Nonparametric resampling of random walks for spectral network clustering

Fabrizio De Vico Fallani,[1] Vincenzo Nicosia,[2] Vito Latora,[2,3] and Mario Chavez[1]
[1]*CNRS UMR-7225, Hôpital de la Pitié-Salpêtrière, Paris, France*
[2]*School of Mathematical Sciences, Queen Mary University of London, Mile End Road, E1 4NS, London, United Kingdom*
[3]*Dipartimento di Fisica e Astronomia, Universitá di Catania, Via S. Sofia 61, 95123, Catania, Italy*
(Received 8 April 2013; revised manuscript received 18 October 2013; published 9 January 2014)

Parametric resampling schemes have been recently introduced in complex network analysis with the aim of assessing the statistical significance of graph clustering and the robustness of community partitions. We propose here a method to replicate structural features of complex networks based on the non-parametric resampling of the transition matrix associated with an unbiased random walk on the graph. We test this bootstrapping technique on synthetic and real-world modular networks and we show that the ensemble of replicates obtained through resampling can be used to improve the performance of standard spectral algorithms for community detection.

DOI: 10.1103/PhysRevE.89.012802 PACS number(s): 89.75.Fb, 02.50.Ga, 05.10.Ln

In the past decade, network science has proven to be a robust and comprehensive framework to investigate, model and understand the structure and function of the complex interaction patterns observed in diverse biological, physical, social, and technological systems [1–4]. One of the most intriguing characteristic of many real-world complex networks is the presence of communities, i.e., tightly knit groups of nodes which exhibit poor connectivity with the rest of the graph [5]. As a matter of fact, many experimental evidences have confirmed that communities are the mesoscale building blocks of complex networks: they usually correspond to functional modules in the brain [6,7], to topical clusters in social and communication networks [8], to metabolic reactions and functional domains in protein interaction networks [9,10], to disciplines and research areas in collaboration networks [5]. Consequently, a lot of effort has been devoted to the identification of efficient algorithms for community detection [11].

A typical problem in complex networks analysis is that a real-world network is just a single observation drawn from an unknown distribution of graphs having certain characteristics [12]. As a consequence, there is no predefined way to assess the statistical variability of any local, mesoscale of global network property, including the presence and composition of communities. A widely used approach to determine the statistical significance of network observables consists in considering random *network ensembles*, i.e., sets of graphs obtained from the original network by keeping fixed some structural properties (e.g., the degree sequence or the clustering coefficient) and rewiring the edges at random [13–15]. In the case of community detection, this approach led to the definition of the modularity function, which quantifies the significance of a given community partition of a graph as the deviation from the average modularity expected in an ensemble of random graphs having the same degree sequence [8]. Another possibility is *parametric bootstrapping*, in which the robustness of a network property is assessed against small perturbations of the graph connectivity [16–18]. This approach relies on the hypothesis that the observed network is representative of a set of graphs (a model) having a certain (*a priori* known) structure. Consequently, the variability of any network observable can be estimated as the deviation from the average of the corresponding model. Many different parametric resampling schemes have been used to assess the robustness of network communities against small connectivity perturbations. However, all these methods require an *a priori* hypothesis about the model to which the network belongs, so that the unbiased statistical assessment of a network partition remains an open challenge [19].

A possible solution to this problem is *non-parametric bootstrapping*, a data-driven technique for providing the statistical confidence of almost any statistical estimate [20,21], based on the generation of repeated observations (replicates) from an unknown population using the available data samples (in our case, a single network) as a starting point. This approach has been successfully employed for several different applications, and in particular to improve the stability and accuracy of clustering algorithms in metric spaces [22].

In this article we propose to use non-parametric bootstrapping to improve the performance of spectral community detection algorithms. The method is based on the construction of replicates of the transition matrix of the network, and on the estimation of an average distance matrix, whose elements correspond to the expected spectral distances between pairs of nodes of the graph, averaged over the ensemble of replicates. Then, the obtained distance matrix is fed into a standard hierarchical clustering algorithm. The idea is that the aggregation of information about different replicates, representative of the unknown ensemble to which the network belongs, should allow to obtain more accurate and robust partitions than the one found from the original observed network [22]. This approach is in the same line of ensemble or consensus clustering methods, which combine several partitions generated by different clustering algorithms—or by different runs of the same algorithm—into a single, more robust partition [17,18,23–25]. We analyze the community partitions obtained by non-parametric bootstrapping in different synthetic and real-world modular networks, and we show that this approach can substantially improve the performances of existing spectral clustering methods.

Spectral clustering for community detection. Let $G(V,E)$ be a connected undirected and unweighted graph with $N = |V|$ nodes and $K = |E|$ edges, and let $A = \{a_{ij}\}$ be the adjacency matrix of G, whose entry $a_{ij} = 1$ if there is an edge connecting

node i and node j, while $a_{ij} = 0$ otherwise. We consider the problem of finding communities of G, i.e., subsets of nodes of G which are more connected internally than with the rest of the graph [8]. Several community detection algorithms are based on mapping each node of G into a point of an appropriate metric space X, so that two nodes i and j having similar structural properties (e.g., similar set of neighbors) are mapped to two points x_i and x_j placed relatively close to each other in X. Then, the nodes are clustered according to the Euclidean distance between their corresponding images in X, so that nodes whose projections are closer in X have a higher probability to be put in the same cluster.

A widely adopted method to map the nodes of a graph into a metric space makes use of spectral properties of matrices associated to G, and in particular of the eigenvectors of the adjacency matrix A or, more frequently, of the transition matrix $P = \{P_{ij}\}$ associated to an unbiased random walk on the graph ($P_{ij} = a_{ij}/k_i$, where $k_i = \sum_j a_{ij}$ is the degree of node i) [6,26–28]. This choice is motivated by the observation that both A and P carry information about the overall structure of the graph. Here we consider the transition matrix P. This matrix is characterized by a set of eigenvalues $\{\lambda_0, \lambda_1, \ldots, \lambda_{N-1}\}$ such that $|\lambda_0| \geqslant |\lambda_1| \geqslant \cdots \geqslant |\lambda_{N-1}|$. Each eigenvalue λ_k is associated to the left and right eigenvectors φ_k and ψ_k, which satisfy $\varphi_k^\intercal P = \lambda_k \varphi_k^\intercal$ and $P\psi_k = \lambda_k \psi_k$, respectively. Thus, it is possible to map node i to the point $x_i \in \mathbb{R}^N$ whose kth coordinate is equal to the the ith component of the kth right eigenvector of P. The distance d_{ij} between two points x_i and x_j, can be written in terms of eigenvectors and eigenvalues of P [29], namely: $d_{ij} = \sqrt{\sum_{k\geqslant 1} \lambda_k^2 (\psi_k(i) - \psi_k(j))^2}$, where $\psi_k(j)$ denotes the jth component of the kth right eigenvector. In general, this distance can be approximated by using only the first β nontrivial eigenvectors and eigenvalues of P [30]:

$$d_{ij} \simeq \sqrt{\sum_{k=1}^{\beta} \lambda_k^2 (\psi_k(i) - \psi_k(j))^2}. \tag{1}$$

The elements $\{d_{ij}\}$ of the matrix D obtained from Eq. (1) represent the distances between each pair of points x_i and x_j in the lower dimensional space $X \equiv \mathbb{R}^\beta$.

Since the terms $\{d_{ij}\}$ represent distances in a metric space, then we can use the matrix D to detect candidate community partitions of G by means of hierarchical clustering, an iterative aggregation algorithm which starts by considering each node as a separate cluster, and successively merges the two clusters having minimal distance according to D [31]. The algorithm stops when all the nodes have been grouped in a single cluster. The hierarchical clustering algorithm produces a dendrogram H, i.e., a tree where each of the $N-1$ internal nodes represents the fusion of two clusters. A horizontal cut of H corresponds to a partition of the graph into a certain number of communities. The quality of each partition $\mathcal{S}$ can be quantified using the modularity function [8], which compares the abundance of edges lying inside each community with respect to a null model. In formula,

$$\mathcal{Q}(\mathcal{S}) = \sum_{s=1}^{N_s} \left[\frac{m_s}{K} - \left(\frac{k_s}{2K} \right)^2 \right], \tag{2}$$

where N_s is the number of clusters in the partition $\mathcal{S}$, K is the total number of edges in the network, m_s is the number of edges between vertices in cluster s, and k_s is the sum of the degrees of the nodes in cluster s. We assume that the best partition in communities of the graph G is the cut of the dendrogram H having maximum modularity.

Clustering through non-parametric bootstrapping. One of the possible limitations of community detection algorithms based on the spectral properties of the transition matrix is that the obtained partition could be sensitive to perturbations of the adjacency matrix of the graph, especially in sparse networks [32–35]. Therefore, we propose to improve the quality of spectral clustering by using information about the average spectral properties of transition matrices obtained by a non-parametric bootstrapping of the observed matrix P.

The authors of Ref. [36] have proposed a generic bootstrap scheme to resample the transition probabilities of a finite state time-invariant Markov chain. Starting from a realization χ of the Markov chain, one constructs the maximum likelihood estimator of the associated transition matrix P as $p_{ij} = \frac{f_{ij}}{f_i}$, where f_{ij} is the observed number of transitions from state i to state j in χ and $f_i = \sum_j f_{ij}$. Then, replicates of the observed transition matrix are obtained by drawing, for each state i, the random variables $\{f_{i1}^*, \ldots, f_{iN}^*\} \sim \text{Multinomial}(f_i; p_{i1}, \ldots, p_{iN})$ according to $\widetilde{P}_{ij} = \frac{f_{ij}^*}{f_i}$. The distribution of $\widetilde{P}$ is then obtained by Monte-Carlo sampling. This approach was shown to be asymptotically valid for approximating the sampling distribution of P [36], and has been also used to assess the confidence intervals of transition probabilities in disease modeling [37].

Since the unbiased random walk on the graph G defined by the transition matrix P is a finite-state time-invariant Markov chain, we can construct a similar resampling scheme in which replicates of the transition matrix P are obtained by randomly drawing the variables $\{f_{i1}^*, \ldots, f_{iN}^*\}$ from a multinomial distribution with probabilities $\{p_{i1}, \ldots, p_{iN}\}$, conditional on the observed degree sequence $\{k_i\}$ of G. It is worth noticing that, in contrast to previous approaches where each link was resampled independently from the others [16,18], here the replicas of the transition probabilities for each node i are drawn from a multinomial distribution, accounting for the observed transitions to other nodes $\{p_{i1}, \ldots, p_{iN}\}$.

Given the transition matrix P of G, we generate B bootstrap transition matrices $\{P_1^*, P_2^*, \ldots, P_B^*\}$. Then, we project each matrix P_b^* into $\mathbb{R}^\beta$ (where β is a tunable parameter), and we estimate the corresponding bootstrap distance matrices D_b^*, whose entry d_{ij}^b is the Euclidean distance between x_i and x_j in $\mathbb{R}^\beta$ according to the mapping induced by P_b^*. Then, we compute the average distance matrix $\widetilde{D}^* = \frac{1}{B}\sum_b D_b^*$, which is expected to be the most consistent (similar) with the central tendency of the population of replicates. The matrix $\widetilde{D}^* = \{\widetilde{d}_{i,j}^*\}$ effectively quantifies the dissimilarity between any pair of vertices of G (the smaller the distance $\widetilde{d}_{i,j}^*$ the more similar are i and j), in terms of the average distance between their projections in $\mathbb{R}^\beta$ across several replicas of P.

We notice that, in principle, the spectral network decomposition based on non-parametric bootstrapping does not rely on modularity, so that any quality function can be used to

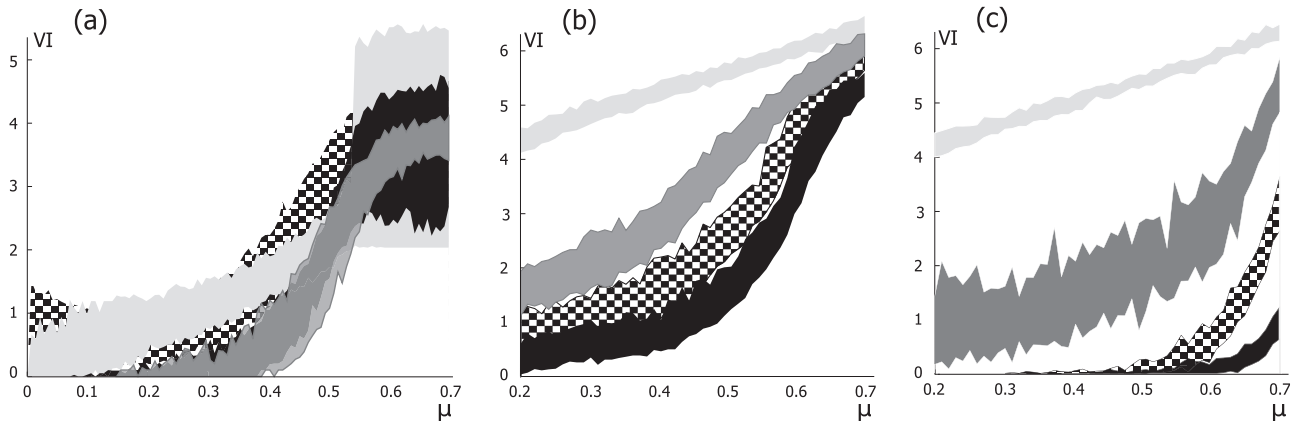

FIG. 1. *Benchmark networks*. The variation of information VI as a function of the proportion of inter-modules links μ in GN graphs (a) and as a function of the mixing parameter μ in LFR500 (b) and LFR2000 (c) graphs. The region inside each curve includes the fifth and the 95th percentiles of VI across R different runs. The four curves in each panel correspond to the optimal partitions obtained using, respectively, the distance matrix D induced by P for $\beta = 1$ (light gray) and $\beta = 10$ (checked pattern), the average distance matrix $\widetilde{D}^*$ over the population of replicates P^* for $\beta = 1$ (black), and modularity optimization on the adjacency matrix A, as described in Refs. [8,40] (dark gray). The network order N, the number of runs R, and the number B of bootstrap realizations for each run are (a) $N = 128$, $R = 100$, and $B = 100$; (b) $N = 500$, $R = 100$, and $B = 100$; and (c) $N = 2000$, $R = 50$, $B = 50$.

determine the best partition in the dendrogram. We also would like to stress that the partitions obtained with resampling-based clustering methods do not necessarily provide the absolute optimum of a given quality function. Instead, non-parametric bootstrapping yields partitions that are the most consistent with the central tendency across different replicates drawn from the same population.

Synthetic networks. We have tested the performance of our approach on two classes of synthetic graphs with tunable modular structure. In the first benchmark (GN), proposed by Girvan and Newman [5], each network consists of $N = 128$ nodes divided into 4 modules of equal size. Pairs of nodes in the same module are connected with probability p_{in}, while nodes belonging to different modules are linked with a probability p_{out}. Parameters are set such that the average degree is kept constant $\langle k \rangle = 16$. By appropriately tuning p_{in} and p_{out} one can set the percentage μ of edges lying between communities. The second class of modular graphs (LFR), proposed by Lancichinetti, Fortunato, and Radicchi [38], accounts for the heterogeneity in the distributions of node degrees and community sizes. In this case, we generated modular graphs with scale-free degree distribution $P(k) \sim k^{-\gamma}$ and community size distribution $P(s) \sim s^{-\eta}$, where $\gamma = 2$ and $\eta = 1$. An appropriate tuning of the model parameters allows to create graphs with a prescribed fraction μ of inter-community edges. We considered graphs having $N = 500$ and $\langle k \rangle = 7$ (LFR500) and graphs with $N = 2000$ nodes and $\langle k \rangle = 28$ (LFR2000).

Since the real partition in communities of these synthetic networks is a-priori known, we can compare the best partition obtained through spectral clustering with the reference one. A widely used measure to compare two different partitions is the variation of information (VI) [39]. In a nutshell, this non-negative metric quantifies how much information is lost and gained in changing from a partition $\mathcal{A}$ to a partition $\mathcal{B}$. It can be estimated by $VI(\mathcal{A},\mathcal{B}) = -\sum_i^{c^{\mathcal{A}}} \sum_j^{c^{\mathcal{B}}} \left(\frac{n_{ij}^{\mathcal{AB}}}{N}\right) \log \frac{n_{ij}^{\mathcal{AB}}}{N} + \frac{n_{ij}^{\mathcal{AB}}}{N} \log \frac{\mathcal{AB}_{ij}/N}{n_i^{\mathcal{A}} n_j^{\mathcal{B}}/N^2}$, where $c^{\mathcal{A}}$ ($c^{\mathcal{B}}$) is the total number of clusters in the partition $\mathcal{A}$ ($\mathcal{B}$), $n_i^{\mathcal{A}}$ ($n_j^{\mathcal{B}}$) is the number of nodes in the ith (jth) cluster of partition $\mathcal{A}$ ($\mathcal{B}$), and $n_{ij}^{\mathcal{AB}}$ is the number of nodes shared by the ith cluster of partition $\mathcal{A}$ and the j^{th} cluster of partition $\mathcal{B}$. Values of VI range from 0, when $\mathcal{A}$ and $\mathcal{B}$ are identical partitions, to $\log N$ when both $\mathcal{A}$ and $\mathcal{B}$ are randomly drawn.

Figure 1 shows the variation of information between the reference partition and the best one obtained through bootstrap-based spectral clustering, as a function of the fraction of inter-community edges μ. The reported results suggest that even when the graphs do not have any more a strong community structure, i.e., when for each node the number of neighbours outside its community is similar with the number of neighbours inside its community, the accuracy of the proposed bootstrap-based method remains pretty high. For GN networks, the accuracy of the bootstrap-based method is comparable to that of a standard modularity optimization algorithm [8,40]. For LFR500 and LFR2000, the non-parametric bootstrap method outperforms the other algorithms, even when we consider an embedding with $\beta = 1$, and exhibits a smaller value of VI up to relatively large values of μ.

The Zachary's karate club network. Figure 2(a) shows the best partition found by the bootstrap-based algorithm ($B = 20000$, $\beta = 1$) in the Zachary's karate club network [41], a paradigmatic example of graph with a strong modular structure. The partition consists of three main modules (black circles, white circles and grey diamonds, respectively) and a small interface community which contains nodes $\{3,9,10,31\}$ (grey ellipses). The distribution of the VI between the partitions obtained through spectral clustering on each single replicate and the one found using $\widetilde{D}^*$ [reported in Fig. 2(b), solid black line] shows that the latter one indeed represents the central tendency of the population of replicates. However, the typical VI between the partition of a single replicate and the one with maximum modularity [Fig. 2(b), grey curve] is higher than that obtained by averaging over all replicates [indicated in Fig. 2(b) by the vertical line]. Notice also that the typical

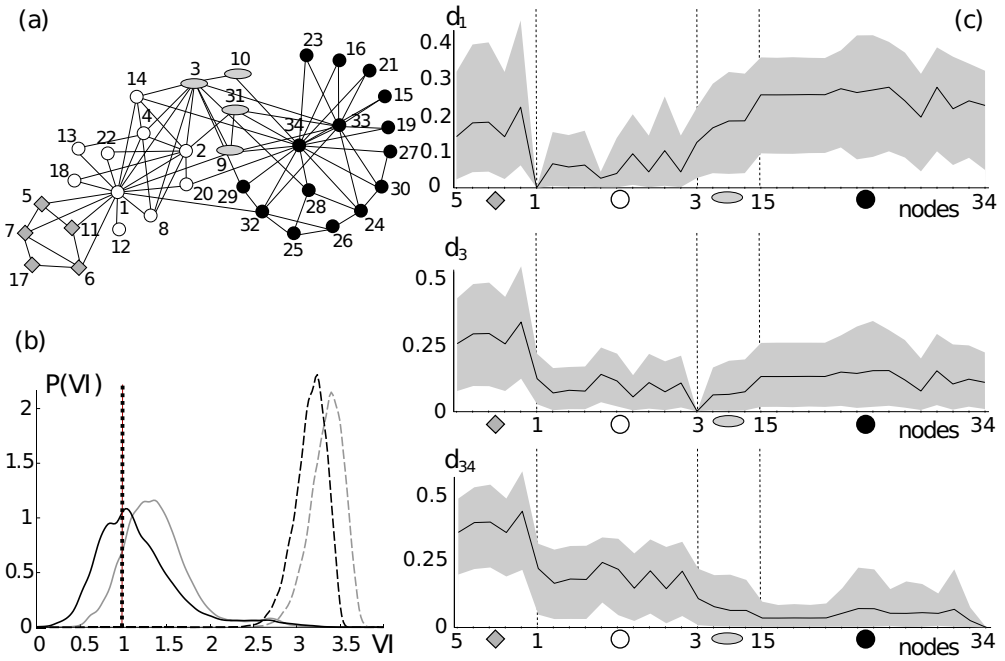

FIG. 2. (a) The best partition of the Zachary Karate club network obtained through non-parametric bootstrap ($B = 20000$ replicas) gives a value of modularity $Q = 0.389$; (b) the distribution of VI across the replicates with respect to the partition induced by $\widetilde{D}^*$ (solid black line) and the partition with maximum modularity (solid grey line); the vertical dotted line indicates the VI between the partition with maximum modularity and the one induced by $\widetilde{D}^*$ ($VI = 0.952$). Dashed lines indicate the distribution of VI for $B = 20000$ random partitions, with respect to the partition induced by $\widetilde{D}^*$ (dashed black line) and the one with maximum modularity (dashed grey line). (c) Spectral distance between node $i = 1$, (top) $i = 3$ (middle), $i = 34$ (bottom), and the rest of the nodes. Gray regions indicate the 0.05th–95th percentiles interval of the bootstrap distribution.

VI between random partitions of the graph and, respectively, the one obtained averaging over all replicates or the one with maximum modularity [respectively the dashed black line and the dashed grey line in Fig. 2(b)] is much larger than that obtained through spectral clustering.

Despite the partition of Fig. 2(a) is not the one with maximum modularity [42], it is worth noticing that most of the nodes put in the interface community (namely, 3, 9, and 10) have been ambiguously classified by many different community detection algorithms [43,44], mostly because assigning them to either the black or the white module has negligible effects on modularity [16]. A more in-depth analysis of the dissimilarity matrix $\widetilde{D}^*$ provides a possible explanation for this fact. In Fig. 2(c) we report the average spectral distance $\widetilde{d}^*_{ij}$ between node 1 (top panel), node 3 (middle panel), and node 34 (bottom panel) and all the other nodes in the graph. As expected, both node 1 and node 34 exhibit a sensibly smaller distance towards the other nodes in their respective natural communities, which is consistent with the fact that 1 and 34 are known to be the centers of these two groups. Conversely, the distance between node 3 and the nodes in the white community is comparable to that between three and the nodes in the black community. This explains why the central tendency of the population of replicates is to place node 3 in a separate community, together with other three nodes having a similar spectral distance pattern.

Concluding remarks. In this work we have shown how the generation of replicates of the transition matrix associated to a graph allows to improve the performance of community detection algorithms based on spectral clustering. In general, we believe that non-parametric bootstrapping techniques, which do not require any assumption about the ensemble of graphs to which a given network belongs, might be successfully employed also to assess the significance of the variability of nodes attributes defined by different random walk parameters (e.g., hitting times or return times), and for the statistical validation of other structural properties defined on different flavors of random walks [45,46].

Acknowledgments. This work was supported by the EU-LASAGNE Project, Contract No. 318132 (STREP). F.D.V.F. is financially supported by the French program "Investissements d'avenir" ANR-10-IAIHU-06.

[1] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. U. Hwang, Phys. Rep. **424**, 175 (2006).
[2] R. Albert and A.-L. Barabasi, Rev. Mod. Phys. **74**, 47 (2002).
[3] A. Barrat, M. Barthlemy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, Cambridge, 2008).
[4] M. Newman, *Networks: An Introduction* (Oxford University Press, Oxford, 2010).
[5] M. Girvan and M. E. J. Newman, Proc. Natl. Acad. Sci. USA **99**, 7821 (2002).
[6] M. Chavez, M. Valencia, V. Navarro, V. Latora, and J. Martinerie, Phys. Rev. Lett. **104**, 118701 (2010).

[7] E. Bullmore and O. Sporns, Nat. Rev. Neurosci. **10**, 186 (2009).
[8] M. E. J. Newman and M. Girvan, Phys. Rev. E **69**, 026113 (2004).
[9] P. F. Jonsson, T. Cavanna, D. Zicha, and P. A. Bates, BMC Bioinf. **7**, 2 (2006).
[10] R. Guimerà and L. A. N. Amaral, Nature (London) **433**, 895 (2005).
[11] S. Fortunato, Phys. Rep. **486**, 75 (2010).
[12] S. Wasserman and K. Faust, *Social Network Analysis* (Cambridge University Press, Cambridge, 1994).
[13] S. Maslov and K. Sneppen, Science **296**, 910 (2002).
[14] E. Ziv, R. Koytcheff, M. Middendorf, and C. Wiggins, Phys. Rev. E **71**, 016110 (2005)
[15] G. Bianconi, P. Pin, and M. Marsili, Proc. Natl. Acad. Sci. USA **106**, 11433 (2009).
[16] D. Gfeller, J. C. Chappelier, and P. De Los Rios, Phys. Rev. E **72**, 056135 (2005)
[17] B. Karrer, E. Levina, and M. E. J. Newman, Phys. Rev. E **77**, 046119 (2008)
[18] M. Rosvall and C. T. Bergstrom, PLoS ONE **5**, e8694 (2010)
[19] A. Mirshahvalad, O. H. Beauchesne, É. Archambault, and M. Rosvall, PLoS ONE **8**, e53943 (2013)
[20] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap* (Chapman and Hall/CRC, New York, 1993).
[21] B. Efron and R. Tibshirani, Statist. Sci. **1**, 54 (1986)
[22] A. Strehl and J. Ghosh, J. Mach. Learn. Res. **3**, 583 (2003).
[23] H. Kwak, Y. Choi, Y.-H. Eom, H. Jeong, and S. Moon, in *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement Conference* (ACM, Chicago, IL, 2009), p. 301.
[24] E.-Y. Kim, D.-U. Hwang, and T.-W. Ko, Phys. Rev. E **85**, 026119 (2012).
[25] A. Lancichinetti and S. Fortunato, Sci. Rep. **2**, 336 (2012).
[26] D. M. Cvetkovic, M. Doob, and H. Sachs, *Spectra of Graphs: Theory and Applications* (Johann Ambrosius Barth Verlag, Heidelberg, 1995).
[27] F. R. K. Chung, *Spectral Graph Theory* (American Mathematical Society, Providence, RI, 1997).
[28] D. Gfeller and P. De Los Rios, Phys. Rev. Lett. **99**, 038701 (2007).
[29] R. R. Coifman and S. Lafon, Appl. Comput. Harmon. Anal. **21**, 5 (2006).
[30] The quality of this approximation depends on the distribution of λ_k. However, empirical evidences suggest that the spectral density of real-world scale-free networks is triangular, with power-law decaying tails (see for instance Ref. [2] for details), so that the spectral distance between any pair of nodes is dominated by the contributions given by eigenvectors associated to the first few largest eigenvalues.
[31] The distance of two clusters C_p and C_q is computed as the average distance between any pair of nodes (i, j) such that $i \in C_p$ and $j \in C_q$. This method is also known as *average linkage*. See Sec. 4.2 in Ref. [11] for details.
[32] A. Capocci, V. Servedio, G. Caldarelli, and F. Colaiori, Physica A **352**, 669 (2005).
[33] F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborová, and P. Zhang, Proc. Natl. Acad. Sci. U.S.A. **110**, 20935 (2013).
[34] M. Mitrović and B. Tadić, Phys. Rev. E **80**, 026123 (2009).
[35] M. E. J. Newman, arXiv:1308.6494.
[36] I. V. Basawa, T. A. Green, W. P. McCormick, and R. L. Taylor, Commun. Stat.-Theory Methods **19**, 1493 (1990).
[37] P. P. Sendi, H. C. Bucher, B. A. Craig, D. Pfluger, and M. Battegay, J. Acquir. Immune Defic. Syndr. Hum. Retrovirol. **20**, 376 (1999).
[38] A. Lancichinetti, S. Fortunato, and F. Radicchi, Phys. Rev. E **78**, 046110 (2008).
[39] M. Meilă, J. Multivariate Anal. **98**, 873 (2007).
[40] M. E. J. Newman, Proc. Natl. Acad. Sci. USA **103**, 8577 (2006).
[41] W. W. Zachary, J. Anthropol. Res. **33**, 452 (1977).
[42] J. Duch and A. Arenas, Phys. Rev. E **72**, 027104 (2005).
[43] D. Li, I. Leyva, J. A. Almendral, I. Sendiña-Nadal, J. M. Buldú, S. Havlin, and S. Boccaletti, Phys. Rev. Lett. **101**, 168701 (2008).
[44] E. Estrada, Chaos **21**, 016103 (2011).
[45] C. Allefeld and S. Bialonski, Phys. Rev. E **76**, 066207 (2007).
[46] R. Sinatra, D. Condorelli, and V. Latora, Phys. Rev. Lett. **105**, 178702 (2010).