POLITECNICO DI TORINO
Repository ISTITUZIONALE

Measuring node centrality when local and global measures overlap

Original
Measuring node centrality when local and global measures overlap / Costantini, L., Sciarra, C., Ridolfi, L., Laio, F.. - In: PHYSICAL REVIEW. E. - ISSN 2470-0045. - 105:4(2022). [10.1103/PhysRevE.105.044317]

Availability:
This version is available at: 11583/2962055 since: 2022-04-26T11:57:06Z

Publisher:
APS

Published
DOI:10.1103/PhysRevE.105.044317

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
AIP postprint versione editoriale con licenza CC BY/Version of Record with CC BY license

Copyright 2022 Author(s). This article is distributed under a Creative Commons Attribution (CC BY) License https://creativecommons.org/licenses/by/4.0/."

(Article begins on next page)

26 November 2026

Measuring node centrality when local and global measures overlap

Lorenzo Costantini,* Carla Sciarra, Luca Ridolfi, and Francesco Laio
DIATI, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129, Turin, Italy

(Received 20 October 2021; revised 2 March 2022; accepted 5 April 2022; published 22 April 2022)

Centrality metrics aim to identify the most relevant nodes in a network. In the literature, a broad set of metrics exists, measuring either local or global centrality characteristics. Nevertheless, when networks exhibit a high spectral gap, the usual global centrality measures typically do not add significant information with respect to the degree, i.e., the simplest local metric. To extract different information from this class of networks, we propose the use of the Generalized Economic Complexity index (GENEPY). Despite its original definition within the economic field, the GENEPY can be easily applied and interpreted on a wide range of networks, characterized by high spectral gap, including monopartite and bipartite network systems. Tests on synthetic and real-world networks show that the GENEPY can shed light about the node centrality, carrying information generally poorly correlated with the node number of direct connections (node degree).

DOI: 10.1103/PhysRevE.105.044317

I. INTRODUCTION

In the last few decades, the use of network theory to unravel systems' features has emerged and spread over different disciplines [1–3], with relevant applications in biology [4], economics [5], epidemics [6], social sciences [7], electrical [8,9], and computer science engineering [10,11]. A network describes the interactions among elements in a system: the elements are called *nodes*, the interactions *edges* or *links* [1–3], and for each system a proper matrix representation exists.

It is crucial to determine the importance of the nodes in a network for understanding the features of the system it represents [12]. Centrality metrics aim to accomplish this task by focusing on the matrix representation of the system. Several centrality measures exist, depending on the chosen criteria to score node importance. Among the most common centrality metrics, one finds the degree and eigenvector centrality [1,2,12]. The degree centrality measures the importance of a node through the number of its connections [1,3], whereas the eigenvector centrality ranks each node accounting for the centrality of the nodes to which it is linked [1,13], and thus considering the entire network structure. In light of such difference of perspective about the network topology, the eigenvector centrality is known to be a global measure of centrality, and the degree a local one, since its centrality analysis is limited to the node neighborhoods [1,12–15]. Other global centrality metrics include closeness [16], betweenness [17], subgraph centrality [18], and total communicability [19], to cite a few.

Notwithstanding the difference in their rationales, high correlation coefficients among the centrality metrics have been observed in high spectral gap networks [14,20–23], namely, those networks presenting a significant difference between the two largest eigenvalues of the matrix representing the system. High spectral gap networks are homogeneous (i.e., with repeated edge patterns) and well connected (i.e., all networks areas can be easily reached moving from one node to the others) [22,24–29]. Several studies have shown that the larger spectral gap, the higher is the Spearman's correlation between both local (hence the degree) and global centrality measures [14,20,30]. In such cases, when comparing the resulting rankings, the degree centrality (i.e., the simplest metrics) is a good proxy for the information provided by the global metrics [21,30]. Therefore, for the networks featuring high spectral gaps, the need emerges for global metrics providing additional information about the node centrality beyond the local one of the degree.

Against this background, we propose to extract such information by applying a centrality measure introduced in the field of economic complexity, namely, the Generalized Economic Complexity index (GENEPY) [31]. The GENEPY metric proved capable of broadening the narrow information that the degree centrality provides about the analysis of the export capacity of countries in the international trade of commodities [31–33]. This feature suggests that the GENEPY framework can be extended to general network systems characterized by high spectral gaps, in order to gain complementary information with respect to the degree centrality. To this aim, after having provided further proof regarding the correlation among common centrality metrics, we present the GENEPY and its application to high spectral gap networks. We refer to bipartite networks in our main results, but these are easily extended also to monopartite ones. In testing the performances of the GENEPY on a wide set of networks, ranging from artificially generated to real-world ones, we show that the information provided by the GENEPY is less related to the degree of the nodes than other metrics, thus shedding light about the node centrality and overcoming the limitations of other metrics.

*Corresponding author: lorenzo.costantini@polito.it

 ©2022 American Physical Society

II. MATHEMATICAL FRAMEWORK

A. Network science definitions and centrality measures

Networks (or graphs) can be mainly classified into two categories: monopartite and bipartite networks. In *monopartite* networks, only one type of interacting entities exist, and the system can either be undirected or directed depending on the presence of directions in the edges. For example, the network representation of the marriage relationships among the Renaissance families in Florence is undirected, since marriage is mutual [1], whereas a directed network suitably represents the citation space of scientific literature [1]. Monopartite networks are described by a square matrix **A**, called an *adjacency* matrix, which is symmetric if the network is undirected, and asymmetric when the graph is directed [1]. The dimension of **A** depends on the number of nodes in the network, and its entries A_{ij} describe the existence of a link between the nodes i and j, either in logical information (1 or 0) or in weighted form. Instead, *bipartite* networks describe systems with two different kinds of interacting entities, and the nodes can be divided into two disjoint subsets U and P. In such a system, the interactions occur only among the elements of different sets [1,34]. The *incidence* matrix **B** describing such systems is rectangular with dimension $N_U \times N_P$, i.e., the number of elements is U and P, respectively. The element B_{ij} is 1 when there exists a link between the node i in U and j in P, and 0 otherwise [1,34]. Examples of bipartite systems are the international trade of goods, where the interacting entities are the exporting countries and corresponding export baskets [32,33], and the plant-pollinator ecosystems, identifying the two sets in the interacting species of flora and fauna [35].

Let $G(V, E)$ be a graph where V and E are the sets of the nodes and edges, respectively (if G is bipartite, then $V = U \cup P$). The neighbors of a node i are all the nodes to which i is connected [1]. The graph is defined as *connected* if, independently from the starting node, it is possible to reach all the other nodes in the network [14,21]. A *walk* of length m is a set of nodes v_1, v_2, ..., v_{m+1} such that for any integer value of r between 1 and m the nodes v_r and v_{r+1} are connected by a (directed) link. A *closed walk* is a walk where $v_1 = v_{m+1}$. A walk without repeated nodes is a *path*, and a closed walk without repeated nodes (except the starting and arriving node) is a *cycle* [14,19,21]. $G'(V', E')$ is a *subgraph* of the graph G if $V' \subseteq V$ and $E' \subseteq E$ [18].

To discuss the correlations among the centrality metrics (including the degree, eigenvector, closeness, betweenness, subgraph centrality, and total communicability) here follows a brief description and interpretation of those centrality metrics. For the sake of conciseness, the description is limited to the general metric rationales, and we refer the reader to [1,2] for further mathematical details (also concerning their tailoring to the different network types). The degree centrality (D) of a node measures its number of connections [1], computed as the sum of all the entries of the matrix referring to that node. The eigenvector centrality (E) scores the nodes according to the number and importance of the nodes to which they are connected. In mathematical terms, the eigenvector centrality equals computing the first eigenvector (the one associated with the largest eigenvalue) of the matrix describing the network at hand [1,13,18,34]. The closeness centrality (C) measures the mean topological distance from one node to the others [1,12,16,30]. The betweenness centrality (B) measures the times a node appears along the shortest paths among vertices [1,15,17]. Supposing a flow of information moving through the system, the betweenness centrality defines the amount of information that moves through a node [12,30]. The subgraph centrality (SG) of the nodes in a network is defined as the weighted sum of all the closed walks (of any length) starting and ending in the same node [18], and the shorter the closed walk, the higher is the associated weight. The subgraph centrality scores the nodes considering their participation in all the subgraphs of the considered network [18]. The total communicability centrality (TC) of a node i is defined as the row sum of the ith row of the exponential of the matrix describing the considered network [19]. Under the assumption of information spreading, this metric quantifies how well a node spreads information to any other node in the network. All the aforementioned metrics, exception made for the degree, consider the entire network structure in the assignment of a centrality score, and hence they are considered as global centrality measures [12–15,19,30,36,37].

The spectral gap, which we will show determines the correlations among these centrality metrics, can be computed for all network kinds. Nevertheless, while the computation of the eigenvalues is straightforward for the square matrices describing undirected networks [21], rectangular matrices need some details. In particular, considering a bipartite network and its incidence matrix **B**, to compute the spectral gap entails considering the existing relationship between the eigenvalues λ of the matrices $\mathbf{B}\mathbf{B}^{\mathrm{T}}$ and $\mathbf{B}^{\mathrm{T}}\mathbf{B}$. The matrices have dimensions $N_U \times N_U$ and $N_P \times N_P$, respectively, and they represent the monopartite projections of the bipartite system. The spectra (i.e., set of all eigenvalues) of these two matrices coincide up to the smallest value n between N_U and N_P [1] (the remaining ones are zero). By ordering the eigenvalues in descending order

$$\lambda_1 > \lambda_2 > \ldots > \lambda_n,$$

we define the relative spectral gap (SG_r) as

$$\mathrm{SG}_r = \frac{\lambda_1 - \lambda_2}{\lambda_1}. \tag{1}$$

The aforementioned definition of the relative spectral gap can be extended to monopartite directed networks substituting **A** for **B**.

B. Artificial bipartite networks with high spectral gap

In their general structure, networks with high spectral gap (typically $\mathrm{SG}_r \geqslant 0.40$ [21]) present: (1) high connectivity among their nodes [22,24–29,38] and (2) high correlation values among several centrality measures [14,20,21,23,39]. Aiming to explore the correlation among different centrality metrics in high spectral gap networks, we artificially generated an ensemble of bipartite networks having this peculiar characteristic. We considered two classes of graphs: (1) networks with a pseudotriangular incidence matrix, which we call pseudotriangular networks (PTNs) [40,41], and (2) bipartite Erdős-Rényi graphs (BERGs) [1,42].

A bipartite network is defined as a PTN if by permuting the rows and columns of its associated incidence matrix **B**, it is possible to identify a separatrix (i.e., a border line) above which the density of links—and, hence, of the nonzero values—is much higher than below it [40]. The separatrix is ideally placed in correspondence of the diagonal of the incidence matrix [40]. Instead, *full triangular matrices*—and corresponding networks—are those matrices in which, with a proper permutation of rows and columns, all the nonzero values lie above the separatrix, while all the zero ones are below it. The matrix describing the international trade of goods [33,41,43] in a given year is an example of a PTN. Other examples include the plant-pollinator system in ecology [44–46], the interbank payment network [47], and the manufacturer-contractor network [48].

For each class, we generated an ensemble of $N_{\text{sim}} = 100$ networks according to the following two procedures. Notice that all generated networks are unweighted (i.e., the weights of the links are all unitary); moreover, in the generation of the networks, we considered only connected graphs.

In defining each network within the PTN ensemble, once the system dimension has been set (i.e., the number of nodes), we started by generating a full triangular incidence matrix, which is then emptied. For each link (i.e., $B_{ij} = 1$), a random edge probability was extracted from a uniform distribution between 0 and 1. The link at hand was maintained if the extracted probability was smaller than a given threshold value β and removed otherwise. In this way, PTNs defined in a wide range of densities were obtained as a function of the threshold value β.

Instead, for the BERG ensemble, starting from an empty matrix with fixed dimensions, we filled the matrix with entries. Using a uniform distribution between 0 and 1, we extracted a random value defining the probability of a link between the node i of set U and j of set P. A network was generated considering the links with a probability value smaller than the threshold value β. In this way, the BERGs also are defined in a wide range of densities, obtained as a function of the threshold value β.

In setting the dimensions of the networks, the elements in U range in the interval $[10, 100]$, with step 15, and those in P in the interval $[100, 500]$ with step 25, thus totaling 119 dimension combinations for the generated artificial network systems. In generating the networks, the link threshold value β for PTNs ranges in the interval $[0.15, 0.85]$ with step 0.10, while for BERG networks, it varies in the range $[0.10, 0.45]$ with step 0.05. The β values are defined in different ranges to generate artificial networks with comparable relative spectral gaps.

These two ensembles define the experimental context wherein analyzing the relationship between spectral gap and the occurrence of correlation among centrality metrics. For each system's dimensions, the PTN and BERG ensembles can be analyzed through the perspective of the threshold value β, determining the density of the networks. In this main text, all results are shown for the ensemble of artificial networks of size 40×250, but similar evaluation and comments arise from the analysis of other network sizes (see Supplemental Material [49], Figs. S1 and S2).

Figure 1 exemplifies two of the generated artificial networks through their incidence matrices, describing a PTN in Fig. 1(a), and a BERG in Fig. 1(d). Figure 1 also shows the Spearman's correlation coefficients between the degree and other centrality metrics (eigenvector, closeness, betweenness, subgraph, and total communicatility) obtained from the analysis of all generated matrices of size 40×250 (the Spearman's correlation metric is a ranking-based one). The correlation is given as a function of the relative spectral gap. As the figure shows, the two ensembles of networks are defined within comparable—although not equal—intervals of relative spectral gap. Moreover, the relative spectral gap is high [46,50,51], and it increases as the threshold value β increases, thus with increasing number of connections in the networks [29] for both PTNs and BERGs. In general, for both network classes, the Spearman's correlation coefficient takes high values, with increasing trend as the relative spectral gap increases (in line with previous studies [14,20]). Therefore, although being global centrality measures, in terms of node ranking the eigenvector, closeness, betweenness, subgraph, and total communicability centrality measures add little information compared to that provided by the degree centrality in both network cases. This result can be further interpreted considering the existing literature. Benzi *et al.* [14] demonstrated that the rankings provided by several centrality measures converge to the ones obtained by applying the degree or eigenvector centrality for high spectral gap networks. As shown in Fig. 1 (as well as in the Supplemental Material [49], Figs. S1 and S2), the rankings obtained from these two metrics also present extremely high Spearman's correlation coefficients. Thus, following the guidelines of Li *et al.* [30], it is possible to approximate the more time-consuming centrality, i.e., the eigenvector one, to simpler one, i.e., the degree, if the correlation between the two metrics is high. Therefore, for the networks at hand, the degree centrality dominates the centrality landscape.

III. BEYOND THE DEGREE: THE GENEPY INDEX

A. Definition

The Generalized Economic Complexity index (GENEPY) was introduced by Sciarra *et al.* [31] to reconcile the contrasting methodologies within the field of economic complexity. It is grounded on the statistical and multidimensional approach to network centrality, which is described in [52]. The methodologies of economic complexity were originally proposed to update the simpler measure of export complexity of countries provided by the degree [32,33]. Following up on the idea of overcoming the limitation of the degree in the economic context, we introduce here the use of the GENEPY to the more general context of high spectral gap networks. The description of the GENEPY framework applied to a general network structure follows, and we refer the reader to [31] for further technical details.

Starting from the incidence matrix of a bipartite network **B**, the GENEPY index is defined by the introduction of a transformation matrix **W**, defined as

$$W_{ij} = \frac{B_{ij}}{k_i k'_j}, \tag{2}$$

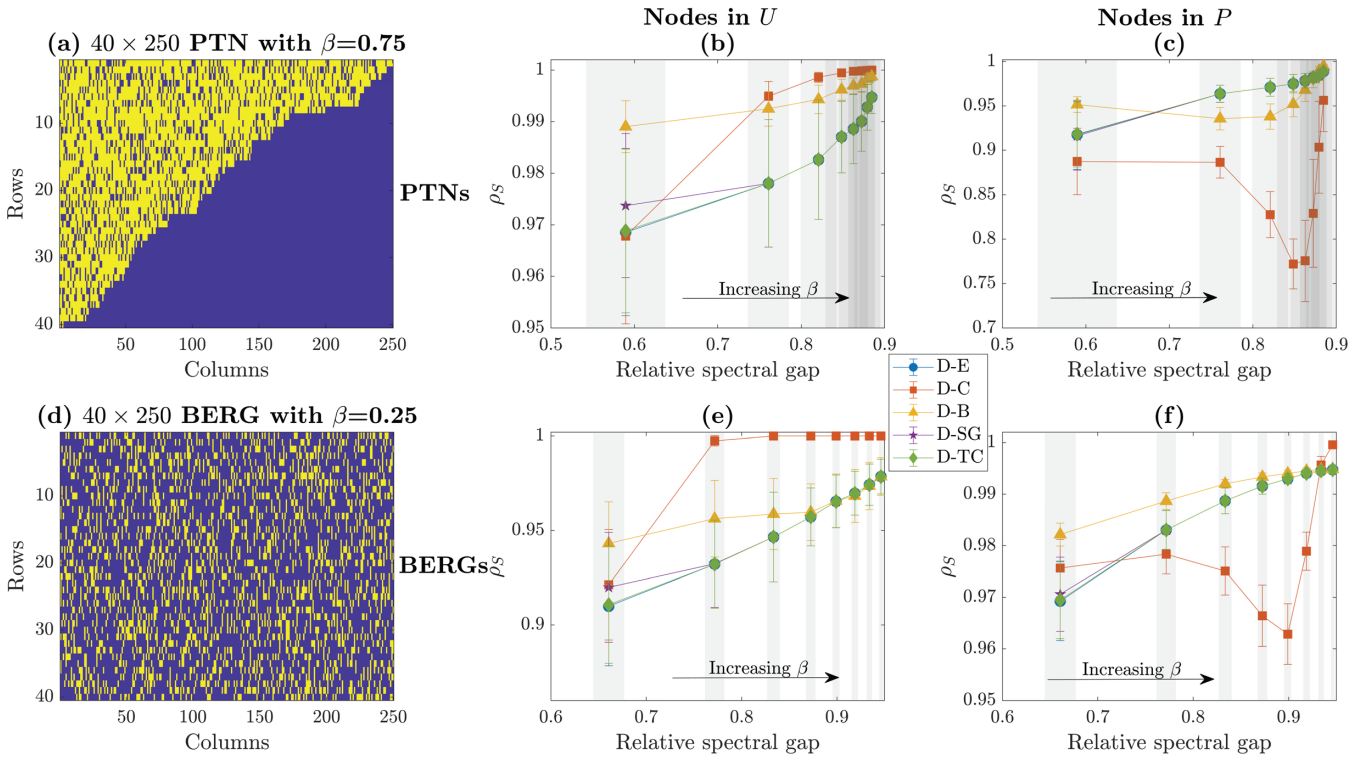


FIG. 1. Examples of two incidence matrices for each class of artificially generated networks of dimension 40 × 250 with a specific value of the parameter β (yellow entries are nonzero values and blue otherwise) and corresponding correlation values between the degree and other centrality metrics as applied to the set of equally sized networks. The Spearman's correlation coefficients (ρ_S) between the degree (D) and the eigenvector (E), closeness (C), betweenness (B), subgraph (SG) centrality, and total communicability (TC) are defined as functions of the relative spectral gap, and the computation of the metrics for both nodes in the U and P sets of the networks is detailed. Top panels refer to pseudotriangular networks (PTNs), and bottom ones to bipartite Erdős-Rényi graphs (BERGs). In panels (b), (c), (e), and (f), each point represents the mean among $N_{\text{sim}} = 100$ networks of the considered class; the whiskers and the shaded regions describe ±1 standard deviation of the correlation values and relative spectral gap, respectively. The arrows indicate the direction in which the threshold values β increase for both PTNs and BERGs.

where $k_i = \sum_{j=1}^{N_P} B_{ij}$ and $k'_j = \sum_{i=1}^{N_U} \frac{B_{ij}}{k_i}$. **W** preserves the network topology **B** describes, with the advantage of partially filtering away the degree-biased information due to the division of the B_{ij} entries for $k_i k'_j$ [53]. The GENEPY index of a generic node in either two sets U and P is the result of the combination of the two largest eigenvalues (ϕ_1 and ϕ_2, with $\phi_1 > \phi_2$) and corresponding eigenvectors ($\mathbf{x}_1$ and $\mathbf{x}_2$) of a *proximity matrix* defined from **W**. For the nodes in U, the proximity matrix **N** is defined as

$$N_{ii^*} = \begin{cases} \sum_j W_{ij} W_{i^* j} & \text{for } i \neq i^*, \\ 0 & \text{for } i = i^*; \end{cases} \tag{3}$$

whereas, for the ones in P, the proximity matrix **G** is defined as

$$G_{jj^*} = \begin{cases} \sum_i W_{ij} W_{ij^*} & \text{for } j \neq j^*, \\ 0 & \text{for } j = j^*. \end{cases} \tag{4}$$

The GENEPY index of a node i is defined as

$$\text{GENEPY}(i) = \left(\sum_{t=1}^{2} \phi_t x_{t,i}^2 \right)^2 + 2 \sum_{t=1}^{2} \phi_t^2 x_{t,i}^2, \tag{5}$$

where $x_{t,i}$ is the ith component of the eigenvector at the dimension $t = 1, 2$ of the proximity matrix at hand (either **N** or **G**). Since **N** and **G** are symmetric square matrices, their eigenvectors are distinctly defined (i.e., the right and left eigenvectors are the same).

The GENEPY index per each set of nodes is derived from the matrices **N** and **G** describing the similarities among the connectivity pattern of the nodes in the network. Such patterns are differently interpreted by each proximity matrix and to compute the GENEPY index for the nodes in U or for those in P provides different information regarding the centrality of the nodes [53]. Specifically, the GENEPY computed onto the proximity matrix **N** ranks the nodes according to two criteria: (1) the number of common neighbors that the considered node shares with other nodes in the network and (2) how many neighbors of the node at hand are also in the neighborhood of only the top ranked nodes (by the GENEPY index computed onto **N**). Instead, the GENEPY computed onto **G** ranks first those nodes that share some neighbors with many other nodes in the network and are also connected to the more isolated nodes. Therefore, we expect the GENEPY computed onto **N** to be better correlated with the degree centrality, since the matrix **N** defines quantitative similarities in the number of connections of the nodes. Conversely, the matrix **G** defines structural similarities in the way the nodes are connected, and thus a lower correlation between the degree of the nodes and the GENEPY index from **G** is expected. To further show the

functioning of the GENEPY index, we provide an example for the bipartite networks in the Appendix, and one considering the monopartite networks in the Supplemental Material [49].

It is worth notice that the matrix **N** (**G**) is the one-mode projection of matrix **W** on U (P), with the diagonal values set to 0. To set the diagonal values to zero entails removing the information about the degree of the nodes, contained in the nominator through the values $\sum_j B_{ij}B_{i^*j}$ for $i^* = i$ [see Eqs. (2)–(4)]. Therefore, setting to 0 these values filters away the information of the degree from the one-mode projection. Another crucial feature of the GENEPY is its multidimensional framework: the first two eigenvectors of the proximity matrices are combined in this metric. They carry different information: $\mathbf{x_1}$ is the eigenvector centrality of the proximity matrix **N** (**G**) that scores the elements in U (P) according to the similarities in their neighborhood (the set of nodes in P connected to a specific node in U and vice versa for the elements in P). In contrast, the second eigenvector $\mathbf{x_2}$ identifies clusters of nodes in U (P) according to their connections [27,31,52,54,55]. A final comment on k'_j is due to the reader. This term describes the degree of an element j in P taking into account the degree and the number of nodes in U to which j is connected. In the transformation from **B** to **N** [or **G**, Eqs. (2)–(4)], the term $\frac{1}{k'_j}$ quantifies how poorly connected is the element j with the nodes in U [31]. As a consequence, the elements of **N** (**G**) associated to a small value of k'_j receive a higher weight. Therefore, the GENEPY ranks the nodes in P from the least to the most central, thus returning an anticentrality score. It follows that, to compare the nodes' importance given by the GENEPY index (anticentrality measure) and the degree (centrality measure) for the nodes in P, the GENEPY rankings should be inverted (e.g., by multiplying each GENEPY value by -1).

B. Tests on artificial networks

Following the correlation analysis of the artificially generated bipartite networks, we also test the performance of the GENEPY index by computing the Spearman's correlation between the proposed metric and the degree, and by comparing it with the degree-eigenvector correlation for both PTNs and BERGs. Other correlation results between the GENEPY and the centrality metrics used in this work can be found in the Supplemental Material, Figs. S3– S7 [49]. Aiming to demonstrate the added value of the use of the GENEPY, we introduce here the Variance Inflation Factor (VIF) [56–58]. The VIF measures the collinearity between two variables (or more than two, in which case the multicollinearity is measured [58]). Collinearity quantifies whether the variable at hand is related through a linear relationship to another one, collinearity being significant for VIF values larger than 5 [57]. The VIF is defined as

$$\mathrm{VIF} = \frac{1}{1-\rho_S^2}, \tag{6}$$

where ρ_S is the Spearman's correlation coefficient between the variables at hand.

We use the VIF to check the collinearity among the rankings obtained by applying the degree, eigenvector centrality, and GENEPY, thus allowing one to quantify the distinct information the GENEPY captures with respect to the degree or eigenvector centrality.

In Fig. 2 we plot the correlation coefficients among the considered centrality measures and the VIF values as a function of the relative spectral gap for PTNs and BERGs (top and bottom panels, respectively) and for both sets U and P (left and right panels, respectively). As the VIF values indicate, the use of the GENEPY in all considered networks (and node sets) allows one to overcome the degree-dominated information. As a general rule of thumb, the VIF from the degree-GENEPY correlation, $\mathrm{VIF}_{D\text{-}GEN}$, is always lower than the one computed on the degree-eigenvector correlation, $\mathrm{VIF}_{D\text{-}E}$. These latter values increase with increasing spectral gap. Although showing the same increasing trends in case of PTNs [Figs. 2(a) and 2(b)], the $\mathrm{VIF}_{D\text{-}GEN}$ values stand below the $\mathrm{VIF}_{D\text{-}E}$ ones. Stable trends as a function of the relative spectral gap can be seen for the BERGs [Figs. 2(c) and 2(d)]. In particular, in these cases the $\mathrm{VIF}_{D\text{-}GEN}$ values are smaller than 5, confirming that the GENEPY rankings are not collinear with those obtained by applying the degree; thus, due to their high correlation with the degree, neither does it with the other centrality metrics (see Figs. S3– S7 of the Supplemental Material [49]).

Looking at the distinction between the two sets of nodes, differences in the correlation values across the sets also emerge [see Figs. 2(a) and 2(c) for the U nodes and Figs. 2(b) and 2(d) for the P nodes] for both the PTN and BERG cases. As expected, higher correlations between the degree and GENEPY are found for the nodes in U rather than for those in P (similar results were obtained for the other network sizes, whose outcomes are given in Figs. S8– S11 in the Supplemental Material [49]).

Notice that the choice of interpreting the nodes in U and P through the matrices **N** and **G**, respectively, is arbitrary in artificial networks since there is no physical meaning of the nodes in the simulations. In fact, we can reverse our interpretation of the system by reversing the analysis of the two sets of nodes, thus computing the GENEPY by transposing the incidence matrix **B** describing the artificial system at hand. Therefore, the sets of nodes U and P would be analyzed by means of the proximity matrices **G** and **N**, respectively. This reversed interpretation of the system cannot be considered in the cases of real-world bipartite systems, where the choice of mapping the U and P node sets in **N** or **G** depends on the nodes' significance in the system. In any case, as we show in Fig. S12 of the Supplemental Material [49], to reverse such an analysis in the artificially generated networks does not change the results presented in Fig. 2. In particular, the GENEPY rankings computed on **N** and **G** are always collinear for PTNs, while they differ for BERGs. However, as shown in Fig. 2, the information provided by the GENEPY differs from the degree or eigenvector centrality.

C. Application to real-world networks

In support of our results on artificially generated networks, we extend the spectral gap-VIF analysis to real-world networks, in both monopartite and bipartite systems. These tests allow one to further comprehend how the GENEPY captures complementary characteristics of the system under study with respect to previously proposed metrics.

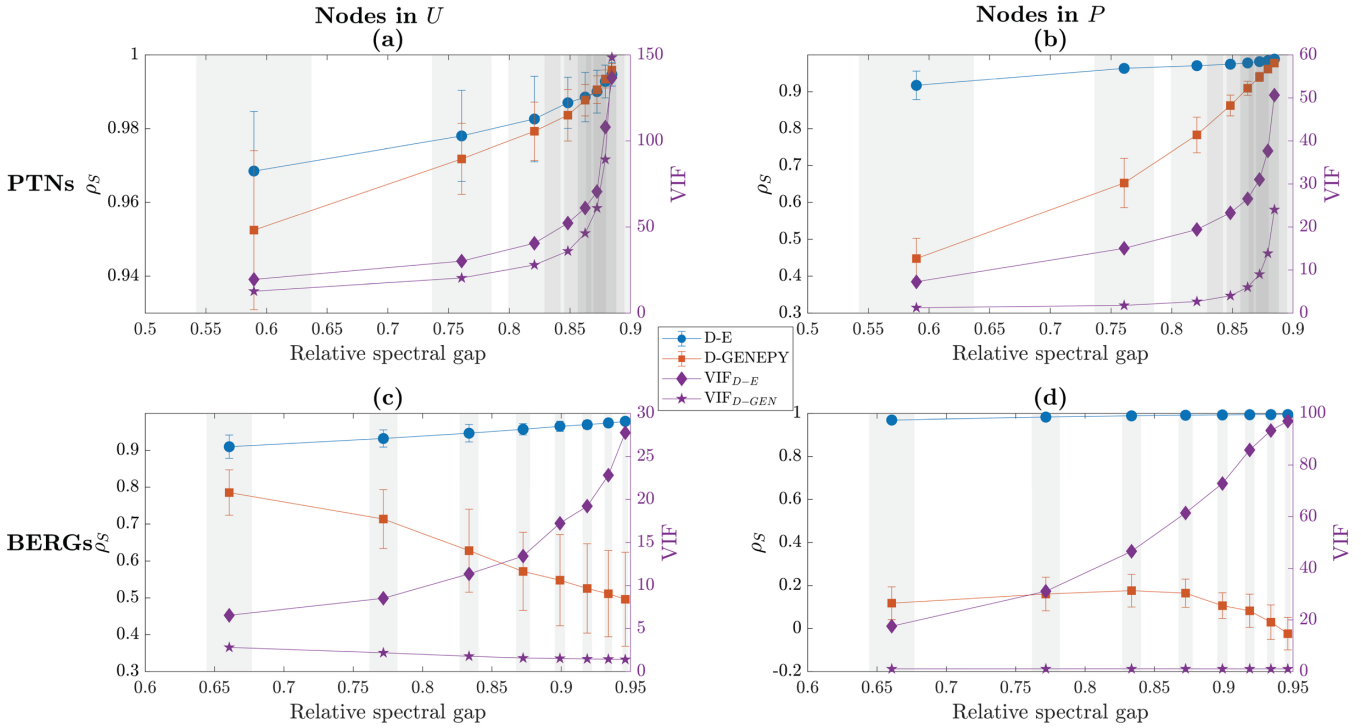


FIG. 2. The Spearman's correlation coefficients among different centrality measures for the pseudotriangular networks (PTNs) (a, b) and bipartite Erdős-Rényi graphs (BERGs) (c, d). The correlation values (ρ_S) between the degree (D) and eigenvector (E) centrality (blue dots), and the degree and GENEPY index (red squares), are shown as functions of the relative spectral gap. The right panels refer to the U set, the left ones to the P set. Each point represents the average value of $N_{\text{sim}} = 100$ realizations, and ± 1 standard deviations are given for both correlation (whiskers) and relative spectral gap (shaded regions) values. In purple, we show the mean variance inflation factor (VIF) computed over the N_{sim} artificial networks between the degree and eigenvector centrality (diamonds) and the degree-GENEPY values (stars).

We start testing the GENEPY performances onto the well-known bipartite system the "Women-Events network," detailing the attendance of women at social events in the late 1930 [34,59]. The system comprises 18 women and 14 events. The relative spectral gap of this network is quite high, 0.54, and the rankings for both kind of entities computed according to the degree, eigenvector, closeness, betweenness, subgraph, and total communicability centrality are collinear, i.e., their VIF values are greater than 5 [see points of the plots in Figs. 3(a) and 3(b) indicated by the vertical black lines]. Since the spectral gap is related to the level of connectivity of the network (as shown by Jun *et al.* [29] and confirmed by our simulations on artificially generated networks), we performed some numerical experiments to further investigate the relationship between centrality metrics and spectral gap. To this aim, we randomly added (and removed) 100 (50) edges one at a time in the incidence matrix of the network. We repeated the link addition and removal process 100 times, thus generating an ensemble of simulations at different number of links in the thus modified network. Therefore, for each ensemble of equal number of links, we evaluated the mean values of the relative spectral gap and the VIF values among the centrality measures considered in this work. Figure 3 presents the results from these simulations for each set of nodes, and it details the VIF values among the degree, GENEPY, and other centrality metrics. As the number of links increases, the relative spectral gap and the VIF values also increase, which is in line with the simulations on the artificially generated networks. In particular, the VIF values computed among the degree and the previously proposed metrics are in general greater than 5.

The comparisons between the GENEPY and the other metrics are presented in Fig. 3(c) for the women and Fig. 3(d) for the events. In both cases, the GENEPY is never collinear to other metrics, highlighting that the information it provides differs from that of the other centrality measures. In computing the GENEPY, the set U details the women (thus the GENEPY is computed onto the proximity matrix $\mathbf{N}$), and the set P the events (i.e., the GENEPY computed onto the proximity matrix $\mathbf{G}$). We can interpret the centrality ranking provided by the GENEPY considering its rationale and the system at hand: a woman is considered more central if she coattends events participating with many other women, some of which exhibit high centrality. Instead, an event is classified as central when it is attended, at the same time, by both more and less socially engaged women, since the participation of high-centrality women increases its centrality score. The corresponding results in correlation are reported in Fig. S13 of the Supplemental Material [49].

Similar results hold for the monopartite example of the Zachary's karate club network [60]. The results, and related comments, are reported in the Supplemental Material, Figs. S14 and S15 [49].

The results on the spectral gap, VIF, and correlation obtained from the Women-Events and Zachary networks are fully confirmed by the tests run on a wide set of real-world network. We considered three different sets of systems: (1)

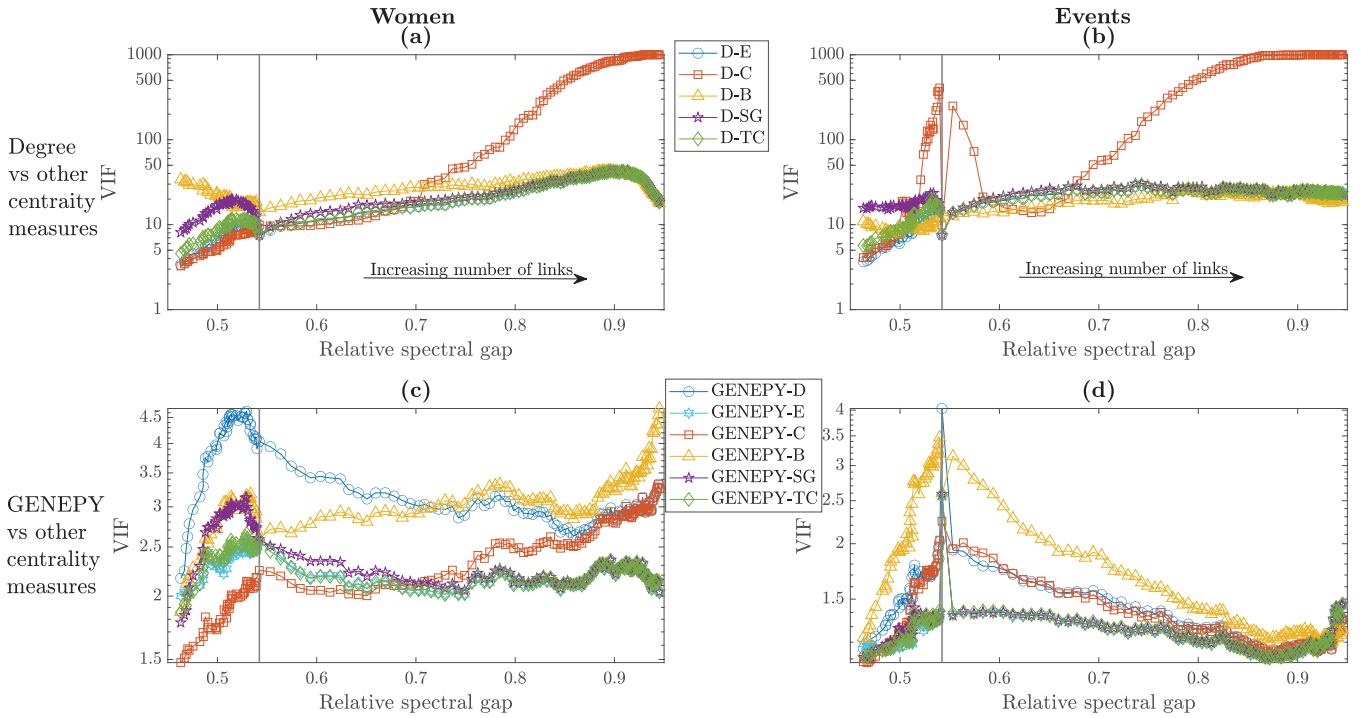


FIG. 3. Values of the VIF computed among the centrality metrics (degree, D; eigenvector, E, blue circles; closeness, C, red squares; betweenness, B, yellow triangles; subgraph centrality, SG, purple pentagrams; and total communicability, TC, green diamonds) applied onto the Women-Events network [34,59] and its simulation with edges' addition and removal. The benchmark for the computation of the VIF values is the degree centrality in the top panels [(a) for the women set and (b) for the events set]; the GENEPY index computed onto **N** in (c); and the GENEPY index computed onto **G** in (d). To modify the spectral gap, 100 (50) links were randomly added (removed), one at a time, to the original network. Each point is the mean, computed at the same number of links in the network, among 100 repetition of the edges' addition and removal process. For the sake of graphical representation, the infinite values of the VIF (corresponding to a Spearman's correlation of 1) were replaced by VIF = 1000 (Spearman's correlation value 0.9995). The black arrows indicate the direction in which the links were added, and the vertical line in each panel highlights the VIF values and relative spectral gap of the original network.

the international trade of goods, taken from the BACI-CEPII dataset [61]; (2) ecological systems, from the Web of Life dataset [62]; and (3) monopartite systems (both undirected and directed) from the SuiteSparse Matrix Collection [63,64]. The variety of these real-world systems allows us to extend the analysis of the GENEPY performances on a broader class of networks, also extending the proposed centrality onto monopartite graphs.

For the sake of comparison with the artificially generated networks, we considered only the networks' topology, thus neglecting any links' weights. In our analysis, we included all connected networks with at least seven nodes in the system. As a result, a total of 284 networks were considered: 23 from the trade systems (years from 1995 to 2017, the nodes in U are in the range $[174, 182]$ and those in P are between 1201 and 1241), 126 ecological networks (nodes in U in the interval $[7, 456]$ while those in P are in the range $[7, 1044]$), and 135 monopartite networks (of which, 104 undirected and 31 directed, with number of nodes between 8 and 6765). The results of these tests are shown in Fig. 4 (the corresponding outcomes in correlation values are in Fig. S16 of the Supplemental Material [49]).

For each centrality metric, the figure shows their VIF values with the GENEPY. The VIF behavior as functions of the relative spectral gap is coherent with what observed for the synthetic graphs and the Women-Events network: the GENEPY index ranks the nodes differently from the metrics previously proposed in network theory. This is also confirmed by the fitting lines reported in the figure; in fact the thick colored lines fitting the VIF values between the GENEPY and the considered centrality metric are always below the thin black lines fitting the degree and the metric indicated in the panel title. These results endorse the outcomes from the analysis on artificial networks about the potential of the GENEPY framework in changing the perspective with respect to the other centrality measures.

Since the plots in Fig. 4 refer to both undirected and directed networks, a further comment is needed to understand the application of the GENEPY framework to monopartite graphs. The GENEPY index provides different results whether the matrix **N** or **G** is used for computation. In particular, due to the similarities of bipartite graphs with directed ones [1], in interpreting the result of the GENEPY framework we associate the analysis related to the set U (i.e., the GENEPY onto **N**) with the outgoing properties of the nodes, and the one for the set P (i.e., the GENEPY onto **G**) with the incoming properties.

As shown in Fig. 4, the lines fitting the VIF values between the degree and the usual centrality measures increase as the the relative spectral gap increases. In particular, over a certain value of the spectral gap, any centrality measure presents collinearity (VIF larger than 5) with the degree. This fact is

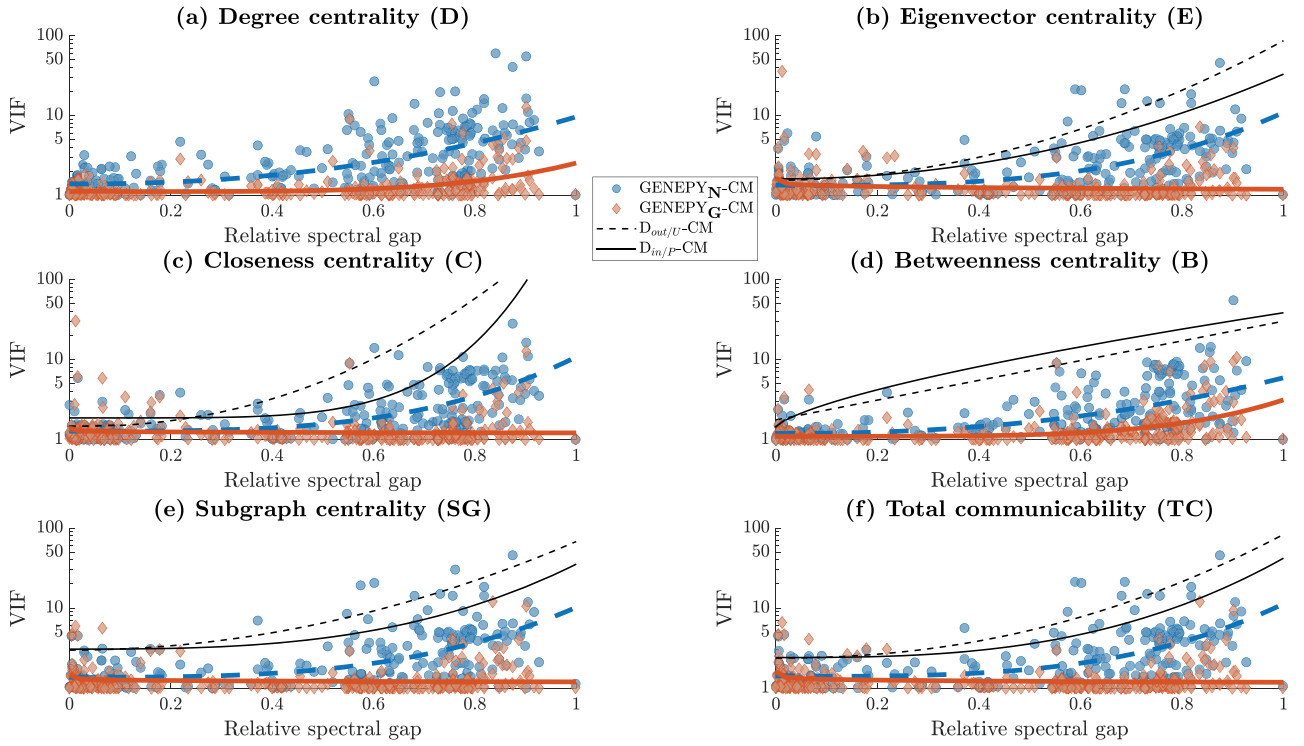


FIG. 4. Comparison of the VIF values between the GENEPY and the degree [D, (a)], eigenvector [E, (b)], closeness [C, (c)], betweenness [B, (d)], subgraph [SG, (e)], and total communicability [TC, (f)] centrality measures as a function of the relative spectral gap for the 284 real-world networks considered in this work. In each panel, the blue circles (red diamond) report the VIF values computed between the GENEPY calculated onto the proximity matrix **N** (**G**) with the Centrality Measure (CM) reported in the panel title. The thick blue dashed (thick red solid) lines fit the logarithm (base 10) of blue circles (red diamonds) with a function $a(\mathrm{SG}_r)^b + c$. The thin dashed and solid black lines describe the fits of VIF values between the degree centrality and the centrality measure reported in the title of each panel. For sake of graphical representation, the y axis is limited to the range [0, 100] to enhance the readability of the points.

in line with the conclusion of Benzi *et al.* [14] for the degree, eigenvector, subgraph, and total communicability centrality measures. In this work, we have enlarged these observations to the closeness and betweenness centrality [see thin black lines in Figs. 4(c) and 4(d), respectively]. Instead, the VIF values between the GENEPY and the other metrics exhibit a different behavior whether they are computed onto **N** or **G**, in any case maintaining a noncollinear node ranking with the other metrics.

The VIF values of the GENEPY computed on **N** are typically larger than those computed on **G**. This is consistent with our previous comments on the rationale behind the GENEPY computed on the proximity matrices **N** and **G**. In fact, because the GENEPY on **N** is related to the number of neighbors a node shares with the others, it is more correlated with the degree (and thus with all the other metrics) than the GENEPY computed on **G**. Furthermore, the fits of the VIF values corresponding to GENEPY on **N** (thick dashed blue lines) exhibit an increasing trend as the spectral gap increases, whereas the behavior of the fits of the VIF values due to the GENEPY on **G** (thick solid red lines) depends on the centrality metric against which the GENEPY is compared. The solid thick lines increase when the GENEPY on **G** is compared with the degree and betweenness centrality, while it remains nearly constant for the other centrality measures. This confirms the fact that GENEPY provides two different centrality perspectives on the nodes in the network in question that are less affected by the spectral gap than the other centrality measures.

IV. CONCLUSIONS

In this work, we propose to use the GENEPY index to shed light on high spectral gap networks. The proposed metric ranks the nodes according to the similarities in their connectivity pattern. We compared the GENEPY performances with other commonly used centrality measures (such as the degree, eigenvector, closeness, and betweenness centrality) onto both synthetic and real-world networks. As we have shown, the previously proposed metrics present high correlation and collinearity with the degree centrality for increasing relative spectral gap. Conversely, the GENEPY metrics and the other centrality measures are less collinear, entailing that the GENEPY can unveil new centrality characteristics of the nodes.

The outcomes of our work demonstrate that the GENEPY centrality offers a complementary approach to the network centrality problem than those previously proposed, which means that the GENEPY has the capability to enlarge the information that can be obtained from an interacting system.

The rankings' correlation analysis is one of the most direct ways to evaluate the differences among several centrality measures, yet other approaches can be employed. For

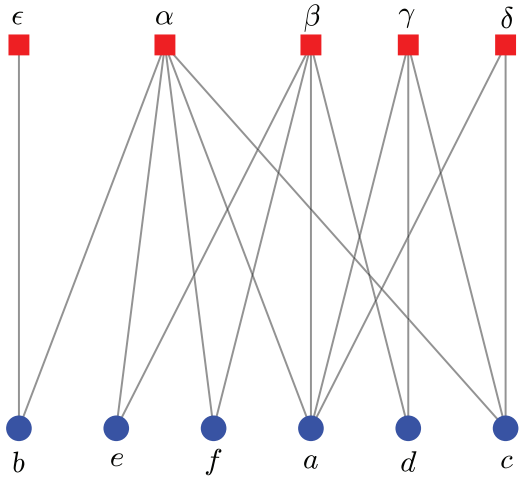


FIG. 5. Representation of the bipartite network used as toy model. The nodes in the set U (indicated with α, β, γ, δ, and ϵ) are represented by red squares, while those in P (a, b, c, d, e, and f) by blue circles. The black solid lines represent the connections among the nodes in the two sets. The nodes in U are ranked according to the GENEPY computed onto **N**, and those in P onto the proximity matrix **G**.

example, trending approaches consider centrality metrics as the explanatory variable in a prediction exercise, where the aim is to proxy a target variable characterizing the system at hand through suitably system-specific functions (e.g., [65,66]). Our results pave the way for future research aimed at comparing the GENEPY and other centrality metrics (in high spectral gap systems) from the predictive capability perspective.

APPENDIX: A TOY MODEL TO EXPLAIN HOW THE GENEPY WORKS

To show how the GENEPY works, let us consider the bipartite network presented in Fig. 5 and described by the following incidence matrix:

$$\mathbf{B} = \begin{array}{c} \\ \alpha \\ \beta \\ \gamma \\ \delta \\ \epsilon \end{array} \begin{array}{c} \begin{array}{cccccc} a & b & c & d & e & f \end{array} \\ \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

This network has a relative spectral gap of 0.77. Aiming at showing the benefits of the proposed metrics, we compare the rankings obtained by applying to this graph the GENEPY index and the degree, eigenvector, closeness, betweenness, subgraph, and total communicability centrality measures.

In Tables I and II, one can see that the rankings obtained from the classical centrality measures (degree, eigenvector, closeness, and betweenness) and those from the eigenproperties of the incidence matrix (subgraph centrality and total communicability) are similar to each other. All metrics agree in ranking the nodes α and the node a for the sets U and P, respectively, as the most central ones in the network. This fact demonstrates that different approaches to solve the centrality problem onto the same network provide about the same information if the spectral gap of the graph is high, as in this case. In contrast, the GENEPY index provides a different node ranking based on the similarities among the connectivity pattern of the nodes, which is described by the proximity matrices **N** and **G** in a different way. Therefore, to compute the GENEPY index onto **N** and **G** provides different information regarding the centrality of the nodes. Specifically, the GENEPY computed onto the proximity matrix **N** considers two aspects of the node connectivity pattern: (1) important nodes have one or more common neighbors (in P) with many nodes of the same set (i.e., U), and (2) relevant nodes share common neighbors only with other important nodes. Considering these two aspects, the node β is the most important element in U, according to the GENEPY centrality, because it is connected to the node a in P, as the nodes α, γ, and δ, and shares neighbors with nodes α and γ in an exclusive way (nodes e, f, and d, respectively). The node α is in the second position of the GENEPY ranking despite having the same number or exclusive connections of the node β. This is

TABLE I. Comparison among the rankings (R) of the nodes in U obtained by applying several centrality measures: degree (D), eigenvector (E), closeness (C), betweenness (B), subgraph centrality (SG), total communicability (TC), and the GENEPY computed onto **N** ($\mathrm{GEN_N}$). The rankings are arranged starting from the most central node (with R = 1) to the least central node (with R = 5). The value in the parentheses indicate the score obtained by applying the considered centrality measure.

R	D	E	C	B	SG	TC	$\mathrm{GEN_N}$
1	$\alpha(5)$	$\alpha(0.64)$	$\alpha(0.06)$	$\alpha(22.7)$	$\alpha(5.98)$	$\alpha(33.0)$	$\beta(0.73)$
2	$\beta(4)$	$\beta(0.53)$	$\beta(0.05)$	$\beta(7.8)$	$\beta(4.71)$	$\beta(26.7)$	$\alpha(0.64)$
3	$\gamma(3)$	$\gamma(0.43)$	$\gamma(0.045)$	$\gamma(4.03)$	$\gamma(3.69)$	$\gamma(22.1)$	$\gamma(0.29)$
4	$\delta(2)$	$\delta(0.34)$	$\delta(0.041)$	$\delta(0.5)$	$\delta(2.73)$	$\delta(17.1)$	$\delta(0.16)$
5	$\epsilon(1)$	$\epsilon(0.07)$	$\epsilon(0.03)$	$\epsilon(0.0)$	$\epsilon(1.60)$	$\epsilon(5.43)$	$\epsilon(0.03)$

TABLE II. Comparison among the rankings (R) of the nodes in P obtained by applying several centrality measures: degree (D), eigenvector (E), closeness (C), betweenness (B), subgraph centrality (SG), total communicability (TC), and the GENEPY computed onto **G** ($\mathrm{GEN_G}$). The rankings are arranged starting from the most central node (with R = 1) to the least central node (with R = 6). The value in the parentheses indicate the score obtained by applying the considered centrality measure. Notice that, for the GENEPY index, the negative scores are given by the recast of the index into a centrality metrics, rather than an anticentrality one.

R	D	E	C	B	SG	TC	$\mathrm{GEN_G}$
1	$a(4)$	$a(0.62)$	$a(0.059)$	$a(10.7)$	$a(5.21)$	$a(30.8)$	$b(-0.019)$
2	$c(3)$	$c(0.45)$	$c(0.053)$	$b(9)$	$c(3.75)$	$c(22.9)$	$c(-0.186)$
3	$b(2)$	$e(0.37)$	$e(0.048)$	$c(5.6)$	$e(2.86)$	$e(18.9)$	$a(-0.318)$
4	$d(2)$	$f(0.37)$	$f(0.048)$	$e(1.77)$	$f(2.86)$	$f(18.9)$	$d(-0.349)$
5	$e(2)$	$d(0.30)$	$b(0.044)$	$f(1.77)$	$d(2.62)$	$d(15.8)$	$e(-0.701)$
6	$f(2)$	$b(0.23)$	$d(0.040)$	$d(1.17)$	$b(2.43)$	$b(13.2)$	$f(-0.701)$

due to the fact that the node α shares the node b as common neighbor with the node ϵ, which is the least central element in U for the GENEPY. Thus, for the GENEPY onto $\mathbf{N}$, the node α loses relevance in the network, and it is below the node β in the ranking. Conversely, the GENEPY centrality computed onto the proximity matrix $\mathbf{G}$ ranks as relevant those nodes that, at the same time, share some neighbors with other nodes (in the same set) and are connected to some more isolated nodes of the other set. Thus, the node b in P is ranked as the most important node; in fact, the node b is connected to the nodes α (as other nodes in P) and ϵ, which is the most isolated node (according to the GENEPY metric).

[1] M. Newman, *Networks* (Oxford University Press, New York, 2018).

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

[3] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, Phys. Rep. **424**, 175 (2006).

[4] E. Alm and A. P. Arkin, Curr. Opin. Struct. Biol. **13**, 193 (2003).

[5] F. Schweitzer, G. Fagiolo, D. Sornette, F. Vega-Redondo, A. Vespignani, and D. R. White, Science **325**, 422 (2009).

[6] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, Rev. Mod. Phys. **87**, 925 (2015).

[7] S. P. Borgatti, A. Mehra, D. J. Brass, and G. Labianca, Science **323**, 892 (2009).

[8] G. A. Pagani and M. Aiello, Physica A **392**, 2688 (2013).

[9] A. E. Motter, S. A. Myers, M. Anghel, and T. Nishikawa, Nat. Phys. **9**, 191 (2013).

[10] W. Liu, C. Liu, Z. Yang, X. Liu, Y. Zhang, and Z. Wei, Commun. Nonlinear Sci. Numer. Simulat. **37**, 249 (2016).

[11] V. Karyotis and M. Khouzani, *Malware Diffusion Models for Modern Complex Networks: Theory and Applications* (Morgan Kaufmann, 2016).

[12] S. P. Borgatti, Social Netw. **27**, 55 (2005).

[13] P. Bonacich, Social Netw. **29**, 555 (2007).

[14] M. Benzi and C. Klymko, SIAM J. Matrix Anal. Appl. **36**, 686 (2015).

[15] A. Salavaty, M. Ramialison, and P. D. Currie, Patterns **1**, 100052 (2020).

[16] L. C. Freeman, D. Roeder, and R. R. Mulholland, Social Netw. **2**, 119 (1979).

[17] L. C. Freeman, Sociometry **40**, 35 (1977).

[18] E. Estrada and J. A. Rodriguez-Velazquez, Phys. Rev. E **71**, 056103 (2005).

[19] M. Benzi and C. Klymko, J. Complex Netw. **1**, 124 (2013).

[20] S. Oldham, B. Fulcher, L. Parkes, A. Arnatkevičiūtė, C. Suo, and A. Fornito, PLoS ONE **14**, e0220061 (2019).

[21] M. Benzi, E. Estrada, and C. Klymko, Linear Algebra Appl. **438**, 2447 (2013).

[22] E. Estrada, EPL (Europhys. Lett.) **73**, 649 (2006).

[23] E. Estrada, Eur. Phys. J. B **52**, 563 (2006).

[24] R. M. Tanner, SIAM J. Algebraic Discrete Meth. **5**, 287 (1984).

[25] N. Alon, Combinatorica **6**, 83 (1986).

[26] S. Hoory, N. Linial, and A. Wigderson, Bull. Am. Math. Soc. **43**, 439 (2006).

[27] E. Estrada, Phys. Rev. E **75**, 016103 (2007).

[28] A. Milanese, J. Sun, and T. Nishikawa, Phys. Rev. E **81**, 046112 (2010).

[29] W. Jun, M. Barahona, T. Yue-Jin, and D. Hong-Zhong, Chin. Phys. Lett. **27**, 078902 (2010).

[30] C. Li, Q. Li, P. Van Mieghem, H. E. Stanley, and H. Wang, Eur. Phys. J. B **88**, 65 (2015).

[31] C. Sciarra, G. Chiarotti, L. Ridolfi, and F. Laio, Nat. Commun. **11**, 1 (2020).

[32] C. A. Hidalgo and R. Hausmann, Proc. Natl. Acad. Sci. USA **106**, 10570 (2009).

[33] A. Tacchella, M. Cristelli, G. Caldarelli, A. Gabrielli, and L. Pietronero, Sci. Rep. **2**, 1 (2012).

[34] S. P. Borgatti and M. G. Everett, Social Netw. **19**, 243 (1997).

[35] S. J. Hegland, J. Dunne, A. Nielsen, and J. Memmott, Biol. Conserv. **143**, 2092 (2010).

[36] P. Bonacich, Social Netw. **13**, 155 (1991).

[37] A. Farahat, T. LoFaro, J. C. Miller, G. Rae, and L. A. Ward, SIAM J. Sci. Comput. **27**, 1181 (2006).

[38] G. Petri, M. Scolamiero, I. Donato, and F. Vaccarino, PLoS ONE **8**, e66506 (2013).

[39] M. Mihail and C. Papadimitriou, in *International Workshop on Randomization and Approximation Techniques in Computer Science*, edited by J. D. P. Rolim and S. Vadhan (Springer, Berlin, Heidelberg, 2002), pp. 254–262.

[40] M. S. Mariani, Z.-M. Ren, J. Bascompte, and C. J. Tessone, Phys. Rep. **813**, 1 (2019).

[41] F. Battiston, M. Cristelli, A. Tacchella, and L. Pietronero, Complexity Econ. **3**, 1 (2014).

[42] P. Erdős and A. Rényi, Publ. Math. Inst. Hung. Acad. Sci **5**, 17 (1960).

[43] R.-J. Wu, G.-Y. Shi, Y.-C. Zhang, and M. S. Mariani, Physica A **460**, 254 (2016).

[44] A. James, J. W. Pitchford, and M. J. Plank, Nature (London) **487**, 227 (2012).

[45] J. Bascompte, P. Jordano, C. J. Melián, and J. M. Olesen, Proc. Natl. Acad. Sci. USA **100**, 9383 (2003).

[46] P. P. Staniczenko, J. C. Kopp, and S. Allesina, Nat. Commun. **4**, 1 (2013).

[47] K. Soramäki, M. L. Bech, J. Arnold, R. J. Glass, and W. E. Beyeler, Physica A **379**, 317 (2007).

[48] S. Saavedra, F. Reed-Tsochas, and B. Uzzi, Nature (London) **457**, 463 (2009).

[49] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.105.044317 for Figs. S1– S16. Refer to Figs. S1– S11 for further details about the correlation among the metrics on artificially generated networks. Figure S12 reports the comparison between the GENEPY computed onto $\mathbf{N}$ and $\mathbf{G}$ for both pseudotriangular networks and bipartite Erdős-Rényi graphs. Figures S13– S16 report the supplemental results for the real-world networks.

[50] J. Friedman, Combinatorica **11**, 331 (1991).

[51] W. Feng and K. Takemoto, Sci. Rep. **4**, 5912 (2014).

[52] C. Sciarra, G. Chiarotti, F. Laio, and L. Ridolfi, Sci. Rep. **8**, 1 (2018).

[53] L. Costantini, C. Sciarra, L. Ridolfi, and F. Laio, in *The 10th International Conference on Complex Networks and Their Applications, Book of Abstracts*, edited by R. M. Benito, H. Cherifi, C. Cherifi, E. Moro, L. M. Rocha, and M. Sales-Pardo (Springer, Madrid, Spain, 2021), pp. 548–551.
[54] P. Mealy, J. D. Farmer, and A. Teytelboym, Sci. Adv. **5**, eaau1705 (2019).
[55] D. Gfeller and P. De Los Rios, Phys. Rev. Lett. **99**, 038701 (2007).
[56] E. R. Mansfield and B. P. Helms, Am. Stat. **36**, 158 (1982).
[57] J. I. Daoud, J. Phys.: Conf. Ser. **949**, 012009 (2017).
[58] A. Alin, WIREs Comp. Stat. **2**, 370 (2010).
[59] A. Davis, B. B. Gardner, and M. R. Gardner, *Deep South: A Social Anthropological Study of Caste and Class* (University of South Carolina Press, Columbia, 2009).
[60] W. W. Zachary, J. Anthropolo. Res. **33**, 452 (1977).
[61] G. Gaulier and S. Zignago, CEPII Working Paper 2010-23, https://ssrn.com/abstract=1994500 (2010).
[62] www.web-of-life.es.
[63] T. A. Davis and Y. Hu, ACM Trans. Math. Softw. (TOMS) **38**, 1 (2011).
[64] https://sparse.tamu.edu.
[65] D. Hartmann, M. R. Guevara, C. Jara-Figueroa, M. Aristarán, and C. A. Hidalgo, World Dev. **93**, 75 (2017).
[66] R. Albert, Plant Cell **19**, 3327 (2007).