

# Characterization of Random Matrix Eigenvectors for Stochastic Block Model

Arun Kadavankandy, Laura Cottatellucci, Konstantin Avrachenkov

► **To cite this version:**

Arun Kadavankandy, Laura Cottatellucci, Konstantin Avrachenkov. Characterization of Random Matrix Eigenvectors for Stochastic Block Model. 49th Asilomar Conference on Signals, Systems, and Computer, Nov 2015, Pacific Grove, CA, United States. 2015. <hal-01258664>

**HAL Id: hal-01258664**

**<https://hal.inria.fr/hal-01258664>**

Submitted on 19 Jan 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Characterization of Random Matrix Eigenvectors for Stochastic Block Model

Arun Kadavankandy

INRIA Sophia Antipolis Méditerranée  
2004 Route des Lucioles BP93  
06902 SOPHIA ANTIPOLIS cedex  
Email: arun.kadavankandy@inria.fr

Laura Cottatellucci

EURECOM  
Mobile Communications Department  
BP193, F-06560 Sophia Antipolis, France  
Email: cottatel@eurecom.fr

Konstantin Avrachenkov

INRIA Sophia Antipolis Méditerranée  
2004 Route des Lucioles BP93  
06902 SOPHIA ANTIPOLIS cedex  
Email: konstantin.avrachenkov@inria.fr

**Abstract**—The eigenvalue spectrum of the adjacency matrix of Stochastic Block Model (SBM) consists of two parts: a finite discrete set of dominant eigenvalues and a continuous bulk of eigenvalues. We characterize analytically the eigenvectors corresponding to the continuous part: the bulk eigenvectors. For symmetric SBM adjacency matrices, the eigenvectors are shown to satisfy two key properties. A modified spectral function of the eigenvalues, depending on the eigenvectors, converges to the eigenvalue spectrum. Its fluctuations around this limit converge to a Gaussian process different from a Brownian bridge. This latter fact disproves that the bulk eigenvectors are Haar distributed.

## I. INTRODUCTION

Large graphs have become an important area of research in recent times, owing to burgeoning interest in the analysis of Facebook™, Twitter™ and other social graphs. A host of interesting problems exists in this field such as those related to community detection and subgraph detection, which in turn find a lot of real-world applications in large social networks. Random graphs (RG) provide an effective and viable way of modeling these large networks. A random graph model proposed to model several observed properties of real internet graphs such as the small-world property, and the scale-free property. Random graphs and algorithms for RGs can be effectively analyzed through the spectral properties of the various random matrices used to represent these RGs, such as the distributions of eigenvalues and eigenvectors of the adjacency matrix or the Laplacian matrix. As an example, for the anomalous subgraph detection problem discussed in [13], and consisting in detecting the presence of a subgraph embedded in a random graph background, Miller et al. [13] propose an algorithm based on the  $L_1$  norm of the centered adjacency matrix eigenvectors.

In this work, we deal with a class of random graphs called the Stochastic Block Model (SBM). It is a class that is especially useful because it models networks with communities. It can be thought of as a generalization of the well-studied Erdős-Rényi random graphs. In an SBM graph with  $n$  nodes and  $M$  communities, each community has  $n/M$  nodes, and the probability of links between any pair of nodes within a given community is uniform and is  $p_i, 1 \leq i \leq M$ . The

probability of links between nodes belonging to different communities is  $p_0, p_i > p_0$ . In [3], the authors derive results about the limiting properties of the eigenvalue distribution of these matrices. It was shown that the eigenvalue spectrum has two important features - a continuous bulk of eigenvalues, and a discrete number of dominant eigenvalues that fall outside the bulk. The dominant eigenvectors are useful in establishing the consistency of the spectral clustering algorithm, which is employed in varied fields to detect underlying community structure in a graph. The analysis of these have spawned a string of works including [6], [8] and the references therein.

In this work, we look at the eigenvectors and eigenvalues of the centered adjacency matrix of SBM. To the best of our knowledge, this area is as yet completely unexplored. Most existing works deal with the eigenvectors of Wigner matrices made up of symmetric random variable entries. At best they can be used to describe the eigenvectors of the Erdős-Rényi graphs, which is a degenerate case of the SBM. The eigenvectors of Wigner matrices are studied in detail in [2] and [11], and the references therein. In the latter, in proving a universality property of the distribution of eigenvectors of hermitian random matrices, the authors show that the eigenvectors of a Wigner matrix are approximately Haar distributed. Their method is to show that a property that holds for orthonormal matrices drawn from a Haar distribution also holds approximately for the normalized eigenvectors of a Wigner matrix under some moment conditions. This partially extends the property of Haar distribution of eigenvectors known to hold for Gaussian Wigner Ensembles such as the Gaussian Unitary Ensemble (GUE) and the Gaussian Orthogonal Ensemble (GOE) [12], to Wigner matrices with general distributions under certain moment conditions.

In [5], by way of showing the convergence of a weighted distribution function of the eigenvalues, with weights depending on eigenvectors, the authors show that the eigenvectors are approximately Haar distributed for a covariance matrix. This was extended to Wigner matrices in [2]. They use familiar tools of random matrix theory such as the Stieltjes transform in their analyzes.

Our contribution is to study the properties analyzed in [2] in the case of centered SBM adjacency matrices. We consider  $Q(x, \mathbf{y})$ , a modified empirical spectral density function of

This work was partly funded by the French Government (National Research Agency, ANR) through the “Investments for the Future” Program reference #ANR-11-LABX-0031-01.

the eigenvalues, where the contribution of each eigenvalue is weighted by the magnitude of the projection of the corresponding eigenvector to an arbitrary, non-random unit vector  $\mathbf{y}$ . First we show that when the link probabilities within communities are different, i.e., the case of asymmetric SBM, the weighted spectral function  $Q(x, \mathbf{y})$  has different limits depending on the unit vector  $\mathbf{y}$ , and we determine the asymptotic limits. From this we conclude that the eigenvectors of the asymmetric SBM are not Haar distributed. In contrast, when the link probabilities within all the communities are the same, the modified empirical spectral distribution  $Q(x, \mathbf{y})$  has the same asymptotic limit as the empirical spectral distribution (e.s.d.) of eigenvalues, irrespective of the chosen unit vector. This is a necessary condition for Haar distribution of eigenvectors. We also show that for a specially chosen unit vector, the fluctuations around its mean form a Gaussian process. However, we also prove that this Gaussian process is not a Brownian bridge, disproving in this way that the bulk eigenvectors of the symmetric SBM are Haar distributed.

*Remark:* Due to space constraints, we state the results without proofs. The proofs will appear in an extended version of this paper.

## II. MODEL, NOTATION AND DEFINITIONS AND ASSUMPTIONS

Let us consider a graph with  $n$  nodes. The adjacency matrix  $\mathbf{A}$  of an undirected graph is an  $n \times n$  symmetric matrix whose  $(i, j)$  element  $A_{ij}$  is 1, if a link exists between node  $i$  and node  $j$ , else it is 0. For an SBM graph,  $A_{ij}$  is a Bernoulli variable with parameter  $p_1(p_2)$ , if the corresponding nodes  $i$  and  $j$  both belong to community 1, i.e.,  $\Omega_1$  (community 2, i.e.,  $\Omega_2$ ) and with parameter  $p_0$  if nodes  $i, j$  belong to different communities. Let  $\mathbf{a}_i$  be the  $i^{\text{th}}$  column of  $\mathbf{A}$ . Since the graph is undirected, or equivalently, the adjacency matrix is symmetric, the  $i^{\text{th}}$  row is  $\mathbf{a}_i^T$ , where the superscript  $T$  denotes transpose. Without loss of generality, we restrict ourselves to the case where the number of communities in the SBM is 2, the number of nodes is even and the nodes are evenly split between the two communities. Additionally, we assume that the nodes are indexed such that node  $i$  belongs to community 1 if  $1 \leq i \leq n/2$ , otherwise it belongs to the second community. Let us denote by  $\mathcal{B}(p_m)$  a Bernoulli probability distribution with parameter  $p_m, m = 0, 1, 2$ ; then,

$$\begin{cases} A_{ij} = A_{ji} \sim \mathcal{B}(p_m), & \text{if } i, j \in \Omega_m \\ A_{ij} = A_{ji} \sim \mathcal{B}(p_0), & \text{if } i \in \Omega_\ell \text{ and } j \in \Omega_m, \ell \neq m. \end{cases} \quad (1)$$

In the following, unless explicitly mentioned,  $p_1 \neq p_2$ . We consider graphs without self-loops, therefore,  $A_{ii} = 0$ . In this work we consider the centered and scaled adjacency matrix  $\tilde{\mathbf{A}} = \gamma(n)(\mathbf{A} - \bar{\mathbf{A}})$ , where  $\gamma(n)$  is a scaling factor meant to make sure the asymptotic eigenvalue distribution has finite support, and  $\bar{\mathbf{A}}$  is the mean of  $\mathbf{A}$  which can be expressed as

$$\bar{\mathbf{A}} = \mathbf{P} \otimes \mathbf{J}_{n/2} \quad (2)$$

where  $\mathbf{P}$  is a  $2 \times 2$  matrix given as:

$$\mathbf{P} = \begin{pmatrix} p_1 & p_0 \\ p_0 & p_2 \end{pmatrix}. \quad (3)$$

and  $\mathbf{J}_{n/2}$  is a  $n/2 \times n/2$  matrix of ones. Let  $\gamma(n) = \sqrt{np_1(1-p_1)}$ . It can be shown that this constant suffices to ensure that  $\lim_{n \rightarrow \infty} \|\tilde{\mathbf{A}}\|_2$  is finite [3]. We use  $\sigma_{ij}^2$  to denote the variance of  $\tilde{A}_{ij}$ . It is easy to see that  $\sigma_{ij}^2 = 1/n$ , for  $i, j$  belonging to community 1,

$$\sigma_{ij}^2 = \frac{p_2(1-p_2)}{n(p_1(1-p_1))}$$

for  $i, j$  in community 2, and

$$\sigma_{ij}^2 = \frac{p_0(1-p_0)}{n(p_1(1-p_1))}$$

otherwise. In this work we consider only dense graphs, i.e., the probabilities are constant functions of  $n$ .

For the real symmetric matrix  $\tilde{\mathbf{A}}$ , by the Spectral Decomposition Theorem we have:

$$\tilde{\mathbf{A}} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T,$$

where  $\mathbf{\Lambda}$  is a diagonal matrix of eigenvalues such that  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$  and  $\mathbf{U}$  is the matrix whose columns are the corresponding eigenvectors  $\mathbf{u}_i$ , such that  $\mathbf{U}\mathbf{U}^T = \mathbf{I}$ ,  $\mathbf{I}$  being the identity matrix.

Let  $\mathbf{e}_k$  denote a unit vector in  $\mathbb{R}^n$  whose  $k^{\text{th}}$  element is 1. We define  $\tilde{\mathbf{R}} \equiv \tilde{\mathbf{A}} - z\mathbf{I}$ . Let  $\tilde{\mathbf{A}}_k$  to denote the matrix obtained from  $\tilde{\mathbf{A}}$ , by setting the  $k^{\text{th}}$  row and column to 0, and define  $\mathbf{R}_k \equiv \tilde{\mathbf{A}}_k - z\mathbf{I}$ . Hence we have:  $\tilde{\mathbf{R}} = \tilde{\mathbf{A}} - z\mathbf{I} = \tilde{\mathbf{A}}_k + \tilde{\mathbf{a}}_k \mathbf{e}_k^T + \mathbf{e}_k \tilde{\mathbf{a}}_k^T - z\mathbf{I} = \tilde{\mathbf{R}}_k + \tilde{\mathbf{a}}_k \mathbf{e}_k^T + \mathbf{e}_k \tilde{\mathbf{a}}_k^T$ . We denote a complex variable by  $z$  such that  $z = u + \sqrt{-1}v, v > 0, u, v \in \mathbb{R}$ . Let  $\chi_{\{A\}}$  be the indicator function for the set  $A$ .

We define the empirical spectral distribution (e.s.d.) of  $\tilde{\mathbf{A}}$  as

$$F_{\tilde{\mathbf{A}}}^-(x) = \frac{1}{n} \sum_{k=1}^n \chi_{\{\lambda_k \leq x\}}, \quad (4)$$

and the Stieltjes transform of  $F_{\tilde{\mathbf{A}}}^-(x)$  for  $z \in \mathbb{C}, \Im z > 0$ , as

$$s_{\tilde{\mathbf{A}}}^-(z) = \int \frac{dF_{\tilde{\mathbf{A}}}^-(x)}{x-z} = \text{trace}(\tilde{\mathbf{A}} - z\mathbf{I})^{-1} = \text{trace}(\tilde{\mathbf{R}}^{-1})$$

In addition, we specify here  $s_\sigma(z)$ , the Stieltjes transform of the semicircle distribution [12] with parameter  $\sigma$  as:

$$s_\sigma(z) = \frac{-1}{z + \sigma^2 s_\sigma(z)} \quad (5)$$

## III. PRELIMINARY RESULTS ON THE SBM

We have the following result for  $F_{\tilde{\mathbf{A}}}^-(x)$ .

**LEMMA 1** [3] *Let  $\tilde{\mathbf{A}}$  be the normalized centered SBM adjacency matrix with  $\gamma(n) = (np_1(1-p_1))^{-1}$ . If for  $m = 0, 1, 2$ ,  $p_m$  are independent of  $n$ , then almost surely, the*

eigenvalue e.d.f. converges weakly to a distribution function whose Stieltjes transform is given by

$$s(z) = c_1(z) + c_2(z) \quad (6)$$

$c_1(z), c_2(z)$  being the unique solution to the system of equations:

$$\begin{aligned} c_1(z) &= \frac{-1/2}{z + \varsigma_1 c_1(z) + \varsigma_0 c_2(z)}, \\ c_2(z) &= \frac{-1/2}{z + \varsigma_0 c_1(z) + \varsigma_2 c_2(z)}, \end{aligned}$$

that satisfies the conditions

$$\Im(c_i(z))\Im(z) > 0 \text{ for } \Im z > 0, \quad (7)$$

for each  $i = 1, 2$ , and  $\varsigma_i = \frac{p_i(1-p_i)}{p_1(1-p_1)}$ ,  $i = 1, 2$ .

#### IV. ASYMPTOTIC RESULTS ON EIGENVECTORS OF SBM

In this section we analyze the asymptotic properties of the eigenvectors of  $\tilde{\mathbf{A}}$ . Following the ideas in [4] we consider the following spectral function

$$Q(x, \mathbf{y}) = \sum_{i=1}^n |\mathbf{u}_i^T \mathbf{y}|^2 \chi_{\{\lambda_i \leq x\}}, \quad x \in (-\infty, \infty), \quad (8)$$

where  $\mathbf{y} \in \mathbb{R}^n$  is an arbitrary deterministic unit vector. Notice that  $Q(x, \mathbf{y}) \geq 0, \forall x$ , and  $\lim_{x \rightarrow -\infty} Q(x, \mathbf{y}) = 0$ , and  $\lim_{x \rightarrow \infty} Q(x, \mathbf{y}) = 1$ . Therefore  $Q(x, \mathbf{y})$  can be thought of as a cumulative distribution function (cdf) of the eigenvalues, with weights determined by the corresponding eigenvectors. In [5], the authors study the above function and observe that if the eigenvectors are Haar-distributed  $Q(x, \mathbf{y})$  satisfies the following two properties:

- **Property I**

$$\lim_{n \rightarrow \infty} |Q(x, \mathbf{y}) - F_{\tilde{\mathbf{A}}}(x)| = 0$$

This property has to be satisfied if  $\mathbf{u}_i$  are uniformly distributed on the unit sphere in  $\mathbb{R}^n$ . Notice that this implies  $|\mathbf{u}_i^T \mathbf{y}|^2 \approx \frac{1}{n}, \forall i$ .

- **Property II**  $\sqrt{\frac{n}{2}}(Q(x, \mathbf{y}) - F_{\tilde{\mathbf{A}}}(x))$  converges to a Brownian Bridge [5].

Indeed, a vector uniformly distributed on the unit sphere in  $\mathbb{R}^n$  is equivalent in distribution to a vector  $\mathbf{z} \in \mathbb{R}^n$  with independent standard Gaussian components normalized such that  $\|\mathbf{z}\|_2 = 1$ .

Instead of analyzing directly  $Q(x, \mathbf{y})$ , we can analyze its Stieltjes transform given as [5]

$$s_Q(z, \mathbf{y}) = \mathbf{y}^T (\tilde{\mathbf{A}} - z\mathbf{I})^{-1} \mathbf{y}. \quad (9)$$

By the Stieltjes inversion formula, the convergence of the Stieltjes transform of a sequence of functions, implies the convergence of the original sequence [12].

#### A. Asymptotic Limit of $Q(x, \mathbf{y})$ for general SBM

In this section we present a result about  $Q(x, \mathbf{y})$  for the special case when  $\mathbf{y} = \mathbf{e}_i$ . The analysis adopts the same method as the one followed in [1]. In this case  $s_Q(z, \mathbf{y}) = (\tilde{\mathbf{A}} - z\mathbf{I})_{ii}^{-1}$ , the diagonal component of the resolvent. Let us denote  $\Psi = (\tilde{\mathbf{A}} - z\mathbf{I})^{-1}$  and the diagonal component as  $\psi_{ii}$ .

It is easy to verify that, for the SBM, the following holds true:

$$\forall i \quad \exists c \text{ such that } \sum_j \sigma_{ij}^2 \leq c. \quad (10)$$

In addition, for  $1 \leq i, j \leq n/2$ ,

$$\sum_k \sigma_{jk}^2 = \sum_k \sigma_{ik}^2, \quad (11)$$

and similarly for  $n/2 + 1 \leq i, j \leq n$ .

We have the following result.

**PROPOSITION 1** For a centered adjacency matrix  $\tilde{\mathbf{A}}$  of SBM with constant probabilities  $p_0, p_1, p_2$ , independent of  $n$ , and  $\mathbf{y} = \mathbf{e}_i, s_Q(z, \mathbf{e}_i)$ , the Stieltjes transform of the spectral function  $Q(x, \mathbf{y})$ , converges in probability as follows.

$$\lim_{n \rightarrow \infty} s_Q(z, \mathbf{e}_i) = \begin{cases} d_1, & \text{if } i \leq \frac{n}{2} \\ d_2, & \text{if } \frac{n}{2} + 1 \leq n, \end{cases} \quad (12)$$

where  $d_1, d_2$  are unique solutions to the following set of fixed point equations

$$\begin{aligned} d_1 &= \frac{1}{-z - \frac{d_1}{2}\varsigma_1 - \frac{d_2}{2}\varsigma_0}, \\ d_2 &= \frac{1}{-z - \frac{d_1}{2}\varsigma_0 - \frac{d_2}{2}\varsigma_2}, \end{aligned}$$

and  $\varsigma_i$  are defined as in Lemma 1.

From the above result, we see that the eigenvectors of a general SBM are not Haar distributed, because the asymptotic limit is a function of the vector  $\mathbf{y}$ . In fact,  $\mathbf{y} = \mathbf{e}_i$ , the asymptotic limit of the spectral function is different when  $i \leq n/2$  and  $i > n/2$ .

We make the following observation as a corollary to Lemma 1 and Proposition 1.

**COROLLARY 1** For an asymmetric SBM with  $p_1 \neq p_2$ ,

$$\lim_{n \rightarrow \infty} \frac{1}{2} (s_Q(z, \mathbf{e}_{k_1}) + s_Q(z, \mathbf{e}_{k_2})) \rightarrow s(z),$$

in probability.

In the following Proposition 2 we present a result for symmetric SBM that holds for any unit vector  $\mathbf{y}$ . More specifically, we show the convergence of the Stieltjes transform  $s_Q(z, \mathbf{y})$  to  $s_\sigma(z)$ , the Stieltjes transform of the semicircle distribution. Ancillary to this proposition is the following lemma.

LEMMA 2 For the centered SBM adjacency matrix  $\tilde{\mathbf{A}}$ , with probabilities  $p_0, p_1, p_2$ ,

$$\begin{aligned} \max_{1 \leq k \leq n/2} E|\alpha_k + 2zc_1(z)|^4 &\rightarrow 0, \\ \max_{n/2 \leq k \leq n} E|\alpha_k + 2zc_2(z)|^4 &\rightarrow 0; \end{aligned}$$

where

$$\alpha_k = \frac{1}{1 + z^{-1} \tilde{\mathbf{a}}_k^T \tilde{\mathbf{R}}_k^{-1} \tilde{\mathbf{a}}_k}. \quad (13)$$

For the symmetric case, we recall that  $c_1(z) = c_2(z) = s_\sigma(z)/2$ . Also when  $p_1 = p_2$ ,

$$\max_{1 \leq k \leq n} |\omega_k + zs_\sigma(z)| \rightarrow 0.$$

By invoking the above lemma, we can prove the following proposition.

PROPOSITION 2 (Symmetric SBM) Let us consider the centered adjacency matrix  $\tilde{\mathbf{A}}$  of SBM, with  $p_1 = p_2$ . For any unit vector  $\mathbf{y}$ , the spectral function  $Q(x, \mathbf{y})$  converges to the cdf of the semicircle law.

### B. Gaussianity of the fluctuations

In this section we focus on the symmetric SBM and verify whether and if not, to what extent Property II is satisfied. The convergence of the process  $\sqrt{n}(\mathbf{y}^T \tilde{\mathbf{R}}^{-1} \mathbf{y} - s_\sigma(z))$  to a Brownian Bridge in distribution is usually shown in two steps:

- The process  $Y_n(z) = \sqrt{n}(\mathbf{y}^T \tilde{\mathbf{R}}^{-1} \mathbf{y} - \mathbf{y}^T E(\tilde{\mathbf{R}})^{-1} \mathbf{y})$  is shown to converge to a Gaussian process of mean zero in distribution;
- $\sqrt{n}(\mathbf{y}^T E(\tilde{\mathbf{R}})^{-1} \mathbf{y} - s_\sigma(z)) \rightarrow 0$ .

In the following proposition we show the convergence of  $\sqrt{n}(\mathbf{y}^T \tilde{\mathbf{R}}^{-1} \mathbf{y} - \mathbf{y}^T E(\tilde{\mathbf{R}})^{-1} \mathbf{y})$  to a Gaussian process for the special case of  $\mathbf{y} = \frac{1}{\sqrt{n}} \mathbf{1}$ , where  $\mathbf{1}$  is a vector of all ones. Note that this should hold for any unit  $\mathbf{y}$ , in order to satisfy Property II.

PROPOSITION 3 For  $\mathbf{y} = \frac{1}{\sqrt{n}} \mathbf{1}$ , and  $p_1 = p_2$ , the process  $Y_n(z)$  converges to a Gaussian process in distribution.

The proof of the above proposition uses Martingale Convergence theorem [14].

In the following proposition we show that even though the process  $Y_n(z)$  is Gaussian, its mean does not converge to  $s_\sigma(z)$ , but instead, there is a bias term which is a function of  $z$ . This shows that the process does not converge to a Brownian bridge, and consequently the eigenvectors of the adjacency matrix of the symmetric SBM are not Haar distributed, even though they display some of the required properties.

PROPOSITION 4 For the centered adjacency matrix  $\tilde{\mathbf{A}}$  of a symmetric SBM with probabilities  $p_0, p_1$ , the fluctuation of

the mean,  $\sqrt{n}(\mathbf{y}^T E \tilde{\mathbf{R}}^{-1} \mathbf{y} - s_\sigma(z))$  does not converge to 0, but it is bounded as follows

$$\left| \sqrt{n}(\mathbf{y}^T E \tilde{\mathbf{R}}^{-1} \mathbf{y} - s_\sigma(z)) \right| \leq C |s_\sigma(z)| + o(1),$$

where  $C$  is a constant that depends on the third moment of the elements of  $\tilde{\mathbf{A}}$ .

## V. SIMULATIONS

In this section we perform simulations to corroborate our theoretical results. First, we consider a realization of a random symmetric SBM of size  $n = 10^3$ , with  $p_0 = 10^{-2}$  and  $p_1 = p_2 = 10^{-1}$ . In Figure 1, we set  $\mathbf{y} = \mathbf{e}_1$  and we plot two histograms of the eigenvalues: the first one denoted by  $dF_{\tilde{\mathbf{A}}}^-(x)$  is obtained giving a unit weight to each eigenvalue falling in a histogram bin  $[x, x + \Delta)$ ; the second one, denoted by  $dQ(x, \mathbf{y})$  is obtained by giving the weight  $|\mathbf{u}_i^T \mathbf{y}|^2$  to an eigenvalue  $\lambda_i \in [x, x + \Delta)$ . In other words, let  $\lambda_i, \lambda_{i+1}, \dots, \lambda_j$  be the  $j - i + 1$  eigenvalues in the bin  $[x, x + \Delta)$ . Then  $dF_{\tilde{\mathbf{A}}}^-(x) = \frac{j-i+1}{n}$ , and  $dQ(x, \mathbf{y}) = \sum_{k=i}^j |\mathbf{u}_k^T \mathbf{y}|^2$ . From Figure 1, it can be seen that both  $dF_{\tilde{\mathbf{A}}}^-(x)$  and  $dQ(x, \mathbf{y})$  approximate the semicircle law very well, consistent with Lemma 1 and Proposition 2, respectively. In Figure 2 we repeat the same experiment as in Figure 1, but for a slightly different setting. In this case  $n = 10^4$ , and  $\mathbf{y} = \frac{1}{\sqrt{n}} \mathbf{1}$ . Although the size of the matrix is an order of magnitude higher, the histogram  $dQ(x, \mathbf{y})$  approximate the semicircle law quite roughly, suggesting a much slower convergence of  $dQ(x, \mathbf{y})$  compared to the case where  $\mathbf{y} = \mathbf{e}_i$ .

Next, we consider an asymmetric SBM with  $n = 10^3$ ,  $p_0 = 10^{-2}$ ,  $p_1 = 0.1$ , and  $p_2 = 0.05$ . In Figure 3, we plot  $dF_{\tilde{\mathbf{A}}}^-(x)$  and  $\frac{1}{2}(dQ(x, \mathbf{e}_1) + dQ(x, \mathbf{e}_{\frac{n}{2}+1}))$ . They match very well, consistently with Corollary 1.

Finally, we aim at validating our theoretical results on the Gaussianity of the eigenvector fluctuations. Then, we generate 4000 independent realizations of a symmetric SBM centered adjacency matrix  $\tilde{\mathbf{A}}$  with  $p_0 = 0.01$ , and  $p_1 = p_2 = 0.1$ . Using these realizations, we calculate the empirical cdf of  $\sqrt{n}dQ(x, \frac{1}{\sqrt{n}} \mathbf{1})$ , for  $x_1 = -1.0538$  and  $x_2 = 1.0489$ . In both Figure 4 and Figure 5, the solid red line show the cdf of the centered variables,  $\sqrt{n}(dQ(x_i, \frac{1}{\sqrt{n}} \mathbf{1}) - E dQ(x, \frac{1}{\sqrt{n}} \mathbf{1}))$ ,  $i = 1, 2$ . We compare them to the cdfs of a zero mean gaussian variable with variance properly adapted to the empirical variance of our processes. These are plotted with solid line with crosses as marker. The perfect match between the solid line and the solid lines with markers confirms the Gaussianity of the perturbations of  $Q(x, \frac{1}{\sqrt{n}} \mathbf{1})$ . The dashed lines in Figures 4 and 5, correspond to the function  $\sqrt{n}(dQ(x, \frac{1}{\sqrt{n}} \mathbf{1}) - dF_{\tilde{\mathbf{A}}}^-(x))$ . The shift of these lines w.r.t the solid lines confirms the presence of the bias pointed out in Proposition 4.

## VI. CONCLUSIONS AND FUTURE WORK

In this paper, we analyzed the bulk eigenvectors of the centered adjacency matrix of SBM graphs. Following a classical approach we studied the spectral function  $Q(x, \mathbf{y})$ , which depends on the eigenvectors and its fluctuations around the e.s.d.

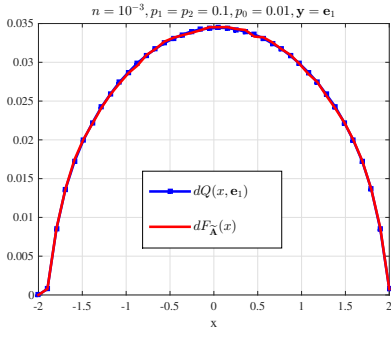


Fig. 1. Asymptotic spectral function for symmetric SBM,  $\mathbf{y} = \mathbf{e}_1$ .

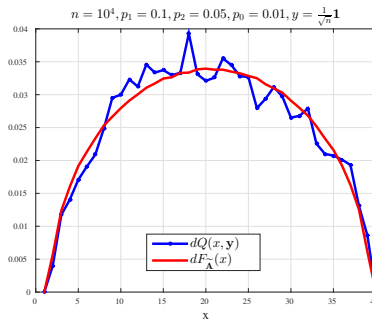


Fig. 2. Asymptotic spectral function for symmetric SBM,  $\mathbf{y} = \frac{1}{\sqrt{n}}\mathbf{1}$ .

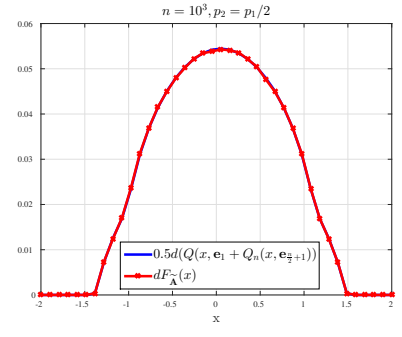


Fig. 3. Asymptotic spectral function for asymmetric SBM,  $\mathbf{y} = \mathbf{e}_1$ .

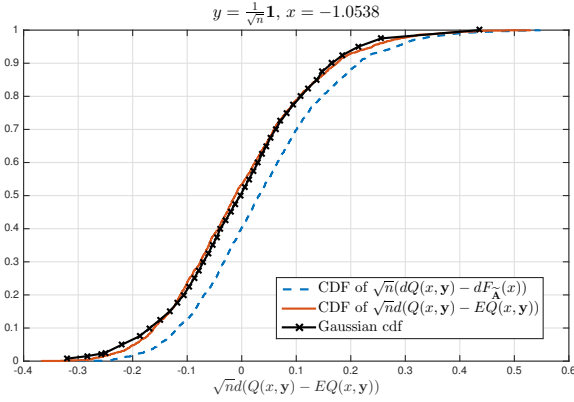


Fig. 4. Gaussianity of the fluctuations of  $Q(x, \mathbf{y})$

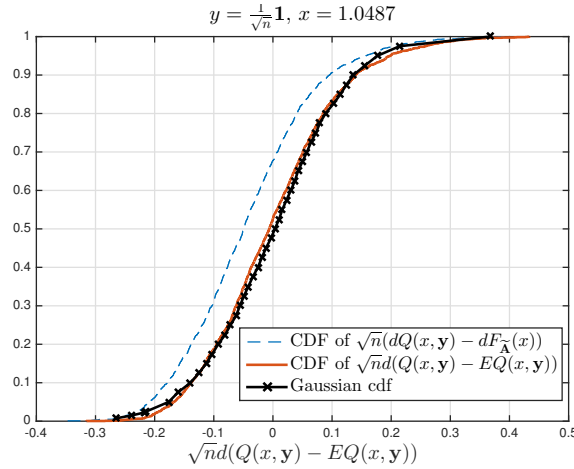


Fig. 5. Gaussianity of the fluctuations of  $Q(x, \mathbf{y})$

$F_{\tilde{A}}(x)$ . We show that for the centered adjacency matrix  $\tilde{\mathbf{A}}$  of symmetric SBM,  $Q(x, \mathbf{y})$  converges almost surely to  $F_{\tilde{A}}(x)$  for any  $\mathbf{y}$ . This suggests that  $|\mathbf{u}_i^T \mathbf{y}| \approx \frac{1}{\sqrt{n}}$ , for any  $\mathbf{y}$ . Additionally we show that the fluctuations  $\sqrt{n}(Q(x, \mathbf{y}) - F_{\tilde{A}}(x))$  converge in distribution to a Gaussian process. However, this process has non-zero mean, and hence is not a Brownian bridge. Therefore, the eigenvectors of the centered SBM

adjacency matrix violates a property required for them to be Haar distributed.

We also consider the eigenvectors of the centered adjacency matrix of the asymmetric SBM, and show that the asymptotic limit of the spectral function  $Q(x, \mathbf{y})$  depends on  $\mathbf{y}$ , as opposed to a matrix with Haar distributed eigenvectors.

In the future, we aim to extend the above result on the fluctuations of  $Q(x, \mathbf{y})$  of centered adjacency matrix  $\tilde{\mathbf{A}}$  of the symmetric SBM to any general  $\mathbf{y}$ .

## REFERENCES

- [1] V. Girko, W. Kirsch, and A. Kutzelnigg, "A necessary and sufficient conditions for the semicircle law. Random Operators and Stochastic Equations," 2(2), 195-202, 1994.
- [2] Z. D. Bai, and G. M. Pan, "Limiting Behavior of Eigenvectors of Large Wigner Matrices," Journal of Statistical Physics, 146(3), 519-549, 2012.
- [3] K. Avrachenkov, L. Cottatellucci, and A. Kadavankandy, "Spectral Properties of Random Matrices for Stochastic Block Model," invited to Proc. of International Workshop on Physics-Inspired Paradigms in Wireless Communications and Networks (PHYSCOMNET) at WiOPT, 29 May 2015.
- [4] J.W. Silverstein, "Weak convergence of random functions defined by the eigenvectors of sample covariance matrices," Ann. Probab. 18, 1990
- [5] Z. D. Bai, B. Q. Miao, and G. M. Pan, "On asymptotics of eigenvectors of large sample covariance matrix," Annals of Probability, 35(4), 1532-1572, 2007
- [6] K. Rohe, S. Chatterjee, and B. Yu, "Spectral clustering and the high-dimensional stochastic blockmodel," Annals of Statistics, 39(4), 1878-1915, 2011
- [7] D. L. Sussman, M. Tang, D. E. Fishkind, and C. E. Priebe, "A consistent dot product embedding for stochastic blockmodel graphs," 17, 2011
- [8] D. E. Fishkind, D.L. Sussman, M. Tang, J.T. Vogelstein and C. E. Priebe, "Consistent adjacency-spectral partitioning for the stochastic block model when the model parameters are unknown," 26. Methodology; Spectral Theory, 2012
- [9] A. Athreya, V. Lyzinski, D.J. Marchette, C.E. Priebe, D.L. Sussman and M. Tang, "A central limit theorem for scaled eigenvectors of random dot product graphs," (1983), 1-15, 2013
- [10] Z. D. Bai, and J. W. Silverstein, "No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices," Annals of probability : 316-345, 1998
- [11] T. Tao, and V. Van, "Random matrices: Universal properties of eigenvectors," Random Matrices: Theory and Applications 1.01: 1150001, 2012
- [12] G.W. Anderson, A. Guionnet, and O. Zeitouni, "An introduction to random matrices," No. 118. Cambridge University Press, 2010.
- [13] B.A. Miller, M.S. Beard, P.J. Wolfe and N.T. Bliss, "A spectral framework for anomalous subgraph detection," Tran. Signal Processing, Aug. 2015.
- [14] P. Billingsley, "Probability and measure," John Wiley & Sons, 2008.
- [15] P. Billingsley, "Convergence of probability measures," John Wiley & Sons, 2013.