Universal data-based method for reconstructing complex networks with binary-state dynamics

Jingwen Li,[1] Zhesi Shen,[1] Wen-Xu Wang,[1,2,*] Celso Grebogi,[3] and Ying-Cheng Lai[3,4,5]

[1]*School of Systems Science, Beijing Normal University, Beijing 100875, China*

[2]*Business School, University of Shanghai for Science and Technology, Shanghai 200093, China*

[3]*Institute for Complex Systems and Mathematical Biology, King's College, University of Aberdeen, Aberdeen AB24 3UE, United Kingdom*

[4]*School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, Arizona 85287, USA*

[5]*Department of Physics, Arizona State University, Tempe, Arizona 85287, USA*

(Received 19 June 2016; revised manuscript received 25 September 2016; published 2 March 2017)

To understand, predict, and control complex networked systems, a prerequisite is to reconstruct the network structure from observable data. Despite recent progress in network reconstruction, binary-state dynamics that are ubiquitous in nature, technology, and society still present an outstanding challenge in this field. Here we offer a framework for reconstructing complex networks with binary-state dynamics by developing a universal data-based linearization approach that is applicable to systems with linear, nonlinear, discontinuous, or stochastic dynamics governed by monotonic functions. The linearization procedure enables us to convert the network reconstruction into a sparse signal reconstruction problem that can be resolved through convex optimization. We demonstrate generally high reconstruction accuracy for a number of complex networks associated with distinct binary-state dynamics from using binary data contaminated by noise and missing data. Our framework is completely data driven, efficient, and robust, and does not require any *a priori* knowledge about the detailed dynamical process on the network. The framework represents a general paradigm for reconstructing, understanding, and exploiting complex networked systems with binary-state dynamics.

DOI: 10.1103/PhysRevE.95.032303

I. INTRODUCTION

Complex networked systems consisting of units with binary-state dynamics are common in nature, technology, and society [1]. In such a system, each unit can be in one of the two possible states, e.g., being active or inactive in neuronal and gene regulatory networks [2], cooperation or defection in networks hosting evolutionary game dynamics [3], being susceptible or infected in epidemic spreading on social and technological networks [4], two competing opinions in social communities [5], etc. The interactions among the units are complex and a state change can be triggered either deterministically (e.g., depending on the states of their neighbors) or randomly. Indeed, deterministic and stochastic state changes can account for a variety of emergent phenomena, such as the outbreak of epidemic spreading [6], cooperation among selfish individuals [7], oscillations in biological systems [8], power blackout [9], financial crisis [10], and phase transitions in natural systems [11]. A variety of models have been introduced to gain insights into binary-state dynamics on complex networks [12], such as the voter models for competition of two opinions [13], stochastic propagation models for epidemic spreading [14], models of rumor diffusion and adoption of new technologies [15], cascading failure models [16], Ising spin models for ferromagnetic phase transition [17], and evolutionary games for cooperation and altruism [18]. A general theoretical approach to dealing with networks hosting binary state dynamics was developed recently [19] based on pair approximation and master equations, providing a good understanding of the effect of the network structure on the emergent phenomena.

In this paper, we address the inverse problem of binary-state dynamics on complex networks, i.e., the problem of reconstructing the network structure and binary dynamics from data. Deciphering the network structure from data has always been a fundamental problem in complexity science, as the structure can determine the type of collective dynamics on the network [20]. More generally, for a complex networked system, reductionism is not effective and it is necessary to reconstruct and study the system as a whole [21]. The importance of network reconstruction has been increasingly recognized and effective methodologies have been developed [22–34]. Of particular relevance to our work is spreading dynamics on complex networks, where the available data are binary: a node is either infected or healthy. In such cases, a recent work [33] demonstrated that the propagation network structure can be reconstructed and the sources of spreading can be detected by exploiting compressive sensing [35–40]. However, for binary state network dynamics, a general reconstruction framework was lacking (prior to the present work). The problem of reconstructing complex networks with binary-state dynamics is extremely challenging, for the following reasons. (i) The switching probability of a node depends on the states of its neighbors according to a variety of functions for different systems, which can be linear, nonlinear, piecewise, or stochastic. If the function that governs the switching probability is unknown, a tremendous difficulty would arise in obtaining a solution of the reconstruction problem. (ii) Structural information is often hidden in the binary states of the nodes in an unknown manner and the dimension of the solution space can be extremely high, rendering impractical (computationally prohibitive) brute-force enumeration of all possible network configurations. (iii) The presence of measurement noise, missing data, and stochastic effects in the switching probability make the reconstruction task even more challenging, calling for the development of effective

*wenxuwang@bnu.edu.cn

methods that are robust against internal and external random effects.

To meet the challenges, we develop a general and robust framework for reconstructing complex networks based solely on the binary states of the nodes without any knowledge about the switching functions. Our idea is centered around developing a general method to linearize the switching functions from binary data. The data-based linearization method is applicable to linear, nonlinear, piecewise, or stochastic switching functions. The method allows us to convert the network reconstruction problem into a sparse signal reconstruction problem for local structures associated with each node. Exploiting the natural sparsity of complex networks, we employ the lasso [41], an L_1 constrained fitting method for statistics and data mining, to identify the neighbors of each node in the network from sparse binary data contaminated by noise. We establish the underlying mechanism that justifies the linearization procedure by conducting tests using a number of linear, nonlinear, and piecewise binary-state dynamics on a large number of model and real complex networks. We find universally high reconstruction accuracy even for small data amount with noise. Because of its high accuracy, efficiency and robustness against noise and missing data, our framework is promising as a general solution to the inverse problem of network reconstruction from binary-state time series, which is key to articulating effective strategies to control complex networks with binary state dynamics using, e.g., the recently developed network controllability frameworks [42–47]. The data-based linearization method is also useful for dealing with general nonlinear systems with a wide range of applications.

II. BINARY-STATE DYNAMICS

We consider a large number of representative binary-state processes on complex networks, which model a plethora of physical, social, and biological phenomena [19]. In such a dynamical process, the state of a node can be 0 (inactive) or 1 (active). In general, the process can be characterized by two switching functions, $F(m,k)$ and $R(m,k)$, which determine the probabilities for a node to change its state from 0 to 1 and vice versa, respectively. The variables in these functions, k and m, are the degree of the node and the number of active neighbors of the node, respectively. The switching functions can be linear, nonlinear, piecewise, bounded, and stochastic for characterizing and generating all kinds of binary-sate dynamical processes occurring on complex networks. Despite the difference among the switching functions, the feature that a node's switching probability depends on its degree and its number of active neighbors is generic. Table I lists the switching functions of different models, and the brief descriptions of each model can be found in the Appendix.

III. RECONSTRUCTION METHOD

Our goal is to articulate a general framework to reconstruct the network structure from binary states of nodes without knowing *a priori* the specific switching functions. A key step is to develop a universal procedure to obtain the linearization of the switching functions from binary data. We demonstrate that this can be accomplished by taking advantage of certain common features of the binary-state dynamics.

TABLE I. Switching functions for various binary-state dynamical processes on complex networks. The function $F(m,k)$ is the probability that a node switches its state from 0 to 1 while $R(m,k)$ represents the probability of the reverse process, where k is the degree of the node, m is the number of neighbors of this node in the active state 1. The models and the other parameters are described under Methods. The parameter values used in the simulations are listed in Supplemental Material Table S1 [48], Sec. 1.

Model	$F(m,k)$	$R(m,k)$
Voter [13]	$\frac{m}{k}$	$\frac{k-m}{k}$
Kirman [49]	$c_1 + dm$	$c_2 + d(k-m)$
Ising Glauber [17,50]	$\frac{1}{1+e^{\frac{\beta}{k}(k-2m)}}$	$\frac{e^{\frac{\beta}{k}(k-2m)}}{1+e^{\frac{\beta}{k}(k-2m)}}$
SIS [14]	$1-(1-\lambda)^m$	μ
Game [3]	$\frac{1}{\alpha+e^{\frac{\beta}{k}[(a-c)(k-m)+(b-d)m]}}$	$\frac{1}{\alpha+e^{\frac{\beta}{k}[(c-a)(k-m)+(d-b)m]}}$
Language [51]	$s(\frac{m}{k})^\alpha$	$(1-s)(\frac{k-m}{k})^\alpha$
Threshold [52]	$\begin{cases}0 & \text{if } m \leqslant M_k \\ 1 & \text{if } m > M_k\end{cases}$	0
Majority vote [53]	$\begin{cases}Q & \text{if } m < k/2 \\ 1/2 & \text{if } m = k/2 \\ 1-Q & \text{if } m > k/2\end{cases}$	$\begin{cases}1-Q & \text{if } m < k/2 \\ 1/2 & \text{if } m = k/2 \\ Q & \text{if } m > k/2\end{cases}$

A. Data-based linearization of switching functions

To proceed, we note that the number of active neighbors at time t can be expressed as

$$m_i(t) = \sum_{j=1, j\neq i}^{N} a_{ij} s_j(t), \tag{1}$$

where $a_{ij} = 1$ if nodes i and j are connected and $a_{ij} = 0$ otherwise, and $s_j(t)$ denotes the state of node j at time step t. In general, the switching probability $P_i^{01}(t)$ for node i to change its state from 0 to 1 at time step t can be written as

$$P_i^{01}(t) = F(m_i(t),k_i) = F\left(\sum_{j=1, j\neq i}^{N} a_{ij} s_j(t), k_i\right), \tag{2}$$

where F is a monotonic function characterizing different dynamical models, e.g., those listed in Table I. In Eq. (2), all the matrix elements a_{ij} $(i,j = 1,\ldots,N)$ that are to be inferred from data characterize the network structure. In general this is a difficult problem, because in Eq. (2), only nodal state $s_j(t)$ is measurable, whereas neither of the quantities k_i and $P_i^{01}(t)$ nor the form of F is known. In fact, not knowing the function F is the main difficulty in reconstructing the adjacency matrix $\{a_{ij}\}$. To overcome this difficulty, we propose a merging process to linearize F, i.e.,

$$F \sim c_i \sum_{j=1, j\neq i}^{N} a_{ij} s_j(t) + d_i, \tag{3}$$

where c_i and d_i are constants associated with node i. Insofar as the linearization is realized, we can solve a_{ij}. The idea of linearization is first proposed and used in Ref. [33], but the mathematical form of F is assumed to be known in that case. It is worth noting that the linearization approach is highly nontrivial and is fundamentally different from that in the

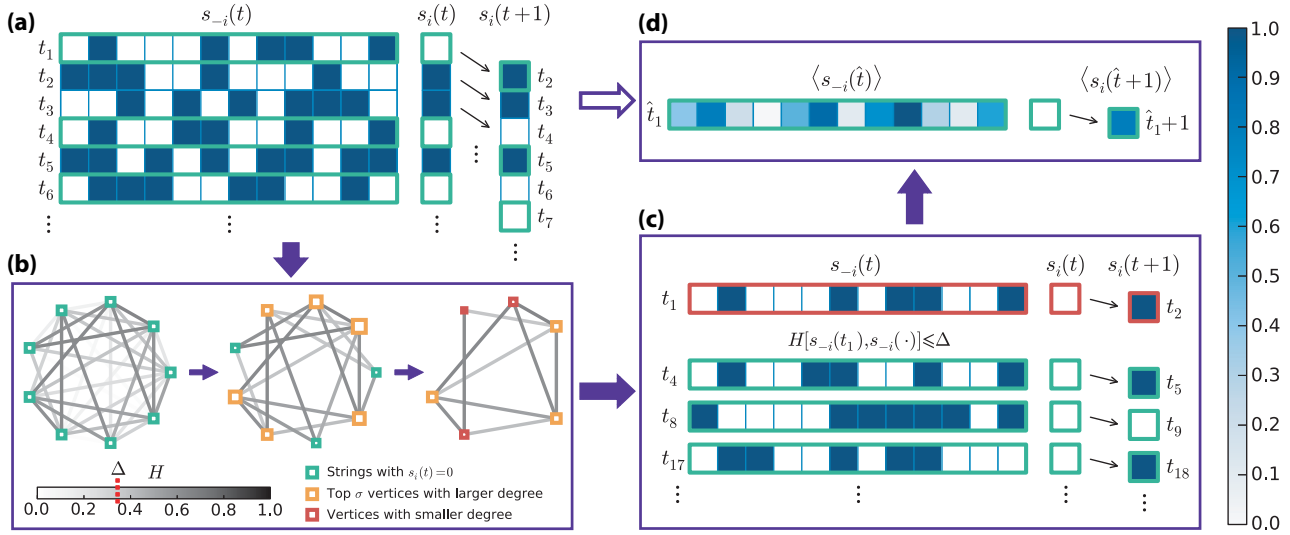


FIG. 1. Schematic illustration of data-based linearization from a merging process. (a) The original binary-state time series, where the dark blue squares denote the 1 state and the white squares denote the 0 state. The variable $s_{-i}(t)$ consists of $s_j(t)$ for all $j \neq i$. Only strings with $s_i(t) = 0$ (highlighted by the green frames) contain useful information for reconstruction. We identify the time steps with $s_i(t) = 0$ and use $s_i(t + 1)$. (b) Method of choosing bases. We first construct a network where the vertices denote strings of $s_{-i}(t)$ with $s_i(t) = 0$ (green squares) and the edges are weighted by the normalized Hamming distance H between the strings. We then eliminate the edges with weight smaller than the threshold Δ. Setting another threshold σ, we obtain the top σ percentage vertices with large degrees (yellow squares) and remove the other vertices together with their edges. Finally, we pick out the vertices with smaller degree (red squares) according to the number of base strings needed for reconstruction. (c) Selection of subordinate strings from a base. We take t_1 as a base $\hat{t}_1$ and calculate H between $s_{-i}(t_1)$ and other strings $s_{-i}(t)$ so as to sort out the time steps satisfying $H[s_{-i}(t_1),s_{-i}(\cdot)] < \Delta$ in this set. (d) Establishing average node states. We calculate the average value $\langle s_{-i}(\hat{t})\rangle$ to represent the state of the data set subject to the base, and the average value $\langle s_i(\hat{t} + 1)\rangle$ to linearize the switching probability $P_i^{01}(t)$ [Eqs. (4)]. The average values are shown in blue. Similarly, we obtain a sequence of $\hat{t}_M$ and the associated average values for reconstructing network structure by employing the lasso to solve $\mathbf{Y}_i = \Phi_i \times \mathbf{X}_i$ (see Methods for details).

standard canonical nonlinear analysis because, in our case, the mathematical form of F is not available, which can be a nonlinear, discrete, and piecewise function.

B. Procedure of dealing with binary-state data

We present the procedure of dealing with binary-state data. The merging-based linearization process enables the probability $P_i^{01}(t)$ to be estimated according to the law of large numbers, from which the solution of a_{ij} can be obtained. In particular, as shown in Fig. 1(a), for an arbitrary node i, we first identify all the time steps with $s_i(t) = 0$ as information about the switching probability $P_i^{01}(t)$ is contained only in the flipping behavior from state 0. To accurately estimate the value of P_i^{01}, we need to collect sufficient time strings, which has the same state as each other. However, we almost cannot find enough such time strings for aggregation because of the dynamical stochasticity. Thus, we relaxed the criterion to finding sufficient similar time strings. In each set of similar time strings, we first pick a base string $s_{-i}(\hat{t})$, and then collect time strings similar to it. Then, the key process comes to selecting base strings that optimize the performance of reconstruction. Figure 1(b) shows our method of selecting the optimal base strings solely based on recorded data. Specifically, we first construct a network whose vertices represent strings composed of $s_j(t)(j \neq i)$ at different time steps when $s_i(t) = 0$, and edges are weighted by the normalized pairwise Hamming distances among the strings. Then, we eliminate edges whose weight is smaller than a threshold, say Δ. Setting another threshold σ, we extract a subnetwork where only the top σ proportion of vertices with largest degree are preserved, while other vertices and their edges are removed. In this way, all remaining strings have a sufficient number of similar strings. Finally, we pick out M vertices with smallest degrees to ensure that the selected base strings are sufficiently different, where M is the number of equations in Eq. (16). The process of selecting base strings ensures us both good estimation for P_i^{01} and dissimilarity among the averaged neighborhood states. For each chosen base string, we use the threshold Δ in the normalized Hamming distance between strings to select a set of subordinate strings that belong to each base string, as shown in Fig. 1(c). A subordinate string is a string whose normalized Hamming distance to the base string is less than the selected threshold Δ. Using the average of $s_j(t)$ to represent the state of node j and the average of $s_i(t + 1)$ to estimate the switching probability $P_i^{01}(t)$ of node i according to the law of large numbers, we obtain $P_i^{01}(t) \approx \langle s_i(\hat{t} + 1)\rangle$, where $\hat{t}$ denotes the time of the base string [see Fig. 1(d)].

The whole process leads to the linearization of F with the following data-based relationship:

$$\langle s_i(\hat{t} + 1)\rangle \approx c_i \sum_{j=1, j\neq i}^{N} a_{ij} \langle s_j(\hat{t})\rangle + d_i, \qquad (4)$$

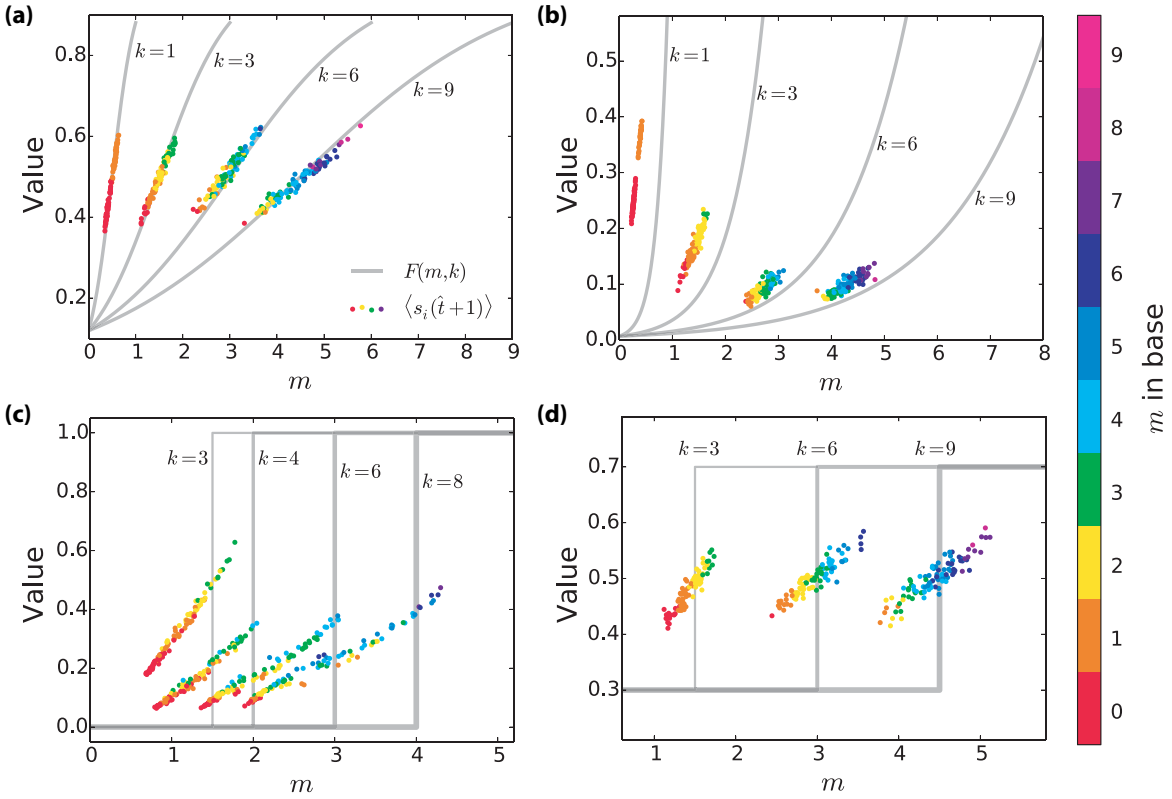


FIG. 2. Data-based linearization procedure for nonlinear and piecewise binary-state dynamics. Linearization of the switching probability function $F(m,k)$ for (a) Ising model, (b) evolutionary game model, (c) threshold model, and (d) majority model. The gray lines represent Eq. (2) with the function $F(m,k)$ from the different models, where k is the node's degree and m is the number of active neighbors. Data points are the result of the linearization procedure from time series, which corresponds to Eq. (4). For the linearized function, m is obtained from $\sum_{j=1,j\neq i}^{N} a_{ij}\langle s_j(\hat{t})\rangle$ and the value of the function is obtained via $\langle s_i(\hat{t}+1)\rangle$. For the data points, each color represents a set of subordinate strings whose base string has m active neighbors. The colors demonstrate that bases with different m values are needed to produce a linear function with a sufficient range of m for reconstruction, which justifies the base selection based on the normalized Hamming distance in Fig. 1. For both nonlinear and piecewise switching functions, a linearized function in the form of Eq. (4) can be generated based entirely on data, which is the key to reconstruction. The data points are obtained from an ER random network of $N = 100$ nodes and average degree $\langle k\rangle = 6$.

where $\langle\cdot\rangle$ is the average over all time t of the subordinate strings within $\hat{t}$. The constant parameter k_i is incorporated into the linear coefficient c_i and the intercept d_i. It is not necessary to estimate the quantities c_i, a_{ij}, and d_i in Eq. (4) separately—it is only necessary to infer value of the product $c_i \times a_{ij}$. In particular, if i and j are not connected, we have $c_i \times a_{ij} = 0$, but a nonzero value of $c_i \times a_{ij}$ means that there is a link between the two nodes. As we will show, the value of d_i can be obtained but this quantity plays little role in the reconstruction.

Figure 2 shows some representative examples to validate the linearization procedure. Four types of dynamics, including two with continuous and nonlinear switching functions and two with discontinuous and piecewise functions, are tested. We see that the switching functions F for different parameter values are linearized, enabling the network structure in the linearized system Eq. (4) to be reconstructed by distinguishing between zero and nonzero values of the reconstructed product $c_i \times a_{ij}$. As compared to the original function F, the range of m in the linearized function typically shrinks considerably as a result of the merging process, as shown in Figs. 2(a) and 2(b). For the discrete piecewise functions in Figs. 2(c) and 2(d), approximately linear functions arise for different parameter values. This is particularly striking, because even given a switching function, it is still difficult to linearize a piecewise function. We have achieve a data-based linearization of nonlinear and piecewise functions without any knowledge *a priori*.

C. Theoretical validation of data-based linearization

We provide an analysis for the completely data-based linearization that gives rise to the general relationship [Eq. (4)] from general binary-state dynamics characterized by the switching probability [Eq. (2)],

For nodes with only one neighbor, the linear relationship Eq. (4) can be rigorously proved. In this scenario, the number of active neighbors is either 0 or 1. Let $P_{\hat{t}}(1)$ denote the proportion of strings with single active neighbors in the set of base $\hat{t}$, and denote the proportion of strings with null active neighbors as $1 - P_{\hat{t}}(1)$. Let the switching probability of null active neighbors and single active neighbors be $f(0)$ and $f(1)$.

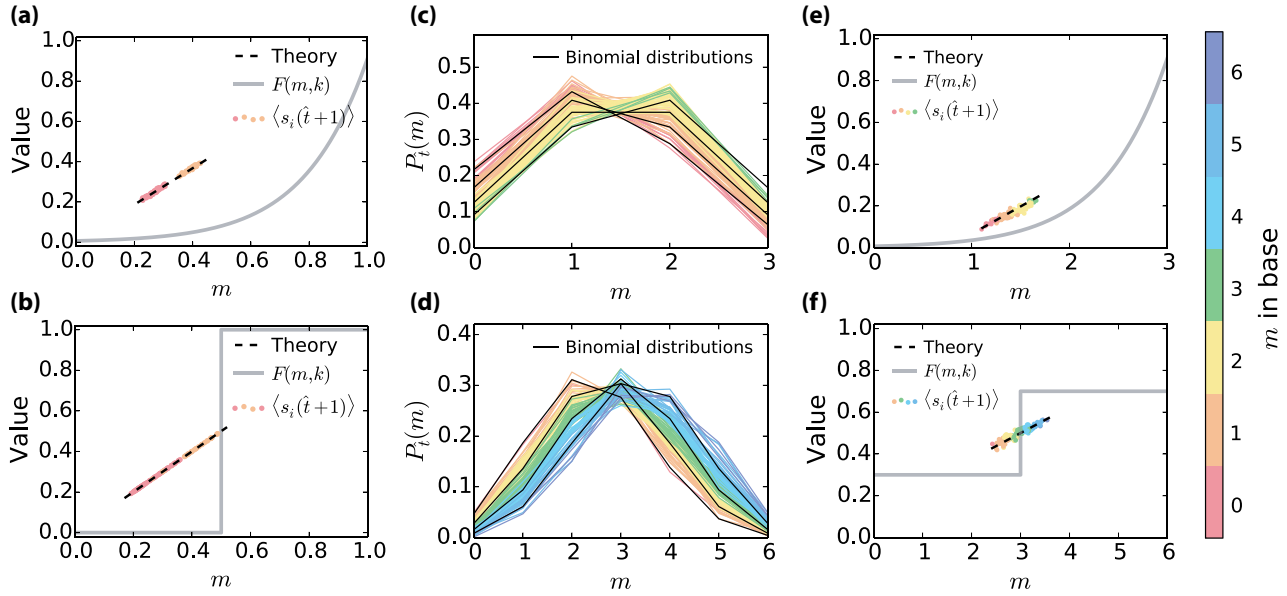


FIG. 3. Theoretical analysis of the data-based linearization. (a, b) Linearization of switching function for nodes with a single neighbor for the game model (a) and the threshold model (b). The gray solid curves are the original switching functions, data points are the results of data-based linearization [Eq. (4)], and the dashed lines are theoretical predictions from Eq. (7). The color of data points represents two sets of subordinate strings whose base string has no active neighbors ($m = 0$) or has a single active neighbor ($m = 1$). For both nonlinear and piecewise switching functions, the theoretical predictions are in exact agreement with data-based linearization, because for $k_i = 1$ the linearization is rigorous without any approximation. (c, d) The distribution of active neighbors m in subordinate strings subject to each base string and binomial distributions for reconstructing node i with $k_i = 3$ for the game model (c) and $k_i = 6$ for the majority model (d), respectively. Each color of curves represents a set of subordinate strings whose base string has m active neighbors. The distribution can be well described by binomial distributions under different success probability in each trial, as exemplified by black curves. There is a good agreement between the distribution of active neighbors in subordinate strings and binomial distributions. (e, f) The original switching function and the linearized function with theoretical prediction based on binomial distribution for the game model (e) and the majority model (f), respectively. The color of data points represents different sets of subordinate strings whose base string has different number of active neighbors m [the same meaning as in (c) and (d)]. The gray curves are the original switching function in the binary-state dynamics. The black dashed lines are the theoretical prediction of the linear relationship through Eq. (15) based on binomial distribution and Taylor linear approximation. The theoretical predictions are in good agreement with numerical results.

Then we have

$$\langle s_i(\hat{t}+1)\rangle \approx \left\langle P_i^{01}(t)\right\rangle = f(0)[1-P_{\hat{t}}(1)] + f(1)P_{\hat{t}}(1) = [f(1)-f(0)]P_{\hat{t}}(1) + f(0) \quad (5)$$

and

$$\sum_{j=1,j\neq i}^{N} a_{ij}\langle s_j(\hat{t})\rangle = P_{\hat{t}}(1). \quad (6)$$

Inserting Eq. (6) into Eq. (5), we have

$$\langle s_i(\hat{t}+1)\rangle \approx [f(1)-f(0)]\sum_{j=1,j\neq i}^{N} a_{ij}\langle s_j(\hat{t})\rangle + f(0), \quad (7)$$

which is a linear form that is subject to Eq. (4), because both $[f(1)-f(0)]$ and $f(0)$ are constants and they are determined by the specific binary-state dynamics.

Figures 3(a) and 3(b) show two representative examples of reconstructing the local structure of a node with one neighbor for the evolutionary game model and the threshold model. We see explicitly linear relationship for both models. With respect to different number of active neighbors in the original bases, two sets of groups are classified.

For nodes with more than one neighbor, the linear relationship can be justified and predicted based on binomial distribution and Taylor linear approximation. For an arbitrary node, say, node i with k neighbors, we will substantiate the linear relationship between $\langle s_i(\hat{t}+1)\rangle$ and $\sum_{j=1,j\neq i}^{N} a_{ij}\langle s_j(\hat{t})\rangle$ resulting from the data-based linearization, where

$$\langle s_i(\hat{t}+1)\rangle \approx \left\langle P_i^{01}(t)\right\rangle = \sum_{m=0}^{k_i} F(m,k_i)P_{\hat{t}}(m), \quad (8)$$

and

$$\sum_{j=1,j\neq i}^{N} a_{ij}\langle s_j(\hat{t})\rangle = \sum_{m=0}^{k_i} m P_{\hat{t}}(m), \quad (9)$$

where $P_{\hat{t}}(m)$ represents the proportion of strings with m active neighbors among all strings that belong to the set of base $\hat{t}$. The key to validating the linear relationship lies in the distribution that $P_{\hat{t}}(m)$ obeys.

Regarding the effect of the merging process as shown in Fig. 1, we hypothesize that $P_{\hat{t}}(m)$ follows binomial distributions with different success probability $p_{\hat{t}}$. We denote the proportion of state 0 in data to be p_0. If the strings are randomly

chosen for each set of a base, $P_{\hat{t}}(m)$ exactly obeys binomial distribution with success probability p_0. However, due to the process of selecting strings that are similar to each set of a base, the distribution will be biased toward the number of active neighbors in the base. Despite the original complex influence of the base and string selections based on Hamming distance, their effects can be simply regarded as selecting a group of strings with similar proportion of state 0 since we actually do not know which the node's neighbors are. This process leads to the success probability that depends on the base string. Figures 3(c) and 3(d) show the comparison between the actual distribution of $P_{\hat{t}}(m)$ obtained from numerical simulations and the binomial distributions with different success probability in each trial in the game and majority model, where the success probability in each trial approximately range from 0.4 to 0.6 because $p_0 \approx 0.5$ in the data. We see that $P_{\hat{t}}(m)$ can be well approximated by binomial distributions with different parameter values, which indeed validates our binomial distribution hypothesis.

Based on the binomial distribution hypothesis, we have

$$P_{\hat{t}}(m) = \mathrm{C}_{k_i}^{m} {p_{\hat{t}}}^{m}(1-p_{\hat{t}})^{k_i-m}. \tag{10}$$

Inserting Eq. (10) into Eq. (8) yields

$$\begin{aligned}\langle s_i(\hat{t}+1)\rangle &\approx \sum_{m=0}^{k_i} F(m,k_i)\mathrm{C}_{k_i}^{m} {p_{\hat{t}}}^{m}(1-p_{\hat{t}})^{k_i-m} \\ &= \sum_{m=0}^{k_i}\mathrm{C}_{k_i}^{m}\sum_{l=0}^{m}\left[(-1)^{m-l}\mathrm{C}_{m}^{l}F(l,k_i)\right]{p_{\hat{t}}}^{m}. \end{aligned}\tag{11}$$

The fact that $p_{\hat{t}}$ fluctuates around p_0 allows us to apply the Taylor series expansion around p_0 to Eq. (11), leading to

$$\begin{aligned}\langle s_i(\hat{t}+1)\rangle \approx{}& \sum_{m=0}^{k_i}\mathrm{C}_{k_i}^{m}\sum_{l=0}^{m}\left[(-1)^{m-l}\mathrm{C}_{m}^{l}F(l,k_i)\right]p_0^{m} \\ &+\sum_{m=0}^{k_i}\mathrm{C}_{k_i}^{m}\sum_{l=0}^{m}\left[(-1)^{m-l}\mathrm{C}_{m}^{l}F(l,k_i)\right]m p_0^{m-1} \\ &\times(p_{\hat{t}}-p_0)+\mathcal{O}(p_{\hat{t}}-p_0). \end{aligned}\tag{12}$$

Omitting the high-order term $O(p_{\hat{t}}-p_0)$, we have

$$\begin{aligned}\langle s_i(\hat{t}+1)\rangle \approx{}& \sum_{m=0}^{k_i}\mathrm{C}_{k_i}^{m}\sum_{l=0}^{m}\left[(-1)^{m-l}\mathrm{C}_{m}^{l}F(l,k_i)\right](1-m)p_0^{m} \\ &+\sum_{m=0}^{k_i}\mathrm{C}_{k_i}^{m}\sum_{l=0}^{m}\left[(-1)^{m-l}\mathrm{C}_{m}^{l}F(l,k_i)\right]m p_0^{m-1}p_{\hat{t}}. \end{aligned}\tag{13}$$

On the other hand, substitute Eq. (10) into Eq. (9) yields

$$\begin{aligned}\sum_{j=1,j\neq i}^{N} a_{ij}\langle s_j(\hat{t})\rangle &= \sum_{m=0}^{k_i} m\mathrm{C}_{k_i}^{m}{p_{\hat{t}}}^{m}(1-p_{\hat{t}})^{k_i-m}. \\ &= k_i p_{\hat{t}}. \end{aligned}\tag{14}$$

Combining Eq. (13) and Eq. (14), we have

$$\begin{aligned}\langle s_i(\hat{t}+1)\rangle \approx{}& \sum_{m=0}^{k_i}\mathrm{C}_{k_i}^{m}\sum_{l=0}^{m}\left[(-1)^{m-l}\mathrm{C}_{m}^{l}F(l,k_i)\right](1-m)p_0^{m} \\ &+\left\{\frac{1}{k_i}\sum_{m=0}^{k_i}\mathrm{C}_{k_i}^{m}\sum_{l=0}^{m}\left[(-1)^{m-l}\mathrm{C}_{m}^{l}F(l,k_i)\right]m p_0^{m-1}\right\} \\ &\times\sum_{j=1,j\neq i}^{N} a_{ij}\langle s_j(\hat{t})\rangle. \end{aligned}\tag{15}$$

Note that all variables in the first term on the right-hand side of Eq. (15) are only determined by the binary-state dynamics and the node degree of i. Hence, the first term corresponding to d_i is a constant with respect to node state s_i. In analogy, all variables in the coefficient of the second term are determined by the binary-state dynamics and the node degree of i as well, indicating the coefficient is a constant corresponding to c_i in Eq. (4). Taken together, we theoretically justified that Eq. (15) is approximately a linear equation in the form of Eq. (4).

Figures 3(e) and 3(f) show the relationship between $\langle s_i(\hat{t}+1)\rangle$ and $\sum_{j=1,j\neq i}^{N} a_{ij}\langle s_j(\hat{t})\rangle$ (namely $\langle m\rangle$) of each set of bases and the linear relationship calculated by using Eq. (15) for the game model and the majority model with nonlinear and piecewise switching dynamics. We see that the theoretical predictions are in good agreement with the results from the merging process for linearization, which strongly validates the data-based linearization for general binary-state dynamics.

It is noteworthy that the key to the success of the data-based linearization lies in selecting similar strings subject to a base and the average over each set of bases. The selection of similar strings accounts for the binomial distribution of active neighbors in a set, and different bases induces different success probability in each trial. Then the average of the binomial distributions leads to the relatively small range of $\langle m\rangle$ compared to the original range in the switching function, allowing us to use Taylor linear approximation. Moreover, high-order terms in the Taylor series expansion contribute little to the binomial distribution, which justifies the low-order approximation. Based on the linear relationship, the reconstruction of local structure can be realized by employing the lasso without requiring the linear coefficients and intercept. In other words, the data-based linearization is generally valid for arbitrary binary-state dynamics without any knowledge of the switching function.

D. Reconstruction of local structure based on the lasso

The linear relationship, Eq. (4), allows us to ascertain the neighbors of any node i from M different values of the base time, e.g., $\hat{t}_1,\ldots,\hat{t}_M$, and their subordinate times. In particular, with respect to $\hat{t}_1,\ldots,\hat{t}_M$, Eq. (4) can be expressed in the matrix form $\mathbf{Y}_i = \Phi_i \times \mathbf{X}_i$ as Eq. (16), where the vector $\mathbf{X}_i$ is to be solved for obtaining the neighbors of i, and the vector $\mathbf{Y}_i$ and the matrix Φ_i can be constructed entirely from binary time series without requiring any other

information:

$$\begin{bmatrix} \langle s_i(\hat{t}_1+1)\rangle \\ \langle s_i(\hat{t}_2+1)\rangle \\ \vdots \\ \langle s_i(\hat{t}_M+1)\rangle \end{bmatrix} = \begin{bmatrix} 1 & \langle s_1(\hat{t}_1)\rangle & \cdots & \langle s_{i-1}(\hat{t}_1)\rangle & \langle s_{i+1}(\hat{t}_1)\rangle & \cdots & \langle s_N(\hat{t}_1)\rangle \\ 1 & \langle s_1(\hat{t}_2)\rangle & \cdots & \langle s_{i-1}(\hat{t}_2)\rangle & \langle s_{i+1}(\hat{t}_2)\rangle & \cdots & \langle s_N(\hat{t}_2)\rangle \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \langle s_1(\hat{t}_M)\rangle & \cdots & \langle s_{i-1}(\hat{t}_M)\rangle & \langle s_{i+1}(\hat{t}_M)\rangle & \cdots & \langle s_N(\hat{t}_M)\rangle \end{bmatrix} \begin{bmatrix} d_i \\ c_i \cdot a_{i1} \\ \vdots \\ c_i \cdot a_{i,i-1} \\ c_i \cdot a_{i,i+1} \\ \vdots \\ c_i \cdot a_{iN} \end{bmatrix}. \tag{16}$$

The natural sparsity of complex networks ensures that, on average, the number of neighbors for a node is much smaller than the network size N, implying that $\mathbf{X}_i$ is typically sparse with most of its elements being zero and the number of nonzero elements is in fact the node degree k_i with $k_i \ll N$. We can then exploit the sparsity to reconstruct $\mathbf{X}_i$ by employing the lasso [41], a convex optimization method for sparse signal reconstruction. The lasso incorporating an L1-norm and an error control term is efficient and robust, enabling a reliable reconstruction of the local network structure as represented by $\mathbf{X}_i$ from a small amount of data. In particular, the problem is to optimize

$$\min_{\mathbf{X}_i} \left\{ \frac{1}{2M} \|\Phi_i \mathbf{X}_i - \mathbf{Y}_i\|_2^2 + \lambda \|\mathbf{X}_i\|_1 \right\}, \tag{17}$$

where $\|\mathbf{X}_i\|_1 = \sum_{j=1, j\neq i}^{N} |x_{ij}|$ is the L_1 norm of $\mathbf{X}_i$ assuring the sparsity of the solution, and the least squares term $\|\Phi_i \mathbf{X}_i - \mathbf{Y}_i\|_2^2$ guarantees the robustness of the solution against noise in data. In Eq. (17), λ is a nonnegative regularization parameter that affects the reconstruction performance in terms of the sparsity of the network, which can be determined by a cross-validation method [62]. An advantage of using the lasso is that M, i.e., the number of bases needed, can be much less than the length of $\mathbf{X}_i$. For each base of each node, the strings included can be collected and calculated from only one set of data sample in the time series, ensuring the sparse data requirement.

After the vector $\mathbf{X}_i$ has been reconstructed, the direct neighbors of node i are simply those associated with nonzero elements in $\mathbf{X}_i$. In the same manner, we can uncover the neighborhoods of all other nodes, so that the full structure of the network can be obtained by matching the neighbors of all nodes.

IV. RECONSTRUCTION PERFORMANCE

A. Measurement indices

To quantify the performance of our reconstruction method, we introduce two standard measurement indices [57], the area under the receiver operating characteristic curve (A_{UROC}) and the area under the precision-recall curve (A_{UPR}). True positive rate (TPR, R_{TP}), false positive rate (FPR, R_{FP}), precision ($\Theta_{\mathrm{precision}}$), and recall ($\Theta_{\mathrm{recall}}$), which are used to calculate A_{UROC} and A_{UPR}, are defined as follows:

$$R_{\mathrm{TP}}(l) = \frac{N_{\mathrm{TP}}(l)}{N_{\mathrm{P}}}, \tag{18}$$

where l is the cutoff in the edge list, $N_{\mathrm{TP}}(l)$ is the number of true positives in the top l predictions in the edge list, and N_{P} is the number of positives in the gold standard.

$$R_{\mathrm{FP}}(l) = \frac{N_{\mathrm{FP}}(l)}{N_{\mathrm{N}}}, \tag{19}$$

where $N_{\mathrm{FP}}(l)$ is the number of false positive in the top l predictions in the edge list, and N_{N} is the number of negatives in the gold standard.

$$\Theta_{\mathrm{precision}}(l) = \frac{N_{\mathrm{TP}}(l)}{N_{\mathrm{TP}}(l) + N_{\mathrm{FP}}(l)} = \frac{N_{\mathrm{TP}}(l)}{l}, \tag{20}$$

$$\Theta_{\mathrm{recall}}(l) = \frac{N_{\mathrm{TP}}(l)}{N_{\mathrm{P}}}, \tag{21}$$

where $\Theta_{\mathrm{recall}}(l)$, which is called sensitivity, is equivalent to $R_{\mathrm{TP}}(l)$. By varying l from 0 to N, two sequences of points $(R_{\mathrm{TP}}(l), R_{\mathrm{FP}}(l))$ and $(\Theta_{\mathrm{recall}}(l), \Theta_{\mathrm{precision}}(l))$ are measured, respectively, and the receiver operating characteristic curve and the precision-recall curve are obtained, as shown in Figs. 3(d) and 3(f). The area under the two curves, denoted as A_{UROC} and A_{UPR}, respectively, represent the reconstruction performance: $A_{\mathrm{UROC}}(A_{\mathrm{UPR}})$ ranges from $A_{\mathrm{UROC}} = 0.5(A_{\mathrm{UPR}} = N_{\mathrm{P}}/2N)$ for random guessing to $A_{\mathrm{UROC}} = 1(A_{\mathrm{UPR}} = 1)$ for perfect reconstructibility.

Because the links of each node are actually identified separately, the A_{UROC} and A_{UPR} are calculated for each node, and we use the mean index values over all the nodes to characterize the reconstruction performance for the whole network.

B. Reconstruction performance affected by network structure and amount of data

We test our method by implementing different dynamical processes on Erdös-Rényi random (ER) [54] (circle), scale-free (SF) [55] (square), small-world (SW) [56] (diamond), and empirical networks. For network reconstruction, knowledge about the switching dynamics and network details is not necessary—only the states of the nodes at different time steps need to be recorded. See Sec. 1 in the Supplemental Material [48] for computational details.

Figure 4 illustrates the reconstruction performance, where Fig. 4(a) shows the element values x_{ij} in the reconstructed neighboring vector $\mathbf{X}_i$ of all nodes for SW and SF networks with the voter model. $n_{\hat{i}}$ is defined as the number of used base strings normalized by network size N. We note that the values of x_{ij} corresponding to actual links are markedly and distinctly greater than those of null connections. Setting a cutoff value

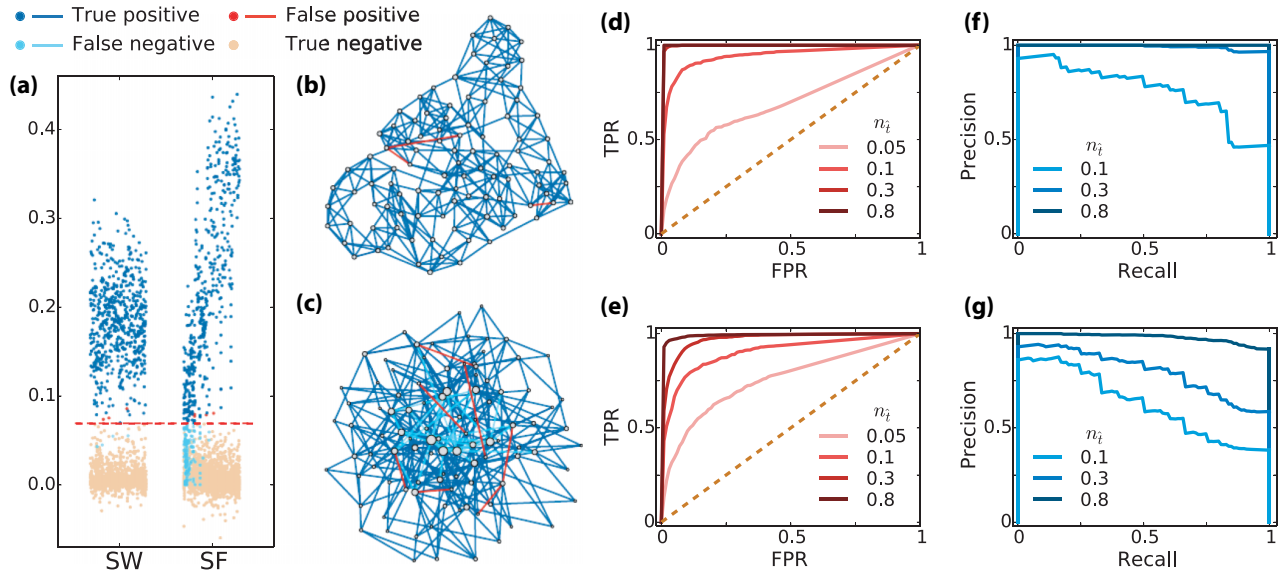


FIG. 4. Reconstruction performance. (a) Reconstructed values of the neighboring vector $\mathbf{X}_i$ for all nodes in SW and SF networks with the voter model, where $N = 100$, $\langle k \rangle = 6$, $n_{\hat{t}} = 0.8$, and the length of time series used is 1.5×10^4. The red dashed line represents the threshold for determining whether a reconstructed value is regarded as representing an actual link (a value larger than the threshold) or a null link (a value smaller than the threshold). The correctly reconstructed links (true positive), falsely reconstructed links (false positive), and missing links (false negative) are represented by the dark blue, red, and light blue points, respectively, while the yellow points indicate the true negative links. (b, c) Visualization of the reconstructed SW and SF networks, respectively. The color legends of the reconstructed links are the same as those in (a). There are more missing links (false negative) in the SF network than in the ER network. (d, e) ROC curves of reconstructed values for SW and SF networks for different values of $n_{\hat{t}}$. (f, g) PR curves of the reconstructed values for SW and SF networks for different values of $n_{\hat{t}}$.

in the gap between the two groups of points in Fig. 4(a), we can separate the actual links from the null connections, enabling a reconstruction of the whole SW network. For the SF network, it is difficult to fully reconstruct the neighbors of the hub nodes, for the following two reasons: (i) in general the linearization procedure works better for small node degree, as shown in Fig. 2; (ii) the lasso-based reconstruction requires smaller data amount and offers better accuracy for sparser vector $\mathbf{X}_i$ associated with small degree nodes. However, for an SF network, a vast majority of the nodes in an SF network are not hubs, which can be precisely reconstructed. The reconstructed SW and SF networks are shown in Figs. 4(b) and 4(c), respectively.

To assess how the number of base strings $\hat{t}$ affects the reconstruction accuracy, we define $n_{\hat{t}}$ to be the number of $\hat{t}$ divided by the network size N to quantify the relative amount of the base strings. As shown in Figs. 4(d)–4(g), the receiver operating characteristic (ROC) and the precision-recall (PR) curves show better performance as $n_{\hat{t}}$ is increased for both SW and SF networks, implying that high accuracy can be achieved for reasonably large values of $n_{\hat{t}}$. Figure 5 shows the A_{UROC} and A_{UPR}

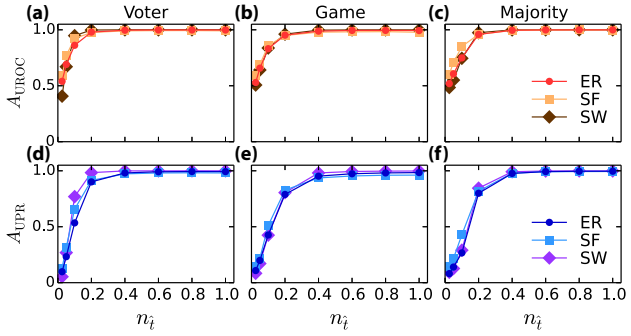


FIG. 5. Reconstruction performance with respect to the number of base strings. (a, b, c) A_{UROC} and (d, e, f) A_{UPR} as functions of the normalized number of base strings $n_{\hat{t}}$ for the voter, game and majority model on ER (circle), SF (square), and SW (diamond) networks. The network size $N = 100$ and $\langle k \rangle = 6$. The length of time series is 1.5×10^4. Other parameter values of binary-state dynamics are shown in Supplemental Material Table S1.

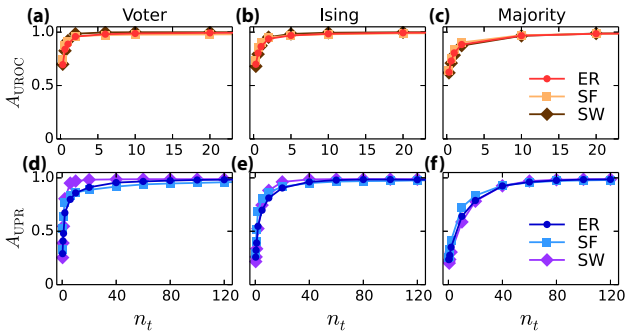


FIG. 6. Reconstruction performance with respect to the length of time series. (a–c) A_{UROC} and (d–f) A_{UPR} as functions of the relative length of time series n_t for the voter, ising and majority model on ER (circle), SF (square), and SW (diamond) networks. The network size $N = 500$ and $\langle k \rangle = 6$. Other parameter values of binary-state dynamics are shown in Supplemental Material Table S1.

TABLE II. A_{UROC} and A_{UPR} measures for various dynamics on a variety of model and empirical networks. The parameter values in the dynamical models are listed in Supplemental Material Table S1. The size and mean degree of ER (circle), SF (square), and SW (diamond) networks are $N = 500$ and $\langle k \rangle = 6$, and the length of time series used is 6×10^4. The length of time series used for empirical networks is 1.5×10^4.

$A_{\mathrm{UROC}}/A_{\mathrm{UPR}}$	Voter	Kirman	Ising	SIS	Game	Language	Threshold	Majority
ER	1.000/0.983	0.999/0.954	1.000/0.982	0.997/0.960	0.999/0.981	0.995/0.934	1.000/0.988	1.000/0.986
SF	0.992/0.959	0.985/0.920	0.998/0.976	0.984/0.924	0.988/0.951	0.986/0.925	0.986/0.985	0.999/0.980
SW	1.000/0.988	1.000/0.982	1.000/0.988	1.000/0.988	1.000/0.988	1.000/0.986	0.994/0.979	1.000/0.987
Dolphins	1.000/0.916	0.997/0.908	0.999/0.911	0.978/0.867	0.993/0.900	0.985/0.870	0.991/0.890	1.000/0.913
Football	0.999/0.884	1.000/0.898	0.999/0.899	0.999/0.884	0.996/0.882	0.992/0.859	0.918/0.637	0.999/0.896
Karate	0.997/0.856	0.969/0.838	0.981/0.836	0.954/0.823	0.984/0.839	0.960/0.803	0.971/0.810	0.996/0.847
Leader	1.000/0.838	0.991/0.912	0.991/0.823	0.968/0.789	0.990/0.818	0.966/0.780	0.970/0.760	0.998/0.832
Polbooks	0.999/0.912	0.991/0.829	0.998/0.908	0.932/0.779	0.986/0.888	0.978/0.857	0.971/0.858	0.999/0.913
Prison	1.000/0.936	0.999/0.896	1.000/0.935	0.992/0.915	0.981/0.909	0.991/0.909	0.999/0.931	1.000/0.935
Santa Fe	0.998/0.967	0.990/0.933	1.000/0.969	0.982/0.937	0.997/0.965	0.996/0.959	0.994/0.961	1.000/0.970

measures as a function of $n_{\hat{\imath}}$ for different dynamical models on ER, SW, and SF networks. Due to the advantage of the lasso for sparse vectors, nearly perfect reconstruction is achieved after $n_{\hat{\imath}}$ exceeds a relatively small critical value, e.g., 0.4.

It is also important to assess how the length of the binary time series affects the reconstruction accuracy and efficiency. We have calculated the A_{UROC} and A_{UPR} measures as a function of the relative time-series length n_t (defined as the total length of time series divided by N) for various dynamical processes on ER, SF, and SW networks. Figure 6 shows the reconstruction performance for voter, Ising, and majority models in combination with different types of networks. We find that A_{UROC} and A_{UPR} rapidly increases as n_t increases. After n_t exceeds a relatively small value, nearly full reconstruction can be achieved, which provides additional evidence for the high efficiency of our reconstruction method (see Supplemental Material [48], Sec. 2 for full results of performance for all models versus n_t). In general, high reconstruction accuracy can be achieved for relatively short time series. We systematically test our method on a variety of model and real networks in combination with eight binary-state dynamics (Table II) and find high values of A_{UROC} and A_{UPR} for all cases.

We explore the effects of network properties such as the average degree $\langle k \rangle$ and the size N on reconstruction performance. As shown in Fig. 7, the reconstruction accuracy decreases as $\langle k \rangle$ increases. The main reason for this result is that the low-order approximation in the data-based linearization is better for smaller node degree. Moreover, with the increase of $\langle k \rangle$, the vector $\mathbf{X}_i$ to be reconstructed will become denser. Note that it usually requires larger amounts of data to reconstruct a denser signal by using the lasso according to the compressive sensing theory. Thus, in general a network with larger $\langle k \rangle$ will be more difficult to be reconstructed. Figure 8 shows the minimum relative length of time series $n_t^{\min}$ to acquire at least 0.95 A_{UROC} and A_{UPR} simultaneously as a function of network size N. We see that $n_t^{\min}$ decreases as N increases, which is because of network sparsity as well. In general, for the same average node degree $\langle k \rangle$, a network with larger size will be sparser, leading to a sparser vector $\mathbf{X}_i$. According to the compressive sensing theory, less data are required for reconstructing a sparser $\mathbf{X}_i$, accounting for the decrease of $n_t^{\min}$ with the increase of N. These results indicate that our reconstruction method is scalable and of practical importance for dealing with large real networked systems.

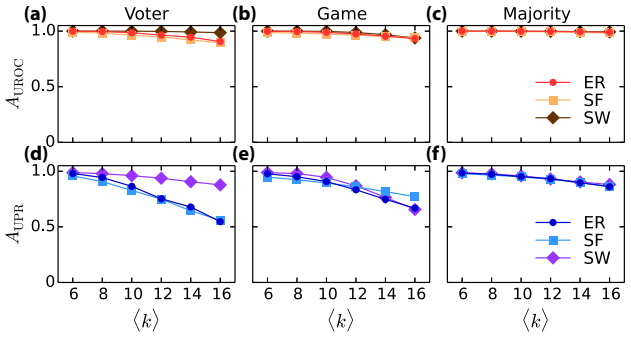


FIG. 7. Reconstruction performance affected by average node degree. (a, b, c) A_{UROC} and (d, e, f) A_{UPR} as functions of the average node degree $\langle k \rangle$ for the voter, game and majority model on ER (circle), SF (square), and SW (diamond) networks. The network size $N = 500$ and relative length of time series $n_t = 100$. Other parameter values of binary-state dynamics are shown in Supplemental Material Table S1.

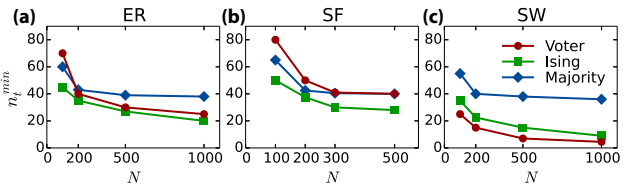


FIG. 8. Reconstruction performance affected by network size. The minimum relative length $n_t^{\min}$ to acquire at least 0.95 A_{UROC} and A_{UPR} simultaneously as a function of network size N for the voter (red circle), Ising (green square), and majority model (blue diamond) on (a) ER, (b) SF, and (c) SW networks. The mean degree of networks is 6. Other parameter values of binary-state dynamics are shown in Supplemental Material Table S1.

TABLE III. Robustness of reconstruction against noise and missing data. A_{UROC} and A_{UPR} measures for voter, game, and majority models on ER, SF, and SW networks for measurement noise $n_f = 10\%$ and the fraction of inaccessible nodes $n_m = 30\%$. The network size is $N = 500$ and the mean degree is $\langle k \rangle = 6$. The length of the time series used is 6×10^4. Details of the parameter values in the dynamical models are listed in Supplemental Material Table S1.

	$n_f = 10\%$			$n_m = 30\%$		
$A_{\mathrm{UROC}}/A_{\mathrm{UPR}}$	Voter	Game	Majority	Voter	Game	Majority
ER	0.995/0.938	0.955/0.707	0.991/0.864	1.000/0.985	0.999/0.983	1.000/0.988
SF	0.983/0.903	0.954/0.800	0.990/0.894	0.995/0.968	0.991/0.957	0.995/0.984
SW	1.000/0.984	0.976/0.741	0.994/0.874	1.000/0.988	1.000/0.988	1.000/0.988

C. Robustness of reconstruction against noise and missing data

In real applications, time series are often contaminated by noise and the data from certain nodes may be lost or inaccessible. To address these practical issues, we test the robustness of our method. Specifically, we instill noise into the time series by randomly flipping a fraction n_f of binary states and assume a fraction n_m of nodes are inaccessible. The results are shown in Table III, where voter, game, and majority models are used as examples of linear, nonlinear, and piecewise dynamics, respectively. Strikingly, we obtain high values of A_{UROC} and A_{UPR} even in the presence of 10% measurement noise or 30% inaccessible nodes, providing strong evidence for the robustness of our framework against noise and missing data. More detailed characterization associated with the results in Table III, i.e., A_{UROC} and A_{UPR} as functions of n_f and n_m, are provided in the Supplemental Material [48], Sec. 3.

V. DISCUSSION

Reconstructing the topological structure and dynamics of complex systems from data is a central issue in both network science and engineering community [22,24,25,27,28,32,58,59]. A framework [29,60,61] of network reconstruction is based on compressive sensing [35–40], a sparse signal recovery method developed in applied mathematics and engineering signal processing. A recent work [33] also demonstrated that compressive sensing can be exploited for network reconstruction in situations where the available time series are polarized (binary), e.g., virus spreading and information diffusion in social and computer networks. While the structure of the virus propagation network and the spreading sources can be obtained, the method is unable to predict the network dynamical systems that generate the binary data.

The contribution of this paper is a general framework to solve the challenging problem of reconstructing complex networks hosting binary-state dynamics, based only on time series without any knowledge of the network structure and the switching functions that generate the binary data. The key to our success is the formulation of a universal data-based linearization method, which is powerful for reconstructing the neighborhood of nodes for any type of nodal dynamics: linear, nonlinear, discontinuous, or stochastic. The natural sparsity of real complex networks allows us to address the local reconstruction as a sparse signal reconstruction problem that can be solved by employing the lasso, a convex optimization method, from small amounts of binary data. The optimization is robust against measurement noise and missing data. Once the neighborhoods of all nodes have been reconstructed, the whole network can be mapped out by assembling all the local structures and making adjustments to ensure consistency. We have validated our framework using a variety of binary-state dynamical models on a number of model and real complex networks. High reconstruction accuracy has been obtained for all cases, even for relatively small amounts of binary data contaminated by noise and when partial data are lost. These results suggest the practical applicability of our framework. In practical applications, instead of evaluating A_{UROC} and A_{UPR}, we often need to distinguish the true links from the nonlinks based on the reconstructed values of $\mathbf{X}$. To accomplish this goal, we can generate a histogram from all the elements of $\mathbf{X}$ and find a appropriate cutting threshold between the two peaks representing links and nonlinks.

While our framework potentially offers a general, completely data-driven approach to reconstructing binary dynamical processes on complex networks, there are still challenges. For example, our framework can deal with various types of switching functions underlying the binary-state dynamics, but in its present form the framework is not applicable to nonmonotonic functions or non-Markovian type of dynamics. Especially, when the switching functions are not monotonic, the data-based linearization would fail due to the violation of the one-to-one correspondence between the switching probability and the number of active neighbors. For non-Markovian dynamics, the merging procedure inherent in our method would fail. To predict the interaction strength among nodes presents another challenge, especially where noise is present and there is missing data. The results reported in this paper suggest strongly that our present framework can serve as a starting point to meet the challenges, eventually leading to a complete and universally applicable solution to the inverse problem of binary network structure and dynamics.

ACKNOWLEDGMENTS

W.-X.W. was supported by NSFC under Grant No. 61573064, No. 61074116, and No. 71631002, as well as the Fundamental Research Funds for the Central Universities, Beijing Nova Programme. Y.-C.L. was supported by ARO under Grant No. W911NF-14-1-0504. W.-X.W. designed research; J.L. and Z.S. performed research; all analyzed data; J.L., W.-X.W., and Y.-C.L. wrote the paper; all edited the paper. The authors declare no competing financial interests.

APPENDIX: DESCRIPTION OF USED BINARY-STATE DYNAMICS

The voter model [13] assumes that a node randomly chooses and then adopts one of its neighbors' state at each time step. The total number of neighbors is its degree k, of which m are active, i.e., they are in state 1. The probabilities that the node will become active and inactive are m/k and $(k-m)/k$, respectively. In the majority-voter model [53], a node tends to align with the majority state of its neighbors, and the probability of misalignment is Q.

In the Kirman's ant colony model [49], a node switches from state 0 to 1 with the probability $F_{k,m} = c_1 + dm$ (with m being the number of active neighbors) and the rate of transition from 1 to 0 is $R_{k,m} = c_2 + d(k-m)$, where the parameters c_1 and c_2 quantify the individual action that is independent of the states of the neighbors and d characterizes the action of copying from neighbors' state.

The Ising model [17] is a classic paradigm to study ferromagnetism at the microscopic level of spins. In the model, a node can assume either one of the two states: spin-up or spin-down. Switching in the state occurs with the probability determined by minimizing the energy (Hamiltonian) of the system. In our study, we chose the transition rates according to the Glauber dynamics [50], as shown in Table I, where the parameter β quantifies the combining effect of temperature and the ferromagnetic-interaction parameter.

The SIS model [14] describes the epidemic process of disease spreading with infection and recovery. Each susceptible individual contracts the disease from each of its infected neighbors at the rate λ, so at each time step a susceptible node with m infected neighbors has the probability $(1-\lambda)^m$ of remaining susceptible. The infection rate is then $1-(1-\lambda)^m$. The recovery rate of an infected node is μ at each time step.

The game model [3] originates from the evolutionary game theory. In a network, each node is a player, and the two states means that the player can take on two different strategies. A player plays with each of his/her neighbors using one chosen strategy at each time step. The profit of a rational player i, when playing with a neighbor j, is characterized by the payoff matrix $\begin{array}{c} \\ s_1 \\ s_2 \end{array}\begin{array}{c} \begin{array}{cc} s_1 & s_2 \end{array} \\ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \end{array}$ where a, b, c, and d are parameters. Different games can be generated by adjusting a, b, c, and b. The payoff of a player is the sum of profit from playing game with all its neighbors. A player switches the strategy with a probability that depends on the payoff it may gain in the next round under the current circumstance by switching its strategy, as illustrated in Table I, where the parameter α qualifies the willingness for an individual to change its strategy according to those of its neighbors, and β is associated with the effect of the expected payoff.

For the language model [51], the two states denote two different language choices of a person. Transition from the primary language to the secondary occurs with the probability that is proportional to the fraction of speakers in the neighbors with the power α, multiplied by the parameter s (or $1-s$) according to the respective language.

The threshold model [52] is a deterministic model, where for each node a certain threshold M_k is set which can be, for example, a function of the node's degree. At each time step, a node becomes active if the number m of its active neighbors exceeds the threshold M_k, and no recovery transformation is permitted.

[1] A. Barrat, M. Barthelemy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, Cambridge, 2008).

[2] A. Kumar, S. Rotter, and A. Aertsen, Spiking activity propagation in neuronal networks: Reconciling different perspectives on neural coding, Nat. Rev. Neuro. **11**, 615 (2010).

[3] G. Szabó and G. Fath, Evolutionary games on graphs, Phys. Rep. **446**, 97 (2007).

[4] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, Epidemic processes in complex networks, Rev. Mod. Phys. **87**, 925 (2015).

[5] J. Shao, S. Havlin, and H. E. Stanley, Dynamic Opinion Model and Invasion Percolation, Phys. Rev. Lett. **103**, 018701 (2009).

[6] C. Granell, S. Gómez, and A. Arenas, Dynamical Interplay Between Awareness and Epidemic Spreading in Multiplex Networks, Phys. Rev. Lett. **111**, 128701 (2013).

[7] F. C. Santos and J. M. Pacheco, Scale-Free Networks Provide a Unifying Framework for the Emergence of Cooperation, Phys. Rev. Lett. **95**, 098104 (2005).

[8] A. Koseska, E. Volkov, and J. Kurths, Oscillation quenching mechanisms: Amplitude vs. oscillation death, Phys. Rep. **531**, 173 (2013).

[9] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin, Catastrophic cascade of failures in interdependent networks, Nature **464**, 1025 (2010).

[10] M. Galbiati, D. Delpini, and S. Battiston, The power to control, Nat. Phys. **9**, 126 (2013).

[11] D. Balcan and A. Vespignani, Phase transitions in contagion processes mediated by recurrent mobility patterns, Nat. Phys. **7**, 581 (2011).

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

[13] V. Sood and S. Redner, Voter Model on Heterogeneous Graphs, Phys. Rev. Lett. **94**, 178701 (2005).

[14] R. Pastor-Satorras and A. Vespignani, Epidemic Spreading in Scale-Free Networks, Phys. Rev. Lett. **86**, 3200 (2001).

[15] C. Castellano, S. Fortunato, and V. Loreto, Statistical physics of social dynamics, Rev. Mod. Phys. **81**, 591 (2009).

[16] A. Bashan, Y. Berezin, S. V. Buldyrev, and S. Havlin, The extreme vulnerability of interdependent spatially embedded networks, Nat. Phys. **9**, 667 (2013).

[17] P. L. Krapivsky, S. Redner, and E. Ben-Naim, *A Kinetic View of Statistical Physics* (Cambridge University Press, New York, 2010).

[18] F. C. Santos, M. D. Santos, and J. M. Pacheco, Social diversity promotes the emergence of cooperation in public goods games, Nature **454**, 213 (2008).

[19] J. P. Gleeson, Binary-State Dynamics on Complex Networks: Pair Approximation and Beyond, Phys. Rev. X **3**, 021004 (2013).

[20] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, Complex networks: Structure and dynamics, Phys. Rep. **424**, 175 (2006).

[21] A.-L. Barabási, The network takeover, Nat. Phys. **8**, 14 (2011).

[22] T. S. Gardner, D. di Bernardo, D. Lorenz, and J. J. Collins, Inferring genetic networks and identifying compound mode of action via expression profiling, Science **301**, 102 (2003).

[23] N. Friedman, Inferring cellular networks using probabilistic graphical models, Science **303**, 799 (2004).

[24] M. Timme, Revealing Network Connectivity from Response Dynamics, Phys. Rev. Lett. **98**, 224101 (2007).

[25] A. Clauset, C. Moore, and M. E. J. Newman, Hierarchical structure and the prediction of missing links in networks, Nature **453**, 98 (2008).

[26] S. Guo, J. Wu, M. Ding, and J. Feng, Uncovering interactions in the frequency domain, PLoS Comput. Biol. **4**, e1000087 (2008).

[27] J. Ren, W.-X. Wang, B. Li, and Y.-C. Lai, Noise Bridges Dynamical Correlation and Topology in Coupled Oscillator Networks, Phys. Rev. Lett. **104**, 058701 (2010).

[28] S. Hempel, A. Koseska, J. Kurths, and Z. Nikoloski, Inner Composition Alignment for Inferring Directed Networks from Short Time Series, Phys. Rev. Lett. **107**, 054101 (2011).

[29] W.-X. Wang, Y.-C. Lai, C. Grebogi, and J.-P. Ye, Network Reconstruction Based on Evolutionary-Game Data via Compressive Sensing, Phys. Rev. X **1**, 021021 (2011).

[30] B. Barzel and A.-L. Barabási, Network link prediction by global silencing of indirect correlations, Nat. Biotechnol. **31**, 720 (2013).

[31] S. Feizi, D. Marbach, M. Médard, and M. Kellis, Network deconvolution as a general method to distinguish direct dependencies in networks, Nat. Biotechnol. **31**, 726 (2013).

[32] G. Caldarelli, A. Chessa, F. Pammolli, A. Gabrielli, and M. Puliga, Reconstructing a credit network, Nat. Phys. **9**, 125 (2013).

[33] Z. Shen, W.-X. Wang, Y. Fan, Z. Di, and Y.-C. Lai, Reconstructing propagation networks with natural diversity and identifying hidden source, Nat. Commun. **5**, 4323 (2014).

[34] X. Han, Z. Shen, W.-X. Wang, and Z. Di, Robust Reconstruction of Complex Networks from Sparse Data, Phys. Rev. Lett. **114**, 028701 (2015).

[35] E. J. Candès, J. K. Romberg, and T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, IEEE Trans. Inf. Theory **52**, 489 (2006).

[36] E. J. Candes, J. K. Romberg, and T. Tao, Stable signal recovery from incomplete and inaccurate measurements, Commun. Pure Appl. Math. **59**, 1207 (2006).

[37] D. L. Donoho, Compressed sensing, IEEE Trans. Inf. Theory **52**, 1289 (2006).

[38] R. G. Baraniuk, Compressive sensing, IEEE Signal Proc. Mag. **24**, 118 (2007).

[39] E. J. Candès and M. B. Wakin, An introduction to compressive sampling, IEEE Signal Proc. Mag. **25**, 21 (2008).

[40] J. Romberg, Imaging via compressive sampling, IEEE Signal Proc. Mag. **25**, 14 (2008).

[41] J. Friedman, T. Hastie, and R. Tibshirani, *The Elements of Statistical Learning* (Springer, Berlin, 2001).

[42] Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabási, Controllability of complex networks, Nature **473**, 167 (2011).

[43] T. Nepusz and T. Vicsek, Controlling edge dynamics in complex networks, Nat. Phys. **8**, 568 (2012).

[44] G. Yan, J. Ren, Y.-C. Lai, C.-H. Lai, and B. Li, Controlling Complex Networks: How Much Energy is Needed? Phys. Rev. Lett. **108**, 218703 (2012).

[45] Z. Yuan, C. Zhao, Z. Di, W.-X. Wang, and Y.-C. Lai, Exact controllability of complex networks, Nat. Commun. **4**, 2447 (2013).

[46] J. Ruths and D. Ruths, Control profiles of complex networks, Science **343**, 1373 (2014).

[47] S. Wuchty, Controllability in protein interaction networks, Proc. Natl. Acad. Sci. USA **111**, 7156 (2014).

[48] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.95.032303 for computational details, dependence of performance on data amount, and robustness against noise and missing data.

[49] A. Kirman, Ants, rationality, and recruitment, Quart. J. Econo. **108**, 137 (1993).

[50] R. J. Glauber, Time-dependent statistics of the Ising model, J. Math. Phys. **4**, 294 (1963).

[51] D. M. Abrams and S. H. Strogatz, Linguistics: Modeling the dynamics of language death, Nature **424**, 900 (2003).

[52] M. Granovetter, Threshold models of collective behavior, Ame. J. Socio. **83**, 1420 (1978).

[53] M. J. de Oliveira, Isotropic majority-vote model on a square lattice, J. Stat. Phys. **66**, 273 (1992).

[54] P. Erdös and A. Rényi, On the evolution of random graphs, Publ. Math. Inst. Hung. Acad. Sci. **5**, 17 (1960).

[55] A.-L. Barabási and R. Albert, Emergence of scaling in random networks, Science **286**, 509 (1999).

[56] D. J. Watts and S. H. Strogatz, Collective dynamics of small-world networks, Nature **393**, 440 (1998).

[57] J. Davis and M. Goadrich, The relationship between precision-recall and roc curves, in *Proceedings of the 23rd International Conference on Machine Learning* (ACM, Pittsburgh, Pennsylvania, USA, 2006), pp. 233–240.

[58] J. Bongard and H. Lipson, Automated reverse engineering of nonlinear dynamical systems, Proc. Natl. Acad. Sci. USA **104**, 9943 (2007).

[59] Z. Levnajić and A. Pikovsky, Network Reconstruction from Random Phase Resetting, Phys. Rev. Lett. **107**, 034101 (2011).

[60] W.-X. Wang, R. Yang, Y.-C. Lai, V. Kovanis, and C. Grebogi, Predicting Catastrophes in Nonlinear Dynamical Systems by Compressive Sensing, Phys. Rev. Lett. **106**, 154101 (2011).

[61] W.-X. Wang, R. Yang, Y.-C. Lai, V. Kovanis, and M. A. F. Harrison, Time-series–based prediction of complex oscillator networks via compressive sensing, Europhys. Lett. **94**, 48006 (2011).

[62] F. Pedregosa *et al.*, Scikit-learn: Machine learning in Python, J. Mach. Learn. Res. **12**, 2825 (2011).

Supplementary Materials for
A universal data based method for reconstructing complex networks with binary-state dynamics

Jingwen Li, Zhesi Shen, Wen-Xu Wang, Celso Grebogi, and Ying-Cheng Lai

1 Computation details

Parameter values in the binary-state dynamics used for network reconstruction are displayed in Supplementary Table S1. The only requirement for choosing the parameter values is that the switching dynamics should be monotonic. Since all the binary-state dynamics are monotonic, there is no specific restriction for the parameter values. Note that several models have convergent behaviors. If the states of nodes converge into a stable state, there will be no more useful information for network reconstruction. If this occurs, we randomly initialize the states of all nodes after a certain period.

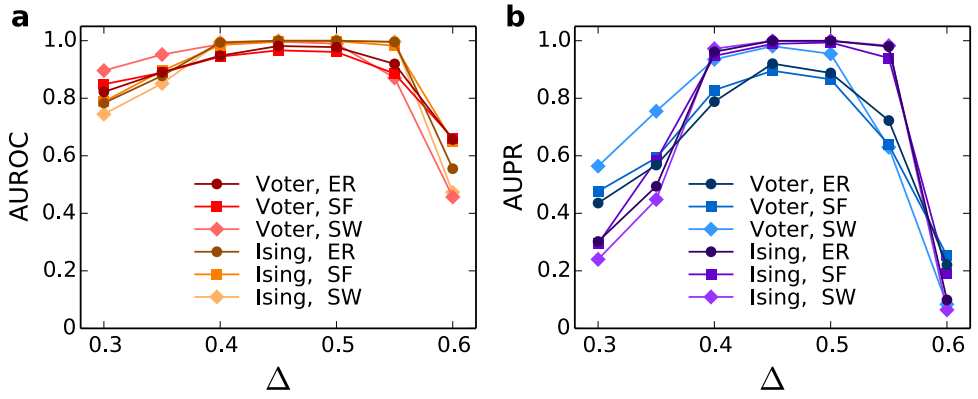
The set of the threshold parameter Δ for realizing the merging process for network reconstruction is independent of network structure and binary-state dynamics. We investigate the dependence of the reconstruction performance on threshold Δ. The results are shown in Supplementary Fig S1. We found that AUROC and AUPR can always reach high values when $0.4 \leqslant \Delta \leqslant 0.55$ in all cases. Thus, we set the threshold Δ to be 0.45 for simplicity.

Regarding the selection of bases, the method is relatively time consuming because it requires calculating the Hamming distance between each pair of strings in different time steps. Hence, to improve computational efficiency, for large-size networks with $N \geqslant 500$, we choose bases randomly instead of using the base-selection method presented in the main text. It reduces accuracy a little in a few cases, but the computational complexity is considerably reduced. Supplementary Figs. S2 (**a**) and (**b**) show the results of reconstruction for Ising and Voter dynamics on ER, WS

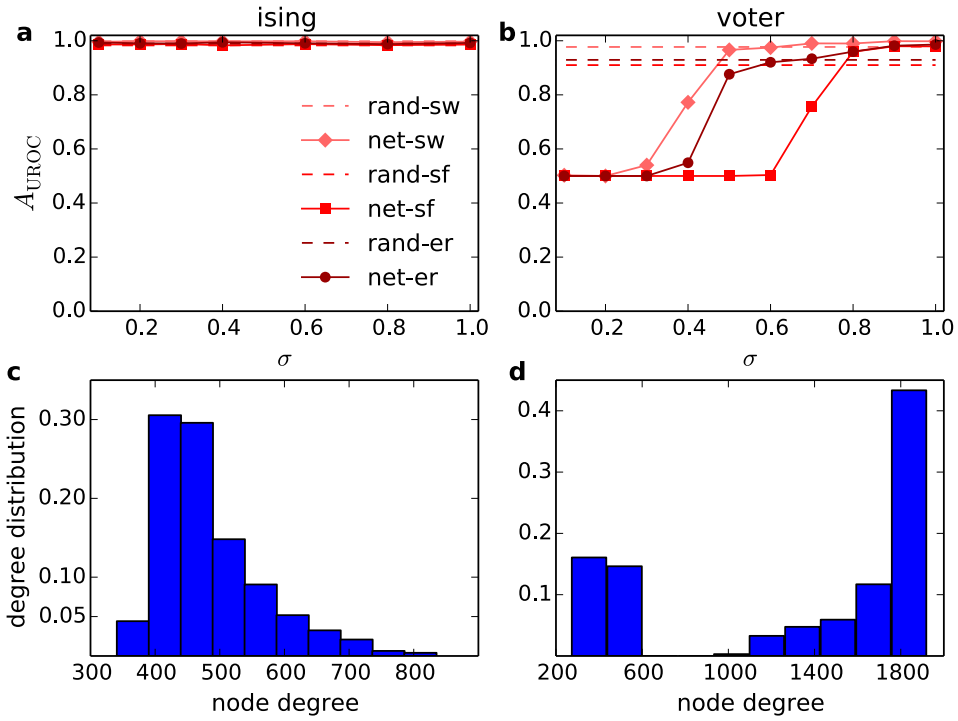
Supplementary Table S1 | Settings in numerical simulations. Parameter values in various binary-state dynamics and the period for initiating node states because of converging to steady state.

Model	Parameters	Convergent	Update period
Voter	—	Yes	100 (5 for N=100)
Kirman	$c_1 = 0.1$, $c_2 = 0.1$, $d = 0.08$	No	—
Ising Gluaber	$\beta = 2$	No	—
SIS	$\lambda = 0.2$, $\mu = 0.5$	No	—
Game	$\alpha = 0.1$, $\beta = 1$, $a = 6$, $b = 5$, $c = 1$, $d = 0$	No	—
Language	$s = 0.5$, $\alpha = 0.7$	No	—
Threshold	$M_k = 2/k$	Yes	5
Majority vote	$Q = 0.3$	Yes	10 (5 for N=100)

and SF networks. We found that for Ising dynamics, the results are almost not affected by the value of σ; but for Voter dynamics, a large value of σ is preferred. The possible reason is that, for dynamics with convergence such as Voter, the time series is dominated by all zeros or all ones. In the similarity networks, the nodes representing all zeros(or ones) time strings densely connect to each other, which leads to a bimodal degree distribution, as shown in Supplementary Fig. S2(**d**). We can see that there are more than 40% nodes in the rightmost bin. Thus, σ should be large enough to exclude these nodes, and then the reconstructing performance will approach high accuracy then, as shown in Fig. S2(**b**). For Ising dynamic, the degree distribution is like a bell shape, and there are no dominant zeros(or ones) time strings, so σ is not a key parameter. We also compare the performance of the networked base selection with the performance of randomly base selection. The networked indeed shows better performance, especially for dynamics with convergence.



Supplementary Figure S1 | Determination of threshold Δ. (**a**) AUROC as a function of threshold parameter Δ for the voter and Ising model on ER, SF and SW networks. (**b**) AUPR as a function of Δ for the two models and three networks. The network size $N = 100$ and $\langle k \rangle = 6$. The length of time series is 1.5×10^4. Other parameters of dynamics are shown in Supplementary Table S1.

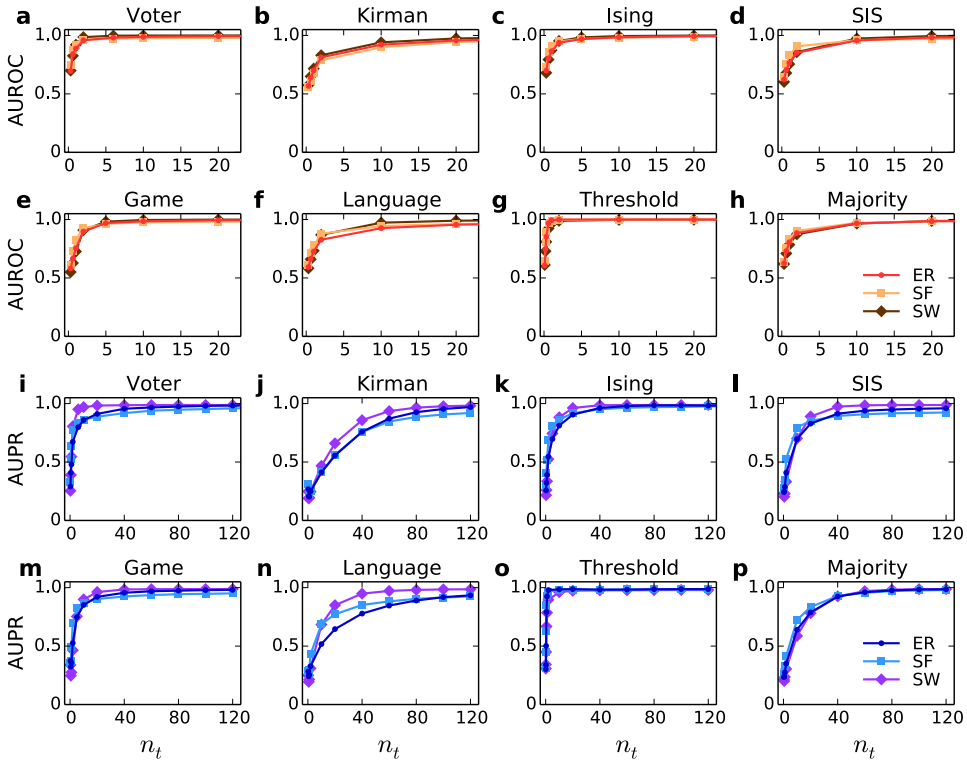


Supplementary Figure S2 | Determination of threshold σ. (**a,b**) AUROC as a function of threshold parameter σ for (**a**) the voter and (**b**) Ising model on ER, SF and SW networks, respectively. The dashed lines are the results of randomly selected bases. (**c**) The degree distribution of the constructed similarity network for Ising dynamic on ER network. (**d**) The degree distribution of the constructed similarity network for Voter dynamic on ER network. The network size $N = 100$ and $\langle k \rangle = 6$. The length of time series is 1.5×10^4. Other parameters of dynamics are shown in Supplementary Table S1.

There is an adjustable parameter λ in the lasso. In general, the parameter is determined by using cross-validation method, such as sklearn.linear_model.LassoCV in python. In terms of the cross-validation method, we obtained the proper value of λ, which is set to be 10^{-4} and 10^{-3} for reconstructing networks with $N \leqslant 500$ and $N = 1000$, respectively, in all reconstructions. All the convex optimizations are implemented in Python(version 2.7) and Sklearn(version 0.14).

2 Dependence of performance on data amount

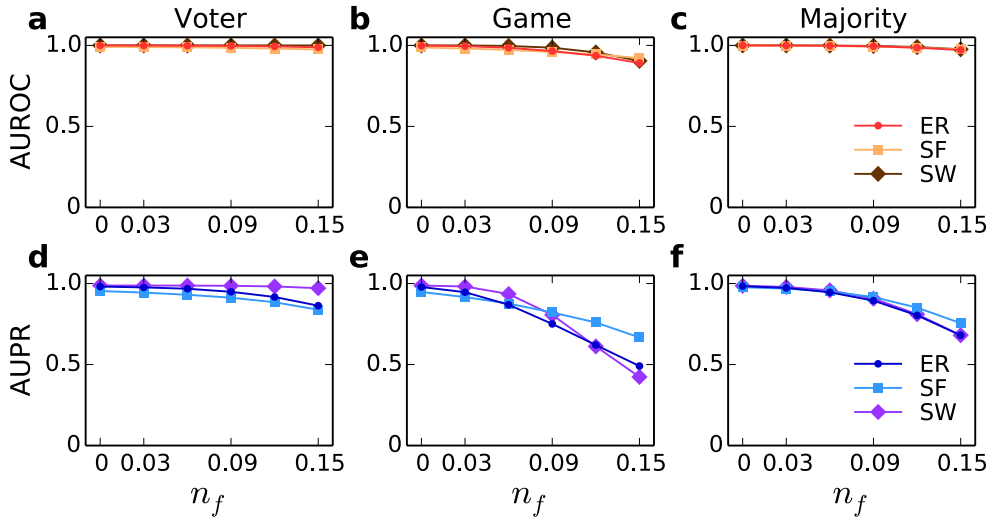
We examine how the length of time series affects reconstruction accuracy. We let n_t denote the ratio of the total length of time series normalized to the network size N. Supplementary Fig. S3 shows the reconstruction performance measured by AUROC and AUPR for various dynamics in combination with different types of networks. We find that AUROC and AUPR rapidly increases as n_t increases. After n_t exceeds a relatively small value, nearly full reconstruction can be achieved, which provides additional evidence for the high efficiency of our method. The results are summarized in Table II in the main text.



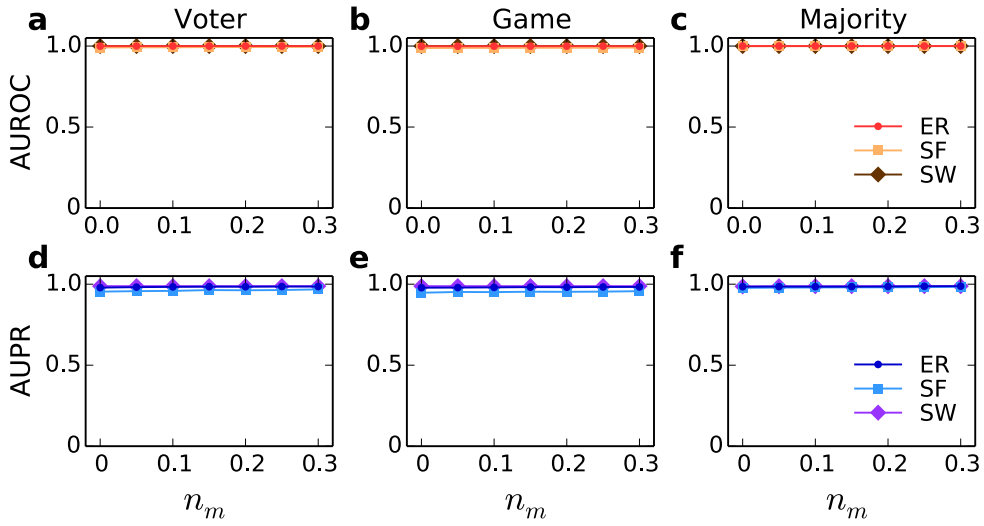
Supplementary Figure S3 | Reconstruction performance with respect to the length of time series. (**a-h**) AUROC and (**i-p**) AUPR as functions of the normalized length of time series n_t for various dynamics on ER, SF and SW networks. The network size $N = 500$ and $\langle k \rangle = 6$. Other parameter values of binary-state dynamics are shown in Supplementary Table S1.

3 Robustness against noise and missing data

Robustness against noise and missing data is important for evaluating the applicability of a method. We consider the scenario of noise-induced wrong records in time series. Specifically, we assume that a fraction n_{f} of binary states are wrong, and flip from 1 to zero or from zero to 1. The presence of unobservable nodes or missing data is quite often in the real situation. We assume that the data of a fraction of nodes, n_{m}, cannot be observed. We investigate the reconstruction accuracy as a function of n_{f} and n_{m}, respectively. As shown in Supplementary Fig. S4 and Supplementary Fig. S5, respectively, we find that high AUROC and AUPR remains in a wide range of n_{f} and n_{m}, providing strong evidence for the robustness of our reconstruction framework against measurement noise and missing data. The results are summarized in Table III in the main text.



Supplementary Figure S4 | Robustness against measurement noise. (**a,b,c**) AUROC and (**d,e,f**) AUPR as functions of the fraction n_f of wrong states in time series for the voter, Ising and majority model on ER, SF and SW networks. Parameters of networks and dynamics are the same as in Supplementary Fig. S3. $n_t = 100$.



Supplementary Figure S5 | Robustness against missing data. (**a,b,c**) AUROC and (**d,e,f**) AUPR as functions of the fraction n_m of unobservable nodes for the voter, Ising and majority model on ER, SF and SW networks. Parameters of networks and dynamics are the same as in Supplementary Fig. S3. $n_t = 100$.