

## Spectral partitioning in equitable graphs

Paolo Barucca

*Department of Banking and Finance, University of Zurich, Zurich, Switzerland  
and London Institute for Mathematical Sciences, London W1K 2XF, United Kingdom*

(Received 24 October 2016; revised manuscript received 24 May 2017; published 30 June 2017)

Graph partitioning problems emerge in a wide variety of complex systems, ranging from biology to finance, but can be rigorously analyzed and solved only for a few graph ensembles. Here, an ensemble of equitable graphs, i.e., random graphs with a block-regular structure, is studied, for which analytical results can be obtained. In particular, the spectral density of this ensemble is computed exactly for a modular and bipartite structure. Kesten-McKay's law for random regular graphs is found analytically to apply also for modular and bipartite structures when blocks are homogeneous. An exact solution to graph partitioning for two equal-sized communities is proposed and verified numerically, and a conjecture on the absence of an efficient recovery detectability transition in equitable graphs is suggested. A final discussion summarizes results and outlines their relevance for the solution of graph partitioning problems in other graph ensembles, in particular for the study of detectability thresholds and resolution limits in stochastic block models.

DOI: [10.1103/PhysRevE.95.062310](https://doi.org/10.1103/PhysRevE.95.062310)

### I. INTRODUCTION

The recent developments of network theory driven by the increasing number of applications in biology, ecology, social systems, economics, and finance [1,2] have stimulated theoretical research in graph theory. In particular, the need to establish the statistical significance of various network metrics and properties in real systems has ignited new results in statistical inference [3], spectral theory of random graphs [4–9], and ensembles of exponential random graphs [10].

In most networks, elements are divided into separate groups, and their behavior will often depend on this division. Finding an optimal partition then allows us to better understand the mesoscopic dynamics of the system and to obtain a more efficient reduced representation in terms of interacting groups. As a consequence, community detection has become a pivotal topic in network science. Stochastic block models (SBM) [11,12] have been introduced to understand, model, and analyze communities and have allowed theoretical insight into the performance and limitations of graph partitioning algorithms. Recently, it has been shown that recovery of communities in SBM displays a detectability transition in the sparse regime; i.e., when the edges are few, communities can be too weak to be identifiable [3,13]. Equitable graphs, the family of graph ensembles analyzed in this work, represent block-regular counterparts of the long-studied SBM: In SBM, edges are drawn independently with a probability depending on the assignment of the two terminal nodes, which results in a Poisson distribution of the number of edges between a pair of groups; in equitable graphs, the number of intrablock and interblock edges are fixed for each node and the graph is the result of a random matching between such edges. This class of random graph models has been analyzed in Ref. [14] in a dense approximation, in Ref. [15] under the name equitable graphs, and under the name microcanonical stochastic block model in Ref. [16]. In Ref. [17], it has been shown that equitable graphs with two equally sized communities have a unique partition almost surely in the large size limit, that an efficient algorithm can be derived in a large region of the ensemble's parameters, and that full recovery of communities can be

obtained starting from a group assignment with an extensive overlap with the original partition. In this paper, spectral theory of random graphs is used to disentangle noise and signal in equitable graphs, and an efficient algorithm for full recovery of communities is proposed. Following the derivation in Ref. [4], a finite set of nonlinear equations for the spectral density is obtained, analogously to Ref. [15]. The solution is found to obey the expected Kesten-McKay's law for regular graphs, allowing us to analytically predict the failure of naive spectral partitioning based on the second eigenvector of the adjacency matrix.

The paper is organized as follows: In Sec. II, equitable graphs are defined and the inference problem is introduced. In Sec. III, a brief introduction of the cavity approach to the spectral density is provided and the general expression for the cavity variances of equitable graphs is derived. In Sec. IV, the expression is solved for modular and bipartite two-community structures and its consequences on spectral clustering are outlined. A general methodology for the block structure inference in equitable graphs based on eigenvectors's extendedness is introduced to overcome the limitation of naive spectral partitioning. Numerical evidence of all results is presented.

Finally, in Sec. V, the relevance of the results with respect to open questions in theory of random graphs is discussed, and possible directions of research both for analytical results in spectral theory and for spectral methodologies for graph partitioning are outlined.

### II. EQUITABLE RANDOM GRAPHS

An ensemble of equitable graphs is defined by a set of vertices  $V$ , a partition  $\mathbf{B} = \{B_a\}_{a=1}^m$  dividing  $V$  in  $m$  nonoverlapping sets of vertices, also called blocks, and a connectivity matrix  $\mathbf{c}$ , a  $m \times m$  matrix of non-negative integer numbers [16]. For the sake of simplicity in the following, I will refer to block  $B_a$  with its corresponding integer index  $a$ . For later use, I also introduce for all nodes the assignments  $g_i$ , such that for each node  $i$  holds  $i \in B_{g_i}$ .

Each graph  $G = (V, E)$  of a random regular block model must satisfy the constraints

$$\forall B_a, B_b \in \mathbf{B} \forall i \in B_a \quad |\{(i, j) \in E \mid j \in B_b\}| = c_{ab}; \quad (1)$$

i.e., the total number of edges of node  $i$  in block  $B_a$  with a vertex in  $B_b$  equals  $c_{ab}$ , for every vertex  $i$  and every pair of blocks  $B_a$  and  $B_b$ . Equation (1) means that all nodes in a given block share the same connection pattern, i.e., number of links with each other block. This condition is stronger than a simple regularity within blocks, i.e., where all nodes in a block only share the same total number of links. In the case of blocks of different sizes,  $|B_a| = N_a$  such that  $\sum_{a=1}^m N_a = |V|$ , then, for the system to have solution, the connectivity matrix  $\mathbf{c}$  and block sizes must obey the relations

$$\forall B_a, B_b \in \mathbf{B} \quad N_a c_{ab} = N_b c_{ba}; \quad (2)$$

i.e., the total number of edges between blocks  $a$  and  $b$  must be uniquely defined.

All graphs satisfying (1) have equal probability in the ensemble.

If I introduce the block degrees  $k_{i \rightarrow a} = |\{(i, j) \in E \mid j \in B_a\}|$ , (1) can be reformulated as follows: The vector of block degrees of each node in a given block equals the row of the connectivity matrix corresponding to the block index, i.e.,  $\forall i \in B_a \quad k_{i \rightarrow B_b} = c_{ab}$ .

Both stochastic block models and equitable graphs are based on an analogous set of parameters, i.e., block assignments and connectivity matrix; nevertheless, there is no trivial mapping between stochastic block models, which are defined via link probabilities and their regular counterpart, which are defined via Eq. (2). A useful analogy to make sense of their relationship is the following: Stochastic block models [11] correspond to random regular block models as the Erdos-Renyi random graphs correspond to the  $k$ -regular random graphs, in the sense that in both cases the randomness which is eliminated from the ensemble is the one given by the (block) degree distribution.

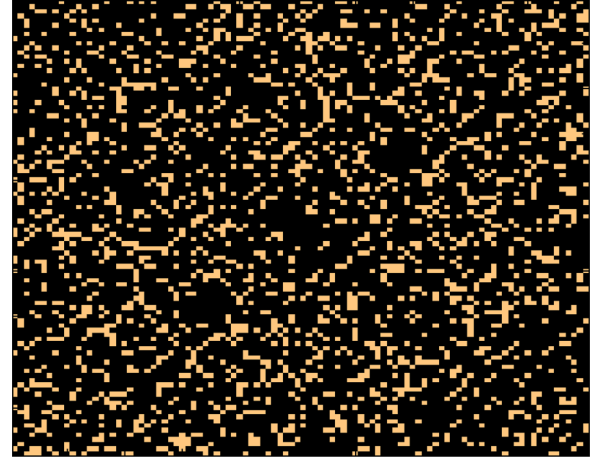
A different ensemble of random graphs with a block structure is the regular stochastic block models, studied in Refs. [3,18], where the probability measure is the same as in stochastic block models but a regularity constraint is imposed to all nodes.

Moreover, the form of the constraints in (2) allow edges to be drawn independently for each pair of blocks, and, for the case of blocks of the same size, it is possible to sample equitable graphs simply by assembling regular graphs: Between each pair of blocks the edges are drawn according to a  $k$ -regular graph, where the value of  $k$  equals the corresponding element of the connectivity matrix, and then the total set of edges is given by the union of the sets of edges for each of the  $m$  regular graphs and  $m(m-1)$  biregular graphs.

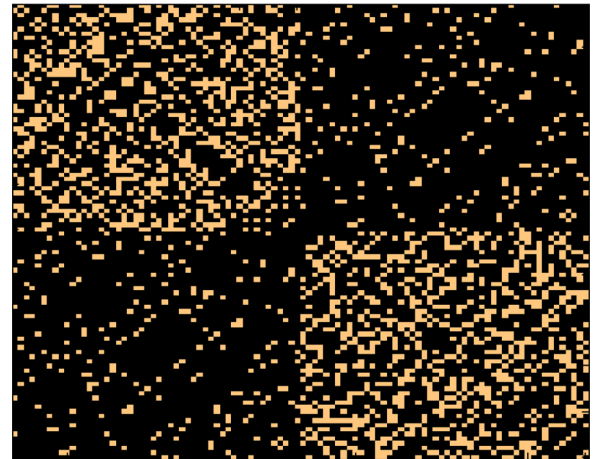
In the latter, the focus will be entirely on the representation of  $G$  in terms of its adjacency matrix  $A = (a_{ij})_{i,j=1}^N$ , where

$$a_{ij} = \begin{cases} 1, & \text{if } (i, j) \in E \\ 0, & \text{otherwise} \end{cases},$$

which allows computation of graph properties in algebraic form and can also be used to visualize the inference problem associated with the graph [Figs. 1(a) and 1(b)]: When



(a)



(b)

FIG. 1. (a) Adjacency matrix of a random regular block model graph with a community structure where no block structure seems to be present even though the connectivity matrix reads,  $\mathbf{c} = [16, 4; 4, 16]$ . (b) Same graph but rows and columns of the adjacency matrix are sorted according to the block structure, which becomes evident.

parameters are unknown, there is no *a priori* criterion to sort indices and the nonzero elements of the adjacency matrix do not display any specific block structure. Once the parameters are known, rows and columns can be sorted according to the block indices and the structure arises in a clear manner.

### A. The inference problem

Given an equitable graph  $G$ , the inference problem consists in reconstructing the parameters, i.e., the partition  $B$  and the connectivity matrix  $\mathbf{c}$ , that generated the graph.

I study inference on this ensemble of random graphs because it allows analysis of the performance of different algorithms in the absence of the noise coming from degree heterogeneity. In SBM, there exists a sharp transition in the assortivity parameter, first conjectured in Ref. [13] and later proved rigorously in a series of works that demonstrate both that asymptotically (i) below such threshold, recovery is information theoretically impossible [19] while (ii) above,

an efficient algorithm finds a partition with positive overlap with the original one [20,21]. In this work, it is shown that the regularity condition, as also found in Ref. [17], substantially changes the detectability properties of the ensemble.

### III. SPECTRAL THEORY

In this section, the spectral properties of random regular block model graphs with blocks of same size are investigated. Analytical results are presented, both for the discrete and the continuous part of the spectrum.

#### A. Discrete part: The signal

Here it is shown how isolated eigenvectors of the adjacency matrix entail exact information on the block structure of equitable graphs. Starting from the secular equation

$$\sum_{j=1}^N a_{ij} u_j = \lambda u_i, \quad (3)$$

an ansatz of *block symmetry* can be made such that nodes in the same block share the same eigencomponent, i.e., for all  $i$  is hypothesized that  $u_i = u_{g_i}$ . Since the number of neighbors between different groups is fixed, it follows that

$$\sum_{b=1}^m c_{ab} u_b = \lambda u_a, \quad (4)$$

which yields the useful conclusion, also pointed out in Refs. [15,17] that each block-symmetric eigenvector of the adjacency matrix corresponds to an eigenvector of the connectivity matrix  $\mathbf{c}$ , and vice versa. These eigenvectors correspond to a finite set of nondensely distributed, at most finitely degenerate eigenvalues. Generally, they can be positioned everywhere in the spectrum and when the block structure is particularly weak they will lie within the bulk of the spectrum. I will refer to them as the community eigenvectors.

#### B. Continuous part: The noise

In this paragraph, statistical physics techniques are used to compute the bulk of the spectrum of the adjacency matrices of equitable graphs. The derivation is entirely equivalent to the one already found in Ref. [15]; here I simply report it in terms of cavity variances [4], for the reader's convenience. Given an ensemble of  $N \times N$  symmetric matrices, the set of eigenvalues of a given adjacency matrix  $A$  is denoted by  $\{\lambda_i^A\}_{i=1}^N$ . The corresponding empirical spectral density is defined as

$$\rho(\lambda; A) = \frac{1}{N} \sum_{i=1}^N \delta(\lambda - \lambda_i^A), \quad (5)$$

which satisfies the identity [22]

$$\rho(\lambda; A) = \frac{2}{\pi} \lim_{\epsilon \rightarrow 0^+} \frac{1}{N} \text{Im} \left[ \frac{\partial}{\partial z} \ln \mathcal{Z}(z; A) \right]_{z=\lambda-i\epsilon}, \quad (6)$$

where  $\text{Im}[z]$  denotes the imaginary part of  $z$  and where  $\ln \mathcal{Z}(z; A)$  is obtained via Gaussian integrals as in Ref. [22],

i.e.,

$$\mathcal{Z}(z; A) = \int \left[ \prod_{i=1}^N \frac{dx_i}{\sqrt{(2\pi)}} e^{-H(x; z, A)} \right] \quad (7)$$

with  $H(x; z, A) = \frac{z}{2} \sum_i x_i^2 - \frac{1}{2} \sum_{i,j} A_{ij} x_i x_j$ . Such formulation yields an expression for the spectral density of any graph of the ensemble in terms of the variances of the Gaussian variables introduced in (7),

$$\rho(\lambda; A) = \frac{1}{\pi} \lim_{\epsilon \rightarrow 0^+} \frac{1}{N} \text{Im} \left[ \sum_i \langle x_i^2 \rangle_z \right]_{z=\lambda-i\epsilon}. \quad (8)$$

In principle, computing variances in (8) is not easier than diagonalizing the adjacency matrix but for sparse graphs an approximation method has been proposed that holds exactly in the large  $N$  limit, the cavity method [8,23].

In the cavity method, conditional probability distributions are introduced for each node and are parametrized by specific variables, i.e., the cavity variances  $\Delta_i^{(j)}$ , each representing the variance of  $x_i$  if its neighbor  $j$  is not taken into account. With such approximation, the following set of self-consistent equations can be derived [4]:

$$\Delta_i^{(j)}(z) = \frac{1}{z - \sum_{l \in \partial i \setminus j} A_{il}^2 \Delta_l^{(i)}(z)}, \quad (9)$$

where  $\partial i$  is the set of neighbor of node  $i$ , i.e.,  $\partial i = \{e \in E | i \in e\}$ . From cavity variances it is possible to compute node variances via the equations

$$\Delta_i(z) = \frac{1}{z - \sum_{l \in \partial i} A_{il}^2 \Delta_l^{(i)}(z)}, \quad (10)$$

which lead to compute the spectral density  $\rho(\lambda; A)$ .

In the case of equitable graphs, the ansatz of *block symmetry* can be made for the cavity variances:

$$\Delta_i^{(j)}(z) = \Delta_{g_i}^{(g_j)}(z). \quad (11)$$

This ansatz, also made in Ref. [15], allows us to perform the summation in the denominator, that consistently turns out to be independent from the individual node, but only from its block, thus reducing the set of equations for cavity variances from a size of order  $N$  (in the sparse case) to the following set of  $m^2$  equations:

$$\Delta_a^{(b)}(z) = \frac{1}{z - \sum_c^m (c_{ac} - \delta_{bc})^+ \Delta_c^{(a)}(z)}, \quad (12)$$

where  $(x)^+ = \max(x, 0)$ . Block variances can then be computed,

$$\Delta_a(z) = \frac{1}{z - \sum_c^m c_{ac} \Delta_c^{(a)}(z)}, \quad (13)$$

and finally the spectral density,

$$\rho(\lambda) = \frac{1}{\pi m} \sum_{a=1}^m \text{Im}[\Delta_a(z)]_{z=\lambda-i\epsilon}. \quad (14)$$

## IV. RESULTS

In this section, I derive the threshold at which naive spectral partitioning fails and introduce a general algorithm based on

the inverse participation ratio (IPR) to solve the inference problem in equitable graphs with two communities.

### A. Modular structures

Graph partitioning, and in particular spectral bisection, is a long-standing problem in graph theory [24–27]. Here for the modular case, two homogenous blocks are analyzed. The blocks share the same size and the connectivity matrix reads

$$\mathbf{c} = \begin{pmatrix} c_{\text{in}} & c_{\text{out}} \\ c_{\text{out}} & c_{\text{in}} \end{pmatrix}, \quad (15)$$

where  $c_{\text{in}}$  and  $c_{\text{out}}$  are non-negative integers such that  $c_{\text{in}} > c_{\text{out}}$ .

In this case, all nodes in this ensemble of equitable graphs share the same total degree,  $c = c_{\text{in}} + c_{\text{out}}$ , so that the ensemble is a subset of  $c$ -regular graphs, also called the regular stochastic block model [17]. For later use, I also define  $r = c_{\text{in}}/c_{\text{out}}$ , and  $\epsilon = 1/r$  is a quantity that has been used to characterize the strength of the assortative structure [13].

Assortative structures have been widely investigated with various approaches: spectral methods [28], modularity maximization [29], belief propagation [13], Markov-chain Monte Carlo methods [30], and other heuristic algorithms [31]. Stochastic block models have been shown to display a detectability transition [13,20].

In this homogenous case, cavity equations can be further simplified: Cavity variances associated to the two blocks can be assumed to be equal,  $\Delta_1^{(b)} = \Delta_2^{(b)}$  for  $b = 1, 2$ , and given the form of the equations, by inspection, it is also possible to look for fully symmetric solutions such that  $\Delta_a^{(b)} = \Delta^{\text{(cav)}}$  for all  $(a, b)$ . This ansatz yields

$$\Delta^{\text{(cav)}}(z) = \frac{1}{z - (c_{\text{out}} + c_{\text{out}}r - 1)\Delta^{\text{(cav)}}}. \quad (16)$$

Now the equation is identical to the one derived in Ref. [4], and analogously carrying out the algebra, the spectral density found is Kesten-McKay's law [32], as found in Ref. [15]:

$$\rho(\lambda) = \frac{c\sqrt{4(c-1) - \lambda^2}}{2\pi(c^2 - \lambda^2)}, \quad (17)$$

where  $c = c_{\text{out}}(1 + r)$  (Fig. 2).

Equation (17) yields the maximal eigenvalue in the bulk,  $\lambda_b^+ = 2\sqrt{c_{\text{in}} + c_{\text{out}} - 1}$ . Community eigenvalues can be easily computed via the characteristic polynomial:

$$(rc_{\text{out}} - \lambda)^2 - c_{\text{out}}^2 = 0. \quad (18)$$

The first eigenvalue  $\lambda_{\text{max}}$  equals the total connectivity  $c$  and its corresponding eigenvector is constant and uninformative. On the other hand, the second community eigenvalue,  $\lambda_{\text{com}} = c_{\text{in}} - c_{\text{out}}$ , is informative and its relationship with  $\lambda_b^+$  is crucial for the inference problem: When  $\lambda_{\text{com}} > \lambda_b^+$  it is simply the second largest eigenvalue and its corresponding eigenvector can be easily and fast computed, but when  $\lambda_{\text{com}} < \lambda_b^+$  then it is no longer the second eigenvalue and its ranking becomes unknown. This is exactly what is found in the numerical simulation in Ref. [14], which are based on the modularity matrix,  $Q$ .

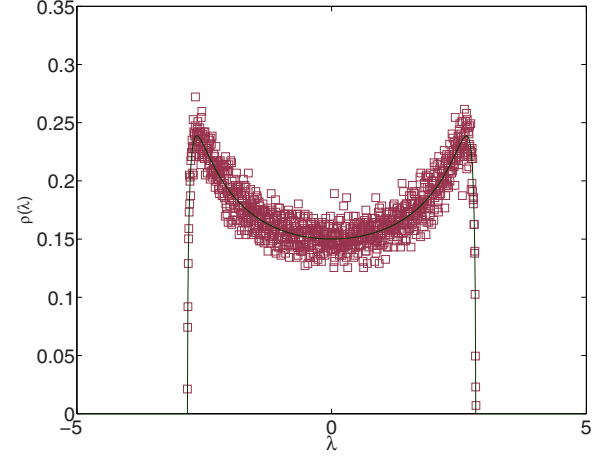


FIG. 2. (Modular case). Spectral density for  $c = 3$  and  $r = 2$  corresponds to Kesten-McKay's law for a  $k$ -regular graph with  $k = 3$ . Squares come from numerical diagonalization of a sample of 100 equitable graphs of size  $N = 1000$ .

Such transition occurs when  $\lambda_{\text{com}} = \lambda_b^+$ , which corresponds to the critical line in the plane  $c - r$  (Fig. 3):

$$r_c = \frac{c + 2\sqrt{c-1}}{c - 2\sqrt{c-1}}. \quad (19)$$

Below the critical line, the community eigenvector gets lost in the bulk and a criterion is needed to identify the right eigenvector.

The solution to this detectability problem in equitable graphs can be found by exploiting the information about the eigenvectors. In fact, the distribution of the eigencomponents of the community eigenvector and of bulk's eigenvectors turn out to be significantly different: From the *block-symmetry* ansatz, the eigenvector corresponding to  $\lambda_{\text{com}}$  is more extended than the typical eigenvector of the bulk.

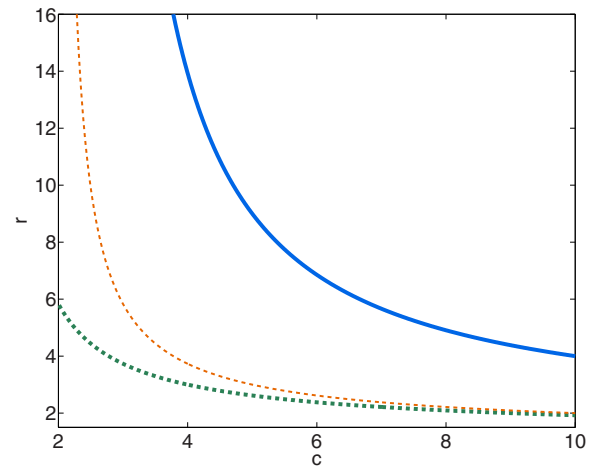


FIG. 3. Critical lines in the plane  $(c, r)$  for equitable graphs (solid line), two-block random regular graphs (dashed line), and SBM (squared line). Above the solid line, standard spectral bisection works for equitable graphs. Below the solid line, naive spectral bisection fails but the IPR-based algorithm succeeds in full recovery.

By looking at a measure of extendedness, such as the inverse participation ratio (IPR) [18,33], it is possible to recognize the informative community eigenvector,  $u_i^{\text{com}}$ , associated to  $\lambda_{\text{com}}$ . In fact, the normalized community eigenvector is block symmetric and all its elements scale like  $1/\sqrt{N}$ , i.e.,  $u_i^{\text{com}} = (\delta_{g_i,1} - \delta_{g_i,2})/\sqrt{N}$ . Consequently,

$$I_4 = \sum_i^N (u_i^{\text{com}})^4 = \frac{N}{2} \sum_a^2 (u_a)^4 = \frac{1}{N}, \quad (20)$$

independently from  $r$ . The inverse participation ratio of the community eigenvector is then  $1/N$  while the random eigenvectors in the bulk have an expected IPR of  $3/N$  [34] and a standard deviation of order  $N^{-3/2}$ , so that the signal-to-noise ratio grows with  $N^{1/2}$ . Therefore, for large graphs, the informative eigenvector remains distinguishable from a typical eigenvector from the bulk.

Then, the inference problem should be solvable for all values of  $r > 1$ , i.e., as long as the signal is actually present, by searching for the most extended eigenvector of the adjacency matrix  $A$  (Fig. 4), excluding the one associated with the uninformative maximum eigenvalue. An analogous approach was followed in Ref. [33] to solve the planted coloring model, where color-symmetric eigenvectors were used to study the convergence of belief propagation in a special class of nontree graphs, obeying a specific regularity condition.

This eigenvector-based solution for the inference problem would also solve the conundrum that arises in Ref. [14]: The detectability threshold for the regular block model is found to be twice as large (see Fig. 3) as the one for the stochastic block model, even though the community structure in the former is partially deterministic (1) while in the latter it is entirely probabilistic. The detectability threshold in Ref. [14] corresponding to (19) only holds for spectral partitioning based on the second eigenvector of the adjacency matrix or, equivalently, on the first eigenvector of the modularity one.

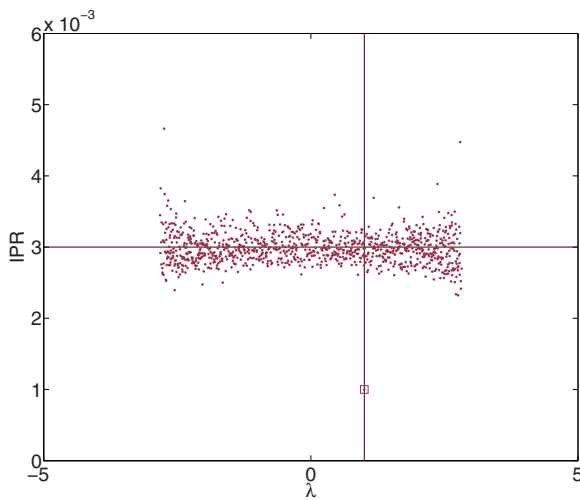


FIG. 4. (Modular case). Inverse participation ratio for each eigenvector in the plane  $\lambda$ -IPR. The eigenvectors of the bulk all share an IPR fluctuating around  $3/N$  while the community eigenvector (square) has an IPR equal to  $1/N$ , which allows us to solve the inference problem also when naive spectral partitioning fails. Parameters are  $c = 3$  and  $r = 2$ .

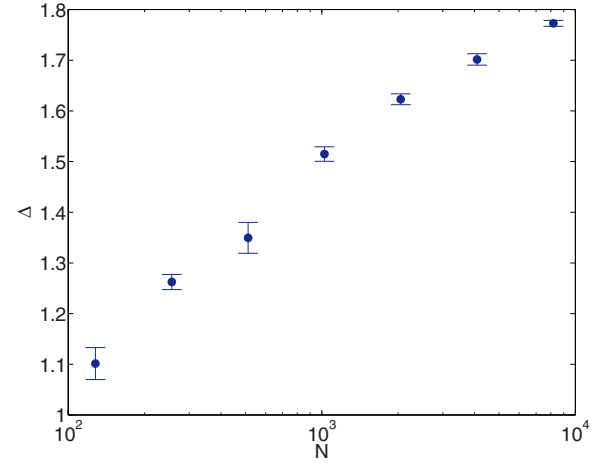


FIG. 5. (Modular case). Relative IPR divergence  $\Delta$ , for graphs of increasing size, ranging from 64 to 8196, each averaged over 10 samples. The most extended random eigenvector in the bulk always remains separate from the significant extended eigenvector associated to the community structure. Parameters are  $c = 9$  and  $r = 2$ .

This result is consistent with the one in Ref. [17], stating that there exists a constant  $d > 0$  such that, for  $c_{\text{in}} > c_{\text{out}} > d$ , the corresponding graph  $G$  almost surely has a unique equitable partition asymptotically. Together, these two results constitute a solid basis for the general conjecture that equitable graphs with two equally sized communities will always admit a full and efficient recovery of the original partition simply as soon as  $c_{\text{in}} > c_{\text{out}} > 0$ , whereas so far it has been rigorously proven only for a limited region of parameters [17].

To validate numerically the performance of the IPR-based algorithm on two-community equitable graphs, extensive simulations were performed varying the connectivity, the assortivity, and the size of the graph. In all cases considered, the algorithm yields an exact recovery, as long as  $c_{\text{in}} > c_{\text{out}}$ .

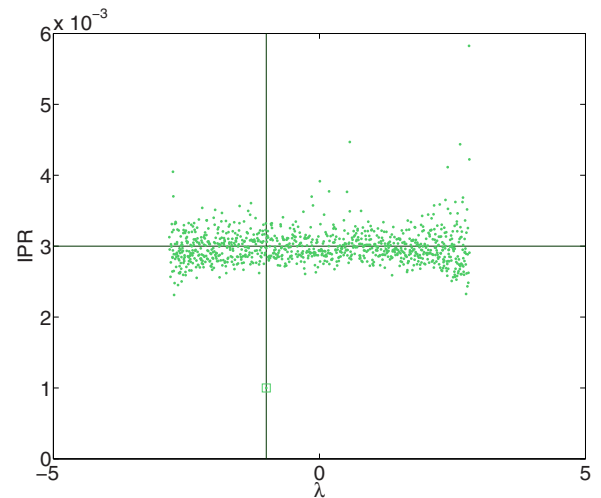


FIG. 6. (Bipartite case) Inverse participation ratio for each eigenvector in the plane  $\lambda$ -IPR. The eigenvectors of the bulk all share an IPR fluctuating around  $3/N$  while the community eigenvector (square) has an IPR equal to  $1/N$ . Parameters are  $c = 3$  and  $r = 2$ .

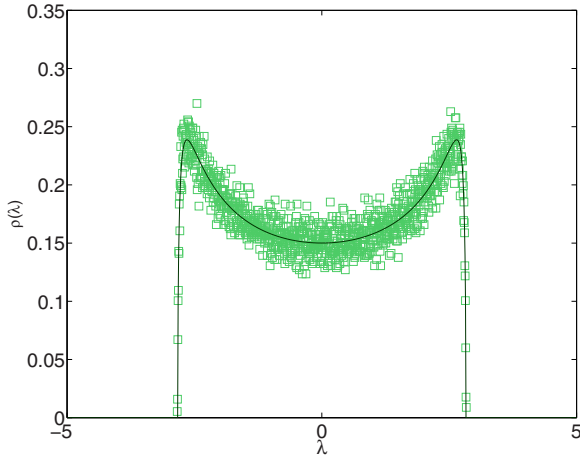


FIG. 7. (Bipartite case) Spectral density for  $c_{in} = 1$  and  $c_{out} = 2$ , corresponding to Kesten-McKay’s law for a  $k$ -regular graph with  $k = 3$ . Squares come from numerical diagonalization of a sample of 100 equitable graphs of size  $N = 1000$ .

To quantify the robustness of the result, a measure of the distance between eigenvectors’s statistics is introduced. IPRs of all eigenvectors (excluding the trivial constant eigenvector associated to the largest eigenvalues,  $\lambda_1 = c$ ) are ordered in increasing order, e.g.,  $I_4^2 = 1/N \leq I_4^3 \leq \dots$ . Then, the relative IPR divergence,  $\Delta$ , is computed as the relative difference between the second smallest IPR in the sequence, constituting the most extended random eigenvector of the bulk and the minimum IPR, associated to the eigenvector that gathers the information on communities, i.e.,  $\Delta = (I_4^3 - I_4^2)/I_4^2$ . Figure 5 shows how  $\Delta$  grows with  $N$ , as fluctuations decrease, and then remains finite for large  $N$ . Since the Gaussian eigenvectors of the bulk have an average IPR of  $3/N$  and the minimum possible IPR is  $1/N$ , the relative IPR divergence is upper bounded by 2 in the asymptotic limit.

**B. Bipartite structures**

In this paragraph, I focus on disassortative equitable graphs, where edges within a block can be present but are always less than edges towards the other block, i.e.,  $c_{in} < c_{out}$ .

The connectivity matrix  $c_{ab}$  considered is the following:

$$\mathbf{c} = \begin{pmatrix} c_{in} & c_{out} \\ c_{out} & c_{in} \end{pmatrix}, \tag{21}$$

where  $c_{in}$  and  $c_{out}$  are non-negative integers such that  $c_{out} > c_{in}$ .

The analysis is entirely analogous to the one put forward for the assortative case (Fig. 6), once the appropriate parallels are drawn: For large  $c_{out}/c_{in}$  the informative community eigenvector corresponds to the lowest eigenvalue,  $\lambda_{com} = c_{in} - c_{out}$ , and the critical line is defined by the condition  $\lambda_{com} = \lambda_b^-$ . Also in this case the cavity equations admit a fully symmetric solution for the variances that leads to Kesten-McKay’s law for the spectral density (Fig. 7).

**V. CONCLUSIONS**

In this paper, equitable graphs [15] have been analyzed via spectral graph theory and graph partitioning theory.

In particular, in the framework of equitable graphs, the picture of the detectability threshold for naive spectral clustering, i.e., only using the eigenvector associated to the second largest eigenvalue of the adjacency matrix, emerges distinctly, as well as the crucial role of the statistics of eigenvectors. Strong analytical and numerical evidence has been provided in support of a new conjecture on the absence of an information-theoretic detectability transition in two-community equitable graphs. Insights from equitable graphs could be used to develop new spectral methods based on both eigenvalues and eigenvectors properties in other graph ensembles.

Future work will deal with the interpolation between standard stochastic block models, regular stochastic block models, and equitable graphs. Further studies will be dedicated to the analysis of heterogeneous and multimodular equitable graph, such as the equitable counterpart of planted partition model, in relation to the problem of resolution limit [35] in modularity maximization and to the generalization of the IPR based algorithm.

**ACKNOWLEDGMENTS**

The author acknowledges support from the FET Project DOLFINS No. 640772, and FET IP Project MULTIPLEX No. 317532. The author would like to thank T. Coolen, T. Martin, F. Lillo, E. Letizia, P. Mazzarisi, and D. Tantari for fruitful discussions.

---

[1] M. E. J. Newman, A. L. Barabasi, and D. J. Watts, *The Structure and Dynamics of Networks* (Princeton University Press, Princeton, NJ, 2006).

[2] D. Easley and J. Kleinberg, *Networks, Crowds, and Markets: Reasoning about a Highly Connected World* (Cambridge University Press, Cambridge, UK, 2010).

[3] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborova, Inference and Phase Transitions in the Detection of Modules in Sparse Networks, *Phys. Rev. Lett.* **107**, 065701 (2011).

[4] T. Rogers, I. P. Castillo, R. Kühn, and K. Takeda, Cavity approach to the spectral density of sparse symmetric random matrices, *Phys. Rev. E* **78**, 031116 (2008).

[5] F. Chung, L. Lu, and V. Vu, Eigenvalues of random power law graphs, *Ann. Combinatorics* **7**, 21 (2003).

[6] L. Erdos, A. Knowles, H. Yau, and J. Yin, Spectral statistics of Erdos-Renyi graphs I: Local semicircle law, *Ann. Probab.* **41**, 2279 (2013).

[7] I. J. Farkas, I. Derenyi, A. L. Barabasi, and T. Vicsek, Spectra of real-world graphs: Beyond the semicircle law, *Phys. Rev. E* **64**, 026704 (2001).

[8] C. Bordenave and M. Lelarge, Resolvent of large random graphs, *Random Struct. Alg.* **37**, 332 (2010).

[9] R. R. Nadakuditi and M. E. J. Newman, Graph Spectra and The Detectability of Community Structure in Networks, *Phys. Rev. Lett.* **108**, 188701 (2012).

[10] G. Robins, P. Pattison, Y. Kalish, and D. Lusher, An introduction to exponential random graph (p\*) models for social networks, *Social Networks* **29**, 173 (2007).

- [11] P. W. Holland, K. B. Laskey, and S. Leinhardt, Stochastic block models: First steps, *Social Networks* **5**, 109 (1983).
- [12] B. Karrer and M. E. J. Newman, Stochastic block models and community structure in networks, *Phys. Rev. E* **83**, 016107 (2011).
- [13] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborova, Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications, *Phys. Rev. E* **84**, 066106 (2011).
- [14] F. Radicchi, Detectability of communities in heterogeneous networks, *Phys. Rev. E* **88**, 010801 (2013).
- [15] M. E. J. Newman and T. Martin, Equitable random graphs, *Phys. Rev. E* **90**, 052824 (2014).
- [16] T. P. Peixoto, Nonparametric Bayesian inference of the microcanonical stochastic block model. *Phys. Rev. E* **95**, 012317 (2017).
- [17] G. Brito, I. Dumitriu, S. Ganguly, C. Hoffman, and L. V. Tran, Recovery and rigidity in a regular stochastic block model, in *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms* (Society for Industrial and Applied Mathematics, Arlington, VA, 2016).
- [18] T. Kawamoto and Y. Kabashima, Limitations in the spectral method for graph partitioning: Detectability threshold and localization of eigenvectors, *Phys. Rev. E* **91**, 062803 (2015).
- [19] E. Mossel, J. Neeman, and A. Sly, Reconstruction and estimation in the planted partition model, *Probab. Theory Relat. Fields* **162**, 431 (2015).
- [20] E. Mossel, J. Neeman, and A. Sly, A proof of the block model threshold conjecture. [arXiv:1311.4115](https://arxiv.org/abs/1311.4115).
- [21] L. Massoulié, Community detection thresholds and the weak Ramanujan property, in *Proceedings of the 46th Annual ACM Symposium on Theory of Computing* (ACM, New York, NY, 2014).
- [22] S. F. Edwards and R. C. Jones, The eigenvalue spectrum of a large symmetric random matrix, *J. Phys. A* **9**, 1595 (1976).
- [23] M. Mézard, G. Parisi, and R. Zecchina, Analytic and algorithmic solution of random satisfiability problems, *Science* **297**, 812 (2002).
- [24] B. Bollobás and A. D. Scott, Exact bounds for judicious partitions of graphs, *Combinatorica* **19**, 473 (1999).
- [25] A. Pothen, H. D. Simon, and K. Liou, Partitioning sparse matrices with eigenvectors of graphs, *SIAM J. Matrix Anal. Appl.* **11**, 430 (1990).
- [26] R. B. Boppana, Eigenvalues and graph bisection: An average-case analysis, in *Proceedings of the 28th Annual IEEE Symposium on Foundations of Computer Science* (IEEE, Los Angeles, CA, 1987).
- [27] S. Fortunato, Community detection in graphs, *Phys. Rep.* **486**, 75 (2010).
- [28] F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborova, and P. Zhang, Spectral redemption in clustering sparse networks, *Proc. Natl. Acad. Sci. USA* **110**, 20935 (2013).
- [29] M. E. J. Newman, Community detection in networks: Modularity optimization and maximum likelihood are equivalent. *Phys. Rev. E* **94**, 052315 (2016).
- [30] P. T. Peixoto, Parsimonious Module Inference in Large Networks, *Phys. Rev. Lett.* **110**, 148701 (2013).
- [31] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, Fast unfolding of communities in large networks, *J. Stat. Mech.* (2008) P10008.
- [32] B. D. McKay, The expected eigenvalue distribution of a large regular graph, *Linear Algebra Appl.* **40**, 203 (1981).
- [33] A. Coja-Oghlan, E. Mossel, and D. Vilenchik, A spectral approach to analyzing belief propagation for 3-colouring combinatorics, *Prob. Comput.* **18**, 881 (2009).
- [34] I. Dumitriu and P. Soumik, Sparse regular random graphs: Spectral density and eigenvectors, *Ann. Probab.* **40**, 2197 (2012).
- [35] S. Fortunato and M. Barthelemy, Resolution limit in community detection, *Proc. Natl. Acad. Sci. USA* **104**, 36 (2007).