Low-rank network decomposition reveals structural characteristics of small-world networks

Victor J. Barranca,[1,*] Douglas Zhou,[2,†] and David Cai[2,3,4,‡]

[1]*Department of Mathematics and Statistics, Swarthmore College, 500 College Avenue, Swarthmore, Pennsylvania 19081, USA*

[2]*Department of Mathematics, MOE-LSC, and Institute of Natural Sciences, Shanghai Jiao Tong University, Dong Chuan Road 800, Shanghai 200240, China*

[3]*Courant Institute of Mathematical Sciences and Center for Neural Science, New York University, 251 Mercer Street, New York, New York 10012, USA*

[4]*NYUAD Institute, New York University, Abu Dhabi, P.O. Box 129188, Abu Dhabi, UAE*

Small-world networks occur naturally throughout biological, technological, and social systems. With their prevalence, it is particularly important to prudently identify small-world networks and further characterize their unique connection structure with respect to network function. In this work we develop a formalism for classifying networks and identifying small-world structure using a decomposition of network connectivity matrices into low-rank and sparse components, corresponding to connections within clusters of highly connected nodes and sparse interconnections between clusters, respectively. We show that the network decomposition is independent of node indexing and define associated bounded measures of connectivity structure, which provide insight into the clustering and regularity of network connections. While many existing network characterizations rely on constructing benchmark networks for comparison or fail to describe the structural properties of relatively densely connected networks, our classification relies only on the intrinsic network structure and is quite robust with respect to changes in connection density, producing stable results across network realizations. Using this framework, we analyze several real-world networks and reveal new structural properties, which are often indiscernible by previously established characterizations of network connectivity.

PACS number(s): 89.75.Fb, 89.75.Hc, 87.85.dq

I. INTRODUCTION

Network models are ubiquitous in characterizing the topology of a wide array of real-world structures that incorporate interacting agents, including author collaborations, neural circuitry, human friendships, and protein synthesis [1–9]. Depending on their connectivity features, these networks can fall into different classes described by their unique graph-theoretic structures [10–12]. In particular, networks with small-world structure are naturally found across a myriad of social, biological, and technological systems [13,14]. Small-world networks exhibit a high degree of clustering and a small average path length between agents, or nodes, possessing advantageous properties of both regularly and randomly connected networks. In light of these properties, small-world networks are particularly efficient in quickly transmitting information at a low cost and therefore often arise in biological as well as engineered systems [13,15].

While simply computing the average clustering coefficient and path length in a network gives an indication of small-worldness, these properties often cannot fairly compare networks of different sizes or connection densities [10,12,16]. For example, densely connected networks trivially have short path lengths and high clustering coefficients, thus making small-world characteristics unable to well distinguish among their connectivity structures and also unable to fairly compare networks of different sizes. Nevertheless, with the increasing prevalence of network models utilized across disciplines, characterizing network connectivity structure beyond classical notions may be key to understanding the relationship between network structure and function. For example, the small-world structure has been observed on both small and large scales in neuronal networks in the brain and has been demonstrated to impact the short-term memory and synchronizability of these networks [17–21]. Similarly, in the case of social networks, the small-world architecture facilitates rapid spread of infectious diseases or information across time-dependent interactions, with high tolerance even in the presence of random attacks [10,13,15]. While small-world networks were originally required to be sparsely connected, since the average number of connections for each node was assumed far less than the size of the network, it is also important to develop tools to characterize the connectivity structure of densely connected networks, which often arise in natural systems [22–26].

The network adjacency (connectivity) matrix provides an informative and computationally efficient graph-theoretic description of both network structure and dynamics [27–29]. However, rather than directly considering the adjacency matrix of a given network, conventional characterizations of small-world properties typically rely on constructing benchmark networks for structural comparison, which can vary broadly in structure across realizations, and thus there remains an important theoretical question of whether alternative measures of small-worldness may rely only on intrinsic network properties [16,30]. In this work, we introduce a novel method of quantifying the structural properties of networks, thereby describing their small-worldness, using the framework of the network adjacency matrix and its low-rank decomposition structure. Low-rank decomposition provides a means of separating a matrix into sparse and

---
*vbarran1@swarthmore.edu

†zdz@sjtu.edu.cn

‡cai@cims.nyu.edu

low-rank components by solving a convex optimization problem [31,32]. Since its discovery, low-rank decomposition has proven useful in numerous applications, such as facial recognition, video surveillance, and matrix completion [33–36].

We develop a method of extending low-rank matrix decompositions to a large set of network connectivity matrices and then use this decomposition to characterize the general structure of networks, e.g., the small-world properties. In addition, our methodology is generalizable to networks varying in both size and connection density. We show that small-world networks may be described by a sum of two matrix components, with each consistently encompassing structurally distinct sets of connections. Generally, the low-rank matrix component captures the highly clustered connections among nodes within clusters, whereas the sparse matrix component captures the relatively few interconnections among clusters. With this intuitive structural partition, we provide a useful technique for separating connections within clusters of highly connected nodes from sparse interconnections between clusters, which holds regardless of node indexing. In many applications, such clusters often form distinct communities, characterized by common functional properties, which are useful for explaining network structure-function relationships [37–40]. Considering that community detection is still relatively challenging for densely connected networks using many conventional techniques [41–45], our low-rank decomposition provides another perspective for identifying sets of community connections in both sparse and dense networks. Unlike conventional measures of small-world properties, the low-rank characterization solves a convex optimization problem and does not require comparison with a constructed benchmark network. Applying our low-rank framework to several diverse real-world data sets, we show that our characterization indeed well measures the small-worldness of classical networks and gives insight into new structural properties not identified by alternative characterizations.

The remainder of the paper is organized as follows. In Sec. II, we briefly review several pertinent aspects of network theory and then formulate our methodology for classifying small-world properties. We introduce in Sec. II A the traditional notions of small-worldness and compare several recent metrics that are commonly used to quantify small-worldness. Moreover, we describe the basic theory of low-rank matrix decompositions, motivating our decomposition technique for network connectivity matrices. Next, in Sec. II B, we formulate our network decomposition and classification framework. We demonstrate in Sec. III the robustness of this new network description, studying the scaling properties and stability of the decomposition. In Sec. IV, we apply our methodology to a diverse set of real-world networks and compare our classification to several conventional small-world metrics. In Sec. V, we discuss implications and possible extensions of this work. Finally, in the Appendix, we show that the network low-rank decomposition gives a consistent method of separating clustered connections from sparse interconnections among clusters regardless of node indexing. We also give details on the augmented Lagrangian method often used in the low-rank decomposition in the Appendix.

II. METHODS

A. Small-world networks and low-rank network decomposition

By considering a network, we are referring to a graphical system composed of a set of nodes interconnected through edges. In this way, the set of all edges between nodes can be represented by an adjacency matrix, $A = (A_{ij})$, with entries determined by the existence and weight of connections between nodes. Assuming the entire structure of the network connectivity is considered, the matrix A is guaranteed square and of size $n \times n$ for a network with n nodes. With respect to the magnitude of adjacency matrix entries, in the case of an unweighted network, if a directed edge connects node j to node i, then $A_{ij} = 1$, and otherwise $A_{ij} = 0$. However, if the network is instead weighted, then connections between nodes are assigned a numeric weight determined by the strength of the connection.

Networks with small-world properties are commonly described through several characterizations of connectivity, which are most notably the average path length and the clustering coefficient. The average path length, l, is the mean shortest path length over all possible pairs of nodes, such that the distance of any path is the sum of the weights corresponding to each edge traversed along the path. Moreover, the clustering coefficient measures the tendency for nodes to form closely connected groups. The clustering coefficient of node i in an unweighted and undirected network is defined by $C_i = (2e_i)/(k_i(k_i - 1))$, where e_i is the number of edges between the neighbors of the node i and k_i is the degree of node i. Likewise, the network clustering coefficient, C, is the average of the clustering coefficients of all individual nodes.

A network with a small value of l and a large value of C allows for quick communication between nodes and therefore rapid spread of information. Networks with this desirable structure were first classified by Watts and Strogatz as small-world networks [13]. In the original definition of a small-world network, it is also required that $1 \ll \ln(n) \ll k \ll n$ for a network with n nodes and mean degree k. The condition that $1 \ll \ln(n) \ll k$ guarantees a randomly connected network is almost fully connected, with a set of edges composing a path between any two nodes in a network. Since $k \ll n$, such a small-world network is also sparsely connected, with each node making a small number of connections compared to the size of the network.

Using the Watts and Strogatz (WS) network construction mechanism, it is possible to build a small-world network by gradually adding more randomness to a regular graph through random rewirings, effectively interpolating between a regular and random network. With a suitable number of rewirings, for approximately $0.01 < p < 0.1$, where p is the rewiring probability for each edge, there is a sufficient number of "shortcuts," or edges connecting distant nodes in the network, such that the average path length of the network is low. In addition, there is still a sufficient number of clusters of highly connected nodes remaining from the ring lattice structure of the original regular graph such that the clustering coefficient is high. Thus, the constructed network achieves the small-world property of both small average path length and high clustering coefficient. For ease of later discussion, in Fig. 1(a), we plot the average path length and clustering coefficient for WS networks
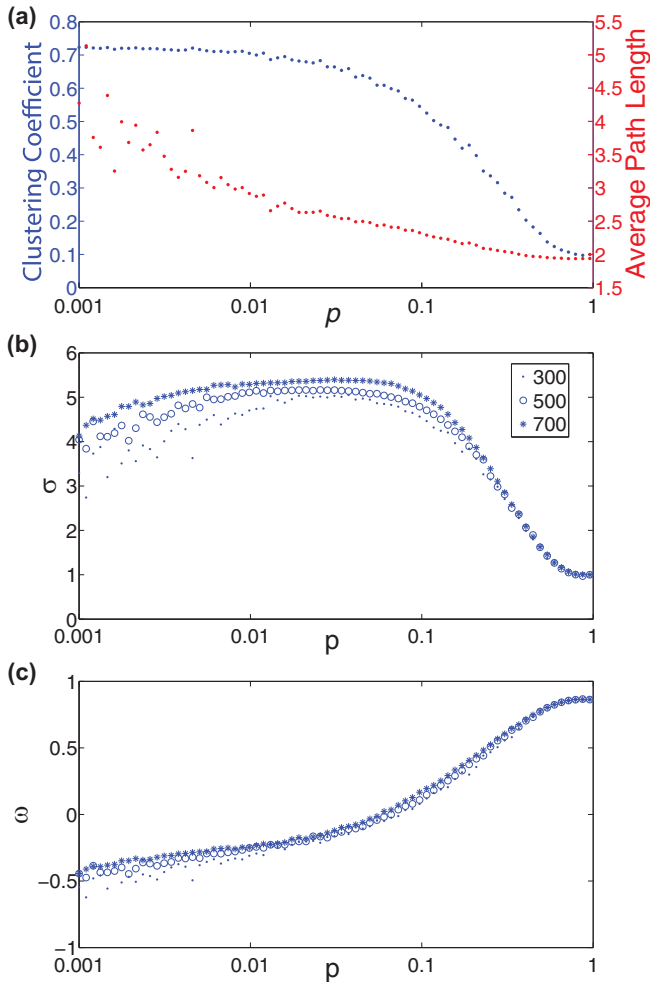
FIG. 1. (Color online) Small-world network characteristics. (a) Average path length (red dots, lower curve) and clustering coefficient (blue dots, upper curve) as a function of rewiring probability, p, for WS networks of 300 nodes with mean degree 30. (b) Small-world characterization σ as a function of p for networks of sizes $n = 300$, 500, and 700 nodes with average degree $0.1n$. (c) Small-world characterization ω as a function of p for the same networks as in (b).

as a function of rewiring probability p. Note that for very small p the networks are more similar to a regular graph with a high clustering coefficient and large average path length, whereas for a large p the networks resemble a random graph with a low clustering coefficient and small average path length.

To answer the question of how small-world a network really is, several parameters have been introduced, focusing on specific combinations of average path length and clustering coefficient. Comparing these statistics for a given network to those of a theoretical randomly connected network with the same number of nodes and average degree, Ref. [16] proposed the measure of small-worldness

$$\sigma = \frac{C/C_{\mathrm{rand}}}{l/l_{\mathrm{rand}}}, \qquad (1)$$

where C_{rand} and l_{rand} are the clustering coefficient and average path length of the random network, respectively. While networks with $\sigma > 1$ are typically considered small world

in Ref. [16], it is important to note that this particular classification may result in an overly loose notion of small-worldness, especially for larger or more densely connected networks, and therefore a larger choice of σ threshold may better agree with the small-world regime, suggested, for example, by Fig. 1(a). In addition, since σ is quite sensitive to changes in C_{rand} and is theoretically unbounded, it is difficult to fairly compare small-world properties of diverse networks using σ alone. These limitations inspired the formulation of a new small-world characterization in Ref. [30]. This alternative description, simultaneously comparing the network statistics to both random and regular networks, is defined as

$$\omega = \frac{l_{\mathrm{rand}}}{l} - \frac{C}{C_{\mathrm{reg}}}, \qquad (2)$$

where C_{reg} is the network clustering coefficient for a regular graph with the same number of nodes and mean degree as the network considered. The parameter ω exists on a bounded interval $-1 \leqslant \omega \leqslant 1$, with positive values reflecting more randomness and negative values a more regular graph. Hence, for a small-world network $\omega \approx 0$. In Figs. 1(b) and 1(c), we plot the dependence of σ and ω respectively for WS networks of various sizes, fixing the mean degree at $n/10$. We note, for a fixed value of p, a relatively large increase in σ and only a slight variation in ω with increasing network size.

Freeing network classifications from comparisons with benchmark regular and random networks, we address the issue of whether small-worldness can be directly determined from the intrinsic network structure. In addition, considering that more densely connected networks trivially exhibit short path lengths and high clustering coefficients, therefore making it difficult to compare their connectivity structure, we seek to provide a characterization of small-worldness that is able to well differentiate among the structure of networks with diverse connection densities. To achieve this, we utilize the framework of low-rank matrix decomposition to further understand network properties.

We motivate our use of low-rank decomposition in analyzing network structure with several observations regarding adjacency matrices for networks with both random and regular characteristics. First, we note that for an n-node unweighted network composed of m disjoint cliques, or sets of vertices in which every possible pair of nodes is connected, the corresponding adjacency matrix, A, will have rank m. Once indexed such that all nodes in each clique are numbered in sequential order, each clique will contribute a block of ones along the diagonal of A and each block will increase the rank of A by 1. If the cliques are large, then $\mathrm{rank}(A) = m < n$, and thus A will be of low rank. Second, if adjacency matrix A has a large set of clusters of highly connected nodes and relatively few interconnections between distinct clusters, as found in a small-world network, then A can be re-expressed as a sum of two matrices, namely one component, L, which contains the connections within clusters, and another component, S, which contains the sparse interconnections between clusters. Combining these two observations, in a network with a small-world-like structure, L will be of relatively low rank and S will be sparse, yielding a low-rank network connectivity matrix decomposition with a natural topological interpretation.

In performing a low-rank decomposition, the aim is to split a given matrix, A, into the sum of two matrices

$$A = L + S, \tag{3}$$

where L is of low rank and S is a sparse matrix containing mostly entries of 0 value. In general, optimizing both the rank of L and sparsity of S is an NP-hard problem, but convex relaxations, such as the principal component pursuit (PCP), converge to an equivalent decomposition while reducing the computational cost dramatically for a large class of matrices [31,33]. This PCP surrogate convex optimization problem is

$$\text{minimize } \|L\|_* + \lambda \|S\|_1, \tag{4a}$$

$$\text{given } L + S = A, \tag{4b}$$

where the nuclear norm $\|L\|_* = \sum_i \sigma_i(L)$ is the sum of all singular values, σ_i, of L and $\|S\|_1 = \sum_{i,j} |S_{ij}|$ is the sum of the absolute values of the elements of S [31]. Intuitively, the number of nonzero singular values indicates the rank of L and the number of nonzero values in S indicates its sparsity. In addition, λ is the sparsity penalization parameter, with values $0 \leqslant \lambda \leqslant 1$, used to balance the minimization of these two terms, such that larger λ requires more sparse S. For a broad class of matrices, A, which are not simultaneously sparse and low rank, there is with large probability a unique solution given that L is not sparse and the sparsity pattern of S is sufficiently random [31,32]. The optimization problem (4) can be solved using a variety of algorithms, including singular value thresholding, augmented Lagrangian, and proximal gradient methods [36,46–48].

B. Low-rank network decomposition method and classification procedure

In this section, we describe a general procedure for decomposing unweighted network connectivity matrices and then using the decomposition to understand the network structure. For an n-node network with connectivity matrix A, we first use low-rank decomposition to compute the low-rank component, $\hat{L}$, and the sparse component, $\hat{S}$, such that $A = \hat{L} + \hat{S}$. In our computations, we use the augmented Lagrangian method to compute the decomposition. Upon determining $\hat{L}$ and $\hat{S}$, we then process each matrix so all entries are either 0s or 1s, reflecting the unweighted nature of the network. To do this, we choose a threshold $\Theta > 0$ and define a processing function, $F : \mathbb{R}^{n \times n} \to \mathbb{R}^{n \times n}$, such that for each entry of $n \times n$ matrix, B,

$$F_{ij}(B) = \begin{cases} 1, & \text{if } B_{ij} > \Theta \\ 0, & \text{otherwise} \end{cases}. \tag{5}$$

To complete the processing, we choose $L = F(\hat{L})$ and $S = F(\hat{S})$, which generally yields a very good recovery of the original adjacency matrix, A, as will be discussed below. Note that for a weighted network, this processing procedure could be easily extended, such as by rounding entries of each matrix to the nearest appropriate weight. However, in this work, we concentrate on cases where networks are treated as unweighted in order to compare their structural properties with respect to WS networks. In our simulations, we choose $\Theta = 0.6$, and since the majority of entries of both $\hat{L}$ and $\hat{S}$ are either 0 or

1, the results are quite insensitive to perturbations in Θ for approximately $0.5 \leqslant \Theta \leqslant 0.7$.

In characterizing a given network, we analyze the properties of each component in the decomposition. For the low-rank component, we compute the normalized rank, $\nu(L)$, which we define for an n-node network as $\nu(L) = \text{Rank}(L)/n$. Hence, $0 \leqslant \nu(L) \leqslant 1$, with higher ν for matrices with higher rank. Similarly, we use the density of the sparse component, namely $\Sigma(S)$, to quantify the percentage of components of S that are nonzero. In this way, $\Sigma(S) = 1$ for a fully connected (with self connections included) network. Moreover, for all decompositions, we choose our sparsity penalization parameter to be

$$\lambda = \frac{1}{\sqrt{\Sigma(A)n}}, \tag{6}$$

where $\Sigma(A)$ is the density of A. The proportionality of λ to $1/\sqrt{n}$ was established by Candés, reflecting that as the network size grows, the penalization of each sparse component should decrease to avoid overpenalizing sparse connections in larger networks [33]. We also multiply by a factor of $1/\sqrt{\Sigma(A)}$ so connectivity matrices with fewer connections are penalized more heavily, preserving the approximate amount of penalization of S across differing connection densities of A. In the absence of this additional penalization factor, networks with sparse connectivity, for example, may result in S capturing all network connections regardless of any distinct network topology. As discussed below, this new choice of λ allows for successful separation of network connections into components L and S, with each capturing connections based on distinct topological features regardless of potential differences in connection density.

In Fig. 2, we construct a set of networks interpolating between regular and random connectivity by using the Watts and Strogatz network construction and varying the rewiring probability p. We plot the dependence of both the normalized rank of L and density of S on p for sparsely connected networks of size 300 nodes with mean degree 30. We observe a clear local minimum in the normalized rank of L, $\nu(L)$, approximately in the small-world regime. In this same regime, we also note a slow increase in $\Sigma(S)$, which is still quite small relative to the case of randomlike networks corresponding to high p. Outside the small-world regime, in the case of more regularly connected or moderately randomly connected networks, we observe that $\nu(L)$ saturates at a relatively high value, near 1. For more regular connectivity, $\Sigma(S)$ is very low, near 0, since L contains almost all network connections. In regimes where S contains nearly no connections, it consequentially has low rank while L is of high rank, contrary to the intuition of the decomposition [see insets in Fig. 2(a) and 2(b) for plots of $\Sigma(L)$ and $\nu(S)$, respectively]. For networks with significantly more random connectivity, the S component instead captures a relatively large number of connections, even more so than L, thereby decreasing $\nu(L)$ while the normalized rank of S, $\nu(S)$, is very high. However, for small-world networks, our intuition requires that the number of connections in the low-rank component be greater than the number of connections in the sparse component, $\Sigma(L) > \Sigma(S)$. Only in the small-world regime does the decomposition well agree with the intuition of the low-rank decomposition, with L exhibiting relatively low rank and S containing a relatively small nonzero number of connections.
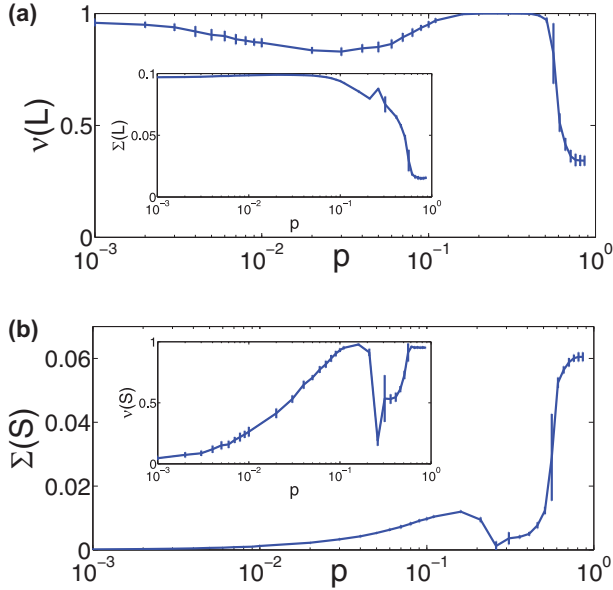
(a)

(b)

FIG. 2. (Color online) Dependence of component structures on the network rewiring probability. (a) Normalized rank of low-rank component $\nu(L)$ as a function of rewiring probability, p, for WS networks of 300 nodes with mean degree 30. (b) Density of sparse component, $\Sigma(S)$, for the same networks as in (a). The inset in (a) depicts the density of low-rank component, $\Sigma(L)$, and the inset in (b) depicts the normalized rank of the sparse component, $\nu(S)$. The mean of these statistics over 20 realizations of the WS network for a given p is plotted with the corresponding standard deviation depicted by the error bars.

FIG. 3. Example network and low-rank decomposition. (a) Adjacency matrix A for a small-world network with $n = 300$ nodes, mean degree 30, and rewiring probability $p = 0.1$. (b) Low rank component L in low-rank network decomposition of A followed by a thresholding process as in Eq. (5). (c) Sparse component S in low-rank network decomposition of A followed by a thresholding process as in Eq. (5). (d) Recovered adjacency matrix $A_r = L + S$. In each plot, black pixels mark connections between nodes of the indicated indices. The relative Frobenius-norm error in the recovered network adjacency matrix in (d) is 0.00089.

To classify a network as small world, we can define intervals $[L_{min}\ L_{max}]$ and $[S_{min}\ S_{max}]$, of $\nu(L)$ and $\Sigma(S)$ values respectively, in which a network exhibits small-world characteristics. If both $\nu(L) \in [L_{min}\ L_{max}]$ and $\Sigma(S) \in [S_{min}\ S_{max}]$, then we classify the network as a small-world network. In this sense, if $\nu(L) \in [L_{min}\ L_{max}]$, then the upper bound guarantees $\nu(L)$ is of low rank, and the lower bound avoids the regime in which the network connectivity is too random and therefore yields decompositions with a very sparse L component. Similarly, if $\Sigma(S) \in [S_{min}\ S_{max}]$, then the upper bound avoids the case in which S is no longer sparse and captures a large number of connections relative to L, and the lower bound guarantees S is not too sparse such that L contains nearly all network connections. With respect to Fig. 2 and our intuition for small-world network structure, networks with appropriately bounded $\nu(L)$ and $\Sigma(S)$ correspond to WS networks with intermediate rewiring probability, and, more generally, correspond to networks with sufficiently many clusters of highly connected nodes such that $\nu(L)$ is low while having a small number of interconnections between clusters composing a relatively sparse S component. The smaller the size of these intervals bounding $\nu(L)$ and $\Sigma(S)$, the stricter the small-world categorization. In addition, since both $\nu(L)$ and $\Sigma(S)$ remain bounded in $[0\ 1]$, this characterization allows for natural comparison of the connectivity structure for networks with different sizes.

The values of $\nu(L)$ and $\Sigma(S)$ also provide information regarding network connectivity features. Generally, networks
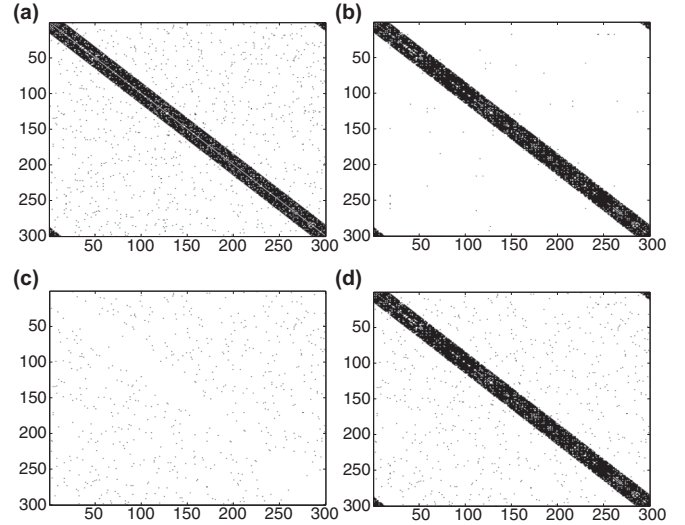
with high $\nu(L)$ and low $\Sigma(S)$, relative to the full network adjacency matrix A, are more regularly connected, whereas networks with low $\nu(L)$ and high $\Sigma(S)$ exhibit characteristics more analogous to randomly connected networks. Thus, without knowing the rewiring probability used in constructing a WS network in Fig. 2 or even the nature of the construction of a more general network, this characterization gives information about the degree of randomness of a given network and also an indication of small-world structure. Such a description only relies on its own properties and does not depend on other constructed regular or random matrices, as required by the existing small-world measures σ and ω. Likewise, our methodology for classifying small-world networks is more general than the original definition of a small-world network in the sense that no particular network size or connection density is required, and hence our characterization may allow for a broader class of networks to be considered small world. The utility and rationale of this broader classification is further addressed in Sec. IV by analyzing several real-world networks of various sizes and connection densities.

In Fig. 3, we compute the network low-rank decomposition for a small-world network with $n = 300$ nodes, mean degree 30, and rewiring probability $p = 0.1$. We compare the original connectivity matrix A, components L and S, and also the quality of the recovered connectivity matrix, $A_r = L + S$, following the decomposition and the thresholding process. Note that since the entries of $\hat{L}$ and $\hat{S}$ in the low-rank network decomposition are thresholded so the recovered components are unweighted, A_r is an approximation of A. Graphically, it is clear that L does indeed well capture the clustered
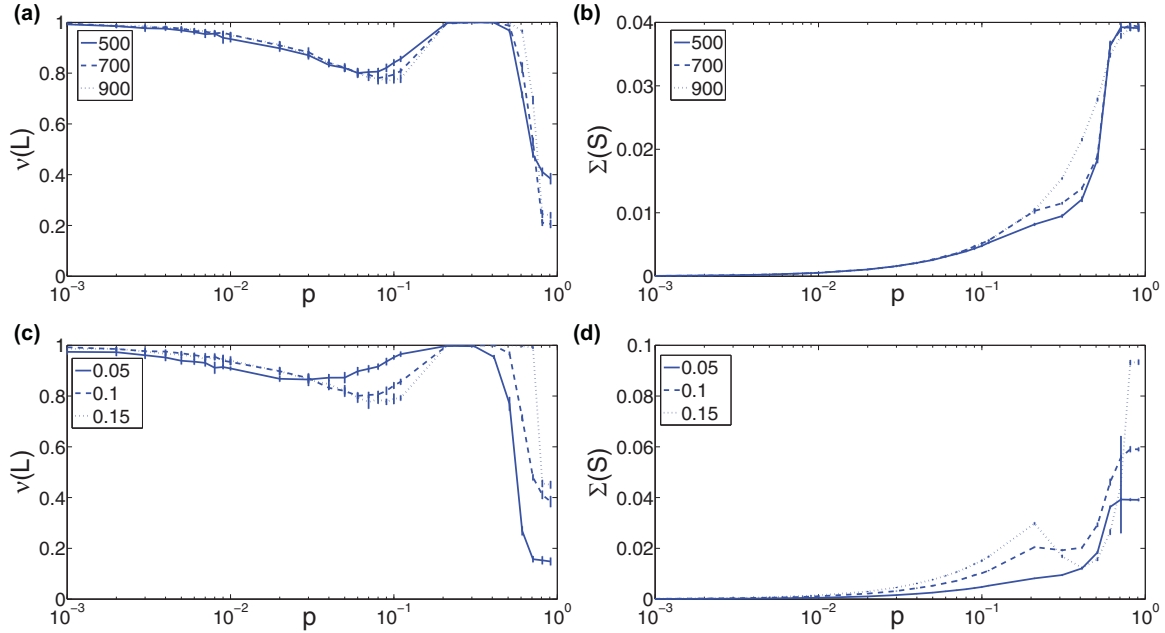
FIG. 4. (Color online) Scaling of low-rank network decomposition. (a) Normalized rank of the thresholded low-rank component L as a function of rewiring probability, p, for WS networks of size $n = 500, 700$, and 900 nodes with mean degree $n/10$. (b) Density of the thresholded sparse component S for the same networks as in (a). (c) Normalized rank of the thresholded low-rank component L as a function of rewiring probability, p, for WS networks of size $n = 500$ nodes with mean degree $0.05n$, $0.1n$, and $0.15n$. (d) Density of the thresholded sparse component S for the same networks as in (c). The mean of these statistics over 20 realizations of the WS network for a given p is plotted with the corresponding standard deviation depicted by the error bars.

connections among nodes near the main diagonal and lower left as well as upper right edges of the connectivity matrix, as shown in Fig. 3(b). Similarly, S primarily contains the sparse interconnections between clusters resulting from rewirings in the WS construction, as shown in Fig. 3(c). The sum $L + S$ depicted in Fig. 3(d) closely resembles A, reflecting that network connectivity is well preserved following the decomposition and subsequent threshold processing of each component. To quantify the error in the recovered connectivity matrix for a given network decomposition, we use the entrywise Frobenius matrix norm defined by $\|B\|_F = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} |B_{ij}|^2}$ for an $n \times n$ matrix B. In the case of Fig. 3, the relative error in the recovered connectivity matrix, $\|A_r - A\|_F / \|A\|_F = 0.00089$, reflecting very small variation in the individual recovered connections.

III. ROBUSTNESS OF LOW-RANK NETWORK DECOMPOSITION

For a network characterization to be general and robust, it should hold over a range of network sizes and connection densities, remaining invariant with respect to the order of indexing of nodes within the network. In Fig. 4, we plot the scaling of the normalized rank of L and density of S for various network sizes and mean degrees. Regardless of network size and the average number of connections, we observe the qualitative features of L and S remain the same. The minimum of $\nu(L)$ appears to slightly decrease with network size and additional connections, leveling off once the network is large enough or there is a sufficient number of connections. These scalings are also quite stable across

network realizations, with a relatively small standard deviation among 20 network realizations corresponding to each p, as depicted by the error bars in Fig. 4. Hence, the decomposition is particularly stable for networks with a large number of nodes or a sufficient number of connections.

It is important to emphasize that the low-rank network decomposition is independent of the indexing of nodes, yielding the same connections between nodes in both the L and S components, only reindexed, regardless of how the nodes are ordered. This independence property demonstrates that the low-rank decomposition characterizes the intrinsic network connectivity between nodes regardless of how one artificially indexes nodes. Hence, $\nu(L)$ and $\Sigma(S)$ will be identical, and therefore the network classification will remain the same, after reindexing. We make this statement more precisely with the reindexing property below, which we demonstrate in the Appendix.

A. Reindexing property

Let the $n \times n$ adjacency matrix A have decomposition $A = L + S$, where L and S are solutions to Eq. (4). Let $\tilde{A}$ be equivalent to A after a sequence of ν reindexings of the nodes, $\{\{i_1, j_1\}, \ldots, \{i_\nu, j_\nu\}\}$, where a relabeling, $\{i_1, j_1\}$, switches the indices of nodes i_1 and j_1. Mathematically, for reindexing k, the corresponding adjacency matrix $A_k = R_{i_k j_k} A_{k-1} R_{i_k j_k}$, where $R_{i_k j_k}$ is the transformation matrix interchanging rows i_k and j_k and A_{k-1} is the $(k-1)$th reindexing of A. If $\tilde{L}$ and $\tilde{S}$ solve Eq. (4) corresponding to reindexed adjacency matrix $\tilde{A}$, then (i) $\nu(L) = \nu(\tilde{L})$ and $\Sigma(S) = \Sigma(\tilde{S})$ and (ii) L can be recovered from $\tilde{L}$ and S can be recovered from $\tilde{S}$ via ν reindexings.

IV. REAL-WORLD NETWORKS

In this section, we illustrate that the low-rank network decomposition yields network classifications that agree quite well with the small-worldness measures, σ and ω, defined by Eqs. (1) and (2), respectively. However, the nature of this novel characterization also underlines important features that may go unnoticed by computing only σ and ω. In Table I, we compare small-world classifications using σ, ω, the normalized rank of L, and the density of S for several real-world networks.

While the description of a network as small-world depends on the classification criterion chosen for any of the three methods described, the general relative degrees of small-worldness suggested by the methods are similar. For example, for the network of collaborations between jazz musicians, the small negative $\omega = -0.11$ value suggests the network has small-world characteristics with a more regular than random structure. At the same time, the large $\sigma = 3.16$ value suggests a small-world structure, and the relatively high but nonsaturated $\nu(L) = 0.78$ with relatively low $\Sigma(S) = 0.0077$ also suggests small-world connectivity with more regularity than randomness.

Based on the WS networks in Fig. 2, choosing small-world intervals corresponding to the low-rank decomposition $[L_{\min} \ L_{\max}] = [0.1 \ 0.9]$ and $[S_{\min} \ S_{\max}] = [0.0001 \ 0.03]$ would be a reasonable choice of bounds for the values of the normalized rank of L and the density of S, respectively. In Ref. [16], it is indicated that networks with $\sigma > 1$ are small world. Similarly, in Ref. [30], it is suggested that $-0.5 \leqslant \omega \leqslant 0.5$ is an appropriate small-world regime. Choosing these bounds for small-world classification, the three methods agree well except in the cases of word adjacencies in "David Copperfield" and university email exchanges. For these two networks, both σ and our low-rank characterization suggest small-worldness, whereas ω classifies these networks as slightly more random than small world. Since the random characteristics are not too prevalent, by choosing a looser bound for ω, the classifications would be identical to the other two methods. Considering that these two networks are also very sparse and the structural properties of sparse networks, such as $C_{\rm rand}$ and $l_{\rm rand}$, for example, can be quite variable across network realizations, it may be the case that most network characterizations are also relatively variable if a network contains too few connections.

For contrast, we consider the flight connection network, which has a similar connection density as the constructed WS networks with $k \approx 0.1n$. While the edge densities of the real-world and WS networks are quite similar, we see that, depending on the specific structure of the set of connections, the small-world classification differs considerably. Since $\omega = -0.15$ in the case of the flight connection network, the connectivity structure appears to be in the small-world regime with slightly more regular than random connectivity. However, a network with the same size and connection density instead arranged more randomly, through edge rewirings, for example, would have much larger positive ω and exhibit a corresponding large increase in $\Sigma(S)$, as evidenced in Figs. 1 and 4, respectively.

Finally, we emphasize that the biological network of connections between cortical areas in the cerebral cortex provides particular insight into the unique utility of the low-rank network decomposition. This particular network describes connectivity between $n = 29$ regions within the six main functional areas of the cortex (occipital, temporal, parietal, frontal, prefrontal, and limbic). Using retrograde tracer injections into the cerebral cortex of the macaque monkey, Ref. [23] uncovered 36% more connections than reported by previous studies due to the relatively long length and low density of these additional edges. This result implies that inter-regional cortical connectivity may be much more dense than previously thought, yielding a network of cortical regions with an edge density of 66% and mean degree $k = 18.48$ upon unweighting the network.

Performing the low-rank network decomposition on this network, we observe that 14 of 16 connections present within the sparse component S are between cortical regions. Since connections within cortical areas are far more dense than interareal connections, we clearly observe the role of the connections in S in linking highly clustered groups of nodes, which in this case are individual cortical areas. Moreover, upon removing the interconnections between cortical areas, identified by S, from the network, we observe a statistical change in the network clustering, with the variance of the clustering coefficients for the individual nodes increasing from 0.0028 to 0.0035. The increase in clustering-coefficient variance after removing the sparse connections, for example, may suggest that the connections in sparse component S help to connect nodes of varying degree and thereby play a role in equalizing clustering across the network.

Considering that such densely connected networks are not small world in the conventional sense, since they violate the sparse connectivity condition, it is important to note that densely connected networks can also be well characterized by the low-rank network decomposition [22]. Densely connected networks typically exhibit short average path lengths and high clustering coefficients. However, since they use more connections to achieve these statistics, they may not be viewed as small-world networks according to the original notion of small-worldness [13]. Nevertheless, considering the prevalence of dense networks in natural systems, such as cerebral cortex, protein, and gene-regulatory networks [23–25], it may be necessary to either extend the definition of small-world networks to incorporate dense connectivity or instead define an appropriate new class of networks as in Ref. [22].

V. DISCUSSION

In this paper, we have identified a new link between small-world network connectivity structure and low-rank matrix decompositions. Formulating a methodology for decomposing network adjacency matrices into low-rank and sparse components, we have developed a useful scheme for determining the small-world characteristics and general structural properties of a network based on the rank and connection density of the decomposition components. This characterization is statistically reliable and makes use of only the intrinsic network structure embodied by the adjacency matrix, avoiding comparison with benchmark networks, as in the case of previous small-world measures [16,30]. In addition, we have

TABLE I. Small-world classification of real-world networks. Comparison of small-world properties for several biological, social, and technological networks. For each network, n is the number of nodes, k is the mean degree, σ is the small-world metric defined in Eq. (1) [16], ω is the small-world metric defined in Eq. (2) [30], $\nu(L)$ is the normalized rank of component L, and $\Sigma(S)$ is the density of component S. In order to compare structural properties among different networks and constructed WS networks, for weighted networks, we construct embedded unweighted networks for our computations, making all connections have equal strength.

Network	n	k	σ	ω	$\nu(L)$	$\Sigma(S)$
Flight connections [49]	500	48	4.66	−0.15	0.78	0.0098
Word adjacencies in "David Copperfield" [50]	112	7.59	2.36	0.70	0.67	0.0067
Email [51]	1133	9.62	22.35	0.57	0.51	0.0015
Brain regions [23]	29	18.48	1.06	−0.12	0.69	0.0190
Caenorhabditis elegans [7,13]	277	14.46	5.23	0.48	0.59	0.0081
Jazz musicians [52]	198	27.70	3.16	−0.11	0.78	0.0077
Network science coauthorship [50]	1589	3.45	303.19	0.18	0.24	0.0004

shown that, independent of node indexing, the decomposition separates connections within highly clustered groups of nodes from relatively sparse interconnections between the clusters. Applying this network decomposition to diverse real-world networks, we have made classifications that agree well with several well-known measures of small-worldness and also identified new structural properties of networks, such as in the case of the cerebral cortex network of monkeys. While conventional small-world descriptors, such as path length and clustering coefficient, are trivially insensitive to structural variations among densely connected networks, the characterization introduced in this work is indeed able to differentiate among topological differences in sparse and dense networks alike.

The results of this work suggest several new directions for research in both matrix decomposition and network theory. The low-rank decomposition theory has been primarily applied to relatively dense matrices using very specific choices of connection penalization [31,53,54]. However, since we suggest a new choice of penalization parameter, i.e., λ in Eq. (6), accounting for changes in matrix connection density, it would be useful for conditions to be developed for which this choice of λ, or one that is similar, will likely yield a viable decomposition. Since different choices of connection penalization often yield diverse decomposition trends and corresponding classifications, an interesting area of future study would be to determine how other choices of λ impact decomposition characteristics. Likewise, it would be informative to study how matrix or network decompositions vary with λ and whether these changes reflect additional matrix properties.

The presence of the same nodes in L and S regardless of indexing is also a particularly important property outside of the current context of consistent network classifications. Since the network decomposition always separates the same clustered connections, in L, from the connections between clusters, in S, this conservation of connections implies that the low-rank network decomposition gives a method of consistently discriminating between connections within clusters and interconnections between clusters. On a similar note, this method of partitioning nodes may be considered a form of community detection as well, identifying groups of nodes with particularly dense connectivity [38,40,43,55].

Using the notion of edge-betweenness introduced in Ref. [40], defined as the number of shortest paths between

distinct nodes passing through a given edge, we observe that the connections in S often exhibit significantly higher edge-betweenness than connections contained in L. Since connections in S tend to connect different clusters, shortest paths will typically run through these edges, thereby yielding high edge-betweenness. Thus, we observe a parallel between community structure and our low-rank network characterization, which agrees well with the intuition for the respective structure of components L and S discussed in Sec. II A. We expect that using our framework to reveal the complete set of clustered connections may be useful in refining community separation algorithms since the nonclustered connections in S may then be disregarded and thereby yield a smaller appropriate set
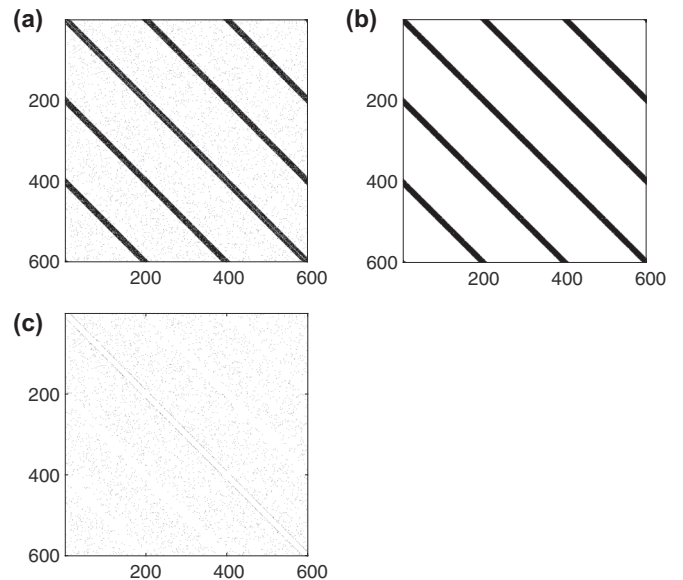

FIG. 5. Example network decomposition in two dimensions. (a) Adjacency matrix A for a small-world network with $n = 600$ nodes, mean degree 60, and rewiring probability $p = 0.1$. In this case, the network is defined over a two-dimensional lattice with periodic boundary conditions and thickness of 3 nodes. (b) Low rank component L in low-rank network decomposition of A followed by a thresholding process as in Eq. (5). (c) Sparse component S in low-rank network decomposition of A followed by a thresholding process as in Eq. (5). In each plot, black pixels mark connections between nodes of the indicated indices.

of connections to be separated. In addition, we have verified that this structural decomposition also holds for networks in higher dimensions. For example, in Fig. 5, we consider the low-rank decomposition for a small-world network on a two-dimensional lattice with periodic boundary conditions and rewiring probability $p = 0.1$. We similarly observe the presence of the clustered connections in L and the sparse rewired connections in S, as in the classic one-dimensional ring lattice case. While these properties of the decomposition components remain identical in higher dimensions, we note that the precise dependence of $\nu(L)$ and $\Sigma(S)$ on p relies on the specific lattice properties in higher dimensions, which may be of interest in future investigations.

In terms of network theory, a comparison of the small-world properties of densely and sparsely connected networks, especially using the terminology introduced in this work, would be a natural direction for future research. Since most real-world networks also evolve in time, space, and sometimes connectivity, understanding the contribution of substructures, such as low-rank network decomposition components, to network behavior marks another important direction for future study.

APPENDIX

1. Reindexing property

In this section of the Appendix, we describe in more detail the reindexing property stated in Sec. III. The justification for the reindexing property can be seen as follows.

Let $\tilde{A}$ be equivalent to the $n \times n$ matrix A after v reindexings, $\{\{i_1, j_1\}, \ldots, \{i_v, j_v\}\}$, with corresponding transformation matrices $\{R_{i_1 j_1}, \ldots, R_{i_v j_v}\} \equiv \{R_1, \ldots, R_v\}$. Thus, $\tilde{A} = R_v \ldots R_1 A R_1 \ldots R_v \equiv T(A)$. Further assume that sets of matrices $\{L, S\}$ and $\{\tilde{L}, \tilde{S}\}$ solve Eq. (4) corresponding to A and $\tilde{A}$, respectively. Likewise, define $\acute{L} = T(L)$ and $\acute{S} = T(S)$. Note that the reindexings can be reversed through a reversed sequence of reindexings, $\{\{i_v, j_v\}, \ldots, \{i_1, j_1\}\}$ since $R_{i_k j_k} R_{i_k j_k} = I$ for $k \in Z^+$, where I is an $n \times n$ identity matrix. Thus, $A = T^{-1}\tilde{A}$, $L = T^{-1}\acute{L}$, and $S = T^{-1}\acute{S}$, where $T^{-1}(A) = R_1 \ldots R_v A R_v \ldots R_1$. Since T and T^{-1} correspond to sequences of elementary row or column operations, we have $\nu(L) = \nu(\acute{L})$, $\Sigma(S) = \Sigma(\acute{S})$, $\|L\|_* = \|\acute{L}\|_*$, and $\|S\|_1 = \|\acute{S}\|_1$ (X1).

We now argue that $\{\acute{L}, \acute{S}\}$ in fact solves Eq. (4) corresponding to $\tilde{A}$. Since the operator T^{-1} is clearly linear, the matrix A can be decomposed into the sum $A = T^{-1}(\tilde{A}) = T^{-1}(\tilde{L}) + T^{-1}(\tilde{S})$. However, since matrices L and S solve Eq. (4), the value of Eq. (4a) using variables $T^{-1}(\tilde{L})$ and $T^{-1}(\tilde{S})$ should be greater than that using variables L and S. From the properties of T in conclusion (X1), the value of Eq. (4a) using variables $\tilde{L}$ and $\tilde{S}$ should also be greater than that using variables $\acute{L}$ and $\acute{S}$ (X2).

On the other hand, $\tilde{A} = T(A) = T(L) + T(S) = \acute{L} + \acute{S}$, and thus the matrix $\tilde{A}$ can be decomposed into the sum between $\acute{L}$ and $\acute{S}$. However, since matrices $\tilde{L}$ and $\tilde{S}$ solve Eq. (4), the value of Eq. (4a) using variables $\acute{L}$ and $\acute{S}$ should be greater than that using variables $\tilde{L}$ and $\tilde{S}$.

Combined with conclusion (X2), we have the value of Eq. (4a) using variables $\tilde{L}$ and $\tilde{S}$ should be equal to that using variables $\acute{L}$ and $\acute{S}$. Therefore, matrices $\acute{L}$ and $\acute{S}$ also solve Eq. (4), and due to the uniqueness of the low-rank decomposition, which holds under broad conditions [31,32], we have $\tilde{L} = \acute{L}$ and $\tilde{S} = \acute{S}$.

2. Augmented Lagrangian method

In this section of the Appendix, we briefly discuss the augmented Lagrangian method useful for performing the low-rank decomposition [47,56]. This algorithm is generally applicable in solving constrained optimization problems in which $f(\mathbf{x})$ is the real-valued function to be minimized and $\mathbf{c_i}(\mathbf{x}) = 0$ is the ith constraint on the optimization problem. The corresponding augmented Lagrangian function to consider in this optimization problem, using a quadratic penalty, is

$$L(\mathbf{x}, \lambda, \mu) = f(\mathbf{x}) + \frac{\mu}{2} \sum_i \mathbf{c_i}(\mathbf{x})^2 + \sum_i \lambda_\mathbf{i} \mathbf{c_i}(\mathbf{x}), \qquad \text{(A1)}$$

where $\lambda_\mathbf{i}$ is the ith estimated Lagrange multiplier and μ is a positive penalization scalar. Below is the pseudocode for one particular implementation of the augmented Lagrangian updating scheme, where $\mathbf{x^k}$, $\lambda^\mathbf{k}$, and μ^k denote the value of $\mathbf{x}$, λ, and μ, respectively, on the kth iteration of the algorithm.

Augmented Lagrangian algorithm

(1) Initialize $\mathbf{x^0}$, $\lambda^\mathbf{0}$, μ^0, and penalization multiplier $\rho \geqslant 1$.
(2) Update solution to optimization problem: $\mathbf{x^{k+1}} = \arg\min_\mathbf{x} L(\mathbf{x}, \lambda^\mathbf{k}, \mu^k)$.
(3) Update Lagrange multipliers: $\lambda^\mathbf{k+1} = \lambda^\mathbf{k} + \mu^k \mathbf{c}(\mathbf{x^{k+1}})$.
(4) Update penalization: $\mu^{k+1} = \rho \mu^k$.
(5) Repeat previous three steps until solution converges.

We remark that in the case of the low-rank decomposition, the constrained optimization problem to consider is given by Eq. (4), and additional details as well as sample code can be found, for example, in Ref. [47]. For an $n \times n$ matrix, the computational complexity for the low-rank matrix decomposition is generally $\mathcal{O}(n^3)$, and we expect that the design of fast algorithms for low-rank decomposition based on particular classes of matrices will be further investigated in future research [31,57,58].

[1] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).

[2] O. Sporns, D. R. Chialvo, M. Kaiser, and C. C. Hilgetag, Trends Cogn. Sci. (Regul. Ed.) **8**, 418 (2004).

[3] J. Scott, *Social Network Analysis: A Handbook* (Sage, Los Angeles, CA, 2000).

[4] F. Ball, D. Mollison, and G. Scalia-Tomba, Ann. Appl. Probab. **7**, 46 (1997).

[5] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A. L. Barabasi, Nature **407**, 651 (2000).

[6] F. Liljeros, Microbes Infect. **5**, 189 (2003).

[7] J. G. White, E. Southgate, J. N. Thomson, and S. Brenner, Philos. Trans. R. Soc. Lond. B **314**, 1 (1986).

[8] Q. Chen, H. Chang, R. Govindan, and S. Jamin, in *Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies, INFOCOM 2002*, Vol. 2 (IEEE Computer Society, New York, 2002), pp. 608–617.

[9] S. N. Dorogovtsev and J. F. F. Mendes, Proc. R. Soc. Lond. B **268**, 2603 (2001).

[10] A. Barabasi and R. Albert, Science **286**, 509 (1999).

[11] P. Erdös and A. Rényi, Publ. Math. Debrecen **6**, 290 (1959).

[12] D. J. Watts, *Small Worlds: The Dynamics of Networks between Order and Randomness* (Princeton University Press, Princeton, NJ, 1999), p. 262.

[13] D. J. Watts and S. H. Strogatz, Nature **393**, 440 (1998).

[14] L. Amaral, A. Scala, M. Barthélémy, and H. Stanley, Proc. Natl. Acad. Sci. USA **97**, 11149 (2000).

[15] V. Latora and M. Marchiori, Phys. Rev. Lett. **87**, 198701 (2001).

[16] M. D. Humphries and K. Gurney, PLoS ONE **3**, e0002051 (2008).

[17] M. P. van den Heuvel, C. J. Stam, M. Boersma, and H. E. Hulshoff Pol, Neuroimage **43**, 528 (2008).

[18] O. Sporns and C. J. Honey, Proc. Natl. Acad. Sci. USA **103**, 19219 (2006).

[19] D. S. Bassett and E. Bullmore, Neuroscientist **12**, 512 (2006).

[20] T. I. Netoff, R. Clewley, S. Arno, T. Keck, and J. A. White, J. Neurosci. **24**, 8075 (2004).

[21] A. Roxin, H. Riecke, and S. A. Solla, Phys. Rev. Lett. **92**, 198101 (2004).

[22] V. J. Barranca, D. Zhou, and D. Cai, Sci. Rep. **5**, 10611 (2015).

[23] N. T. Markov, M. Ercsey-Ravasz, D. C. Van Essen, K. Knoblauch, Z. Toroczkai, and H. Kennedy, Science **342**, 1238406 (2013).

[24] V. Baldazzi, D. Ropers, Y. Markowicz, D. Kahn, J. Geiselmann, and H. de Jong, PLoS Comput. Biol. **6**, e1000812 (2010).

[25] V. Spirin and L. Mirny, Proc. Natl. Acad. Sci. USA **100**, 12123 (2003).

[26] A. Clauset, C. Moore, and M. E. Newman, Nature **453**, 98 (2008).

[27] M. Newman, *Networks: An Introduction* (Oxford University Press, Oxford, 2010).

[28] E. Estrada and D. J. Higham, SIAM Rev. **52**, 696 (2010).

[29] J. G. Restrepo, E. Ott, and B. R. Hunt, Phys. Rev. E **76**, 056119 (2007).

[30] Q. K. Telesford, K. E. Joyce, S. Hayasaka, J. H. Burdette, and P. J. Laurienti, Brain Connect. **1**, 367 (2011).

[31] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. S. Willsky, in *Proceedings of the 47th Annual Allerton Conference on Communication, Control, and Computing, Allerton'09* (IEEE Press, Piscataway, NJ, 2009), p. 962.

[32] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. S. Willsky, SIAM J. Optim. **21**, 572 (2011).

[33] E. J. Candès, X. Li, Y. Ma, and J. Wright, J. ACM **58**, 1 (2011).

[34] R. Mazumder, T. Hastie, and R. Tibshirani, J. Mach. Learn. Res. **11**, 2287 (2010).

[35] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan, Proc. IEEE **98**, 1031 (2010).

[36] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma, IEEE Trans. Pattern Anal. Mach. Intell. **34**, 2233 (2012).

[37] A. Arenas, A. Diaz-Guilera, and C. J. Perez-Vicente, Phys. Rev. Lett. **96**, 114102 (2006).

[38] M. E. J. Newman, Nat. Phys. **8**, 25 (2011).

[39] E. A. Leicht and M. E. J. Newman, Phys. Rev. Lett. **100**, 118703 (2008).

[40] M. Girvan and M. E. J. Newman, Proc. Natl. Acad. Sci. USA **99**, 7821 (2002).

[41] S. Fortunato and M. Barthelemy, Proc. Natl. Acad. Sci. USA **104**, 36 (2007).

[42] A. Clauset, M. E. J. Newman, and C. Moore, Phys. Rev. E **70**, 066111 (2004).

[43] S. Fortunato, Phys. Rep. **486**, 75 (2010).

[44] M. E. J. Newman, Phys. Rev. E **69**, 066133 (2004).

[45] M. E. Newman, Proc. Natl. Acad. Sci. USA **103**, 8577 (2006).

[46] Z. Lin, A. Ganesh, J. Wright, L. Wu, M. Chen, and Y. Ma, in *Intl. Workshop on Comp. Adv. in Multi-Sensor Adapt. Processing*, Vol. 61 (IEEE Press, Aruba, Dutch Antilles, The Netherlands, 2009), p. 213.

[47] Z. Lin, M. Chen, and Y. Ma, arXiv:1009.5055 (2010).

[48] J.-F. Cai, E. J. Candés, and Z. Shen, SIAM J. Optim. **20**, 1956 (2010).

[49] J. Marcelino and M. Kaiser, PLOS Curr. Influenza. **4** (2012), doi: 10.1371/4f8c9a2e1fca8.

[50] M. E. J. Newman, Phys. Rev. E **74**, 036104 (2006).

[51] R. Guimerà, L. Danon, A. Díaz-Guilera, F. Giralt, and A. Arenas, Phys. Rev. E **68**, 065103(R) (2003).

[52] P. Gleiser and L. Danon, Adv. Complex Syst. **6**, 565 (2003).

[53] B. Hutchinson, M. Ostendorf, and M. Fazel, in *13th Annual Conference of the International Speech Communication Association, Interspeech 2012* (ISCA, Portland, OR, 2012), p. 1676.

[54] P.-A. Savalle, E. Richard, and N. Vayatis, in *Proceedings of the 29th International Conference on Machine Learning* (Omnipress, Edinburgh, Scotland, UK, 2012), p. 1351.

[55] B. Ball, B. Karrer, and M. E. J. Newman, Phys. Rev. E **84**, 036103 (2011).

[56] C. Wu, X.-C. Tai *et al.*, SIAM J. Imag. Sci. **3**, 300 (2010).

[57] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Vol. 3 (JHU Press, Baltimore, 2012).

[58] F. R. Bach and M. I. Jordan, in *Proceedings of the 22nd International Conference on Machine Learning* (ACM, New York, 2005), pp. 33–40.