Spectral properties of networks with community structure

Sanjeev Chauhan,* Michelle Girvan, and Edward Ott
Department of Physics, University of Maryland, College Park, Maryland 20742, USA
(Received 17 May 2009; published 24 November 2009)

In this paper, we discuss the eigenspectra of networks with community structure. It is shown that in many cases, the spectrum of eigenvalues of the adjacency matrix of a network with community structure gives a clear indication of the number of communities in the network. In particular, for a network with N nodes and N_c communities, there will typically be N_c eigenvalues that are significantly larger than the magnitudes of all the other $(N-N_c)$ eigenvalues. We discuss this property as well as its use and limitations for determining N_c.

DOI: 10.1103/PhysRevE.80.056114 PACS number(s): 89.75.Hc, 02.10.Ud

I. INTRODUCTION

Many real complex networks are characterized by the presence of community structure; i.e., there are groups of network nodes that have relatively stronger relationship with nodes in their own group than with nodes outside. Such structures can have significant influence on the functional characteristics of the network. There has been considerable research on developing techniques for finding community structure [1–3], and this continues to be an active area of research. Many community finding algorithms are based on the concept of modularity [4–7], which divides a network into communities by maximizing this quantity.

Spectral properties of the Laplacian matrix of networks with communities have also been studied quite intensively. These properties can be used to detect community structure in complex networks [8,9]. There has been work that uses synchronization dynamics to find community structure and relate it to the spectral information of the Laplacian matrix [10,11]. The eigenspectra of undirected "real-world" networks without community structure has been studied in Refs. [12,13]. Here, by eigenspectrum of a network we mean the spectrum of its adjacency matrix. Farkas *et al.* [12] studied the spectral density of the sparse uncorrelated random graphs, the small-world graph and the scale-free graph, and their deviation from the well know semicircle law [14,15]. Goh *et al.* [13] analyzed the eigenspectra and eigenvectors of the evolving Barabasi-Albert scale-free networks [16–18]. Random uncorrelated graphs have been used by physicists to study various physical phenomenon, and much work has been done exploring the spectral properties of such matrices [19].

To our knowledge, the eigenspectra of networks, directed or undirected, with community structure has gained little or no attention. The objective of this paper is to study the spectral properties of network adjacency matrix with community structure. In particular, we propose a method for finding the *number* of communities in a network from the eigenspectrum of the network adjacency matrix.

Any given network can be represented by its adjacency matrix A. In the case of unweighted networks treated here, $A_{ij}=1$ if there is a link from node j to node i, and $A_{ij}=0$ otherwise, where $i,j=1,2,\ldots,N$, and N is the number of network nodes. In the case of directed (undirected) networks, in general, $A_{ij}\neq A_{ji}(A_{ij}=A_{ji})$. Our interest is primarily in the case, where N is large and A is sparse. As we shall show, the eigenspectrum of the adjacency matrix of a network with communities has the interesting property that it has multiple eigenvalues that are well separated from the rest of the eigenvalues. Our main point in this paper is that in many cases, the number of such eigenvalues often gives a clear indication of the number of communities in the network.

The organization of this paper is as follows. As background, in Sec. II, we discuss the pattern formed by plots in the complex plane of the eigenvalues of the adjacency matrix of a network with no communities, illustrating the generic occurrence of a cloud of $(N-1)$ eigenvalues of magnitude substantially less than the maximum eigenvalue, which is real and positive. Section III discusses the eigenspectra of networks with communities. We show how the number of communities can be obtained from the eigenspectra of the network adjacency matrix. In Sec. IV, we apply our method to some real-world networks. In Sec. V, we discuss the limitations of our method.

II. EIGENVALUE SPECTRA OF NETWORKS WITHOUT COMMUNITIES

A. Perron-Frobenius eigenvalue

The Perron-Frobenius theorem for matrices with non-negative entries implies that the eigenvalue of A of largest magnitude, here denoted λ_*, is real and positive [20]. As an example, Fig. 1(a) shows a plot of the location of all the eigenvalues of the adjacency matrix of a $N=500$ node Erdos-Renyi directed network with $\langle d_{in}\rangle=\langle d_{out}\rangle=20$, where $\langle\ldots\rangle$ denotes the average over all nodes $(i=1,2,3,\ldots,500)$ and $d^i_{in}(d^i_{out})$ denotes the number of incoming (outgoing) network links at node i [these numbers are also called the in degree (out degree) of node i]. Note that since every out link originating from a node is also an in link for some other node, we necessarily have $\langle d_{in}\rangle=\langle d_{out}\rangle$; thus, we use the notation $\langle d\rangle$ to denote both $\langle d_{in}\rangle$ and $\langle d_{out}\rangle$. For the example in Fig. 1(a), we have taken d_{in} and d_{out} at a node to be uncorrelated. By uncorrelated in/out degrees, we mean that the joint in-degree/out-degree probability distribution function $\tilde{P}(d_{in},d_{out})$, giving the probability of (d_{in},d_{out}) at a randomly chosen node, factors

*sanjeevk@umd.edu

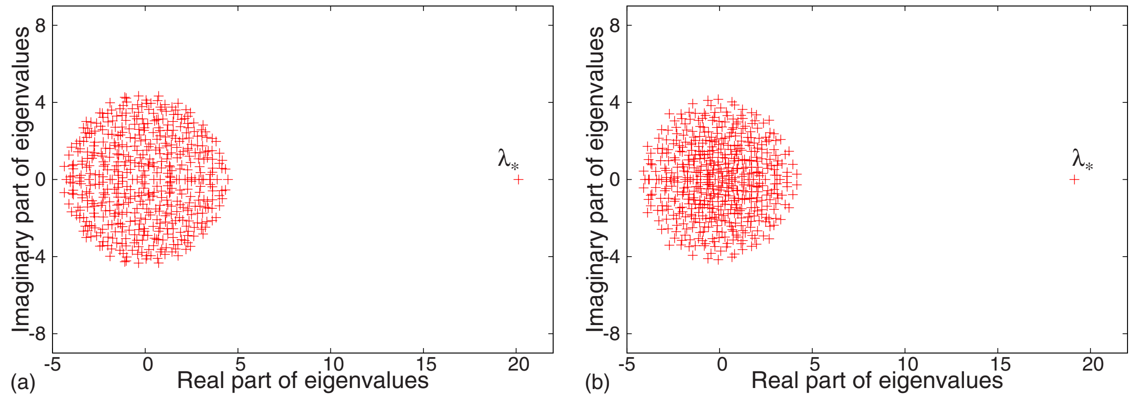

FIG. 1. (Color online) Plots of the real and imaginary parts of the adjacency matrix eigenvalues for computer-generated directed networks with no community structure. The largest eigenvalue λ_* can be seen outside the cloud of the rest of the eigenvalues. (a) Erdos-Renyi network with $N=500$, $\langle d\rangle=20$. (b) scale-free network with $N=500$, $\gamma=2.5$, and $\langle d\rangle=20$.

$$\tilde{P}(d_{in},d_{out}) = P_{in}(d_{in})P_{out}(d_{out}), \tag{1}$$

and, as a consequence, $\langle d_{in}d_{out}\rangle=\langle d_{in}\rangle\langle d_{out}\rangle=\langle d\rangle^2$.

We see in Fig. 1(a) that there is a single real positive eigenvalue $\lambda_*\cong 20$, while all the other 499 eigenvalues fall in a circular cloud centered approximately at the origin and entirely enclosed within a radius denoted λ_0, of about 4, which is substantially less than the maximum eigenvalue $\lambda_*\cong 20$. Thus, there is a large gap between the Perron-Frobenius eigenvalue λ_* and the other eigenvalues. Assuming that, aside from the in-degree/out-degree correlation at a node, the network correlations are otherwise random, the mean-field approximation to λ_* is (see Ref. [21])

$$\lambda_* \cong \frac{\langle d_{in}d_{out}\rangle}{\langle d\rangle}. \tag{2}$$

For an uncorrelated case, i.e., $\langle d_{in}d_{out}\rangle=\langle d\rangle^2$, as in Fig. 1, the mean-field approximation to λ_* is $\lambda_*\cong\langle d\rangle$, in agreement with the numerically found value. On the other hand, as shown in Sec. II B, the root-mean-square radius of the cloud has an upper bound given by $\langle d\rangle^{1/2}$, thus, explaining the separation of λ_* from the other eigenvalues. Figure 1(b) is a plot similar to that in Fig. 1(a), but for the case of a scale-free network with degree distribution as in Eq. (1) with $P_{in}(d)=P_{out}(d)\sim d^{-2.5}$; as for the case illustrated in Fig. 1(a), the network is again randomly connected with $N=500$, $\langle d\rangle=20$. Again, we see a strong separation between the Perron-Frobenius eigenvalue and the cloud formed by the other 499 eigenvalues.

In networks that are undirected (i.e., $A_{ij}=A_{ji}$), all eigenvalues are real, but a similar result still often applies. All the non-Perron-Frobenius eigenvalues lie in an interval approximately centered at zero with root-mean-square radius which, as shown in next section, scales no stronger than $\langle d\rangle^{1/2}$, and $\lambda_*-\lambda_0$ can be large. As can be seen from Eq. (2),

$$\lambda_* \cong \frac{\langle d^2\rangle}{\langle d\rangle} \tag{3}$$

for undirected networks. Note that $\frac{\langle d^2\rangle}{\langle d\rangle}>\langle d\rangle$ (by the Schwartz inequality). References [12,13] have also given some results concerning separation between the largest eigenvalue and the bulk of eigenvalue cloud for certain undirected networks.

B. Size of the cloud of non-Perron-Frobenius eigenvalues

In the case of undirected Erdos-Renyi networks, the semicircle law predicts the size of the eigenvalue cloud as $\sim 2\sqrt{Np(1-p)}$ [12], where p is the probability of connection between two nodes. The distribution of eigenvalues in the cloud in this case is symmetric. For undirected scale-free networks, the spectral density deviates from the semicircle law. It resembles a symmetric triangle-like distribution with power-law tail of the density of the eigenvalues [12,13].

For any given network, directed or undirected, we now show that the root-mean-square radius of the cloud of non-Perron-Frobenius eigenvalues has an upper bound given by $\langle d\rangle^{1/2}$, independent of whether the degrees are correlated or not. Since A has entries either 1 or 0 for the edges, the trace of A^TA, where A^T is the transpose of A, is equal to the total number of directed edges, say M, in the network,

$$\mathrm{Tr}(A^TA) = M. \tag{4}$$

The matrix A can be expressed in Schur decomposition form [22] as

$$A = UQU^*, \tag{5}$$

where U is a unitary matrix and U^* denotes its conjugate transpose. Q is an upper triangular matrix, which can be written as $D+T$, where D is a diagonal matrix with the eigenvalues of A being the diagonal entries, and T is a strictly upper triangular matrix. From this, since A is real,

$$A^* = A^T = UQ^*U^*. \tag{6}$$

Thus, for $\mathrm{Tr}(A^TA)$, we obtain

$$\mathrm{Tr}(A^TA) = \mathrm{Tr}(UQ^*U^*UQU^*) = \mathrm{Tr}(Q^*Q), \tag{7}$$

where we have used the fact that trace is invariant under a similarity transformation and U is unitary.

In Eq. (7), $\mathrm{Tr}(Q^*Q)$ is equal to $\mathrm{Tr}(T^*T)+\Sigma_{k=1}^N|\lambda_k|^2$. Since $\mathrm{Tr}(T^*T)$ is real and positive, Eq. (4) yields

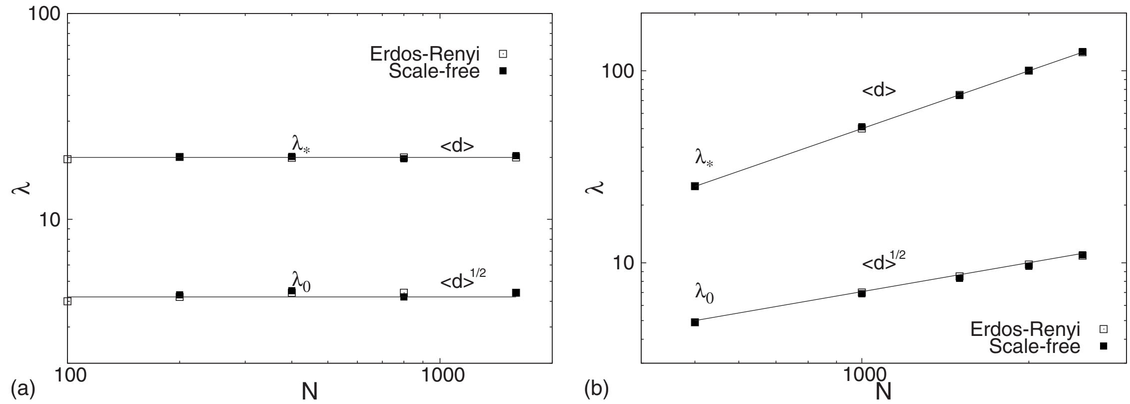

FIG. 2. Plot of the largest eigenvalue λ_* and the actual radius λ_0 of the eigenvalue cloud for networks with no communities versus the number of nodes in the network. All networks are directed with no degree correlations. (a) Erdos-Renyi and scale-free networks with constant degree $\langle d\rangle=20$. (b) Erdos-Renyi and scale-free networks with degree increasing in proportion to N such that $\langle d\rangle/N=0.05$. In plots (a) and (b), the data points for Erdos-Renyi and scale-free networks overlap.

$$\sum_{k=1}^{N}|\lambda_k|^2 \le M. \tag{8}$$

For large N,

$$\langle|\lambda_k|^2\rangle_{k\neq 1} \le (M-\lambda_*^2)/N, \tag{9}$$

where $\langle\ldots\rangle_{k\neq 1}$ denotes the average over all eigenvalues with $\lambda_1\equiv\lambda_*$ not included. The equality holds when the network is undirected. Since, in large sparse networks, $M\gg\lambda_*$ and $M=N\langle d\rangle$, we get an upper bound on the root-mean-square radius of the eigenvalue cloud as

$$\langle|\lambda_k|^2\rangle_{k\neq 1}^{1/2} \le (M/N)^{1/2}=\langle d\rangle^{1/2}. \tag{10}$$

Figure 2 shows a plot of the largest eigenvalue λ_* and the actual radius of the cloud λ_0, with changing network sizes for random computer-generated, directed, and in/out-degree-uncorrelated networks. Plots for both Erdos-Renyi and scale-free networks are shown. Figure 2(a) is for the case where $\langle d\rangle=20$ is held constant as N increases. Figure 2(b) is for the case where $\langle d\rangle/N=1/20$ is held fixed as N increases. The upper solid lines in Figs. 2(a) and 2(b) correspond to $\langle d\rangle$, while the lower ones correspond to $\langle d\rangle^{1/2}$. We see that $\lambda_*\cong\langle d\rangle$ for uncorrelated directed networks, in agreement with Eq. (2). The actual radius of the cloud (not the root-mean-square radius), on the other hand, for this particular case of uncorrelated directed networks is approximately equal to $\langle d\rangle^{1/2}$. Thus, we see that, for the cases shown, the largest eigenvalue is well separated from the cloud of the rest of the eigenvalues, and, as the average degree of the network increases, the separation between them increases. Figure 3 shows a similar plot for undirected networks. In this case too, we see the large separation between λ_* and λ_0. All scale-free networks considered in Figs. 2 and 3 have degree distribution $P_{in}(d)=P_{out}(d)\sim d^{-\gamma}$, with the exponent $\gamma=2.5$.

We note that, although we have only presented illustrative numerical results for random networks, we have also conducted extensive tests for networks with other structures (e.g., assortative and disassortative networks) obtaining similar results.

C. Shape of the cloud of non-Perron-Frobenius eigenvalues

For a network with zero or few number of bidirected edges, the cloud of non-Perron-Frobenius eigenvalues is circular. Here, by a bidirected edge we mean a pair of directed

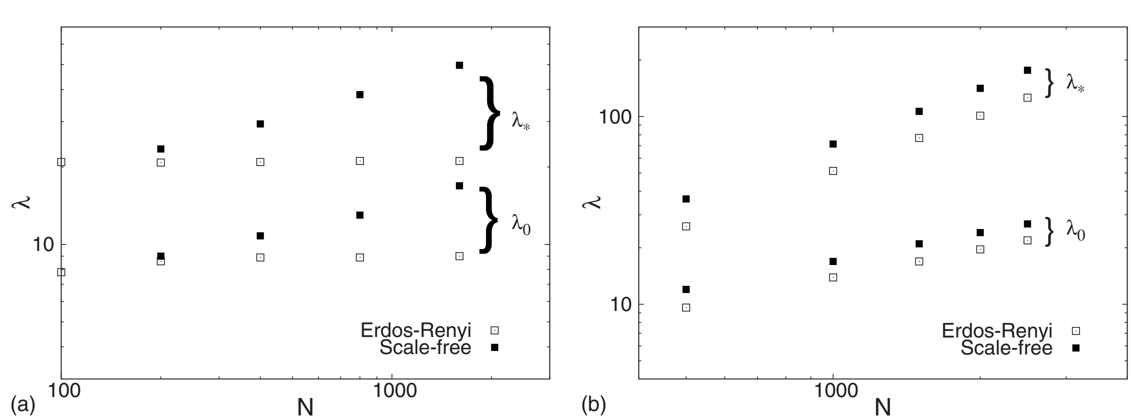

FIG. 3. Plot of the largest eigenvalue and the actual radius of the cloud for networks with no communities versus the number of nodes in the network. All networks are undirected. (a) Erdos-Renyi and scale-free networks with constant degree $\langle d\rangle=20$. (b) Erdos-Renyi and scale-free networks with $\langle d\rangle/N=0.05$.

edges corresponding to $A_{ij}=A_{ji}=1$ for nodes i and j. However, for a network where the number of bidirected edges is comparable to M, numerical computations show that the cloud shape becomes elliptical. In the limiting case where we have all bidirected edges, i.e., the case of undirected networks ($A_{ij}=A_{ji}$), the cloud collapses to a line interval on the real axis. This transition from circle to ellipse to line interval can be understood by considering the trace of A^2, which is equal to the sum of the squares of the eigenvalues of A. Topologically, the trace of A^2 is equal to the number of directed cycles of length two, which in turn equals twice the number of bidirected edges in the network. Thus,

$$\sum_{k=1}^{N} \{[\mathrm{Re}(\lambda_k)]^2 - [\mathrm{Im}(\lambda_k)]^2\} = 2M_b. \tag{11}$$

where M_b is the number of bidirected edges in the network. In the above equation, we have used the fact that complex eigenvalues occur in conjugate pairs. Now for the networks with no self-loops, $\langle \mathrm{Re}(\lambda_k)\rangle = \langle \mathrm{Im}(\lambda_k)\rangle = 0$, since $\mathrm{Tr}(A)=0$. Thus, for $M_b \gg \lambda_*^2$, the difference in the spread of real and imaginary parts of the eigenvalues in the cloud is given by

$$\sigma^2[\mathrm{Re}(\lambda_k)]_{k\neq 1} - \sigma^2[\mathrm{Im}(\lambda_k)]_{k\neq 1} \cong \frac{2M_b}{N}, \tag{12}$$

where $\sigma^2[.]$ denotes the variance of the corresponding entries. The size of the term on the right-hand side of Eq. (12) determines the ellipticity of the eigenvalue cloud for networks with zero or very small number of self loops. Thus, the ellipticity of the eigenvalue cloud measures the number of pairs of nodes in the network that have direct mutual relationship with each other (i.e., are joined by bidirected links). In the normalize form, for the large sparse networks, we can write the ellipticity of the eigenvalue cloud as $\frac{2M_b}{M}$, which has the property that $0 \leq \frac{2M_b}{M} \leq 1$.

In general, the distribution of eigenvalues in the cloud of non-Perron-Frobenius need not be symmetric and the cloud may be asymmetric. This happens when the odd moments ($\mathcal{M}_j, j=3,5,7,\ldots$) of the graph spectral density are nonzero, where

$$\mathcal{M}_j = \frac{1}{N}\sum_{k=1}^{N} \lambda_k^j = \frac{1}{N}\mathrm{Tr}(A^j). \tag{13}$$

Topologically, $\mathrm{Tr}(A^j)$ counts the number of j-hop closed paths in the network. Farkas *et al.* [12] considered the case of undirected small-world networks in which $\mathcal{M}_3$ is high because of the high value of clustering (density of graph triangles). Accordingly, they find high skewness in the spectral density of the small-world graphs.

III. NETWORKS WITH COMMUNITIES

In order to see how the phenomenon of Fig. 1 (i.e., the appearance of λ_* well outside the cloud of other eigenvalues) is affected by the presence of community structure, we give several numerical examples in Sec. III B. Analytical results, describing the behavior of largest eigenvalues observed in Sec. III B, are given in Sec. III C. Before presenting our numerical results in Sec. III B, we give our method of generating directed networks with community structure.

A. Generating directed networks with communities

In our numerical experiments in Sec. III B, we consider two types of networks. One of them is the Erdos-Renyi-type directed network with communities with random placement of both within community and between community links. The second type of network is the scale-free network with communities with power-law degree distribution. To generate Erdos-Renyi-type directed networks with communities, we divide the N nodes in the network into the desired number of communities, say N_c. Communities could have equal or unequal number of nodes as required. Elements A_{ij} of the adjacency matrix, corresponding to links between nodes within the same community, are set to 1 with some chosen probability (else they are zero), while elements corresponding to links between nodes in different communities are made 1 with some other smaller probability. By changing these probabilities, we can tune the strength of community structure and the average degree in the network.

To generate scale-free directed networks with community structure, we again start by dividing the nodes into the desired number of communities. For making connection between nodes in the same community, we generate power-law degree distribution $P(d) \propto d^{-\gamma}$ for both the in degrees and the out degrees of the nodes in the community. Say the kth community has N_k nodes. We generate N_k numbers using the formula [23]

$$b(m + m_0 - 1)^{-1/(\gamma-1)}, \tag{14}$$

for $m=1,2,3,\ldots,N_k$. Here, the constants b and m_0 determine the maximum degree and node averaged degree. We randomly assign these N_k numbers to the N_k nodes in community k and call these assigned numbers the target within community in degree of the corresponding node i. We denote this number $t_{i,in}^k$. We then repeat the random assignment of these numbers and call the result the target within community out degree of node i, $t_{i,out}^k$. Note that $t_{i,in}^k$ and $t_{i,out}^k$ are assigned independently at random, so that they are uncorrelated. From these target degree sequences, we obtain the (i,j)th entry of the adjacency matrix A, where i and j are in community k, by setting $A_{i,j}=1$ with probability

$$p_{ij}^k = \frac{1}{M_k} t_{i,in}^k t_{j,out}^k, \tag{15}$$

where M_k is the target number of edges between nodes in community k. Note that $M_k = \Sigma_i t_{i,in}^k = \Sigma_i t_{i,out}^k$. Links between communities are assigned in a similar manner. For example, say we want to generate links pointing from nodes in community l to nodes in community k. For N_k nodes in community k, we generate N_k numbers using Eq. (14). We assign these N_k numbers to nodes in community k and call them the target in links from nodes in community l to nodes in community k, $t_{i,in}^{kl}$ for the ith node in community k. We repeat this procedure to get target out links from nodes in community l to nodes in community k, $t_{j,out}^{kl}$ for the jth node in community

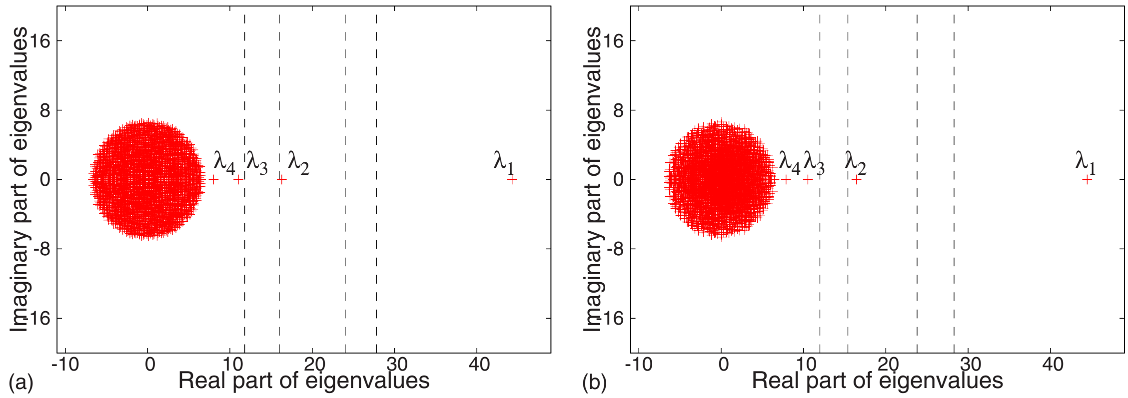

FIG. 4. (Color online) Plot of real and imaginary parts of eigenvalues of computer-generated directed networks with unequal-sized communities. (a) Erdos-Renyi-type network and (b) scale-free network. The average number of within community and between community links are equal in the two cases. We see four eigenvalues corresponding to four communities outside the cloud of rest of the eigenvalues.

l. For a link from node j to node i, we then use the probability,

$$p_{ij}^{kl} = \frac{1}{M_{kl}} t_{i,in}^{kl} t_{j,out}^{kl}, \tag{16}$$

where M_{kl} is the target number of between community links pointing from nodes in community l to nodes in community k. While generating these target degrees, we choose our constants b and m_0 in Eq. (14) such that $M_{kl}=\Sigma_i t_{i,in}^{kl} \approx \Sigma_j t_{j,out}^{kl}$. We repeat this procedure for all pairs of communities. While assigning the target values for the number of links to each node, we assign higher $t_{i,in}^{kl}$ and $t_{j,out}^{kl}$ to nodes with higher $t_{i,in}^{k}$ and $t_{j,out}^{l}$, respectively. Similarly, nodes with smaller within community target links get smaller between community target links. Using this procedure, we get power-law distribution for both within community and between community in degrees and out degrees.

B. Numerical results

In this section, we will verify numerically that when the network has N_c communities, the eigenvalue plot shows N_c eigenvalues outside the cloud of non-Perron-Frobenius eigenvalues. We consider two cases of networks with $N=2000$ nodes consisting of four communities: case (i) the communities have different sizes; $N_c=700$, 600, 400, and 300; case (ii) all the communities are of equal size, $N_c=500$ for each of the four communities.

For the case where the average degree of nodes in a community is proportional to the number of nodes in a community, case (i) leads to the situation where the largest eigenvalues of communities that are "disconnected" (i.e., there are no between community links) are nondegenerate, while for case (ii) the largest eigenvalues will be approximately degenerate. Figure 4 shows the eigenvalue plot for a computer-generated Erdos-Renyi-type network and for a scale-free network for case (i). Figure 4(a) is for the Erdos-Renyi-type network, and Fig. 4(b) is for the scale-free network with $\gamma=2.5$ in Eq. (14). For the Erdos-Renyi-type network used to get the eigenvalue plot in Fig. 4(a), the probability of connection between pairs of nodes within same community was 0.04. With this, the average degree of nodes in a community is proportional to the number of nodes in the community. For between community edges, the probability of connection between pairs of nodes was 0.015. With these parameters, the sum of the number of edges within all communities equals the number of all between community edges. The average degree of nodes in the network is $\langle d \rangle \approx 44$. For generating the scale-free network for the plot in Fig. 4(b), the number of edges within communities and between pairs of communities was the same as the number of edges for the Erdos-Renyi-type network described above. For within community links, the maximum degree in the sequence from Eq. (14) was one fifth of the total number of nodes in the community. For between community links for a pair of communities, the maximum degree was one tenth of the number of nodes in the smallest community from the pair.

In both cases in Fig. 4, it is evident that there are four real positive eigenvalues that occur outside a circular-shaped cloud formed by the remaining 1996 non-Perron-Frobenius eigenvalues. For comparison, we indicate by vertical dashed lines the four largest (real) eigenvalues that would result if the between community links of these networks were removed. For the smallest community, the number of in links (and also out links) from other communities was approximately twice the number of within community links. In this case, we still see the perturbed largest eigenvalue of this community outside the cloud of non-Perron-Frobenius eigenvalues.

Figure 5 shows the eigenvalue plot of a computer-generated Erdos-Renyi-type [Fig. 5(a)] and a scale-free [Fig. 5(b)] network with $\gamma=2.5$ in Eq. (14) for case (ii). For Erdos-Renyi-type and scale-free networks, the network generation parameters are chosen such that the nodes on an average have 20 within community in/out links and 20 in/out links to nodes not in their community. Again, it is clearly evident that there are four eigenvalues occurring outside the cloud of 1996 non-Perron-Frobenius eigenvalues. If the between community links of these networks are removed, the four eigenvalues are nearly degenerate with an average value indicated by the vertical dashed line. In both Figs. 5(a) and 5(b), we see that three of the eigenvalues outside the cloud cluster tightly together, while the larger of the four eigenval-

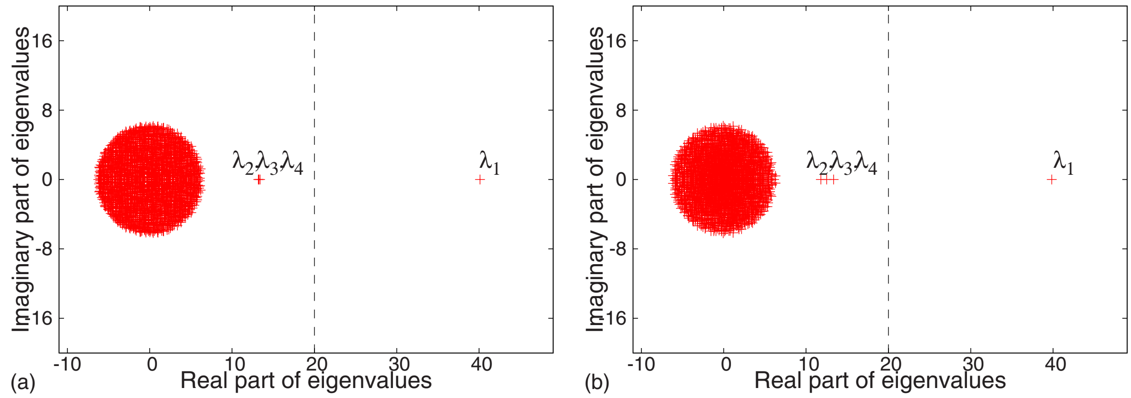

FIG. 5. (Color online) Plot of the real and imaginary parts of the eigenvalues of the adjacency matrix of computer-generated directed networks with four equal-sized communities. The four largest eigenvalues can be seen outside the cloud formed by the rest of the eigenvalues. (a) Erdos-Renyi-type network and (b) scale-free network.

ues outside the cloud has a substantially bigger value. This largest eigenvalue is always real and positive (by Perron-Frobenius theorem). The triplet of other three larger eigenvalues, in general, could have a complex-conjugate pair. Furthermore, when we take the average of these four eigenvalues, this average turns out to be very nearly equal to the degenerate value obtained with the between community connections removed. This observed structure will be explained further in our analysis in Sec. III C 2.

C. Perturbation theory

As verified numerically in the section above, when the network has N_c communities, the eigenvalue plot shows N_c eigenvalues outside the cloud formed by the rest of the eigenvalues. In order to understand this, consider the simple limiting case of a directed network with multiple communities, where all the links exist within the communities, and there are no links between communities. In this case, with the proper labeling of the nodes, the adjacency matrix shows block-diagonal structure [i.e., there are N_c blocks along the matrix diagonal with $A_{ij} \equiv 0$ for (i,j) not in a block]. Thus, the eigenvalues of the adjacency matrix are simply the union of the eigenvalues of the individual blocks. Hence, a plot of the real and imaginary parts of the eigenvalues of the adjacency matrix then has the largest eigenvalues of each of the communities outside the cloud of its other eigenvalues. In addition, these eigenvalues outside their community clouds are all positive and real. In the case where the smallest community Perron-Frobenius eigenvalue exceeds the largest of the radii of the community clouds, the adjacency matrix of the whole network will have N_c Perron-Frobenius eigenvalues outside the aggregate cloud formed by the individual community clouds. Furthermore, we claim that when links between communities are added, provided that the number of added links is not too great, the eigenspectrum still shows that the number of eigenvalues outside the cloud corresponds to the number of communities N_c.

In order to analytically address the above claim, we will use the perturbation theory by considering links between communities as a perturbation to the adjacency matrices of originally disconnected communities. First we consider the case of networks that have nondegenerate largest eigenvalues of disconnected communities, which corresponds to case (i) in Sec. III B. Following that, we consider networks that have degenerate (or nearly degenerate) largest eigenvalues of disconnected communities, which corresponds to case (ii) in Sec. III B.

1. Nondegenerate case

In this section, we analyze the case of networks that have nondegenerate largest eigenvalues of disconnected communities. We will show that the largest eigenvalues of the disconnected communities have lowest nonzero perturbative correction of second order when the addition of between community links is treated as a perturbation.

Consider the case of networks that have N_c unequal-sized communities, each having unequal (i.e., nondegenerate) largest eigenvalues when the communities are disconnected. Let A denote the adjacency matrix of such a network. With proper labeling of the nodes, the matrix A will have block matrix structure with $N_c \times N_c$ number of blocks. Blocks on the diagonal correspond to the adjacency matrices of the individual communities, while the off-diagonal blocks correspond to the perturbation (connections between communities). Let us denote by (I,J) the block of A (Fig. 6). When $I=J$, the block is the adjacency matrix of community I, while if $I \neq J$ then $A_{(I,J)}$ corresponds to the block of the adjacency

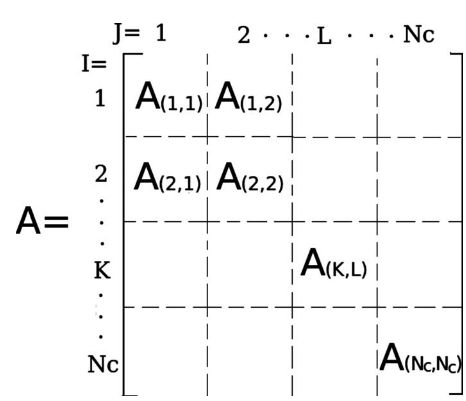

FIG. 6. Adjacency matrix in block matrix form.

matrix in which links pointing from community J to community I are stored. Now, let us write A as

$$A = A_0 + \delta A, \tag{17}$$

where A_0 is a matrix whose diagonal block elements are the diagonal block elements of A and whose off-diagonal block elements are zero. δA is a matrix with zeros on its diagonal blocks and with off-diagonal block elements being the off-diagonal blocks of A. For the case where between community connections are sufficiently sparser than within community link, we regard δA as a perturbation to A_0.

We denote the N_c nondegenerate largest eigenvalues of A_0 by λ_{*k}, where $k=1,2,\ldots,N_c$. Let U_k be the right eigenvector of A_0, corresponding to the eigenvalue λ_{*k}, where entries in U_k are zero except for those elements corresponding to community k in A_0. Write the perturbations of U_k and λ_{*k} due to δA in Eq. (17), as

$$U_k' = U_k + \delta U_{k,1} + \delta U_{k,2}, \tag{18}$$

$$\lambda_{*k}' = \lambda_{*k} + \delta\lambda_{*k,1} + \delta\lambda_{*k,2}, \tag{19}$$

where the subscripts 1 and 2 denote first- and second-order corrections. Letting V_k denote the left eigenvector of A_0, corresponding to the eigenvalue λ_{*k}, and multiplying $AU_k'=\lambda_{*k}'U_k'$ from the left by V_k, we obtain

$$\delta\lambda_{*k,1} + \delta\lambda_{*k,2} = V_k \delta A\, \delta U_{k,1}, \tag{20}$$

where we have made use of $V_k \delta A U_k=0$, which follows from the facts that δA is zero on its diagonal blocks, while both V_k and U_k are nonzero only for their entries corresponding to community k. Since we assume δA to be small, the right-hand side of Eq. (20) is of second order and, hence, $\delta\lambda_{*k,1}$ is zero. Therefore, the lowest nonzero correction to the largest eigenvalues is of second order,

$$\delta\lambda_{*k,2} = V_k \delta A\, \delta U_{k,1}. \tag{21}$$

This shows that the largest eigenvalues of disconnected communities that have nondegenerate largest eigenvalues are perturbed more weakly than the perturbation applied.

First-order correction $\delta U_{k,1}$ to the eigenvector U_k can be found by again considering $AU_k'=\lambda_{*k}'U_k'$. Keeping terms up to first order, we get

$$\delta U_{k,1} = (\lambda_{*k} - A_0)^{-1} \delta A\, U_k, \tag{22}$$

when $(\lambda_{*k}-A_0)$ is invertible. Throughout this section, we have assumed that the eigenvectors of A_0 are normalized such that $V_k U_k=1, \forall k$.

We tested our calculations, specifically Eq. (21), by comparing with actual eigenvalues of some computer-generated Erdos-Renyi-type directed networks with four unequal-sized communities. In Fig. 7, we show comparison between actual and predicted four largest eigenvalues of the network adjacency matrix with increasing between community links. The networks have $N=2000$ with 700, 600, 400, and 300 nodes in each community. The probability of connection between pairs of nodes in the same community was 0.037, which gives $\langle d\rangle \approx 20$ for the whole network, when there are no between community links. When there are nonzero links between communities, to get an estimate of the four perturbed largest eigenvalues, we numerically calculate the four largest eigenvalues of the disconnected communities and add the lowest-order correction given by Eq. (21). As can be seen, our perturbation calculation predicts the four largest eigenvalues well when the number of between community links is small. The radius of the cloud (the symbol ⋇) given in Fig. 7 is the actual radius of the disk of the non-Perron-Frobenius eigenvalues found by numerically calculating all the eigenvalues of the network adjacency matrix. Figure 7 also shows that when the number of between community links is large, we can still see the actual perturbed largest eigenvalue of the smallest community outside the cloud of the non-Perron-Frobenius eigenvalues.

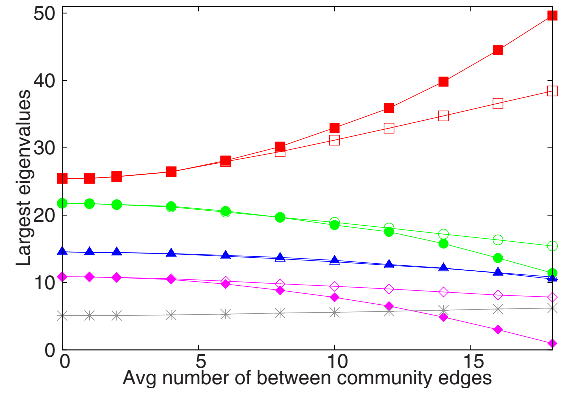

FIG. 7. (Color online) Comparison of the actual and predicted four largest eigenvalues with increasing between community edges for Erdos-Renyi-type directed networks with four unequal-sized communities. Squares (■, □) correspond to λ_{*1}', circles (●, ○) to λ_{*2}', triangles (▲, △) to λ_{*3}', and diamonds (♦, ◇) correspond to λ_{*4}'. Open symbols correspond to actual values while the filled ones are the estimated values calculated using the second-order perturbation theory. The symbol ⋇ shows the actual radius of the non-Perron-Frobenius eigenvalue cloud. All data points are averaged over 20 simulated networks. Error bars are smaller than the symbol sizes. Lines are just guide for the eyes.

2. Degenerate case

We now consider the case of networks that have N_c initially disconnected equal-sized communities, each with N/N_c nodes and with similar number of within community edges. In this case, each of these disconnected communities will have approximately equal largest eigenvalues. We denote this approximately common eigenvalue by λ_*. As the perturbation is applied by adding between community links, we find that (N_c-1) of the N_c perturbed largest eigenvalues will become approximately equal and smaller than the remaining perturbed largest eigenvalue (as an example, see in Fig. 5 for the case $N_c=4$). The perturbation of these eigenvalues is such that the mean distance of all these N_c largest eigenvalues from their initial value is zero. The adjacency matrix A will have block matrix structure with $N_c \times N_c$ number of blocks of equal sizes of dimension $N/N_c \times N/N_c$. As before, we write $A=A_0+\delta A$, with A_0 and δA being same as described in Eq. (17).

Let us write a right eigenvector, say U', of A, which corresponds to one of the perturbed largest eigenvalues as

$$U' = \sum_{k=1}^{N_c} \alpha_k U_k + \delta U, \quad (23)$$

where α_k's are the coefficients to be determined, δU is a higher-order correction, and U_k denotes the right eigenvector of the block matrix A_k, corresponding to its maximum eigenvalue. All the blocks in matrix A_k are zero except for the diagonal block corresponding to community k. As a consequence, the entries in U_k will be zero except for those elements corresponding to community k in A_0. We regard δU as small since we regard the perturbation to be small. Note that as in Eq. (18), the perturbed eigenvector U' in Eq. (23) does not have subscript k, corresponding to community k, because we will have N_c such eigenvectors for different sets of coefficients α_k. We again denote by V_k the left eigenvector of A_0, corresponding to the maximum eigenvalue of A_k, and assume that the eigenvectors of A_0 are normalized such that $V_k U_k = 1$.

Multiplying $A_0 + \delta A$ from right by U' and from left by V_l, and keeping terms up to the first order, we get

$$\sum_{k \neq l} y_{lk} \alpha_k = \alpha_l (\lambda'_* - \lambda_*), \quad (24)$$

where λ'_* is the perturbed eigenvalue and $y_{lk} = V_l \delta A U_k$. For N_c different V_l eigenvectors, we will have N_c such equations corresponding to $l = 1, 2, \ldots, N_c$ in Eq. (24).

For the case in which we have equal-sized communities that have similar number of within and between community links with the same degree distribution (similar perturbation for all the communities), all the y_{lk} coefficients are approximately the same. Thus, to simplify our calculation and to get qualitative results, we assume that $y_{lk} = y \, \forall \, l, k$ with $y > 0$. Equation (24) is an eigenvalue problem of the form $C\alpha = \lambda'_* \alpha$ with $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_{N_c}]^T$, and

$$C = y \tilde{\mathbf{1}}_{N_c} + (\lambda_* - y) \mathbf{1}_{N_c}, \quad (25)$$

where $\tilde{\mathbf{1}}_{N_c}$ is a $N_c \times N_c$ matrix all of whose entries are ones, while $\mathbf{1}_{N_c}$ is the $N_c \times N_c$ identity matrix. The eigenvectors of C are thus the eigenvectors of $\tilde{\mathbf{1}}_{N_c}$. One such eigenvector is $[111\ldots.1]^T$ corresponding to an eigenvalue of C equal to $\lambda_* + (N_c - 1)y$. The other $N_c - 1$ eigenvectors of $\tilde{\mathbf{1}}_{N_c}$ correspond to the $N_c - 1$-dimensional space of vectors $[\alpha_1, \alpha_2, \ldots, \alpha_{N_c}]^T$ such that $\Sigma_i \alpha_i = 0$. For all these vectors, the eigenvalue of $\tilde{\mathbf{1}}_{N_c}$ is zero, corresponding to $N_c - 1$ degenerate eigenvalues of C given by $\lambda_* - y$. This suggests that there is a largest perturbed eigenvalue, approximately given by $\lambda_* + (N_c - 1)y$, which is larger than the rest of the $N_c - 1$ degenerate eigenvalues, approximately given by $\lambda_* - y$, which are clumped together (as can be seen in Fig. 5 for $N_c = 4$). Note that the average of the N_c perturbed eigenvalues is λ_* (the unperturbed degenerate eigenvalue). It can be shown with simple argument that y scales no stronger than N.

Figure 8 gives comparison between actual eigenvalues and our calculations of this section. The networks considered

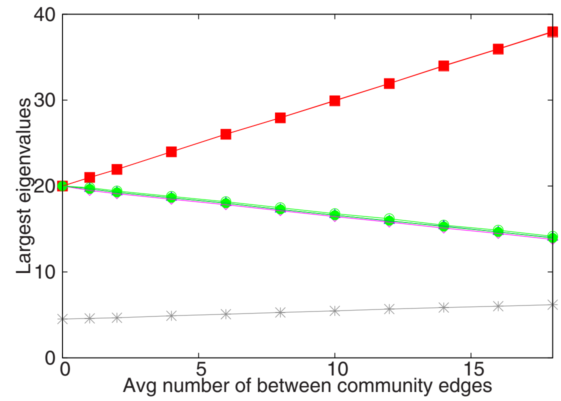

FIG. 8. (Color online) The actual and predicted four largest eigenvalues with increasing between community edges for Erdos-Renyi-type directed networks with four equal-sized communities. Squares (■ for predicted, □ for actual) correspond to λ_1. Data points for predicted values overlap with the actual ones. The symbols ⋇ show the actual radius of the non-Perron-Frobenius eigenvalue cloud. Rest of the data points correspond to to λ_2, λ_3, and λ_4, which are all approximately equal. Here, the data points for predicted λ_2, λ_3, and λ_4 lie on top of each other and overlap with the actual ones. All data points are averaged over 20 simulated networks. Error bars are smaller than the symbol sizes. Lines are just guide for the eyes.

are the Erdos-Renyi-type directed networks with four equal-sized communities. The networks have $N = 2000$ with 500 nodes in each community. The within community link probability is 0.04, which gives $\langle d \rangle \approx 20$ when the communities are disconnected. The estimate of the perturbed four largest eigenvalues was calculated by numerically finding the unperturbed largest eigenvalues and using the estimate of y, which is calculated by averaging over all the 12 possible y_{lk}'s. As can be seen in Fig. 8, our calculations agree well with the actual values.

Thus, our numerical results of Sec. III B seem to be quite well explained by our perturbation results of the present section (Sec. III C) even though the "perturbations" for the numerical examples of Sec. III are not small (e.g., Figs. 4 and 5).

D. Discussion

Based on our perturbation analysis, one might suspect that since the unperturbed eigenvectors U_k and V_k have nonzero element values only for their community entries, it might be possible to use the eigenvectors of the adjacency matrix A to obtain the communities and not just their number. In initiating our research reported here, we were originally motivated by this possibility. However, as described below, we found this to be problematic.

In Sec. III C 1, the eigenvectors of the matrix A, corresponding to the largest eigenvalues of communities, are denoted by U'_k. When δA is small, the entries in eigenvector U'_k, that are labeled by nodes belonging to nodes in community k, will have larger magnitude compared to entries labeled by nodes not in community k. For a given node i, by comparing entries labeled by node i in eigenvectors U'_k, for

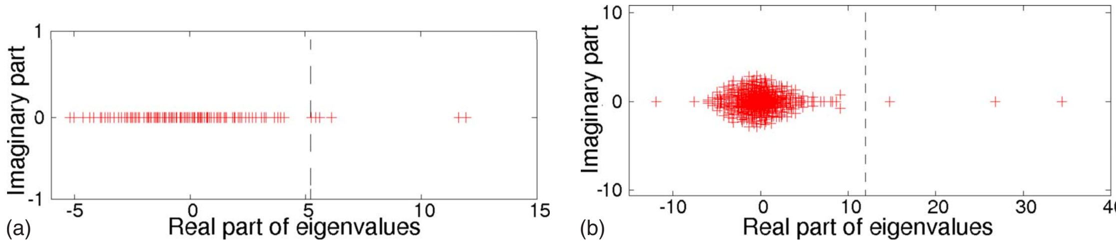

FIG. 9. (Color online) Plots of the real and imaginary parts of the eigenvalues of adjacency matrix of real networks. (a) Political books network. (b) Political blogs network.

$k=1,2,\ldots,N_c$, we can assign node i to the group of nodes that have the largest magnitude of the corresponding entry in the same eigenvector.

Our experimentation with this method on some computer-generated networks shows that the method works pretty well when the eigenvalues of disconnected communities are non-degenerate and the perturbation is not too large. This method fails, however, when the maximum eigenvalues of disconnected communities are too close and the perturbations are too large. When the maximum eigenvalues of disconnected communities are nearly degenerate, an indication of the difficulty is provided by Eq. (23), which shows that the perturbed eigenvector can have almost equal contribution from all the unperturbed U_k eigenvectors ($k=1,2,\ldots,N_c$).

IV. APPLICATION TO REAL NETWORKS

We now test our prediction on two real networks for which we show eigenvalue plots in Fig. 9. The networks considered are the political books network [24] and the political blogs network [25]. These two examples are convenient for our purpose since we naturally have the division of the network into two major groups based on the left/liberal or right/conservative orientation of the book or the blog. The political books network is an undirected network. The nodes represent books on politics available from the online retailer Amazon.com. There is an edges between two nodes when the same buyer(s) buys books represented by the nodes. The Political blogs network, on the other hand, is the compilation of network data on US political weblogs as recorded by Adamic and Glance [25]. It is a directed network where the edges represent hyperlinks between the weblogs on US politics.

The total number of nodes in the political books network is 105. We show the eigenvalue plot of the adjacency matrix of the political books network in Fig. 9(a). Since it is an undirected network, the adjacency matrix is symmetric, and all eigenvalues are consequently real. We "estimate" the size of the eigenvalue cloud by the magnitude of the most negative eigenvalue. The vertical dashed line in Fig. 9(a) corresponds to this value. Consistent with the prediction in this paper, we see that there are two eigenvalues substantially to the left of this dashed line ($\lambda=11.9$, 11.6).

The political blogs network is a relatively larger network as compared to the political books network. It is a directed network with 1224 nodes. The eigenvalues of the adjacency matrix of this network are in general complex since this is a directed network [Fig. 9(b)]. The cloud of eigenvalues is substantially contracted towards the real axis. For this network, we have $M_b=2307$ and $M=19022$. The difference in the spread of real and imaginary parts of eigenvalue cloud [left-hand side of Eq. (12) with two largest eigenvalues excluded] is 2.22. Again we estimate the cloud size from the magnitude of the most negative eigenvalue (vertical dashed line). The two eigenvalues of magnitudes 34.5 and 26.9, corresponding to the two communities, can be seen separated from the rest of the cloud by a large amount.

In Figs. 9(a) and 9(b), we see that there are few eigenvalues that lie just outside (to the right of) the vertical dashed line. These eigenvalues lying close to the vertical dashed line cannot be said to belong to any particular community with any degree of certainty. For networks where the eigenvalue cloud is symmetric, as can be seen for the computer-generated networks considered in this paper, the size of the cloud can be well estimated by looking at the eigenvalue of the largest magnitude with the negative real part. However, for many real networks, as discussed in Sec. II C, the eigenvalue cloud may not be symmetric. For the political books network, we calculated the clustering coefficient given in Ref. [26], which we found to be relatively high (a value of 0.348). For the political blogs network, we found relatively high values of first few odd moments of the spectral density, an order of magnitude higher, compared to the randomly generated scale-free networks with similar degree distribution and two communities. These findings suggest that the clouds are right skewed and should actually extend past the vertical dashed line.

V. LIMITATIONS IN DETERMINING THE NUMBER OF COMMUNITIES

The method we propose in this paper for finding the number of communities works best when the node average degrees within communities are of the same order. Limitation to this method occurs when one or more of the communities are much smaller compared to the largest community or when a community has sparser within community connections compared to other communities. In particular, even in the absence of perturbation [$\delta A=0$ in Eq. (17)], the maximum eigenvalue of the smaller community can lie inside the cloud of non-Perron-Frobenius eigenvalues of the largest community. This puts a limitation on the sizes of the communities that can be detected using our method. For example, in the simplest case where the in and out degrees are uncorrelated and $\delta A=0$, this happens when, $\langle d\rangle_s \le \langle d\rangle_l^{1/2}$, where $\langle d\rangle_s$ is the average degree of a smaller community and $\langle d\rangle_l$ is the average degree of the largest community. In the case of network communities, where the average degree of nodes is proportional to the number of nodes within communities, this condition roughly translates to the statement that when $N_s \le N_l^{1/2}$, we will not be able to detect smaller communities with N_s nodes when the number of nodes in one of the largest community is N_l.

As discussed in Sec. II C, in case of networks that have nonzero odd moments of the spectral density, the cloud of non-Perron-Frobenius eigenvalues may not be symmetric. As can happen in small-world networks without community structure with large clustering, the largest eigenvalue of the network adjacency matrix may not be well separated from the cloud of the non-Perron-Frobenius eigenvalues [12]. In case of networks with community structure, the skewed eigenvalue cloud may even overlap with the largest eigenvalues of the smaller communities. Thus, we may not be able to see them well separated from the eigenvalue cloud.

VI. CONCLUSIONS

We studied the eigenspectra of adjacency matrix of large sparse networks. The eigenspectrum gives a clear indication of the number of "dominant" communities in the networks in certain cases. Here, by dominant we mean the communities whose eigenvalues lie outside the cloud of the non-Perron-Frobenius eigenvalues. We examine the eigenvalues of the network adjacency matrix and infer the number of communities by finding the number of eigenvalues falling outside a typically occurring dense cloud of eigenvalues. For the example of uncorrelated in/out degrees, we argued that there is a large gap between the non-Perron-Frobenius eigenvalues and the Perron-Frobenius eigenvalue. Owing to this large gap (also seen more generally with in/out degree correlation and assortative/disassortative networks), we can determine the number of communities in a network, even when the community structure is not strong.

In this paper, we have not specified exactly the radius of the eigenvalue cloud. While there are results on the spectral density of the eigenvalue cloud for Erdos-Renyi and scale-free undirected networks when the distribution of eigenvalues is symmetric, we still need to deal with the case when the odd moments of the spectral density are nonzero, resulting in an asymmetric eigenvalue distribution.

Finding the number of communities from the eigenvalue plot could be helpful in some community finding algorithms (as in Ref. [27]), where the number of communities is an input to the algorithm. The method has a limitation based on the relative sizes of the communities, and, in general, it may miss smaller or weaker communities (Sec. V). Further limitations in determining the number of communities from the eigenvalue plot may occur when the eigenspectra are highly skewed because of nonzero odd moments.

ACKNOWLEDGMENTS

We thank the anonymous referee for pointing out some important references and for making useful comments that helped improve the paper.

[1] M. Girvan and M. E. J. Newman, Proc. Natl. Acad. Sci. U.S.A. **99**, 7821 (2002).
[2] M. E. J. Newman, Eur. Phys. J. B **38**, 321 (2004).
[3] L. Danon, A. Diaz-Guilera, J. Duch, and A. Arenas, J. Stat. Mech.: Theory Exp. (2005) P09008.
[4] M. E. J. Newman, Proc. Natl. Acad. Sci. U.S.A. **103**, 8577 (2006).
[5] J. Duch and A. Arenas, Phys. Rev. E **72**, 027104 (2005).
[6] R. Guimerà and L. A. N. Amaral, Nature (London) **433**, 895 (2005).
[7] E. A. Leicht and M. E. J. Newman, Phys. Rev. Lett. **100**, 118703 (2008).
[8] L. Donetti and M. A. Munoz, J. Stat. Mech.: Theory Exp. (2004) P10012.
[9] A. Capocci, V. D. P. Servedio, G. Caldarelli, and F. Colaiori, Physica A **352**, 669 (2005).
[10] A. Arenas, A. Diaz-Guilera, and C. J. Perez-Vicente, Phys. Rev. Lett. **96**, 114102 (2006).
[11] J. A. Almendral and A. Diaz-Guilera, New J. Phys. **9**, 187 (2007).
[12] I. J. Farkas, I. Derenyi, A.-L. Barabasi, and T. Vicsek, Phys. Rev. E **64**, 026704 (2001).
[13] K.-I. Goh, B. Kahng, and D. Kim, Phys. Rev. E **64**, 051903 (2001).
[14] M. L. Mehta, *Random Matrices*, 2nd ed. (Academic, New York, 1991).
[15] F. Juhasz, *Algebraic Methods in Graph Theory* (North-Holland, Amsterdam, 1981), pp. 313–316.
[16] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[17] A.-L. Barabasi, R. Albert, and H. Jeong, Physica A **272**, 173 (1999).
[18] A.-L. Barabasi, R. Albert, and H. Jeong, Physica A **281**, 69 (2000).
[19] M. Bauer and O. Golinelli, J. Stat. Phys. **103**, 301 (2001).
[20] C. R. MacCluer, SIAM Rev. **42**, 487 (2000).
[21] J. G. Restrepo, E. Ott, and B. R. Hunt, Phys. Rev. E **76**, 056119 (2007).
[22] G. H. Golub and C. F. V. Loan, *Matrix Computations* (The Johns Hopkins University Press, Baltimore, 1996).
[23] J. G. Restrepo, E. Ott, and B. R. Hunt, Phys. Rev. Lett. **97**, 094102 (2006).
[24] V. Krebs (unpublished); www.orgnet.com
[25] L. A. Adamic and N. Glance, in *The Political Blogosphere and the 2004 U.S. Electron: Divided They Blog*, Proceedings of the 3rd International Workshop on Link discovery (AMC, New York, 2005). p. 36–43.
[26] A. Barrat and M. Weigt, Eur. Phys. J. B **13**, 547 (2000).
[27] S. Chauhan, M. Girvan, and E. Ott (unpublished); e-print arXiv:0911.2735.