Network Reconstruction Based on Evolutionary-Game Data via Compressive Sensing

Wen-Xu Wang,1,2 Ying-Cheng Lai,2,3 Celso Grebogi,3 and Jieping Ye4

1Department of Systems Science, School of Management and Center for Complexity Research, Beijing Normal University, Beijing 100875, China
2School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, Arizona 85287, USA
3Institute for Complex Systems and Mathematical Biology, King’s College, University of Aberdeen, Aberdeen AB24 3UE, UK
4School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, Tempe, Arizona 85287, USA

(Received 31 August 2011; published 21 December 2011)

Evolutionary games model a common type of interactions in a variety of complex, networked, natural systems and social systems. Given such a system, uncovering the interacting structure of the underlying network is key to understanding its collective dynamics. Based on compressive sensing, we develop an efficient approach to reconstructing complex networks under game-based interactions from small amounts of data. The method is validated by using a variety of model networks and by conducting an actual experiment to reconstruct a social network. While most existing methods in this area assume oscillator networks that generate continuous-time data, our work successfully demonstrates that the extremely challenging problem of reverse engineering of complex networks can also be addressed even when the underlying dynamical processes are governed by realistic, evolutionary-game type of interactions in discrete time.

DOI: 10.1103/PhysRevX.1.021021 Subject Areas: Complex Systems, Nonlinear Dynamics, Statistical Physics

In many fields of science and engineering, one encounters the situation where the system of interest is composed of networked elements, called nodes, but the pattern of the node-to-node interaction or the network topology is totally unknown. It is desirable and of significant interest to uncover the network topology based on time series of certain observable quantities extracted from experiments or observations. Examples of potential applications abound: reconstruction of gene-regulatory networks based on expression data in systems biology [1–4], extraction of various functional networks in the human brain from activation data in neuroscience [5–8], and uncovering organizational networks based on discrete data or information in social science and homeland defense. In the past few years, the problem of network reconstruction has received growing attention [9–16]. Most existing works were based, however, on networks of oscillators whose dynamics are mathematically described by coupled, continuous differential equations. In particular, either some knowledge about the dynamical evolution of the underlying networked system is needed [9–11] or long, oscillatory signals in continuous time are required [12–16]. The advantage of availing oneself of continuous-time data is lost for networks in social, economic, and even biological sciences where node-to-node interactions are governed by evolutionary-game types of dynamics [17–21]. In addition to being discrete, the available data may be sporadic and the amount may be small. To our knowledge, the problem of reconstructing the full topology of a network based on discrete and “rare” data remains outstanding [22].

In this paper, we articulate a general method of addressing the problem of how to uncover network topology using evolutionary-game data based on compressive sensing, a recently developed paradigm for sparse-signal reconstruction [23–28] with broad applications ranging from image compression/reconstruction to the analysis of large-scale sensor-network data. Although convex optimization, of which compressive sensing is one type, has been used to reconstruct coupled-oscillator networks [9–11], we shall show the advantages of compressive sensing, such as its small data requirement, in solving the general inverse problem of network reconstruction based on either continuous [29,30] or discrete data. We propose a mathematical framework to convert the problem of uncovering network topology into that of sparse-signal reconstruction. In a typical game, agents use different strategies in order to gain the maximum payoff. Generally, the strategies can be divided into two types: cooperation and defection. We will show that, even when the available information about each agent’s strategy and payoff is limited, our compressive-sensing-based method can yield precise knowledge of the node-to-node interaction pattern in a highly efficient manner. We validate our method by (1) extensive numerical computations using model complex networks and evolutionary games, and (2) an actual social experiment in which participants forming a friendship network play a typical game to generate short sequences of strategy and payoff data. The high prediction accuracy achieved and the unique requirement of an extremely small data set make our method particularly suitable for potential applications...
that reveal “hidden” networks embedded in various social, economic, and biological systems.

In an evolutionary game, at any time, a player can choose one of two strategies (S): cooperation (C) or defection (D), which can be expressed as $S(C) = (1, 0)^T$ and $S(D) = (0, 1)^T$, where $T$ stands for “transpose.” The payoffs of the two players in a game are determined by their strategies and the payoff matrix of the specific game. For example, for the prisoner’s-dilemma game (PDG) [31] and the snowdrift game (SG) [32], the payoff matrices are

$$P_{PDG} = \begin{pmatrix} 1 & 0 \\ b & 0 \end{pmatrix} \quad \text{or} \quad P_{SG} = \begin{pmatrix} 1 & 1 - r \\ 0 & 0 \end{pmatrix},$$

where $b$ ($1 < b < 2$) and $r$ ($0 < r < 1$) are parameters characterizing the temptation to defect. When a defector encounters a cooperator, the defector gains payoff $b$ in the PDG and payoff $1 + r$ in the SG, but the cooperator gains the “sucker” payoff 0 in the PDG and payoff $1 - r$ in the SG. At each time step, all agents play the game with their neighbors and gain payoffs. For agent $i$, the payoff is

$$G_i = \sum_{j \in \Gamma_i} S_j^T P S_j,$$  

where $S_i$ and $S_j$ denote the strategies of agents $i$ and $j$ at the time and the sum is over the neighbor-connection set $\Gamma_i$ of $i$. After obtaining its payoff, an agent updates its strategy to maximize its payoff at the next round. Possible mathematical rules to capture an agent’s decision-making process include the best-take-over rule [31], the Fermi equation [33], and payoff-difference-determined updating probability [34]. To be concrete, we use the Fermi rule in our simulations of evolutionary-game dynamics and generate time series accordingly, which is defined as follows. After a player $i$ randomly chooses a neighbor $j$, $i$ adopts $j$’s status $S_j$ with the probability [33]

$$W(S_i \rightarrow S_j) = \frac{1}{1 + \exp((G_i - G_j)/\kappa)},$$

where $\kappa$ characterizes the stochastic uncertainties in the game dynamics. For example, $\kappa = 0$ corresponds to absolute rationality where the probability is 0 if $G_j < G_i$ and 1 if $G_i < G_j$, and $\kappa \rightarrow \infty$ corresponds to completely random decision making. The probability $W$ thus characterizes the bounded rationality of agents in society and the natural selection based on relative fitness in evolution.

The goal of compressive sensing is to reconstruct a vector $X \in \mathbb{R}^N$ from linear measurements $Y$ about $X$ in the form

$$Y = \Phi \cdot X,$$

where $Y \in \mathbb{R}^M$ and $\Phi$ is an $M \times N$ matrix. The striking feature of compressive sensing is that the number of measurements is much less than the number of components of the unknown vector, i.e., $M \ll N$. Accurate reconstruction can be achieved by solving the following convex-optimization problem:

$$\min ||X||_1 \text{ subject to } Y = \Phi \cdot X,$$

where $||X||_1 = \sum_{i=1}^N |X_i|$ is the $L_1$ norm of vector $X$. Solutions to the convex-optimization problem are available [23–28]. Convex optimization based on $L_1$ norm has been used for solving network-construction problems in oscillator networks [9–11]. Here, we shall show that the compressive-sensing approach provides a solution to network-construction problems (other than oscillator networks) based on the small amount of data from evolutionary games.

The key to solving the network-reconstruction problem lies in the relationship between the agents’ payoffs and strategies. The interactions among agents in the network can be characterized by an $N \times N$ adjacency matrix $A$ with elements $a_{ij} = 1$ if agents $i$ and $j$ are connected, and $a_{ij} = 0$ otherwise. The payoff of agent $x$ can be expressed by

$$G_x(t) = a_{1x} S_1^T(t) \cdot P \cdot S_1(t) + \cdots + a_{x,x-1} S_{x-1}^T(t) \cdot P \cdot S_{x-1}(t) + a_{x,x+1} S_{x+1}^T(t) \cdot P \cdot S_{x+1}(t) + \cdots + a_{N,x} S_N^T(t) \cdot P \cdot S_N(t),$$

where $a_{ix}$ ($i = 1, \ldots, x - 1, x + 1, \ldots, N$) represents a possible connection between agent $x$ and its neighbor $i$; $a_{ix} S_i^T(t) \cdot P \cdot S_i(t)$ ($i = 1, \ldots, x - 1, x + 1, \ldots, N$) stands for the possible payoff of agent $x$ from the game with $i$ (if there is no connection between $x$ and $i$, the payoff is zero because $a_{xx} = 0$); and $t = 1, \ldots, m$ is the number of rounds that all agents play the game with their neighbors. This relation provides us with a base to construct the vector $G_x$ in a proper compressive-sensing framework to obtain a solution of the neighbor-connection vector $A_x$ of agent $x$. In particular, we write $\Phi_x =$

$$\begin{pmatrix} F_{x1}(t_1) & \cdots & F_{xx-1}(t_1) & F_{xx+1}(t_1) & \cdots & F_{xn}(t_1) \\ F_{x2}(t_2) & \cdots & F_{xx-1}(t_2) & F_{xx+1}(t_2) & \cdots & F_{xn}(t_2) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ F_{xm}(t_m) & \cdots & F_{xx-1}(t_m) & F_{xx+1}(t_m) & \cdots & F_{xn}(t_m) \end{pmatrix},$$

$$G_x = (G_x(t_1), G_x(t_2), \ldots, G_x(t_m))^T,$$

and

$$A_x = (a_{x1}, \ldots, a_{xx-1}, a_{xx+1}, \ldots, a_{xN})^T,$$

where $F_{xy}(t_i) = S_y^T(t_i) \cdot P \cdot S_x(t_i)$. The vectors $G_x$, $A_x$, and matrix $\Phi_x$ satisfy

$$G_x = \Phi_x \cdot A_x.$$

The sparsity of $A_x$ makes the compressive-sensing framework applicable. The vector $G_x$ can be obtained directly from the payoff data. Since $S_y^T(t_i)$ and $S_x(t_i)$ in $F_{xy}(t_i)$ come from data and $P$ is known, the matrix $\Phi_x$ can be calculated from the strategy data. The vector $A_x$ can thus
be predicted based solely on the time series. Note that the self-interaction term $a_{xx}$ is not included in the vector $A$ and that the self-interaction column $[F_{xx}(t_1), \cdots, F_{xx}(t_m)]^T$ is excluded from the matrix $\Phi$. In a similar fashion, the neighbor-connection vectors of all other agents can be predicted, yielding the network adjacency matrix $A = (A_1, A_2, \cdots, A_N)$.

We first use model complex networks to demonstrate our method by implementing PDG and SG on three types of complex networks: random [35], small-world [36], and scale-free [37]. Time series of strategies and payoffs are recorded during the system’s evolution toward the steady state; they are used for uncovering the topology of the interaction network. To quantify the performance of our method in terms of the amount of required measurements for different game types and network structures, we introduce the success rates of existent links (SREL) and nonexistent links (SRNL). If the predicted value of an element of the adjacency matrix $A$ is close to 1, the corresponding link is deemed to exist. If the value is close to zero, the prediction is that there is no link. In practice, we assign a small threshold, e.g., 0.1, so that the range of existent links is $1 \pm 0.1$ and the range of nonexistent links is $0 \pm 0.1$. Any value outside the two intervals is regarded as a failure of prediction. For a single player, SREL is defined as the ratio of the number of successfully predicted neighbor-connection links to the number of actual neighbors, and SRNL is similarly defined. We then average over all nodes to obtain the values of SREL and SRNL for the entire network. The reason for treating the success rates for existent and nonexistent links separately lies in the sparsity of the underlying complex network, where the number of nonexistent links is usually much larger than the number of existent links. The choice of the threshold does not affect the values of the success rates, insofar as it is not too close to 1, nor too close to zero.

The success rates of prediction for two types of games and three types of network topologies are shown in Fig. 1. The length of the time series is represented by the number of measurements collected during the temporal evolution normalized by the number $N$ of agents; e.g., a value of 1 means that the number of used measurements equals $N$. For all combinations of game dynamics and network topologies examined, a perfect success rate can be achieved with an extremely small amount of data. For example, for random and small-world networks, the length of data required for achieving a 100% success rate is between 0.3 and 0.4. This value is slightly larger (about 0.5) for scale-free networks, due to the presence of hubs whose connections are much denser than those of most nodes, although their neighbor-connection vectors are still sparse. Figure 1 thus demonstrates that our method is both accurate and efficient. The requirement of an exceptionally small amount of data is particularly important for situations where only rare information is available. From this standpoint, evolutionary games are suitable to simulate

![Graphs showing success rates for different network types and link conditions](image)

FIG. 1. Success rates of inferring three types of networks: random, small-world and scale-free, with PDG and SG dynamics. The network size $N$ is 100. Each data point is obtained by averaging over 10 network realizations. For each realization, measurements are randomly picked from a time series of temporary evolution. The error bars denote the standard deviations. The payoff parameters for the PDG and the SG are $b = 1.2$ and $r = 0.7$, respectively. We have also systematically tested other values of $b$ and $r$ and obtained similar success rates. The average node degrees of all used networks are fixed to 6 and the noise parameter $\kappa = 0.1$.  

021021-3
such situations as meaningful data can be collected only during the transient phase before the system reaches its steady state, and game dynamics are typically fast to converge so that the transients are short. In addition, the robustness of our method has been tested in situations where the time series are contaminated by noise. For example, we have studied the case where random noise of amplitude up to 30% b (where b is the parameter characterizing temptation to defect) is added to the payoffs of PDG. When the amount of used data exceeds 0.4, the success rate approaches 100% for random networks. Similar performance has been achieved for small-world and scale-free networks. The immunity to random noise seen in our method is not surprising, as compressive sensing represents an optimization scheme that is fundamentally resilient to noise. In contrast, another type of noise, noise \( \kappa \) in the strategy-updating process, plays a positive role in network reconstruction because this kind of noise can increase the relaxation time toward one of the absorbing states (all C or all D), thereby providing more information for successful reconstruction.

To measure the efficiency of our method in reconstructing network structure with respect to different network sizes \( N \), we systematically investigate the dependence of the minimum required data for a successful reconstruction on \( N \). Without loss of generality, we define the threshold \( T_d \) (minimum amount) of data required for accurate reconstruction when the success rates of SREL and SRNL reach 99% in Fig. 1. As shown in Fig. 2, as the network size increases, a relatively less amount of data is required to precisely identify the links in the network. The reason is that, for a network with complex topology, as its size is increased, the sparsity condition can be satisfied more readily, rendering compressive sensing more efficient. We have also examined other types of networks, such as random and small-world networks. The results are qualitatively the same as that for the scale-free networks. This observation suggests that our method is particularly efficient for large and sparse networks based on rare measurable information, and this efficiency is facilitated by the characteristics of the compressive-sensing approach.

Our method can be generalized straightforwardly to weighted networks with inhomogeneous node-to-node interactions. Using weights to characterize various interaction strengths, we define the weighted adjacency matrix \( w \) as

\[
w_{ij} = \begin{cases} w_{ij} > 1, & \text{if } i \text{ connects to } j, \\ 0, & \text{otherwise.} \end{cases}
\]

In the context of evolutionary games on networks, the weight \( w_{ij} \) characterizes the situation of aggregate investment. In particular, for both players, more investments in general will lead to more payoffs. Given the link weights, the weighted payoff \( G_i^w \) of an arbitrary individual is given by

\[
G_i^w = \sum_{j \in G_i} w_{ij} S_i^T \cdot P \cdot S_j,
\]

where \( G_i \) denotes the neighbor set of \( i \). Under evolutionary-game dynamics, the weighted-network structure is taken into account by the weighted payoff \( G_i^w \). To uncover such a network from data, we need the weighted payoff vector \( G_i^w \), matrix \( \Phi_i \), and weighted neighbor-connection vector \( W_x \) for an arbitrary individual \( x \). The vectors \( G_i^w \) and \( W_x \) are given by

\[
G_i^w = (G_i^w(t_1), G_i^w(t_2), \cdots, G_i^w(t_m))^T,
\]

\[
W_x = (w_{x,1}, \cdots, w_{x,x-1}, w_{x,x+1}, \cdots, w_{x,N})^T.
\]

Similar to unweighted networks, we have

\[
G_i^w = \Phi_i \cdot W_x,
\]

where \( W_x \) can be predicted from the strategy and payoff data. The prediction accuracy can be conveniently characterized by various prediction errors, which are defined separately for link weights and nonexistent links with zero weight. In particular, the relative error of a link weight is defined as the ratio of the absolute difference between the predicted weight and the true weight to the true weight. The average error over all link weights is the prediction error \( E_w \). However, a relative error for a zero-weight (nonexistent) link cannot be defined, so we use the absolute error \( E_\alpha \). Figure 3 shows the prediction errors for PDG dynamics on a scale-free network with random link weights chosen uniformly from the interval \([1.0, 6.0]\). We observe that the prediction errors decrease fast as the number of measurements is increased. As the relative data size exceeds about 0.4, the two types of prediction errors approach essentially zero, indicating that all link weights have been successfully predicted without failure and redundancy, despite that the link weights are random. We have also examined random and specific settings.
small-world networks and observed that, to achieve the same level of accuracy, the requirement for data can be somewhat relaxed as compared with scale-free networks.

We next present an example to uncover a real social network. In the experiment, 22 participants from Arizona State University played PDG together iteratively and, in each round, each player was allowed to change his or her strategies to optimize the payoff. The payoff parameter is set (arbitrarily) to be $b = 1.2$. The player who had the highest normalized payoff (original payoff divided by the number of neighbors) summed over time was the winner and was rewarded. During the experiment, each player was allowed to communicate only with his or her direct neighbors for strategy updating. Prior to the experiment, there was a social tie (link) between two players if they had already been acquainted with each other; otherwise, there was no link. Among the 22 players, two withdrew before the experiment was completed, so they were treated as isolated nodes. The network structure is illustrated in Fig. 4(a). It exhibits typical features of social networks, such as the appearance of a large density of triangles and a core consisting of 4 players (nodes 5, 11, 13, and 16), which is fully connected within and has more links than other nodes in the network. The core essentially consists of players who were responsible for recruiting other players to participate in the experiment. Each of the 20 players who completed the experiment played 31 rounds of games, and each recorded his or her own strategy and payoff at each time, which represented the available data base for prediction. The data used for each prediction run was randomly picked from this data base. The preexisting friendship ties among the participants tend to favor cooperation and preclude the system from being trapped in the social dilemma for a small number of rounds of games. However, for a long run, a full-defection state may occur. In this sense, the recorded data were taken during the transient dynamical phase and are thus suitable for network reconstruction.

The results are shown in Fig. 4(b). We see that the social network can be successfully uncovered, despite the complicated decision-making process of each individual during the experiment. Compared to the simulation results, a larger data set (with a relative size of about 0.6) is needed for a perfect prediction of social ties. This can be attributed to the relative smaller size and denser connections in the social network than in model networks.

An interesting phenomenon is that the winner picked in terms of the normalized payoff had only two neighbors, in contrast to the players with the largest node degree, whose normalized payoffs are approximately at the average level, as shown in Fig. 4(c). In addition, the payoffs of players with smaller node degrees are highly nonuniform, while those with higher node degrees show a smaller difference. This suggests that players with high degree may not act as leaders due to their low average normalized payoffs. This experimental finding is in striking agreement with numerical predictions in the literature about the relationship between individuals’ normalized payoffs and their node degrees [38–40]. We also observe from experimental data that a typical player with a large number of neighbors failed to stimulate the neighbors to follow his or her

FIG. 3. Prediction errors $E_w$ in link weights and $E_z$ in nonexistent links for PDG on weighted scale-free networks. The network size is 100, and the weights follow a uniform distribution, ranging from 1.0 to 6.0. Each value of the prediction error is obtained using 10 independent network realizations. Other parameters are the same as for Fig. 1.

FIG. 4. (a) Structure of the experimental social network. (b) Success rates of uncovering the network topology and (c) normalized payoff of each player as a function of node degrees. The sizes of the red nodes in (a) denote their node degrees, and the two light gray nodes (for players who did not complete the experiment) are isolated without any interactions with other nodes. The 10 independent realizations used in calculating the average success rates were randomly chosen from the database of 31 rounds of games.
strategies. This observation suggests that hubs may not be as influential as expected in social networks. However, this finding should not be interpreted as a counterexample to the leader’s role in evolutionary games [34,41], since the network based on friendship may violate the absolute selfishness assumption of players who tend to be reciprocal with each other.

In summary, we proposed a general method based on compressive sensing to uncover interaction networks based on evolutionary-game data. The method was validated for complex networks of different topologies and a real social network. For all cases considered, as the number of data points exceeds a low critical value depending on the sparsity of the underlying network, the prediction errors approach zero rapidly, without or with noise in the data. To our knowledge, no previous method can match our method in terms of the accuracy and efficiency when only a small set of discrete data is available.

Our method, besides being fully applicable to complex networks governed by evolutionary-game-type interactions, can be applied in other contexts where the dynamical processes are discrete in time and the amount of available data is small. For example, inferring gene-regulatory networks from sparse experimental data is a problem of paramount importance in systems biology. For such an application, Eq. (6) should be replaced by the Hill equation, which models generic interactions among genes. In an expansion using base functions specifically suited for gene-regulatory interactions, a compressive-sensing framework, mathematically represented by Eq. (4), may be established. The underlying reverse-engineering problem can then be solved. A challenge that must be overcome is to represent the Hill function by an appropriate mathematical expansion so that the form of compressive sensing can be met.

We thank Dr. R. Yang, Dr. H.-X. Yang, and Dr. R.-R. Liu for valuable discussions. This work was supported by AFOSR under Grant No. FA9550-10-1-0083, by NSF under Grants No. CDI-1026710 and No. BECS-1023101, and by NSFC under Grant No. 11105011.


