}\sum_{i,j}\mathbb{E}A_{ij}} - \sqrt{\frac{1}{n^2}\sum_{i,j}A_{ij}}\right| > \frac{\eta(m_\mu + o(1))\sqrt{\frac{1}{n^2}\sum_{i,j}A_{ij}}}{3}, \frac{1}{n^2}\sum_{i,j}A_{ij} > \psi\right)\\
&+ \mathbb{P}\left(\left|\sqrt{\frac{1}{n^2}\sum_{i,j}\mathbb{E}A_{ij}} - \sqrt{\frac{1}{n^2}\sum_{i,j}A_{ij}}\right| > \frac{\eta m_\mu\sqrt{\frac{1}{n^2}\sum_{i,j}A_{ij}}}{3} + o(1), \frac{1}{n^2}\sum_{i,j}A_{ij} \leq \psi\right)\\
&\overset{(3)}{\leq} \underbrace{\mathbb{P}\left(\left|\sqrt{\frac{1}{n^2}\sum_{i,j}\mathbb{E}A_{ij}} - \sqrt{\frac{1}{n^2}\sum_{i,j}A_{ij}}\right| > \frac{\eta m_\mu\sqrt{\psi}}{3} + o(1)\right)}_{B_1} + \underbrace{\mathbb{P}\left(\frac{1}{n^2}\sum_{i,j}A_{ij} \leq \psi\right)}_{B_2}
\end{aligned} \tag{53}$$

where in the inequality (1) we have used the fact that $n^{-1}\sum_{j=1}^n A_{ij} \leq 1$; in the inequality (2) $\psi > 0$ is any constant smaller than m_μ^2 and in the inequality (3) we have used $\sqrt{n^{-2}\sum_{i,j}A_{ij}} > \sqrt{\psi}$ and the fact that the probability of the intersection between two events is always smaller than the probability of one of those events. It then remains to control B_1 and B_2. For B_2 we have

$$\mathbb{P}\left(\frac{1}{n^2}\sum_{i,j}A_{ij} \leq \psi\right) \leq \exp\left[-\frac{n(m_\mu^2 - \psi)^2}{2\left(\sigma^2 + (m_\mu^2 - \psi)/3\right)} + o(1)\right] \tag{54}$$

where the inequality follows from Bernstein's inequality with the similar arguments as previously. Finally for the term B_1 we have

$$
\begin{aligned}
&\mathbb{P}\left(\left|\sqrt{\frac{1}{n^2}\sum_{i,j}\mathbb{E}A_{ij}}-\sqrt{\frac{1}{n^2}\sum_{i,j}A_{ij}}\right|>\frac{\eta m_\mu\sqrt{\psi}}{3}+o(1)\right)\\
&=\mathbb{P}\left(\left|\frac{\frac{1}{n^2}\sum_{i,j}\mathbb{E}A_{ij}-\frac{1}{n^2}\sum_{i,j}A_{ij}}{\sqrt{\frac{1}{n^2}\sum_{i,j}\mathbb{E}A_{ij}}+\sqrt{\frac{1}{n^2}\sum_{i,j}A_{ij}}}\right|>\frac{\eta m_\mu\sqrt{\psi}}{3}+o(1)\right)\\
&\overset{(1)}{\leq}\mathbb{P}\left(\left|\frac{1}{n^2}\sum_{i,j}\mathbb{E}A_{ij}-\frac{1}{n^2}\sum_{i,j}A_{ij}\right|>\frac{\eta m_\mu\sqrt{\psi}(m_\mu+\sqrt{\psi})}{3}+o(1)\right)+\mathbb{P}\left(\frac{1}{n^2}\sum_{i,j}A_{ij}\leq\psi\right)\\
&\overset{(2)}{\leq}\exp\left[-\frac{n\psi\left[\eta m_\mu(m_\mu+\sqrt{\psi})\right]^2}{18\left(\sigma^2+\eta m_\mu\sqrt{\psi}(m_\mu+\sqrt{\psi})/9\right)}+o(1)\right]+\exp\left[-\frac{n(m_\mu^2-\psi)^2}{2\left(\sigma^2+(m_\mu^2-\psi)/3\right)}+o(1)\right]
\end{aligned}
\tag{55}
$$

where in the inequality (1) of Equation (55) we have used the same arguments as in the inequalities (2) − (3) of Equation (52) and in the inequality (2) we have used Bernstein's inequality along with Equation (54). From Equations (52)(54)(55), we conclude that $\sum_{n=1}^{\infty}\sum_{i=1}^{n}\mathbb{P}(\hat{q}_i-q_i>\eta)<\infty$ since $m_\mu^2-\psi>0$. It follows the same lines to show that $\sum_{n=1}^{\infty}\sum_{i=1}^{n}\mathbb{P}(q_i-\hat{q}_i>\eta)<\infty$ which concludes the proof.

Appendix C. First deterministic equivalents

Let $\mathbf{Q}_z^\alpha=\left(\bar{\mathbf{X}}-z\mathbf{I}_n\right)^{-1}$ with $\bar{\mathbf{X}}$ a symmetric random matrix having independent entries $\bar{X}_{ij}$ which are Gaussian random variables with zero mean and variance $\frac{\sigma_{ij}^2}{n}$. For short, we shall denote $\mathbf{Q}_z^\alpha$ by $\mathbf{Q}$. We want to find a deterministic equivalent $\bar{\mathbf{Q}}$ of $\mathbf{Q}$ in the sense that $\frac{1}{n}\operatorname{tr}\mathbf{C}\mathbf{Q}-\frac{1}{n}\operatorname{tr}\mathbf{C}\bar{\mathbf{Q}}\to 0$ and $\mathbf{d}_1^\mathsf{T}(\mathbf{Q}-\bar{\mathbf{Q}})\mathbf{d}_2\to 0$ almost surely, for all deterministic Hermitian matrix $\mathbf{C}$ and deterministic vectors $\mathbf{d}_i$ of bounded norms (spectral norm for matrices and Euclidian norm for vectors). To this end, we will evaluate $\mathbb{E}(\mathbf{Q})$ since using Lemma 26, one can show that $n^{-1}\operatorname{tr}(\mathbf{C}\mathbf{Q})$ and $\mathbf{a}^\mathsf{T}\mathbf{Q}\mathbf{b}$ concentrate respectively around $n^{-1}\operatorname{tr}(\mathbf{A}\mathbb{E}Q)$ and $\mathbf{d}_1^\mathsf{T}\mathbb{E}\mathbf{Q}\mathbf{d}_2$ for all bounded norm matrix $\mathbf{C}$ and vectors $\mathbf{d}_1,\mathbf{d}_2$. For the computations, we use standard Gaussian calculus introduced in (Pastur et al., 2011). Using the resolvent identity (for two invertible matrices $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{A}^{-1}-\mathbf{B}^{-1}=-\mathbf{A}^{-1}(\mathbf{A}-\mathbf{B})\mathbf{B}^{-1}$), one has

$$\mathbf{Q}=\frac{1}{z}\bar{\mathbf{X}}\mathbf{Q}-\frac{1}{z}\mathbf{I}_n. \tag{56}$$

We then first compute $\mathbb{E}(\bar{\mathbf{X}}\mathbf{Q})$. By writing $\bar{X}_{il}=\frac{\sigma_{il}}{\sqrt{n}}Z_{il}$ where Z_{il} is a random variable with zero mean and unit variance, we thus have

$$\mathbb{E}(\bar{\mathbf{X}}\mathbf{Q})_{ij}=\sum_{l=1}^{n}\frac{\sigma_{il}}{\sqrt{n}}\mathbb{E}(Z_{il}Q_{lj}).$$

By applying Stein's Lemma (Lemma 25 in Section A), we have

$$\begin{aligned}\mathbb{E}(Z_{il}Q_{lj}) &= \mathbb{E}\left(\frac{\partial(\bar{\mathbf{X}}-z\mathbf{I})^{-1}_{lj}}{\partial Z_{il}}\right)\\ &= \mathbb{E}\left(-(\bar{\mathbf{X}}-z\mathbf{I})^{-1}\frac{\partial \bar{\mathbf{X}}}{\partial Z_{il}}(\bar{\mathbf{X}}-z\mathbf{I})^{-1}\right)_{lj}\\ &= \mathbb{E}\left(-(\bar{\mathbf{X}}-z\mathbf{I})^{-1}\frac{\sigma_{il}}{\sqrt{n}}(\mathbf{E}_{il}+\mathbf{E}_{li})(\bar{\mathbf{X}}-z\mathbf{I})^{-1}\right)_{lj}\end{aligned}$$

where $\mathbf{E}_{il}$ is the matrix with all entries equal to 0 but the entry (i,l) which is equal to 1. Using simple algebra, we have

$$\left((\bar{\mathbf{X}}-z\mathbf{I})^{-1}\mathbf{E}_{il}(\bar{\mathbf{X}}-z\mathbf{I})^{-1}\right)_{lj} = (\bar{\mathbf{X}}-z\mathbf{I})^{-1}_{li}(\bar{\mathbf{X}}-z\mathbf{I})^{-1}_{lj}$$

and

$$\left((\bar{\mathbf{X}}-z\mathbf{I})^{-1}\mathbf{E}_{li}(\bar{\mathbf{X}}-z\mathbf{I})^{-1}\right)_{lj} = (\bar{\mathbf{X}}-z\mathbf{I})^{-1}_{ll}(\bar{\mathbf{X}}-z\mathbf{I})^{-1}_{ij}.$$

We thus get

$$\mathbb{E}(\bar{\mathbf{X}}\mathbf{Q})_{ij} = \sum_{l=1}^{n} -\frac{\sigma_{il}^2}{n}\left(\mathbb{E}\left[Q_{li}Q_{lj}\right]+\mathbb{E}\left[Q_{ll}Q_{ij}\right]\right).$$

Going back to (56), we thus have

$$\begin{aligned}\mathbb{E}(Q_{ij}) &= -\frac{1}{z}\sum_{l=1}^{n}\frac{\sigma_{il}^2}{n}\mathbb{E}\left[Q_{li}Q_{lj}\right] - \frac{1}{z}\sum_{l=1}^{n}\frac{\sigma_{il}^2}{n}\mathbb{E}\left[Q_{ll}Q_{ij}\right] - \frac{1}{z}\delta_{ij}\\ &= -\frac{1}{z}\mathbb{E}\left[\mathbf{Q}\frac{\mathbf{\Sigma}_i}{n}\mathbf{Q}\right]_{ij} - \frac{1}{z}\mathbb{E}\left[Q_{ij}\operatorname{tr}\left(\frac{\mathbf{\Sigma}_i\mathbf{Q}}{n}\right)\right] - \frac{1}{z}\delta_{ij} \qquad (57)\end{aligned}$$

where $\mathbf{\Sigma}_i = \mathcal{D}\left(\sigma_{ij}^2\right)_{j=1}^{n}$. Since the goal is to retrieve $\mathbb{E}(Q_{ij})$, the following lemma allows to split $\mathbb{E}\left[Q_{ij}\operatorname{tr}\left(\frac{\mathbf{\Sigma}_i\mathbf{Q}}{n}\right)\right]$ into $\mathbb{E}\left[Q_{ij}\right]$ and $\mathbb{E}\left[\operatorname{tr}\left(\frac{\mathbf{\Sigma}_i\mathbf{Q}}{n}\right)\right]$.

Lemma 28 *For $\mathbf{Q} = (\bar{\mathbf{X}} - z\mathbf{I}_n)^{-1}$ and $\mathbf{\Sigma}_i = \mathcal{D}(\sigma_{ij}^2)_{j=1}^{n}$, where $\bar{\mathbf{X}}$ is a symmetric random matrix having independent entries (up to the symmetry) of zero mean and variance $\frac{\sigma_{ij}^2}{n}$, we have*

$$\mathbb{E}\left[Q_{ij}\operatorname{tr}\left(\frac{\mathbf{\Sigma}_i\boldsymbol{Q}}{n}\right)\right] = \mathbb{E}\left[Q_{ij}\right]\mathbb{E}\left[\operatorname{tr}\left(\frac{\mathbf{\Sigma}_i\boldsymbol{Q}}{n}\right)\right] + o(1).$$

Proof For two real random variables x and y, by Cauchy-Shwarz's inequality,

$$\left|\mathbb{E}\left[(x-\mathbb{E}(x))(y-\mathbb{E}(y))\right]\right| \leq \sqrt{\operatorname{Var}(x)}\sqrt{\operatorname{Var}(y)}$$

which, for $x = \operatorname{tr}\left(\frac{\mathbf{\Sigma}_i\mathbf{Q}}{n}\right)$ and $y = Q_{ij} - \mathbb{E}(Q_{ij})$ gives

$$\left|\mathbb{E}\left[Q_{ij}\operatorname{tr}\left(\frac{\mathbf{\Sigma}_i\mathbf{Q}}{n}\right)\right] - \mathbb{E}\left[Q_{ij}\right]\mathbb{E}\left[\operatorname{tr}\left(\frac{\mathbf{\Sigma}_i\mathbf{Q}}{n}\right)\right]\right| \leq \sqrt{\operatorname{Var}(x)}\sqrt{\operatorname{Var}(y)}$$

since $\mathbb{E}(y)$ is equal to 0 in that case. Using Nash Poincaré inequality (Lemma 26 in Section A), one can show that $\mathrm{Var}(x) = \mathcal{O}\left(\frac{1}{n^2}\right)$(Hachem et al., 2007). Additionally, $\forall i, j$ and $z \in \mathbb{C}^+$, $|Q_{ij}| \leq \frac{1}{|\Im(z)|}$. This finally implies that $\mathbb{E}\left[Q_{ij} \operatorname{tr}\left(\frac{\mathbf{\Sigma}_i \mathbf{Q}}{n}\right)\right] - \mathbb{E}\left[Q_{ij}\right] \mathbb{E}\left[\operatorname{tr}\left(\frac{\mathbf{\Sigma}_i \mathbf{Q}}{n}\right)\right] = \mathcal{O}(n^{-1})$. ■

Since $\Im(-z - \mathbb{E}\operatorname{tr}(\frac{\mathbf{\Sigma}_i \mathbf{Q}}{n})) < -\Im(z)$ for $z \in \mathbb{C}^+$, $-z - \mathbb{E}\operatorname{tr}(\frac{\mathbf{\Sigma}_i \mathbf{Q}}{n})$ does not vanish asymptotically. Going back to $\mathbb{E}(Q_{ij})$ in Equation (57), we may then write

$$\mathbb{E}(Q_{ij}) = \frac{\mathbb{E}\left[\mathbf{Q}\frac{\mathbf{\Sigma}_i}{n}\mathbf{Q}\right]_{ij} + \delta_{ij}}{-z - \mathbb{E}\left[\operatorname{tr}\left(\frac{\mathbf{\Sigma}_i \mathbf{Q}}{n}\right)\right]} + \mathcal{O}(n^{-1}). \tag{58}$$

Multiplying Equation (58) by $\frac{\sigma_{ki}^2}{n}$, taking $j = i$, summing over i and scaling by n, we get

$$\operatorname{tr}\mathbb{E}\left(\frac{\mathbf{\Sigma}_k \mathbf{Q}}{n}\right) = \sum_{i=1}^{n}\left(\frac{\mathbb{E}\left[\frac{\mathbf{\Sigma}_k}{n}\mathbf{Q}\frac{\mathbf{\Sigma}_i}{n}\mathbf{Q}\right]_{ii} + \frac{\sigma_{ki}^2}{n}}{-z - \mathbb{E}\left[\operatorname{tr}\left(\frac{\mathbf{\Sigma}_i \mathbf{Q}}{n}\right)\right]}\right) + \mathcal{O}(n^{-1}).$$

Using a similar approach to the proof of Lemma 28, we can show that $\sum_{i=1}^n \mathbb{E}\left[\frac{\mathbf{\Sigma}_k}{n}\mathbf{Q}\frac{\mathbf{\Sigma}_i}{n}\mathbf{Q}\right]_{ii} = \mathcal{O}(n^{-1})$. We thus have

$$\frac{1}{n}\operatorname{tr}\mathbb{E}\left(\mathbf{\Sigma}_k \mathbf{Q}\right) = \sum_{i=1}^{n} \frac{\frac{\sigma_{ki}^2}{n}}{-z - \frac{1}{n}\mathbb{E}\left[\operatorname{tr}\left(\mathbf{\Sigma}_i \mathbf{Q}\right)\right]} + o(1).$$

By using standard techniques (Hachem et al., 2007), one can show that the unique solution $e_i(z)$ to $e_i(z) = \frac{1}{n}\sum_{j=1}^n \frac{\sigma_{ij}^2}{-z - e_j(z)}$ is such that $\frac{1}{n}\operatorname{tr}\mathbb{E}\left(\mathbf{\Sigma}_i \mathbf{Q}\right) - e_i(z) \xrightarrow{\text{a.s.}} 0$. Going back to Equation (58), we can thus write for large n

$$\mathbb{E}\left[(-z\mathbf{I} - \mathcal{D}(e_i(z)))\mathbf{Q}\right]_{ij} = \mathbb{E}\left[\mathbf{Q}\frac{\mathbf{\Sigma}_i}{n}\mathbf{Q}\right]_{ij} + \delta_{ij} + o(1). \tag{59}$$

Let us denote $\mathbf{\Xi} = -z\mathbf{I} - \mathcal{D}(e_i(z))$. Since $-z - \mathbb{E}\operatorname{tr}\left(\frac{\mathbf{\Sigma}_i \mathbf{Q}}{n}\right)$ is away from zero for $z \in \mathbb{C}^+$ so is $-z - e_i(z)$ and thus $\mathbf{\Xi}$ is invertible and bounded. For large n, we can write for a given deterministic matrix $\mathbf{C}$ of bounded norm

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{n}\operatorname{tr}\mathbf{C}\mathbf{Q}\right] &= \frac{1}{n}\sum_{i,j}(\mathbf{C}\mathbf{\Xi}^{-1})_{ji}\mathbb{E}\left(\mathbf{\Xi}\mathbf{Q}\right)_{ij} \\
&\overset{(1)}{=} \frac{1}{n}\sum_{i,j}(\mathbf{C}\mathbf{\Xi}^{-1})_{ji}\left(\mathbb{E}\left[\mathbf{Q}\frac{\mathbf{\Sigma}_i}{n}\mathbf{Q}\right]_{ij} + \delta_{ij}\right) + o(1) \\
&= \frac{1}{n}\operatorname{tr}\mathbb{E}\left(\mathbf{C}\mathbf{\Xi}^{-1}\mathbf{Q}\frac{\mathbf{\Sigma}_i}{n}\mathbf{Q}\right) + \frac{1}{n}\operatorname{tr}(\mathbf{C}\mathbf{\Xi}^{-1}) + o(1)
\end{aligned}$$

where (1) follows from Equation (59). We can then prove that $\frac{1}{n}\operatorname{tr}\mathbb{E}\left(\mathbf{C}\mathbf{\Xi}^{-1}\mathbf{Q}\frac{\mathbf{\Sigma}_i}{n}\mathbf{Q}\right) = \mathcal{O}(n^{-1})$ using a similar approach to the proof of Lemma 28. Hence for large n

$$\mathbb{E}\left[\frac{1}{n}\operatorname{tr}\mathbf{C}\mathbf{Q}\right] = \frac{1}{n}\operatorname{tr}(\mathbf{C}\mathbf{\Xi}^{-1}) + o(1). \tag{60}$$

Similarly, for any vectors $\mathbf{a}$, $\mathbf{b}$ of bounded norms, we may write

$$\begin{aligned}\mathbb{E}\left[\mathbf{a}^*\mathbf{Q}\mathbf{b}\right] &= \sum_{i,j}(\mathbf{a}^*\boldsymbol{\Xi}^{-1})_i\mathbb{E}\left(\boldsymbol{\Xi}\mathbf{Q}\right)_{ij}\mathbf{b}_j \\ &= \sum_{i,j}(\mathbf{a}^*\boldsymbol{\Xi}^{-1})_i\mathbb{E}\left[\mathbf{Q}\frac{\boldsymbol{\Sigma}_i}{n}\mathbf{Q}\right]_{ij}\mathbf{b}_j + \mathbf{a}^*\boldsymbol{\Xi}^{-1}\mathbf{b} + o(1).\end{aligned}$$

We also have that $\sum_{i,j}(\mathbf{a}^*\boldsymbol{\Xi}^{-1})_i\mathbb{E}\left[\mathbf{Q}\frac{\boldsymbol{\Sigma}_i}{n}\mathbf{Q}\right]_{ij}\mathbf{b}_j = \mathcal{O}(n^{-1})$. This can be proved similarly to the proof of Lemma 28. Hence,

$$\mathbb{E}\left[\mathbf{a}^*\mathbf{Q}\mathbf{b}\right] = \mathbf{a}^*\boldsymbol{\Xi}^{-1}\mathbf{b} + o(1). \tag{61}$$

Appendix D. Second deterministic equivalents

Our goal is to find a deterministic equivalent to the random quantity $\mathbf{Q}^\alpha_{z_1}\boldsymbol{\Xi}\mathbf{Q}^\alpha_{z_2}$ for any diagonal deterministic matrix $\boldsymbol{\Xi}$ where we recall that $\mathbf{Q}^\alpha_{z_1} = \left(\frac{\bar{\mathbf{X}}}{\sqrt{n}} - z_1\mathbf{I}_n\right)^{-1}$ with $\bar{\mathbf{X}}$ defined previously in Appendix C. The proof follows the same techniques as the proof of the first deterministic equivalent $\mathbf{Q}^\alpha_z$ in Appendix C but here, the resolvent identity is either applied on $\mathbf{Q}^\alpha_{z_1}$ or $\mathbf{Q}^\alpha_{z_2}$. The technical details will be omitted as the key techniques have already been developped in Appendix C. For the sake of readability, we will denote $\mathbf{Q}^\alpha_{z_1} \equiv \mathbf{Q}_1$ and $\mathbf{Q}^\alpha_{z_2} \equiv \mathbf{Q}_2$. As in Appendix C, we will evaluate $\mathbb{E}(\mathbf{Q}_1\boldsymbol{\Xi}\mathbf{Q}_2)$. By the resolvent identity, we have

$$\begin{aligned}\mathbb{E}(\mathbf{Q}_1\boldsymbol{\Xi}\mathbf{Q}_2)_{ij} &= -\frac{1}{z_1}\mathbb{E}(\boldsymbol{\Xi}\mathbf{Q}_2)_{ij} + \frac{1}{z_1}\mathbb{E}(\mathbf{X}\mathbf{Q}_1\boldsymbol{\Xi}\mathbf{Q}_2)_{ij} \\ &= -\frac{1}{z_1}\Xi_{ii}\mathbb{E}(\mathbf{Q}_2)_{ij} + \frac{1}{z_1}\mathbb{E}\sum_{k,l}X_{ik}(Q_1)_{kl}\Xi_{ll}(Q_2)_{lj}.\end{aligned}$$

We have from Lemma 25, $\mathbb{E}\sum_{k,l}X_{ik}(Q_1)_{kl}\Xi_{ll}(Q_2)_{lj} = \sum_{k,l}\frac{\sigma_{ik}}{\sqrt{n}}\Xi_{ll}\mathbb{E}\frac{\partial[(Q_1)_{kl}(Q_2)_{lj}]}{\partial Z_{ik}}$. By expanding all terms and all calculus done, we obtain

$$\begin{aligned}\mathbb{E}(\mathbf{Q}_1\boldsymbol{\Xi}\mathbf{Q}_2)_{ij} = -\frac{1}{z_1}\mathbb{E}(\boldsymbol{\Xi}\mathbf{Q}_2)_{ij} - \frac{1}{z_1}\sum_{k,l}\frac{\sigma^2_{ik}}{n}\Xi_{ll}\mathbb{E}\Bigg[&\underbrace{(Q_1)_{ki}(Q_1)_{kl}(Q_2)_{lj}}_{(1)} + \underbrace{(Q_1)_{kk}(Q_1)_{il}(Q_2)_{lj}}_{(2)} \\ &+ \underbrace{(Q_1)_{kl}(Q_2)_{li}(Q_2)_{kj}}_{(3)} + \underbrace{(Q_1)_{kl}(Q_2)_{lk}(Q_2)_{ij}}_{(4)}\Bigg].\end{aligned}$$

Asymptotically, the non vanishing terms are (2) and (4) so that

$$\begin{aligned}\mathbb{E}(\mathbf{Q}_1\boldsymbol{\Xi}\mathbf{Q}_2)_{ij} &= -\frac{1}{z_1}\mathbb{E}(\boldsymbol{\Xi}\mathbf{Q}_2)_{ij} - \frac{1}{z_1}\sum_{k,l}\frac{\sigma^2_{ik}}{n}\Xi_{ll}\mathbb{E}\left[(Q_1)_{kk}(Q_1)_{il}(Q_2)_{lj} + (Q_1)_{kl}(Q_2)_{lk}(Q_2)_{ij}\right] \\ &\quad + o(1) \\ &= -\frac{1}{z_1}\mathbb{E}(\boldsymbol{\Xi}\mathbf{Q}_2)_{ij} - \frac{1}{z_1}\frac{1}{n}\mathbb{E}\left[\operatorname{tr}(\boldsymbol{\Sigma}_i\mathbf{Q}_1)(\mathbf{Q}_1\boldsymbol{\Xi}\mathbf{Q}_2)_{ij}\right] - \frac{1}{z_1}\frac{1}{n}\mathbb{E}\left[\operatorname{tr}(\boldsymbol{\Sigma}_i\mathbf{Q}_2\boldsymbol{\Xi}\mathbf{Q}_1)(\mathbf{Q}_2)_{ij}\right] \\ &\quad + o(1).\end{aligned} \tag{62}$$

Similarly to what was done in the proof of of Lemma 28, we can show that $\mathbb{E}\frac{1}{n}\operatorname{tr}(\boldsymbol{\Sigma}_i\mathbf{Q}_1)\mathbb{E}(\mathbf{Q}_1\boldsymbol{\Xi}\mathbf{Q}_2)_{ij} = \mathbb{E}\left(\frac{1}{n}\operatorname{tr}(\boldsymbol{\Sigma}_i\mathbf{Q}_1)\right)\mathbb{E}\left((\mathbf{Q}_1\boldsymbol{\Xi}\mathbf{Q}_2)_{ij}\right)+o(1)$. We can then write from (62)

$$\mathbb{E}\left(\left(\mathbf{I}_n + \frac{1}{z_1}\mathcal{D}\left(\frac{1}{n}\operatorname{tr}(\boldsymbol{\Sigma}_i\mathbf{Q}_1)\right)_{i=1}^n\right)\mathbf{Q}_1\boldsymbol{\Xi}\mathbf{Q}_2\right) = \\ -\frac{1}{z_1}\mathbb{E}\left(\boldsymbol{\Xi} + \mathcal{D}\left(\frac{1}{n}\operatorname{tr}(\boldsymbol{\Sigma}_i\mathbf{Q}_2\boldsymbol{\Xi}\mathbf{Q}_1)\right)_{i=1}^n\right)\mathbf{Q}_2 + o(1). \tag{63}$$

From (63) and the result of Lemma 15, this entails

$$\mathbb{E}\left(\mathbf{Q}_1\boldsymbol{\Xi}\mathbf{Q}_2\right) \longleftrightarrow \bar{\mathbf{Q}}_1\boldsymbol{\Xi}\bar{\mathbf{Q}}_2 + \bar{\mathbf{Q}}_1\mathcal{D}\left(\mathbb{E}\frac{1}{n}\operatorname{tr}(\boldsymbol{\Sigma}_i\mathbf{Q}_2\boldsymbol{\Xi}\mathbf{Q}_1)\right)_{i=1}^n\bar{\mathbf{Q}}_2. \tag{64}$$

Every object in (64) is known but $\mathbb{E}\frac{1}{n}\operatorname{tr}(\boldsymbol{\Sigma}_i\mathbf{Q}_2\boldsymbol{\Xi}\mathbf{Q}_1)$ which we need to evaluate now. By left-multiplying (64) by $\boldsymbol{\Sigma}_i$ and taking the normalized trace, we get

$$\mathbb{E}\frac{1}{n}\operatorname{tr}(\boldsymbol{\Sigma}_i\mathbf{Q}_2\boldsymbol{\Xi}\mathbf{Q}_1) = \frac{1}{n}\operatorname{tr}\left(\boldsymbol{\Sigma}_i\bar{\mathbf{Q}}_1\boldsymbol{\Xi}\bar{\mathbf{Q}}_2\right) + \mathbb{E}\frac{1}{n}\operatorname{tr}\left(\boldsymbol{\Sigma}_i\bar{\mathbf{Q}}_1\mathcal{D}\left(\frac{1}{n}\operatorname{tr}(\boldsymbol{\Sigma}_i\mathbf{Q}_2\boldsymbol{\Xi}\mathbf{Q}_1)\right)_{i=1}^n\bar{\mathbf{Q}}_2\right). \tag{65}$$

By denoting $f_i = \frac{1}{n}\mathbb{E}\left(\operatorname{tr}(\boldsymbol{\Sigma}_i\mathbf{Q}_2\boldsymbol{\Xi}\mathbf{Q}_1)\right)$, Equation (65) leads to

$$\mathbf{f} = \left\{\frac{1}{n}\operatorname{tr}\left(\boldsymbol{\Sigma}_i\bar{\mathbf{Q}}_1\boldsymbol{\Xi}\bar{\mathbf{Q}}_2\right)\right\}_{i=1}^n + \frac{1}{n}\left\{\left(\bar{\mathbf{Q}}_2\boldsymbol{\Sigma}_i\bar{\mathbf{Q}}_1\right)_{jj}\right\}_{i,j=1}^n\mathbf{f}$$

which finally entails

$$\mathbf{f} = \left(\mathbf{I}_n - \frac{1}{n}\left\{\left(\bar{\mathbf{Q}}_2\boldsymbol{\Sigma}_i\bar{\mathbf{Q}}_1\right)_{jj}\right\}_{i,j=1}^n\right)^{-1}\frac{1}{n}\left\{\bar{\mathbf{Q}}_2\boldsymbol{\Sigma}_i\bar{\mathbf{Q}}_1\right\}_{i,j=1}^n\operatorname{diag}(\boldsymbol{\Xi}).$$

To complete the proof of Lemma 20, we need to show that $\operatorname{Var}\left(\frac{1}{n}\operatorname{tr}(\mathbf{Q}_1\boldsymbol{\Xi}\mathbf{Q}_2)\right)$ and $\operatorname{Var}\left(\frac{1}{n}\operatorname{tr}(\boldsymbol{\Sigma}_i\mathbf{Q}_2\boldsymbol{\Xi}\mathbf{Q}_1)\right)$ are asymptotically summable so that by the Borell Cantelli Lemma, $\frac{1}{n}\operatorname{tr}(\mathbf{Q}_1\boldsymbol{\Xi}\mathbf{Q}_2)$ and $\frac{1}{n}\operatorname{tr}(\boldsymbol{\Sigma}_i\mathbf{Q}_2\boldsymbol{\Xi}\mathbf{Q}_1)$ converge respectively almost surely to their expectations. Those follow directly by using Nash Poincare inequality (Lemma 26) similarly to what was done in the proof of Lemma 28.

References

Adamic, L. A. and Glance, N. (2005). The political blogosphere and the 2004 us election: divided they blog. In *Proceedings of the 3rd international workshop on Link discovery*, pages 36–43. ACM.

Ajanki, O., Erdos, L., and Krüger, T. (2015). Quadratic vector equations on complex upper half-plane. *arXiv preprint arXiv:1506.05095*.

Bai, Z.-D. and Silverstein, J. W. (1998). No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Annals of probability*, pages 316–345.

Baik, J., Ben Arous, G., and Péché, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Annals of Probability*, pages 1643–1697.

Baik, J. and Silverstein, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *Journal of Multivariate Analysis*, 97(6):1382–1408.

Benaych-Georges, F. and Nadakuditi, R. R. (2012). The singular values and vectors of low rank perturbations of large rectangular random matrices. *Journal of Multivariate Analysis*, 111:120–135.

Billingsley, P. (1995). Probability and measure. wiley series in probability and mathematical statistics.

Boucheron, S., Lugosi, G., and Massart, P. (2013). *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press.

Chapon, F., Couillet, R., Hachem, W., and Mestre, X. (2012). The outliers among the singular values of large rectangular random matrices with additive fixed rank deformation. *arXiv preprint arXiv:1207.0471*.

Chen, Y., Li, X., and Xu, J. (2015). Convexified modularity maximization for degree-corrected stochastic block models. *arXiv preprint arXiv:1512.08425*.

Chung, F. R. (1997). *Spectral graph theory*, volume 92. American Mathematical Soc.

Coja-Oghlan, A. and Lanka, A. (2009). Finding planted partitions in random graphs with general degree distributions. *SIAM Journal on Discrete Mathematics*, 23(4):1682–1714.

Couillet, R., Benaych-Georges, F., et al. (2016). Kernel spectral clustering of large dimensional data. *Electronic Journal of Statistics*, 10(1):1393–1454.

Decelle, A., Krzakala, F., Moore, C., and Zdeborová, L. (2011). Inference and phase transitions in the detection of modules in sparse networks. *Physical Review Letters*, 107(6):065701.

Duch, J. and Arenas, A. (2005). Community detection in complex networks using extremal optimization. *Physical review E*, 72(2):027104.

Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, 486(3):75–174.

Gao, C., Ma, Z., Zhang, A. Y., and Zhou, H. H. (2016). Community detection in degree-corrected block models. *arXiv preprint arXiv:1607.06993*.

Guimera, R., Sales-Pardo, M., and Amaral, L. A. N. (2004). Modularity from fluctuations in random graphs and complex networks. *Physical Review E*, 70(2):025101.

Gulikers, L., Lelarge, M., and Massoulié, L. (2015). A spectral method for community detection in moderately-sparse degree-corrected stochastic block models. *arXiv preprint arXiv:1506.08621*.

Hachem, W., Loubaton, P., Mestre, X., Najim, J., and Vallet, P. (2013). A subspace estimator for fixed rank perturbations of large random matrices. *Journal of Multivariate Analysis*, 114:427–447.

Hachem, W., Loubaton, P., Najim, J., et al. (2007). Deterministic equivalents for certain functionals of large random matrices. *The Annals of Applied Probability*, 17(3):875–930.

Hartigan, J. A. and Wong, M. A. (1979). Algorithm as 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1):100–108.

Hastings, M. B. (2006). Community detection as an inference problem. *Physical Review E*, 74(3):035102.

Jin, J. et al. (2015). Fast community detection by score. *The Annals of Statistics*, 43(1):57–89.

Karrer, B. and Newman, M. E. (2011). Stochastic blockmodels and community structure in networks. *Physical Review E*, 83(1):016107.

Krzakala, F., Moore, C., Mossel, E., Neeman, J., Sly, A., Zdeborová, L., and Zhang, P. (2013). Spectral redemption in clustering sparse networks. *Proceedings of the National Academy of Sciences*, 110(52):20935–20940.

Lei, J., Rinaldo, A., et al. (2015). Consistency of spectral clustering in stochastic block models. *The Annals of Statistics*, 43(1):215–237.

Lyzinski, V., Sussman, D. L., Tang, M., Athreya, A., Priebe, C. E., et al. (2014). Perfect clustering for stochastic blockmodel graphs via adjacency spectral embedding. *Electronic Journal of Statistics*, 8(2):2905–2922.

Nadakuditi, R. R. and Newman, M. E. (2012). Graph spectra and the detectability of community structure in networks. *Physical review letters*, 108(18):188701.

Newman, M. (2013). Spectral community detection in sparse networks. *arXiv preprint arXiv:1308.6494*.

Newman, M. (2016). Community detection in networks: Modularity optimization and maximum likelihood are equivalent. *arXiv preprint arXiv:1606.02319*.

Newman, M. E. (2004). Fast algorithm for detecting community structure in networks. *Physical review E*, 69(6):066133.

Newman, M. E. (2006a). Finding community structure in networks using the eigenvectors of matrices. *Physical review E*, 74(3):036104.

Newman, M. E. (2006b). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582.

Newman, M. E. and Leicht, E. A. (2007). Mixture models and exploratory analysis in networks. *Proceedings of the National Academy of Sciences*, 104(23):9564–9569.

Ng, A. Y., Jordan, M. I., Weiss, Y., et al. (2002). On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems*, 2:849–856.

Ng, S. K., Krishnan, T., and McLachlan, G. J. (2012). The em algorithm. In *Handbook of computational statistics*, pages 139–172. Springer.

Pastur, L. A., Shcherbina, M., and Shcherbina, M. (2011). *Eigenvalue distribution of large random matrices*, volume 171. American Mathematical Society Providence, RI.

Qin, T. and Rohe, K. (2013). Regularized spectral clustering under the degree-corrected stochastic blockmodel. In *Advances in Neural Information Processing Systems*, pages 3120–3128.

Saade, A., Krzakala, F., and Zdeborová, L. (2014). Spectral clustering of graphs with the bethe hessian. In *Advances in Neural Information Processing Systems*, pages 406–414.

Silverstein, J. W. and Bai, Z. (1995). On the empirical distribution of eigenvalues of a class of large dimensional random matrices. *Journal of Multivariate analysis*, 54(2):175–192.

Tiomoko Ali, H. and Couillet, R. (2016). Performance analysis of spectral community detection in realistic graph models. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP16)*.

Tiomoko Ali, H. and Couillet, R. (2017). Improved spectral community detection algorithm in large dense graphs. `https://github.com/hafizTiomoko/improved_spectral_community_detection`.

Zhao, Y., Levina, E., Zhu, J., et al. (2012). Consistency of community detection in networks under degree-corrected stochastic block models. *The Annals of Statistics*, 40(4):2266–2292.