

Heterogeneous multidimensional scaling for complex networks

Qi Xuan, Xiaodi Ma, Chenbo Fu*, Hui Dong, Guijun Zhang and Li Yu

*Department of Automation
Zhejiang University of Technology
Hangzhou 310023, P. R. China
cbfu@zjut.edu.cn

Received 6 February 2014

Accepted 22 June 2014

Published 22 July 2014

Many real-world networks are essentially heterogeneous, where the nodes have different abilities to gain connections. Such networks are difficult to be embedded into low-dimensional Euclidean space if we ignore the heterogeneity and treat all the nodes equally. In this paper, based on a newly defined heterogeneous distance and a generalized network distance under the constraints of network and triangle inequalities, respectively, we propose a new heterogeneous multidimensional scaling method (HMDS) to embed different networks into proper Euclidean spaces. We find that HMDS behaves much better than the traditional multidimensional scaling method (MDS) in embedding different artificial and real-world networks into Euclidean spaces. Besides, we also propose a method to estimate the appropriate dimensions of Euclidean spaces for different networks, and find that the estimated dimensions are quite close to the real dimensions for those geometrical networks under study. These methods thus can help to better understand the evolution of real-world networks, and have practical importance in network visualization, community detection, link prediction and localization of wireless sensors.

Keywords: Geometrical network; scale-free network; Euclidean space; wireless network; graph visualization.

PACS Nos.: 89.75.Hc, 89.75.Fb, 89.75.Da.

1. Introduction

Many complex systems in nature and society are described by networks, ranging from biological systems, e.g. protein–protein networks^{1,2} and brain networks,^{3,4} to sociological systems, e.g. collaboration networks^{5,6} and communication networks.^{7,8} It is being proved that networks as discrete spaces are useful to study various dynamics, such as material and information flows,^{9,10} reaction-diffusion dynamics,^{11–13} synchronization and consensus^{14–17} etc. On the other hand, Euclidean geometry is still considered as the basis of many sciences. In fact, discrete network space and continuous Euclidean space are different metric spaces¹⁸ with different definitions of distance under several constraints including the well-known triangle inequality. For

example, the network distance between two nodes is always defined as the smallest number of connections included in any path between them¹⁹ while the Euclidean distance between two points is defined as the length of the line segment connecting them. Both the metric spaces have their respective advantages in certain situations, i.e. the different status of individuals can be identified by simply comparing their local structural properties in network spaces,^{8,20} while many individual behaviors can be predicted by designing machine learning algorithms^{21–23} in Euclidean space. Thus, it is of much scientific significance to establish a bridge between the two, i.e. transform the description of a system from one space to the other, in order to better understand the system and use different methods together to solve some problems more efficiently.

There are several ways to generate a network given a set of points in some Euclidean space, e.g. two points are connected if and only if their Euclidean distance is smaller than a predefined threshold.²⁴ However, by comparison, it is much more difficult to embed a network into some low-dimensional Euclidean space, i.e. specialize the positions for all the nodes in the Euclidean space, satisfying that the Euclidean distance between two nodes is smaller than some threshold if and only if they are connected in the network. Solution of the latter has practical significance in the areas such as network visualization,²⁵ community detection,²⁶ link prediction²⁷ and localization of wireless sensors.²⁸ Traditionally, researchers tend to use the multidimensional scaling (MDS) method²⁹ to solve this problem, where the distance between two nodes in the network (it can be the average shortest path length) need to be firstly defined. By adopting MDS, Venna and Kaski³⁰ visualized gene interaction networks and Bonabeau³¹ visualized the work network in a technical laboratory and the California tidal flat community food web.

However, MDS treats all the nodes equally in a network, while many real networks have quite heterogeneous structure, i.e. the nodes in these networks have different abilities to gain connections. For example, in a wireless network,^{28,32,33} different sensors may have different sensing radius and two sensors are linked only when their sensing ranges overlap with each other. As a result, it is straightforward to expect that, statistically, the sensors of larger sensing radius have more neighbors. Another similar example is the Apollonian network,³⁴ where two spheres are linked if the distance between their centers are equal to the sum of their radiuses, so that the spheres of larger radius also tend to have more neighbors. Recently, Caldarelli *et al.*³⁵ proposed a network model where each node has a fitness and two nodes are linked with a probability depending on their fitness. These studies suggest that the diversity of individual abilities in a system may lead to the diversity of node degrees, and thus account for the emergence of heterogeneous networks.^{7,36–38} In other words, the observable structural properties of the nodes in a network can partly reflect their intrinsic abilities to gain connections, and thus can be used to improve the existent embedding methods.

In this paper, based on MDS, we propose a heterogeneous multidimensional scaling (HMDS) method by utilizing the information of node degree to better embed

networks into low-dimensional Euclidean spaces. We first generalize the network distance, and then introduce several kinds of relationships between individuals, one of which is used to define a new heterogeneous distance between two nodes by considering their different abilities to establish connections. Note that since such individual abilities may be unknown in some real cases and they are partly reflected in structural properties, we use their node degrees in the network, rather than the abilities themselves, to calculate the pairwise heterogeneous distances in practice. Our main finding is that HMDS indeed behaves much better than MDS in embedding artificial Apollonian networks and real-world geometrical and social networks.

The rest of the paper is organized as follows. In Sec. 2, the traditional MDS to embed networks into Euclidean spaces is introduced. In Sec. 3, the network distance is generalized and further extended in heterogeneous networks by considering the different individual abilities to gain connections, based on which the HMDS method is proposed. Then, both MDS and HMDS are tested on artificial and real networks in Sec. 4. The work is finally concluded in Sec. 5.

2. MDS for Networks

An undirected network is denoted by $G = (V, E)$ with $V = \{v_1, v_2, \dots, v_N\}$ and $E \subset V \times V$ as its node and link sets, respectively. The network structure is always represented by the adjacency matrix $A = [a_{ij}]_{N \times N}$ with $a_{ij} = 1$ if $(v_i, v_j) \in E$ and $a_{ij} = 0$ otherwise. In general, the distance between a pair of nodes is defined as the shortest path length between them in the network. Thus, we also have a distance matrix $D = [d_{ij}]_{N \times N}$ with each element d_{ij} equal to the minimum number of links in a path between nodes v_i and v_j in the network. Provided by the distance matrix, MDS is formulated as an optimization problem,²⁹ where the coordinates of the nodes, denoted by $x_i, i = 1, 2, \dots, N$, are found by minimizing the cost function:

$$\min_{x_1, x_2, \dots, x_N} \sum_{i < j} (\|x_i - x_j\| - d_{ij})^2. \tag{1}$$

Denote by $B = [b_{ij}]_{N \times N}$ the centralized inner product matrix with its elements defined as

$$b_{ij} = \frac{1}{2} \left(-d_{ij}^2 + \frac{1}{N} \sum_{j=1}^N d_{ij}^2 + \frac{1}{N} \sum_{i=1}^N d_{ij}^2 - \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N d_{ij}^2 \right). \tag{2}$$

Let $\lambda_i, i = 1, 2, \dots, r$, the top r largest positive eigenvalues of the matrix B , satisfying $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > 0$ and e_1, e_2, \dots, e_r the corresponding unit eigenvectors, i.e. $\|e_r\| = 1$. The optimal coordinate matrix is simply defined as

$$X = [\sqrt{\lambda_1}e_1, \sqrt{\lambda_2}e_2, \dots, \sqrt{\lambda_r}e_r], \tag{3}$$

with each row, denoted by x_i , representing the coordinates of node v_i in r -dimensional Euclidean space.

Once a network is embedded into r -dimensional Euclidean space, each pair of nodes v_i and v_j will have a real Euclidean distance between them, defined as

$$d_{ij}^* = \|x_i - x_j\|. \quad (4)$$

For simplicity, denote by $D^* = [d_{ij}^*]_{N \times N}$ the embedded distance matrix, based on which an embedded network $G^* = (V, E^*)$ can be constructed by connecting $|E|$ pairs of nodes with the smallest Euclidean distances, so that it has the same number of links as the original network. Denoting by $A^* = [a_{ij}^*]_{N \times N}$ the adjacent matrix of the embedded network, the ratio of the links in the original network that do not exist in the embedded network, named as the relative embedding error for the network, is then calculated by

$$\epsilon = \frac{\sum_{i=1}^N \sum_{j=i+1}^N |a_{ij} - a_{ij}^*|}{2 \sum_{i=1}^N \sum_{j=i+1}^N a_{ij}}. \quad (5)$$

We will use this formula to measure how well a network is embedded into certain low-dimensional Euclidean space by adopting different methods.

3. Methods

In reality, different individuals may have different abilities to establish connections with others, which results in many real-world heterogeneous networks.^{7,36–38} For example, it is quite natural to consider that, statistically, the countries of larger area have more neighbors in country neighborhood network,³⁸ and the sensors of larger sensing radius may be connected to more others in wireless networks^{28,32,33} if they are uniformly distributed in Euclidean space. In these networks, whether there is a link or not between a pair of nodes is not only determined by the distance between their respective centers but also depends on their abilities to influence each other. However, as we know, the traditional MDS, as well as many other generalized versions,^{39–41} does not differentiate the nodes based on their abilities to gain connections, instead, they treat all the nodes equally when embedding a network into Euclidean space. Here, we first introduce a method to generalize the network distance; then, based on a newly defined heterogeneous distance, we further extend the generalized network distance in heterogeneous networks, based on which we propose HMDS to embed such networks more efficiently. Note that the individual abilities to gain connections can be reflected by some of the structural properties, e.g. node degree, in the network. Therefore, this method can still be applied even when only the network topology is known *a priori*.

3.1. Generalized network distance

In general, the distance between a pair of nodes v_i and v_j in a network G should be defined under the constraints of the following two inequalities:

- *Network inequality*: The distances between pairwise unconnected nodes must be larger than those between pairwise connected nodes.

- *Triangle inequality:* For every triple of nodes $\{v_i, v_j, v_k\}$, the distance between each pair of them cannot be larger than the sum of the distances between the other two pairs, i.e. it must satisfy $d_{ij} \leq d_{ik} + d_{jk}$, with the condition $d_{ik} \leq d_{jk} \leq d_{ij}$.

Denoting by l the shortest path length between a pair of nodes, we generalize the network distance between them as

$$d(l) = 1 + \eta(l - 1), \tag{6}$$

with the parameter $\eta \in (0, 1]$, based on which we propose the first theorem:

Theorem 1. *The generalized network distances between pairwise nodes, defined by Eq. (6), satisfy both the network and triangle inequalities.*

Proof. The *network inequality* is satisfied because Eq. (6) suggests that $d(l) = 1$ if the two nodes are connected, i.e. $l = 1$; and $d(l) > 1$ if they are not connected, i.e. $l > 1$; thus the generalized network distances between pairwise unconnected nodes must be larger than those between pairwise connected nodes. Meanwhile, for every triple of nodes $\{v_i, v_j, v_k\}$ with the ordered shortest path lengths $l_{ik} \leq l_{jk} \leq l_{ij}$, we have $d_{ik} \leq d_{jk} \leq d_{ij}$ and $l_{ij} \leq l_{ik} + l_{jk}$, based on which we get

$$\begin{aligned} d_{ik} + d_{jk} - d_{ij} &= 1 + \eta(l_{ik} - 1) + 1 + \eta(l_{jk} - 1) - 1 - \eta(l_{ij} - 1) \\ &\geq 1 - \eta \geq 0, \end{aligned} \tag{7}$$

thus, the *triangle inequality* is also satisfied. *Theorem 1* is proved. □

3.2. HMDS for networks

We now further extend the generalized network distance in heterogeneous networks. Denote by $S = [s_{ij}]_{N \times N}$ the similarity matrix among the nodes in the network G with its elements defined as

$$s_{ij} = \frac{f(\theta_i, \theta_j)}{d_{ij}}, \tag{8}$$

where $\theta_i > 0$ represents the ability of node v_i to gain connections. Considering that the pairwise nodes with larger similarities are more likely to be connected, then, whether a pair of nodes are connected or not is determined not only by the distance between them but also by their respective abilities. Note that different systems may have different formations of function f , as shown in Fig. 1: for geometrical bordered networks, we have $f(\theta_i, \theta_j) = \theta_i + \theta_j$, where θ_i represents the radius of a bounded domain; for wireless networks, we have $f(\theta_i, \theta_j) = \max(\theta_i, \theta_j)$, where θ_i represents the sensing radius of a sensor; for the interactions between the super large objects, such like stars in the universe, we have $f(\theta_i, \theta_j) = \sqrt{\theta_i \times \theta_j}$, where θ_i represents the mass of an object and the corresponding s_{ij}^2 is proportional to the gravitation between the two involving objects.

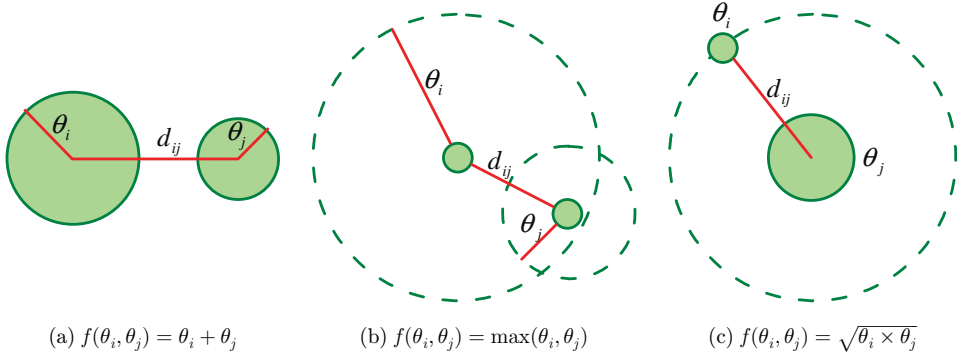


Fig. 1. (Color online) Different forms of function f to characterize different kinds of relationships. (a) For geometrical bordered networks, $f(\theta_i, \theta_j) = \theta_i + \theta_j$ with θ_i being the radius of a bounded domain. (b) For wireless networks, $f(\theta_i, \theta_j) = \max(\theta_i, \theta_j)$ with θ_i being the sensing radius of a sensor. (c) For the interactions between the stars in the universe, $f(\theta_i, \theta_j) = \sqrt{\theta_i \times \theta_j}$ with θ_i being the mass of a star.

In this paper, we focus on one of these relationships, i.e. $f(\theta_i, \theta_j) = \theta_i + \theta_j$, based on which we propose the heterogeneous distance matrix $H = [h_{ij}]_{N \times N}$ with its elements defined as

$$h_{ij} = \frac{1}{s_{ij}} = \frac{d_{ij}}{\theta_i + \theta_j}. \quad (9)$$

It is reasonable to expect that h_{ij} is a better metric than d_{ij} to measure whether there might be a link between a pair of nodes in Euclidean space, since Eq. (9) indicates that higher connecting abilities of the involving nodes will lead to smaller heterogeneous distance and thus benefit the establishment of the link between them. Denote by l_{ij}^h the heterogeneous shortest path length between a pair of nodes v_i and v_j , which minimizes the sum of the abilities of the constituent nodes in the path (including v_i and v_j). Then, the generalized distance in heterogeneous networks is further extended as

$$d_{ij} = (1 - 2\eta)(\theta_i + \theta_j) + 2\eta l_{ij}^h. \quad (10)$$

This generalization can be understood by considering each node v_i as a rigid sphere with its radius equal to θ_i . Their centers are arranged in a straight line with only the surfaces of neighbors touching each other when $\eta = 1$. In this case, we have $d_{ij} = 2l_{ij}^h - \theta_i - \theta_j$ meaning the distance between the centers of v_i and v_j . And all the rigid spheres touch each other when $\eta = 0$. In this case, we have $d_{ij} = \theta_i + \theta_j$. From Eqs. (9) and (10), we have

$$h_{ij} = \frac{(1 - 2\eta)(\theta_i + \theta_j) + 2\eta l_{ij}^h}{\theta_i + \theta_j}. \quad (11)$$

Based on these definitions, we can treat the network and triangle inequalities separately, as described by the following two theorems.

Theorem 2. *The heterogeneous distances satisfy the network inequality, i.e. the pairwise unconnected nodes have larger heterogeneous distances than the pairwise connected nodes.*

Proof. We have $h_{ij} = 1$ if v_i and v_j are connected, i.e. $l_{ij}^h = \theta_i + \theta_j$, and $h_{ij} > 1$ if they are not connected, i.e. $l_{ij}^h > \theta_i + \theta_j$. Thus *Theorem 2* is proved. \square

Theorem 3. *The generalized network distance in heterogeneous networks satisfy the triangle inequality, i.e. for every triple of nodes $\{v_i, v_j, v_k\}$, it satisfies $d_{ij} \leq d_{ik} + d_{jk}$, with the condition $d_{ik} \leq d_{jk} \leq d_{ij}$.*

Proof. Since it must be satisfied that $l_{ij}^h \leq l_{ik}^h + l_{jk}^h - \theta_k$, by Eq. (10), we have

$$\begin{aligned} d_{ik} + d_{jk} - d_{ij} &= (1 - 2\eta)(\theta_i + \theta_k) + 2\eta l_{ik}^h + (1 - 2\eta)(\theta_j + \theta_k) + 2\eta l_{jk}^h \\ &\quad - (1 - 2\eta)(\theta_i + \theta_j) - 2\eta l_{ij}^h \\ &\geq 2(1 - 2\eta)\theta_k + 2\eta\theta_k \\ &= 2(1 - \eta)\theta_k \geq 0, \end{aligned} \tag{12}$$

which indicates that the triangle inequality is satisfied in this case. Thus *Theorem 3* is proved. \square

Then, the new distance matrix D with its elements defined by Eq. (10) is used to calculate the centralized inner product matrix and further the coordinates of the nodes in certain Euclidean space by Eqs. (2) and (3), respectively. Denote by $H^* = [h_{ij}^*]_{N \times N}$ the embedded heterogeneous distance matrix with its elements calculated by

$$h_{ij}^* = \frac{\|x_i - x_j\|}{\theta_i + \theta_j}, \tag{13}$$

which is used to define the embedded network $G^* = (V, E^*)$, where the $|E|$ pairs of nodes with the smallest embedded heterogeneous distances are linked based on the new network inequality. Still denoted by A^* the adjacent matrix of the embedded network, the relative embedding error is then calculated by Eq. (5).

4. Embedding Results

In this part, MDS and HMDS are tested on two kinds of networks, i.e. the artificial Apollonian networks and real geometrical and social networks. We first introduce these networks and then show the embedding results.

4.1. Artificial and real-world networks

Apollonian networks³⁴ may be the most famous artificial geometrical networks, where nodes are solid spheres of different scales and two nodes are linked if the corresponding spheres touched on surface. This network is rooted from the problem of space-filling packing of spheres according to the ancient Greek mathematician Apollonius of Perga. In two-dimensional Euclidean space, three circles touch each

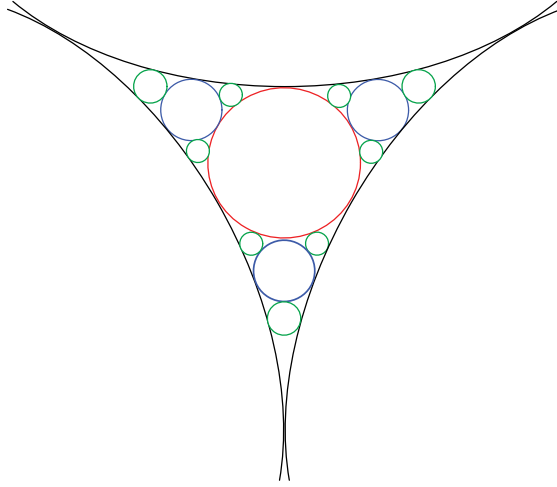


Fig. 2. (Color online) A two-dimensional Apollonian packing of disks. In the Apollonian network, consider each disk as a node, and two nodes are linked if the corresponding disks touch each other.

other and the hole between them is filled by the circle that touches all the three, forming again three smaller holes that are then filled by three smaller circles in the same way, and so forth, as shown in Fig. 2. This procedure can be further generalized to higher dimensions.⁴² Interestingly, such networks have particularly rich properties, such as scale-free, small-world, high clustering and so on.

Here, the Apollonian networks are established based on the iterative mechanism proposed by Zhang *et al.*,⁴² with their topological structures controlled by two parameters, i.e. the dimension D and the number of iterations T . The total numbers of nodes and edges are respectively calculated by

$$\begin{aligned}
 N &= D + 1 + \sum_{t=1}^T (D + 1)^{t-1} \\
 &= D + 1 + \frac{(D + 1)^T - 1}{D},
 \end{aligned} \tag{14}$$

$$\begin{aligned}
 |E| &= D + 1 + \sum_{t=1}^T (D + 1)^t \\
 &= D + 1 + \frac{(D + 1)^{T+1} - D - 1}{D}.
 \end{aligned} \tag{15}$$

The six tested Apollonian networks are generated by the model with the parameters $D = 2, 3$ and $T = 4, 5, 6$, respectively.

Besides, we also collected six real-world networks which are briefly introduced as follows. The country neighborhood network³⁸ was obtained from the current world map, where nodes represent countries and two nodes are linked if the corresponding countries are neighbors on territory. Here, only the giant cluster on the Asia, Europe

Table 1. The basic structural properties including the number of nodes N , the number of links $|E|$, the maximum and minimum degrees k_{\max} and k_{\min} , the heterogeneity $H = \langle k^2 \rangle / \langle k \rangle^2$, the average clustering coefficient C and the average shortest path length L for the six artificial Apollonian networks, generated by the model proposed by Zhang *et al.* with different parameters, and the six real-world networks.

Networks	N	$ E $	k_{\max}	k_{\min}	H	C	L
Apollonian							
$D = 2, T = 4$	43	123	24	3	1.74	0.80	2.15
$D = 2, T = 5$	124	366	48	3	2.38	0.81	2.57
$D = 2, T = 6$	367	1095	96	3	3.29	0.82	2.96
$D = 3, T = 4$	89	346	54	4	2.64	0.85	2.02
$D = 3, T = 5$	345	1370	164	4	5.31	0.88	2.32
$D = 3, T = 6$	1369	5466	488	4	11.4	0.88	2.58
Real-world							
Country	130	276	14	1	1.29	0.52	6.70
USAir	332	2126	139	1	3.46	0.63	2.74
Neuron	297	2148	134	1	1.80	0.29	2.46
Karate	34	78	17	1	1.69	0.57	2.41
Dolphin	62	159	12	1	1.33	0.25	3.11
Jazz	198	2747	100	1	1.40	0.62	2.24

and Africa continent is considered. The US airline network⁴³ was obtained from the National Transportation Atlas database in 1997, where nodes represent airports and links denote airlines between them. The neural network of *C. elegans*⁴⁴ was obtained from Newman's network data page,⁴⁵ where nodes represent neurons and two neurons are linked if there is at least one synapse or gap junction between them. The karate-club social network was collected by Zachary,⁴⁶ where nodes represent members of a karate club at a US university, and two members are linked if they are friends. The dolphin social network was compiled by Lusseau *et al.*,⁴⁷ where nodes represent dolphins in a community living off Doubtful Sound, New Zealand and two dolphins are associated if they present similar swimming patterns, including direction and movement, frequently. And the jazz network⁴⁸ was obtained from The Red Hot Jazz Archive digital database, where nodes represent the bands that performed between 1912 and 1940 and two bands are linked if they have at least one musician in common. Several basic structural properties²⁰ for the six artificial Apollonian networks and the six real-world networks are presented in Table 1.

4.2. Tests on artificial networks

As we can see, the main difference between MDS and HMDS is that the former treats the nodes in a network equally while the latter treats them differently according to their intrinsic abilities to gain connections. Such abilities may be reflected by some of their local topological properties in the network, e.g. node degree. Thus, here we simply set $\theta_i = k_i$ with k_i being the degree of node v_i in the network, considering that, in many real cases, only the network topology is known *a priori*.

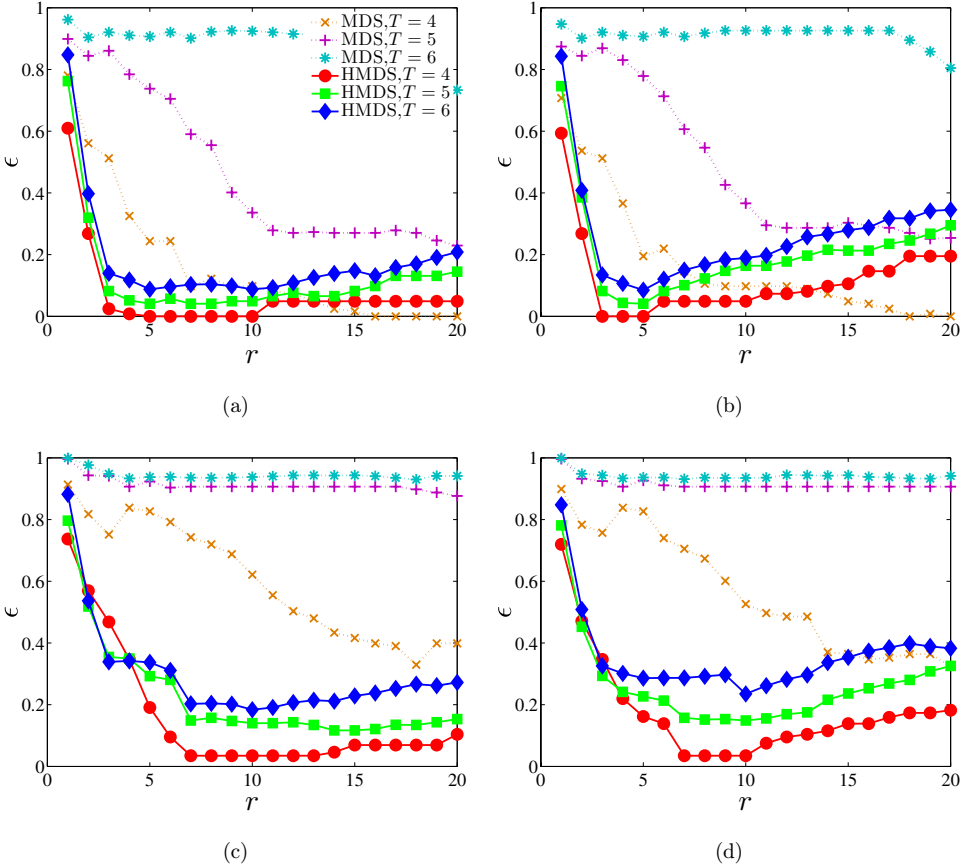


Fig. 3. (Color online) The relative embedding errors as functions of the dimensions of Euclidean spaces by MDS and HMDS for Apollonian networks with different parameters (a) $D = 2$ and $\eta = 0.5$, (b) $D = 2$ and $\eta = 1$, (c) $D = 3$ and $\eta = 0.5$ and (d) $D = 3$ and $\eta = 1$.

MDS and HMDS are first tested on Apollonian networks with different parameters $D = 2, 3$ and $T = 4, 5, 6$. The relative embedding errors as functions of the dimensions of Euclidean spaces for different cases are shown in Fig. 3. We find that HMDS behaves far better than MDS in each case, and MDS loses its efficiency when the size of Apollonian network is relatively large, i.e. the relative embedding error is close to 1 when the network size $N \geq 345$ ($D = 2, T = 6$, or $D = 3, T \geq 5$). In fact, by adopting HMDS, the relative embedding errors decrease most when $r \leq 3$ as the dimension of Euclidean space increases, which indicates that the Apollonian networks can be embedded into low-dimensional Euclidean spaces by this method. In fact, by HMDS, the Apollonian network can be always embedded into low-dimensional Euclidean spaces, i.e. $r \leq 3$ in this experiment, when $0.5 \leq \eta \leq 1$. However, the relative embedding errors increase fast when $\eta \rightarrow 0$.

Recently, several methods were proposed to identify the fractal dimensions of real-world networks based on their topological structure,^{49–51} which can be used to measure the self-similarities of these networks and thus help to inspire new models.⁵² However, there is still lack of methods to determine the appropriate dimension of the Euclidean space to embed a network, which is important because it tells the number of the most significant characters that influence the connections between individuals. Here, we use the minimum dimension satisfying

$$g(r) = \epsilon_r - \epsilon_{r+1} < \delta, \quad (16)$$

where ϵ_r is the relative embedding error in r -dimensional Euclidean space and δ is a predefined threshold, to estimate the appropriate dimension r_a . For example, r_a is equal to 3 for all the Apollonian networks with different parameters if the threshold is set to $\delta = 0.1$. Such results fit the real dimensions of these Apollonian networks quite well, considering that here we just use node degrees to estimate the abilities of individuals to gain connections and use their generalized network distances to estimate the real Euclidean distances between them.

When a network is embedded into an Euclidean space, each node v_i has its coordinates x_i , then the embedded distances and the embedded heterogeneous distances between pairwise nodes by adopting MDS and HMDS can be calculated by Eqs. (4) and (13), respectively. Since the embedded networks and further the relative embedding errors are totally determined by the embedded distances for MDS or the embedded heterogeneous distances for HMDS, it is also interesting to investigate the difference between the embedded distance (or embedded heterogeneous distance) distributions of pairwise connected and unconnected nodes in the original networks. In particular, when the networks are embedded into three-dimensional Euclidean space, the embedded distance distributions by adopting MDS for pairwise connected and unconnected nodes in Apollonian networks with different network parameters $D = 2, T = 6$ and $D = 3, T = 6$ are shown in Figs. 4(a) and 4(c), respectively, where the centers of the distributions of pairwise connected and unconnected nodes are very close to each other. On the other hand, the corresponding embedded heterogeneous distance distributions by adopting HMDS are shown in Figs. 4(b) and 4(d), respectively, where the centers of the distributions of pairwise connected and unconnected nodes are much farther from each other. Here, the parameter of generalized network distance is set the same, i.e. $\eta = 0.5$. These indicate that, by MDS, there lacks appropriate thresholds of distance to distinguish the pairwise connected and unconnected nodes, as a result, the relative embedding errors are surprisingly large in this case, i.e. $\epsilon > 0.9$; while by HMDS, we can indeed find such appropriate thresholds of heterogeneous distance, leading to the much smaller relative embedding errors.

4.3. Tests on real-world networks

Besides, we also test MDS and HMDS on the six real-world networks, including country neighborhood network, airline network in the US, neural network of

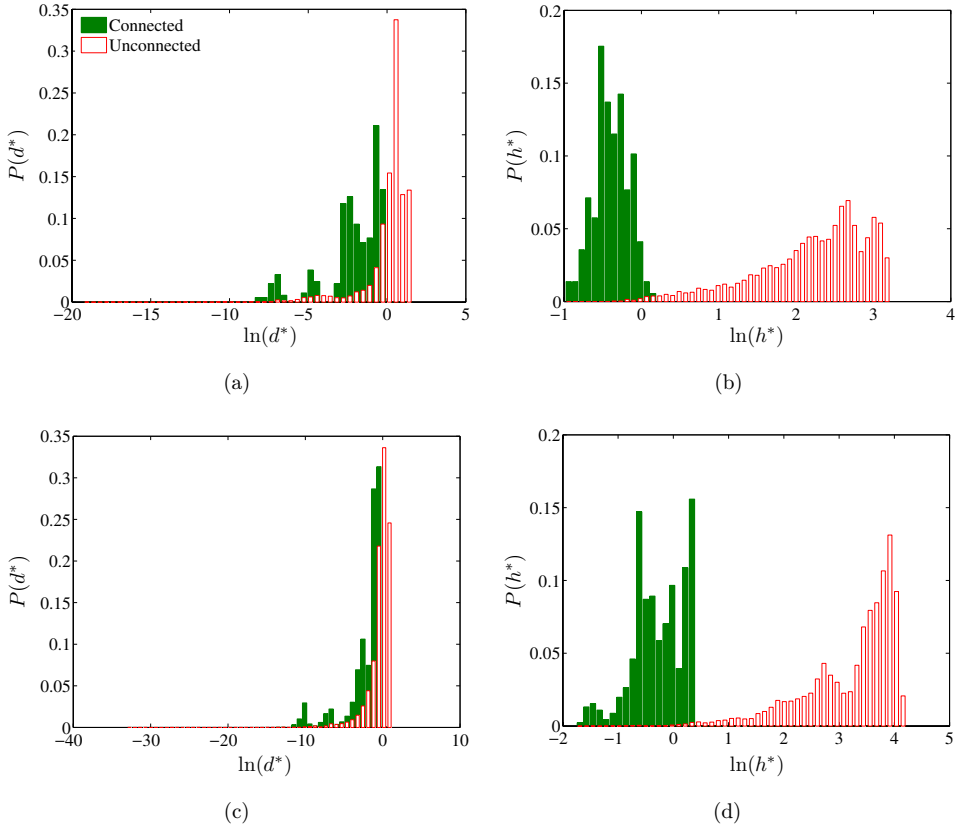


Fig. 4. (Color online) The embedded distance and the embedded heterogeneous distance distributions for pairwise connected and unconnected nodes in Apollonian networks with (a) $D = 2$ by adopting MDS, (b) $D = 2$ by adopting HMDS, (c) $D = 3$ by adopting MDS and (d) $D = 3$ by adopting HMDS. The other parameters are set the same, i.e. $T = 6$ and $\eta = 0.5$. Here, the networks are embedded into three-dimensional Euclidean space.

C. elegans, karate-club social network, dolphin social network and jazz social network, as presented in Table 1. The first three are geometrical networks, where the nodes have explicit coordinates in the respective Euclidean spaces, while the others are social networks, where the nodes represent people or dolphins.

The relative embedding errors as functions of the dimensions of Euclidean spaces by MDS and HMDS with different distance parameters $\eta = 0.1, 0.5, 1$ for the six real-world networks are shown in Figs. 5(a)–5(f). Similarly, HMDS behaves much better than MDS in embedding all the six real-world networks into relatively low-dimensional Euclidean space when $\eta \geq 0.5$, and this superiority seems more remarkable on the three geometrical networks and the karate-club social network with relatively large heterogeneity. Besides, we find that the distance parameter η has little effect on MDS but does significantly influence HMDS. To compare them in more detail, the relative embedding errors obtained by MDS and HMDS with $\eta = 0.5$ for the six

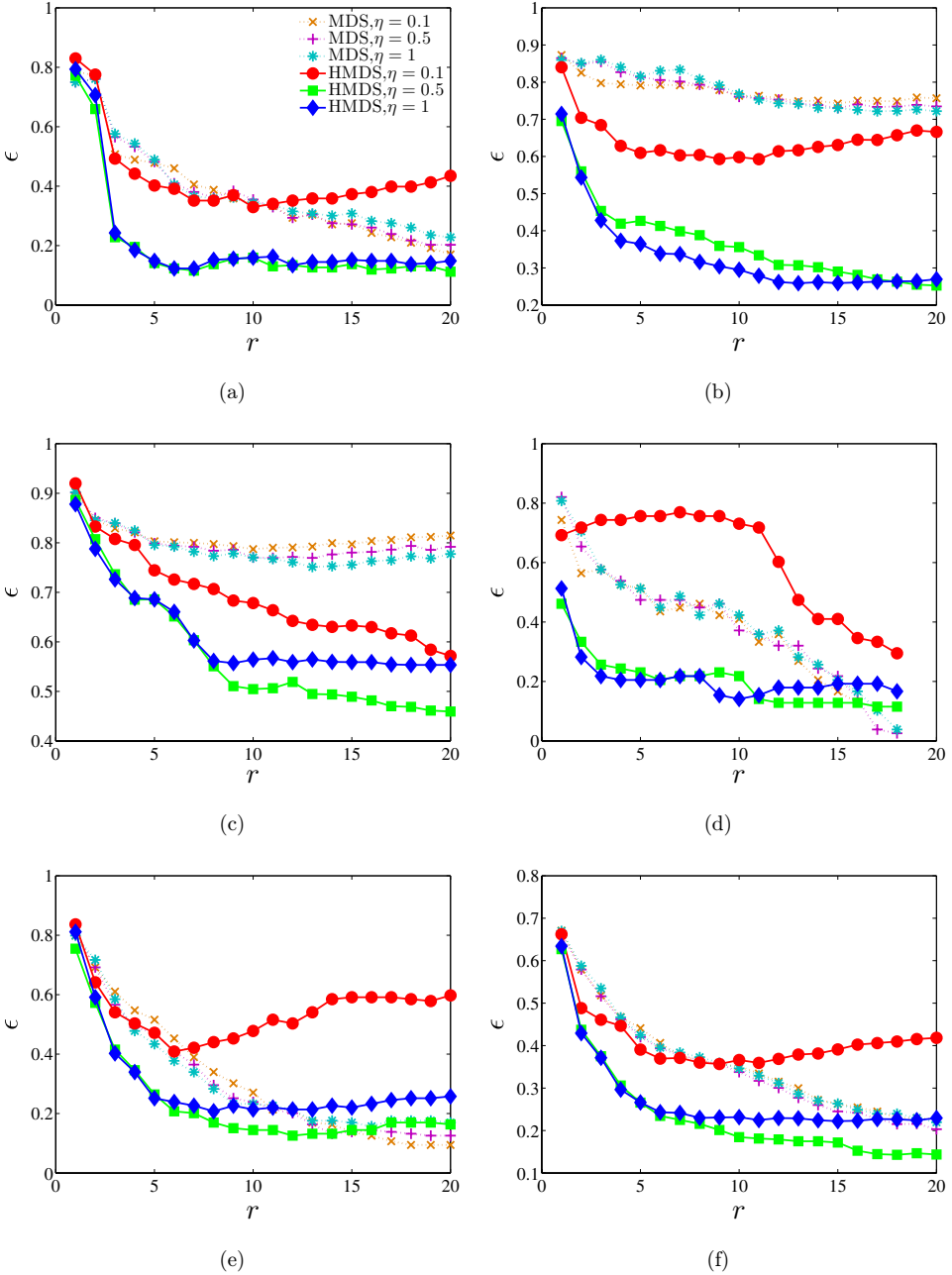


Fig. 5. (Color online) The relative embedding errors as functions of the dimensions of Euclidean spaces by MDS and HMDS with different distance parameters $\eta = 0.1, 0.5, 1$ for the six real-world networks: (a) country neighborhood network, (b) airline network in the US, (c) neural network of *C. elegans*, (d) karate-club social network, (e) dolphin social network and (f) jazz social network.

Table 2. The relative embedding errors for the six real-world networks by MDS and HMDS in different low-dimensional spaces, i.e. $r \leq 8$, with the distance parameters set to $\eta = 0.5$.

Dimension	1	2	3	4	5	6	7	8
MDS								
Country	0.75	0.76	0.57	0.53	0.48	0.41	0.38	0.36
USAir	0.86	0.85	0.85	0.83	0.81	0.81	0.80	0.79
Neuron	0.90	0.85	0.84	0.82	0.80	0.79	0.79	0.78
Karate	0.82	0.65	0.58	0.54	0.47	0.47	0.47	0.45
Dolphin	0.81	0.69	0.57	0.49	0.47	0.41	0.36	0.30
Jazz	0.66	0.58	0.52	0.46	0.42	0.39	0.38	0.37
HMDS								
Country	0.77	0.66	0.23	0.20	0.14	0.12	0.12	0.14
USAir	0.70	0.56	0.45	0.42	0.43	0.41	0.40	0.39
Neuron	0.89	0.81	0.74	0.69	0.69	0.65	0.60	0.55
Karate	0.46	0.33	0.26	0.24	0.23	0.21	0.22	0.22
Dolphin	0.75	0.57	0.42	0.35	0.26	0.21	0.20	0.17
Jazz	0.63	0.44	0.37	0.31	0.27	0.23	0.23	0.22

real-world networks in different low-dimensional Euclidean spaces, i.e. $r \leq 8$, are presented in Table 2.

By calculating the error decreasing functions $g(r)$ for the six real-world networks, as represented by Eq. (16), we estimate the dimension of Euclidean space for each network with different values of threshold δ . In particular, the estimated dimension r_a for the six real-world networks with different threshold δ from 0.01 to 0.1 are presented in Table 3. For the three geometrical networks including country neighborhood network, US airline network and neural network of *C. elegans*, the estimated dimensions r_a are equal to 3, 3 and 4, respectively, for a wide range of threshold δ . These results are close to the dimensions of their latent Euclidean spaces, i.e. 2, 2 and 3. The slight differences between the estimated and real dimensions may also be caused by the approximations for the individual abilities to gain connections and the Euclidean distances in HMDS. Meanwhile, it is shown that different social networks can also be embedded into the Euclidean spaces of relatively low dimensions, e.g. $r_a \leq 6$, which indicates that whether a social connection is established or

Table 3. The estimated dimensions r_a for the six real-world networks with different threshold δ from 0.01 to 0.1, by HMDS with the distance parameter set to $\eta = 0.5$.

Threshold ($\delta, \times 10^{-2}$)	1	2	3	4	5	6	7	8	9	10
Country	6	5	5	3	3	3	3	3	3	3
USAir	4	4	4	3	3	3	3	3	3	3
Neuron	4	4	4	4	4	3	3	1	1	1
Karate	6	3	3	3	3	3	3	2	2	2
Dolphin	6	6	6	6	6	5	3	3	3	3
Jazz	6	6	6	5	4	4	2	2	2	2

not between a pair of individuals in these networks may be just determined by a small number of factors.

5. Conclusions

In this paper, we proposed a new HMDS, by integrating the local structural properties of nodes into the traditional MDS, to embed heterogeneous networks into low-dimensional Euclidean space more effectively. The experiments show that HMDS indeed behaves much better than MDS in embedding artificial Apollonian networks and real-world networks into low-dimensional Euclidean spaces. Besides, we also provide a simple method to estimate the appropriate dimension of the Euclidean space to embed a network. This dimension indicates the number of significant factors that influence the connections between individuals in the network.

Note that, due to the insufficient information of many real-world networks on their node abilities to gain connections, here we just simply adopt the node degrees to estimate such abilities and then use their linear combination to define the heterogeneous and generalized network distances in heterogeneous networks. This treatment may cause the slight difference between the estimated and real dimensions of the embedding networks. Thus, our method can be expanded at least in the following two directions in the future: first, the intrinsic abilities of nodes to gain connections can be better estimated by using more local structural properties; and second, new definitions of heterogeneous and generalized network distances between pairwise nodes under the constraints of network and triangle inequalities, respectively, can be proposed by considering the different kinds of relationships in different systems. For some special cases such as wireless networks, the knowledge about the intrinsic abilities of individuals and the function to establish links can certainly facilitate the application of HMDS, and this method can also be used together with other localization methods to further improve the embedding results.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 61004097 and 61273212), the China Scholarship Council (CSC), the China Postdoctoral Science Foundation (Grant No. 2014M551770) and the Zhejiang Provincial Natural Science Foundation of China (Grant No. LQ12F03014).

References

1. J. Seebacher and A.-C. Gavin, *Cell* **144**, 1000 (2011).
2. M. Vidal, M. E. Cusick and A.-L. Barabási, *Cell* **144**, 986 (2011).
3. P. E. Vértes, A. F. Alexander-Bloch, N. Gogtay, J. N. Giedd, J. L. Rapoport and E. T. Bullmore, *Proc. Natl. Acad. Sci. USA* **109**, 5868 (2012).
4. V. M. Eguiluz, D. R. Chialvo, G. A. Cecchi, M. Baliki and A. V. Apkarian, *Phys. Rev. Lett.* **94**, 018102 (2005).

5. Q. Xuan and V. Filkov, in *Handbook of Human Computation* (Springer, New York, 2013), p. 791.
6. Q. Xuan and V. Filkov, in *Proc. 36th Int. Conf. Software Engineering*, (ACM, Hyderabad, 2014), p. 222.
7. Q. Xuan, F. Du and T.-J. Wu, *Chaos* **19**, 023101 (2009).
8. Q. Xuan, M. Gharehyazie, P. T. Devanbu and V. Filkov, in *Proc. 2012 Int. Conf. Social Informatics* (IEEE, Washington, 2012), p. 78.
9. Q. Xuan, F. Du, Y. Li and T.-J. Wu, *IEEE T. Autom. Sci. Eng.* **8**, 442 (2011).
10. P. S. Dodds, *Phys. Rev. Lett.* **104**, 048702 (2010).
11. Q. Xuan, F. Du, T.-J. Wu and G. Chen, *Phys. Rev. E* **82**, 046116 (2010).
12. Q. Xuan, F. Du, H. Dong, L. Yu and G. Chen, *Phys. Rev. E* **84**, 036101 (2011).
13. Q. Xuan, F. Du, L. Yu and G. Chen, *Phys. Rev. E* **87**, 032809 (2013).
14. W. Yu, G. Chen and J. Lü, *Automatica* **45**, 429 (2009).
15. W. Yu, G. Chen, Z. Wang and W. Yang, *IEEE Trans. Syst., Man, Cyber. (Part B)* **39**, 1568 (2009).
16. H. X. Hu, W. Yu, Q. Xuan, L. Yu and G. Xie, *IET Control Theory A* **7**, 1196 (2013).
17. H. X. Hu, A. Liu, Q. Xuan, L. Yu and G. Xie, *Syst. Control Lett.* **62**, 1125 (2013).
18. V. Bryant, *Etric Spaces: Iteration and Application* (Cambridge University Press, Cambridge, 1985).
19. Q. Xuan, Y. Li and T. J. Wu, *Physica A* **388**, 1257 (2009).
20. L. D. F. Costa, F. A. Rodrigues, G. Travieso and P. R. V. Boas, *Adv. Phys.* **56**, 167 (2007).
21. A. K. Jain, *Pattern Recogn. Lett.* **31**, 651 (2010).
22. H. Klock and J. M. Buhmann, *Pattern Recogn.* **33**, 651 (2000).
23. H. Fei, R. Jiang, Y. Yang, B. Luo and J. Huan, in *Proc. 20th ACM Int. Conf. Information and Knowledge Management*, (ACM, Glasgow, 2011), p. 995.
24. C. Herrmann, M. Barthélemy and P. Provero, *Phys. Rev. E* **68**, 026128 (2003).
25. I. Herman, G. Melançon and M. S. Marshall, *IEEE Trans. Vis. Comput. Graph.* **6**, 24 (2002).
26. J. Duch and A. Arenas, *Phys. Rev. E* **72**, 027104 (2005).
27. L. Lü and T. Zhou, *Physica A* **390**, 1150 (2011).
28. N. Li and J. C. Hou, *IEEE/ACM T. Netw.* **13**, 1313 (2005).
29. T. F. Cox and M. A. A. Cox, *Multidimensional Scaling*, 2nd edn. (Chapman and Hall/CRC, New York, 2001).
30. J. Venna and S. Kaski, in *Proc. European Symp. Artificial Neural Networks*, (D-side Publication Bruges, 2006), p. 557.
31. E. Bonabeau, *Inform. Sci.* **143**, 159 (2002).
32. M. Cardei, S. Yang and J. Wu, *IEEE Trans. Parallel Dis. Syst.* **19**, 545 (2008).
33. X. Han, X. Cao, E. L. Lloyd and C.-C. Shen, *IEEE Trans. Mobile Comput.* **9**, 643 (2010).
34. J. P. K. Doye and C. P. Massen, *Phys. Rev. E* **71**, 016128 (2005).
35. G. Caldarelli, A. Capocci, P. D. L. Rios and M. A. Muñoz, *Phys. Rev. Lett.* **89**, 258702 (2002).
36. Q. Xuan, Y. Li and T.-J. Wu, *Phys. Rev. E* **73**, 036105 (2006).
37. Q. Xuan, Y. Li and T.-J. Wu, *Physica A* **378**, 561 (2007).
38. Q. Xuan and T.-J. Wu, *Phys. Rev. E* **79**, 046106 (2009).
39. A. M. Bronstein, M. M. Bronstein and R. Kimmel, *Proc. Natl. Acad. Sci. USA* **103**, 1168 (2006).
40. F. K. W. Chan and H. C. So, *IEEE Trans. Sig. Proc.* **57**, 4548 (2009).
41. H. Jamali-Rad and G. Leus, *IEEE Trans. Sig. Proc.* **60**, 4485 (2012).
42. Z. Zhang, F. Comellas, G. Fertin and L. Rong, *J. Phys. A: Math. Gen.* **39**, 1811 (2006).

43. V. Batagelj and A. Mrvar, <http://vlado.fmf.uni-lj.si/pub/networks/data/> (2007).
44. D. J. Watts and S. H. Strogatz, *Nature* **393**, 440 (1998).
45. M. Newman, Network data, <http://www-personal.umich.edu/mejn/netdata/> (2011).
46. W. W. Zachary, *J. Anth. Res.* **33**, 452 (1977).
47. D. Lusseau, *Proc. R. Soc. Lond. B* **270**, S186 (2003).
48. P. M. Gleiser and L. Danon, *Adv. Comp. Syst.* **6**, 565 (2003).
49. C. Song, S. Havlin and H. A. Makse, *Nature* **433**, 392 (2005).
50. K.-I. Goh, G. Salvi, B. Kahng and D. Kim, *Phys. Rev. Lett.* **96**, 018701 (2006).
51. L. K. Gallos, C. Song and H. A. Makse, *Physica A* **386**, 686 (2007).
52. C. Song, S. Havlin and H. A. Makse, *Nat. Phys.* **2**, 275 (2006).