PHYSICAL REVIEW E **94**, 052303 (2016)

Roles of mixing patterns in the network reconstruction

Qiang Guo,[1] Guang Liang,[1] Jia-Qi Fu,[1] Jing-Ti Han,[2] and Jian-Guo Liu[1,2,*]
[1] *Research Center of Complex Systems Science, University of Shanghai for Science and Technology, Shanghai 200093, People's Republic of China*
[2] *Data Science and Cloud Service Research Centre, Shanghai University of Finance and Economics, Shanghai 200433, People's Republic of China*
(Received 31 March 2016; published 4 November 2016)

Compressive sensing is an effective way to reconstruct the network structure. In this paper, we investigate the effect of the mixing patterns, measured by the assortative coefficient, on the performance of network reconstruction. First, we present a model to generate networks with different assortativity coefficients, then we reconstruct the network structure by using the compressive sensing method. The experimental results show that when the assortativity coefficient $r = 0.2$, the accuracy of the network reconstruction reaches the maximum value, which suggests that the compressive sensing is more effective for uncovering the links of social networks. Moreover, the accuracy of the network reconstruction will be higher as the network size increases.

DOI: 10.1103/PhysRevE.94.052303

I. INTRODUCTION

The compressive sensing has drawn considerable attention in the network reconstruction [1–5]. For example, the compressive sensing can reconstruct a propagation network in terms of the available time series [6], and could uncover the network topology using limited information based on evolutionary-game data, achieving the high accuracy results even with small data set information [7]. The compressive sensing method has already been applied to many research areas, such as biology networks [8], social networks [9], technology networks, and so on.

Generally speaking, networks could present distinct statistical and functional properties with different degree-mixing patterns, which is measured by the assortativity coefficient [10–13]. When a network is assortative, the nodes with close degrees tend to connect with each other. On the contrary, in the disassortative networks, the large-degree nodes tend to connect with small-degree ones [14]. Previous studies suggested that the social networks are much more likely to be assortative while the technology and biology networks tend to be disassortative. For instance, physics coauthorship network and company directors network are assortative, while the internet and neural network are disassortative [11,14]. As one widely used network reconstruction method, there is limited information about the effect of the network mixing pattern on the compressive sensing performance.

In this paper, we present a model to generate a series of networks with different assortativity coefficients [15,16]. Then we introduce the evolution game to generate the signals for compressive sensing, and use the different sizes of evolutionary-game data to reconstruct networks [17,18]. Comparing with the original networks, the successfully regenerated rate could be calculated. The experimental results show that the successfully regenerated rates of assortative networks are higher than disassortative networks, which suggests that the compressive sensing is an effective way for regenerating the social network structures. Furthermore, we investigate the effects of the network sizes on the performance of compressive sensing, and the experimental results show that the network regenerated rates will increase as the network size increases. More importantly, the high network regenerated rates also can be achieved even if there is a lot of noise in the data information.

II. METHOD

Evolutionary games model a common type of interaction in a variety of complex systems [7,11], it has been used to reconstruct the network based on compressive sensing [19–22]. The dynamical evolution of the underlying networked system and the signals in continuous time are required, as well these node-to-node interactions are governed by evolutionary-game types of dynamics. In an evolutionary game, at any time, a player has two strategies (S): cooperation (C) or defection (D), we use the $\mathbf{S}(C) = (1,0)^T$ and $\mathbf{S}(D) = (0,1)^T$ as the strategy matrices, and they will get different payoff with different strategies. The payoff matrix $\mathbf{G}$ of the prisoner dilemma (PD) game is

$$\mathbf{G} = \begin{pmatrix} I & P \\ H & Q \end{pmatrix}. \tag{1}$$

With different strategies, the player will get different payoff U, the payoff of mutual cooperators is I, and mutual defectors is P. While defectors will gain H when another one is a cooperator, and the cooperator gains payoff Q. Here we choose the values that $I = 1, H = 1.2 (1 < H < 2), P = Q = 0$ [17]. In a specific network, all nodes play the game with their neighbors at each step, and at the end of one round the payoff U_i of node i is

$$U_i = \sum_{j \in \Gamma_i} \mathbf{S}_i^T \cdot \mathbf{G} \cdot \mathbf{S}_j, \tag{2}$$

where $\mathbf{S}_i$ and $\mathbf{S}_j$ are the strategies of node i and j, respectively, and Γ_i is the neighbor set of i. Each node will obtain the payoffs after one round of game then update their strategies in the next round. In order to maximize its benefits in the next round, every node will update their strategies according to its neighbors' and its own payoffs last round. In this paper, we

*liujg004@ustc.edu.cn

use Fermi rule in the experiment to change the strategies of the nodes, which is defined as follows [23]. The node i randomly chooses a neighbor j, then i will adopt j's strategy $\mathbf{S}_j$ with the probability

$$P(\mathbf{S}_i \leftarrow \mathbf{S}_j) = \frac{1}{1+\exp[(U_i - U_j)/\kappa]}, \quad (3)$$

where κ reflects agents' selection rationalities in a game. For example, $\kappa = 0$ corresponds to absolute rationality where the probability is 0 if $U_i > U_j$ and 1 if $U_i < U_j$, and $\kappa \to \infty$ corresponds to completely random decision making. The probability thus indicates selection art as well as attitude of agents toward relative fitness in evolution.

The compressive sensing method could be implemented in the following way:

$$\mathbf{Y} = \Phi \cdot \mathbf{X}. \quad (4)$$

Based on Eq. (4), one can reconstruct the network relationship $\mathbf{X} \in R^N$ with the time series data $\mathbf{Y} \in R^M$ and $\Phi \in R^{M\times N}$, where $M \ll N$. Accurate results can be achieved by solving the following convex optimization problem:

$$\begin{aligned} &\min \|\mathbf{X}\|_1 \\ &\text{subject to } \mathbf{Y} = \Phi \cdot \mathbf{X}, \end{aligned} \quad (5)$$

where $\min \|\mathbf{X}\|_1 = \Sigma_{i=1}^N |\mathbf{X}_i|$ is the L_1 norm of vector $\mathbf{X}$. The optimization problem solved by Candès *et al.* [1,2,23,24] has been used for solving the network reconstruction problem [25,26]. Convex optimization based on L_1 norm has been used for solving network-construction problems in oscillator networks. To ensure the restricted isometry property [2], we orthogonalize each column of Φ. We can obtain an optimal solution by the primal-dual interior point method (L1EQ − PD) [27] to ensure the results are reliable. Solving the convex optimization, we can reconstruct the network structure with the PD game information. In this paper, $\mathbf{Y}$ is the payoff matrix set at different rounds:

$$\mathbf{Y}_i = [U_i(t_1), U_i(t_2), \ldots, U_i(t_m)]^T, \quad (6)$$

$$U_i(t) = \sum_j a_{ij} \cdot \mathbf{S}_i^T(t) \cdot \mathbf{G} \cdot \mathbf{S}_j(t), \quad (7)$$

where $U_i(t)$ is the payoff of node i at time t, and t is the number of rounds that node i play the game with other nodes $j(j = 1, \ldots, i-1, i+1, \ldots, n)$, and $t = 1,2, \ldots, m$, where n is the number of nodes, m is the length of time series. Here a_{ij} is the relationship between i and j, $a_{ij} = 1$ means that node i and j are connected, otherwise $a_{ij} = 0$. And Φ_i is the payoff matrix of node i which is defined as follows:

$$\Phi_i = \begin{pmatrix} F_{i1}(t_1) & F_{i,2}(t_1) & \cdots & F_{in}(t_1) \\ F_{i1}(t_2) & F_{i,2}(t_2) & \cdots & F_{in}(t_2) \\ \vdots & \vdots & \vdots & \vdots \\ F_{i1}(t_m) & F_{i,2}(t_m) & \cdots & F_{in}(t_m) \end{pmatrix}, \quad (8)$$

$$F_{ij} = \mathbf{S}_i^T(t) \cdot \mathbf{G} \cdot \mathbf{S}_j(t), \quad (9)$$

where F_{ij} can be calculated based on Eq. (9). Since nodes cannot play games with themselves, so F_{ii} equals 0 in the model. Then we use Eq. (4) to reconstruct the network with the data information:

$$\mathbf{A}_i = (a_{i1}, a_{i2}, \ldots, a_{in})^T, \quad (10)$$

where $\mathbf{A}_i$ reflect whether the node i is connect with other nodes. The sparsity of $\mathbf{A}_i$ makes the compressive sensing framework applicable, vectors $\mathbf{Y}_i, \Phi_i$, and $\mathbf{A}_i$ satisfy

$$\mathbf{Y}_i = \Phi \cdot \mathbf{A}_i. \quad (11)$$

By using the compressive sensing method, the network structure $\mathbf{A} = (\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_n)$ can be achieved one by one. Here the payoff data $\mathbf{Y}_i$ can be collected during the PD game, simultaneously, the strategies $\mathbf{S}_i(t)$ can also be recorded to calculate the matrix Φ_i. With the sparsity of $\mathbf{A}_i$, accurate results can be achieved from part of time series data information ($\eta \equiv m/n$, where η is the size of data which is used to network reconstruction, m is the number of accessible time instances in the time series).

In this paper, we present a model to generate networks with tunable assortativity coefficients. Given an initial connected network with n_0 nodes, a new node with e_0 edges is connected with the existing node i in terms of the probability p_i, which is defined as follows:

$$p_i = \frac{k_i^\beta}{\sum_j k_j^\beta}, \quad (12)$$

where k_i is the degree of the node i, β is a tunable parameter to generate different mixing patterns. In this way, we could generate networks with n nodes and e edges, regarding to different assortativity coefficients. Here we choose $n_0 = 5$ and $e_0 = 5$.

Then we implement PD game on the networks, time series of payoffs and strategies are recorded during the process. To measure the performance of the network reconstruction, we introduce the *Success Rate* (SR includes SR_{io} and SR_{ir}) to measure the accuracy of network reconstruction, which are defined as follows:

$$\begin{aligned} \text{SR}_{io} &= \frac{1}{n}\sum_{i=1}^n \left(\frac{\mathbf{X}_{ir} \bigcap \mathbf{X}_{io}}{\mathbf{X}_{io}}\right), \\ \text{SR}_{ir} &= \frac{1}{n}\sum_{i=1}^n \left(\frac{\mathbf{X}_{ir} \bigcap \mathbf{X}_{io}}{\mathbf{X}_{ir}}\right), \end{aligned} \quad (13)$$

where $\mathbf{X}_{io}$ are the neighbors that node i has in the test set which is constructed by the network model introduced above, $\mathbf{X}_{ir}$ are the neighbors that node i has in the training set which is reconstructed by compressive sensing.

III. EXPERIMENTAL RESULTS

We generate a group of networks with 100 nodes and 480 links with different assortative coefficient. As shown in Fig. 1(a) represents the accuracy of the network construction, measured by success rate SR. When the assortativity coefficients range from −0.4 to 0.2, as the assortative coefficient r increases, the value of SR would increase correspondingly. Especially SR reaches the maximum value when the assortativity coefficient $r = 0.2$, then it decreases. This tendency can be found with different η regarding to different success rate measured by SR_{io} or SR_{ir}. For example, Fig. 1(a) shows that

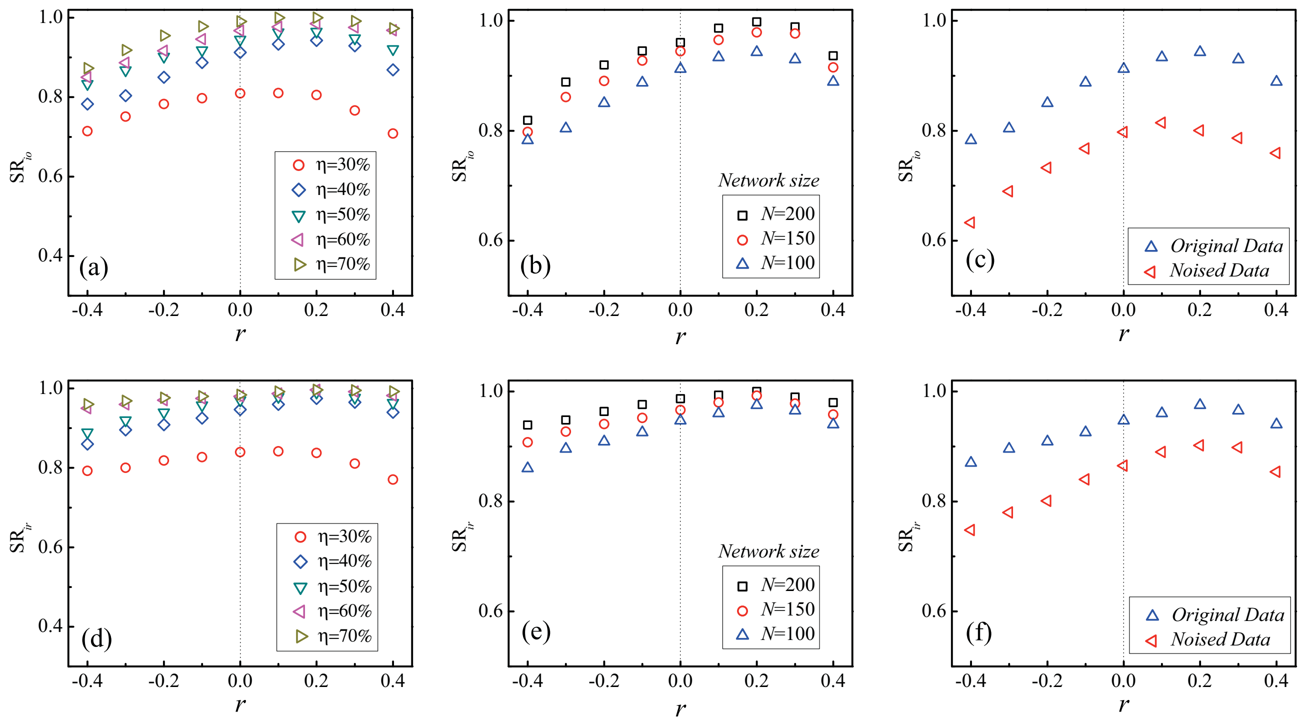

FIG. 1. Success rates (SR_{io} and SR_{ir}) of networks with different assortativity coefficients. (a) and (d) are the correlation between the assortative coefficient r and the accuracy SR, where the network size is 100, η is the size of data information which used to reconstruct the network, and $\kappa = 0.1$. From which one can find that the accuracy SR increases as the assortative coefficient r increases and reach the maximum value when $r = 0.2$, then SR decreases. This tendency can be found no matter which the η is. Each data point achieved by averaging over ten independent runs for each network. (b) and (e) are the performance of the success rata SR for different network sizes, where the network sizes, respectively, are 100, 150, 200, and the $\eta = 40\%$. As the network size increases, SR tends to be larger. The SR of assortative network is higher than the ones in disassortative networks when the size of network remains unchanged, this tendency also can be found when the network size increase. (c) and (f) test the SR of network reconstruction for the PD on different degree-mixing networks without measurement noise (original data), and with the Gaussian noise $\varepsilon \backsim [\mathcal{N}(0,0.1^2)]$ (noised data). The network size is 100, and $\eta = 40\%$.

when the $\eta = 40\%$, SR_{io} tends to be higher as the assortativity coefficient increases, and when $r = 0.2$, success rate SR_{io} can reach 0.95. In addition, when $\eta = 70\%, 60\%, 50\%$, a nearly perfect success rate SR_{io} can be achieved as the assortativity coefficient r closes to 0.2. In the same condition, more accurate results can be found in Fig. 1(d) when success rates are measured by SR_{ir}. The experiment results of Fig. 1(a) and 1(d) indicate that the performance of compressive sensing for assortative networks is better than the ones of disassortative ones.

In addition, we investigate the performance of compressive sensing for different network size. As shown in Fig. 1(b) and 1(e), we can find that the SR of the assortative network is higher than the ones of disassortative networks, which indicates that the success rate has the same tendency even though the network size increases. On the other hand, the network size can influence the network reconstruction. The experiment results indicate that as the network size increases, more accuracy results can be achieved. For instance, in Fig. 1(b), with the assortative network $r = 0.2$, success rate SR_{io} increases from 0.95 to 1 as the network size increase from 100 to 200. Furthermore, we add Gaussian noises $[\mathcal{N}(0,0.1^2)]$ in the original data ($\mathbf{Y}' = \mathbf{Y} + \varepsilon$, where $\mathbf{Y}'$ is the noise measurement) to measure the stability of compressive sensing on network reconstruction. As shown in Fig. 1(c) and 1(f), we randomly extract 40% data containing the noises to reconstruct the network. The experimental results show that for the noised data, the compressive sensing could also identify the missing links accurately.

Moreover, we test the results for four empirical networks: dolphins network, David Copperfield network, jazz musicians network, and residence hall network. Dolphins network is a network of bottlenose dolphins, the nodes are the bottlenose dolphins, an edge indicates a frequent association. David Copperfield network is the undirected network of common noun and adjective adjacencies from the novel "David Copperfield" by English 19th century writer Charles Dickens. A node represents either a noun or an adjective. An edge connects two words that occur in adjacent positions. Jazz musicians network is the collaboration network between jazz musicians. Each node is a jazz musician and an edge denotes that two musicians have played together in a band. Residence hall network contains friendship ratings between residents living at a residence hall. Then we use the compressive sensing method to reconstruct these networks with the PD model.

As shown in Table I, we reconstruct the network structure by compressive sensing method with 40% data information generated by evolutionary game model. The results of

TABLE I. Basic statistical properties and the empirical results of four networks are presented in the table. N is the number of nodes, E denotes the number of links, r is the assortativity coefficient of networks, and the success rates (SR_{io} and SR_{ir}) represent the performance of network reconstruction.

Network	N	E	r	SR_{io}	SR_{ir}
Dolphins	62	159	−0.0436	0.7565	0.7846
David Copperfield	112	425	−0.1293	0.8555	0.8624
Jazz musicians	198	2742	0.0202	0.8686	0.8939
Residence hall	217	2672	0.0960	0.9809	0.9936

compressive sensing on network reconstruction indicate that as the assortativity coefficient increases from −0.1293 to 0.09596, more accurate results can be achieved from 0.8555 to 0.9809. In addition, the SR (SR_{io} and SR_{ir}) of the dolphins network and David Copperfield network indicates that the network size also can improve the performance of compressive sensing on network reconstruction from 0.7565 to 0.8555. These empirical results can verify the characteristic of compressive sensing on network reconstruction.

IV. CONCLUSION AND DISCUSSIONS

To conclusion, in this paper we investigated the effect of the mix patterns, measured by the assortative coefficient, on the network reconstruction based on compressive sensing. First, we proposed a model to generate a series of networks with tunable assortativity coefficients. Then we reconstructed the networks from disassortative networks to assortative networks with part of data information, which is randomly extracted from the PD game. The experimental results indicate that with the same η, the success rates tend to be higher when the networks become assortative. Specifically speaking, when the assortativity coefficient $r = 0.2$, 95% of links could be uncovered by using the compressive sensing method, which suggests that compressive sensing is an effective way for uncovering the missing links of social networks. Furthermore, we investigated the effect of the assortative coefficient r on compressive sensing in different network sizes. The experimental results show that the SR of the assortative network is higher than the ones on the disassortative network even though the network size increases. We also found that with the increasing of network sizes, SR can be more accurate. The reason may lie in the fact that, based on our network model, as the network size increases, the networks become more sparse, which made the compressive sensing more efficient [9]. Finally, four empirical network reconstructions verified our conclusion with the compressive sensing method.

In summary, degree-mixing patterns have enormous impact on network reconstruction based on compressive sensing which is of great research importance on practical application. However, in this paper we only investigate the effect of degree-mixing patterns and network size on compressive sensing, while the sparsity of network and information data are also important elements in network reconstruction. The network configuration can also influence the performance of compressive sensing in network reconstruction which is beyond the current scope of this paper. And the training set size is also one of the key factors that affect the network reconstruction. So the useful method mentioned above which uncovers the topology of the network is significant to building up a framework for understanding the compressive sensing on network reconstruction.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (Nos. 71371125, 61374177, and 71271126), the Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning, the Shuguang Program Project of Shanghai Educational Committee (Grant No. 14SG42).

[1] E. Candès and M. Wakin, IEEE Signal Process. Mag. **25**, 21 (2008).
[2] E. Candès, J. Romberg, and T. Tao, IEEE Trans. Inf. Theory **52**, 489 (2006).
[3] S. G. Shandilya and M. Timme, New J. Phys. **13**, 013004 (2011).
[4] M. Timme, Phys. Rev. Lett. **98**, 224101 (2007).
[5] D. Napoletani and T. D. Sauer, Phys. Rev. E **77**, 026103 (2008).
[6] Z.-S. Shen, W.-X. Wang, Y. Fan, Z.-R. Di, and Y.-C. Lai, Nat. Commun. **5**, 4323 (2014).
[7] W.-X. Wang, Y.-C. Lai, C. Grebogi, and J.-P. Ye, Phys. Rev. X **1**, 021021 (2011).
[8] C. Y. Hwan, J. W. Gray, and C. J. Tomlin, BMC. Bio. **15**, 400 (2014).
[9] X. Han, Z.-S. Shen, W.-X. Wang, and Z.-R. Di, Phys. Rev. Lett. **114**, 028701 (2015).
[10] M. Chavez, D. U. Hwang, J. Martinerie, and S. Boccaletti, Phys. Rev. E **74**, 066107 (2006).
[11] Z.-H. Rong, X. Li, and X.-F. Wang, Phys. Rev. E **76**, 027101 (2007).
[12] R. Xulvi-Brunet and I. M. Sokolov, Phys. Rev. E **70**, 066102 (2004).
[13] M.-H. Li, S.-G. Guan, C.-S. Wu, X.-F. Gong, K. Li, J.-S. Wu, Z.-R. Di, and C.-H. Lai, Sci. Rep. **4**, 4861 (2014).
[14] M. E. J. Newman, Phys. Rev. Lett. **89**, 208701 (2002).
[15] Q. Guo, T. Zhou, J.-G. Liu, W.-J. Bai, B.-H. Wang, and M. Zhao, Phys. A **371**, 814 (2006).
[16] A. L. Barabási and R. Albert, Science **286**, 509 (1999).
[17] M. A. Nowak and R. M. May, Nature **359**, 826 (1992).
[18] C. Hauert and M. Doebeli, Nature **428**, 643 (2004).
[19] E. Candès, *Proceedings of the International Congress of Mathematicians, Madrid, Spain*, 2006.
[20] J. Romberg, IEEE Signal Process. Mag. **25**, 14 (2008).
[21] R. G. Baraniuk, IEEE Signal Process. Mag. **24**, 118 (2007).
[22] M. Timme and J. Casadiego, J. Phys. A **47**, 343001 (2014).
[23] G. Szabó and C. Tõke, Phys. Rev. E **58**, 69 (1998).
[24] E. Candès, J. Romberg, and T. Tao, Pure Appl. Math. **59**, 1207 (2006).
[25] D. Donoho, IEEE Trans. Inf. Theory **52**, 1289 (2006).
[26] W.-X. Wang, R. Yang, Y.-C. Lai, V. Kovanis, and C. Grebogi, Phys. Rev. Lett. **106**, 154101 (2011).
[27] S. Mehrotra, SIAM J. Optim. **2**, 575 (1992).