Node matching between complex networks

Qi Xuan* and Tie-Jun Wu
Department of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China
(Received 10 November 2008; revised manuscript received 26 March 2009; published 4 August 2009)

Revealing corresponding identities of a dedicated individual in several different complex systems is a common task in many areas, and this task is transferred to a node matching problem among complex networks in this paper. A feasible node matching algorithm based on network structure is proposed. Through solving node matching problems on different types of networks by our algorithm, it is revealed that the structure of the networks under study may significantly influence the final matching results. For example, it is found that higher matching precision can be obtained on random networks with moderate density of links, and the results on small-world networks are always better than those on random or regular networks. Moreover, in scale-free networks, it seems that hub nodes play dominant roles, i.e., better matching results can be expected by selecting nodes with larger degrees as the revealed matched nodes. These findings will help us design more efficient node matching algorithm in the future.

DOI: 10.1103/PhysRevE.80.026103 PACS number(s): 89.75.Hc, 89.65.Ef, 89.70.Eg, 89.75.Kd

I. INTRODUCTION

In the last decades, more and more complex systems are modeled by complex networks [1–9], and a series of research work, from network measuring [10], modeling [11–13], and dynamic analyzing [14] to network optimizing [15,16], have been done to study these collected complex network data. However, most of these researches always focus on just one target complex network and pay little attention to the interaction between different networks, which must be inappropriate in such a strongly connected world [17].

Take language networks, for example. Each language has its network structure by considering its words as nodes and co-occurrence of words in sentences as links [3]. Naturally, such a network must not be isolated. In fact, there are thousands of languages in the world today. Furthermore, each of these languages has its evolution history, that is, there were also lots of ancient languages in history. These language networks interact with each other all the time as their carriers, i.e., people, move from one place to another and communicate with each other day by day. Another more modern example is about communication networks. In the past, people could only communicate with each other face to face or through letters. With the advancement of communication techniques, especially the appearance of the internet, nowadays, people can communicate with each other promptly through more communication tools, e.g., phone, email, blog, BBS, and so on. Their log data record human interaction and can be transferred to corresponding communication networks [5–9]. Obviously, each of these communication networks is just a part of the global social network, and different communication networks must be overlapped because a person in the social network is usually active in several communication networks with different identities, e.g., phone numbers in phone networks [5,6], email addresses in email networks [7], user IDs in blog or BBS networks [8,9], and so on.

As is known to us, a concept can be expressed by different words in different language networks, and a person may be active in several communication networks with different identities. Namely, direct interactions among complex networks may be caused by the various roles of the same individual in different complex networks. So it will be of much importance to reveal the same individual based on the data collected from different complex networks, e.g., finding the words expressing the same concept in different language networks or finding the identities denoting the same person in different communication networks, etc. The solution of such node matching problems cannot only help us understand the coevolution rules among different complex networks so as to build more appropriate models for them but also has its direct applications, such as translating ancient writing, outlining the global social network by combining various communication networks, as is shown in Fig. 1(a), and so on.

The matching problems in different areas may have their dedicated solving methods. For example, an ancient writing researcher may use his knowledge about semantics to infer the concept of those ancient words. However, in this paper, we are only interested in and mainly focus on the following question: does the structure of the networks under study provide extra information in solving the node matching problem, i.e., finding matched nodes representing the same individual in different complex networks? The answer to this question is positive with the fact that, in this world, when two individuals have relationship in one complex network, they will be more likely to "contact" each other in another complex network if they both appear in that network. In other words, an individual shows similar behavioral patterns in different complex networks and such patterns can be partly reflected by its local structural properties in corresponding networks. For example, we investigated 24 QQ (the most famous Instant Messenger in China) users and revealed that, averagely speaking, they share telephone numbers of about 60% of their QQ friends, i.e., when two individuals communicate with each other in the QQ network, they will contact each other in the telephone network with a quite high probability close to 0.6. Such a result is also consistent with a public-opinion poll [18] that over 60% QQ

*Author to whom correspondence should be addressed; crestxq@hotmail.com

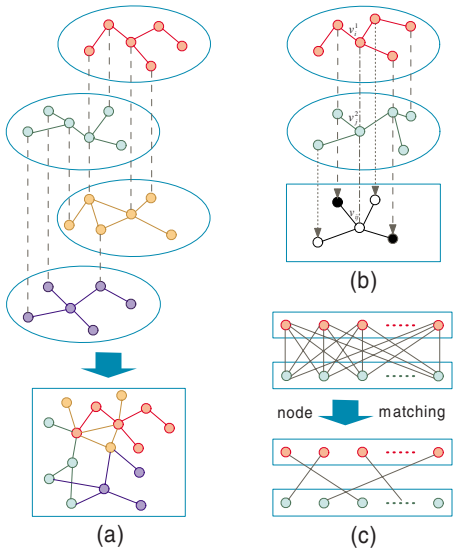

FIG. 1. (Color online) (a) Nodes in different networks representing the same individual are named as matched nodes that are connected by the brown (dark) dashed lines in this figure. The node matching problem is to reveal all pairs of matched nodes in different networks so as to outline the global social/language network. (b) The overlapped projection of the local worlds of the two nodes v_i^1 and v_j^2 from different networks. v_i^1 and v_j^2 have the common projection v_{ij}, and each pair of revealed matched nodes [connected by the brown (dark) dashed lines] around the nodes v_i^1 and v_j^2 also have the common projection denoted by a filled node in the projective plane, as a result, v_{ij} has four neighbors. Then the similarity between the node v_i^1 and the node v_j^2 can be calculated by the ratio of the number of the filled neighbors to the number of neighbors of v_{ij}, i.e., $S(v_i^1, v_j^2)=2/4=0.5$, in this figure. (c) The similarities between nodes belonging to different networks (excluding the revealed matched nodes) can be represented by a bipartite graph, where each link has a weight denoting the similarity between its two end nodes. Then the optimal matching problem for such a weighted bipartite graph is to find a set of nonadjacent weighted links to maximize the sum of their weights (similarities).

users are inclined to communicate with their acquaintances, e.g., families, schoolmates, etc., in the QQ network. The numbers of the total QQ friends as well as the numbers of the QQ friends with telephone numbers for the 24 investigated QQ users are plotted in Fig. 2.

In this paper, we think that an individual may behave similarly if it appears in different complex networks. The similarity between a pair of nodes belonging to different networks is calculated by their connections to several pairs of preliminarily revealed matched nodes. Then the node matching problem between two different networks is transferred to a maximum weighted bipartite matching problem [19], which can be solved by many well-known matching algorithms in graph theory [20–22]. The method is applied to different types of networks and some interesting results have been obtained. For example, it is found that higher matching precision can be derived on random networks with moderate density of links and the matching results on small-world networks are always better than those on random or regular networks. Moreover, it seems that better matching results could be derived by selecting those nodes with larger degree

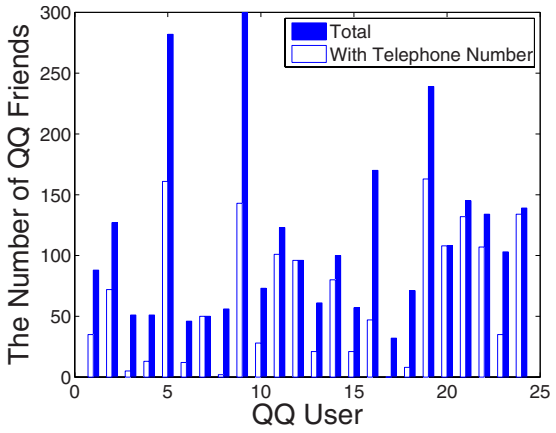

FIG. 2. (Color online) The numbers of the total QQ friends as well as the numbers of the QQ friends with telephone numbers for the 24 investigated QQ users. It is revealed that the ratio of the number of the QQ friends with telephone numbers to the number of the total QQ friends for all the investigated 24 QQ users is equal to $1574/2702=0.58$.

as the revealed matched nodes (the input of the algorithm), which is especially remarkable for scale-free networks.

Recently, a large number of texts written in different languages have been described as complex language networks [3,23]. It is found that many of them share several similar properties, such as small world, scale-free, and so on. Moreover, Amancio *et al.* [24] adopted several network measurements to evaluate the performance of different machine translations and found that some measurements return similar results when a language network was translated to another one. This finding suggests that the node matching algorithm proposed in this paper can be used to automatically translate texts provided some of their words have been translated correctly in the beginning.

The rest of the paper is organized as follows. In the next section, the node matching problem is defined and a node matching algorithm based on network structure is proposed. Then in Sec. III, the algorithm is adopted to solve the node matching problems for three different types of networks, i.e., the random networks, the small-world networks, and the scale-free networks, where the influences of the network structure on the matching results will be also carefully studied. Finally, the paper is concluded in Sec. IV.

II. NODE MATCHING ALGORITHM

Generally, a matching in a graph is a set of pairwise nonadjacent links, that is, no two links in the set share a common node [19]. Matching problems are often concerned with bipartite graphs, especially with weighted bipartite graphs where each link has an associated value. A maximum weighted bipartite matching is defined as a perfect matching (every node in the graph is incident to exactly one link of the matching) where the sum of the values of the links on the matching have a maximal value. It should be noted that if the graph is not completely bipartite, missing links are inserted with value of zero. Finding such a matching is known as the assignment problem [25].

The maximum weighted bipartite matching problem can be solved by the Bellman-Ford (BF) algorithm [20] with the time complexity $O(N^4)$ if there are N nodes in the focused graph. When the link weights are non-negative, the BF algorithm can be replaced by the Dijkstra algorithm with a lower time complexity $O(N^3)$. Besides, Kuhn developed a combinatorial optimization algorithm to solve the same problem. He named it Hungarian algorithm in honor of the work of König and Egerváry on which it is based [21]. A few years later, Munkres reviewed the algorithm and observed that its complexity is strongly polynomial. Since then the algorithm has been known remarkably also as the Kuhn-Munkres (KM) algorithm [22]. Similarly, the time complexity of the original KM algorithm was $O(N^4)$, and Edmonds and Karp, and independently Tomizawa noticed that it could be also modified to achieve an $O(N^3)$ running time.

Here we focus our work on node matching between different networks, which has received little attention in the literature. The problem can be transferred to a maximum weighted bipartite matching problem after calculating the similarities between nodes belonging to two different target networks, and thus can be solved by those well-known matching algorithms. Naturally, in a network, a node can be characterized by its topological properties [10], e.g., degree, clustering coefficient, betweenness, and so on. Therefore, it seems possible to reconstruct some types of networks so as to match nodes between them only from that topological information provided that a comprehensive set of such measurements is adopted [26]. However, recent research shows that real-world complex networks themselves are always highly symmetric [27], i.e., there are always lots of nodes sharing the same neighbors in a network, which will even prevent us from distinguishing them in one network only by considering their topological properties, not to mention matching them between different complex networks. Moreover, the focused complex networks are always just overlapped but not completely identical [18], i.e., the matched nodes may have a little difference in their local structural properties. Therefore, in most cases, it is unpractical to match nodes between different networks without considering other *a priori* knowledge.

Fortunately, there are always a part of matched nodes having been revealed in real-world complex networks, e.g., in communication networks, individuals may leave their email addresses or phone numbers when they register blog or BBS accounts. Moreover, still take QQ, for example: like MSN, QQ has its own email system, i.e., each QQ account is automatically assigned with a unique QQ email address, and the same public-opinion poll [18] shows that about 16% QQ users take the QQ email as their preferred email facility, which suggests that at least 16% pairs of matched nodes between the QQ network and the email network have been already revealed.

In the present paper, the revealed matched nodes as *a priori* knowledge as well as the network topological information will be adopted to design the node matching algorithm. The node matching problem then can be described as node matching problem between two different networks: two networks under study are denoted by $G_1=(V_1,E_1)$ and $G_2=(V_2,E_2)$, where $V_i=\{v_1^i,v_2^i,\ldots,v_{N_i}^i\}$ and E_i represent the node set and the link set of network i $(i=1,2)$, respectively. Without loss of generality, the M pairs of matched nodes $v_j^1 \leftrightarrow v_j^2$ in these two different networks are defined as $\{v_1^i,v_2^i,\ldots,v_M^i\} \subseteq V_i$ $(i=1,2)$ with $M \leq \min\{N_1,N_2\}$, while P_r $(P_r<M)$ pairs of them have been already revealed, named as revealed matched nodes, and denoted by $\{v_1^i,v_2^i,\ldots,v_{P_r}^i\} \subset V_i$ $(i=1,2)$. Then the problem is can we design a method to find the other $M-P_r$ pairs of matched nodes in these two distinct networks by using the information of the network structures of G_1 and G_2 and the revealed matched nodes?

A feasible node matching algorithm is presented by the following three steps:

(i) Similarity calculation: the similarity between two nodes belonging to different networks can be measured by the number of pairs of revealed matched nodes around them, e.g., the number of common friends they contact with in different communication networks, where a common friend is denoted by a pair of revealed matched nodes in corresponding communication networks. Then the similarity between nodes v_i^1 and v_j^2 can be calculated by Eq. (1),

$$S(v_i^1,v_j^2)=\frac{n_M(v_i^1,v_j^2)}{n_L(v_i^1)+n_L(v_j^2)-n_M(v_i^1,v_j^2)}, \tag{1}$$

where $n_M(v_i^1,v_j^2)$ denotes the number of pairs of revealed matched nodes (v_k^1,v_k^2) where v_i^1 and v_j^2 are mutually connected, i.e., v_i^1 is connected to v_k^1 and v_j^2 is connected to v_k^2, in the corresponding networks, and $n_L(v_i^1)$ [or $n_L(v_j^2)$] represents the number of links connected to the node v_i^1 (or v_j^2) in the network G_1 (or G_2). Equation (1) guarantees that the similarity between two nodes belonging to the different networks has the normalized value in [0, 1], and can be visually illustrated by the overlapped projection of the local worlds of the two nodes v_i^1 and v_j^2, as is shown in Fig. 1(b). In this figure, v_i^1 and v_j^2 have the common projection v_{ij}, and each pair of revealed matched nodes [connected by the brown (dark) dashed links] around the nodes v_i^1 and v_j^2 also have the common projection denoted by a filled node in the projective plane. As a result, v_{ij} has four neighbors. Then the similarity between the node v_i^1 and the node v_j^2 can be calculated by the ratio of the number of the filled neighbors to the number of neighbors of v_{ij}, i.e., $S(v_i^1,v_j^2)=2/4=0.5$.

(ii) Node matching: the similarities between nodes belonging to different networks (excluding the revealed matched nodes) can be represented by a bipartite graph $G_b=(U_1,U_2,W)$, where $U_i=\{v_{P_r+1}^i,v_{P_r+2}^i,\ldots,v_{N_i}^i\}$ $(i=1,2)$, and W denotes the set of links weighted by the similarities S between these two groups of nodes. Then the node matching problem between G_1 and G_2 can be transferred to a maximum matching problem [19] for the bipartite graph G_b, i.e., under the assumption $N_1 \leq N_2$, finding a set of nonadjacent weighted links $\{w_1,w_2,\ldots,w_{N_1-P_r}\}$ to maximize the sum of their weights $\Sigma_{i=1}^{N_1-P_r} s_i$, as is shown in Fig. 1(c). This maximum weighted bipartite matching problem can be solved by the KM algorithm [22]. It should be noted that, although the KM algorithm was developed for the case $N_1=N_2$, it could be also feasible in the case $N_1<N_2$ through factitiously adding N_2-N_1 isolated nodes in G_1. Then N_2-P_r pairs of matched nodes can be derived by the KM algorithm, where

N_2-N_1 pairs of them containing factitious nodes are rejected automatically while the other N_1-P_r pairs as well as their similarities are reserved.

(iii) Matched pairs selection: if the value of M is known *a priori*, the only thing we need to do is to sort N_1-P_r pairs of matched nodes by their attached similarities, then select the top $M-P_r$ pairs with larger similarities as the final pairs of matched nodes. However, the value of M may be unknown in real-world complex networks. In such a situation, a threshold $\theta \in [0,1)$ must be provided and those pairs of matched nodes with similarities larger than θ then are selected as the final pairs of matched nodes. Statistically, a proper value of the threshold θ should satisfy Eq. (2),

$$\frac{1}{M_\theta}\sum_{s_i>\theta} s_i \approx \frac{1}{P_r}\sum_{i=1}^{P_r} S(v_i^1, v_i^2), \tag{2}$$

where M_θ denotes the number of selected pairs of matched nodes with similarities larger than θ.

III. MATCHING EXPERIMENTS

In order to test the node matching algorithm proposed in the last section, two interactional complex networks with M pairs of matched nodes should be created at first. Here, for convenience, the parameters are set to be $N_1=N_2=M=N$. Generally, there are two ways to create a pair of interactional networks, as is shown in Figs. 3(a) and 3(b), respectively, both of which may work in reality. One way is that the pair of interactional networks G_1 and G_2 are evolved from a common original network; in other words, they are derived from the same network (obtained by some model) through random rewiring. The other ways are introduced as follows:

(i) Networks initialization: two networks G_1 and G_2 with N nodes, respectively, are created by the same rule, where all the nodes are randomly matched, i.e., N pairs of randomly matched nodes $v_i^1 \leftrightarrow v_i^2$ are provided.

(ii) Interaction: if v_i^1 (or v_i^2) and v_j^1 (or v_j^2) is connected in G_1 (or G_2) while v_i^2 (or v_i^1) and v_j^2 (or v_j^1) is not connected in G_2 (or G_1), then connect v_i^2 (or v_i^1) and v_j^2 (or v_j^1) with probability η_1 (or η_2).

In this paper, the second way is adopted to create pairs of tested interactional networks. Then P_r $(P_r<N)$ pairs of matched nodes are selected as the revealed matched nodes. So there will be totally $N-P_r$ pairs of nodes needed to be further matched. If P_c $(P_c \le N-P_r)$ pairs of them are revealed correctly, then the matching precision ϕ of the algorithm can be calculated by Eq. (3),

$$\phi = \frac{P_c}{N-P_r}. \tag{3}$$

A. Node matching between random networks

Considering that G_1 and G_2 are all P. Erdős and A. Rényi (ER) random networks [28] and $\eta_1=\eta_2=\eta$, then the three parameters including the connection probability ρ of each random network, the interactional degree η, and the sample ratio $\gamma=P_r/N$, i.e., the proportion of the randomly selected

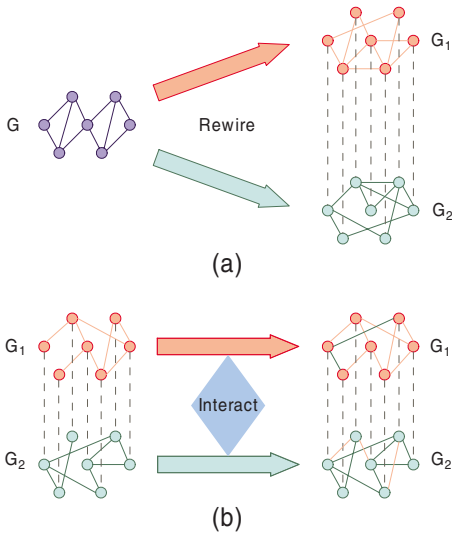

FIG. 3. (Color online) Two ways to create a pair of interactional networks. (a) The pair of interactional networks G_1 and G_2 are derived from the same original network (obtained by some model) through random rewiring. The corresponding nodes are matched [connected by brown (dark) dashed lines]. (b) In the beginning, two networks G_1 and G_2 are obtained by the same model, where all the nodes are randomly matched [connected by brown (dark) dashed lines]. Then the pair of interactional networks G_1 and G_2 are derived by interacting with each other, i.e., two nonlinked nodes in the network G_1 are connected by a green (dark) line with probability η_2 if their corresponding matched nodes in G_2 are linked while two nonlinked nodes in G_2 are connected by a red (gray) line with probability η_1 if their corresponding matched nodes in G_1 are linked. η_1 and η_2 are named as interactional degree. In this paper, the second way is adopted to create pairs of tested interactional networks.

revealed matched nodes, will influence the final matching precision ϕ. Generally, the matching precision ϕ will increase monotonically as the interactional degree η or the sample ratio γ increases while the influence of the connection probability ρ on the matching precision is more complicated.

When $\rho=0$, all of the nodes in G_1 and G_2 are isolated, that is, following Eq. (1), the similarity $S(v_i^1, v_j^2)=0$ must be satisfied for each pair of nodes v_i^1 and v_j^2, which must further result in a zero matching precision. As the densities (defined by the connection probability ρ) of the networks increase, a fixed number of randomly selected revealed matched nodes can provide more detailed structural information for other nodes in the networks. As a result, the matching precision ϕ will be improved. However, too many links (i.e., $\rho \to 1$) may cause random networks getting more symmetric, which, on the contrary, will decrease the matching precision ϕ. For instance, when $\rho=1$, each node in G_1 or G_2 is connected to all pairs of revealed matched nodes, that is, following Eq. (1), each pair of nodes v_i^1 and v_j^2 will have the same similarity $S(v_i^1, v_j^2)=P_r/(N-1)$, which will also result in a zero matching precision.

In this experiment, the number of nodes in each random network is set to be $N=100$, and the relationships between the matching precision ϕ and the connection probability ρ

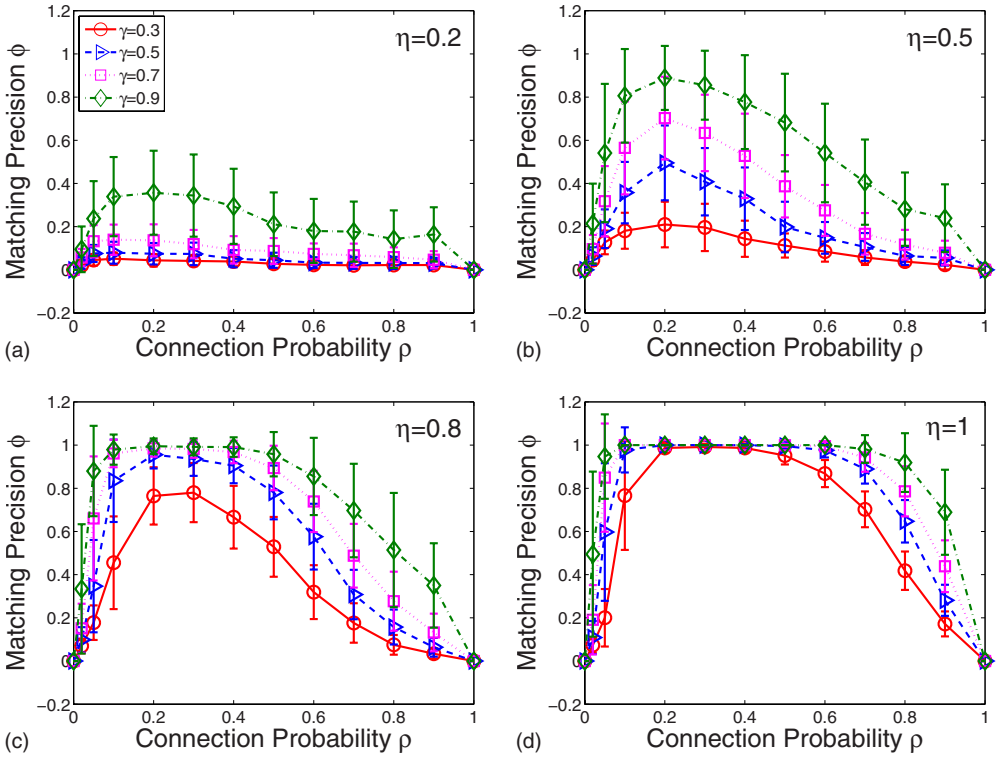

FIG. 4. (Color online) The relationships between the matching precision ϕ and the connection probability ρ for various sample ratios γ=0.3,0.5,0.7,0.9 and different interactional degrees η=0.2,0.5,0.8,1. For each parameter set $\{\rho, \eta, \gamma\}$, the experiment is implemented on 100 different pairs of random networks, each of which has N=100 nodes.

for various sample ratios γ=0.3,0.5,0.7,0.9 and different interactional degrees η=0.2,0.5,0.8,1 are shown in Figs. 4(a)–4(d). It is found that, as long as the interactional degree $\eta>0$, the node matching algorithm will take effect although the matching precision ϕ is quite low when the value of η is much smaller. For example, if η=0.2, the matching precision ϕ is always lower than 0.5 even when there is a very high proportion of the revealed matched nodes, i.e., γ=0.9, as is shown in Fig. 4(a). Moreover, as is expected, the polarization of the connection probability, i.e., $\rho\to 0$ or $\rho\to 1$, will also result in a quite low matching precision, as is shown in Figs. 4(a)–4(d).

Furthermore, suppose the maximal matching precision is obtained when $\rho=\rho_{max}(\eta,\gamma)$ for each pair of η and γ, as is shown in Figs. 4(a)–4(d), it could be found that the inequality $0<\rho_{max}(\eta,\gamma)<0.3$ is always satisfied. More interestingly, $\rho_{max}(\eta,\gamma)$ gets larger as the interactional degree η increases, i.e., the peaks of the curves shift rightward in Fig. 4 from (a) to (d). In order to provide a clearer result, the relationship between ρ_{max} and η for a constant sample ratio γ =0.3 is shown in Fig. 5(a). Figure 5(b) shows that, when η $\geq$0.2 [the much larger standard deviation makes the value of ρ_{max} at η=0.1 unreliable, as a result, it is excluded in Fig. 5(b)], ρ_{max} increases linearly as η increases, i.e., ρ_{max} =0.11η+0.18, with a relatively small fitting error. In parallel to this, the relationship between ρ_{max} and γ for a constant interactional degree η=0.5 is shown in Fig. 6(a). The extra small slope of the red (dark) fitted line in this figure suggests that ρ_{max} is almost independent of the sample ratio γ.

B. Node matching between small-world networks

As more and more real-world complex networks are collected and analyzed, it is widely believed that real-world networks are always far different from ER random networks. In fact, it is found that many of those real-world complex networks not only have small average shortest path length as ER random networks but also are highly clustered as regular networks. In order to explain this phenomenon, Watts and Strogatz (WS) proposed a simple small-world network model by introducing randomness into a regular network through a rewiring process [1]. In the WS model, the tradeoff between an ER random network and a regular network can be adjusted by the rewiring probability p, and for some intermediate values of p, the network presents small-world property, i.e., the network has a small average shortest path length, and at the same time, is highly clustered.

Obviously, in the WS model, the rewiring probability p plays an important role in determining the structure of the network. Thereby, it is always very interesting to investigate the influence of the rewiring probability p on the network dynamic properties. For example, Walsh [29] found that it is always more difficult (time consuming) to search in a small-world network ($0<p<1$) than in a regular network (p=0) or a random network (p=1). In this section, we will study the influence of the rewiring probability p on the node matching problem, and try to reveal if the higher matching precision can be obtained on small-world networks than on regular networks or random networks.

Because the randomly node matching as well as the interaction between the different regular networks has the similar

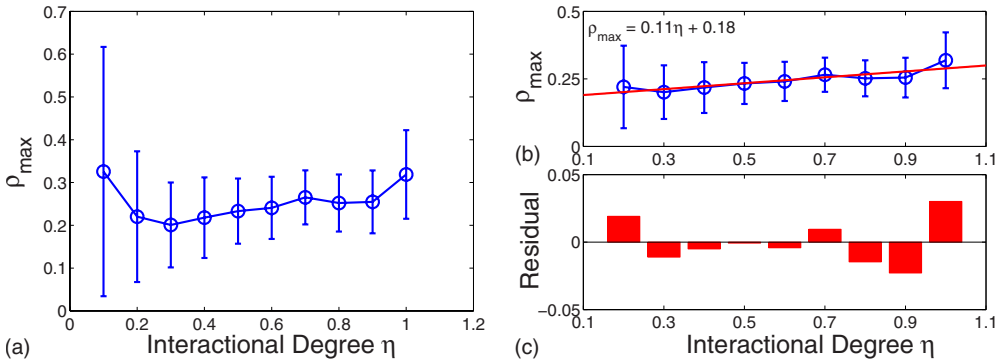

FIG. 5. (Color online) (a) The right value of the connection probability ρ_{max}, where the highest matching precision is derived (if the highest matching precision is derived at several values of ρ, one of them is randomly selected as ρ_{max}), as a function of the interactional degree η when the sample ratio is fixed to $\gamma=0.3$. For each parameter set $\{\rho,\eta\}$, the experiment is implemented on 100 different pairs of random networks, each of which has $N=100$ nodes. It is shown that ρ_{max} has a much larger standard deviation when $\eta=0.1$. (b) When $\gamma=0.3$ and $\eta\geq 0.2$, there seems to be a linear relationship between ρ_{max} and η, i.e., $\rho_{max}=0.11\eta+0.18$, fitted by the red (dark) line with a relatively small mean squared error (MSE) equal to 2.54×10^{-4}. (c) The corresponding residual plot of the fit.

effect as the rewiring process on creating small-world networks, in order to clearly present the influence of the rewiring probability p on the matching precision ϕ, the networks G_1 and G_2 are set to be totally the same, i.e., they are just two copies of the network G derived by the WS model with a rewiring probability p, and a pair of copies of the same node in G is considered to be a pair of matched nodes. Namely, if v_i^1 and v_j^1 is connected in G_1, v_i^2 and v_j^2 must be also connected in G_2, and vice versa.

Each small-world network in the experiment has $N=100$ nodes as well as an average degree $\langle k\rangle=10$, and the relationships between the matching precision and the rewiring probability p are depicted in Figs. 7(a)–7(d) for various sample ratios $\gamma=0.2,0.3,0.4,0.5$, respectively. Generally, much lower matching precision is derived by KM algorithm on regular networks ($p=0$) than on small-world networks ($0<p<1$) or random networks ($p=1$). Moreover, when $\gamma=0.2$ and $\gamma=0.3$, as is shown in Figs. 7(a) and 7(b), the matching precision climbs rapidly at the beginning as the randomness increases and gets to its peak at about $p=0.3$, then it drops slowly until $p=1$. This finding reveals that, at proper values of the sample ratio γ, the matching results on small-world networks are indeed a little better than those on regular networks or random networks. However, in this experiment, for each rewiring probability $p\in[0,1]$, Eqs. (4) and (5) must be satisfied,

$$\gamma\to 0\Rightarrow\phi\to 0, \tag{4}$$

$$\gamma\to 1\Rightarrow\phi\to 1. \tag{5}$$

As a result, the slight superiority of the small-world networks over the random networks on the node matching problem will soon disappear as the sample ratio γ further decreases or increases, which is partially verified by Figs. 7(c) and 7(d) for $\gamma=0.4$ and $\gamma=0.5$, respectively.

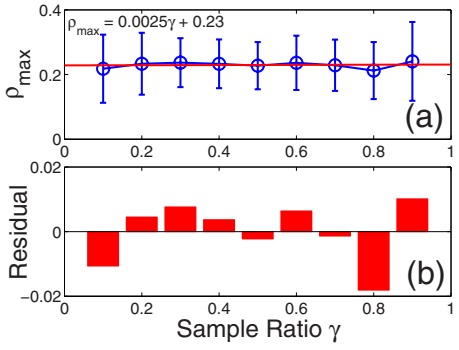

FIG. 6. (Color online) (a) The right value of the connection probability ρ_{max}, where the highest matching precision is derived (if the highest matching precision is derived at several values of ρ, one of them is randomly selected as ρ_{max}), as a function of the sample ratio γ when the interactional degree is fixed to $\eta=0.5$. For each parameter set $\{\rho,\gamma\}$, the experiment is implemented on 100 different pairs of random networks, each of which has $N=100$ nodes. The extra small slope of the red (dark) fitted line (with the MSE equal to 7.86×10^{-5}) suggests that ρ_{max} almost keeps constant when γ changes from 0.1 to 0.9. (b) The corresponding residual plot of the fit.

C. Node matching between scale-free networks

Besides the small-world property, many real-world complex networks also share a heterogeneous structure characterized by a power-law degree distribution. In these scale-free networks, there are always a few hub nodes obtaining most links while lots of other nodes have very few links. The first scale-free network model is proposed by Barabási and Albert (BA) [2] with the following modeling process: start with a small number m_0 of nodes connected with each other, add a new node at every time step, and connect it to m ($m\leq m_0$) different nodes that are selected with a probability linearly proportional to their degrees, then, after T time steps, a scale-free network with $N=m_0+T$ nodes is created.

Generally, the hub nodes play much more important roles in the dynamics of scale-free networks than other nodes. For instance, as is revealed by Albert *et al.* [30], once those hub nodes are attacked, the average shortest path length of a scale-free network will increase quickly; as a result, the com-

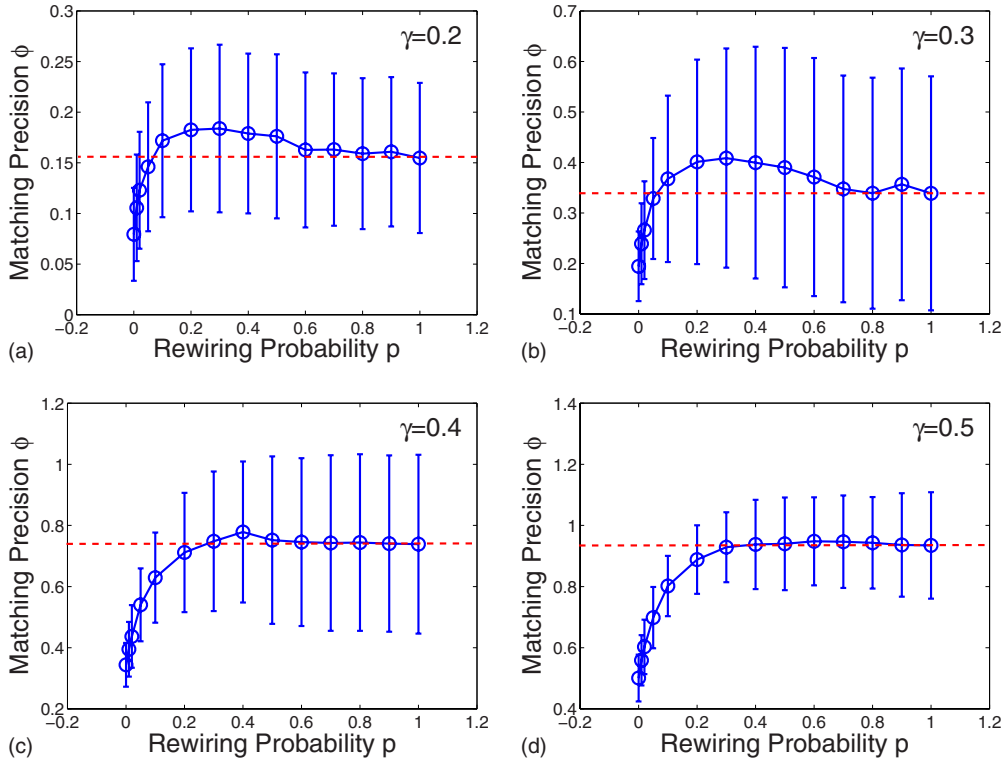

FIG. 7. (Color online) The relationships between the matching precision ϕ and the rewiring probability p for small-world networks under various sample ratios $\gamma=0.2,0.3,0.4,0.5$. For each parameter set $\{p,\gamma\}$, the experiment is implemented on 1000 different pairs of small-world networks, each of which has $N=100$ nodes and an average degree $\langle k\rangle=10$. The red (dark) dashed lines denote the matching precision of the algorithm on the random networks when $p=1$.

munication efficiency of the network will be largely weakened. For the node matching problem between scale-free networks, the hub nodes also have their outstanding status. Based on the interactional model proposed in this paper, denoting the degree of v_i^1 by d_i^1 and the degree of v_j^2 by d_j^2, if they are randomly selected as a pair of matched nodes, then, averagely speaking, there are $d_i^1 d_j^2/N$ other pairs of matched nodes around them before the interaction. Moreover, after the interaction, the degree of v_i^1 and that of v_j^2 can be calculated by Eqs. (6) and (7), respectively,

$$\widetilde{d_i^1} = d_i^1 + d_j^2\left(1 - \frac{d_i^1}{N}\right)\eta_2, \tag{6}$$

$$\widetilde{d_j^2} = d_j^2 + d_i^1\left(1 - \frac{d_j^2}{N}\right)\eta_1. \tag{7}$$

Also, the number of pairs of other matched nodes around the matched nodes v_i^1 and v_j^2 after the interaction can be calculated by Eq. (8),

$$F_{ij} = d_j^2\left(1 - \frac{d_i^1}{N}\right)\eta_2 + d_i^1\left(1 - \frac{d_j^2}{N}\right)\eta_1 + \frac{d_i^1 d_j^2}{N}. \tag{8}$$

Equations (6)–(8) can be simplified to Eqs. (9)–(11), respectively, because real-world complex networks always have a very huge number of nodes as well as a relatively small average degree,

$$\widetilde{d_i^1} \approx d_i^1 + \eta_2 d_j^2, \tag{9}$$

$$\widetilde{d_j^2} \approx d_j^2 + \eta_1 d_i^1, \tag{10}$$

$$F_{ij} \approx \eta_1 d_i^1 + \eta_2 d_j^2. \tag{11}$$

Then Eq. (12) can be derived,

$$F_{ij} \approx \begin{cases} \dfrac{\eta_1(1-\eta_2)}{1-\eta_1\eta_2}\widetilde{d_i^1} + \dfrac{\eta_2(1-\eta_1)}{1-\eta_1\eta_2}\widetilde{d_j^2}, & \eta_1\eta_2 < 1, \\ \dfrac{1}{2}\widetilde{d_i^1} + \dfrac{1}{2}\widetilde{d_j^2}, & \eta_1\eta_2 = 1. \end{cases} \tag{12}$$

Following the formulation of F_{ij}, statistically, there are more pairs of unrevealed matched nodes around a pair of revealed matched nodes v_i^1 and v_j^2 with larger F_{ij}, i.e., in order to improve the final matching precision ϕ, a natural strategy is to sort all the pairs of matched nodes by F_{ij} in descending order, then selecting the top P_r pairs of matched nodes as the revealed matched nodes. However, this strategy seems unpractical because, in reality, it is supposed that the matched nodes are unknown beforehand. Fortunately, Eq. (12) indicates a substitute way, i.e., selecting nodes with larger degree in one target network, revealing their matched nodes in the other target network, then these pairs of matched nodes are set to the revealed matched nodes of the algorithm. Particu-

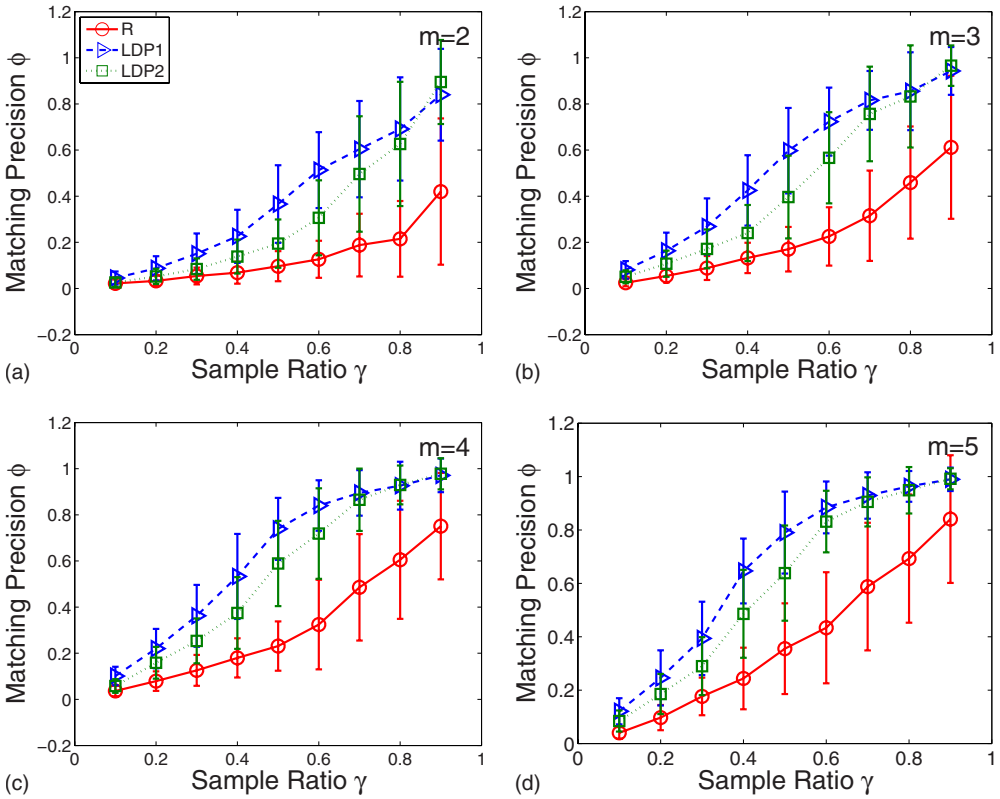

FIG. 8. (Color online) The relationships between the matching precision ϕ and the sample ratio γ by adopting the three different revealed matched nodes selection rules, i.e., R, LDP1, and LDP2, for scale-free networks with different parameters. (a) With parameters $N=100$, $m=m_0=2$. (b) With parameters $N=100$, $m=m_0=3$. (c) With parameters $N=100$, $m=m_0=4$. (d) With parameters $N=100$, $m=m_0=5$. For each parameter set $\{\gamma,m,m_0\}$ and each selection rule, the experiment is implemented on 100 different pairs of scale-free networks. It is shown that, generally, better matching results can be derived by selecting revealed matched nodes following the degree based rules LDP1 and LDP2 than following the random selection rule R. Moreover, in most cases, LDP1 is prior to LDP2 when $\eta_1 > \eta_2$ in this experiment.

larly, three feasible revealed matched node selection rules are proposed as follows:

(i) Random (R): P_r nodes in G_1 (or G_2) are randomly selected, and these nodes as well as their matched nodes in G_2 (or G_1) are selected as the revealed matched nodes.

(ii) Large degree priority in G_1 (LDP1): the nodes in the network G_1 are sorted by their degree in descending order, and their top P_r as well as their matched nodes in G_2 are selected as the revealed matched nodes.

(iii) Large degree priority in G_2 (LDP2): the nodes in the network G_2 are sorted by their degree in descending order, and their top P_r as well as their matched nodes in G_1 are selected as the revealed matched nodes.

Equation (12) suggests that, based on the interactional model proposed in this paper, statistically, better matching results can be always derived by selecting revealed matched nodes following the degree based rules LDP1 and LDP2 than following the random selection rule R. Moreover, LDP1 would be prior to LDP2 when $\eta_1 > \eta_2$, and vice versa. It should be noted that LDP1 (or LDP2) still takes effect even when $\eta_1=0$ (or $\eta_2=0$) as long as $\eta_1+\eta_2>0$ just because, in such a situation, Eq. (13) [or Eq. (14)] must be satisfied,

$$F_{ij} \approx \eta_2 \widetilde{d_j^2} = \widetilde{d_i^1} - d_i^1, \tag{13}$$

$$F_{ij} \approx \eta_1 \widetilde{d_i^1} = \widetilde{d_j^2} - d_j^2. \tag{14}$$

In this experiment, the two networks G_1 and G_2 are both created by the BA model with $N=100$ nodes. In order to reveal the different matching results caused by adopting the different selection rules LDP1 and LDP2, the interaction parameters η_1 and η_2 are set to be different, i.e., $\eta_1=0.9$ and $\eta_2=0.1$. Then, the relationships between the matching precision ϕ and the sample ratio γ for various selection rules and different values of $m=2,3,4,5$ are shown in Figs. 8(a)–8(d). It is clearly shown that, in most cases, as is expected, the best matching results can be derived by adopting LDP1, while both the degree based rules LDP1 and LDP2 are prior to the random selection rule R. That is, if possible, revealing those small numbers of hub nodes in one target network as well as their matched nodes in the other target network beforehand as the *a priori* knowledge will indeed increase the matching precision of the node matching algorithm between scale-free networks.

IV. CONCLUSIONS

Real-world complex networks are considered to be interacted with each other when an individual appears in these networks with different identities. It is very interesting for

network researchers to distinguish these identities of the same individual through their topological information, which is especially significant in sociology, linguistics, and biology. However, such node matching problem between different complex networks receives relatively little attention in the literature although matching itself is well defined in graph theory and was carefully studied a hundred years ago.

In this paper, we designed a method to calculate the similarities between nodes of different networks through their connections to several pairs of preliminarily revealed matched nodes and transferred the node matching problem between two different networks to a maximum weighted bipartite matching problem that can be solved by many well-known matching algorithms in the graph theory. Through solving node matching problems on different types of networks by our algorithm, we found that the structure of the networks may largely influence the final matching results. For instance, it seems that hub nodes play dominant roles in the node matching algorithm, i.e., better matching results could be always derived by selecting nodes with large degrees as revealed matched nodes. Recent research on language networks shows that some topological properties did not change much when a language network was translated to another one. This finding suggests that the node matching algorithm proposed in this paper can be used to automatically translate texts provided some of their words have been translated correctly in the beginning.

However, the node matching algorithm proposed in this paper has computational complexity $O(N^3)$, which hinders its applications in real-world networks with a quite large number of nodes. Hence, it will be of much importance to design a more practical node matching algorithm with a lower computational complexity in the future.

ACKNOWLEDGMENTS

We would like to thank all the members in our research group in the Department of Control Science and Engineering, Zhejiang University at Yuquan Campus, for the valuable discussion about the ideas presented in this paper. This work has been supported by China Postdoctoral Science Foundation (Grant No. 20080441256) and the National 973 Program of China (Grant No. 2002CB312200).

[1] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[2] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[3] R. Ferrer i Cancho and R. V. Solé, Proc. R. Soc. London, Ser. B **268**, 2261 (2001).
[4] A. E. Motter, A. P. S. de Moura, Y.-C. Lai, and P. Dasgupta, Phys. Rev. E **65**, 065102(R) (2002).
[5] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási, Proc. Natl. Acad. Sci. U.S.A. **104**, 7332 (2007).
[6] K. Dasgupta, R. Singh, and B. Viswanathan, ACM International Conference Proceeding Series, EDBT'08 (ACM, Nantes, France, 2008), Vol. 261, p. 668–677.
[7] M. E. J. Newman, S. Forrest, and J. Balthrop, Phys. Rev. E **66**, 035101(R) (2002).
[8] N. F. A. Hasan and L. A. Adamic, ICWSM'07, Colorado, USA (unpublished).
[9] T. Furukawa, M. Ishizuka, Y. Matsuo, I. Ohmukai, and K. Uchiyama, *Analyzing Reading Behavior by Blog Mining* (AAAI, Vancouver, Canada, 2007).
[10] L. D. F. Costa, F. A. Rodrigues, G. Traviewso, and P. R. V. Boas, Adv. Phys. **56**, 167 (2007).
[11] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A.-L. Barabási, Science **297**, 1551 (2002).
[12] A. E. Motter, T. Nishikawa, and Y.-C. Lai, Phys. Rev. E **68**, 036105 (2003).
[13] Q. Xuan, Y. Li, and T.-J. Wu, Phys. Rev. E **73**, 036105 (2006).
[14] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, Phys. Rep. **424**, 175 (2006).
[15] L. Donetti, P. I. Hurtado, and M. A. Muñoz, Phys. Rev. Lett. **95**, 188701 (2005).
[16] L. Zhao, T. H. Cupertino, K. Park, and Y.-C. Lai, Chaos **17**, 043103 (2007).
[17] M. Kurant and P. Thiran, Phys. Rev. Lett. **96**, 138701 (2006).
[18] http://www.00100.cc/report.aspx?wid=4149
[19] http://en.wikipedia.org/wiki/Matching
[20] http://en.wikipedia.org/wiki/Bellman-Ford_algorithm
[21] D. B. West, *Introduction to Graph Theory*, 2nd ed. (China Machine Press, Beijing, 2004).
[22] http://en.wikipedia.org/wiki/Hungarian_algorithm
[23] R. Ferrer i Cancho, R. V. Solé, and R. Köhler, Phys. Rev. E **69**, 051915 (2004).
[24] D. R. Amancio, L. Antiqueira, T. A. S. Pardo, L. F. Costa, O. N. Oliveira, Jr., and M. G. V. Nunes, Int. J. Mod. Phys. C **19**, 583 (2008).
[25] http://en.wikipedia.org/wiki/Assignment_problem
[26] S. Zhu, Y. Okuno, G. Tsujimoto, and H. Mamitsuka, Syst. Biol. **1**, P16 (2007).
[27] Y. Xiao, M. Xiong, W. Wang, and H. Wang, Phys. Rev. E **77**, 066108 (2008).
[28] P. Erdős and A. Rényi, Publ. Math. (Debrecen) **6**, 290 (1959).
[29] T. Walsh, *Proceedings of the Sixteenth International Joint conference on Artificial Intelligence, IJCAI'99*, (Morgan Kaufmann, San Francisco, 1999), p.1172–1177.
[30] R. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **406**, 378 (2000).