

## Accepted Manuscript

Data based identification and prediction of nonlinear and complex dynamical systems

Wen-Xu Wang, Ying-Cheng Lai, Celso Grebogi

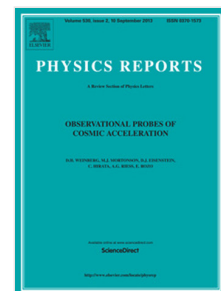
PII: S0370-1573(16)30134-X

DOI: <http://dx.doi.org/10.1016/j.physrep.2016.06.004>

Reference: PLREP 1910

To appear in: *Physics Reports*

Accepted date: 10 June 2016



Please cite this article as: W.-X. Wang, Y.-C. Lai, C. Grebogi, Data based identification and prediction of nonlinear and complex dynamical systems, *Physics Reports* (2016), <http://dx.doi.org/10.1016/j.physrep.2016.06.004>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# Data Based Identification and Prediction of Nonlinear and Complex Dynamical Systems

Wen-Xu Wang<sup>a,b</sup>, Ying-Cheng Lai<sup>c,d,e</sup>, and Celso Grebogi<sup>e</sup>

a School of Systems Science, Beijing Normal University, Beijing, 100875, China

b Business School, University of Shanghai for Science and Technology, Shanghai 200093, China

c School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, Arizona 85287, USA

d Department of Physics, Arizona State University, Tempe, Arizona 85287, USA

e Institute for Complex Systems and Mathematical Biology, King's College, University of Aberdeen, Aberdeen AB24 3UE, UK

## Abstract

The problem of reconstructing nonlinear and complex dynamical systems from measured data or time series is central to many scientific disciplines including physical, biological, computer, and social sciences, as well as engineering and economics. The classic approach to phase-space reconstruction through the methodology of delay-coordinate embedding has been practiced for more than three decades, but the paradigm is effective mostly for low-dimensional dynamical systems. Often, the methodology yields only a topological correspondence of the original system. There are situations in various fields of science and engineering where the systems of interest are complex and high dimensional with many interacting components. A complex system typically exhibits a rich variety of collective dynamics, and it is of great interest to be able to detect, classify, understand, predict, and control the dynamics using data that are becoming increasingly accessible due to the advances of modern information technology. To accomplish these tasks, especially prediction and control, an accurate reconstruction of the original system is required.

Nonlinear and complex systems identification aims at inferring, from data, the mathematical equations that govern the dynamical evolution and the complex interaction patterns, or topology, among the various components of the system. With successful reconstruction of the system equations and the connecting topology, it may be possible to address challenging and significant problems such as identification of causal relations among the interacting components and detection of hidden nodes. The “inverse” problem thus presents a grand challenge, requiring new paradigms beyond the traditional delay-coordinate embedding methodology.

The past fifteen years have witnessed rapid development of contemporary complex graph theory with broad applications in interdisciplinary science and engineering. The combination of graph, information, and nonlinear dynamical systems theories with tools from statistical physics, optimization, engineering control, applied mathematics, and scientific computing enables the development of a number of paradigms to address the problem of nonlinear and complex systems reconstruction. In this Review, we review the recent advances in this forefront and rapidly evolving field, with a focus on compressive sensing based methods. In particular, compressive sensing is a paradigm developed in recent years in applied mathematics, electrical engineering, and nonlinear physics to reconstruct sparse signals using only limited data. It has broad applications ranging from image compression/reconstruction to the analysis of large-scale sensor networks, and it has become a powerful technique to obtain high-fidelity signals for applications where sufficient observations are not available. We will describe in detail how compressive sensing can be exploited to address a diverse array of problems in data based reconstruction of nonlinear and complex networked systems. The problems include identification of chaotic systems and prediction of catastrophic bifurcations, forecasting future attractors of time-varying nonlinear systems, reconstruction of complex networks with oscillatory and evolutionary game dynamics, detection of hidden nodes, identification of chaotic elements in neuronal networks, and reconstruction of complex geospatial networks and nodal positioning. A number of alternative methods, such as those based on system response to external driving, synchronization, noise-induced dynamical correlation, will also be discussed. Due to the high relevance of network reconstruction to biological sciences, a special Section is devoted to a brief survey of the current methods to infer biological networks. Finally, a number of open problems including control and controllability of complex nonlinear dynamical networks are discussed.

The methods reviewed in this Review are principled on various concepts in complexity science and engineering such as phase transitions, bifurcations, stabilities, and robustness. The methodologies have the potential to significantly improve our ability to understand a variety of complex dynamical systems ranging from gene regulatory systems to social networks towards the ultimate goal of controlling such systems.

# Contents

<b>1</b>	<b>Introduction</b>	<b>5</b>
1.1	Existing works on data based reconstruction of nonlinear dynamical systems . . . .	5
1.2	Existing works on data based reconstruction of complex networks and dynamical processes . . . . .	6
1.3	Compressive sensing based reconstruction of nonlinear and complex dynamical systems . . . . .	7
1.4	Plan of this review . . . . .	8
<b>2</b>	<b>Compressive sensing based nonlinear dynamical systems identification</b>	<b>12</b>
2.1	Introduction to the compressive sensing algorithm . . . . .	12
2.2	Mathematical formulation of systems identification based on compressive sensing .	14
2.3	Reconstruction and identification of chaotic systems . . . . .	15
2.4	Predicting catastrophe in nonlinear dynamical systems . . . . .	18
2.4.1	Predicting catastrophic bifurcations based on compressive sensing . . . . .	18
2.4.2	Using compressive sensing to predict tipping points in complex systems . .	19
2.5	Forecasting future states (attractors) of nonlinear dynamical systems . . . . .	22
<b>3</b>	<b>Compressive sensing based reconstruction of complex networked systems</b>	<b>29</b>
3.1	Reconstruction of coupled oscillator networks . . . . .	29
3.2	Reconstruction of complex networks with evolutionary-game dynamics . . . . .	35
3.3	Detection of hidden nodes in complex networks . . . . .	42
3.3.1	Principle of detecting hidden nodes based on compressive sensing . . . . .	42
3.3.2	Mathematical formulation of compressive sensing based detection of a hidden node . . . . .	44
3.3.3	Examples of hidden node detection in the presence of noise . . . . .	48
3.4	Identifying chaotic elements in neuronal networks . . . . .	57
3.4.1	Basic procedure of identifying chaotic neurons . . . . .	59
3.4.2	Example: identifying chaotic neurons in the FitzHugh-Nagumo (FHN) network . . . . .	60
3.5	Data based reconstruction of complex geospatial networks and nodal positioning .	63
3.6	Reconstruction of complex spreading networks from binary data . . . . .	70
3.6.1	Mathematical formulation . . . . .	71
3.6.2	Reconstructing complex spreading networks: examples . . . . .	77
<b>4</b>	<b>Alternative methods for reconstructing complex, nonlinear dynamical networks</b>	<b>83</b>
4.1	Reconstructing complex networks from response dynamics . . . . .	83
4.2	Reconstructing complex networks via system clone . . . . .	84

4.3	Network reconstruction via phase-space linearization . . . . .	85
4.4	Reconstruction of oscillator networks based on noise induced dynamical correlation . . . . .	86
4.5	Reverse engineering of complex systems . . . . .	91
<b>5</b>	<b>Inference approaches to reconstruction of biological networks</b>	<b>94</b>
5.1	Correlation based methods . . . . .	94
5.1.1	Value based methods . . . . .	94
5.1.2	Rank based methods . . . . .	95
5.2	Causality based measures . . . . .	96
5.3	Information-theoretic based methods . . . . .	98
5.3.1	Mutual information . . . . .	98
5.3.2	Maximal information coefficient . . . . .	99
5.3.3	Maximum entropy . . . . .	100
5.4	Bayesian network . . . . .	100
5.5	Regression and resampling . . . . .	100
5.6	Supervised and semi-supervised methods . . . . .	101
5.7	Transfer and joint entropies . . . . .	101
5.8	Distinguishing between direct and indirect interactions . . . . .	102
<b>6</b>	<b>Discussions and future perspectives</b>	<b>104</b>
6.1	Localization of diffusion sources in complex networks . . . . .	105
6.2	Data based reconstruction of complex networks with binary-state dynamics . . . . .	106
6.3	Universal structural estimator and dynamics approximator for complex networks . . . . .	108
6.4	Controlling nonlinear and complex dynamical networks . . . . .	109

# 1 Introduction

An outstanding problem in interdisciplinary science is nonlinear and complex systems identification, prediction, and control. Given a complex dynamical system, the various types of dynamical processes are of great interest. The ultimate goal in the study of complex systems is to devise practically implementable strategies to control the collective dynamics. A great challenge is that the network structure and the nodal dynamics are often unknown but only limited measured time series are available. To control the system dynamics, it is imperative to map out the system details from data. Reconstructing complex network structure and dynamics from data, the inverse problem, has thus become a central issue in contemporary network science and engineering [1–38]. There are broad applications of the solutions of the network reconstruction problem, due to the ubiquity of complex interacting patterns arising from many systems in a variety of disciplines [39–42].

## 1.1 Existing works on data based reconstruction of nonlinear dynamical systems

The traditional paradigm of nonlinear time series analysis is the delay-coordinate embedding method, the mathematical foundation of which was laid by Takens more than three decades ago [43]. He proved that, under fairly general conditions, the underlying dynamical system can be faithfully reconstructed from time series in the sense that a one-to-one correspondence can be established between the reconstructed and the true but unknown dynamical systems. Based on the reconstruction, quantities of importance for understanding the system can be estimated, such as the relative weights of deterministcity and stochasticity of the underlying system, its dimensionality, the Lyapunov exponents, and unstable periodic orbits that constitute the skeleton of the invariant set responsible for the observed dynamics.

There exists a large body of literature on the application of the delay-coordinate embedding technique to nonlinear/chaotic dynamical systems [44, 45]. A pioneering work in this field is Ref. [46]. The problem of determining the proper time delay was investigated [47–54], with a firm theoretical foundation established by exploiting the statistics for testing continuity and differentiability from chaotic time series [55–59]. The mathematical foundation for the required embedding dimension for chaotic attractors was laid in Ref. [60]. There were works on the analysis of transient chaotic time series [61–65], on the reconstruction of dynamical systems with time delay [66], on detecting unstable periodic orbits from time series [67–72], on computing the fractal dimensions from chaotic data [53, 54, 73–78], and on estimating the Lyapunov exponents [79–85].

There were also works on forecasting nonlinear dynamical systems [23, 86–108]. A conventional approach is to approximate a nonlinear system with a large collection of linear equations in different regions of the phase space to reconstruct the Jacobian matrices on a proper grid [101, 102, 104] or fit ordinary differential equations to chaotic data [103]. Approaches based on chaotic synchronization [106] or genetic algorithms [107, 108] to parameter estimation were also investigated. In most existing works, short-term predictions of a dynamical system can be achieved by employing the classical delay-coordinate embedding paradigm [43, 44]. For nonstationary systems, the method of over-embedding was introduced [109] in which the time-varying parameters were treated as independent dynamical variables so that the essential aspects of determinism of the

underlying system can be restored. A recently developed framework based on compressive sensing was able to predict the *exact* forms of both system equations and parameter functions based on available time series for stationary [23] and time-varying dynamical systems [25].

## 1.2 Existing works on data based reconstruction of complex networks and dynamical processes

Data based reconstruction of complex networks in general is deemed to be an important but difficult problem and has attracted continuous interest, where the goal is to uncover the full topology of the network based on simultaneously measured time series [1–38]. There were previous efforts in nonlinear systems identification and parameter estimation for coupled oscillators and spatiotemporal systems, such as the auto-synchronization method [106]. There were also works on revealing the connection patterns of networks. For example, methods were proposed to estimate the network topology controlled by feedback or delayed feedback [6,9,21]. Network connectivity can be reconstructed from the collective dynamical trajectories using response dynamics [20]. The approach of random phase resetting was introduced to reconstruct the details of the network structure [18]. For neuronal systems, there was a statistical method to track the structural changes [28,32]. Some earlier methods required more information about the network than just data. For example, the following two approaches require *complete* information about the dynamical processes, e.g., equations governing the evolutions of *all* nodes on the network. (1) In Ref. [6], the detailed dynamics at each node is assumed to be known. A replica of the network, or a computational model of this “target” network, can then be constructed, with the exception that the interaction strengths among the nodes are chosen randomly. It has been demonstrated that in situations where a Lyapunov function for the network dynamics exists, the connectivity of the model network converges to that of the target network [6]. (2) In Ref. [8], a Kuramoto-type of phase dynamics [110,111] on the network is assumed, where a steady-state solution exists. By linearizing the network dynamics about the steady-state solution, the associated Jacobian matrix can be obtained, which reflects the network topology and connectivity. Besides requiring complete information about the nodal dynamics, the amount of computations required tends to increase dramatically with the size of the network [6]. For example, suppose the nodal dynamics is described by a set of differential equations. For a network of size  $N$ , in order for its structure to be predicted, the number of differential equations to be solved typically increases with  $N$  as  $N^2$ .

For nonlinear dynamical networks, there was a method [10] based on chaotic time-series analysis through estimating the elements of the Jacobian matrix, which are the mutual partial derivatives of the dynamical variables on different nodes in the network. A statistically significant entry in the matrix implies a connection between the two nodes specified by the row and the column indices of that entry. Because of the mathematical nature of the Jacobian matrix, i.e., it is meaningful only for infinitesimal tangent vectors, linearization of the dynamics in the neighborhoods of the reconstructed phase-space points is needed, for which constrained optimization techniques [112–118] were found to be effective [10]. Estimating the Jacobian matrices, however, has been a challenging problem in nonlinear dynamics [81,105] and its reliability can be ensured but only for low-dimensional, deterministic dynamical systems. The method [10] appeared thus to be limited to small networks with sparse connections.

While many of the earlier works required complete or partial information about the intrinsic dy-



namics of the nodes and their coupling functions, completely data-driven and model-free methods exist. For example, The global climate network was reconstructed using the mutual information method, enabling energy and information flow in the network to be studied [14]. The sampling bias of DNA sequences in viruses from different regions can be used to reveal the geospatial topologies of the influenza networks [16]. Network structure can also be obtained by calculating the causal influences among the time series based on the Granger causality [119] method [5, 33, 120–122], the overarching framework [123], the transfer entropy method [29], or the method of inner composition alignment [19]. However, such causality based methods are unable to reveal information about the nodal dynamical equations. In addition, there were regression-based methods [124] for systems identification based on the least squares approximation through the Kronecker-product representation [125], which would require large amounts of data.

In systems biology, reverse engineering of gene regulatory networks from expression data is a fundamentally important problem, and it attracts a tremendous amount of interest [126–134]. The wide spectrum of methods for modeling genetic regulatory networks can be categorized based on the level of details with which the genetic interactions and dynamics are modeled [135]. One of the classical mathematical formalisms used to model the dynamics of biological processes is differential equations, which can capture the dynamics of each component in a system at a detailed level [136–138]. A major limitation of this approach is its overwhelming complexity and the resulting computational requirement, which limits its applicability to small-scale systems. In contrast, Boolean network models assume that the states of components in the system are binary and the state transitions are governed by logic operations [139–141]. Since the gene expressions are usually described by their expression levels and the interaction patterns between genes may not be logic operations, in some cases the Boolean network models may not be biologically appropriate. Another class of methods explicitly model biological systems as a graph in which the vertices represent basic units in the system and the edges characterize the relationships between the units. The graph itself can be constructed either by directly comparing the measurements for the vertices based on certain metric, such as the Euclidean distance, mutual information, or correlation coefficient [130, 142, 143], or by some probabilistic approaches for Bayesian network learning [144–146]. However, inferring regulatory interactions based on Bayesian networks is an intractable problem [147, 148]. The linear regression models for learning regulatory networks assume that expression level of a gene can be approximated by a linear combination of the expressions of other genes [128, 129, 133, 149], and such models form a middle ground between the models based on differential equations and Boolean logic.

### 1.3 Compressive sensing based reconstruction of nonlinear and complex dynamical systems

A recent line of research [22–24, 30, 31, 35–38] exploited compressive sensing [113–118]. The basic principle is that the dynamics of many natural and man-made systems are determined by smooth enough functions that can be approximated by finite series expansions. The task then becomes that of estimating the coefficients in the series representation of the vector field governing the system dynamics. In general, the series can contain high order terms, and the total number of coefficients to be estimated can be quite large. While this is a challenging problem, if most coefficients are zero (or negligible), the vector constituting all the coefficients will be sparse. In addition, a generic fea-

ture of complex networks in the real world is that they are sparse [42]. Thus for realistic nonlinear dynamical networks, the vectors to be reconstructed are typically sparse, and the problem of sparse vector estimation can then be solved by the paradigm of compressive sensing [113, 114, 116–118] that reconstructs a sparse signal from limited observations. Since the observation requirements can be relaxed considerably as compared to those associated with conventional signal reconstruction schemes, compressive sensing has evolved into a powerful technique to reconstruct sparse signal from small amounts of observations that are much less than those required in conventional approaches. Compressive sensing has been introduced to the field of network reconstruction for discrete time and continuous time nodal dynamics [23, 24], for evolutionary game dynamics [22], for detecting hidden nodes [31, 36], for predicting and controlling synchronization dynamics [30], and for reconstructing spreading dynamics based on binary data [37]. Compressive sensing also finds applications in quantum measurement science, e.g., to exponentially reduce the experimental configurations required for quantum tomography [150].

#### 1.4 Plan of this review

This Review presents the recent advances in the forefront and rapidly evolving field of nonlinear and complex dynamical systems identification and prediction. Our focus will be on the compressive sensing based approaches. Alternative approaches will also be discussed, which include noised-induced dynamical mapping, perturbations, reverse engineering, synchronization, inner composition alignment, global silencing, Granger causality, and alternative optimization algorithms.

In **Sec. 2**, we first introduce the principle of compressive sensing and discuss nonlinear dynamical systems identification and prediction. We next discuss a compressive sensing based approach to predicting catastrophes in nonlinear dynamical systems under the assumption that the system equations are completely unknown and only time series reflecting the evolution of the dynamical variables of the system are available. We then turn to time-varying nonlinear dynamical systems, motivated by the fact that systems with one or a few parameters varying slowly with time are of considerable interest in many areas of science and engineering. In such a system, the attractors in the future can be characteristically different from those at the present. To predict the possible future attractors based on available information at the present is thus a well-defined and meaningful problem, which is challenging especially when the system equations are not known but only time-series measurements are available. We review a compressive-sensing based method for time-varying systems. This framework allows us to reconstruct the system equations and the time dependence of parameters based on limited measurements so that the future attractors of the system can be predicted through computation.

**Section 3** focuses on compressive sensing based reconstruction of complex networked systems. The following problems will be discussed in detail.

*Reconstruction of coupled oscillator networks.* The basic idea is that the mathematical functions determining the dynamical couplings in a physical network can be expressed by power-series expansions. The task is then to estimate all the nonzero coefficients, which can be accomplished by exploiting compressive sensing [24].

*Reconstruction of social networks based on evolutionary-game data via compressive sensing.* Evolutionary games are a common type interactions in a variety of complex networked, natural

and social systems. Given such a system, uncovering the interacting structure of the underlying network is key to understanding its collective dynamics. We discuss a compressive sensing based method to uncover the network topology using evolutionary-game data. In particular, in a typical game, agents use different strategies in order to gain the maximum payoff. The strategies can be divided into two types: cooperation and defection. It was shown [22] that, even when the available information about each agent's strategy and payoff is limited, the compressive-sensing based approach can yield precise knowledge of the node-to-node interaction patterns in a highly efficient manner. In addition to numerical validation of the method with model complex networks, we discuss an actual social experiment in which participants forming a friendship network played a typical game to generate short sequences of strategy and payoff data. The high prediction accuracy achieved and the unique requirement of extremely small data set suggest that the method can be appealing to potential applications to reveal "hidden" networks embedded in various social and economic systems.

*Detecting hidden nodes in complex networks from time series.* The power of science lies in its ability to infer and predict the existence of objects from which no direct information can be obtained experimentally or observationally. A well known example is to ascertain the existence of black holes of various masses in different parts of the universe from indirect evidence, such as X-ray emissions. In complex networks, the problem of detecting *hidden* nodes can be stated, as follows. Consider a network whose topology is completely unknown but whose nodes consist of two types: one accessible and another inaccessible from the outside world. The accessible nodes can be observed or monitored, and we assume that time series are available from each node in this group. The inaccessible nodes are shielded from the outside and they are essentially "hidden." The question is, can we infer, based solely on the available time series from the accessible nodes, the existence and locations of the hidden nodes? Since no data from the hidden nodes are available, nor can they be observed directly, they act as some sort of "black box" from the outside world. Solution of the network hidden node detection problem has potential applications in different fields of significant current interest. For example, to uncover the topology of a terrorist organization and especially, various ring leaders of the network is a critical task in defense. The leaders may be hidden in the sense that no direct information about them can be obtained, yet they may rely on a number of couriers to operate, which are often subject to surveillance. Similar situations arise in epidemiology, where the original carrier of a virus may be hidden, or in a biology network where one wishes to detect the most influential node from which no direct observation can be made. We discuss in detail a compressive sensing based method [31, 36] to ascertain hidden nodes in complex networks and to distinguish them from various noise sources.

*Identifying chaotic elements in complex neuronal networks.* We discuss a completely data-driven approach [35] to reconstructing coupled neuronal networks that contain a small subset of chaotic neurons. Such chaotic elements can be the result of parameter shift in their individual dynamical systems, and may lead to abnormal functions of the network. To accurately identify the chaotic neurons may thus be necessary and important, for example, for applying appropriate controls to bring the network to a normal state. However, due to couplings among the nodes the measured time series even from non-chaotic neurons would appear random, rendering inapplicable traditional nonlinear time-series analysis, such as the delay-coordinate embedding method, which yields information about the global dynamics of the entire network. The method to be discussed is based on compressive sensing. In particular, identifying chaotic elements can be formulated as

a general problem of reconstructing the nodal dynamical systems, network connections, and all coupling functions as well as their weights.

*Data based reconstruction of complex geospatial networks, nodal positioning, and detection of hidden node.* Complex geospatial networks with components distributed in the real geophysical space are an important part of the modern infrastructure. Given a complex geospatial network with nodes distributed in a two-dimensional region of the physical space, can the locations of the nodes be determined and their connection patterns be uncovered based solely on data? In realistic applications, time series/signals can be collected from a single location. A key challenge is that the signals collected are necessarily time delayed, due to the varying physical distances from the nodes to the data collection center. We discuss a compressive sensing based approach [38] that enables reconstruction of the full topology of the underlying geospatial network and more importantly, accurate estimate of the time delays. A standard triangularization algorithm can then be employed to find the physical locations of the nodes in the network. A hidden source or threat, from which no signal can be obtained, can also be detected through accurate detection of all its neighboring nodes. As a geospatial network has the feature that a node tends to connect with geophysically nearby nodes, the localized region that contains the hidden node can be identified.

*Reconstructing complex spreading networks with natural diversity and identifying hidden source.* Among the various types of collective dynamics on complex networks, propagation or spreading dynamics is of paramount importance as it is directly relevant to issues of tremendous interests such as epidemic and disease outbreak in the human society and virus spreading on computer networks. We discuss a theoretical and computational framework based on compressive sensing to reconstruct networks of arbitrary topologies, in which spreading dynamics with heterogeneous diffusion probabilities take place [37]. The approach enables identification with high accuracy of the external sources that trigger the spreading dynamics. Especially, a full reconstruction of the stochastic and inhomogeneous interactions presented in the real-world networks can be achieved from small amounts of polarized (binary) data, a virtue of compressive sensing. After the outbreak of diffusion, hidden sources outside the network, from which no direct routes of propagation are available, can be ascertained and located with high confidence. This represents essentially a new paradigm for tracing, monitoring, and controlling epidemic invasion in complex networked systems, which will be of value to defending and preserving the systems against disturbances and attacks.

**Section 4** presents a number of alternative methods for reconstructing complex and nonlinear dynamical networks. These are methods based on (1) response of the system to external driving, (2) synchronization (via system clone), (3) phase-space linearization, (4) noise induced dynamical correlation, and (5) automated reverse engineering. In particular, for the system response based method (1), the basic idea was to measure the collective response of the oscillator network to an external driving signal. With repeated measurements of the dynamical states of the nodes under sufficiently independent driving realizations, the network topology can be recovered [8]. Since complex networks are generally sparse, the number of realizations of external driving can be much smaller than the network size. For the synchronization based method, the idea was to design a replica or a clone system that is sufficiently close to the original network without requiring knowledge about network structure [6]. From the clone system, the connectivities and interactions among nodes can be obtained directly, realizing network reconstruction. For the phase-space linearization method,  $L_1$ -minimization in the phase space of a networked system is employed to reconstruct the

topology without knowledge of the self-dynamics of nodes and without using any external perturbation to the state of nodes [10]. For the method based on noise-induced dynamical correlation, the principle was based on exploiting the rich interplay between nonlinear dynamics and stochastic fluctuations [13, 15, 27]. Under the condition that the influence of noise on the evolution of infinitesimal tangent vectors in the phase space of the underlying networked dynamical system is dominant, it can be argued [15, 27] that the dynamical correlation matrix that can be computed readily from the available nodal time series approximates the network adjacency matrix, fully unveiling the network topology. For the automated reverse engineering method, the approach was based on problem solving using partitioning, automated probing and snipping [151], a process that is akin to genetic algorithm. Each of the five methods will be described in reasonable details.

Network reconstruction was pioneered in biological sciences. **Section 5** is devoted to a concise survey of the approaches to reconstructing biological networks. Those include methods based on correlation, causality, information-theoretic measures, Bayesian inference, regression and resampling, supervision and semi-supervision, transfer and joint entropies, and distinguishing between direct and indirect interactions.

Finally, in **Sec. 6**, we offer a general discussion of the field of data-based reconstruction of complex networks and speculate on a number of open problems. These include: localization of diffusion sources in complex networks, reconstruction of complex networks with binary state dynamics, possibility of developing a universal framework of structural estimator and dynamics approximator for complex networks, and a framework of control and controllability for complex nonlinear dynamical networks.

## 2 Compressive sensing based nonlinear dynamical systems identification

### 2.1 Introduction to the compressive sensing algorithm

Compressive sensing is a paradigm developed in recent years by applied mathematicians and electrical engineers to reconstruct sparse signals using only limited data [113–118]. The observations are measured by linear projections of the original data on a few predetermined, random vectors. Since the requirement for the observations is considerably less comparing to conventional signal reconstruction schemes, compressive sensing has been deemed as a powerful technique to obtain high-fidelity signal especially in cases where sufficient observations are not available. Compressive sensing has broad applications ranging from image compression/reconstruction to the analysis of large-scale sensor networks [117, 118].

Mathematically, the problem of compressive sensing is to reconstruct a vector  $\mathbf{a} \in \mathbb{R}^N$  from linear measurements  $\mathbf{X}$  about  $\mathbf{a}$  in the form:

$$\mathbf{X} = \mathcal{G} \cdot \mathbf{a} \quad (1)$$

where  $\mathbf{X} \in \mathbb{R}^M$  and  $\mathcal{G}$  is an  $M \times N$  matrix. By definition, the number of measurements is much less than that of the unknown signal, i.e,  $M \ll N$ . Accurate recovery of the original signal is possible through the solution of the following convex optimization problem [113, 114]:

$$\min \|\mathbf{a}\|_1 \quad \text{subject to} \quad \mathcal{G} \cdot \mathbf{a} = \mathbf{X}, \quad (2)$$

where

$$\|\mathbf{a}\|_1 = \sum_{i=1}^N |\mathbf{a}_i| \quad (3)$$

is the  $l_1$  norm of vector  $\mathbf{a}$ .

The general principle of solving the convex optimization problem Eq. (2) can be described briefly, as follows. By introducing a new variable vector  $\mathbf{u} \in \mathbb{R}^N$ , Eq. (2) can be recast into a linear constraint minimization problem [113, 114]:

$$\min \sum_{i=1}^N u_i \quad \text{subject to} \quad \begin{cases} a_i - u_i \leq 0 \\ -a_i - u_i \leq 0 \\ \mathcal{G} \cdot \mathbf{a} = \mathbf{X} \end{cases} \quad (4)$$

Defining  $\mathbf{z} = [\mathbf{a}^T, \mathbf{u}^T]^T$ , we can rewrite Eq. (4) as

$$\langle \mathbf{c}_0, \mathbf{z} \rangle \quad \text{subject to} \quad \begin{cases} f_i(\mathbf{z}) \leq 0 \\ f'_i(\mathbf{z}) \leq 0 \\ \mathcal{G}_0 \cdot \mathbf{z} = \mathbf{X} \end{cases}, \quad (5)$$

where  $f_i(\mathbf{z}) = \langle \mathbf{c}_i, \mathbf{z} \rangle$  and  $f'_i(\mathbf{z}) = \langle \mathbf{c}'_i, \mathbf{z} \rangle$  ( $\langle \cdot \rangle$  denotes the inner product of the two vectors). Here,  $\mathbf{c}_0, \mathbf{c}_i, \mathbf{c}'_i \in \mathbb{R}^{2N}$ ,  $\mathcal{G}_0$  is a  $M \times 2N$  matrix,  $(\mathbf{c}_0)_j = 0$  for  $j \leq N$  and  $(\mathbf{c}_0)_j = 1$  for  $j > N$ ;  $(\mathbf{c}_i)_j = 1$  for

$j = i$ ,  $(\mathbf{c}_i)_j = -1$  for  $j = N + i$ ;  $(\mathbf{c}'_i)_j = 1$  for  $j = i$ ,  $(\mathbf{c}'_i)_j = -1$  for  $j = N + i$ ;  $\mathcal{G}_0 = [\mathcal{I}^{M \times N}, \mathcal{G}]$ . To solve the linear constraint minimization problem in Eq. (5), one can use the Karush-Kuhn-Tucker conditions [152], i.e., at the optimal point  $\mathbf{z}^*$ , there exist vectors  $\mathbf{v}^* \in \mathbb{R}^M$ ,  $\mathbf{q}^* \in \mathbb{R}^N$ ,  $\mathbf{q}'^* \in \mathbb{R}^N$ , where  $(\mathbf{q}^*)_j \geq 0$  and  $(\mathbf{q}'^*)_j \geq 0$  for  $j = 1, \dots, N$ , such that the following hold:

$$\begin{aligned} \mathbf{c}_0 + \mathcal{G}_0^T \cdot \mathbf{v}^* + \sum_i q_i^* \mathbf{c}_i + \sum_i q_i'^* \mathbf{c}'_i &= \mathbf{0}, \\ q_i^* f_i(\mathbf{z}^*) &= 0, i = 1, \dots, N, \\ q_i'^* f'_i(\mathbf{z}^*) &= 0, i = 1, \dots, N, \\ \mathcal{G}_0 \cdot \mathbf{z}^* &= \mathbf{X}, \\ f_i(\mathbf{z}^*) &\leq 0, \\ f'_i(\mathbf{z}^*) &\leq 0. \end{aligned} \quad (6)$$

Equation (6) can be solved by the classical Newton method in the valid solution set determined by the inequality constraints

$$\{q_i \geq 0, q_i' \geq 0, f_i(\mathbf{z}) \leq 0, f'_i(\mathbf{z}) \leq 0\}, \quad (7)$$

where a point  $(\mathbf{z}, \mathbf{v}, \mathbf{q}, \mathbf{q}')$  in this set is called as an interior point. Define a residual vector for all the equality conditions in Eq. (6) as  $\mathbf{r} = [\mathbf{r}_1^T, \mathbf{r}_2^T, \mathbf{r}_3^T, \mathbf{r}_4^T]^T$ , with  $\mathbf{r}_i$  ( $i = 1, \dots, 4$ ) given by where

$$\begin{aligned} \mathbf{r}_1 &= \mathbf{c}_0 + \mathcal{G}_0^T \cdot \mathbf{v} + \sum_i q_i \mathbf{c}_i + \sum_i q_i' \mathbf{c}'_i, \\ \mathbf{r}_2 &= -\lambda \cdot \mathbf{f}, \\ \mathbf{r}_3 &= -\lambda' \cdot \mathbf{f}', \\ \mathbf{r}_4 &= \mathcal{G}_0 \cdot \mathbf{z} - \mathbf{X}, \end{aligned} \quad (8)$$

where  $\mathbf{f} = [f_1(\mathbf{z}), f_2(\mathbf{z}), \dots, f_N(\mathbf{z})]^T$ ,  $\mathbf{f}' = [f'_1(\mathbf{z}), f'_2(\mathbf{z}), \dots, f'_N(\mathbf{z})]^T$ ,  $\lambda$  and  $\lambda'$  are diagonal matrices with  $(\lambda)_{ii} = q_i$  and  $(\lambda')_{ii} = q_i'$ . To find the solution to Eq. (6), one linearizes the residual vector  $\mathbf{r}$  using the Taylor expansion about the point  $(\mathbf{z}, \mathbf{v}, \mathbf{q}, \mathbf{q}')$ , which gives

$$\mathbf{r}(\mathbf{z} + \Delta \mathbf{z}, \mathbf{v} + \Delta \mathbf{v}, \mathbf{q} + \Delta \mathbf{q}, \mathbf{q}' + \Delta \mathbf{q}') = \mathbf{r}(\mathbf{z}, \mathbf{v}, \mathbf{q}, \mathbf{q}') + \mathcal{J}(\mathbf{z}, \mathbf{v}, \mathbf{q}, \mathbf{q}') \cdot \begin{pmatrix} \Delta \mathbf{z} \\ \Delta \mathbf{v} \\ \Delta \mathbf{q} \\ \Delta \mathbf{q}' \end{pmatrix}, \quad (9)$$

where  $\mathcal{J}$  is the Jacobian matrix of  $\mathbf{r}$  given by

$$\mathcal{J} = \begin{pmatrix} \mathbf{0} & \mathcal{G}_0^T & \mathcal{C}^T & \mathcal{C}'^T \\ -\lambda \cdot \mathcal{C} & \mathbf{0} & -\mathcal{F} & \mathbf{0} \\ -\lambda' \cdot \mathcal{C}' & \mathbf{0} & \mathbf{0} & -\mathcal{F}' \\ \mathcal{G}_0 & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}, \quad (10)$$

the  $N \times 2N$  matrices  $\mathcal{C}$  and  $\mathcal{C}'$  have  $\mathbf{c}_i$  and  $\mathbf{c}'_i$  as rows, respectively,  $\mathcal{F}$  and  $\mathcal{F}'$  are diagonal matrices with  $(\mathcal{F})_{ii} = f_i(\mathbf{z})$  and  $(\mathcal{F}')_{ii} = f'_i(\mathbf{z})$ . The steepest descent direction can be obtained by setting zero the left-hand side of Eq. (9), which gives

$$\begin{pmatrix} \Delta \mathbf{z} \\ \Delta \mathbf{v} \\ \Delta \mathbf{q} \\ \Delta \mathbf{q}' \end{pmatrix} = -\mathcal{J}^{-1} \cdot \mathbf{r}. \quad (11)$$

With the descent direction so determined, to solve Eq. (6), one can update the solution by  $\mathbf{z} = \mathbf{z} + s\Delta\mathbf{z}$ ,  $\mathbf{v} = \mathbf{v} + s\Delta\mathbf{v}$ ,  $\mathbf{q} = \mathbf{q} + s\Delta\mathbf{q}$ ,  $\mathbf{q}' = \mathbf{q}' + s\Delta\mathbf{q}'$  with step length  $s$ , where  $s$  should be chosen to guarantee that  $(\mathbf{z} + s\Delta\mathbf{z}, \mathbf{v} + s\Delta\mathbf{v}, \mathbf{q} + s\Delta\mathbf{q}, \mathbf{q}' + s\Delta\mathbf{q}')$  is an interior point of the valid solution set Eq. (7). Iterating this procedure gives the reconstructed sparse signal  $\mathbf{a}$ .

## 2.2 Mathematical formulation of systems identification based on compressive sensing

The inverse problem of identifying nonlinear dynamical systems can be cast in the framework of compressive sensing so that optimal solutions can be obtained even when the number of base coefficients to be estimated is large and/or the amount of available data is small. Consider systems described by

$$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}), \quad (12)$$

where  $\mathbf{x} \in R^m$  represents the set of externally accessible dynamical variables and  $\mathbf{F}$  is a smooth vector function in  $R^m$ . The  $j$ th component of  $\mathbf{F}(\mathbf{x})$  can be represented as a power series:

$$[\mathbf{F}(\mathbf{x})]_j = \sum_{l_1=0}^n \sum_{l_2=0}^n \cdots \sum_{l_m=0}^n (a_j)_{l_1, \dots, l_m} \cdot x_1^{l_1} x_2^{l_2} \cdots x_m^{l_m}, \quad (13)$$

where  $x_k$  ( $k = 1, \dots, m$ ) is the  $k$ th component of the dynamical variable, and the scalar coefficient of each product term  $(a_j)_{l_1, \dots, l_m} \in R$  is to be determined from measurements. Note that the terms in Eq. (13) are all possible products of different components with different powers, and there are  $(1+n)^m$  terms in total.

Without loss of generality, one can examine one dynamical variable of the system. (Procedures for other variables are similar.) For example, to construct the measurement vector  $\mathbf{X}$  and the matrix  $\mathbf{G}$  for the case of  $m = 3$  (dynamical variables  $x$ ,  $y$ , and  $z$ ) and  $n = 3$ , one has the following explicit equation for the first dynamical variable:

$$[\mathbf{F}(\mathbf{x})]_1 \equiv (a_1)_{0,0,0} x^0 y^0 z^0 + (a_1)_{1,0,0} x^1 y^0 z^0 + \cdots + (a_1)_{3,3,3} x^3 y^3 z^3.$$

Denote the coefficients of  $[\mathbf{F}(\mathbf{x})]_1$  by  $\mathbf{a}_1 = [(a_1)_{0,0,0}, (a_1)_{1,0,0}, \dots, (a_1)_{3,3,3}]^T$ . Assuming that measurements of  $\mathbf{x}(t)$  at a set of time  $t_1, t_2, \dots, t_w$  are available, one can write

$$\mathbf{g}(t) = [x(t)^0 y(t)^0 z(t)^0, x(t)^0 y(t)^0 z(t)^1, \dots, x(t)^3 y(t)^3 z(t)^3]$$

such that  $[\mathbf{F}(\mathbf{x}(t))]_1 = \mathbf{g}(t) \cdot \mathbf{a}_1$ . From the expression of  $[\mathbf{F}(\mathbf{x})]_1$ , one can choose the measurement vector as

$$\mathbf{X} = [\dot{x}(t_1), \dot{x}(t_2), \dots, \dot{x}(t_w)]^T,$$

which can be calculated from time series. Finally, one obtains the following equation in the form  $\mathbf{X} = \mathcal{G} \cdot \mathbf{a}_1$ :

$$\begin{pmatrix} \dot{x}(t_1) \\ \dot{x}(t_2) \\ \vdots \\ \dot{x}(t_w) \end{pmatrix} = \begin{pmatrix} \mathbf{g}(t_1) \\ \mathbf{g}(t_2) \\ \vdots \\ \mathbf{g}(t_w) \end{pmatrix} (\mathbf{a}_1). \quad (14)$$



To ensure the restricted isometry property [113, 114], one normalizes  $\mathcal{G}$  by dividing elements in each column by the  $L_2$  norm of that column:  $(\mathcal{G}')_{ij} = (\mathcal{G})_{ij}/L_2(j)$  with  $L_2(j) = \sqrt{\sum_{i=1}^M [(\mathcal{G})_{ij}]^2}$ , so that  $\mathbf{X} = \mathcal{G}' \cdot \mathbf{a}'_1$ . After the normalization,  $\mathbf{a}'_1 = \mathbf{a}_1 \cdot L_2$  can be determined via some standard compressive-sensing algorithm [113, 114]. As a result, the coefficients  $\mathbf{a}_1$  are given by  $\mathbf{a}'_1/L_2$ . To determine the set of power-series coefficients corresponding to a different dynamical variable, say  $y$ , one simply replaces the measurement vector by  $\mathbf{X} = [\dot{y}(t_1), \dot{y}(t_2), \dots, \dot{y}(t_w)]^T$  and use the same matrix  $\mathcal{G}$ . This way all coefficients  $\mathbf{a}_1$ ,  $\mathbf{a}_2$ , and  $\mathbf{a}_3$  of three dimensions can be estimated.

## 2.3 Reconstruction and identification of chaotic systems

The problem of predicting dynamical systems based on time series has been outstanding in nonlinear dynamics because, despite previous efforts [44] in exploiting the delay-coordinate embedding method [43, 46] to decode the topological properties of the underlying system, how to accurately infer the underlying *nonlinear system equations* remains largely an unsolved problem. In principle, a nonlinear system can be approximated by a large collection of linear equations in different regions of the phase space, which can indeed be accomplished through reconstructing the Jacobian matrices on a proper grid that covers the phase-space region of interest [101, 105]. However, the accuracy and robustness of the procedure are challenging issues, including the difficulty with the required computations. For example, in order to be able to predict catastrophic bifurcations, local reconstruction of a large set of linearized dynamics is not sufficient but rather, an accurate prediction of the underlying nonlinear equations themselves is needed.

In 2011 it was proposed [23] that compressive sensing provides a powerful method for data based nonlinear systems identification, based on the principle that it is possible to fully reconstruct dynamical systems from time series because the dynamics of many natural and man-made systems are determined by functions that can be approximated by series expansions in a suitable base. The task is then to estimate the coefficients in the series representation. In general, the number of coefficients to be estimated can be large but many of them may be zero (the sparsity condition). According to the conventional wisdom this would be a difficult problem as a large amount of data is required and the computations involved can be extremely demanding. However, compressive sensing [113–118]. provides a viable solution to the problem, where the basic principle is to reconstruct a sparse signal from small amount of observations, as measured by linear projections of the original signal on a few predetermined vectors.

In Ref. [23], a number of classic chaotic systems were used to demonstrate the compressive sensing based approach. Here we quote one example, the Hénon map [153], a classical model that has been used to address many fundamental issues in chaotic dynamics. The prediction of the map equations resembles that of a vector field. The map is given by:  $(x_{n+1}, y_{n+1}) = (1 - ax_n^2 + y_n, bx_n)$ , where  $a$  and  $b$  are parameters. For  $b = 0.3$ , the map exhibits periodic and chaotic attractors for  $a < a_c \approx 1.42625$ , where  $a_c$  is the critical parameter value for a boundary crisis [154–156], above which almost all trajectories diverge. Assuming power-series expansions up to order 3 in the map equations, the authors were able to identify the map coefficients with high accuracy using only a few data points. Figure 1 shows the distributions of the estimated power-series coefficients, where extremely narrow peaks about zero indicate that a large number of the coefficients are effectively zero, which correspond to nonexistent terms in the map equations. Coefficients that are not in-

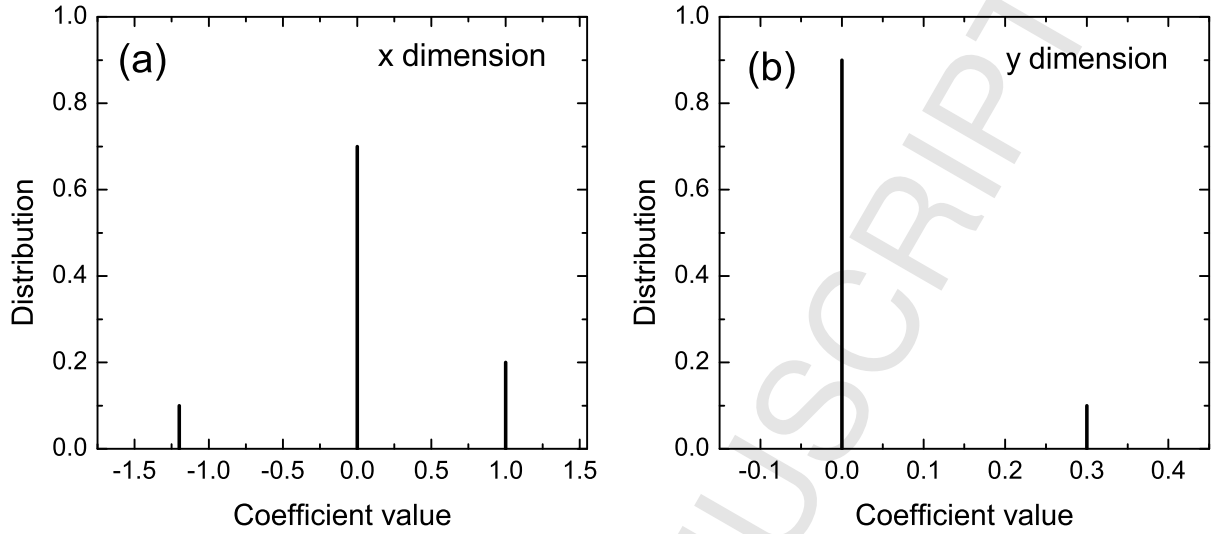
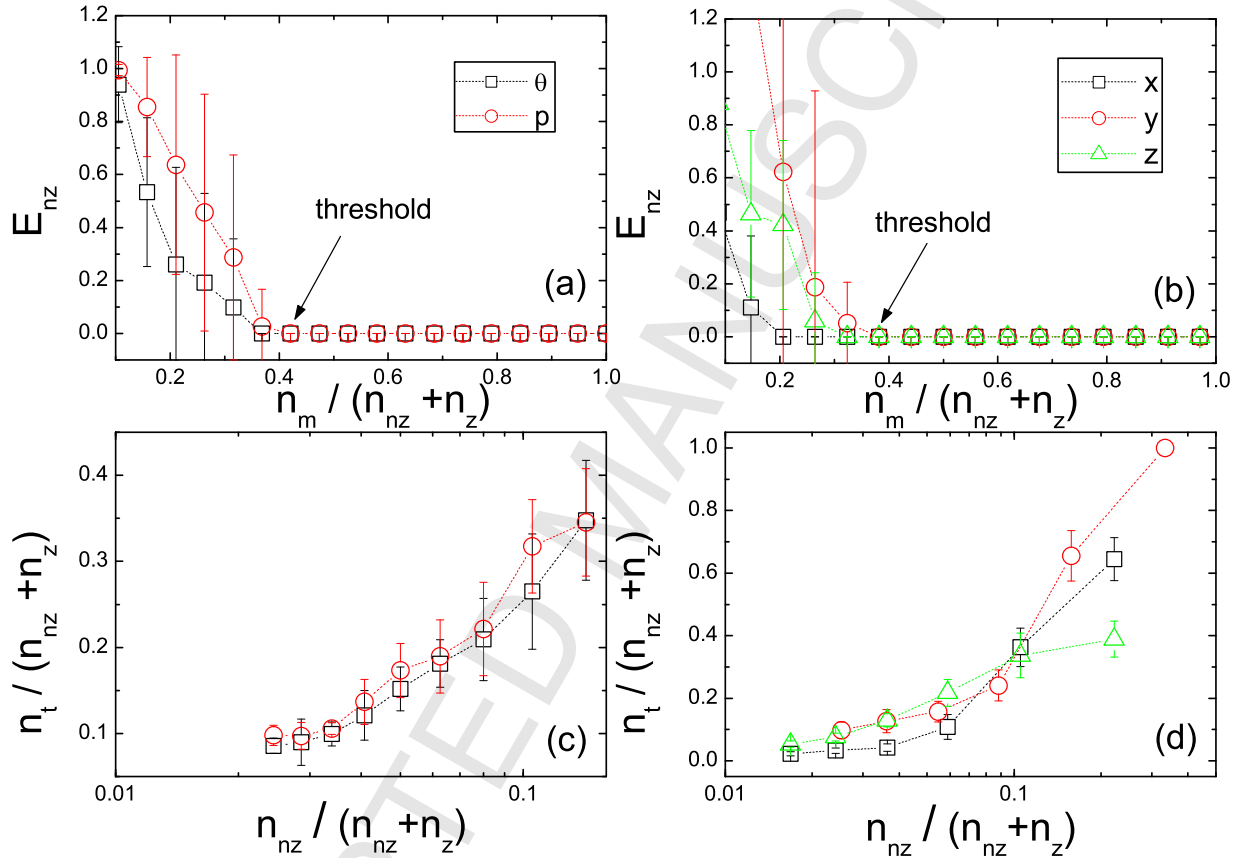


Figure 1: **An example of chaotic systems identification based on compressive sensing.** For the Hénon map, in (a)  $x$  dimension and (b)  $y$  dimension, distributions of the predicted values of ten power-series coefficients up to order 3: constant,  $y$ ,  $y^2$ ,  $y^3$ ,  $x$ ,  $xy$ ,  $xy^2$ ,  $x^2$ ,  $x^2y$  and  $x^3$ . From Ref. [23] with permission.

cluded in the zero peak correspond then to the existent terms and they determine the predicted map equations. Note that, to predict correctly the map equations, the number of required data points is extremely low. Similar results were obtained [23] for the classic standard map [157, 158], the chaotic Lorenz system [159], and the chaotic Rössler oscillator [160].

To quantify the performance of the compressive sensing based systems identification method with respect to the amount of required data, the prediction errors were calculated [23], which are defined separately for nonzero (existing) and zero terms in the dynamical equations. The relative error of a nonzero term is defined as the ratio to the true value of the absolute difference between the predicted and true values. The average over the errors of all terms in a component is the prediction error  $E_{nz}$  of nonzero terms for the component. In contrast, the absolute error  $E_z$  is used for zero terms. Figures 2(a) and 2(b) show  $E_{nz}$  as a function of the ratio of the number  $n_m$  of measurements to the total number  $n_{nz} + n_z$  of terms to be predicted, for the standard map and the Lorenz system, respectively. Note that, for the standard map, it is necessary to explore alternative bases of expansion so that the sparsity condition can be satisfied. A practical strategy is that, assuming that a rough idea about the basic physics of the underlying dynamical system is available, one can choose a base that is compatible with the knowledge. In the case of the standard map, a base including the trigonometric functions can be chosen. The results in Figs. 2(a) and 2(b) indicate that, when the number  $n_m$  of measurements exceeds a threshold  $n_t$ ,  $E_{nz}$  becomes effectively zero. For convenience, one can define  $n_t$  by using the small threshold value  $E_{nz} = 10^{-3}$  so that  $n_t$  is the minimum number of required measurements for an accurate prediction. Figures 2(a) and 2(b) show that  $n_t$  is much less than  $n_{nz} + n_z$  if  $n_{nz}$ , the number of nonzero terms, is small. The performance of the method can thus be quantified by the threshold with respect to the numbers of measurements and terms to be predicted. As shown in Figs. 2(c) and 2(d) for the standard map and the Lorenz system, respectively, as the nonzero terms become sparser among all terms to be predicted (characterized by a decrease in  $n_{nz}/(n_{nz} + n_z)$  when  $n_{nz} + n_z$  is increased), the ratio of



**Figure 2: Performance characterization of compressive sensing based chaotic systems identification.** Prediction errors  $E_{nz}$  in dynamical equations as a function of the ratio of the number  $n_m$  of measurements to the total number  $n_{nz} + n_z$  of terms to be predicted for (a) the standard map and (b) the Lorenz system. The ratio of the threshold  $n_t$  to  $n_{nz} + n_z$  for different equations as a function of the ratio  $n_{nz} / (n_{nz} + n_z)$  for (c) the standard map and (d) the Lorenz system. In (a) and (b),  $n_{nz} + n_z$  is 20 and 35, respectively. The error bars represent the standard deviations obtained from 30 independent realizations. In (c) and (d),  $n_{nz} + n_z$  can be adjusted by the order of the power series. In (c), the points range from order 3 to order 11, and in (d) from order 2 to order 7, and  $E_{nz}$  and  $E_z$  exhibit the same threshold. From Ref. [23] with permission.

the threshold  $n_t$  to the total number of terms  $n_{nz} + n_z$  becomes smaller. These results demonstrated the advantage of the compressive-sensing based method to predict dynamical systems, i.e., high accuracy and extremely low required measurements. In general, to predict a nonlinear dynamical system as accurately as possible, many reasonable terms should be assumed in the expansions, insofar as the percentage of nonzero terms is small so that the sparsity condition of compressive sensing is met.

There are situations where the system is high-dimensional or stochastic. A possible solution is to employ the Bayesian inference to determine the system equations. In general the computational challenge associated with the approach can be formidable, but the power-series or more general expansion based compressive-sensing method may present an effective strategy to overcome the difficulty.

## 2.4 Predicting catastrophe in nonlinear dynamical systems

### 2.4.1 Predicting catastrophic bifurcations based on compressive sensing

Nonlinear dynamical systems, in their parameter space, can often exhibit catastrophic bifurcations that ruin the desirable or “normal” state of operation. Consider, for example, the phenomenon of crisis [154] where, as a system parameter is changed, a chaotic attractor collides with its own basin boundary and is suddenly destroyed. After the crisis, the state of the system is completely different from that associated with the attractor before the crisis. Suppose that the state before the crisis is normal and desirable and the state after the crisis is undesirable or destructive. The crisis can thus be regarded as a *catastrophe* that one strives to avoid at all cost. Catastrophic events, of course, can occur in different forms in all kinds of natural and man-made systems. A question of paramount importance is how to predict catastrophes in advance of their possible occurrences. This is especially challenging when the equations of the underlying dynamical system are unknown and one must then rely on measured time series or data to predict any potential catastrophe.

Compressive sensing based nonlinear systems identification provides an approach to forecasting potential catastrophic bifurcations [23]. Assume that an accurate model of the system is not available, i.e., the system equations are unknown, but the time evolutions of the key variables of the system can be accessed through monitoring or measurements. The method [23] thus consists of three steps: (i) predicting the dynamical system based on time series, (ii) identifying the parameters of the system, and (iii) performing a computational bifurcation analysis using the predicted system equations to locate potential catastrophic events in the parameter space so as to determine the likelihood of system’s drifting into a catastrophe regime. In particular, if the system operates at a parameter setting close to such a critical bifurcation, catastrophe is imminent as a small parameter change or a random perturbation can push the system beyond the bifurcation point. Consider the concrete case of crises in nonlinear dynamical systems. Once a complete set of system equations has been predicted and the parameters have been identified, one proceeds to examine the available parameter space. It should be noted that, to explore the multi-parameter space of a dynamical system can be challenging, but this can often lead to the discovery of new phenomena in dynamics. Examples are the phenomenon of double crises in systems with two bifurcation parameters [161] and the hierarchical structures in the parameter space [162, 163].

In Ref. [23], a number of examples were given in which the bifurcation diagrams computed

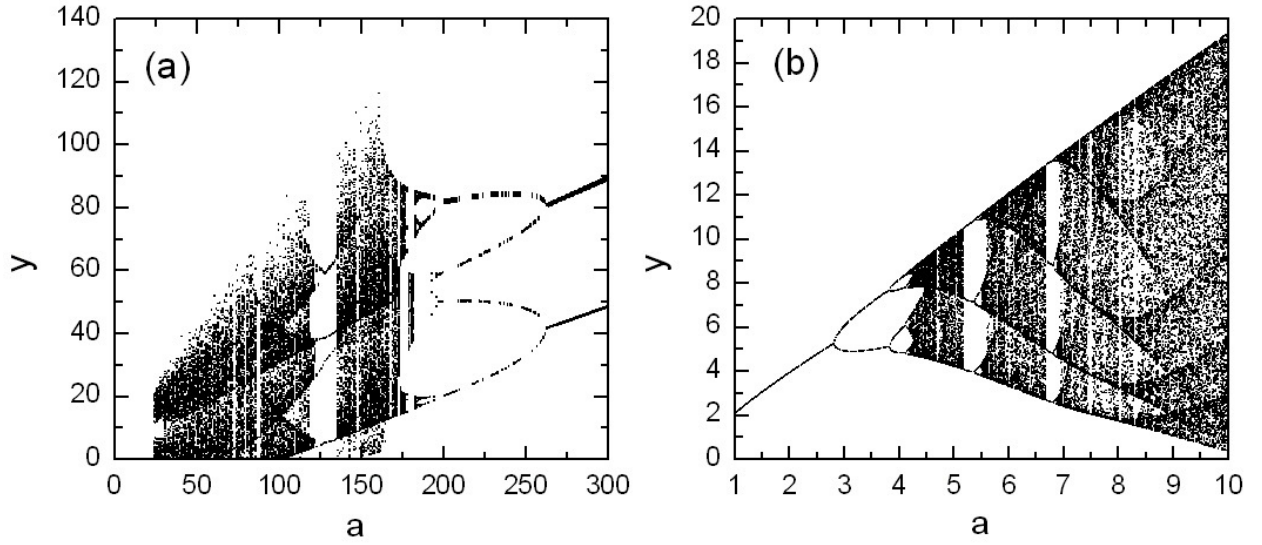
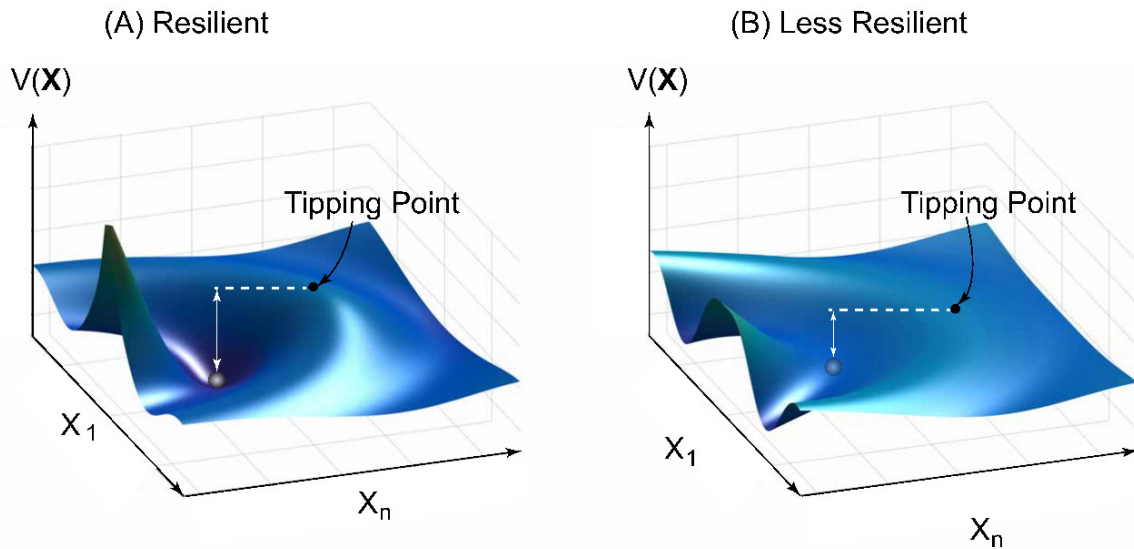


Figure 3: **Bifurcation diagrams predicted by the compressive sensing based method of non-linear systems identification.** (a) A predicted Lorenz system given by  $\dot{x} = 10.000548307148881 \times y - 10.001897147696283 \times x$ ,  $\dot{y} = x(a - 1.000933186801829 \times z) - 1.000092963203845 \times y$ ,  $\dot{z} = 0.999893761636553 \times xy - 2.666477325955504 \times z$ . (b) A predicted Rössler system given by  $\dot{x} = -0.999959701293536 \times y - 0.999978902248041 \times z$ ,  $\dot{y} = 1.000004981649221 \times x + 0.200005996113158 \times y$ ,  $\dot{z} = 0.199997011085648 + 0.999999156496251 \times z(x - a)$ . In both cases,  $n_m = 18$  and  $n_{nz} + n_z = 35$ . From Ref. [23] with permission.

from the predicted system equations agree well with those from the original systems, so all possible critical bifurcation points can be predicted accurately based on time series only. Figures 3(a) and 3(b) illustrate a predicted bifurcation diagram from the chaotic Lorenz and Rössler systems, respectively.

#### 2.4.2 Using compressive sensing to predict tipping points in complex systems

It is increasingly recognized that a variety of complex dynamical systems ranging from ecosystems and the climate to economical, social, and infrastructure systems can exhibit *tipping points* at which irreversible transition from a normal functioning state to a catastrophic state occurs [164, 165]. Examples of such transitions are blackouts in the power grids, sudden emergence of massive jamming in urban traffic systems, the shutdown of the thermohaline circulation in the North Atlantic [166], sudden extinction of species in ecosystems [167–170], and the occasional switches of shallow lakes from clear to turbid waters [171]. In fact, the seemingly abrupt transitions are the consequence of *gradual* changes in the system which can, for example, be caused by slow drifts in the external conditions. To understand the dynamical properties of the system near a tipping point, to predict the tendency for the system to drift toward the tipping point, to issue early warnings, and finally to apply control to reverse or to slow down the trend, are significant but extremely challenging research problems. Compounded with the difficulty is the fact the complex systems are often *interdependent and non-stationary*. For example, the evolution of an ecosystem depends on human behaviors, which in turn affects the well being of the human society. Infrastructure systems such as the power grids and communication networks are interdependent upon each other [172, 173], both



**Figure 4: Relationship between tipping point and resilience.** For a real world complex dynamical system, the potential landscape is time-dependent or *non-stationary*: there are slow but ultimately large drifts in the system parameters and equations leading to catastrophic collapse, and there is fast but small random perturbation. The origins of these drifts and perturbation are discussed in the text.

being affected by human activities (social system). The concept of interdependence is prevalent in many disciplines. At the present there is little understanding of tipping points in interdependent complex systems in terms of their emergence and dynamical properties.

In a dynamical system, the existence of one or several tipping points is intimately related to the concept of *resilience* [174–179], which can be understood heuristically by resorting to the intuitive picture of a ball moving in a valley under gravity, as shown in Fig. 4. To the right of the valley there is a hill, or a potential barrier in terms of classical mechanics. The downhill side to the right of the barrier corresponds to a catastrophic behavior. Normal functioning of the system is represented as the confinement of ball's motion within the valley. If the valley is sufficiently deep (or the height of the barrier is sufficiently large), as shown in Fig. 4(a), there will be little probability for the ball to move across the top of the barrier towards catastrophic behavior, implying that the system is more resilient to random noise or external disturbances. However, if the barrier height is small, as shown in Fig. 4(b), the system is less resilient as small perturbation can push the ball over to the left side of the barrier. The top of the barrier thus corresponds to a tipping point, across which from the left the system will essentially collapse. In mechanics, the situation can be formulated using a potential function so that, mathematically, the motion of the ball can be described by the Hamilton's equations [180]. Given a dynamical system, if the potential landscape can be determined, it will be possible to locate the tipping points. In systems biology, the potential function is called the Waddington landscape [181], which essentially determines the biological paths for cell development and differentiation [182, 183] - the *landscape metaphor*. In the past few years, a quantitative approach to mapping out the potential landscape for gene circuits or gene regulatory networks has been developed [184–187]. In nonlinear dynamical systems, a similar concept exists - *quasipotential* [188–190].

Because of non-stationarity, for a complex system in the real world, Fig. 4 in fact represents

only a “snapshot” of the potential landscape. For complex dynamical systems, the potential landscape must necessarily be *time-dependent or non-stationary*, which completely governs the emergence of tipping points and the global dynamics of the system. There are two types of time-varying disturbances to the landscape: (1) slow but eventually large changes in parameters and system equations and (2) fast but small random perturbation. The physical origins of these disturbances can be argued, as follows.

Take, for example, the challenge of adding distributed renewable energy generation to the power grid. The time scale for changes is years. Over that time period, significant new renewables can be added, and yet the precise timing, location, and amount of distributed renewable energy generation is unpredictable, because it is not possible to know how social decisions (in terms of regulation, business planning, and consumer choice) will play out. (A similar situation occurs for climate change: slow, long-term, secular, and nonlinear changes in climatic averages would occur, which will impact generation via water availability and temperature, etc. At the same time, weather remains a high frequency pattern on top of these slow changes.) Clearly the current grid is largely stable vis-a-vis existing distributions of weather variables, but how will that change in the future? For example, in the future, charging networks may comprise mainly slow charging stations [191] as the widespread use of fast charging stations would raise the power demand in the electrical grid [192], causing tremendous difficulties for the managers to operate the grid. There can also be changing regulatory incentives or management structures. All these can lead to *social non-stationarity*. Mathematically, the social factors can be modeled by adiabatic changes to the system, which affect the potential landscape on time scales that are slower than that of the intrinsic dynamics.

At short time scales, precise hourly patterns of electricity generation may fluctuate due to changes in wind and cloud activity - a type of *environmental non-stationarity*. There can also be *technological non-stationarity* such as shifting demand patterns associated with consumer electronics, plug-in hybrid and electric vehicles, etc. These occur on short time scale, which can be modeled as random perturbation or noise to the system.

Referring to Fig. 4, the topographic landscape metaphor of resilience, one can see that, if the system is stationary, there is a fixed threshold across which the system will collapse. In this case, the system resilience can simply be characterized by the distance to the threshold. However, for a non-stationary system, it is not possible to measure the threshold distance by establishing the absolute positions of the ball (system) and the tipping point and then computing the difference between them. Instead, one must attempt to estimate the differences directly in real time. Two open questions are: is it possible to determine the non-stationary landscape so as to predict the emergence and the locations of the time-dependent tipping points? Can human intervention or control strategies be developed to prevent or significantly slow down the system’s evolution toward a catastrophic tipping point?

At the present there are no answers to these questions of the grand challenge nature, but we wish to argue that compressive sensing based nonlinear systems identification can provide insights into the fundamental issues pertinent to these questions. Consider, for example, a complex power grid system, in which the time series such as the voltage and power at all generator nodes are available, as well as social interaction data that are typically polarized (e.g., binary). There were recent efforts in reconstructing complex networks of nonlinear oscillators based on continuous time series [24] and in uncovering epidemic propagation networks using binary data [37] (to be discussed



in Sec. 3). In principle, these approaches can be combined to deal with nonstationary complex systems. Specifically, large but slowly varying physical non-stationarity can be modeled through the appearance of additional, concrete mathematical terms involving voltage, phase, and current variables, or through the disappearance of certain terms. Social non-stationarity can be modeled by functions of Boolean variables that generate polarized data or can be reconstructed using these data. It would then be possible to establish a mathematical framework combining reconstruction methods for continuous time series and polarized data. Potentially, this could represent an innovative and concrete approach to incorporating social data into a complex physical/technological system and assessing, quantitatively, the effect of social non-stationarity on the dynamical evolution of the system.

## 2.5 Forecasting future states (attractors) of nonlinear dynamical systems

A dynamical system in the physical world is constantly subject to random disturbance or adiabatic perturbation. Broadly speaking, there are two types of perturbations: stochastic or deterministic. Stochastic disturbances (or noise) can typically be described by random processes and they do not alter the intrinsic structure of the underlying equations of the system. Deterministic perturbations, however, can cause the system equations or parameters to vary with time. Suppose the perturbations are adiabatic, i.e.,  $T_i$ , the time scale of the intrinsic dynamics of the system, is much smaller than  $T_e$ , the time scale of the external perturbation. In this case, some “asymptotic states” or “attractors” of the system can still be approximately defined in a time scale that is much larger than  $T_i$  but smaller than  $T_e$ . When the dynamics in such a time interval is examined in a long run, the attractor of the system will depend on time. Often, one is interested in forecasting the “future” asymptotic states of the system. Take the climate system as an example. The system is under random disturbances, but adiabatic perturbations are also present, such as  $\text{CO}_2$  injected into the atmosphere due to human activities, the level of which tends to increase with time. The time scale for appreciable increase in the  $\text{CO}_2$  level to occur (e.g., months or years) is much larger than the intrinsic time scales of the system (e.g., days). The climate system is thus an adiabatically time-varying, nonlinear dynamical system. It is of interest to forecast what the future attractors of the system might be in order to determine whether it will behave as desired or sustainably. The issue of sustainability is, of course, critical to many other natural and man-made systems as well. To be able to forecast the future states of such systems is essential to assessing their sustainability.

It was demonstrated that compressive sensing can be exploited for predicting the future states (attractors) of adiabatically time varying dynamical systems [25]. The general problem statement is: given a nonlinear dynamical system whose equations or parameters vary adiabatically with time, but otherwise are completely unknown, can one predict, based solely on measured time series, the future asymptotic attractors of the system? To be concrete, consider the following dynamical system:

$$d\mathbf{x}/dt = \mathbf{F}[\mathbf{x}, \mathbf{p}(t)], \quad (15)$$

where  $\mathbf{x}$  is the dynamical variable of the system in the  $d$ -dimensional phase space and  $\mathbf{p}(t) = [p_1(t), \dots, p_K(t)]$  denotes  $K$  independent, time-varying parameters of the system. Assume that both the form of  $\mathbf{F}$  and  $\mathbf{p}(t)$  are unknown but at time  $t_M$ , the end of the time interval during which measurements are taken, the time series  $\mathbf{x}(t)$  for  $t_M - T_M \leq t \leq t_M$  are available, where  $T_M$  denotes the measurement time window. The idea was to predict, using compressive sensing, the precise



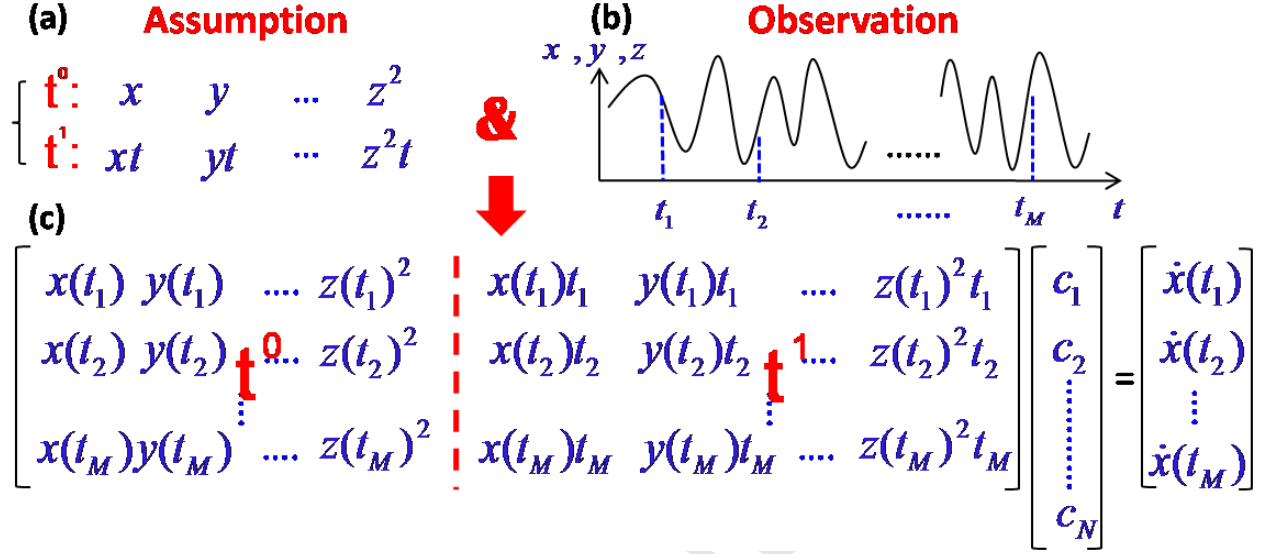


Figure 5: **Illustration of compressive sensing based scheme for predicting future attractors:** the problem of identifying time-varying nonlinear dynamical systems can be solved by treating time as an independent expansion variable and invoking compressive sensing. From Ref. [25] with permission.

mathematical forms of  $\mathbf{F}$  and  $\mathbf{p}(t)$  based on the available time series at  $t_M$  so that the evolution and the likely attractors of the system for  $t > t_M$  can be computationally assessed and anticipated [25]. The predicted forms of  $\mathbf{F}$  and  $\mathbf{p}(t)$  at time  $t_M$  would contain errors that in general will increase with time. In addition, for  $t > t_M$  new perturbations can occur to the system so that the forms of  $\mathbf{F}$  and  $\mathbf{p}(t)$  may be further changed. It is thus necessary to execute the prediction algorithm frequently using time series available at the time. In particular, the system could be monitored at all times so that time series can be collected, and predictions should be carried out at  $t_i$ 's, where  $\dots > t_i > \dots > t_{M+2} > t_{M+1} > t_M$ . For any  $t_i$ , the prediction algorithm is to be performed based on available time series in a suitable window prior to  $t_i$ .

To formulate the problem of predicting time-varying dynamical systems in the framework of compressive sensing, one can expand all components of the time-dependent vector field  $\mathbf{F}[\mathbf{x}, \mathbf{p}(t)]$  in Eq. (15) into a power series in terms of both dynamical variables  $\mathbf{x}$  and time  $t$ . The  $i$ th component  $\mathbf{F}[\mathbf{x}, \mathbf{p}(t)]_i$  of the vector field can be written as

$$\sum_{l_1, \dots, l_m=1}^n [(\alpha_i)_{l_1, \dots, l_m} x_1^{l_1} \dots x_m^{l_m} \cdot \sum_{w=0}^v (\beta_i)_w t^w] \equiv \sum_{l_1, \dots, l_m=1}^n \sum_{w=0}^v (c_i)_{l_1, \dots, l_m; w} x_1^{l_1} \dots x_m^{l_m} \cdot t^w, \quad (16)$$

where  $x_k$  ( $k = 1, \dots, m$ ) is the  $k$ th component of the dynamical variable and  $c_i$  is the  $i$ th component of the coefficient vector to be determined. Assume that the time evolution of each term can be approximated by the power series expansion in time, i.e.,  $\sum_{w=0}^v (\beta_i)_w t^w$ . The power-series expansion allows us to cast Eq. (15) into the standard form of compressive sensing, Eq. (1). In principle, if every combined scalar coefficient  $(c_i)_{l_1, \dots, l_m; w}$  associated with the corresponding term in Eq. (16) can be determined from time series for  $t \leq t_M$ , the vector field component  $[\mathbf{F}(\mathbf{x}, \mathbf{p}(t))]_i$  becomes known. Repeating the procedure for all components, the entire vector field for  $t > t_M$  can be found.

To explain the compressive sensing based method in an intuitive way, one can consider the special case where the number of components of the dynamical variables is  $m = 3$  ( $x$ ,  $y$ , and  $z$ ),

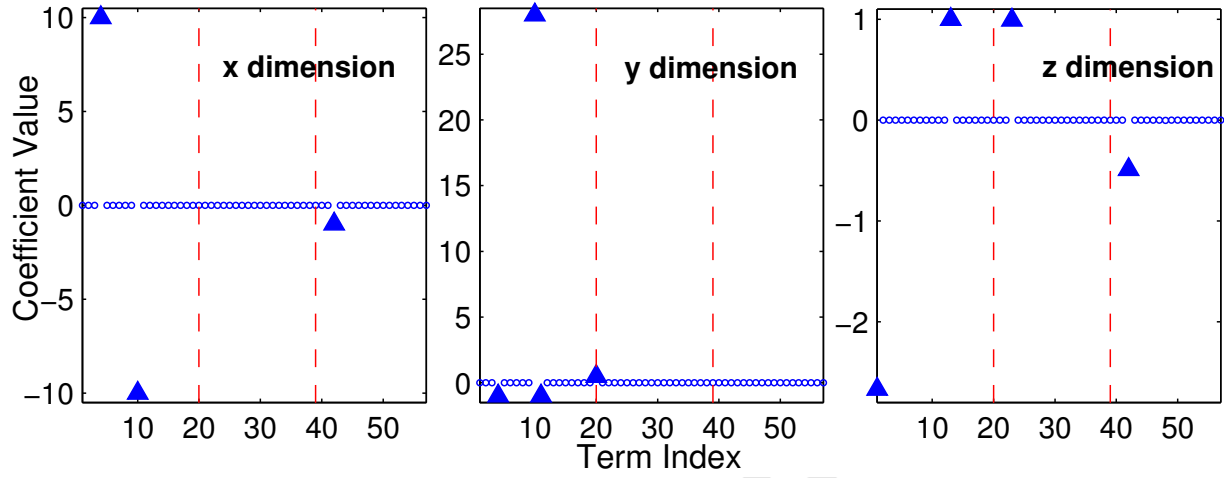


Figure 6: **An example of predicting time-varying chaotic systems.** For the time-varying Lorenz chaotic systems Eq. (17) with  $k_1(t) = -t^2$ ,  $k_2(t) = 0.5t$ ,  $k_3(t) = t$ , and  $k_4(t) = -0.5t^2$ , predicted values of coefficients of power-series terms versus the term index for the  $x$ -,  $y$ -, and  $z$ -equations, where solid triangles and open circles denote nonzero and zero coefficients, respectively. Note that, the number  $M$  of data points used for prediction is about 50% of the total number  $N$  of unknown coefficients in each power-series expansion. From Ref. [25] with permission.

the order of the power series is  $l_1 + l_2 + l_3 \leq 2$ , and the maximum power of time  $t$  in Eq. (16) is  $v = 1$ , i.e., only the  $t^0$  and  $t^1$  terms are included. Focusing on one dynamical variable, say  $x$ , the total number of terms in the power-series expansion is 20, as specified in Fig. 5(a). Let the measurements  $x(t)$ ,  $y(t)$ , and  $z(t)$  be taken at times  $t_1, t_2, \dots, t_M$ , as shown in Fig. 5(b). The values of all 20 power-series terms at these time instants can then be obtained, as shown in Fig. 5(c), where all the terms are divided into two blocks according to the distinct powers of the time variable  $t$ :  $t^0$  and  $t^1$ . The projection matrix  $\mathcal{G}$  in Eq. (1) thus consists of these two blocks. (In the general case where higher powers of the time variable is involved,  $\mathcal{G}$  would contain a corresponding number of blocks.) The components of vector  $\mathbf{X}$  in Eq. (1) are the first derivatives  $dx/dt$  evaluated at  $t_1, t_2, \dots, t_M$ , which can also be approximated by the measured time series  $x(t)$  at these times. As shown in Fig. 5(c), Eq. (16) for this simple example can be written in the form of Eq. (1). To ensure sparsity, one can assume many terms in the power-series expansion up to some high order  $n$  so that the total number of terms in Eq. (16),  $N$ , will be quite large. As a result,  $\mathbf{a}$  is high-dimensional but most of its components are zero. The number of measurements,  $M$ , needs not be as large as  $N$ :  $M \ll N$ . Another requirement of compressive sensing is the restricted isometric property that can be guaranteed by normalizing the matrix  $\mathcal{G}$  and by using linear-programming based signal-recovery algorithms [113–118]. To determine the set of power-series coefficients corresponding to a different dynamical variable, say  $y$ , one simply replaces the measurement vector by  $\mathbf{X} = [\dot{y}(t_1), \dot{y}(t_2), \dots, \dot{y}(t_M)]^T$ . The matrix  $\mathcal{G}$ , however, remains the same. The problem of forecasting *time-varying* nonlinear dynamical systems then fits perfectly into the compressive-sensing paradigm.

As a proof of principle, the authors of Ref. [25] used the classical Lorenz chaotic system [159] as an example by incorporating explicit time dependence in a number of additional terms. The

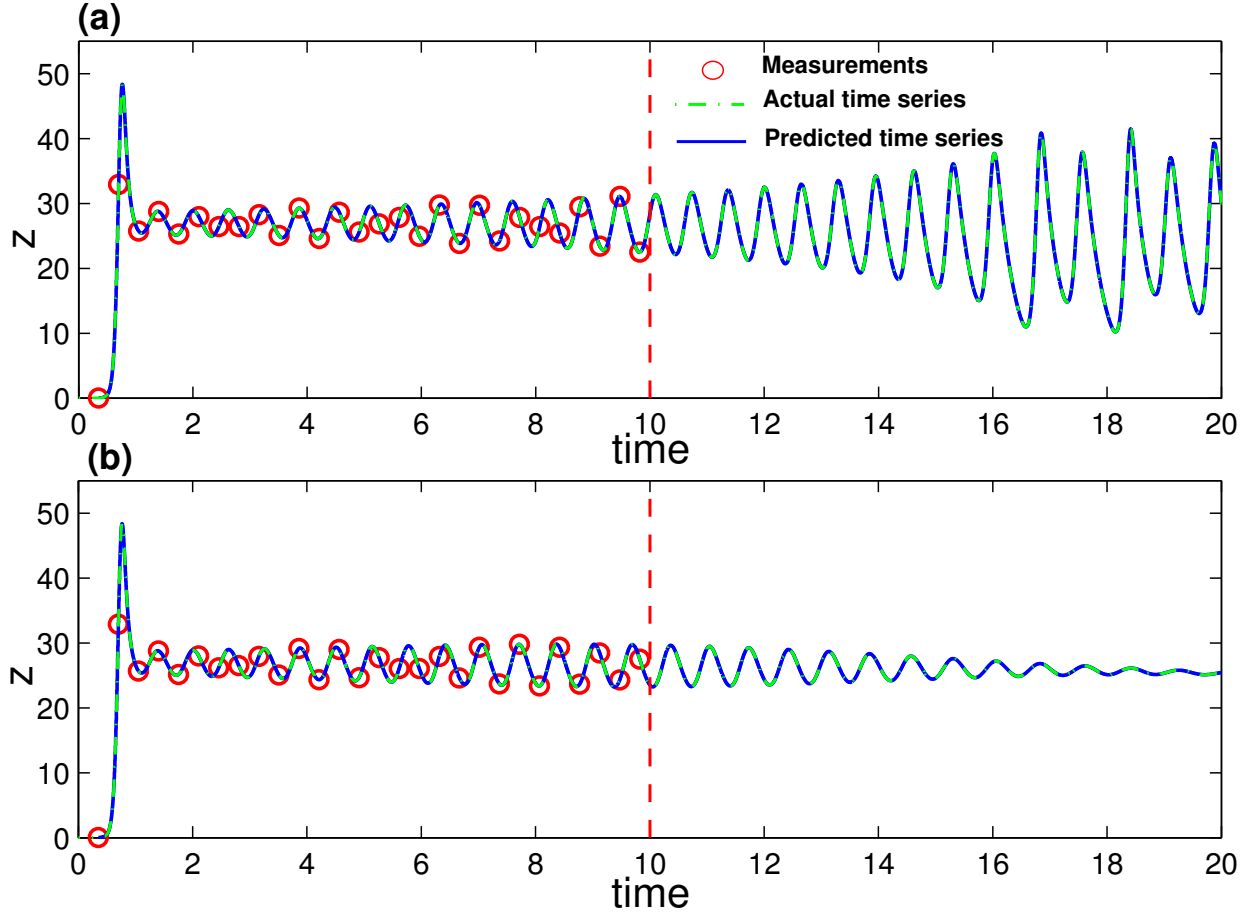
modified Lorenz system is given by

$$\begin{aligned}\dot{x} &= -10(x - y) + k_1(t)y, \\ \dot{y} &= 28x - y - xz + k_2(t)z, \\ \dot{z} &= -(8/3)z + xy + [k_3(t) + k_4(t)]y,\end{aligned}\tag{17}$$

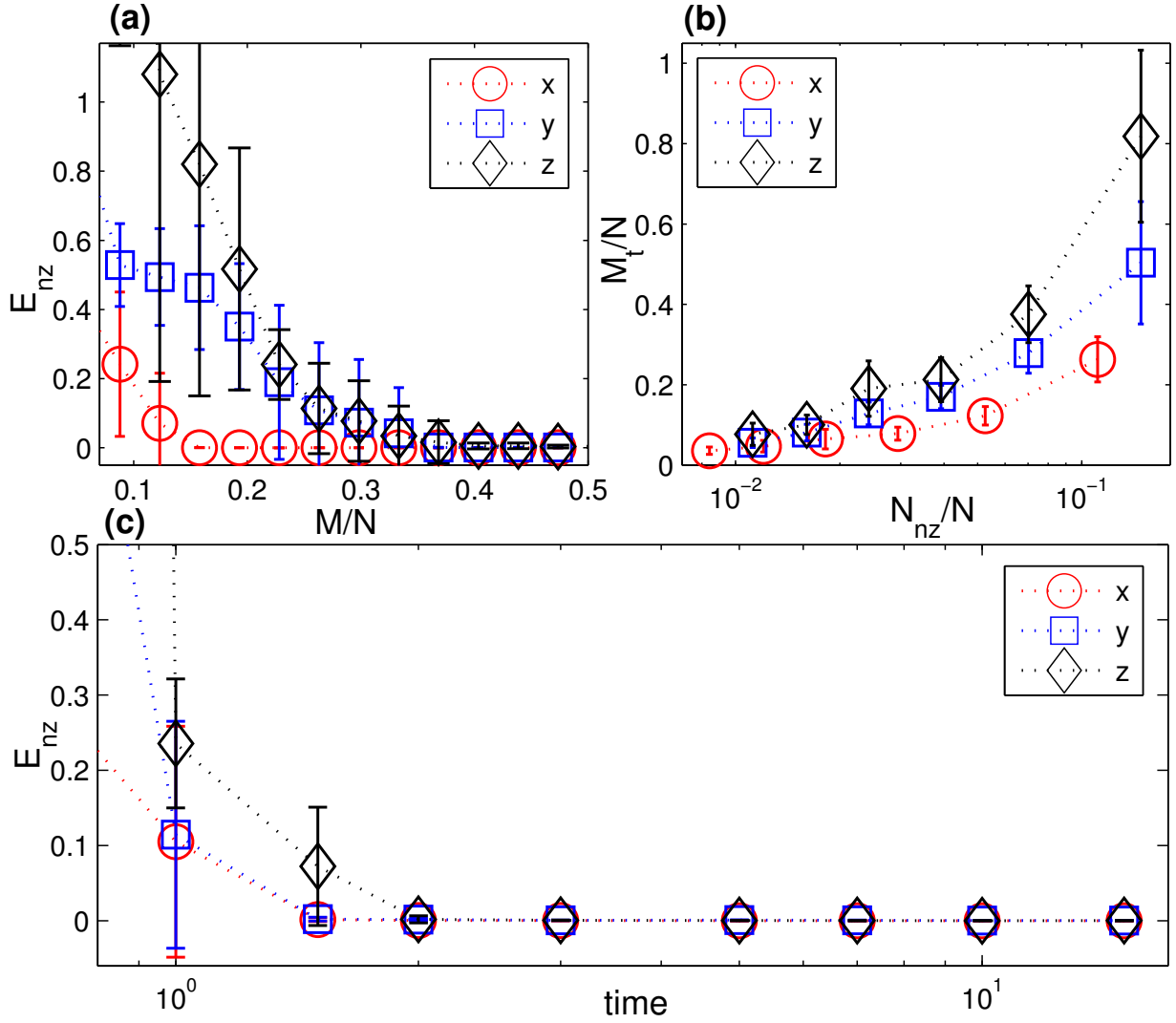
where  $k_1(t) = -t^2$ ,  $k_2(t) = 0.5t$ ,  $k_3(t) = t$ , and  $k_4(t) = -0.5t^2$ . Suppose that the system equations are unknown but only measured time series  $x(t)$ ,  $y(t)$ , and  $z(t)$  in a finite time interval are available. The number of dynamical variables is  $m = 3$  and we choose the orders of the power-series expansions in the three variables according to  $l_1 + l_2 + l_3 \leq 3$ . The maximum power in the time dependence is chosen to be  $\nu = 2$  so that explicit time-dependent terms  $t^0$ ,  $t^1$  and  $t^2$  are considered. The total number of coefficients to be predicted is then  $(\nu + 1) \sum_{i=1}^3 (i + 1)(i + 2)/2 = 57$ . (Note that, using low-order power-series expansions in both the dynamical variables and time is solely for facilitating explanation and presentation of results, while the forecasting principle is the same for realistic dynamical systems where much higher orders may be needed.) Figure 6 shows the predicted coefficient values versus the term index for all three dynamical variables, where in each panel, solid triangles and open circles denote the predicted non-zero and zero coefficients, respectively, and the red dashed dividing lines indicate the terms associated with different powers of the time variable, i.e.,  $t^0$ ,  $t^1$  and  $t^2$  (from left to right). The meaning of these results can be explained by using any one of the dynamical variables. For example, for the  $x$ -component of the vector field, the prediction algorithm gives only 3 nonzero coefficients. By identifying the corresponding values of the term index, one can see that they correspond to the two terms without explicit time dependence:  $y$ ,  $x$ , and the term that contains explicit such dependence:  $t^2y$ , respectively. A comparison of the predicted nonzero coefficient values with the actual ones in the original Eq. (17) indicates that the method works remarkably well. Similar results were obtained for the  $y$  and  $z$  components of the vector field.

When the vector field  $\mathbf{F}[\mathbf{x}, \mathbf{p}(t)]$  of the underlying dynamical system has been predicted, one can solve Eq. (15) numerically to assess the state variables at any future time and the asymptotic attractors. Figures 7(a,b) present one example, where a forecasted time series calculated from the predicted vector field is shown, together with the values of the corresponding dynamical variable from the actual Lorenz system at a number of time instants. The two cases shown are where the parameter functions  $k_i(t)$  ( $i = 1, \dots, 4$ ) are all zero and time-varying, respectively. Excellent agreement was again obtained, indicating the power of the method to predict the future states and attractors of time-varying dynamical systems. The interpretation and implication of Figs. 7(a,b) are the following. Note that  $t = 0$  and  $t = 10$  correspond to the beginning and the end of the measurement time window  $[t_1, t_M]$ , respectively. For the original classical Lorenz system without time-varying parameters, the asymptotic attractor is chaotic, as can be seen from Fig. 7(a). However, as external perturbations are turned on at  $t = 0$ , there are four time-varying parameters in the system for  $t > 0$ . In this case, the asymptotic attractor becomes a fixed-point, as can be seen from the asymptotic behavior  $z \rightarrow \text{constant}$  in Fig. 7(b). In both cases, by using limited amount of measurements, namely, available time series in the window  $[t_1, t_M]$ , one obtains quite accurate forecasting results. The result exemplified in Fig. 7(b) is especially significant, as it indicates that the future state and attractors of time-varying dynamical systems can be accurately predicted based on limited data availability.

An error and performance analysis was carried out in Ref. [25] by using the indicators  $E_{nz}$  and



**Figure 7: Predicting future attractors of time varying dynamical systems.** For the time-varying Lorenz system Eq. (17), predicted time series and measured values of the dynamical variable  $z(t)$  for (a) time-independent case where  $k_i(t) = 0$  ( $i = 1, \dots, 4$ ) and (b) time-dependent case where  $k_1(t) = -0.01t^2$ ,  $k_2(t) = 0.01t$ ,  $k_3(t) = 0.01t$ , and  $k_4(t) = -0.01t^2$ . Red circles denote the measurements used for prediction, while the green dash and blue solid lines represent the actual and predicted time series, respectively. In both panels,  $t = 0$  and  $t = 10$  correspond to the beginning and the end of the measurement window, i.e.,  $t_1$  and  $t_M$ , respectively. From Ref. [25] with permission.



**Figure 8: Performance analysis of compressive sensing based identification of time varying nonlinear dynamical systems.** For the time-varying Lorenz system as in Fig. 6, (a) prediction errors  $E_{nz}$  as a function of the ratio  $M/N$ , and (b) ratio  $M_t/N$  as a function of the ratio  $N_{nz}/N$ , where  $N$  can be increased by using higher-order power-series expansions (e.g., up to 7). (c) With fixed number of measurement  $M$ , prediction errors  $E_{nz}$  as a function of the length of measurement window. The error bars are obtained from 20 independent realizations of the prediction algorithm. The prediction errors  $E_z$  associated with non-existent terms show similar behaviors. From Ref. [25] with permission.

$E_z$ , the prediction error for existent and non-existent terms, respectively. Figure 8(a) shows, for the time-varying Lorenz system,  $E_{nz}$  versus the ratio of the number  $M$  of measurements to the total number  $N$  of terms to be predicted. For all dynamical variables, one observes that, as  $M$  exceeds a threshold value  $M_t$ ,  $E_{nz}$  becomes effectively zero, where  $M_t$  can be defined quite arbitrarily, e.g., the minimum number of measurements required to achieve  $E_{nz} = 10^{-3}$ . The data requirement for accurate prediction can then be assessed by examining how  $M_t$  depends on the sparsity of the coefficient vector to be predicted, which can be defined as the ratio of the number  $N_{nz}$  of the nonzero terms to the total number  $N$  of terms to be predicted. Note that,  $N$  or the ratio  $N_{nz}/N$  can be adjusted by varying the order of the assumed power series. From Fig. 8(b), one can see that, as  $N_{nz}/N$  is decreased (e.g., by increasing  $N$ ) so that the vector to be predicted becomes more sparse, the ratio  $M_t/N$  also decreases. In particular, for the smallest value of  $N_{nz}/N$  examined, where  $N = 357$ , only about 5% of the data points are needed for accurate prediction, despite the time-varying nature of the underlying dynamical system. Figure 8(c) shows the prediction errors with respect to different length of the measurement window for a fixed number of data points. It can be seen that, when the length exceeds a certain (small) value so that the time series extends to the whole attractor in the phase space,  $E_{nz}$  approaches zero rapidly.

Dynamical systems are often driven by time-periodic forces, such as the classical Duffing system [193]. In such a case, it is necessary to explore alternative bases of expansion with respect to the time variable other than power series to ensure the sparsity condition. A realistic strategy to choose a suitable expansion base is to make use of the basic physics underlying the dynamical system of interest. Insofar as an appropriate base can be chosen so that the coefficient vector to be predicted is sparse, the compressive sensing based methodology is applicable.

### 3 Compressive sensing based reconstruction of complex networked systems

Compressive sensing has recently been introduced to the field of network reconstruction for continuous time coupled oscillator networks [23, 24], for evolutionary game dynamics on networks [22], for detecting hidden nodes [31, 36], for predicting and controlling synchronization dynamics [30, 194], for reconstructing spreading dynamics based on binary data [37], and for reconstructing complex geospatial networks through estimating the time delays [38]. In this Section we shall the methodologies and the main results.

#### 3.1 Reconstruction of coupled oscillator networks

We describe a compressive sensing based framework that enables a full reconstruction of coupled oscillator networks whose vector fields consist of a limited number of terms in some suitable base of expansion [24]. There are two facts that justify the use of compressive sensing. First, complex networks in the real world are typically sparse [39, 42, 195]. Second, the mathematical functions determining the dynamical couplings in a physical network can be expressed by power-series expansions. The task is then to estimate all the nonzero coefficients. Since the underlying coupling functions are unknown, the power series can contain high-order terms. The number of coefficients to be estimated can therefore be quite large. However, the number of nonzero coefficients may be only a few so that the vector of coefficients is effectively sparse. Because the network structure as well as the dynamical and coupling functions are sparse, compressive sensing stands out as a feasible framework for full reconstruction of the network topology and dynamics.

A complex oscillator network can be viewed as a high-dimensional dynamical system that generates oscillatory time series at various nodes. The dynamics of a node can be written as

$$\dot{\mathbf{x}}_i = \mathbf{F}_i(\mathbf{x}_i) + \sum_{j=1, j \neq i}^N C_{ij} \cdot (\mathbf{x}_j - \mathbf{x}_i), \quad (i = 1, \dots, N), \quad (18)$$

where  $\mathbf{x}_i \in R^m$  represents the set of externally accessible dynamical variables of node  $i$ ,  $N$  is the number of accessible nodes, and  $C_{ij}$  is the coupling matrix between the dynamical variables at nodes  $i$  and  $j$  denoted by

$$C_{ij} = \begin{pmatrix} c_{ij}^{1,1} & c_{ij}^{1,2} & \dots & c_{ij}^{1,m} \\ c_{ij}^{2,1} & c_{ij}^{2,2} & \dots & c_{ij}^{2,m} \\ \dots & \dots & \dots & \dots \\ c_{ij}^{m,1} & c_{ij}^{m,2} & \dots & c_{ij}^{m,m} \end{pmatrix}. \quad (19)$$

In  $C_{ij}$ , the superscripts  $kl$  ( $k, l = 1, 2, \dots, m$ ) stand for the coupling from the  $k$ th component of the dynamical variable at node  $i$  to the  $l$ th component of the dynamical variable at node  $j$ . For any two nodes, the number of possible coupling terms is  $m^2$ . If there is at least one nonzero element in the matrix  $C_{ij}$ , nodes  $i$  and  $j$  are coupled and, as a result, there is a link (or an edge) between them in the network. Generally, more than one element in  $C_{ij}$  can be nonzero. Likewise, if all the elements of  $C_{ij}$  are zero, there is no coupling between nodes  $i$  and  $j$ . The connecting topology and

the interaction strengths among various nodes of the network can be predicted if we can reconstruct all the coupling matrices  $C_{ij}$  from time-series measurements.

Generally, the compressive sensing based method consists of the following two steps. First, one rewrites Eq. (18) as

$$\dot{\mathbf{x}}_i = [\mathbf{F}_i(\mathbf{x}_i) - \sum_{j=1, j \neq i}^N C_{ij} \cdot \mathbf{x}_i] + \sum_{j=1, j \neq i}^N C_{ij} \cdot \mathbf{x}_j, \quad (20)$$

where the first term in the right-hand side is exclusively a function of  $\mathbf{x}_i$ , while the second term is a function of variables of other nodes (couplings). We define the first term to be  $\mathbf{f}_i(\mathbf{x}_i)$ , which is unknown. In general, the  $k$ th component of  $\mathbf{f}_i(\mathbf{x}_i)$  can be represented by a power series of order up to  $n$ :

$$[\mathbf{f}_i(\mathbf{x}_i)]_k \equiv \left[ \mathbf{F}_i(\mathbf{x}_i) - \sum_{j=1, j \neq i}^N C_{ij} \cdot \mathbf{x}_i \right]_k = \sum_{l_1=0}^n \sum_{l_2=0}^n \cdots \sum_{l_m=0}^n [(\alpha_i)_k]_{l_1, \dots, l_m} [(\mathbf{x}_i)_1]^{l_1} [(\mathbf{x}_i)_2]^{l_2} \cdots [(\mathbf{x}_i)_m]^{l_m}, \quad (21)$$

where  $(\mathbf{x}_i)_k$  ( $k = 1, \dots, m$ ) is the  $k$ th component of the dynamical variable at node  $i$ , the total number of products is  $(1+n)^m$ , and  $[(\alpha_i)_k]_{l_1, \dots, l_m} \in \mathbb{R}^m$  is the coefficient scalar of each product term, which is to be determined from measurements as well. Note that the terms in Eq. (21) are all possible products of different components with different power of exponents. As an example, for  $m = 2$  (the components are  $x$  and  $y$ ) and  $n = 2$ , the power series expansion is  $\alpha_{0,0} + \alpha_{1,0}x + \alpha_{0,1}y + \alpha_{2,0}x^2 + \alpha_{0,2}y^2 + \alpha_{1,1}xy + \alpha_{2,1}x^2y + \alpha_{1,2}xy^2 + \alpha_{2,2}x^2y^2$ .

Second, one rewrites Eq. (20) as

$$\dot{\mathbf{x}}_i = \mathbf{f}_i(\mathbf{x}_i) + C_{i1} \cdot \mathbf{x}_1 + C_{i2} \cdot \mathbf{x}_2 + \cdots + C_{iN} \cdot \mathbf{x}_N \quad (22)$$

The goal is to estimate the various coupling matrices  $C_{ij}$  ( $j = 1, \dots, i-1, i+1, \dots, N$ ) and the coefficients of  $\mathbf{f}_i(\mathbf{x}_i)$  from sparse time-series measurements. According to the compressive sensing theory, to reconstruct the coefficients of Eq. (22) from a small number of measurements, most coefficients should be zero - the sparse signal requirement. To include as many coupling forms as possible, one expands each term  $C_{ij}\mathbf{x}_j$  in Eq. (22) as a power series in the same form of  $\mathbf{f}_i(\mathbf{x}_i)$  but with different coefficients:

$$\dot{\mathbf{x}}_i = \mathbf{f}_1(\mathbf{x}_1) + \mathbf{f}_2(\mathbf{x}_2) + \cdots + \mathbf{f}_N(\mathbf{x}_N). \quad (23)$$

This setting not only includes many possible coupling forms but also ensures that the sparsity condition is satisfied so that the prediction problem can be formulated in the compressive-sensing framework. For an arbitrary node  $i$ , information about node-to-node coupling, or about the network connectivity, is contained completely in  $\mathbf{f}_j$  ( $j \neq i$ ). For example, if in the equation of  $i$ , a term in  $\mathbf{f}_j$  ( $j \neq i$ ) is not zero, there then exists coupling between  $i$  and  $j$  with the strength given by the coefficient of the term. Subtracting the coupling terms  $-\sum_{j=1, j \neq i}^N C_{ij} \cdot \mathbf{x}_j$  from  $\mathbf{f}_i$  in Eq. (21), which is the sum of coupling coefficients of all  $\mathbf{f}_j$  ( $j \neq i$ ), the nodal dynamics  $\mathbf{F}_i(\mathbf{x}_i)$  can be obtained. That is, once the coefficients of Eq. (23) have been determined, the node dynamics and couplings among the nodes are all known.

The formulation of the method can be understood in a more detailed and concrete manner by focusing on one component of the dynamical variable at all nodes in the network, say component 1.



(Procedures for other components are similar.) For each node, one first expands the corresponding component of the vector field into a power series up to power  $n$ . For a given node, due to the interaction between this component and other  $(m-1)$  components of the vector field, there are  $(n+1)^m$  terms in the power series. The number of coefficients to be determined for each individual nodal dynamics is thus  $(n+1)^m$ . Now consider a specific node, say node  $i$ . For every other node in the network, possible couplings from node  $i$  indicates the need to estimate another set of  $(n+1)^m$  power-series coefficients in the functions of  $\mathbf{f}_j(\mathbf{x}_j)$ . There are in total  $N(n+1)^m$  coefficients that need to be determined. The vector  $\mathbf{a}$  to be determined in the compressive sensing framework contains then  $N(n+1)^m$  components. For example, to construct the measurement vector  $\mathbf{X}$  and the matrix  $\mathcal{G}$  for the case of  $m=3$  (dynamical variables  $x$ ,  $y$ , and  $z$ ) and  $n=3$ , one obtains the following explicit dynamical equation for the first component of the dynamical variable of node  $i$ :

$$\begin{aligned} \Gamma_i(x_i) = & (a_i)_{000} \cdot x_i^0 y_i^0 z_i^0 + \cdots + (a_i)_{003} \cdot x_i^0 y_i^0 z_i^3 \\ & + (a_i)_{010} \cdot x_i^0 y_i^1 z_i^0 + \cdots + (a_i)_{100} \cdot x_i^1 y_i^0 z_i^0 + \cdots + (a_i)_{333} \cdot x_i^3 y_i^3 z_i^3. \end{aligned} \quad (24)$$

We can denote the coefficients of  $\Gamma_i(x_i)$  by  $\mathbf{a}_i = [(a_i)_{000}, (a_i)_{001}, \dots, (a_i)_{333}]^T$ . Assuming that measurements  $\mathbf{x}_i(t)$  ( $i = 1, \dots, N$ ) at a set of time  $t_1, t_2, \dots, t_M$  are available, one denotes

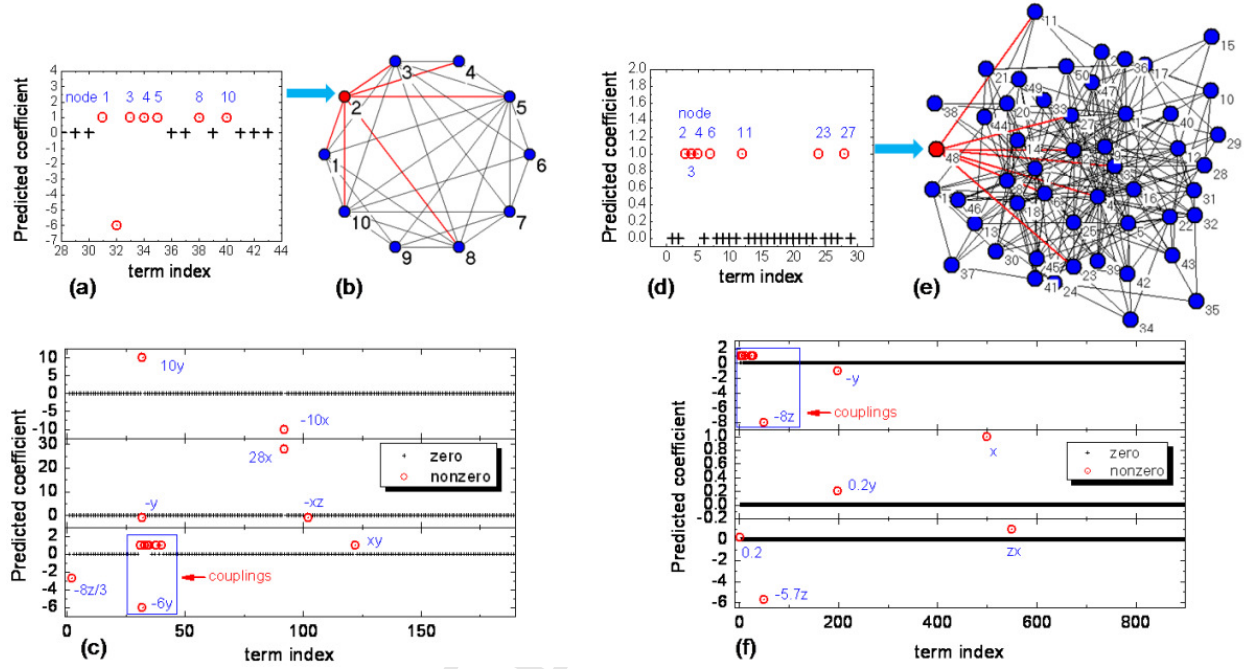
$$\mathbf{g}_i(t) = [x_i(t)^0 y_i(t)^0 z_i(t)^0, x_i(t)^0 y_i(t)^0 z_i(t)^1, \dots, x_i(t)^3 y_i(t)^3 z_i(t)^3], \quad (25)$$

such that  $\Gamma_i[x_i(t)] = \mathbf{g}_i(t) \cdot \mathbf{a}_i$ . According to Eq. (24), the measurement vector can be chosen as  $\mathbf{X} = [\dot{x}_i(t_1), \dot{x}_i(t_2), \dots, \dot{x}_i(t_M)]^T$ , which can be calculated from time series. Finally, one obtains the following equation in the form  $\mathbf{X} = \mathcal{G} \cdot \mathbf{a}$ :

$$\begin{pmatrix} \dot{x}_i(t_1) \\ \dot{x}_i(t_2) \\ \vdots \\ \dot{x}_i(t_M) \end{pmatrix} = \begin{pmatrix} \mathbf{g}_1(t_1) & \mathbf{g}_2(t_1) & \cdots & \mathbf{g}_N(t_1) \\ \mathbf{g}_1(t_2) & \mathbf{g}_2(t_2) & \cdots & \mathbf{g}_N(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{g}_1(t_M) & \mathbf{g}_2(t_M) & \cdots & \mathbf{g}_N(t_M) \end{pmatrix} \cdot \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_N \end{pmatrix}.$$

To ensure the restricted isometry property [113–118], one can normalize the coefficient vector by dividing the elements in each column by the  $L_2$  norm of that column:  $(\mathcal{G})_{ij} = (\mathcal{G})_{ij}/L_2(j)$  with  $L_2(j) = \sqrt{\sum_{i=1}^M [(\mathcal{G})_{ij}]^2}$ . After  $\mathbf{a}$  is determined via some standard compressive-sensing algorithm, the coefficients are given by  $\mathbf{a}/L_2$ . To determine the set of power-series coefficients corresponding to a different component of the dynamical variable, say component 2, one simply replaces the measurement vector by  $\mathbf{X} = [\dot{y}_i(t_1), \dot{y}_i(t_2), \dots, \dot{y}_i(t_M)]^T$  and use the same matrix  $\mathcal{G}$ . This way all coefficients can be estimated. After the equations of all components of  $i$  are determined, one can repeat this process for all other nodes to reconstruct the whole system.

The working of the method was illustrated by using networks of coupled chaotic Lorenz and Rössler oscillators as examples [24]. The classical Lorenz and Rössler systems are given by  $[\dot{x}, \dot{y}, \dot{z}] = 10(y-x), x(28-z) - y, xy - (8/3)z]$  and  $[\dot{x}, \dot{y}, \dot{z}] = -y - z, x + 0.2y, 0.2 + z(x - 5.7)]$ , respectively. Since  $m=3$ , the power series of  $x$ ,  $y$  and  $z$  can be chosen such that  $l_1 + l_2 + l_3 \leq 3$ . The total number of the coefficients to be estimated is then  $N \sum_{i=1}^3 (i+1)(i+2)/2 + 1 = 19N + 1$ , where  $i = l_1 + l_2 + l_3$  ranges from 1 to 3. Random and scale-free network topologies were studied [24]. In particular, the Lorenz oscillator network was chosen to be a Erdős-Rényi type of homogeneous random network [196], generated by assuming a small probability of link for any



**Figure 9: Examples of compressive sensing based network reconstruction: networks of coupled Lorenz and Rössler chaotic oscillators.** (a) Predicted coupling terms in the  $z$  variable for node #2 in a random network (b) of 10 coupled chaotic Lorenz oscillators. (c) Predicted terms in the dynamics of node #2 and couplings between node #2 and other nodes. (d) Predicted coupling terms in the  $x$ -variable for node #48 in a scale-free network (e) of 50 coupled chaotic Rössler oscillators. (f) Predicted terms in the dynamics of node #48 and couplings between node #48 and other nodes. In (c) and (f), top to bottom panels: predicted terms with coefficients in the  $x$ ,  $y$  and  $z$  variables, where the corresponding true values of the existent terms are marked. In (a) and (d), the node numbers corresponding to the existing terms are marked and the coupling forms are  $c^{3,2}$  and  $c^{1,3}$ , respectively. Here, “term index” refers to the order of to be predicted coefficients appearing in Eq. (23) and in the vector  $\mathbf{a}$  of Eq. (26). The average degrees  $\langle k \rangle$  for the random and scale-free networks are 6 and 10, respectively. The number of data points used for prediction is 140 and the time interval for data collection is  $\Delta t = 1$ . From Ref. [24] with permission.

pair of nodes. The coupling between nodal dynamics was assumed to occur between the  $y$  and the  $z$  variables in the Lorenz equations, leading to the following coupling matrix:  $c_{ij}^{3,2} = 1$  if nodes  $i$  and  $j$  are connected and  $c_{ij}^{3,2} = 0$  otherwise. The Rössler oscillator network was assumed to be a Barabási-Albert type of scale-free network [197] with a heterogeneous degree distribution. The coupling scheme is  $c_{ij}^{1,3} = 1$  for link between  $i$  and  $j$ . Both types of network structures are illustrated schematically in Fig. 9. Time series were generated by integrating the whole networked system with time step  $h = 10^{-4}$  for  $6 \times 10^6$  steps. However, the number of “measured” data points required for the method to be successful can be orders of magnitude less than  $6 \times 10^6$ , a *fundamental advantage of compressive-sensing method*. Specifically, random measurements were collected from the integrated time series and the number of elements in each row of the matrix  $\mathbf{G}$  is given by  $N(n+1)^m$ .

Figure 9 shows some representative results. For the random Lorenz network, the inferred coefficients were shown of node #2 associated with both the couplings with other nodes [Fig. 9(a)] and those with its own dynamics [Fig. 9(a)]. The term index is arranged from low to high values, corresponding to the order from low to high node number. The predicted coupling strengths between node #2 and others are shown in Fig. 9(a), where each term according to its index corresponds to a specific node. Nonzero terms belonging to nodes other than node #2 indicate inter-node couplings. The predicted interactions with nonzero coefficients (the value is essentially unity) are in agreement with the neighbors of node #2 in the sample random network in Fig. 9(b). The term 32 related to  $-6y$  is the coupling strength from node #2, which equals the sum of the coupling strengths from the other connected nodes. Figure 9(c) displays the inferred coefficients for both nodal dynamics and coupling terms in the three components  $x$ ,  $y$  and  $z$ . All predicted terms with nonzero coefficients are in agreement with those in the equations of the dynamics of node #2, together with the inter-node coupling terms  $c^{3,2}$ .

Figure 9(d) shows the predicted links between node #48 and others in a Rössler oscillator network with a scale-free structure. All existing couplings have been accurately inferred, as compared to the structure presented in Fig. 9(e), even though the interaction patterns among nodes are heterogeneous. Both the detected local dynamical and coupling terms associated with node #48 are indicated in Fig. 9(f), where in the  $x$  component, the term  $-8z$  is the combination of the local-dynamical term  $-z$  and the coupling of node #48 with 7 neighboring nodes. Since all the couplings have been successfully detected, the local-dynamical term  $-z$  in the  $x$  component can be separated from the combination so that all terms of node #48 are predicted. Similar results were obtained for all other nodes, leading to a complete and accurate reconstruction of the underlying complex networked system [24].

A performance analysis was carried out [24], with the result that the number of required data points is much smaller than the number of terms in the power series function, a main advantage of the compressive-sensing technique. Insofar as the number of data points exceeds a critical value, the prediction errors are effectively zero, indicating the robustness of the reconstruction. One empirical observation was that, if the sampling frequency is high, the number of data points is not able to cover the dynamics in the whole phase space. In order to obtain a faithful prediction of the whole system, the sampling frequency must be sufficiently low. Another important question was how the structural properties of the network affect the prediction precision. A calculation [24] of the dependence of the prediction error on the average degree  $\langle k \rangle$  and the network size  $N$  indicated that, regardless of the network size, insofar as the network connections are sparse, the prediction

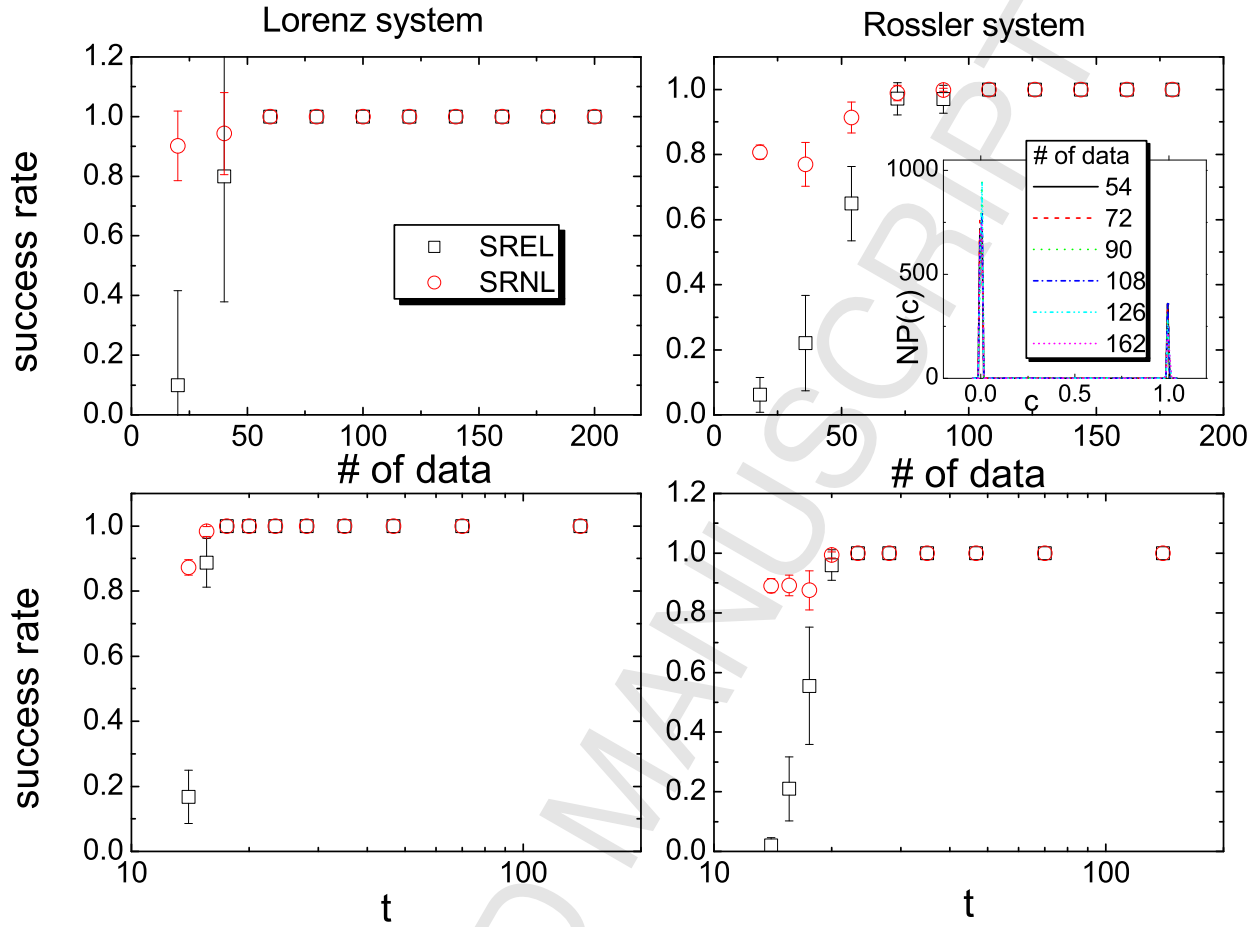


Figure 10: **Reconstruction accuracy of full network topology.** Success rates of existing links SREL and nonexistent links SRNL as a function of the number of data points and length  $t$  of time series for random Lorenz and Rössler networks. The inset in the upper-right panel shows the distribution of coupling strength in the Rössler network for different numbers of data points. From Ref. [24] with permission.

errors remain to be quite small, providing further support for the robustness of the compressive-sensing based method.

With respect to reconstruction of the network topology, computations demonstrated [24] that, despite the small prediction errors, all existing links in the original network can be predicted extremely reliably. The performance of the method for predicting network structures can be quantified through the success rates for existing links (SREL) and nonexistent links (SRNL), defined to be the ratio between the number of successfully predicted links and total number of links and the ratio of the number of correctly predicted nonexistent links to the total number of nonexistent links, respectively. Figure 10 shows the success rates versus the number of data points and  $t$  for both random Lorenz and Rössler oscillator networks. It can be seen that, when the number of data points and  $t$  are sufficiently large, both SREL and SRNL reach 100%. The inset in the upper-right panel shows the distribution of the coupling strengths in the Rössler network. For all tested numbers of data points ( $> 54$ ), there exist two sharp peaks centered at  $c = 0$  and  $c = 1.0$ , corresponding to the absence of coupling and the existing coupling of strength 1.0, respectively. The narrowness of the

Nodal dynamics	Lorenz		Rössler	
Errors	$E_{nz}$	$E_z$	$E_{nz}$	$E_z$
Dolphin social network [198]	$3.6 \times 10^{-3}$	$7.4 \times 10^{-6}$	$2.8 \times 10^{-3}$	$2.2 \times 10^{-7}$
Friendship network of karate club [199]	$3.3 \times 10^{-3}$	$3.6 \times 10^{-6}$	$1.9 \times 10^{-3}$	$2.7 \times 10^{-7}$
Network of political book purchases [24]	$3.8 \times 10^{-3}$	$8.3 \times 10^{-5}$	$2.8 \times 10^{-3}$	$2.1 \times 10^{-7}$
Electric circuit networks [200]	$4.4 \times 10^{-3}$	$5.8 \times 10^{-6}$	$2.9 \times 10^{-3}$	$2.7 \times 10^{-7}$
The neural network of C. Elegans [201]	$1.1 \times 10^{-3}$	$1.9 \times 10^{-5}$	$2.8 \times 10^{-3}$	$2.2 \times 10^{-7}$

Table 1: Prediction errors  $E_{nz}$  and  $E_z$  for five real-world networks. From Ref. [24] with permission.

two peaks in the distribution makes it feasible to distinguish existing links (with nonzero coupling strength) from nonexistent links (effectively with zero coupling). This provides an explanation for the 100% success rates shown in Fig. 10.

The method was also tested [24] on a number of real-world networks, ranging from social to biological and technological networks, where the nodal dynamics were assumed to be of the Lorenz and Rössler types. For five real-world networks, the prediction errors are shown in Table 1. It can be seen that all errors are small.

### 3.2 Reconstruction of complex networks with evolutionary-game dynamics

Many complex dynamical systems in biology, social science, and economics can be mathematically modeled by evolutionary games [202–206]. For example, in recent years evolutionary game models played an important role in addressing the biodiversity problem through microscopic modeling and the mechanism of species competition and coexistence at the level of individual interactions [207–234]. It was demonstrated [22] that compressive sensing can be exploited to reconstruct the full topology of the underlying network based on evolutionary game data. In particular, 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. It was shown [22] that, even when the available information about each agent's strategy and payoff is limited, the compressive-sensing based method can yield precise knowledge about the node-to-node interaction patterns in a highly efficient manner. The basic principle was further demonstrated by using an actual social experiment in which participants forming a friendship network played a typical game to generate short sequences of strategy and payoff data.

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  $\mathbf{S}(C) = (1, 0)^T$  and  $\mathbf{S}(D) = (0, 1)^T$ . 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) [235] and the snowdrift games (SG) [236], the payoff matrices are

$$\mathcal{P}_{PDG} = \begin{pmatrix} 1 & 0 \\ b & 0 \end{pmatrix} \quad \text{or} \quad \mathcal{P}_{SG} = \begin{pmatrix} 1 & 1-r \\ 1+r & 0 \end{pmatrix}, \quad (26)$$

respectively, 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

$$P_i = \sum_{j \in \Gamma_i} \mathbf{S}_i^T \cdot \mathcal{P} \cdot \mathbf{S}_j, \quad (27)$$

where  $\mathbf{S}_i$  and  $\mathbf{S}_j$  denote the strategies of agents  $i$  and  $j$  at the time and the sum is over the neighboring set  $\Gamma_i$  of  $i$ . After obtaining its payoff, an agent updates its strategy according to its own and neighbors' payoffs, attempting 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 [235], the Fermi equation [237], and payoff-difference-determined updating probability [238]. In the computational study of evolutionary game dynamics, the Fermi rule has been commonly used, which is defined, as follows. After a player  $i$  randomly chooses a neighbor  $j$ ,  $i$  adopts  $j$ 's status  $\mathbf{S}_j$  with the probability [237]:

$$W(\mathbf{S}_i \leftarrow \mathbf{S}_j) = \frac{1}{1 + \exp[(P_i - P_j)/\kappa]}, \quad (28)$$

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

The key to solving the network-reconstruction problem lies in the relationship between agents' payoffs and strategies. The interactions among agents in the network can be characterized by an  $N \times N$  adjacency matrix  $\mathcal{A}$  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

$$P_x(t) = a_{x1} \mathbf{S}_x^T(t) \cdot \mathcal{P} \cdot \mathbf{S}_1(t) + \cdots + a_{x,x-1} \mathbf{S}_x^T(t) \cdot \mathcal{P} \cdot \mathbf{S}_{x-1}(t) + a_{x,x+1} \mathbf{S}_x^T(t) \cdot \mathcal{P} \cdot \mathbf{S}_{x+1}(t) + \cdots + a_{xN} \mathbf{S}_x^T(t) \cdot \mathcal{P} \cdot \mathbf{S}_N(t), \quad (29)$$

where  $a_{xi}$  ( $i = 1, \dots, x-1, x+1, \dots, N$ ) represents a possible connection between agent  $x$  and its neighbor  $i$ ,  $a_{xi} \mathbf{S}_x^T(t) \cdot \mathcal{P} \cdot \mathbf{S}_i(t)$  ( $i = 1, \dots, x-1, x+1, \dots, N$ ) 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  $a_{xi} = 0$ ), and  $t = 1, \dots, m$  is the number of round that all agents play the game with their neighbors. This relation provides a base to construct the vector  $\mathbf{P}_x$  and matrix  $\mathcal{G}_x$  in a proper compressive-sensing framework to obtain solution of the neighboring vector  $\mathbf{A}_x$  of agent  $x$ . In particular, one can write

$$\begin{aligned} \mathbf{P}_x &= (P_x(t_1), P_x(t_2), \dots, P_x(t_m))^T, \\ \mathbf{A}_x &= (a_{x1}, \dots, a_{x,x-1}, a_{x,x+1}, \dots, a_{xN})^T, \end{aligned} \quad (30)$$

and  $\mathcal{G}_x =$

$$\begin{pmatrix} F_{x1}(t_1) & \cdots & F_{x,x-1}(t_1) & F_{x,x+1}(t_1) & \cdots & F_{xN}(t_1) \\ F_{x1}(t_2) & \cdots & F_{x,x-1}(t_2) & F_{x,x+1}(t_2) & \cdots & F_{xN}(t_2) \\ \vdots & \cdots & \vdots & \vdots & \vdots & \vdots \\ F_{x1}(t_m) & \cdots & F_{x,x-1}(t_m) & F_{x,x+1}(t_m) & \cdots & F_{xN}(t_m) \end{pmatrix},$$

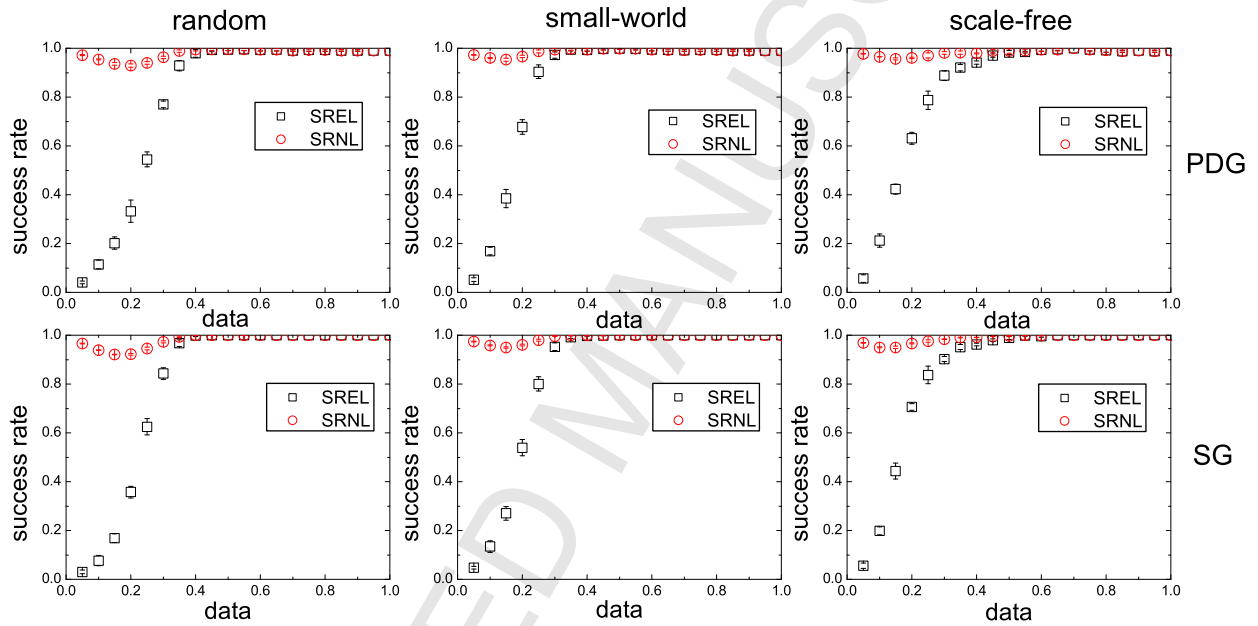
where  $F_{xy}(t_i) = \mathbf{S}_x^T(t_i) \cdot \mathcal{P} \cdot \mathbf{S}_y(t_i)$ . The vectors  $\mathbf{P}_x$ ,  $\mathbf{A}_x$  and matrix  $\mathcal{G}_x$  satisfy

$$\mathbf{P}_x = \mathcal{G}_x \cdot \mathbf{A}_x, \quad (31)$$

where  $\mathbf{A}_x$  is sparse due to the sparsity of the underlying complex network, making the compressive-sensing framework applicable. Since  $\mathbf{S}_x^T(t_i)$  and  $\mathbf{S}_y(t_i)$  in  $F_{xy}(t_i)$  come from data and  $\mathcal{P}$  is known, the vector  $\mathbf{P}_x$  can be obtained directly while the matrix  $\mathcal{G}_x$  can be calculated from the strategy and payoff data. The vector  $\mathbf{A}_x$  can thus be predicted based solely on the time series. Since the self-interaction term  $a_{xx}$  is not included in the vector  $\mathbf{A}_x$  and the self-column  $[F_{xx}(t_1), \dots, F_{xx}(t_m)]^T$  is excluded from the matrix  $\mathcal{G}_x$ , the computation required for compressive sensing can be reduced. In a similar fashion, the neighboring vectors of all other agents can be predicted, yielding the network adjacency matrix  $\mathcal{A} = (\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_N)$ .

The method was first tested [22] by implementing PDG and SG on three types of standard complex networks: random [196], small-world [201] and scale-free [197]. In particular, time series of strategies and payoffs were recorded during the system's evolution towards the steady state, which were used for uncovering the interaction network topology. The performance of the method can be quantified in terms of the amount of required measurements for different game types and network structures through the success rates of existent links (SREL) and non-existent links (SRNL). If the predicted value of an element of the adjacency matrix  $\mathcal{A}$  is close to 1, the corresponding link can be deemed to exist. If the value is close to zero, the prediction is that there is no link. In practice, a small threshold can be assigned, e.g., 0.1, so that the ranges of the existent and non-existent links are  $1 \pm 0.1$  and  $0 \pm 0.1$ , respectively. Any value outside the two intervals is regarded as a prediction failure. For a single player, SREL is defined as the ratio of the number of successfully predicted neighboring links to the number of actual neighbors, and SRNL is similarly defined. Averaging over all nodes leads to the values of SREL and SRNL for the entire network. The reason for treating the success rates for existent and non-existent links separately lies in the sparsity of the underlying complex network, where the number of non-existent 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 one, 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. 11. 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., the value of one means that the number of used measurements equals  $N$ . For all combinations of game dynamics and network topologies examined, perfect success rate can be achieved with extremely low amount of data. For example, for random and small-world networks, the length of data required for achieving 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 most nodes, although their neighboring vectors are still sparse. Figure 11 thus demonstrates that the method is both accurate and efficient. The exceptionally low data requirement is particularly important for situations where only rare information is available. From this standpoint, evolutionary games are suitable to simulate 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 convergent so that the transients are short. In addition, the robustness of the method was tested in situations where the time series are contaminated by noise [22]. For example, the case was studied where random noise of amplitude up to 30% is added to the payoffs of PDG. When the



**Figure 11: Reconstructing the full topology of complex networks hosting evolutionary game dynamics.** 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 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. Other values of  $b$  and  $r$  were also tested, yielding similar success rates. The average node degrees of all used networks are fixed to 6 and the noise parameter  $\kappa = 0.1$ . From Ref. [22] with permission.



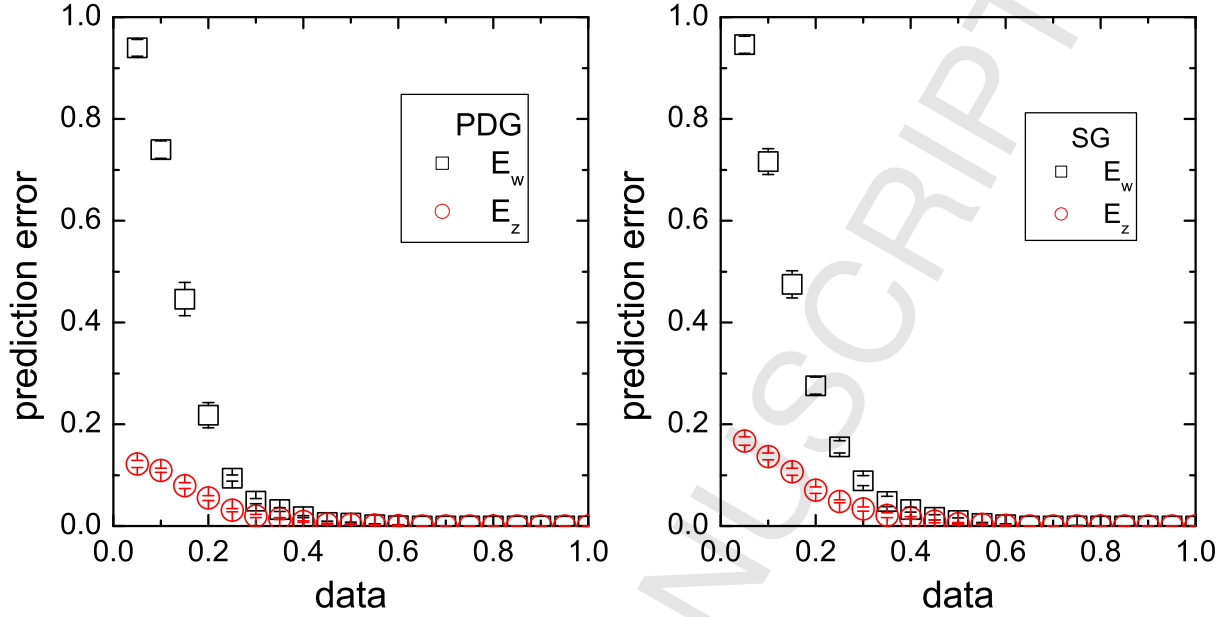


Figure 12: **Reconstructing the full topology of weighted complex networks hosting evolutionary game dynamics.** Prediction errors  $E_w$  in link weights and  $E_z$  in non-existent links for PDG on weighted scale-free networks. The network size is 100 and the weights follow a uniform distribution, 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 Fig. 11. From Ref. [22] with permission.

size of the data exceeds 0.4, the success rate approaches 100% for random networks. Similar performance was achieved for small-world and scale-free networks. The noise immunity embedded in the 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 of the fact that this kind of noise can increase the relaxation time towards one of the absorbing states (all C or all D), thereby providing more information for successful reconstruction.

The method can be generalized straightforwardly to weighted networks with inhomogeneous node-to-node interactions. Using weights to characterize various interaction strengths, the elements of the weighted adjacency matrix  $\mathcal{W}$  can be defined as

$$W_{ij} = \begin{cases} w_0 > 1, & \text{if } i \text{ connects to } j \\ 0, & \text{otherwise.} \end{cases} \quad (32)$$

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  $P_i^w$  of an arbitrary individual is given by

$$P_i^w = \sum_{j \in \Gamma_i} w_{ij} \mathbf{S}_i^T \cdot \mathcal{P} \cdot \mathbf{S}_j, \quad (33)$$

where  $\Gamma_i$  denotes the neighboring set of  $i$ . With the evolutionary-game dynamics, the weighted network structure is taken into account by the weighted payoff  $P_i^w$ . To uncover such a network

from data, the weighted payoff vector  $\mathbf{P}_x^w$ , matrix  $\mathcal{G}_x$ , and the weighted neighboring vector  $\mathbf{W}_x$  for an arbitrary individual  $x$  are needed. The vectors  $\mathbf{P}_x^w$  and  $\mathbf{W}_x$  are given by

$$\begin{aligned}\mathbf{P}_x^w &= (P_x^w(t_1), P_x^w(t_2), \dots, P_x^w(t_m))^T, \\ \mathbf{W}_x &= (W_{x1}, \dots, W_{x,x-1}, W_{x,x+1}, \dots, W_{xN})^T.\end{aligned}\quad (34)$$

Similar to unweighted networks, one has

$$\mathbf{P}_x^w = \mathcal{G}_x \cdot \mathbf{W}_x, \quad (35)$$

where  $\mathbf{W}_x$  can be calculated 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 non-existent 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 latter. The average error over all link weights is the prediction error  $E_w$ . However, a relative error for a zero (non-existent) weighted link cannot be defined, so it is necessary to use the absolute error  $E_z$ . Figure 12 shows the prediction errors for PDG dynamics on a scale-free network with random link weights chosen uniformly from the interval  $[1.0, 6.0]$ . It can be seen 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. Random and small-world networks were also tested with the finding that, to achieve the same level of accuracy, the requirement for data can be somewhat relaxed as compared with scale-free networks.

The compressive sensing based reconstruction framework can be applied to real social networks. Especially, an experimental was reported in Ref. [22], where 22 participants from Arizona State University played PDG together iteratively and, at each round each player was allowed to change his/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 rewarded. During the experiment, each player was allowed to communicate only with his/her direct neighbors for strategy updating. Prior to experiment, there was a social tie (link) between two players if they had already been acquainted to 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. 13(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 he/she recorded his/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 pre-existed friendship ties among the participants tend to favor cooperation and preclude the system from being trapped in the social dilemma, due to the relatively short data streams. 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 were thus suitable for network reconstruction. The results are shown in Fig. 13(b). It can be seen that the social network was successfully uncovered, despite the complicated process

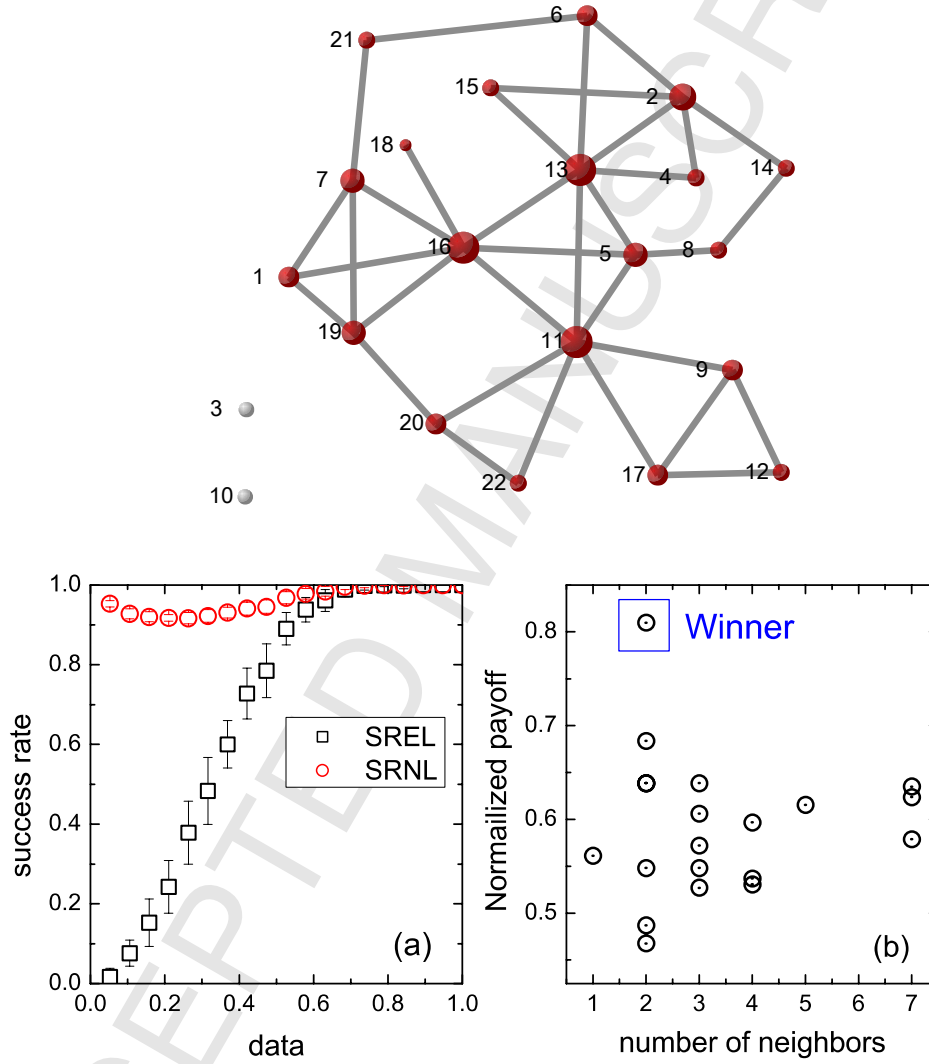


Figure 13: **Reconstructing an experimental social network using evolutionary game dynamics.** (a) Structure of experimental social network. (b) Success rates of uncovering the network topology and (c) normalized payoff of player as a function of node degrees. The 10 independent realizations used in calculating the average success rates were randomly chosen from the data base of 31 rounds of games. From Ref. [22] with permission.

of individual's decision making during the experiment. Compared to the simulation results, larger data set (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 were approximately at the average level, as shown in Fig. 13(c). In addition, the payoffs of players of smaller degrees were highly non-uniform, while those of higher degrees showed smaller difference. This suggests that players of high degree may not act as leaders due to their average normalized payoffs. This experimental finding was in striking agreement with numerical predictions in literature about the relationship between individuals' normalized payoffs and their node degrees [239]. It was also observed from experimental data that a typical player with a large number of neighbors failed to stimulate their neighbors to follow his/her strategies, suggesting that hubs may not be as influential in social networks. However, this finding should not be interpreted as a counter-example to the leader's role in evolutionary games [238,240], since the network based on friendship may violate the absolute selfish assumption of players who tend to be reciprocal with each other.

The method, besides being fully applicable to complex networks governed by evolutionary-game type of interactions, can be applied to 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 [241–244]. For such an application, Eq. (29) 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 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 sparsity requirement for compressive sensing can be met.

### 3.3 Detection of hidden nodes in complex networks

#### 3.3.1 Principle of detecting hidden nodes based on compressive sensing

When dealing with an unknown complex networked system that has a large number of interacting components organized hierarchically, curiosity demands that we ask the following question: are there hidden objects that are not accessible from the external world? The problem of inferring the existence of hidden objects from observations is quite challenging but it has significant applications in many disciplines of science and engineering. Here the meaning of “hidden” is that no direct observation of or information about the object is available so that it appears to the outside world as a black box. However, due to the interactions between the hidden object and other observable components in the system, it may be possible to utilize “indirect” information to infer the existence of the hidden object and to locate its position with respect to objects that can be observed. The difficulty to develop effective solutions is compounded by the fact that the indirect information on which any method of detecting hidden objects relies can be subtle and sensitive to changes in the system or in the environment. In particular, in realistic situations noise and random disturbances are present. It is conceivable that the “indirect” information can be mixed up with that due to noise

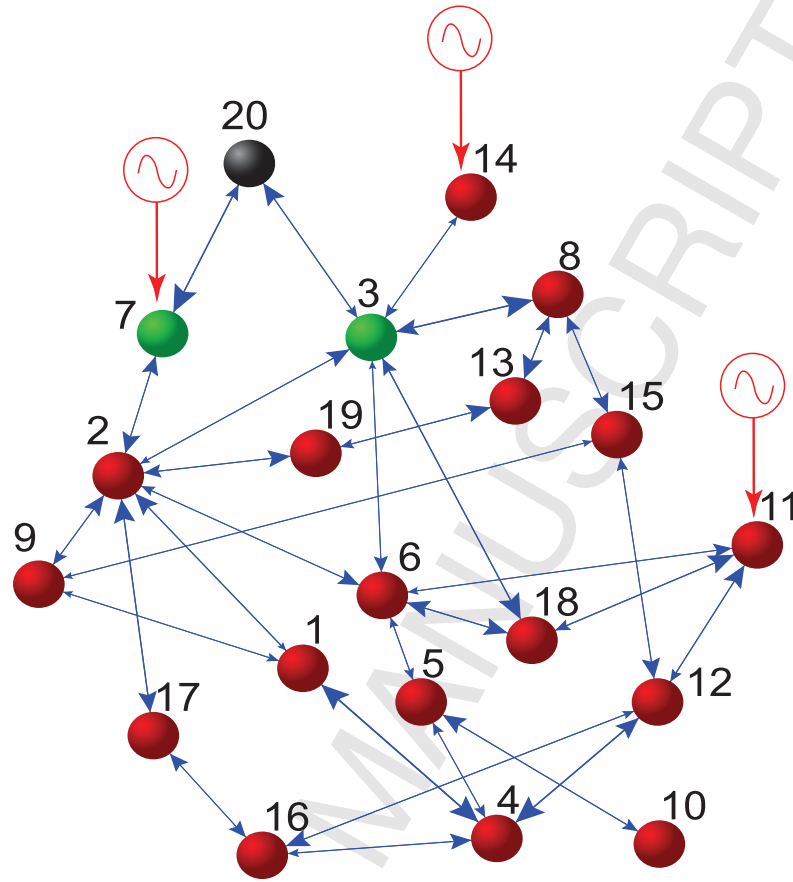


Figure 14: **Illustration of the principle of hidden node detection in a concrete setting.** An example of a complex network with a hidden node. Time series from all nodes except hidden node #20 can be measured, which can be detected when its immediate neighbors, nodes #3 and #7 are unambiguously identified. Nodes #7, #11, and #14 are driven by local noise sources. From Ref. [36] with permission.

or be severely contaminated. The presence of noise thus poses a serious challenge to detecting hidden nodes, and some effective “noise-mitigation” method must be developed.

One approach to addressing the problem of detecting a hidden node [31, 36] was based on compressive sensing. The basic principle principle is that the existence of a hidden node typically leads to “anomalies” in the quantities that can be calculated or deduced from observation. Simultaneously, noise, especially local random disturbances applied at the nodal level, can also lead to large variance in these quantities. This is so because, a hidden node is typically connected to a few nodes in the network that are accessible to the external world, and a noise source acting on a particular node in the network may also be regarded as some kind of hidden object. Thus, the key to any detection methodology is to identify and *distinguish* the effects of hidden nodes on detection measure from those due to *local* noise sources.

The key issues associated with detection of hidden nodes can be explained in a concrete manner by using the network setting shown schematically in Fig. 14, where there are 20 nodes, the couplings among the nodes are weighted, and the entire network is in a noisy environment, but a number of nodes also receive relatively strong random driving, e.g., nodes 7, 11, and 14. Assume

an oscillator network so that the nodal dynamics are described by nonlinear differential equations, and that time series can be measured simultaneously from all nodes in the network except one, labeled as #20, which is a hidden node. The task of ascertaining the presence and locating the position of the hidden node are equivalent to identifying its immediate neighbors, which are nodes #3 and #7 in Fig. 14. Note that, in order to be able to detect the hidden node based on information from its neighboring nodes, the interactions between the hidden node and its neighbors must be directional from the former to the latter or be bidirectional. Otherwise, if the coupling is solely from the neighbors to the hidden node, the dynamics of the neighboring nodes will not be affected by the hidden node and, consequently, time series from the neighboring nodes will contain absolutely no information about the hidden node, rendering it undetectable. The action of local noise source on a node is naturally directional, i.e., from the source to the node.

In Ref. [31], it was demonstrated that, when the compressive-sensing paradigm is applied to uncovering the network topology [24], the predicted linkages associated with nodes #3 and #7 are typically anomalously dense, and this piece of information is basically what is needed to identify them as the neighboring nodes of the hidden node. In addition, when different segments of the measurement data are used to reconstruct the coupling weights for these two nodes, the reconstructed exhibit significantly larger variances than those associated with other nodes. However, the predicted linkages associated with the nodes driven by local noise sources can exhibit behaviors similar to those due to the hidden nodes, leading to uncertainty in the detection of the hidden node. This issue is critical to developing algorithms for real-world applications. One possible solution [36] was to exploit the principle of *differential signal* to investigate the behavior of the predicted link weights as a function of the data used in the reconstruction. Due to the advantage of compressive sensing, the required data amount can be quite small and, hence, even if the method requires systematic increase in the data amount, it would still be reasonably small. It was argued that demonstrated [36] that, when the various ratios of the predicted weights associated with all pairs of links between the possible neighboring nodes and the hidden node are examined, those associated with the hidden nodes and nodes under strong local noise show characteristically distinct behaviors, rendering unambiguous identification of the neighboring nodes of the hidden node. Any such ratio is essentially a kind of differential signal, because it is defined with respect to a pair of edges.

### 3.3.2 Mathematical formulation of compressive sensing based detection of a hidden node

#### Compressive-sensing based method to uncover weighted network dynamics and topology.

Consider the typical setting of a complex network of  $N$  coupled oscillators in a noisy environment. The dynamics of each individual node, when it is isolated from other nodes, can be described as  $\dot{\mathbf{x}}_i = \mathbf{F}_i(\mathbf{x}_i) + \xi\eta_i$ , where  $\mathbf{x}_i \in \mathbb{R}^m$  is the vector of state variables,  $\eta_i$  are an  $m$ -dimensional vector whose entries are independent Gaussian random variables of zero mean and unit variance, and  $\xi$  denotes the noise amplitude. A weighted network can be described by

$$\dot{\mathbf{x}}_i = \mathbf{F}_i(\mathbf{x}_i) + \sum_{j=1, j \neq i}^N \mathcal{W}_{ij} \cdot [\mathbf{H}(\mathbf{x}_j) - \mathbf{H}(\mathbf{x}_i)] + \xi\eta_i, \quad (36)$$

where  $\mathcal{W}_{ij} \in \mathbb{R}^{m \times m}$  is the coupling matrix between node  $i$  and node  $j$ , and  $\mathbf{H}$  is the coupling function. Defining

$$\mathbf{F}'_i(\mathbf{x}_i) \equiv \mathbf{F}_i(\mathbf{x}_i) - \mathbf{H}(\mathbf{x}_i) \cdot \sum_{j=1, j \neq i}^N \mathcal{W}_{ij},$$

one has

$$\dot{\mathbf{x}}_i = \mathbf{F}'_i(\mathbf{x}_i) + \sum_{j=1, j \neq i}^N \mathcal{W}_{ij} \cdot \mathbf{H}(\mathbf{x}_j) + \xi \eta_i, \quad (37)$$

i.e., all terms directly associated with node  $i$  have been grouped into  $\mathbf{F}'_i(\mathbf{x}_i)$ . One can then expand  $\mathbf{F}'_i(\mathbf{x}_i)$  into the following form:

$$\mathbf{F}'_i(\mathbf{x}_i) = \sum_{\gamma} \tilde{\mathbf{a}}_i^{(\gamma)} \cdot \tilde{\mathbf{g}}_i^{(\gamma)}(\mathbf{x}_i), \quad (38)$$

where  $\tilde{\mathbf{g}}_i^{(\gamma)}(\mathbf{x}_i)$  are a set of orthogonal and complete base functions chosen such that the coefficients  $\tilde{\mathbf{a}}_i^{(\gamma)}$  are sparse. While the coupling function  $\mathbf{H}(\mathbf{x}_i)$  can be expanded in a similar manner, for simplicity we assume that they are linear:  $\mathbf{H}(\mathbf{x}_i) = \mathbf{x}_i$ . This leads to

$$\dot{\mathbf{x}}_i = \sum_{\gamma} \tilde{\mathbf{a}}_i^{(\gamma)} \cdot \tilde{\mathbf{g}}_i^{(\gamma)}(\mathbf{x}_i) + \sum_{j=1, j \neq i}^N \mathcal{W}_{ij} \cdot \mathbf{x}_j + \xi \eta_i, \quad (39)$$

where all the coefficients  $\tilde{\mathbf{a}}_i^{(\gamma)}$  and weights  $\mathcal{W}_{ij}$  need to be determined from time series  $\mathbf{x}_i$ . In particular, the coefficient vector  $\tilde{\mathbf{a}}_i^{(\gamma)}$  determines the nodal dynamics and the weighted matrices  $\mathcal{W}_{ij}$ 's give the full topology and coupling strength of the entire network.

Suppose simultaneous measurements are available of all state variables  $\mathbf{x}_i(t)$  and  $\mathbf{x}_i(t + \delta t)$  at  $M$  different uniform instants of time at interval  $\Delta t$  apart, where  $\delta t \ll \Delta t$  so that the derivative vector  $\dot{\mathbf{x}}_i$  can be estimated at each time instant. Equation (39) for all  $M$  time instants can then be written in a matrix form with the following measurement matrix:

$$\mathcal{G}_i = \begin{pmatrix} \tilde{\mathbf{g}}_i(t_1) & \mathbf{x}_1(t_1) & \cdots & \mathbf{x}_k(t_1) & \cdots & \mathbf{x}_N(t_1) \\ \tilde{\mathbf{g}}_i(t_2) & \mathbf{x}_1(t_2) & \cdots & \mathbf{x}_k(t_2) & \cdots & \mathbf{x}_N(t_2) \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ \tilde{\mathbf{g}}_i(t_M) & \mathbf{x}_1(t_M) & \cdots & \mathbf{x}_k(t_M) & \cdots & \mathbf{x}_N(t_M) \end{pmatrix}, \quad (40)$$

where the index  $k$  in  $\mathbf{x}_k(t)$  runs from 1 to  $N$ ,  $k \neq i$ , and each row of the matrix is determined by the available time series at one instant of time. The derivatives at different time can be written in a vector form as  $\mathbf{X}_i = [\dot{\mathbf{x}}_i(t_1), \dots, \dot{\mathbf{x}}_i(t_M)]^T$ , and the coefficients from the functional expansion and the weights associated with all links in the network can be combined concisely into a vector  $\mathbf{a}_i$  as

$$\mathbf{a}_i = [\tilde{\mathbf{a}}_i, \mathbf{W}_{1i}, \dots, \mathbf{W}_{i-1,i}, \mathbf{W}_{i+1,i}, \dots, \mathbf{W}_{N,i}]^T, \quad (41)$$

where  $[\cdot]^T$  denotes the transpose. For a properly chosen expansion base and a general complex network whose connections are sparse, the vector  $\mathbf{a}_i$  to be determined is sparse as well. Finally, Eq. (39) can be written as

$$\mathbf{X}_i = \mathcal{G}_i \cdot \mathbf{a}_i + \xi \eta_i. \quad (42)$$

In the absence of noise or if the noise amplitude is negligibly small, Eq. (42) represents a linear equation but the dimension of the unknown coefficient vector  $\mathbf{a}_i$  can be much larger than that of  $\mathbf{X}_i$ , and the measurement matrix will have many more columns than rows. In order to fully recover the network of  $N$  nodes with each node having  $m$  components, it is necessary to solve  $N \times m$  such equations. Since  $\mathbf{a}_i$  is sparse, insofar as its number of non-zero coefficients is smaller than the dimension of  $\mathbf{X}_i$ , the vector  $\mathbf{a}_i$  can be uniquely and efficiently determined through compressive sensing.

**Detection of hidden node.** A meaningful solution of Eq. (42) based on compressive sensing requires that the derivative vector  $\mathbf{X}_i$  and the measurement matrix  $\mathcal{G}_i$  be entirely known which, in turn, requires time series from all nodes. In this case, the information available for reconstruction of the complex networked system is deemed to be *complete* [31, 36]. In the presence of a hidden node, for its immediate neighbors, the available information will not be complete in the sense that some entries of the vector  $\mathbf{X}_i$  and the matrix  $\mathcal{G}_i$  are unknown. Let  $h$  denote the hidden node. For any neighboring node of  $h$ , the vector  $\mathbf{X}_i$  and the matrix  $\mathcal{G}_i$  in Eq. (42) now contain unknown entries at the locations specified by the index  $h$ . For any other node not in the immediate neighborhood of  $h$ , Eq. (42) is unaffected. When compressive-sensing algorithm is used to solve Eq. (42), there will then be large errors in the solution of the coefficient vector  $\mathbf{a}_i$  associated the neighboring nodes of  $h$ , regardless of the amount of data used. In general, the so-obtained coefficient vector  $\mathbf{a}_i$  will not appear sparse. Instead, most of its entries will not be zero, a manifestation of which is that the node would appear to have links with almost every other node in the network. In contrast, for nodes not in the neighborhood of  $h$ , the corresponding errors will be small and can be reduced by increasing the data amount, and the corresponding coefficient vector will be sparse. It is this observation which makes identification of the neighboring nodes of the hidden node possible in a noiseless or weak-noise situation [31].

The need and the importance to distinguish the effects of hidden node from these of noise can be better seen by separating the term associated with  $h$  in Eq. (39) from those with other accessible nodes in the network. Letting  $l$  denote a node in the immediate neighborhood of the hidden node  $h$ , we have

$$\mathbf{X}_l = \mathcal{G}_l' \cdot \mathbf{a}_l' + (\mathcal{W}_{lh} \cdot \mathbf{x}_h + \xi \eta_l), \quad (43)$$

where  $\mathcal{G}_l'$  is the new measurement matrix that can be constructed from all available time series. While background noise may be weak, the term  $\mathcal{W}_{lh} \cdot \mathbf{x}_h$  can in general be large in the sense that it is comparable in magnitude with other similar terms in Eq. (39). Thus, when the network is under strong noise, especially for those nodes that are connected to the neighboring nodes of the hidden node, the effects of hidden node on the solution can be entangled with those due to noise. In addition, if the coupling strength from the hidden node is weak, it would be harder to identify the neighboring nodes. For example, a hidden node in a network with Gaussian weight distribution will be harder to detect, due to the finite probability of the occurrence of near zero weights. When the coupling strength is comparable or smaller than the background noise amplitude, the corresponding link cannot be detected.

**Method to distinguish hidden nodes from local noise sources.** The basic idea to distinguish the effects of hidden node and of local noise sources [36] is based on the following consideration.



Take two neighboring nodes of the hidden node, labeled as  $i$  and  $j$ . Because the hidden node is a common neighbor of nodes  $i$  and  $j$ , the couplings from the hidden node should be approximately proportional to each other, with the proportional constant determined by the ratio of their link weights with the hidden node. When the dynamical equations of nodes  $i$  and  $j$  are properly normalized, the terms due to the hidden node tend to cancel each other, leaving the normalization constant as a single unknown parameter that can be estimated subsequently. This parameter is the *cancellation ratio*, denoted as  $\Omega_{ij}$ . As the data amount is increased,  $\Omega_{ij}$  tends to its true value. Practically one then expects to observe a systematic change in the estimated value of the ratio as data used in the compressive-sensing algorithm is increased from some small to relatively large amount. If only local noise sources are present, the ratio should show no systematic change with the data amount. Thus the distinct behaviors of  $\Omega_{ij}$  as the amount of data is increased provides a way to distinguish the hidden node from noise and, at the same time, to ascertain the existence of the hidden node.

For simplicity, assume that all coupled oscillators share the same local coupling configuration and that each oscillator is coupled to any of its neighbors through one component of the state vector only. Thus, each row in the coupling matrix  $\mathcal{W}_{ih}$  associated with a link between node  $i$  and  $h$  has only one non-zero element. Let  $p$  denote the component of the hidden node coupled to the first component of node  $i$ , the dynamical equation of which can then be written as

$$[\dot{\mathbf{x}}_i]_1 = \left[ \sum_{\gamma} \tilde{\mathbf{a}}_i^{(\gamma)} \cdot \tilde{\mathbf{g}}_i^{(\gamma)}(\mathbf{x}_i) \right]_1 + \left[ \sum_{k \neq i, h}^N \mathcal{W}_{ik} \cdot \mathbf{x}_k \right]_1 + w_{ih}^{1p} \cdot [\mathbf{x}_h]_p + \xi \eta_i, \quad (44)$$

where  $[\mathbf{x}_h]_p$  denotes the time series of the  $p$ th component of the hidden node, which is unavailable, and  $w_{ih}^{1p}$  is the coupling strength between the hidden node and node  $i$ . The dynamical equation of the first component of neighboring node  $j$  of the hidden node has a similar form. Letting

$$\Omega_{ij} = w_{ih}^{1p} / w_{jh}^{1p}, \quad (45)$$

be the cancellation ratio, multiplying  $\Omega_{ij}$  to the equation of node  $j$ , and subtracting from it the equation for node  $i$ , one obtains

$$\begin{aligned} [\dot{\mathbf{x}}_i]_1 &= \Omega_{ij} [\dot{\mathbf{x}}_j]_1 + \sum_{\gamma} \tilde{\mathbf{a}}_i^{(\gamma)} \cdot \tilde{\mathbf{g}}_i^{(\gamma)}(\mathbf{x}_i) + \sum_{k \neq i, h} w_{ik}^{1p} [\mathbf{x}_k]_p - \Omega_{ij} \sum_{\gamma} \tilde{\mathbf{a}}_j^{(\gamma)} \cdot \tilde{\mathbf{g}}_j^{(\gamma)}(\mathbf{x}_j) \\ &\quad - \Omega_{ij} \sum_{k \neq j, h} w_{jk}^{1p} [\mathbf{x}_k]_p + (w_{ih}^{1p} - \Omega_{ij} w_{jh}^{1p}) \cdot [\mathbf{x}_h]_p + \xi \eta_i - \Omega_{ij} \xi \eta_j. \end{aligned} \quad (46)$$

It can be seen that the terms associate with  $[\mathbf{x}_h]_p$  vanish and all deterministic terms on the left-hand side of Eq. (46) are known, which can then be solved by the compressive-sensing method. From the coefficient vector so estimated, one can identify the coupling of nodes  $i$  and  $j$  to other nodes, except for the coupling between themselves since such terms have been absorbed into the nodal dynamics, and the couplings to their common neighborhood are degenerate in Eq. (46) and cannot be separated from each other. Effectively, one has combined the two nodes together by introducing the cancellation ratio  $\Omega_{ij}$ .

As a concrete example, consider the situation where each oscillator has three independent dynamical variables, named as  $x$ ,  $y$  and  $z$ . For the nodal and coupling dynamics polynomial expansions of order up to  $n$  can be chosen. The  $x$  component of the nodal dynamics  $[\mathbf{F}'_i(\mathbf{x}_i)]_x$  for node  $i$

is:

$$[\mathbf{F}'_i(\mathbf{x}_i)]_x = \sum_{l_x=0}^n \sum_{l_y=0}^n \sum_{l_z=0}^n [a_{l_x l_y l_z}]_x \cdot x_i^{l_x} y_i^{l_y} z_i^{l_z},$$

and the coupling from other node  $k$  to the  $x$  component can be written as

$$C_{ik}^x = w_{ik}^{xx} \cdot x_k + w_{ik}^{xy} \cdot y_k + w_{ik}^{xz} \cdot z_k,$$

where  $w_{ik}^{xy}$  denotes the coupling weight from the  $y$  component of node  $k$  to the  $x$  component of node  $i$ , and so on. The nodal dynamical terms in the matrix  $G_i$  are

$$[\tilde{\mathbf{g}}_i]_x = [x_i^0 y_i^0 z_i^0, x_i^1 y_i^0 z_i^0, \dots, x_i^n y_i^n z_i^n],$$

and the corresponding coefficients are  $[a_{l_x l_y l_z}]_x$ . The vector of coupling weights is

$$[\mathcal{W}_{ij}]_x = [w_{ij}^{xx}, w_{ij}^{xy}, w_{ij}^{xz}].$$

Equation (46) becomes

$$\begin{pmatrix} \dot{x}_i(t_1) \\ \dot{x}_i(t_2) \\ \vdots \\ x_i(t_M) \end{pmatrix} \approx \begin{pmatrix} \dot{x}_j(t_1) & 1 & [\tilde{\mathbf{g}}_i(t_1)]_x & [\tilde{\mathbf{g}}_j(t_1)]_x & x_1(t_1) & \cdots & z_N(t_1) \\ \dot{x}_j(t_2) & 1 & [\tilde{\mathbf{g}}_i(t_2)]_x & [\tilde{\mathbf{g}}_j(t_2)]_x & x_1(t_2) & \cdots & z_N(t_2) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ \dot{x}_j(t_M) & 1 & [\tilde{\mathbf{g}}_i(t_M)]_x & [\tilde{\mathbf{g}}_j(t_M)]_x & x_1(t_M) & \cdots & z_N(t_M) \end{pmatrix} \cdot \begin{pmatrix} \Omega_{ij} \\ c \\ \tilde{\mathbf{a}}'_i \\ -\Omega_{ij} \cdot \tilde{\mathbf{a}}'_j \\ w_{i1}^{xx} - \Omega_{ij} w_{j1}^{xx} \\ \vdots \\ w_{iN}^{xz} - \Omega_{ij} w_{jN}^{xz} \end{pmatrix},$$

where  $c$  is the sum of constant terms from the dynamical equations of nodes  $i$  and  $j$ , and  $\tilde{\mathbf{a}}'_i$  is the coefficient vector to be estimated which excludes all the constants. Using compressive sensing to solve this equation, one can recover the cancellation ratio  $\Omega_{ij}$  and the equations of node  $i$ . When  $\Omega_{ij}$  is known the dynamics of node  $j$  can be recovered from the coefficient vector  $-\Omega_{ij} \cdot \tilde{\mathbf{a}}'_j$ .

In Ref. [36], an analysis and discussions were provided about the possible extension of the method to systems of characteristically different nodal dynamics and/or with multiple hidden nodes. In particular, it was shown that the method can be readily adopted to network systems whose nodal dynamics are not described by continuous-time differential equations but by discrete-time processes such as evolutionary-game dynamics. In such a case, the derivatives used for continuous-time systems can be replaced by the agent payoffs. The cancellation factors can then be calculated from data to differentiate the hidden nodes from local noise sources. It was also shown that, under certain conditions with respect to the coupling patterns between the hidden nodes and their neighboring nodes, the cancellation factors can be estimated even when there are multiple, entangled hidden nodes in the network.

### 3.3.3 Examples of hidden node detection in the presence of noise

The methodology of hidden node detection in the presence of noise can be illustrated using coupled oscillator networks. (Results from evolutionary-game dynamical networks can be found in

Ref. [36].) As discussed in Sec. 3.1, given such a networked system, one can use compressive sensing to uncover all the nodal dynamical equations and coupling functions [24]. The expansion base needs to be chosen properly so that the number of non-zero coefficients is small as compared with the total number  $N_t$  of unknown coefficients. All  $N_t$  coefficients constitute a coefficient vector to be estimated. The amount of data used can be conveniently characterized by  $R_m$ , the ratio of the number  $M$  of data points used in the reconstruction, to  $N_t$ .

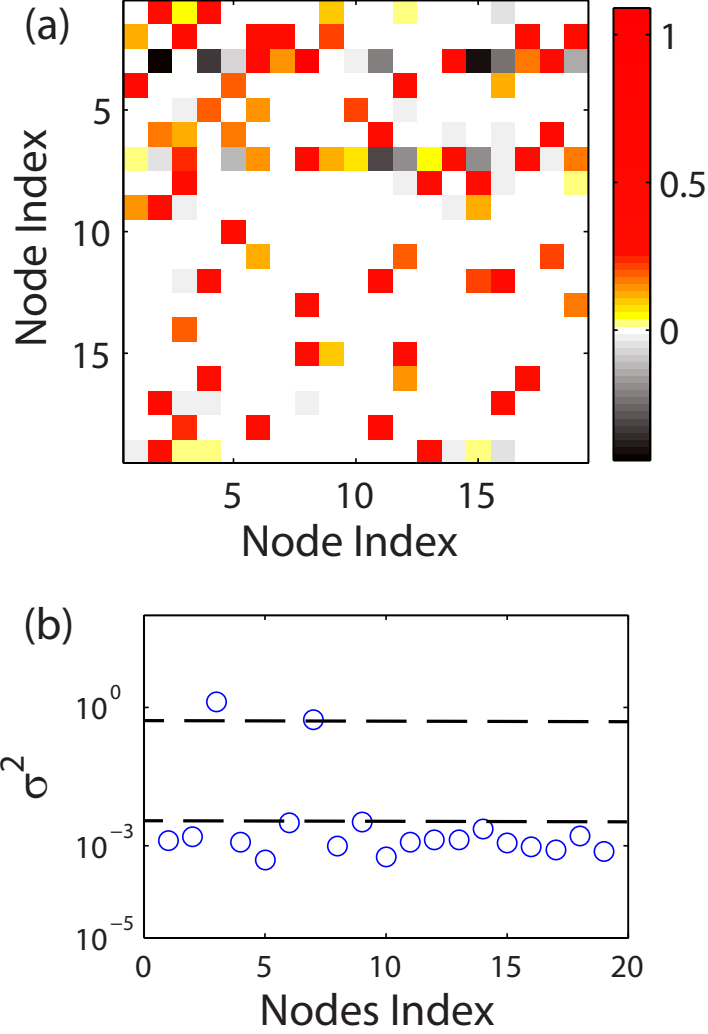
The method to differentiate hidden nodes and noise was tested [36] using random networks of nonlinear/chaotic oscillators, where the nodal dynamics were chosen to be those from the Rössler oscillator [160],

$$[\dot{x}_i, \dot{y}_i, \dot{z}_i] = [-y_i - z_i, x_i + 0.2y_i, 0.2 + z_i(x_i - 5.7)],$$

which exhibits a chaotic attractor. The size of the network was varied from 20 to 100, and the probability of connection between any two nodes is 0.04. The network link weights are equally distributed in  $[0.1, 0.5]$  (arbitrary). Background noise of amplitude  $\xi$  was applied (independently) to every oscillator in the network, with amplitude varying from  $10^{-4}$  to  $5 \times 10^{-3}$ . The noise amplitude is thus smaller than the average coupling strength of the network. The tolerance parameter  $\epsilon$  in the compressive sensing algorithm can be adjusted in accordance with the noise amplitude [36]. Time series are generated by using the standard Heun's algorithm [245] to integrate the stochastic differential equations. To approximate the velocity field, a third-order polynomial expansion was used in the compressive-sensing formulation. (In Ref. [36], more examples can be found using network systems of varying sizes, different weight distributions and topologies, and alternative nodal dynamics.)

**Illustration of hidden node detection.** Consider the network in Fig. 14, where only background noise is present and there are no local noise sources. Linear coupling between any pair of connected nodes is from the  $z$ -component to the  $x$ -component in the Rössler system. From the available time series (nodes #1-19), the coefficient vector can be solved using compressive sensing. In particular, for node  $i$ , the terms associated with couplings from the  $z$ -components of other nodes appear in the  $i$ th row of the coupling matrix. As shown in Fig. 15(a), when the data amount is  $R_m = 0.7$ , the network's coupling matrix can be predicted. The predicted links and the associated weights are sparse for all nodes except for nodes #3 and #7, the neighbors of the hidden node. While there are small errors in the predicted weights due to background noise, the predicted couplings for the two neighbors of the hidden node, which correspond to the 3rd and the 7th row in the coupling matrix, appear to be from almost all other nodes in the network and some coupling strength is even negative. Such anomalies associated with the predicted coupling patterns of the neighboring nodes of the hidden node cannot be removed by increasing the data amount. Nonetheless, it is precisely these anomalies which hint at the likelihood that these two "abnormal" nodes are connected with a hidden node.

While abnormally high connectivity predicted for a node is likely indication that it belongs to the neighborhood of the hidden node, in complex networks there are hub nodes with abnormally large degrees, especially for scale-free networks [197]. In order to distinguish a hidden node's neighboring node from some hub node, one can exploit the variance of the predicted coupling constants, which can be calculated from different segments of the available data sets. Due to the intrinsically low-data requirement associated with compressive sensing, the calculation of the variance is feasible because any reasonable time series can be broken into a number of segments,



**Figure 15: Example of hidden node detection without local noise sources.** For the network in Fig. (14), (a) predicted coupling matrix for all nodes except node #20. Time series from nodes #1 to #19 are available, while node #20 is hidden. The predicted weights are indicated by color coding and the amount of data used is  $R_m = 0.7$ . The abnormally dense patterns in the 3rd and 7th rows suggest that nodes #3 and #7 are the immediate neighbors of the hidden node. (b) Variance  $\sigma^2$  of the predicted coefficients for all accessible nodes, which is calculated using 20 independent reconstructions based on different segments of the data. The variances associated with nodes #3 and #7 are apparently much larger than those with the other nodes, confirming that these are the neighboring nodes of the hidden node. There is a definite gap between the values of the variance associated with neighboring and non-neighboring nodes of the hidden node, as indicated by the two horizontal dashed lines in (b). From Ref. [36] with permission.

and prediction can be made from each data segment. For nodes not in the neighborhood of the hidden node, the variances are small as the predicted results hardly change when different segments of the time series are used. However, for the neighboring nodes of the hidden node, due to lack of complete information needed to construct the measurement matrix, the variance values can be much larger. Figure 15(b) shows the variance  $\sigma^2$  in the predicted coupling strength for all 19 accessible nodes. It can be seen that the values of the variance for the neighboring nodes of the hidden node, nodes #3 and #7, are at or above the upper dashed line and are significantly larger than those associated with all other nodes that all fall below the lower dashed line. This indicates strongly that they are indeed the neighboring nodes of the hidden node. The gap between the two dashed lines can be taken as a quantitative measure of the detectability of the hidden node. The larger the gap, the more reliable it is to distinguish the neighbors of the hidden node from the nodes that not in the neighborhood. The results in Fig. 15 thus indicate that the location of the hidden node in the network can be reliably inferred in the presence of weak background noise. The size of the gap, or the hidden-node detectability depends on the system details. A systematic analysis of the detectability measure was done [36], where it was found that the variance due to the hidden node is mainly determined by the strength of its coupling with the accessible nodes in the network. It was also found that system size and network topology have little effect on the hidden-node detectability. It is worth emphasizing that the detectability relies also on successful reconstruction of all nodes that are not in the neighborhood of the hidden node, which determine the lower dashed line in Fig. 15.

The reliability of the reconstruction results can be quantified by investigating how the prediction errors in the link weights of all accessible nodes, except the predicted neighbors of the hidden node, change with the data amount. For an existent link, one can use the normalized absolute error  $E_{nz}$ , the error in the estimated weight with respect to the true one, normalized by the value of the true link weight. Figure 16 shows the results for  $N = 100$ . The link weights are uniformly distributed in the interval  $[0.1, 0.5]$  and the background noise amplitude is  $\xi = 10^{-3}$ . The tolerance parameter in the compressive-sensing algorithm was set to be  $\varepsilon = 0.5$ , which is optimal for this noise amplitude. (Details of determining the optimal tolerance parameter for different values of the background noise amplitude can be found in Ref. [36].) It can be seen that for  $R_m > 0.4$ ,  $E_{nz}$  decreases to the small value of about 0.01, which is determined by the background noise level. As  $R_m$  is increased further, the error is bounded by a small value determined by the noise amplitude, indicating that the reconstruction is robust. Although the value of  $E_{nz}$  does not decrease further toward zero due to noise, the prediction results are reliable in the sense that the predicted weights and the real values agree with each other, as shown in the inset of Fig. 16, a comparison of the actual and the predicted weights for all existent links. All the predicted results are in the vicinities of the corresponding actual values, as indicated by a heavy concentration of the dots along the diagonal line. The central region in the dot distribution has brighter color than the marginal regions, confirming that vast majority of the predicted results are accurate. In Ref. [36], it was further shown that robust reconstruction can be achieved regardless of the network size, connection topology and weight distributions, insofar as sufficient data are available.

The error measure  $E_{nz}$  to characterize the accuracy of the reconstruction is similar to  $z$ -scores, or the standard score in statistics, with the minor difference being that the  $z$ -scores use the standard derivatives of the distribution to normalize the raw scores, while the exact values were used [36] in the examples. In realistic applications the exact values are usually not available, so it is necessary

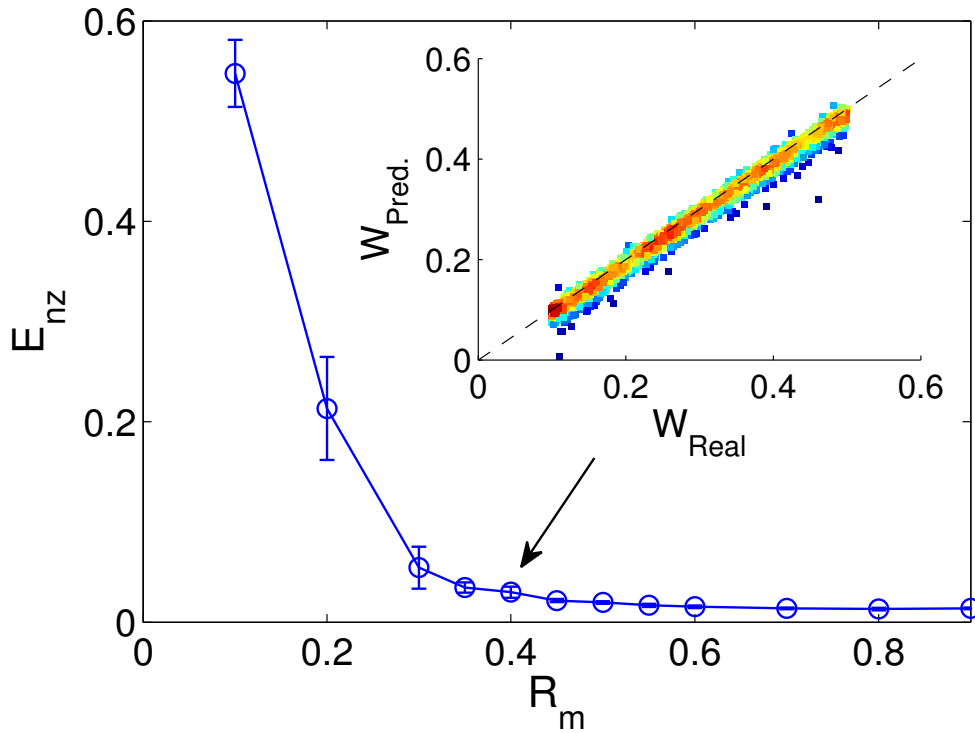


Figure 16: **Error analysis of hidden node detection in the presence of weak background noise.** For random networks of size  $N = 100$  with uniform weight distribution in  $[0.1, 0.5]$ , prediction error  $E_{nz}$  associated with nonzero coefficients of the dynamical equations of all nodes except for the neighboring nodes of the hidden node, as a function of normalized data amount  $R_m$ . The background noise amplitude is  $\xi = 10^{-3}$  for all nodes. All data points are obtained from 10 independent realizations. Inset is a comparison of the predicted and actual weights for all existent links. Each dot represents one such link, and its  $x$ -value is the actual weight while the  $y$ -value is the corresponding predicted result. The color for each dot is determined by the dot density around it, while bright color signifies high density. The arrow indicates the value of  $R_m$  used in the comparison study. The tolerance of the compressive-sensing algorithm is set to be  $\varepsilon = 0.5$ . From Ref. [36] with permission.

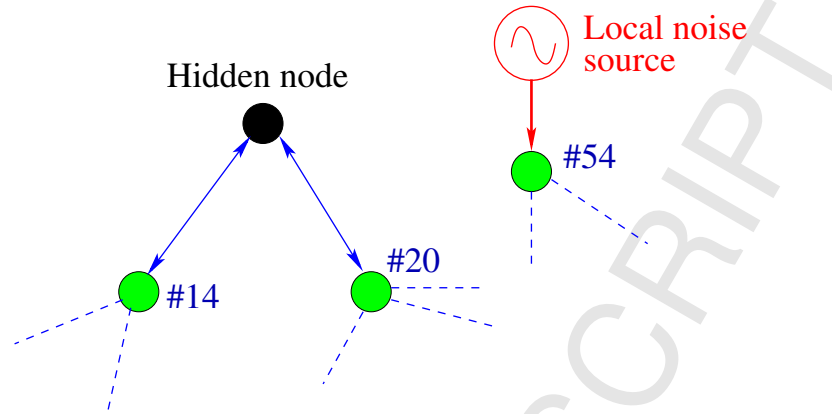


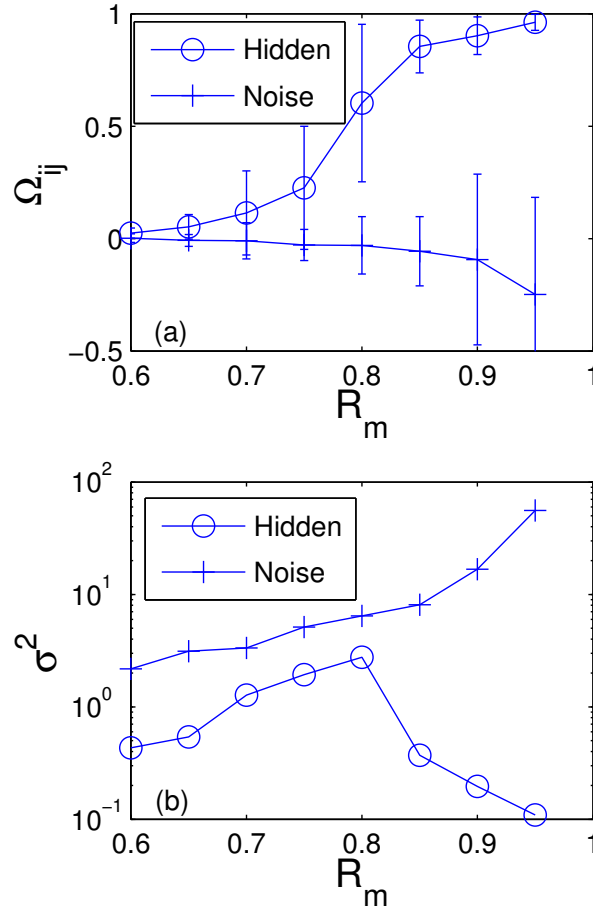
Figure 17: **Typical local structure of a hidden node.** Schematic illustration of a hidden node and its coupling configuration with two neighbors in a random network of  $N = 61$  nodes, where 60 are accessible. A strong noise source is present at node #54. From Ref. [36] with permission.

to use the  $z$ -score measure.

It should be emphasized that there are two types of “dense” connections: one from reconstruction and another intrinsic to the network. In particular, in the two-dimensional representation of the reconstruction results [e.g., Fig. 15(A)], the neighboring nodes of the hidden node typically appear densely linked to many other nodes in the network. These can be a result of lack of incomplete information (i.e., time series) due to the hidden node (in this case, there is indeed a hidden node), or the intrinsic dense connection pattern associated with, for example, a hub node in a scale-free network. The purpose of examining the variances of the reconstructed connections from independent data segments is for distinguishing these two possibilities. Extensive computations [36] indicated that a combination of the dense connection and large variance can ascertain the existence of hidden node with confidence.

**Differentiating hidden node from local noise sources.** When strong noise sources are present at certain nodes, the predicted coupling patterns of the neighboring nodes of these nodes will show anomalies. (Here the meaning of the term “strong” is that the amplitudes of the random disturbances are order-of-magnitude larger than that of background noise.) The method based on the cancellation ratio was demonstrated [36] to be effective at distinguishing hidden nodes from local noise sources, insofar as the hidden node has at least two neighboring nodes not subject to such disturbances. To be concrete, consider a network of  $N = 61$  coupled chaotic Rössler oscillators, which has 60 accessible nodes and one hidden node (#61) that is coupled to two neighbors: nodes #14 and #20, as illustrated schematically in Fig. 17. Assume a strong noise source is present at node #54. It was found [36] that the reconstructed weights match their true values to high accuracy. It was also found that the reconstructed coefficients including the ratio  $\Omega_{ij}$  are all constant and invariant with respect to different data segments, a strong indication that the pair of nodes are the neighboring nodes of the same hidden node, thereby confirming its existence.

When there are at least two accessible nodes in the neighborhood of the hidden node that are not subject to strong noisy disturbance, such as nodes #14 and #20, as the data amount  $R_m$  is increased towards 100%, the cancellation ratio should also increase and approach unity. This behavior is



**Figure 18: Distinguishing hidden nodes from local noise sources.** For the network described in Fig. 17, (a) Predicted values of the cancellation ratio  $\Omega_{ij}$  obtained from the differential signal of two neighboring nodes of the hidden node (#14 and #20, indicated by circles) and from the differential signal of nodes #14 and #54, where the latter is driven by noise of amplitude  $\xi = 10^{-2}$  (crosses). (b) Average variance values of the predicted local coefficient vectors for the two combinations. The background noise amplitude is  $\xi = 10^{-5}$ . The results were obtained from 20 independent realizations. From Ref. [36] with permission.



shown as the open circles in Fig. 18(a). However, when a node is driven by a local noise source, regardless of whether it is in the neighborhood of the hidden node, the cancellation ratio calculated from this node and any other accessible node in the network exhibits a characteristically different behavior. Consider, for example, nodes #14 and #54. The reconstructed connection patterns of these two nodes both show anomalies, as they appear to be coupled with all other nodes in the network. In contrast to the case where the pair of nodes are influenced by the hidden node only, here the cancellation ratio does not exhibit any appreciable increase as the data amount is increased, as shown with the crosses in Fig. 18(a). In addition, the average variance values of the predicted coefficient vectors of the two nodes exhibit characteristically different behaviors, depending on whether any one node in the pair is driven by strong noise or not. In particular, for the node pair #14 and #20, since neither is under strong noise, the average variance will decrease toward zero as  $R_m$  approaches unity, as shown in Fig. 18(b) (open circles). In contrast, for the node pair #14 and #54, the average variance will increase with  $R_m$ , as shown in Fig. 18(b) (crosses). This is because, when one node is under strong random driving, the input to the compressive-sensing algorithm will be noisy, so its performance will deteriorate. However, compressive sensing can perform reliably when the input data are “clean,” even when they are sparse. Increasing the data amount beyond a threshold is not necessarily helpful, but longer and noisier data sets can degrade significantly the performance. The results in Figs. 18(a,b) thus demonstrate that the cancellation ratio between a pair of nodes, in combination with the average variance of the predicted coefficient vectors associated with the two nodes, can effectively distinguish a hidden node from a local noise source. If there are more than one hidden node or there is a cluster of hidden nodes, the procedure to estimate the cancellation factors is similar but requires additional information about the neighboring nodes of the hidden nodes. The cancellation-factor based method can be extended to network systems with nodal dynamics not of the continuous-time type, such as evolutionary-game dynamics [36].

**Multiple entangled hidden nodes.** When there are multiple hidden nodes in a network, there is a possibility that an identified node is connected to more than one hidden node. For example, node  $b$  in all three panels of Fig. 19 is affected by hidden nodes  $H1$  and  $H2$ . The cancellation factor(s) can still be estimated if each hidden node in the network has at least two nodes (otherwise it can be treated as a local noise source).

For simplicity, first consider the situation of two entangled hidden nodes in that they have overlapping neighborhoods. Some possible coupling patterns are shown in Fig. 19. In panel (A), the two hidden nodes share three NH nodes but with different coupling strength. In panels (B) and (C), the hidden nodes share two or one common node(s). If the two hidden nodes do not share any nodes, the cancellation factors can be estimated independently using the same method as for the situation of one hidden node. Another extreme case is that the two hidden nodes have and only have two identical nodes, which is equivalent to the case of one hidden node.

A procedure was developed [36] to estimate the cancellation factors for the situation in Fig. 19(A). The procedure can be extended to the other two cases in a straightforward manner by setting zero the weights from the hidden nodes to nodes  $a$  and (or)  $b$ . For any of the three neighboring nodes  $i \in [a, b, c]$ , its dynamical equation can be expanded as

$$[\dot{\mathbf{x}}_i]_1 = [\sum_{\gamma} \tilde{\mathbf{a}}_i^{(\gamma)} \cdot \tilde{\mathbf{g}}_i^{(\gamma)}(\mathbf{x}_i)]_1 + [\sum_{k \neq i, H_1, H_2}^N \mathcal{W}_{ij} \cdot \mathbf{x}_j]_1 + w_{i, H_1}^{1p} \cdot [\mathbf{x}_{H_1}]_p + w_{i, H_2}^{1p} \cdot [\mathbf{x}_{H_2}]_p + \xi \eta_i. \quad (47)$$

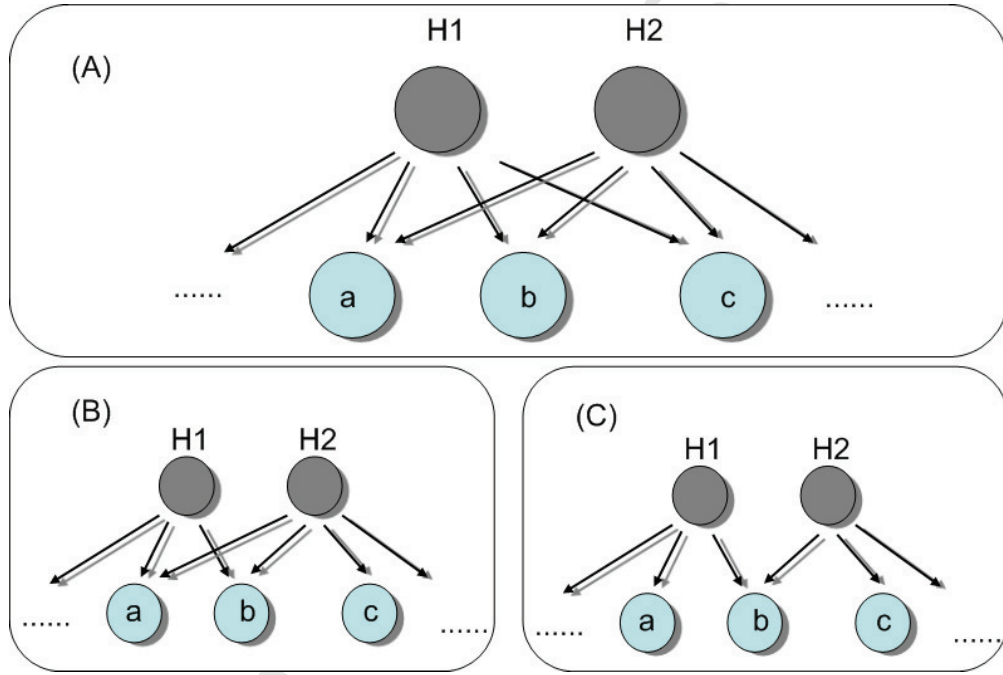


Figure 19: **Schematic illustration of configurations of two entangled hidden nodes.** The dark nodes, H1 and H2 represent the two hidden nodes, and the light blue nodes, a, b and c are accessible nodes. In all panels, each hidden node has at least two neighboring accessible nodes. In panel (A), all three accessible nodes are jointly coupled to the two hidden nodes. In panel (B), two accessible nodes are shared by the hidden nodes, while in panel (C) only one accessible node is coupled to both hidden nodes. From Ref. [36] with permission.

To cancel the hidden-node effect in one node, e.g., node  $b$ , one needs the time series of the two other non-hidden (NH) nodes,  $a$  and  $c$ , so as to cancel the coupling terms from the two hidden nodes. Let the corresponding cancellation factors be  $\Omega_{ba}$  and  $\Omega_{bc}$ . A new dynamical equation without the interferences from the hidden nodes can then be obtained:

$$\begin{aligned} [\dot{\mathbf{x}}_a]_1 &= \Omega_{ba}[\dot{\mathbf{x}}_b]_1 - \Omega_{ca}[\dot{\mathbf{x}}_c]_1 \\ &= \left[ \sum_{\gamma} \tilde{\mathbf{a}}_a^{(\gamma)} \cdot \tilde{\mathbf{g}}_a^{(\gamma)}(\mathbf{x}_a) \right]_1 - \Omega_{ba} \cdot \left[ \sum_{\gamma} \tilde{\mathbf{a}}_b^{(\gamma)} \cdot \tilde{\mathbf{g}}_b^{(\gamma)}(\mathbf{x}_b) \right]_1 - \Omega_{ca} \cdot \left[ \sum_{\gamma} \tilde{\mathbf{a}}_c^{(\gamma)} \cdot \tilde{\mathbf{g}}_c^{(\gamma)}(\mathbf{x}_c) \right]_1 \\ &+ [w_{a,H_1}^{1p} - \Omega_{ba}w_{b,H_1}^{1p} - \Omega_{ca}w_{c,H_1}^{1p}] \cdot [\mathbf{x}_{H_1}]_p + [w_{a,H_2}^{1p} - \Omega_{ba}w_{b,H_2}^{1p} - \Omega_{ca}w_{c,H_2}^{1p}] \cdot [\mathbf{x}_{H_2}]_p \\ &+ \left[ \sum_{j \neq i, H_1, H_2}^N (\mathcal{W}_{aj} - \Omega_{ba}\mathcal{W}_{bj} - \Omega_{ca}\mathcal{W}_{cj}) \cdot \mathbf{x}_j \right]_1 + \xi(\eta_a - \Omega_{ba}\eta_b - \Omega_{ca}\eta_c), \end{aligned} \quad (48)$$

where it is assumed that the coefficients associated with the coupling terms from the hidden nodes are zero. This can be achieved when the equation

$$\begin{pmatrix} w_{b,H_1} & w_{c,H_2} \\ w_{b,H_2} & w_{c,H_2} \end{pmatrix} \cdot \begin{pmatrix} \Omega_{ba} \\ \Omega_{ca} \end{pmatrix} \equiv \mathcal{M}_w \cdot \begin{pmatrix} \Omega_{ba} \\ \Omega_{ca} \end{pmatrix} = \begin{pmatrix} w_{a,H_1} \\ w_{a,H_2} \end{pmatrix} \quad (49)$$

holds and has only one trivial solution. The couplings to the hidden nodes should thus satisfy the condition  $\det(\mathcal{M}_w) = 2$ . One can then estimate the cancellation factors using Eq. (48), which is free of influence from the hidden nodes.

To demonstrate the procedure, a small chaotic Rössler oscillator network of five nodes was tested, as illustrated in Fig. 19(A). A reconstruction procedure was carried out for #b, utilizing the time series from all three NH nodes. Since the unknown coefficients are highly correlated and dense in this small system, the least squares method can be chosen for reconstruction with relative data amount  $R_m = 1.2$ . The predicted and the actual results are shown in the Fig. 20. In panel (A), terms #1 and #2 denote the two cancellation factors. The nodal dynamics for node #b, #a and node #c are listed in order. Panel (B) shows the entangled couplings in the network. All predicted terms match well with the actual ones.

When an NH node is coupled with  $K$  ( $K \geq 3$ ) hidden nodes, their successful detection requires that every hidden node be connected with two or more NH nodes. Then  $K$  cancellation factors can be estimated when the coupling weights satisfy the condition  $\det(\mathcal{M}_w) = K$ , where the elements  $m_{ih}$  in  $\mathcal{M}_w$  correspond to the coupling terms from the  $h$ th hidden node to the  $i$ th NH node.

### 3.4 Identifying chaotic elements in neuronal networks

Compressive sensing can be exploited to identify a subset of chaotic elements embedded in a network of nonlinear oscillators from time series. The oscillators, when isolated, are not identical in that their parameters are different, so dynamically they can be in distinct regimes. For example, all oscillators can be described by differential equations of the same mathematical form, but with different parameters. Consider the situation where only a small subset of the oscillators are chaotic, and the remaining oscillators are in dynamical regimes of regular oscillations. Due to the mutual couplings among the oscillators, the measured time series from most oscillators would appear random. The challenge is to identify the small subset of originally (“truly”) chaotic oscillators.

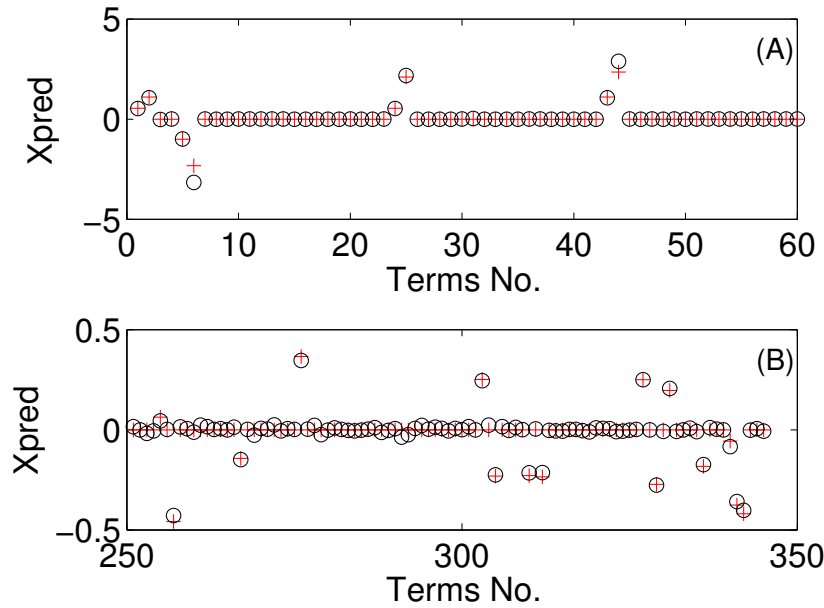


Figure 20: **Demonstration of detection of multiple entangled hidden nodes.** For a small oscillator network with two entangled hidden nodes, comparison of the predicted coefficients (black circles) and the actual ones (red crosses) obtained by applying the method of cancellation factor. In panel (A), coefficients from No. 1 to No. 60 correspond to these for the dynamical equations. Panel (B) specifies the predicted results for the coupling functions. From Ref. [36] with permission.

The problem of identifying chaotic elements from a network of coupled oscillators arises in biological systems and biomedical applications. For example, for a network of coupled neurons that exhibit regular oscillations in a normal state, the parameters of each isolated neuron are in regular regime. Under external perturbation or slow environmental influences the parameters of some neurons can drift into a chaotic regime. When this occurs the whole network would appear to behave chaotically, which may correspond to certain disease. The coupling and nonlinearity stipulate that irregular oscillations at the network level can emerge even if only a few oscillators have gone “bad.” It is thus desirable to be able to pin down the origin of the ill-behaved oscillators - the few chaotic neurons among a large number of healthy ones.

One might attempt to use the traditional approach of time-delayed coordinate embedding to reconstruct the phase space of the underlying dynamical system [43, 44, 46] and then to compute the Lyapunov exponents [79, 82]. However, for a network of nonlinear oscillators, the phase-space dimension is high and an estimate of the largest Lyapunov exponent would only indicate if the *whole* coupled system is chaotic or nonchaotic, depending on the sign of the estimated exponent. In principle, using time series from any specific oscillator(s) would give qualitatively the same result. Thus, the traditional approach cannot give an answer as to which oscillators are chaotic when isolated.

Recently, a compressive sensing based method was developed [35] to address the problem of identifying a subset of ill-behaved chaotic elements from a network of nonlinear oscillators, majority of them being regular. In particular, for a network of coupled, mixed nonchaotic and chaotic neurons, it was demonstrated that, by formulating the reconstruction task as a compressive

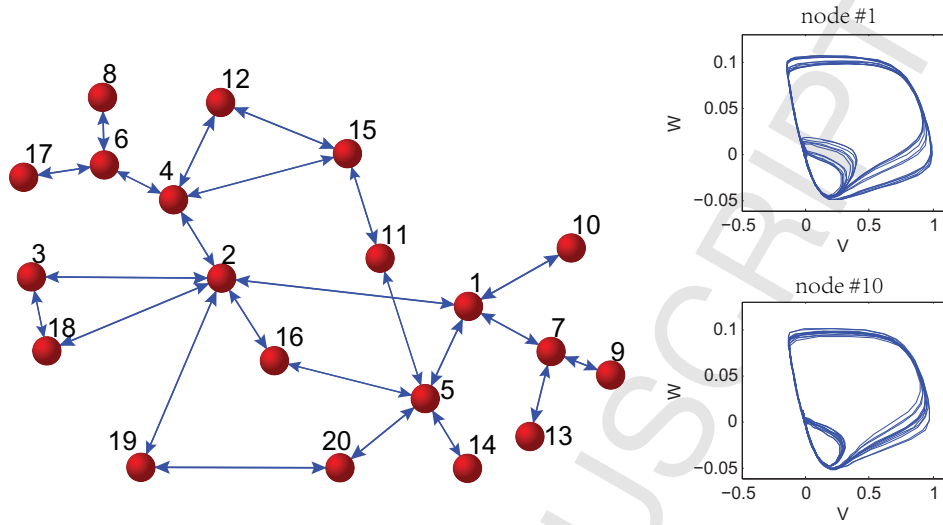


Figure 21: **A small neuronal network in which most dynamical units are regular and only a few are chaotic.** (a) Schematic illustration of the network, where the dynamics of each neuron is mathematically described by the Fitzhugh-Nagumo (FHN) equations. (b,c) Dynamical trajectories of two neurons from the coupled system, one being chaotic when isolated and another regular, respectively. The trajectories give little hint as to which one is originally chaotic and which one is regular, due to the coupling. Specifically, neuron #1 is originally chaotic (by setting parameter  $a = 0.42$  in the FHN equation), while all other neurons are regular (their values of the corresponding parameter in the FHN equation are chosen uniformly from the interval  $[0.43, 0.45]$ ). From Ref. [35] with permission.

sensing problem, the system equations and the coupling functions as well as all the parameters can be obtained accurately from sparse time series. Using the reconstructed system equations and parameters for each and every neuron in the network and setting all the coupling parameters to zero, a routine calculation of the largest Lyapunov exponent can unequivocally distinguish the chaotic neurons from the nonchaotic ones.

### 3.4.1 Basic procedure of identifying chaotic neurons

Figure 21(a) shows schematically a representative coupled neuronal network. Consider a pair of neurons, one chaotic and another nonchaotic when isolated (say #1 and #10, respectively). When they are embedded in a network, due to coupling, the time series collected from both will appear random and qualitatively similar, as shown in Figs. 21(b) and 21(c). It is visually quite difficult to distinguish the time series and to ascertain which node is originally chaotic and which is regular. The difficulty is compounded by the fact that the detailed coupling scheme is not known *a priori*. Suppose that the chaotic behavior leads to undesirable function of the network and is to be suppressed. A viable and efficient method is to apply small pinning controls [246–249] to the relatively few chaotic neurons to drive them into some regular regime. (An implicit assumption is that, when all neurons are regular, the collective dynamics is regular. That is, the uncommon but not unlikely situation that a network systems of coupled regular oscillators would exhibit chaotic behaviors is excluded.) Accurate identification of the chaotic neurons is thus key to implementing the pinning control strategy.

Given a neuronal network, the task is thus to locate all neurons that are *originally* chaotic and neurons that are *potentially* likely to enter into a chaotic regime when they are isolated from the other neurons or when the couplings among the neurons are weakened. The compressive sensing based approach [35] consists of two steps. Firstly, the framework is employed to estimate, from measured time series only, the parameters in the FHN equation for each neuron, as well as the network topology and various coupling functions and weights. This can be done by expanding the nodal dynamical equations and the coupling functions into some suitable mathematical base as determined by the specific knowledge about the actual neuronal dynamical system, and then casting the problem into that of determining the sparse coefficients associated with various terms in the expansion. The nonlinear systems identification problem can then be solved using the standard compressive sensing algorithm. Secondly, all coupling parameters are set to zero so that the dynamical behaviors of each and every *individual* neuron can be analyzed by calculating the Lyapunov exponents. These with a positive largest exponent are identified as chaotic.

A typical time series from a neuronal network consists of a sequence of spikes in the time evolution of the cell membrane potential. It was demonstrated [35] that the compressive sensing based reconstruction method works well even for such spiky time series. The dependence of reconstruction accuracy on data amount was analyzed to verify that only limited data are required to achieve high accuracy in reconstruction.

### 3.4.2 Example: identifying chaotic neurons in the FitzHugh-Nagumo (FHN) network

The FHN model, a simplified version of the biophysically detailed Hodgkin-Huxley model [250], is a mathematical paradigm for gaining significant insights into a variety of dynamical behaviors in real neuronal systems [251, 252]. For a single, isolated neuron, the corresponding dynamical system is described by the following two-dimensional, nonlinear ordinary differential equations:

$$\begin{aligned}\frac{dV}{dt} &= \frac{1}{\delta}[V(V-a)(1-V)-W], \\ \frac{dW}{dt} &= V-W-b+S(t),\end{aligned}\tag{50}$$

where  $V$  is the membrane potential,  $W$  is the recover variable,  $S(t)$  is the driving signal (e.g., periodic signal),  $a$ ,  $b$ , and  $\delta$  are parameters. The parameter  $\delta$  is chosen to be infinitesimal so that  $V(t)$  and  $W(t)$  are “fast” and “slow” variables, respectively. Because of the explicitly time-dependent driving signal  $S(t)$ , Eq. (50) is effectively a three-dimensional dynamical system, in which chaos can arise [155]. For a network of FHN neurons, the equations are

$$\begin{aligned}\frac{dV_i}{dt} &= \frac{1}{\delta}[V_i(V_i-a)(1-V_i)-W_i] + \sum_{j=1}^N c_{ij}(V_j-V_i) \\ \frac{dW_i}{dt} &= V_i-W_i-b+S(t),\end{aligned}\tag{51}$$

where  $c_{ij}$  is the coupling strength (weight) between the  $i$ th and the  $j$ th neurons (nodes). For  $c_{ij} = c_{ji}$ , the interactions between any pair of neurons are symmetric, leading to a symmetric adjacency matrix for the network. For  $c_{ij} \neq c_{ji}$ , the network is asymmetrically weighted.

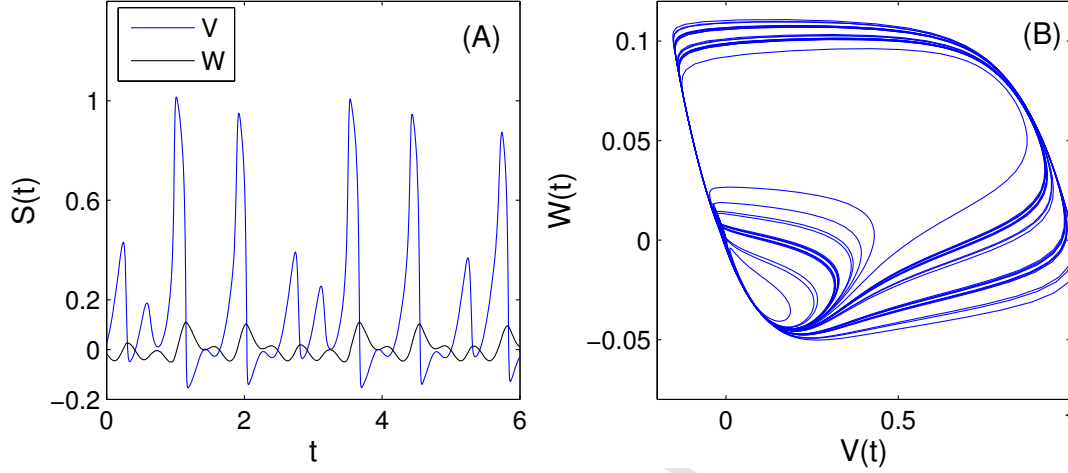


Figure 22: **Dynamical behavior of a chaotic neuron.** (a) Chaotic time series of the membrane potential  $V$  and the recovery variable  $W$  from a single neuron for  $a = 0.42$ , and (b) the corresponding dynamical trajectory. From Ref. [35] with permission.

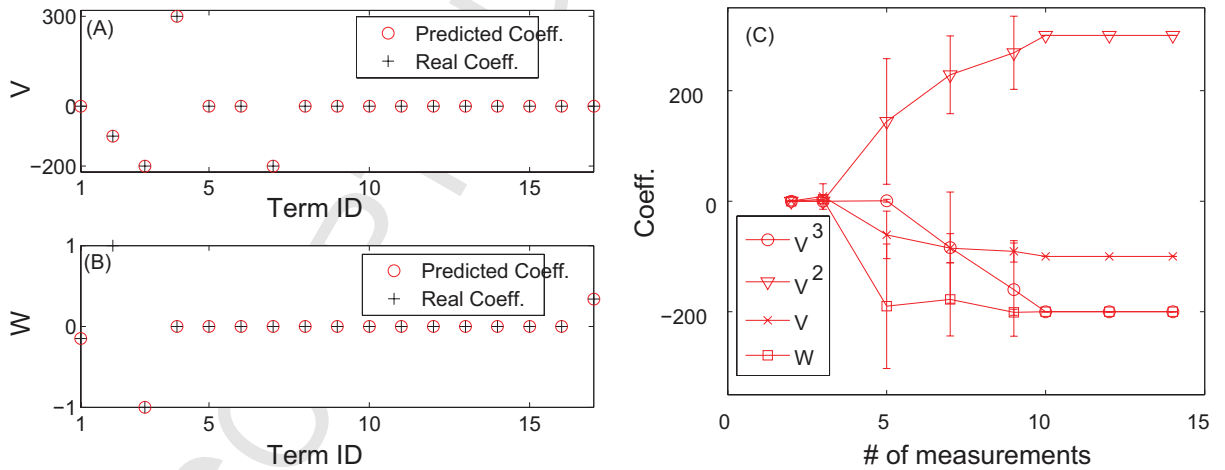


Figure 23: **Identification of parameters of a single FHN neuron using compressive sensing.** (a,b) Predicted coefficients and comparison with the actual parameter values in the dynamical equations of variables  $V$  and  $W$ . The number of data points used was 12. (c) Predicted parameters for a single neuron as the number of data points is increased. The sampling interval is  $\Delta t = 0.05$ . All results are averaged over 10 independent time series. From Ref. [35] with permission.

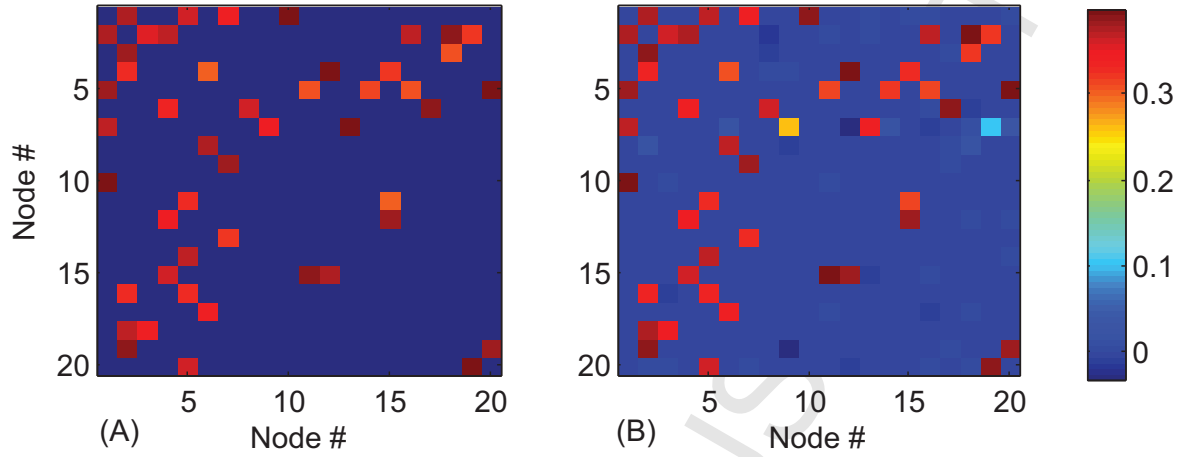


Figure 24: **Reconstruction of an FHN network.** For the network in Fig. 21(a), (a) actual and (b) estimated weighted adjacency matrix. The normalized data amount used in the reconstruction is  $R_m = 0.7$ . From Ref. [35] with permission.

Consider the FHN model with sinusoidal driving:  $S(t) = r \sin \omega_0 t$ . The model parameters are  $r = 0.32$ ,  $\omega_0 = 15.0$ ,  $\delta = 0.005$ , and  $b = 0.15$ . For  $a = 0.42$ , an individual neuron exhibits chaos. Representative chaotic time series and the corresponding dynamical trajectory are shown in Fig. 22. Reconstruction of an isolated neuron can be done by setting zero all coupling terms in network. For this purpose power series of order 4 can be used as the expansion base [35] so that there are 17 unknown coefficients to be determined. Three consecutive measurements are sampled at time interval  $\tau = 0.05$  apart and a standard two-point formula can be used to extrapolate the derivatives. From a random starting point, 12 data points were generated. Results of reconstruction are shown in Figs. 23(a) and 23(b) for variables  $V$  and  $W$ , respectively. The last two coefficients associated with each variable represent the strength of the driving signal. Since only the variable  $W$  receives a sinusoidal input, the last coefficient in  $W$  is nonzero. By comparing the positions of the nonzero terms and the previously assumed vector form  $\mathbf{g}_i(t)$ , one can fully reconstruct the dynamical equations of any isolated neuron. In particular, from Figs. 23(a) and 23(b) it can be seen that all estimated coefficients agree with their respective true values. Figure 23(c) shows how the estimated coefficients converge to the true values as the number of data points is increased. It can be seen that, for over 10 data points, all the parameters associated with a single FHN neuron can be faithfully identified.

Next consider the network of coupled FHN neurons as schematically shown in Fig. 21(a), where the coupling weights among various pairs of nodes are uniformly distributed in the interval  $[0.3, 0.4]$ . The network topology is random with connection probability  $p = 0.04$ . From time series the compressive sensing matrix for each variable of all nodes can be reconstructed. Since the couplings occur among the variables  $V$  of different neurons, the strengths of all incoming links can be found in the unknown coefficients associated with different  $V$  variables. Extracting all coupling terms from the estimated coefficients gives all off-diagonal terms in the weighted adjacency matrix. Figure 24 shows the reconstructed adjacency matrix as compared with the real one for  $R_m = 0.7$ , where  $R_m$  is the relative number of data points normalized by the total number



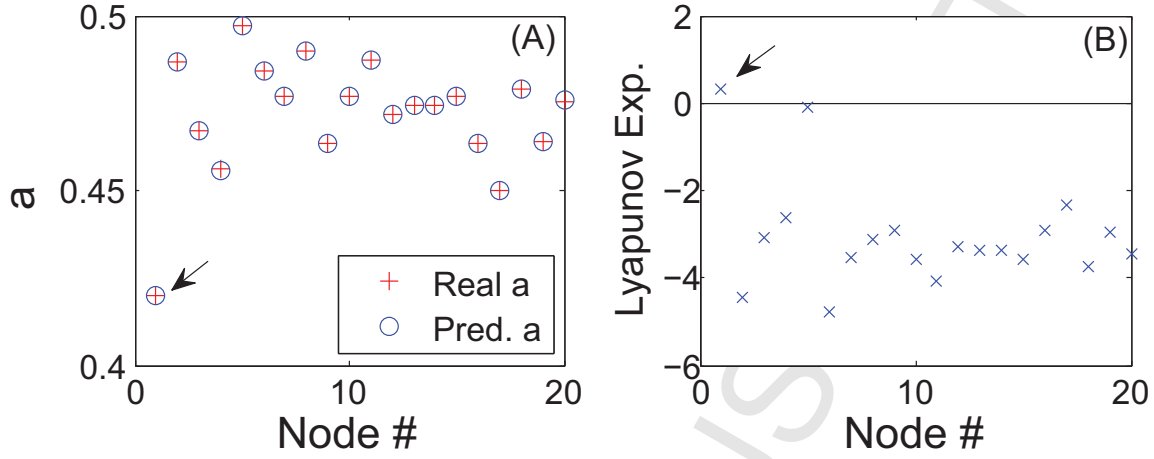


Figure 25: **Identification of chaotic neurons.** (a) Estimated values of the parameter  $a$  for different neurons (red circles), as compared with the actual values (black crosses). The random network size is  $N = 20$  with connection probability  $p = 0.04$ . The normalized data amount used in reconstruction is  $R_m = 0.7$ . (b) Largest Lyapunov exponents calculated from the reconstructed system equations. The reference line denotes null value. From Ref. [35] with permission.

of unknown coefficients. It can be seen that the compressive sensing based method can predict all links correctly, in spite of the small errors in the predicted weight values. The errors are mainly due to the fact that there are large coefficients in the system equations but the coupling weights are small.

Using the weighted adjacency matrix, one can identify the coupling terms in the network vector function so as to extract the terms associated with each isolated nodal velocity field. The value of parameter  $a$  can then be identified and the largest Lyapunov exponent can be calculated for each individual neuron. The results are shown in Figs. 25(a) and 25(b). It can be seen that, for this example, neuron #1 has a positive largest exponent while the largest exponents for all others are negative, so #1 is identified as the only chaotic neuron among all neurons in the network.

### 3.5 Data based reconstruction of complex geospatial networks and nodal positioning

Complex geospatial networks with components distributed in the real geophysical space are an important part of the modern infrastructure. Examples include large scale sensor networks and various subnetworks embedded in the Internet. For such a network, often the set of active nodes depends on time: the network can be regarded as static only in relatively short time scale. For example, in response to certain breaking news event, a social communication network within the Internet may emerge, but the network will dissolve itself after the event and its impacts fade away. The connection topologies of such networks are usually unknown but in certain applications it is desirable to uncover the network topology and to *determine the physical locations of various nodes* in the network. Suppose time series or signals can be collected from the nodes. Due to the

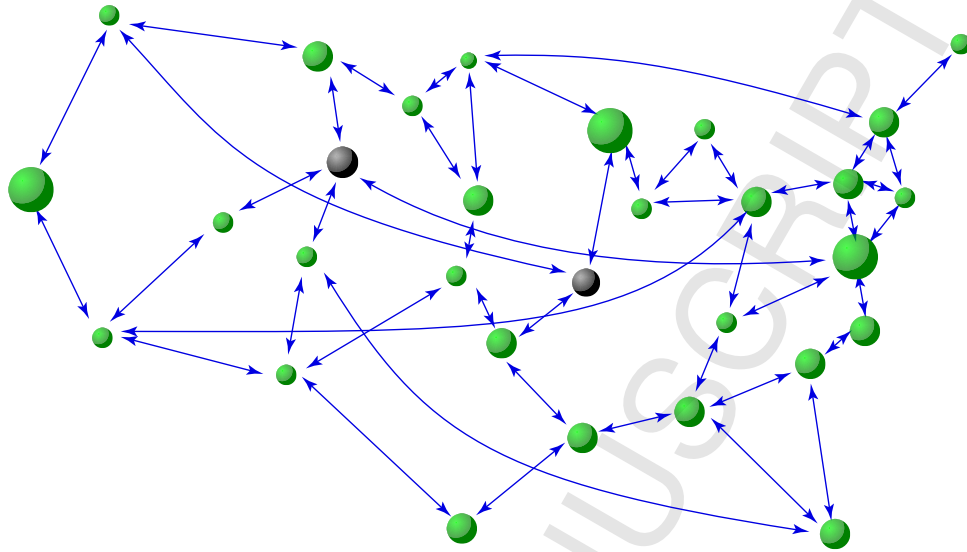


Figure 26: **A schematic illustration of a complex geospatial network.** The connection topology, the positions of the nodes in the physical space, and nodal dynamical equations all are unknown *a priori*, but only time series from the nodes can be collected at a single node in the network (e.g., a data collecting center). The challenges are to reconstruct the dynamical network, to locate the precise position of each node, and to detect hidden nodes, all based solely on time series with inhomogeneous time delays. The green circles denote “normal” nodes and the dark circles indicate hidden nodes. From Ref. [38] with permission.

distributed physical locations of the nodes, the signals are time delayed. An illustrative example of a complex geospatial network is shown in Fig. 26, where there is a monitoring center that collects data from nodes at various locations, but their precise geospatial coordinates are unknown. The normal nodes are colored in green. There are also hidden nodes that can potentially be the sources of threats (e.g., those represented by dark circles). Is it possible to uncover the network topology, estimate the time delays embedded in the signals from different nodes, and then determine their physical locations? Can the existence of a hidden node be ascertained and its actual geophysical location be determined?

A recent work [38] showed that these questions can be addressed by using the compressive sensing based reconstruction paradigm. In particular, the time delays of the dynamics at various nodes can be estimated using time series collected from *a single location*. Note that there were previous methods of finding time delays in complex dynamical systems, e.g., those based on synchronization [253], Bayesian estimation [254], and correlation between noisy signals [27]. The compressive sensing based method provides an alternative approach that has the advantages of generality, high efficiency, low data requirement, and applicability to large networks. It was demonstrated [38] that the method can yield estimates of the nodal time delays with reasonable accuracy. After the time delays are obtained, the actual geospatial locations of various nodes can be determined by using, e.g., a standard triangular localization method [255], given that the locations of a small subset of nodes are known. Hidden nodes can also be detected. These results can potentially be useful for applications such as locating sensors in wireless networks and identifying/detecting/anticipating potential geospatial threats [256], an area of importance and broad

interest.

**Reconstruction of time delays using compressive sensing.** Consider a continuous-time oscillator network with time delayed couplings [257–259], where for every link, the amount of delay is proportional to the physical distance of this link. The time delayed oscillator network model has been widely used in studying neuronal activities [258] and food webs in ecology [259]. Mathematically, the system can be written as

$$\dot{\mathbf{x}}_i = \mathbf{F}_i[\mathbf{x}_i(t)] + \sum_{j=1, j \neq i}^N \mathcal{W}_{ij} \cdot [\mathbf{x}_j(t - \tau_{ij}) - \mathbf{x}_i(t)], \quad (52)$$

for  $i = 1, \dots, N$ , where  $\mathbf{x}_i \in \mathbb{R}^m$  is the  $m$ -dimensional state variable of node  $i$  and  $\mathbf{F}_i[\mathbf{x}_i(t)]$  is the vector field for its isolated nonlinear nodal dynamics. For a link  $l_{ij}$  connecting nodes  $i$  and  $j$ , the interaction weight is given by the  $m \times m$  weight matrix  $\mathcal{W}_{ij} \in \mathbb{R}^{m \times m}$  with its element  $w_{ij}^{p,q}$  representing the coupling from the  $q$ th component of node  $j$  to the  $p$ th component of node  $i$ . For simplicity, we assume only one component of  $w_{ij}^{p,q}$  is non-zero and denotes it as  $w_{ij}$ . The associated time delay is denoted as  $\tau_{ij}$ . For a modern geospatial network, the speed of signal propagation is quite high in a proper medium (e.g., optical fiber). The time delay can thus be assumed to be small and the Taylor expansion can be used to express the delay coupling terms to the first order, e.g.,  $x_i(t - \tau_{ji}) \approx x_i(t) - \tau_{ji}\dot{x}_i$ , where  $\dot{x}_i$  is the time derivative. When the coupling function between any pair of nodes is linear, in a suitable mathematical basis constructed from the time series data,  $\tilde{\mathbf{g}}^{(\gamma)}[\mathbf{x}_i(t)]$ , the coupling and time delayed terms, together with the nodal dynamical equations, can be expanded into a series. The task is to estimate all the expansion coefficients. Assume linear coupling functions and causality so that all  $\tau_{ij}$  ( $i, j = 1, \dots, N$ ) are positive (for simplicity). All terms directly associated with node  $i$  can be regrouped into  $\mathbf{F}'_i[\mathbf{x}_i(t)]$ , where

$$\mathbf{F}'_i[\mathbf{x}_i(t)] \equiv \mathbf{F}_i[\mathbf{x}_i(t)] - \mathbf{x}_i(t) \cdot \sum_{j=1, j \neq i}^N \mathcal{W}_{ij}, \quad (53)$$

and  $\mathbf{F}'_i[\mathbf{x}_i(t)]$  has been expanded into the following series form:

$$\mathbf{F}'_i[\mathbf{x}_i(t)] = \sum_{\gamma} \tilde{\alpha}^{(\gamma)} \cdot \tilde{\mathbf{g}}^{(\gamma)}[\mathbf{x}_i(t)], \quad (54)$$

with  $\tilde{\mathbf{g}}^{(\gamma)}[\mathbf{x}_i(t)]$  representing a suitably chosen set of orthogonal and complete base functions so that the coefficients  $\tilde{\alpha}^{(\gamma)}$  are sparse. The time delayed variable  $\mathbf{x}_j(t - \tau_{ij})$  can be expanded as

$$\mathbf{x}_j(t - \tau_{ij}) \approx \mathbf{x}_j(t) - \tau_{ij}\dot{\mathbf{x}}_j(t). \quad (55)$$

All the coupling terms with inhomogeneous time delays associated with node  $i$  can then be written as

$$\left[ \sum_{j=1, j \neq i}^N \mathcal{W}_{ij} \mathbf{x}_j(t - \tau_{ij}) \right]_p \equiv \sum_{j=1, j \neq i}^N [\mathcal{B}_{ij} \cdot \mathbf{x}_j(t) + C_{ij} \cdot \dot{\mathbf{x}}_j(t)], \quad (56)$$

where  $\mathcal{B}_{ij} = \mathcal{W}_{ij}$  and  $C_{ij} = -\mathcal{W}_{ij}\tau_{ij}$ . In the compressive sensing framework, Eq. (52) can then be written in the following compact form:

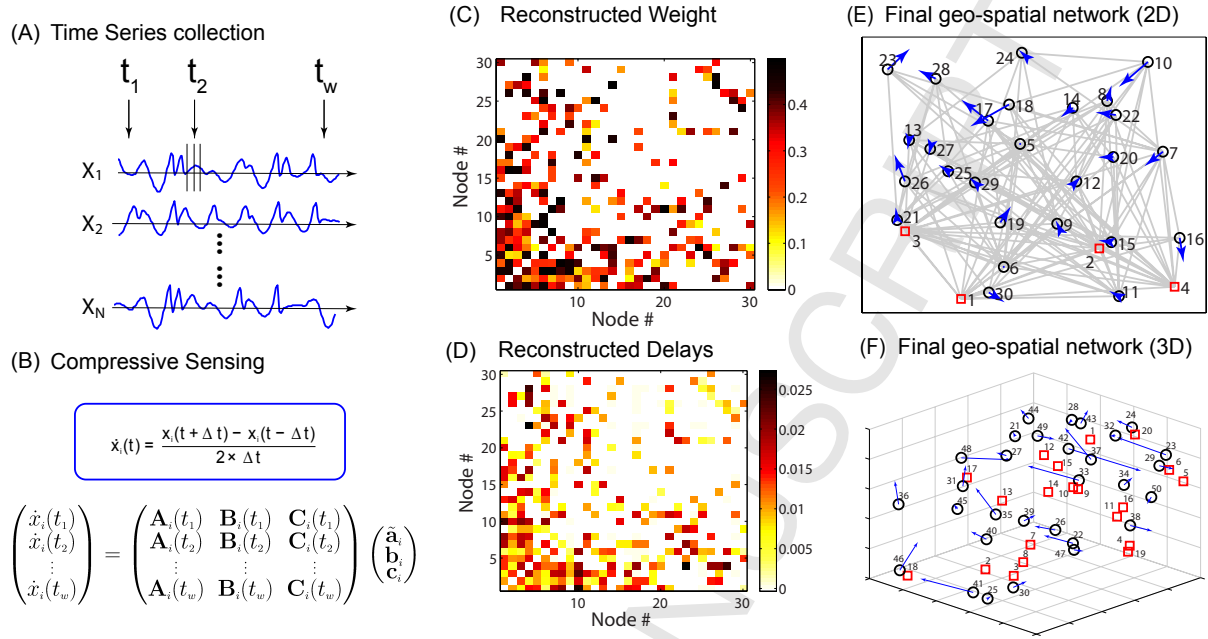
$$\dot{\mathbf{x}}_i(t) = \sum_{\gamma} \tilde{\alpha}^{(\gamma)} \cdot \tilde{\mathbf{g}}^{(\gamma)}[\mathbf{x}_i(t)] + \sum_{j=1, j \neq i}^N [\mathcal{B}_{ij} \cdot \mathbf{x}_j(t) + C_{ij} \cdot \dot{\mathbf{x}}_j(t)], \quad (57)$$

which is a set of linear equations for data collected at different time  $t$ , where  $\tilde{\alpha}^{(\gamma)}$ ,  $\mathcal{B}_{ij}$  and  $C_{ij}$  are to be determined. If the unknown coefficient vectors can be reconstructed accurately, one has complete information about the nodal dynamics as represented by  $\mathbf{F}'[\mathbf{x}(t)]$ , the topology and interacting weights of the underlying network as represented by  $\mathcal{W}_{ij}$ , as well as the time delays associated with the nonzero links because of the relations  $\mathcal{W}_{ij} = \mathcal{B}_{ij}$ . Note that, if the coupling form is nonlinear, the relationship between the delay term  $\mathbf{c}_{ij}$  and the time delays  $\tau_{ij}$  would be hard to interpret, especially when the exact coupling form is not known.

**Nodal positioning and reconstruction of geospatial network.** After obtaining the time delays, one can proceed to determining the actual positions of all nodes. If time series are collected simultaneously from all nodes at the data collecting node, the estimated coupling delay  $\tau_{ij}$  associated with the link  $l_{ij}$  is proportional to the physical distance  $d_{ij} = d_{ji}$  of the link. However, in reality strictly synchronous data collection is not possible. For example, if the signals are collected, e.g., at a location  $s$  outside the network with varying time delays  $\tau_{si}$ , the estimated delays associated with various links in the network are no longer proportional to the actual distances. The varying delays due to asynchronous data collection can be canceled and the distances can still be estimated as  $d_{ij} = (c/2)(\tau_{ij} + \tau_{ji})$ , where  $\tau_{ij}$  is the signal delay associated with node  $j$  from the reconstruction of node  $i$ , vice versa for  $\tau_{ji}$ , and  $c$  is the signal propagation speed.

When the mutual distances between the nodes have been estimated, one can determine their actual locations, e.g., by using the standard triangular localization algorithm [255]. This method requires that the positions of  $N_B$  reference nodes be known, the so-called beacon nodes. Starting from the beacon nodes, the triangulation algorithm makes use of the distances to these reference nodes to calculate the Cartesian coordinates of the detected nodes. The beacon node set can then be expanded with the newly located nodes. Nodes that are connected to the new beacon set, each with more than three links in the two-dimensional space, can be located. The process continues until the locations of all nodes have been determined, or no new nodes can be located. The choice of the proper initial beacon set to fully reconstruct the network depends on the network topology. For example, for a scale-free network, one can choose nodes that were firstly added during the process of network generation as the initial beacon set, thereby guaranteeing that all nodes can be located using the procedure. An empirical rule is then to designate the largest degree nodes as the initial beacon node set.

More specifically, given the positions of  $k$  reference nodes (or beacon nodes)  $(x_k, y_k)$ , and their distances  $d_{i,1}, d_{i,2}, \dots, d_{i,k}$  to the target node  $i$ , one can calculate the position of node  $i$  using the triangular localization method [255], for  $k$  larger than the space dimension. In general, it is necessary to solve the least squares optimization problem  $\mathcal{H} \cdot \mathbf{x}_i = \mathbf{b}$ , where  $\mathbf{x}_i = [x_i, y_i]^T$  is the position of node  $i$ , and  $\mathcal{H} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k]^T$  is the position matrix corresponding to the set of beacon nodes, where  $\mathbf{b} = 0.5 \times [D_1, D_2, \dots, D_k]^T$  and  $D_k = d_{ik}^2 - y_k^2 + x_k^2$ . To locate the positions of all nodes in the network, one starts with a small set of beacon nodes whose actual positions are known. Initially one can locate the nodes that are connected to at least three nodes in the set of beacon nodes, insofar as the three reference nodes are not located on a straight line. When this is done, the newly located nodes can be added into the set of reference nodes and the neighboring nodes can be located through the new set of beacon nodes. This process can be iterated until the positions of all nodes are determined or no more qualified neighboring nodes can be found. For a general network, such an initial beacon set may not be easily found. A special case is



**Figure 27: Illustration of compressive sensing based method to reconstruct a complex geospatial network from time series.** (A) For any node, time series of its dynamical variables are collected at  $w$  different instants of time. (B) The corresponding derivatives are approximated using the standard first order Gaussian method, which are needed in constructing the compressive sensing equations. (C,D) An example of link weights and time delays obtained from the reconstructed coefficient vectors, respectively. (E) Given the positions of four beacon nodes (marked as red rectangles), the locations of the remaining nodes (marked as black circles) are determined by using the standard triangularization method. The blue arrows indicate the estimation errors, which point from the actual to the estimated positions. The various coupling terms are illustrated using gray lines. There are in total 30 nodes in the network, connecting with each other via the scale-free topology. The average outgoing degree is five. The amount of data used is  $R_m = 0.5$ . (F) Following the same procedure, a geospatial network of 50 nodes which are distributed in a three-dimensional cube of unit length is reconstructed. There are 20 beacon nodes and the normalized amount of data used is  $R_m = 0.55$ . For clarity, the links between these nodes are not shown. From Ref. [38] with permission.

scale-free networks, for which the initial beacon set can be chosen as the nodes with the largest degrees. For a random network, one can also choose the nodes of the largest degree as the initial beacon node set, and use a larger beacon set to locate most of the nodes in the network. For an arbitrary network topology, the following simple method was proposed [38] to select the set of beacon nodes: estimate the distances from one node to all other unconnected nodes using the weighted shortest distance and then proceed with the triangular localization algorithm. There are alternative localization algorithms based on given distances, e.g, the multidimensional scaling method [260, 261].

**An example of reconstruction of a complex geospatial network.** A numerical example was presented in Ref. [38] to demonstrate the reconstruction of complex geospatial networks, in which all nodes were assumed to be distributed in a two-dimensional square, or a three-dimensional cube of unit length. The network topology was scale free [197] or random [196], and its size was

varied. A nonlinear oscillator was placed at each node, e.g., the Rössler oscillator ( $[\dot{x}, \dot{y}, \dot{z}] = [-y - z, x + 0.2y, 0.2x + z(x - 0.2)]$ ). The coupling weights were asymmetric and uniformly distributed in the interval  $[0.1, 0.5]$ . A small threshold was assigned to the estimated weight as  $w_0 = 0.05$  (somewhat arbitrary), where if the estimated weight is larger (smaller) than  $w_0$ , the corresponding link is regarded as existent (nonexistent). As a result, the following holds:  $d_{ij} = c \cdot \tau_{ij} = c \cdot \tau_{ji}$ , and the parameter  $c$  was chosen to be 100 (arbitrarily). Linear coupling functions were chosen for any pair of connected nodes, where the interaction occurs between the  $z$ -variable of one node and the  $x$ -variable of another. The time series used to reconstruct the whole network system were acquired by integrating the coupled delayed differential-equation system [262] with step size  $5 \times 10^{-5}$ . The vector fields of the nodal dynamics were expanded into a power series of order  $l_x + l_y + l_z \leq 3$ . The derivatives required for the compressive sensing formulation were approximated from time series by the standard first order Gaussian method. The data requirement was characterized by  $R_m$ , the ratio of the number of data points used to the total number of unknown coefficients to be estimated. The beacon nodes were chosen to be those having the largest degrees in the network, and their positions were assumed to be known.

Figure 27 summarizes the major steps required for reconstructing a complex geospatial network using compressive sensing, where  $N = 30$  nodes connecting with each other in a scale-free manner are randomly distributed in a two-dimensional square. Oscillatory time series are collected from each node, from which the compressive sensing equations can be obtained, as shown in Figs. 27(A) and 27(B). The reconstructed coefficients for the nodal dynamical equations contain the coupling weights  $B_{ij} = w_{ij}$  and the delay terms  $C_{ij} = -w_{ij} \times \tau_{ij}$ . The links with reconstructed weights larger than the threshold  $w_0$  are regarded as actual (existent) links, for which the time delays  $\tau_{ij}$  can be estimated as  $\tau_{ij} = -C_{ij}/w_{ij}$ . Repeating this procedure for all nodes, the weighted adjacency matrix (which defines the network topology) and the time delay matrix can be determined. The estimated adjacency matrix and the time delays are displayed in Figs. 27(C) and 27(D), respectively, which match well with those of the actual network. Note that the reconstructed time delays are symmetric with respect to the link directions, as shown in Fig. 27(D), which is correct as they depend only on the corresponding physical distances. With the estimated time delays, the four largest degree nodes, node #1 ~ #4, are chosen as the beacon nodes, so that the locations of all remaining nodes can be determined. The fully reconstructed geospatial network is shown in Fig. 27(E), where the red rectangles indicate the locations of the beacon nodes. The black circles denote the actual locations of the remaining nodes and the heads of the blue arrows indicate their estimated positions (shorter arrows mean higher estimation accuracy). The amount of data used is relatively small:  $R_m = 0.5$ .

A detailed performance analysis was provided in Ref. [38] where the issue of positioning accuracy was addressed. Specifically, to locate all nodes in a two-dimensional space requires knowledge of the positions of at least three nodes (minimally four nodes in the three-dimensional space). Due to noise, the required number of beacon nodes will generally be larger. Since node positioning is based on time delays estimated from compressive sensing, which contain errors, the number of required beacon nodes is larger than three even in two dimensions. The positioning accuracy can be quantified [38] by using the normalized error  $M_r$ , defined as the medium distance error between the estimated and actual locations for all nodes (except the beacon nodes), normalized by the distributed length  $L$ . Figure 28 shows  $M_r$  versus the fraction  $R_B$  of the beacon nodes. The reconstruction parameters are chosen such that the errors in the time delay estimation is  $D_{nz} \approx 0.12$ . For small values of  $R_B$ , the positioning errors are large. Reasonable positioning errors are obtained

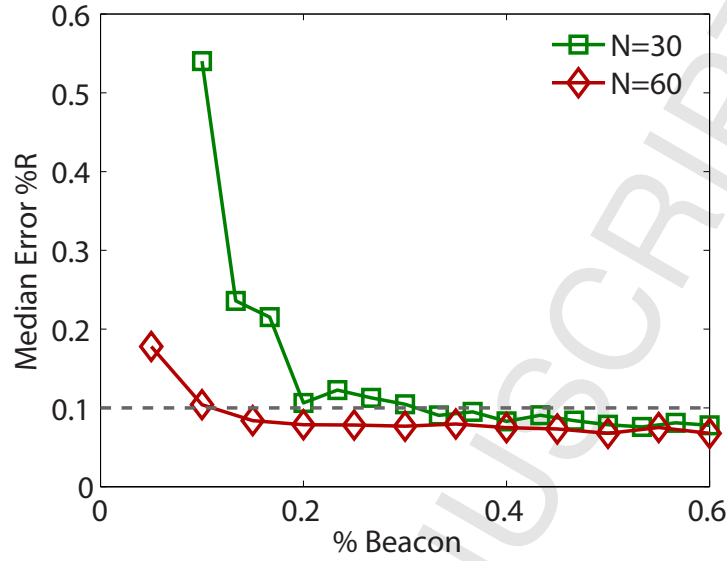


Figure 28: **Positioning errors.** Normalized positioning error  $M_r$ , defined as the medium absolute estimated distance error normalized by the distributed length  $L$ , as a function of the fraction  $R_B$  of the beacon nodes. The networks have the scale-free topology with the average outgoing degree  $k = 5$ . Two values of the network size are used:  $N = 30$  and  $N = 60$ . The beacon nodes are chosen as these having the largest degrees. The time delays are estimated using the data amount  $R_m = 0.5$ , for which the average error is  $D_{nz} \approx 0.12$ . The results were obtained by averaging over 10 independent network realizations. From Ref. [38] with permission.

when  $R_B$  exceeds, say, 0.2.

The compressive sensing based approach can be used to ascertain the existence of a hidden node and to estimate its physical location in a complex geospatial network [38]. To detect a hidden node, it is necessary to identify its neighboring nodes [31, 36]. For an externally accessible node, if there is a hidden node in its neighborhood, the corresponding entry in the reconstructed adjacency matrix will exhibit an abnormally dense pattern or contain meaningless values. In addition, the estimated coefficients for the dynamical and coupling functions of such an abnormal node typically exhibit much larger variations when different data segments are used, in comparison with those associated with normal nodes that do not have hidden nodes in their neighborhoods. The mathematical formulation of the method to uncover a hidden node in complex geospatial networks can be found in Ref. [38]. For the network of size  $N = 30$  in Fig. 28, initially, there are only 29 time series, one from each of the normal node, and it is not known *a priori* that there would be a hidden node in the network. The network was reconstructed to obtain the estimated weights and time delays, as shown in Fig. 29(A). One can see that the connection patterns of some nodes are relatively dense and the values of the weights and time delays are meaningless (e.g., negative values), giving the first clue that these nodes may be the neighboring nodes of a hidden node. To confirm that this is indeed the case, the available time series are divided into a number of segments based on the criterion that the data requirement for reconstruction is satisfied for each segment. As shown in Fig. 29(B), extraordinarily large variances in the estimated coefficients associated with the abnormal nodes arise. Combining results from Figs. 29(A) and 29(B), one can claim with confidence



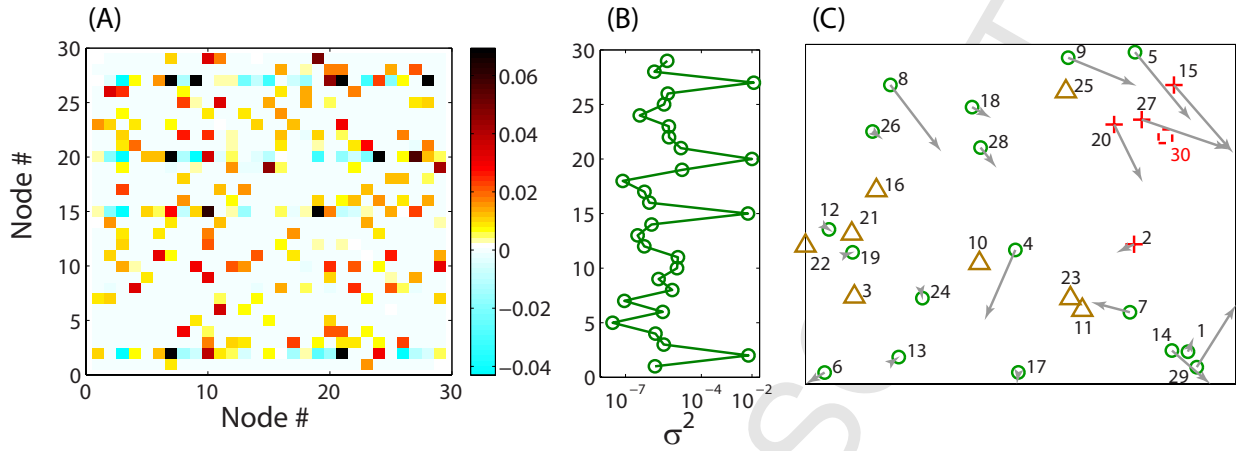


Figure 29: **Detection of hidden nodes in geospatial networks.** For a random network of  $N = 30$  nodes, illustration of detecting a hidden node (#30). (A) Reconstructed time delays using time series from 29 externally accessible nodes. (B) Average variance in the reconstructed incoming coupling delays calculated from different segments of the available time series. (C) Estimated positions of all accessible nodes in comparison with the respective actual positions, and the location of the hidden node. The triangles denote the beacon nodes, whose positions are known *a priori*. The green circles denote “normal” nodes without any hidden node in their immediate neighborhoods, while the crosses are direct neighbors of the hidden node. The actual position of the hidden node #30 is marked as a dashed square. From Ref. [38] with permission.

that the four nodes are indeed in the immediate neighborhood of a hidden node, ascertaining its existence in the network. The method also works if there are more than one hidden node, given that they do not share common neighboring nodes.

### 3.6 Reconstruction of complex spreading networks from binary data

An important class of collective dynamics is virus spreading and information diffusion in social and computer networks [263–270]. We discuss here the problem of reconstructing the network hosting the spreading process and identifying the source of spreading using limited measurements [37]. This is a challenging problem due to (1) the difficulty in predicting and monitoring mutations of deadly virus and (2) absence of epidemic threshold in heterogeneous networks [271–274]. The problem is directly relevant to affairs of significant current interest such as rumor propagation in the online virtual communities, which can cause financial loss or even social instabilities (an example being the 2011 irrational and panicked acquisition of salt in southeast Asian countries caused by the nuclear leak in Japan). In such a case, identifying the propagation network for controlling the dynamics is of great interest.

A significant challenge in reconstructing a spreading network lies in the nature of the available time series: they are polarized, despite stochastic spreading among the nodes. Indeed, the link pattern and the probability of infection are encrypted in the binary status of the individuals, infected or not. There were recent efforts in addressing the inverse problem of some special types of complex propagation networks [275,276]. For example, for diffusion process originated from a single source, the routes of diffusion from the source constitute a tree-like structure. If information about the early stage of the spreading dynamics is available, it would be feasible to decode all branches



that reveal the connections from the source to its neighbors, and then to their neighbors, and so on. Taking into account the time delays in the diffusion process enables a straightforward inference of the source in a complex network through enumerating all possible hierarchical trees [275, 276]. However, if no immediate information about the diffusion process is available, the tree-structure based inference method is inapplicable, and the problem of network reconstruction and locating the source becomes intractable, hindering control of diffusion and delivery of immunization. The loss of knowledge about the source is common in real situations. For example, passengers on an international flight can carry a highly contagious disease, making certain airports the immediate neighbors of the hidden source, which would be difficult to trace. In another example, the source could be migratory birds coming from other countries or continents. A general data-driven approach, applicable in such scenarios, is an outstanding problem in network science and engineering.

Recently, a compressive sensing based framework was developed to reconstruct complex spreading networks based on binary data [37]. Since the dynamics of epidemic propagation are typically highly stochastic with binary time series, the standard power series expansion method to fit the problem into the compressive sensing paradigm (as discussed in preceding sections) is not applicable, notwithstanding the methods of alternative sparsity enforcing regularizers and convex optimization used in Ref. [277] to infer networks. The idea in Ref. [37] was then to develop a scheme to implement the highly nontrivial transformation associated with the spreading dynamics in the paradigm of compressive sensing. Two prototypical models of epidemic spreading on model and real-world (empirical) networks were studied: the classic susceptible-infected-susceptible (SIS) dynamics [263] and the contact process (CP) [278, 279]. Inhomogeneous infection and recovery rates as representative characteristics of the natural diversity were incorporated into the diffusion dynamics to better mimic the real-world situation.

The basic assumption is then that only binary time series can be measured, which characterize the status of any node, infected or susceptible, at any time after the outbreak of the epidemic. The source that triggers the spreading process is assumed to be externally inaccessible (hidden). In fact, one may not even realize its existence from the available time series. The method developed in Ref. [37] enables, based on relatively small amounts of data, a full reconstruction of the epidemic spreading network with nodal diversity and successful identification of the immediate neighboring nodes of the hidden source (thereby ascertaining its existence and uniquely specifying its connections to nodes in the network). The framework was validated with respect to different amounts of data generated from various combinations of the network structures and dynamical processes. High accuracy, high efficiency and applicability in a strongly stochastic environment with measurement noise and missing information are the most striking characteristics of the framework [37]. As a result, broad applications can be expected in addressing significant problems such as targeted control of virus spreading in computer networks and rumor propagation on social networks.

### 3.6.1 Mathematical formulation

**Spreading processes.** The SIS model [263] is a classic epidemic model to study a variety of spreading behaviors in social and computer networks. Each node of the network represents an individual and links are connections along which the infection can propagate to others with certain probability. At each time step, a susceptible node  $i$  in state 0 is infected with rate  $\lambda_i$  if it is connected

to an infected node in state 1. [If  $i$  connects to more than one infected neighbor, the infection probability  $P^{01}$  is given by Eq. (59) below.] At the same time, infected nodes are continuously recovered to be susceptible at the rate  $\delta_i$ . The CP model [278, 279] describes, e.g., the spreading of infection and competition of animals over a territory. The main difference between SIS and CP dynamics lies in the influence on a node's state from its vicinity. In both SIS and CP dynamics,  $\lambda_i$  and  $\delta_i$  depend on the individuals' immune systems and are selected from a Gaussian distribution characterizing the natural diversity. Moreover, a hidden source is regarded as infected at all time.

### Mathematical formulation of reconstruction from binary data based on compressive sensing.

Assume that the disease starts to propagate from a fraction of the infected nodes. The task is to locate any hidden source based solely on binary time series after the outbreak of infection. The state of an arbitrary node  $i$  is denoted as  $S_i$ , where

$$S_i = \begin{cases} 0, & \text{susceptible;} \\ 1, & \text{infected.} \end{cases} \quad (58)$$

Due to the characteristic difference between the SIS dynamics and CP, it is useful to treat them separately.

SIS dynamics. The probability  $P_i^{01}(t)$  of an arbitrary node  $i$  being infected by its neighbors at time  $t$  is

$$P_i^{01}(t) = 1 - (1 - \lambda_i)^{\sum_{j=1, j \neq i}^N a_{ij} S_j(t)}, \quad (59)$$

where  $\lambda_i$  is the infection rate of  $i$ ,  $a_{ij}$  stands for the elements of the adjacency matrix ( $a_{ij} = 1$  if  $i$  connects to  $j$  and  $a_{ij} = 0$  otherwise),  $S_j(t)$  is the state of node  $j$  at  $t$ , and the superscript 01 denotes the change from the susceptible state (0) to the infected state (1). At the same time, the recovery probability of  $i$  is  $P_i^{10}(t) = \delta_i$ , where  $\delta_i$  is the recovery rate of node  $i$  and the superscript 10 denotes the transition from the infected state to the susceptible state. Equation (59) can be rewritten as

$$\ln[1 - P_i^{01}(t)] = \ln(1 - \lambda_i) \cdot \sum_{j=1, j \neq i}^N a_{ij} S_j(t). \quad (60)$$

Suppose measurements at a sequence of times  $t = t_1, t_2, \dots, t_m$  are available. Equation (60) leads to the following matrix form  $\mathbf{X}_{m \times 1} = \mathbf{G}_{m \times (N-1)} \cdot \mathbf{a}_{(N-1) \times 1}$ :

$$\begin{bmatrix} \ln[1 - P_i^{01}(t_1)] \\ \ln[1 - P_i^{01}(t_2)] \\ \vdots \\ \ln[1 - P_i^{01}(t_m)] \end{bmatrix} = \begin{bmatrix} S_1(t_1) & \cdots & S_{i-1}(t_1) & S_{i+1}(t_1) & \cdots & S_N(t_1) \\ S_1(t_2) & \cdots & S_{i-1}(t_2) & S_{i+1}(t_2) & \cdots & S_N(t_2) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ S_1(t_m) & \cdots & S_{i-1}(t_m) & S_{i+1}(t_m) & \cdots & S_N(t_m) \end{bmatrix} \begin{bmatrix} \ln(1 - \lambda_i) a_{i1} \\ \vdots \\ \ln(1 - \lambda_i) a_{i, i-1} \\ \ln(1 - \lambda_i) a_{i, i+1} \\ \vdots \\ \ln(1 - \lambda_i) a_{iN} \end{bmatrix},$$

between node  $i$  and all other nodes, and it is sparse for a general complex network. It can be seen that, if the vector  $\mathbf{X}_{m \times 1}$  and the matrix  $\mathbf{G}_{m \times (N-1)}$  can be constructed from time series,  $\mathbf{a}_{(N-1) \times 1}$  can then be solved by using compressive sensing. The main challenge here is that the infection probabilities  $P_i^{01}(t)$  at different times are not given directly by the time series of the nodal states. A

heuristic method to estimate the probabilities can be devised [37] by assuming that the neighboring set  $\Gamma_i$  of the node  $i$  is known. The number of such neighboring nodes is given by  $k_i$ , the degree of node  $i$ , and their states at time  $t$  can be denoted as

$$S_{\Gamma_i}(t) \equiv \{S_1(t), S_2(t), \dots, S_{k_i}(t)\}. \quad (61)$$

In order to approximate the infection probability, one can use  $S_i(t) = 0$  so that at  $t + 1$ , the node  $i$  can be infected with certain probability. In contrast, if  $S_i(t) = 1$ ,  $S_i(t + 1)$  is only related with the recovery probability  $\delta_i$ . Hence, it is insightful to focus on the  $S_i(t) = 0$  case to derive  $P_i^{01}(t)$ . If one can find two time instants:  $t_1, t_2 \in T$  ( $T$  is the length of time series), such that  $S_i(t_1) = 0$  and  $S_i(t_2) = 0$ , one can calculate the normalized Hamming distance  $H[S_{\Gamma_i}(t_1), S_{\Gamma_i}(t_2)]$  between  $S_{\Gamma_i}(t_1)$  and  $S_{\Gamma_i}(t_2)$ , defined as the ratio of the number of positions with different symbols between them and the length of string. If  $H[S_{\Gamma_i}(t_1), S_{\Gamma_i}(t_2)] = 0$ , the states at the next time step,  $S_i(t_1 + 1)$  and  $S_i(t_2 + 1)$ , can be regarded as i.i.d Bernoulli trials. In this case, using the law of large numbers, one has

$$\lim_{l \rightarrow \infty} \frac{1}{l} \sum_{v=1}^l S_i(t_v + 1) \rightarrow P_i^{01}(\hat{t}_\alpha), \quad \forall t_v, S_i(t_v) = 0, \quad H[S_{\Gamma_i}(\hat{t}_\alpha), S_{\Gamma_i}(t_v)] = 0. \quad (62)$$

A more intuitive understanding of Eq. (62) is that, if the states of  $i$ 's neighbors are unchanged, the fraction of times of  $i$  being infected by its neighbors over the entire time period will approach the actual infection probability  $P_i^{01}$ . Note, however, that the neighboring set of  $i$  is unknown and to be inferred. A strategy is then to artificially enlarge the neighboring set  $S_{\Gamma_i}(t)$  to include all nodes in the network except  $i$ . Denote

$$S_{-i}(t) \equiv \{S_1(t), S_2(t), \dots, S_{i-1}(t), S_{i+1}(t), \dots, S_N(t)\}. \quad (63)$$

If  $H[S_{-i}(t_1), S_{-i}(t_2)] = 0$ , the condition  $H[S_{\Gamma_i}(t_1), S_{\Gamma_i}(t_2)] = 0$  can be ensured. Consequently, due to the nature of the i.i.d Bernoulli trials, application of the law of large numbers leads to

$$\lim_{l \rightarrow \infty} \frac{1}{l} \sum_{v=1}^l S_i(t_v + 1) \rightarrow P_i^{01}(\hat{t}_\alpha), \quad \forall t_v, S_i(t_v) = 0, \quad H[S_{-i}(\hat{t}_\alpha), S_{-i}(t_v)] = 0.$$

Hence, the infection probability  $P_i^{01}(\hat{t}_\alpha)$  of a node at  $\hat{t}_\alpha$  can be evaluated by averaging over its states associated with zero normalized Hamming distance between the strings of other nodes at some time associated with  $\hat{t}_\alpha$ . In practice, to find two strings with absolute zero normalized Hamming distance is unlikely. It is then necessary to set a threshold  $\Delta$  so as to pick the suitable strings to approximate the law of large numbers, that is

$$\frac{1}{l} \sum_{v=1}^{l \gg 1} S_i(t_v + 1) \simeq \frac{1}{l} \sum_{v=1}^{l \gg 1} P_i^{01}(t_v), \quad \forall t_v, S_i(t_v) = 0, \quad H[S_{-i}(\hat{t}_\alpha), S_{-i}(t_v)] < \Delta, \quad (64)$$

where  $S_{-i}(\hat{t}_\alpha)$  serves as a base for comparison with  $S_{-i}(t)$  at all other times and  $(1/l) \sum_{v=1}^{l \gg 1} P_i^{01}(t_v) \simeq P_i^{01}(\hat{t}_\alpha)$ . Since  $H[S_{-i}(\hat{t}_\alpha), S_{-i}(t_v)]$  is not exactly zero, there is a small difference between  $P_i^{01}(\hat{t}_\alpha)$  and  $P_i^{01}(t_v)$  ( $v = 1, \dots, l$ ). It is thus useful to consider the average of  $P_i^{01}(t_v)$  for all  $t_v$  to obtain

$P_i^{01}(\hat{t}_\alpha)$ , leading to the right-hand side of Eq. (64). Let

$$\begin{aligned}\langle S_i(\hat{t}_\alpha + 1) \rangle &= (1/l) \sum_{v=1}^{l \gg 1} S_i(t_v + 1) \\ \langle P_i^{01}(\hat{t}_\alpha) \rangle &= (1/l) \sum_{v=1}^{l \gg 1} P_i^{01}(t_v).\end{aligned}$$

In order to reduce the error in the estimation, one can implement the average on  $S_{-i}(t)$  over all selected strings using Eq. (64). The averaging process is with respect to the nodal states  $S_{j,j \neq i}(t)$  on the right-hand side of the modified dynamical Eq. (60). Specifically, averaging over time  $t$  restricted by Eq. (64) on both sides of Eq. (60) yields

$$\langle \ln[1 - P_i^{01}(t)] \rangle = \ln(1 - \lambda_i) \sum_{j=1, j \neq i}^N a_{ij} \langle S_j(t) \rangle.$$

For  $\lambda_i$  small compared with insignificant fluctuations, the following approximation holds:

$$\ln[1 - \langle P_i^{01}(t) \rangle] \simeq \langle \ln[1 - P_i^{01}(t)] \rangle,$$

which leads to

$$\ln[1 - \langle P_i^{01}(t) \rangle] \simeq \ln(1 - \lambda_i) \sum_{j=1, j \neq i}^N a_{ij} \langle S_j(t) \rangle.$$

Substituting  $\langle P_i^{01}(\hat{t}_\alpha) \rangle$  by  $\langle S_i(\hat{t}_\alpha + 1) \rangle$ , one finally gets

$$\ln[1 - \langle S_i(\hat{t}_\alpha + 1) \rangle] \simeq \ln(1 - \lambda_i) \cdot \sum_{j=1, j \neq i}^N a_{ij} \langle S_j(\hat{t}_\alpha) \rangle. \quad (65)$$

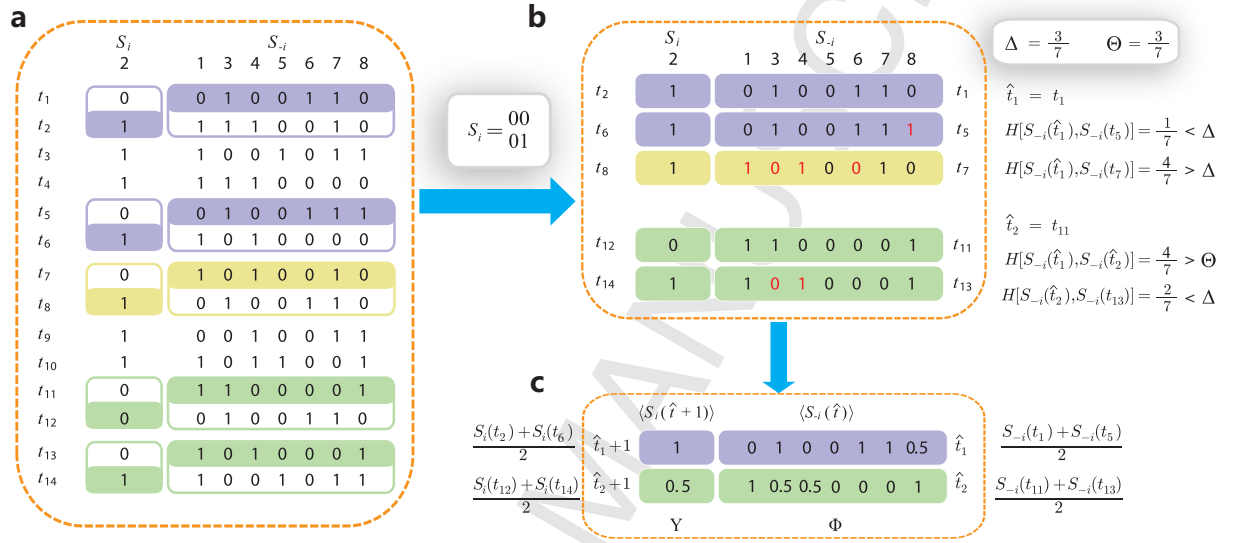
While the above procedure yields an equation that bridges the links of an arbitrary node  $i$  with the observable states of the nodes, a single equation does not contain sufficient structural information about the network. The second step is then to derive a sufficient number of linearly independent equations required by compressive sensing to reconstruct the local connection structure. To achieve this, one can choose a series of base strings at a number of time instants from a set denoted by  $T_{base}$ , in which each pair of strings satisfy

$$H[S_{-i}(\hat{t}_\beta), S_{-i}(\hat{t}_\alpha)] > \Theta, \quad \forall \hat{t}_\alpha, \hat{t}_\beta \in T_{base}, \quad (66)$$

where  $\hat{t}_\alpha$  and  $\hat{t}_\beta$  correspond to the time instants of two base strings in the time series and  $\Theta$  is a threshold. For each string, the process of establishing the relationship between the nodal states and connections can be repeated, leading to a set of equations at different values of  $\hat{t}_\alpha$  in Eq. (65). This process finally gives rise to a set of reconstruction equations in the matrix form:

$$\begin{bmatrix} \ln[1 - \langle S_i(\hat{t}_1 + 1) \rangle] \\ \ln[1 - \langle S_i(\hat{t}_2 + 1) \rangle] \\ \vdots \\ \ln[1 - \langle S_i(\hat{t}_m + 1) \rangle] \end{bmatrix} = \begin{bmatrix} \langle S_1(\hat{t}_1) \rangle & \cdots & \langle S_{i-1}(\hat{t}_1) \rangle & \langle S_{i+1}(\hat{t}_1) \rangle & \cdots & \langle S_N(\hat{t}_1) \rangle \\ \langle S_1(\hat{t}_2) \rangle & \cdots & \langle S_{i-1}(\hat{t}_2) \rangle & \langle S_{i+1}(\hat{t}_2) \rangle & \cdots & \langle S_N(\hat{t}_2) \rangle \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \langle S_1(\hat{t}_m) \rangle & \cdots & \langle S_{i-1}(\hat{t}_m) \rangle & \langle S_{i+1}(\hat{t}_m) \rangle & \cdots & \langle S_N(\hat{t}_m) \rangle \end{bmatrix} \begin{bmatrix} \ln(1 - \lambda_i) a_{i1} \\ \vdots \\ \ln(1 - \lambda_i) a_{i,i-1} \\ \ln(1 - \lambda_i) a_{i,i+1} \\ \vdots \\ \ln(1 - \lambda_i) a_{iN} \end{bmatrix}, \quad (67)$$

where  $\hat{t}_1, \hat{t}_2, \dots, \hat{t}_m$  correspond to the time associated with  $m$  base strings and  $\langle \cdot \rangle$  denote the average over all satisfied  $t$ . The vector  $\mathbf{X}_{m \times 1}$  and the matrix  $\mathcal{G}_{m \times (N-1)}$  can then be obtained based solely on time series of nodal states and the vector  $\mathbf{a}_{(N-1) \times 1}$  to be reconstructed is sparse, rendering applicable the compressive sensing framework. As a result, *exact reconstruction* of all neighbors of node  $i$  from relatively small amounts of observation can be achieved. In a similar fashion the neighboring vectors of all other nodes can be uncovered from time series, enabling a full reconstruction of the whole network by matching the neighboring sets of all nodes.



**Figure 30: Schematic illustration of construction of  $\mathbf{X}$  and  $\mathcal{G}$  from binary time series.** (a) 14 snapshots of data at the time instants  $t_1$ - $t_{14}$  of 8 nodes in a sample network, where  $S_i$  is the time series of node 2 and  $S_{-i}$  denotes the strings of other nodes at different times. The neighborhood of node 2 is to be reconstructed. Only the pairs 00 and 01 in the time series of  $S_i$  ( $i = 2$ ) and the corresponding  $S_{-i}$  contain useful information about the network, as marked by different colors. (b) Since  $S_i(t+1)$  is determined by the neighbors of  $i$  and  $S_{-i}(t)$ , one can sort out  $S_i(t+1)$  and  $S_{-i}(t)$  in the colored sections of the time series in (a). With the threshold parameters  $\Delta = 3/7$  and  $\Theta = 3/7$ , the normalized Hamming distance between each pair of strings  $S_{-i}(t)$  can be calculated, leading to two base strings at  $\hat{t}_1 = t_1$  and  $\hat{t}_2 = t_{11}$  with  $H[S_{-i}(\hat{t}_1), S_{-i}(\hat{t}_2)] > \Theta$ . The colored strings can be separated into two groups that are led by the two base strings, respectively. In each group, the normalized Hamming distances  $H[S_{-i}(\hat{t}_\alpha), S_{-i}(t_v)]$  between the base string and other strings are calculated and the difference from  $S_{-i}(\hat{t}_\alpha)$  in each string is marked by red. Using parameter  $\Delta$ , in the group led by  $S_{-i}(\hat{t}_1)$ ,  $S_{-i}(t_5)$  and  $S_i(t_6)$  are preserved, because of  $H[S_{-i}(\hat{t}_1), S_{-i}(t_5)] < \Delta$ . In contrast,  $S_{-i}(t_7)$  and  $S_i(t_8)$  are disregarded because  $H[S_{-i}(\hat{t}_1), S_{-i}(t_7)] > \Delta$ . In the group led by  $S_{-i}(\hat{t}_2)$ , due to  $H[S_{-i}(\hat{t}_2), S_{-i}(t_{13})] < \Delta$ , the string is preserved. The two sets of remaining strings marked by purple and green can be used to yield the quantities required by the reconstruction formula. (Note that different base strings are allowed to share some strings, but for simplicity, this situation is not illustrated here.) (c) The average values  $\langle S_i(\hat{t}_\alpha + 1) \rangle$  and  $\langle S_{-i}(\hat{t}_\alpha) \rangle$  used to extract the vector  $\mathbf{X}$  and the matrix  $\mathcal{G}$  in the reconstruction formula, where  $\langle S_{-i}(\hat{t}_1) \rangle = [S_{-i}(t_1) + S_{-i}(t_5)]/2$ ,  $\langle S_{-i}(\hat{t}_2) \rangle = [S_{-i}(t_{11}) + S_{-i}(t_{13})]/2$ ,  $\langle S_i(\hat{t}_1 + 1) \rangle = [S_i(t_2) + S_i(t_6)]/2$ , and  $\langle S_i(\hat{t}_2 + 1) \rangle = [S_i(t_{12}) + S_i(t_{14})]/2$  based on the remaining strings marked in different colors. Compressive sensing can be used to reconstruct the neighboring vector  $\mathbf{a}$  of node 2 from  $\mathbf{X}$  and  $\mathcal{G}$  from  $\mathbf{X} = \mathcal{G} \cdot \mathbf{a}$ . From Ref. [37] with permission.

CP dynamics. The infection probability of an arbitrary node  $i$  is given by

$$P_i^{01}(t) = \lambda_i \sum_{j=1, j \neq i}^N a_{ij} S_j(t) / k_i, \quad (68)$$

where  $k_i$  is the degree of the node  $i$ , and the recovery probability is  $P_i^{10}(t) = \delta_i$ . In close analogy to the SIS dynamics, one has

$$\langle S_i(\hat{t}_\alpha + 1) \rangle \simeq \langle P_i^{01}(\hat{t}_\alpha) \rangle = \frac{\lambda_i \sum a_{ij} \langle S_j(\hat{t}_\alpha) \rangle}{k_i}. \quad (69)$$

One can then choose a series of base strings using a proper threshold  $\Theta$  to establish a set of equations, expressed in the matrix form  $\mathbf{X}_{m \times 1} = \mathcal{G}_{m \times (N-1)} \cdot \mathbf{a}_{(N-1) \times 1}$ , where  $\mathcal{G}$  has the same form as in Eq. (67), but  $\mathbf{X}$  and  $\mathbf{a}$  are given by

$$\begin{aligned} \mathbf{X} &= [\langle S_i(\hat{t}_1 + 1) \rangle, \langle S_i(\hat{t}_2 + 1) \rangle, \dots, \langle S_i(\hat{t}_m + 1) \rangle]^T, \\ \mathbf{a} &= \left[ \frac{\lambda_i}{k_i} a_{i1}, \dots, \frac{\lambda_i}{k_i} a_{i,i-1}, \frac{\lambda_i}{k_i} a_{i,i+1}, \dots, \frac{\lambda_i}{k_i} a_{iN} \right]^T. \end{aligned} \quad (70)$$

The reconstruction framework based on building up the vector  $\mathbf{X}$  and the matrix  $\mathcal{G}$  is schematically illustrated in Fig. 30. It is noteworthy that the framework can be extended to directed networks in a straightforward fashion due to the feature that the neighboring set of each node can be independently reconstructed. For instance, the neighboring vector  $\mathbf{a}$  can be defined to represent a unique link direction, e.g., incoming links. Inference of the directed links of all nodes yields the full topology of the entire directed network.

**Inferring inhomogeneous infection rates.** The values of the infection rate  $\lambda_i$  of nodes can be inferred after the neighborhood of each node has been successfully reconstructed. The idea roots in the fact that the infection probability of a node approximated by the frequency of being infected calculated from time series is determined both by its infection rate and by the number of infected nodes in its neighborhood. An intuitive picture can be obtained by considering the following simple scenario in which the number of infected neighbors of node  $i$  does not change with time. In this case, the probability of  $i$  being infected at each time step is fixed. One can thus count the frequency of the 01 and 00 pairs embedded in the time series of  $i$ . The ratio of the number of 01 pairs over the total number of 01 and 00 pairs gives approximately the infection probability. The infection rate can then be calculated by using Eqs. (59) and (68) for the SIS and CP dynamics, respectively. In a real-world situation, however, the number of infected neighbors varies with time. The time-varying factor can be taken into account by sorting out the time instants corresponding to different numbers of the infected neighbors, and the infection probability can be obtained at the corresponding time instants, leading to a set of values for the infection rate whose average represents an accurate estimate of the true infection rate for each node.

To be concrete, considering all the time instants  $t_v$  associated with  $k_I$  infected neighbors, one

can denote

$$\begin{aligned} S_i^{(k_I)} &= (1/l) \sum_{v=1}^l S_i(t_v + 1), \text{ for all } t_v, \\ \sum_{j \in \Gamma_i} S_j(t_v) &= k_I, \\ S_i(t_v) &= 0, \end{aligned}$$

where  $\Gamma_i$  is the neighboring set of node  $i$ ,  $k_I$  is the number of infected neighbors, and  $S_i^{(k_I)}$  represents the average infected fraction of node  $i$  with  $k_I$  infected neighbors. Given  $S_i^{(k_I)}$ , one can rewrite Eq. (59) by substituting  $S_i^{(k_I)}$  for  $P_i^{01}(t)$  and  $\lambda_i^{(k_I)}$  for  $\lambda_i$ , which yields

$$\lambda_i^{(k_I)} = 1 - \exp \left[ \ln (1 - S_i^{(k_I)}) / k_I \right].$$

The estimation error can be reduced by averaging  $\lambda_i^{(k_I)}$  with respect to different values of  $k_I$ , as follows:

$$\lambda_i^{\text{true}}(\text{SIS}) \approx \langle \lambda_i^{(k_I)} \rangle = \frac{1}{N_{\lambda_i}} \sum_{k_I \in \lambda_i} \lambda_i^{(k_I)}, \quad (71)$$

where  $\lambda_i$  denotes the set of all possible infected neighbors during the epidemic process and  $N_{\lambda_i}$  denotes the number of different values of  $k_I$  in the set. Analogously, for CP, one can evaluate  $\lambda_i^{\text{true}}$  from Eq. (68) as

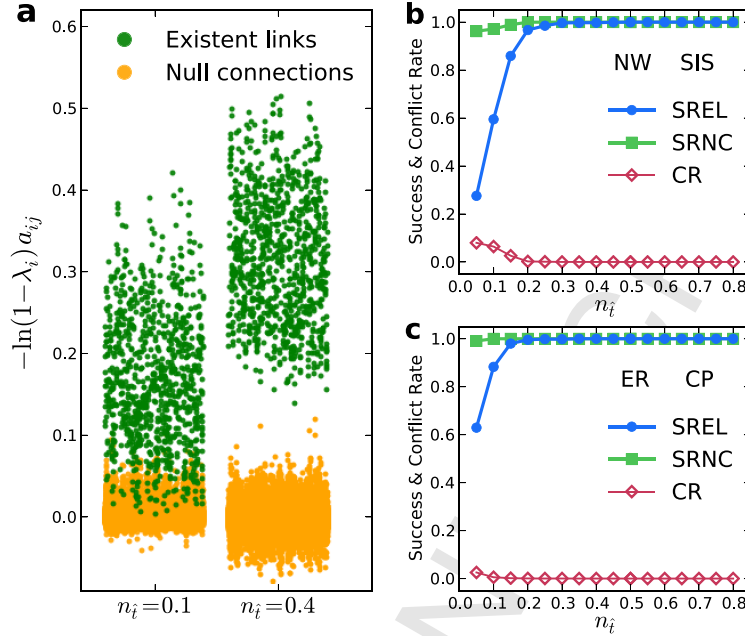
$$\lambda_i^{\text{true}}(\text{CP}) \approx \langle \lambda_i^{(k_I)} \rangle = \frac{1}{N_{\lambda_i}} \sum_{k_I \in \lambda_i} \frac{S_i^{(k_I)} k_i}{k_I} \quad (72)$$

where  $k_i = \sum_{j=1}^N a_{ij}$  is the node degree of  $i$ . After all the links of  $i$  have been successfully reconstructed,  $S_i^{(k_I)}$  can be obtained from the time series in terms of the satisfied  $S_i(t_v + 1)$ , allowing one to infer  $\lambda_i^{\text{true}}$  via Eqs. (71) and (72).

The method so described for estimating the infection rates is applicable to any type of networks insofar as the network structure has been successfully reconstructed [37].

### 3.6.2 Reconstructing complex spreading networks: examples

**Reconstructing networks and inhomogeneous infection and recovery rates.** A key performance indicator of the binary data based reconstruction framework is the number of base strings (equations) for a variety of diffusion dynamics and network structures. It is necessary to calculate the success rates for existent links (SREL) and null connections (SRNC), corresponding to non-zero and zero element values in the adjacency matrix, respectively, in terms of the number of base strings. The binary nature of the network dynamical process and data requires that the strict criterion be imposed [37], i.e., a network is regarded to have been fully reconstructed if and only if both success rates reach 100%. The sparsity of links makes it necessary to define SREL and SRNC separately. Since the reconstruction method is implemented for each node in the network, SREL and SRNC can be defined with respect to each individual node and, the two success rates for the entire network are the respective averaged values over all nodes. The issue of trade-off can also



**Figure 31: Performance of reconstructing complex spreading networks from binary data.** (a) Element values  $\ln(1-\lambda_i)a_{ij}$  of the vector  $\mathbf{X}$  times -1 for different fraction  $n_i$  of the base strings for SIS dynamics. (b-c), success rate (SREL and SRNC) and conflict rate (CR) of reconstruction as a function of  $n_i$  for SIS dynamics on Newman-Watts (NW) small-world networks (b) and CP dynamics on Erdős-Rényi (ER) random networks (c), respectively. For the SIS dynamics, the parameters were  $\Theta = 0.25$ ,  $\Delta = 0.45$ , the infection and recovery rates  $\lambda_i$  and  $\delta_i$  were randomly distributed in the ranges  $(0.2, 0.4)$  and  $(0.4, 0.6)$ , respectively. For the CP dynamics, the parameters were  $\Theta = 0.35$ ,  $\Delta = 0.45$ ,  $\lambda_i$  and  $\delta_i$  were randomly distributed in the ranges  $(0.7, 0.9)$  and  $(0.2, 0.4)$ , respectively. The network size  $N$  is 200 with average node degree  $\langle k \rangle = 4$ . The results were obtained by ensemble averaging over 10 independent realizations. The success rate was determined by setting a proper threshold. From Ref. [37] with permission.

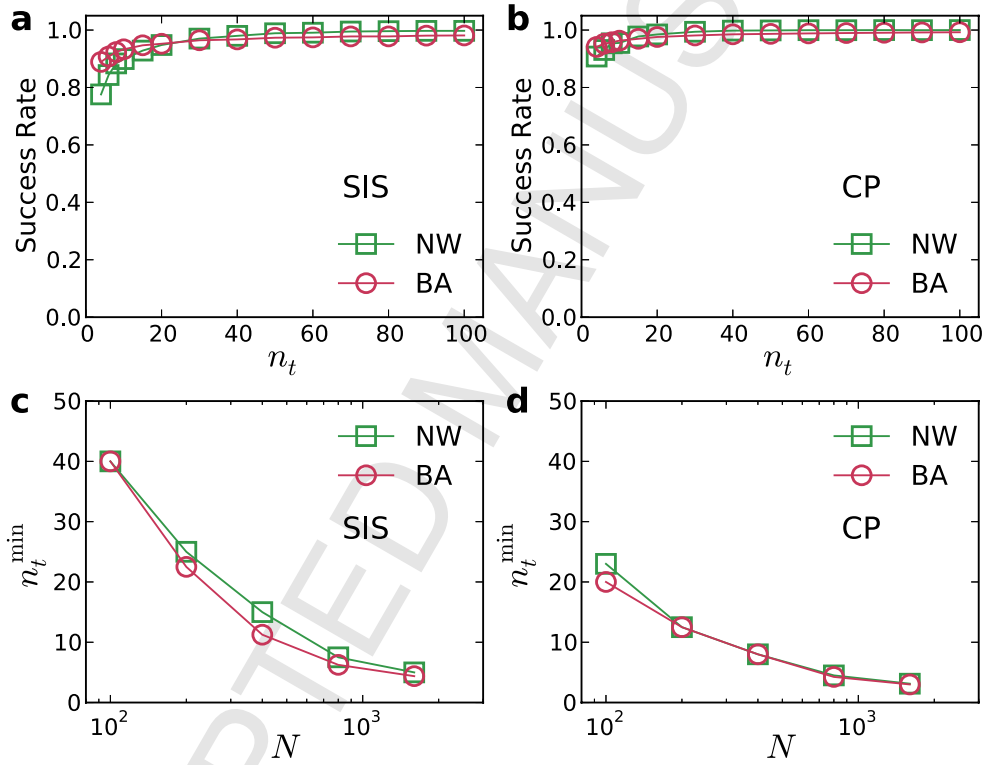
be considered in terms of the true positive rate (TPR - for correctly inferred links) and the false positive rate (FPR - for incorrectly inferred links).

In Ref. [37], a large number of examples of reconstruction were presented. Take an example where there is no hidden source, and binary time series can be obtained by initiating the spreading process from a fraction of infected nodes. Figure 31(a) shows the reconstructed values of the components of the neighboring vector  $\mathbf{X}$  of all nodes. Let  $n_i$  be the number of base strings normalized by the total number of strings. For small values of  $n_i$ , e.g.,  $n_i = 0.1$ , the values of elements associated with links and those associated with null connections (actual zeros in the adjacency matrix) overlap, leading to ambiguities in the identification of links. In contrast, for larger values of  $n_i$ , e.g.,  $n_i = 0.4$ , an explicit gap emerges between the two groups of element values, enabling correct identification of all links by simply setting a threshold within the gap. The success rates (SREL and SRNC) as a function of  $n_i$  for SIS and CP on both homogeneous and heterogeneous networks are shown in Figs. 31(b,c), where nearly perfect reconstruction of links are obtained, insofar as  $n_i$  exceeds a relatively small value - an advantage of compressed sensing. The exact reconstruction is robust in the sense that a wide range of  $n_i$  values can yield nearly 100% success rates. The reconstruction method is effective for tackling real networks in the absence of any *a priori* knowledge



about its topology.

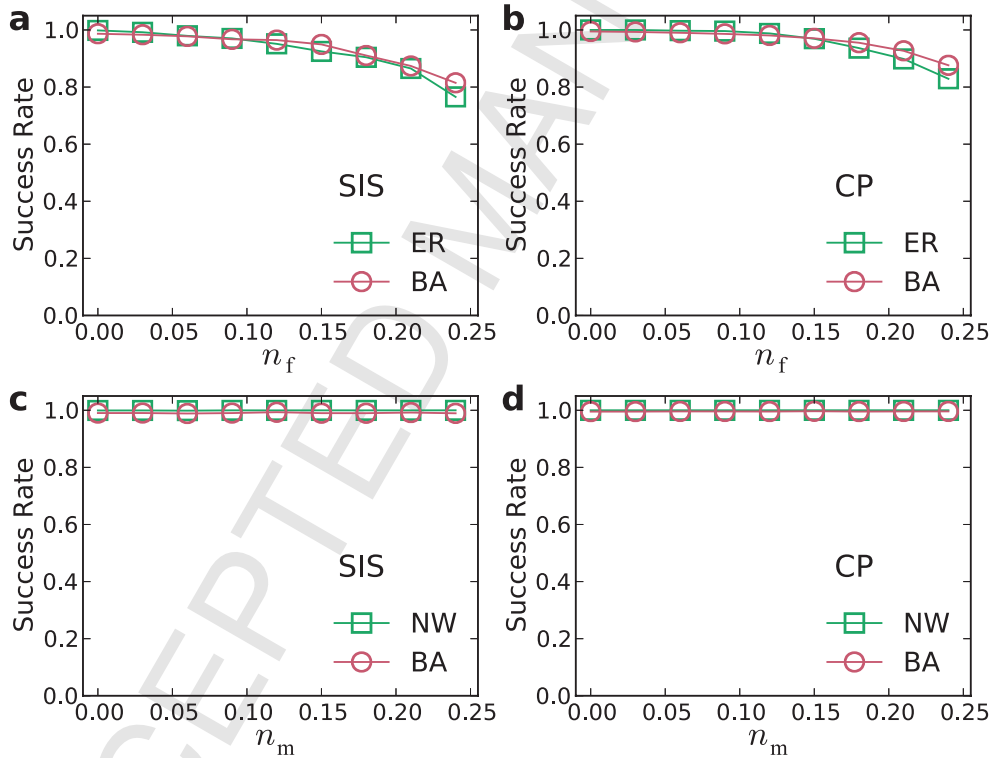
Since the network is to be reconstructed through the union of all neighborhoods, one may encounter “conflicts” with respect to the presence/absence of a link between two nodes as generated by the reconstruction results centered at the two nodes, respectively. Such conflicts will reduce the accuracy in the reconstruction of the entire network. The effects of edge conflicts can be characterized by analyzing the consistency of mutual assessment of the presence or absence of link between each pair of nodes, as shown in Figs. 31(b,c). It can be seen that inconsistency arises for small values of  $n_t$  but vanishes *completely* when the success rates reach 100%, indicating perfect consistency among the mutual inferences of nodes and consequently guaranteeing accurate reconstruction of the entire network.



**Figure 32: Effect of the length of the binary time series and network size on reconstruction.** (a-b) Success rate as a function of the relative length  $n_t$  of time series for SIS (a) and CP (b) in combination with NW and BA scale free networks. (c-d) the minimum relative length  $n_t^{\min}$  that assures at least 95% success rate as a function of the network size  $N$  for SIS (c) and CP (d) dynamics on NW and BA networks, where the success rate is taken as the geometric average over SREL and SRNC. For the SIS dynamics,  $\lambda_i \in (0.1, 0.3)$  and  $\delta \in (0.2, 0.4)$ . For the CP dynamics,  $\lambda_i \in (0.7, 0.9)$  and  $\delta \in (0.3, 0.5)$ . In (a) and (b), the network size  $N$  is 500. The other parameters are the same as in Fig. 31. Note that  $n_t \equiv t/N$ , where  $t$  is the absolute length of time series, and  $n_t^{\min} \equiv t_{\min}/N$ , where  $t_{\min}$  is the minimum absolute length of time series required for at least 95% success rate. From Ref. [37] with permission.

While the number of base strings is relatively small compared with the network size, it is necessary to have a set of strings at different time with respect to a base string to formulate the mathematical reconstruction framework. How the length of the time series affects the accuracy of reconstruction was studied [37]. Figures 32(a,b) show the success rates as a function of the relative

length  $n_t$  of time series for SIS and CP dynamics on both homogeneous and heterogeneous networks, respectively, where  $n_t$  is the total length of time series from the beginning of the spreading process divided by the network size  $N$ . The results demonstrate that, even for very small values of  $n_t$ , most links can already be identified, as reflected by the high values of the success rate shown. Figures 32(c,d) show the minimum length  $n_t^{\min}$  required to achieve at least 95% success rate for different network size. For both SIS and CP dynamics on networks,  $n_t^{\min}$  decreases considerably as  $N$  is increased. This seemingly counterintuitive result is due to the fact that different base strings can share strings at different times to enable reconstruction. In general, as  $N$  is increased,  $n_t$  will increase accordingly. However, a particular string can belong to different base strings with respect to the threshold  $\Delta$ , accounting for the slight increase in the absolute length of the time series and the reduction in  $n_t^{\min}$ . The dependence of the success rate on the average node degree  $\langle k \rangle$  for SIS and CP on different networks was investigated as well [37]. The results in Figs. 31 and 32 demonstrate the high accuracy and efficiency of the compressive sensing based reconstruction method based on small amounts of binary data.



**Figure 33: Reconstruction against noise and inaccessible nodes.** (a-b) Success rate as a function of the fraction  $n_f$  of the flipped states induced by noise in time series for SIS (a) and CP (b) dynamics on ER and BA networks. (c-d) success rate as a function of the fraction  $n_m$  of nodes that are externally inaccessible for SIS (c) and CP (d) dynamics on NW and BA networks. The network size  $N$  is 500 and  $\langle k \rangle = 4$ . The other parameters are the same as in Fig. 31. From Ref. [37] with permission.

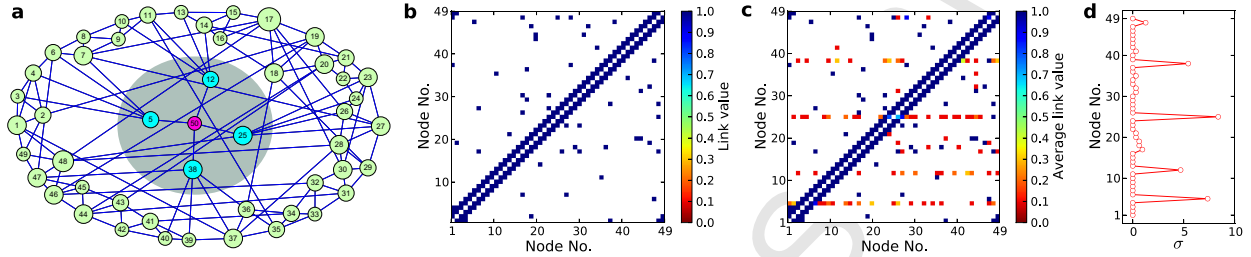
In practice, noise is present and it is also common for time series from certain nodes to be missing, so it is necessary to test the applicability of the method under these circumstances. Figures 33(a,b) show the dependence of the success rate on the fraction  $n_f$  of states in the time series that flip due to noise for SIS and CP dynamics on two types of networks, respectively. It can

be seen that the success rates are hardly affected, providing strong support for the applicability of the reconstruction method. For example, even when 25% of the nodal states flip, about 80% success rates can still be achieved for both dynamical processes and different network topologies. Figures 33(c,d) present the success rate versus the fraction  $n_m$  of unobservable nodes, the states of which are externally inaccessible. It can be seen that the high success rate remains mostly unchanged as  $n_m$  is increased from zero to 25%, a somewhat counterintuitive but striking result. It was found that [37], in general, missing information can affect the reconstruction of the neighboring vector, as reflected by the reduction in the gap between the success rates associated with the actual links and null connections. However, even for high values of  $n_f$ , e.g.,  $n_f = 0.3$ , there is still a clear gap, indicating that a full recovery of all links is achievable. Taken together, the high accuracy, efficiency and robustness against noise and missing information provide strong credence for the validity and power of the framework for binary time-series based network reconstruction.

Having reconstructed the network structure, one can estimate the infection and recovery rates of individuals to uncover their diversity in immunity. This is an essential step to implement target vaccination strategy in a population or on a computer network to effectively suppress/prevent the spreading of virus at low cost, as a large body of literature indicates that knowledge about the network structure and individual characteristics is sufficient for controlling the spreading dynamics [280–283]. An effective method was proposed [37] to infer the individuals' infection rates  $\lambda_i$  based solely on the binary time series of the nodal states after an outbreak of contamination. In particular, after all links have been successfully predicted,  $\lambda_i$  can be deduced from the infection probabilities that can be approximated by the corresponding infection frequencies. These probabilities depend on both  $\lambda_i$  and the number of infected neighbors. The reproduced infection rates  $\lambda_i$  of individuals for both SIS and CP dynamics on different networks were in quite good agreement with the true values with small prediction errors. An error analysis revealed uniformly high accuracy of the method [37]. The inhomogeneous recovery rates  $\delta_i$  of nodes can be predicted from the binary time series in a more straightforward way, because  $\delta_i$ 's do not depend on the nodal connections. Thus the framework is capable of predicting characteristics of nodal diversity in terms of degrees and infection and recovery rates based solely on binary time series of nodal states.

**Locating the source of spreading.** Assume that a hidden source exists outside the network but there are connections between it and some nodes in the network. In practice, the source can be modeled as a special node that is always infected. Starting from the neighborhood of the source, the infection originates from the source and spreads all over the network. One first collects a set of time series of the nodal states except the hidden source. As discussed in detail in Sec. 3.3, the basic idea of ascertaining and locating the hidden source is based on missing information from the hidden source when attempting to reconstruct the network [31, 36]. In particular, in order to reconstruct the connections belonging to the immediate neighborhood of the source accurately, time series from the source are needed to generate the matrix  $\mathcal{G}$  and the vector  $\mathbf{X}$ . But since the source is hidden, no time series from it are available, leading to reconstruction inaccuracy and, consequently, anomalies in the predicted link patterns of the neighboring nodes. It is then possible to detect the neighborhood of the hidden source by identifying any abnormal connection patterns [31, 36], which can be accomplished by using different data segments. If the inferred links of a node are stable with respect to different data segments, the node can be deemed to have no connection with the hidden source; otherwise, if the result of inferring a node's links

varies significantly with respect to different data segments, the node is likely to be connected to the hidden source. The standard deviation of the predicted results with respect to different data segments can be used as a quantitative criterion for the anomaly. Once the neighboring set of the source is determined, the source can be precisely located topologically.



**Figure 34: An example of locating an external hidden source from time series.** (a) Hidden source treated as a special node (in red) is connected to four nodes in the network (blue). The time series of other nodes except the source (No. 50) after the outbreak of an epidemic were assumed to be available. (b) True adjacency matrix of the NW network with identical link weights to facilitate a comparison with the reconstructed adjacency matrix. (c) Reconstructed adjacency matrix from a number of segments in time series. The four neighboring nodes of the source are predicted to be densely linked to other nodes, as indicated by the average value of the elements in the four rows corresponding to these nodes. (d) Structural variance  $\sigma$  of each node. The four neighboring nodes of the source exhibit much larger values of  $\sigma$  than those from the other nodes, providing unequivocal evidence that they belong to the immediate neighborhood of the hidden source. From Ref. [37] with permission.

An example [37] is shown in Fig. 34, where a hidden source is connected with four nodes in the network [Fig. 34(a)], as can be seen from the network adjacency matrix [Fig. 34(b)]. The reconstruction framework was implemented on each accessible node by using different sets of data in the time series. For each data set, the neighbors of all nodes were predicted, generating the underlying adjacency matrix. Averaging over the elements corresponding to each location in all the reconstructed adjacency matrices leads to Fig. 34(c), in which each row corresponds to the mean number of links in a node's neighborhood. The inferred links of the immediate neighbors of the hidden source exhibit anomalies. To quantify the anomalies, the structural standard deviation  $\sigma$  was calculated from different data segments, where  $\sigma$  associated with node  $i$  is defined through the  $i$ th row in the adjacency matrix as

$$\sigma_i = \frac{1}{N} \sum_{j=1}^N \sqrt{\frac{1}{g} \sum_{k=1}^g (a_{ij}^{(k)} - \langle a_{ij} \rangle)^2}, \quad (73)$$

where  $j$  denotes the column,  $a_{ij}^{(k)}$  represents the element value in the adjacency matrix inferred from the  $k$ th group of the data,  $\langle a_{ij} \rangle = (1/g) \sum_{k=1}^g a_{ij}^{(k)}$  is the mean value of  $a_{ij}$ , and  $g$  is the number of data segments. Applying Eq. (73) to the reconstructed adjacency matrices gives the results in Fig. 34(d), where the values of  $\sigma$  associated with the immediate neighboring nodes of the hidden source are much larger than those from others (which are essentially zero). A threshold can be set in the distribution of  $\sigma_i$  to identify the immediate neighbors of the hidden source.

## 4 Alternative methods for reconstructing complex, nonlinear dynamical networks

### 4.1 Reconstructing complex networks from response dynamics

Probing into a dynamical system in terms of its response to external driving signals is commonly practiced in biological sciences. This approach is particularly useful for studying systems with complex interactions and dynamical behaviors. The basic idea of response dynamics was previously exploited for reconstructing complex networks of coupled phase oscillators [8]. Through measuring the collective response of the oscillator network to an external driving signal, the network topology can be recovered through repeated measurement of the dynamical states of the nodes, provided that the driving realizations are sufficiently independent of each other. Since complex networks are generally sparse, the number of realizations of external driving can be much smaller than the network size.

A network of coupled phase oscillators with a complex interacting topology is described by

$$\dot{\phi}_i = \omega_i + \sum_{j=1}^N J_{ij} f_{ij}(\phi_j - \phi_i) + I_{i,m}, \quad (74)$$

where  $\phi_i(t)$  is the phase of oscillator  $i$  at time  $t$ ,  $\omega_i$  is its natural frequency,  $J_{ij}$  is the coupling strength from oscillator  $j$  to oscillator  $i$  (weighted adjacency matrix, where  $J_{ij} = 0$  indicates the absence of a link from  $j$  to  $i$ ), and  $f_{ij}$ 's are the pairwise coupling functions among the connected oscillators. The nodes on which the external driving signals are imposed are specified as  $I_{i,m}$ , where  $m = 0$  indicates no driving signal. When driving is present, the collective frequency is

$$\Omega_m = \omega_i + \sum_{j=1}^N J_{ij} f_{ij}(\phi_{j,m} - \phi_{i,m}) + I_{i,m}. \quad (75)$$

The frequency difference between the driven and non-driven systems is

$$D_{i,m} = \sum_{j=1}^N J_{ij} [f_{ij}(\phi_{j,m} - \phi_{i,m}) - f_{ij}(\phi_{j,0} - \phi_{i,0})]. \quad (76)$$

For sufficiently small signal strength, the phase dynamics of the oscillators are approximately those of the original. We can approximate  $f_{ij}(x) = f'_{ij}(\Delta_{ji,0}) + O(x^2)$  and let  $\theta_{j,m}$  denote  $\phi_{j,m} - \phi_{j,0}$ . After a number of experiments, the network configuration converges to

$$\mathcal{D} = \hat{\mathcal{J}} \cdot \theta, \quad (77)$$

where  $\hat{\mathcal{J}}_{ij}$  is the Laplacian matrix with

$$\hat{\mathcal{J}}_{ij} = \begin{cases} J_{ij} f'_{ij}(\Delta_{ji,0}), & \text{for } i \neq j, \\ -\sum_{k \neq i} J_{ik} f'_{ik}(\Delta_{ki,0}), & i = j. \end{cases} \quad (78)$$

Note that the phase and frequency difference matrices  $\theta$  and  $\mathcal{D}$  are measurable. Thus, the network structure characterized by the matrix  $\hat{\mathcal{J}}$  can be solved by  $\hat{\mathcal{J}} = \mathcal{D} \cdot \theta^{-1}$ . The reconstruction method

based on the response dynamics not only can reveal links among the nodes but also can provide a quantitative estimate of the interaction strengths represented in the matrix  $\hat{J}$ . In general the number of experimental realizations required is  $M = N$  to ensure a unique solution of the matrix  $\hat{J}$  to represent the network structure. However, most real networks are sparse in the sense that the degree of a node is usually much less than the network size, i.e.,  $K \ll N$ . To obtain the matrix  $\hat{J}$  with the least number of links is effectively a constraint for reducing the required data amount through some optimization algorithms. Specifically, constraint (77) can be used to parameterize the family of admissible matrices through  $(N - M)N$  parameters,  $P_{ij}$ , in terms of a singular value decomposition of  $\Theta^T = \mathcal{U}$  in a standard way, where the singular matrix  $\mathcal{S}$  of dimension  $M \times N$  contains the singular values on the diagonal. The coupling matrices can be reformulated to be  $\hat{J} = \mathcal{D} \cdot \mathcal{U} \cdot \tilde{\mathcal{S}} \cdot \mathcal{V}^T + \mathcal{P} \cdot \mathcal{V}$ , where  $P_{ij}$  is set to be zero for  $j \leq M$ ,  $\tilde{S}_{ij} = \delta_{ij}/\sigma_i$  if  $\sigma_i > 10^{-4}$  and  $\tilde{S}_{ii} = 0$  if  $\sigma_i \leq 10^{-4}$ . Finally, with respect to the parameter matrix  $\mathcal{P}$ , the incoming links of any node  $i$  can be inferred by minimizing the 1-norm of the  $i$ th row vector of  $\hat{J}$  as

$$\|\hat{J}_i\|_1 := \sum_{j=1; j \neq i}^N |\hat{J}_{ij}|. \quad (79)$$

Numerical tests [34] showed that the number  $M$  of the required experimental realizations can be substantially smaller than  $N$  to yield reasonable reconstruction results.

A detailed description of the response dynamics based approach to network reconstruction can be found in Ref. [34]. In general, the reconstruction method is applicable to networked systems whose behaviors are dominated by the linearized dynamics about some stable state. An issue concerns how the reconstruction method can be extended to networked systems characterized by more than one variable, i.e., systems beyond phase coupled oscillators. Another issue is whether it would be possible to reconstruct a network from time series without any perturbation to the nodal dynamics. Addressing these issues is of both theoretical and practical importance.

## 4.2 Reconstructing complex networks via system clone

Exploiting synchronization between a driver and a response system through feedback control led to a method for reconstructing complex networks [6]. The basic idea was to design a replica or a clone system that is sufficiently close to the original network without requiring knowledge about the network structure. From the clone system, the connectivities and interactions among the nodes can be obtained directly, realizing the goal of network reconstruction.

The conventional method to deal with driver-response systems is to design proper feedback control to synchronize the state of the response system with that of the driver system. To achieve reconstruction, the elements in the adjacency matrix characterizing the network topology of the clone system are treated as the variables to be synchronized with those of the original system. With an appropriate feedback control, the adjacency matrix of the clone system can be “forced” to converge to the unknown adjacency matrix in the original system due to synchronization. For this approach to be effective, the local dynamics of each node needs to be known. In addition, the local dynamical and the coupling functions need to be Lipschitz continuous [6].

More details of the synchronization approach can be described, as follows. Consider a net-

worked system described by a set of differential equations

$$\dot{\mathbf{x}}_i = \mathbf{f}(\mathbf{x}_i) + \sum_{j=1}^N a_{ij} \mathbf{h}_j(\mathbf{x}_j), \quad (80)$$

where  $\mathbf{x}_i$  denotes the state of node  $i$  ( $i = 1, \dots, N$ ),  $a_{ij}$  is the  $ij$ th element of the adjacency matrix  $\mathcal{A}$ ,  $\mathbf{f}(\mathbf{x}_i)$  represents the local dynamics of node  $i$ , and  $\mathbf{h}_j$  is the coupling function. In order to realize stable synchronization, the functions  $\mathbf{f}_i$  and  $\mathbf{h}_i$  are required to be Lipschitzian for all nodes. To design a clone system requires that the functions  $\mathbf{f}_i$  and  $\mathbf{h}_i$  be known *a priori*. The clone system under feedback control can be written as

$$\begin{aligned} \dot{\mathbf{y}} &= \mathbf{f}_i(\mathbf{y}_i) + \sum_{j=1}^N b_{ij} \mathbf{h}_j(\mathbf{y}_j) + \mathbf{q}_i + \mathbf{u}_i, \\ \dot{b}_{ij} &= -\gamma_{ij} \mathbf{h}_j(\mathbf{y}_j) \cdot \mathbf{e}_i, \end{aligned} \quad (81)$$

where  $\gamma_{ij}$  are positive coefficients used to control the evolution of  $b_{ij}$  so as to “copy” the dynamics of the original system,  $\mathbf{q}_i$  represents the modeling errors, and  $\mathbf{e}_i \equiv \mathbf{y}_i - \mathbf{x}_i$  denotes the difference between the states of the clone and the original systems. To ensure that the clone system can approach the original system, the following Lyapunov function can be exploited:

$$\dot{\Omega} = \sum_{i=1}^N \dot{e}_i^2 + \sum_{i=1}^N \sum_{j=1}^N (1/\gamma_{ij}) (b_{ij} - a_{ij})^2. \quad (82)$$

The Lipschitzian and stability constraints give rise to a negative Lyapunov exponent and the feedback control as

$$\mathbf{u}_i = -k_1 \mathbf{e}_i - \frac{1}{4\epsilon_1} \delta^2 \mathbf{e}_i, \quad (83)$$

where  $k_1 > L_1 + NL_2$ ,  $L_1$  and  $L_2$  are the positive constants in the Lipschitzian constraint. It can be proven [6] that feedback control in the form of Eq. (83) in combination with the Lipschitzian constraint can guarantee that the Lyapunov exponent is zero or negative so that the clone system converges to the original system with small errors. Numerical tests in [6] demonstrated the working of the synchronization-based reconstruction method.

In an alternative approach [20], it was argued that if the local dynamics of the nodes are available, a direct solution of the network topology is possible without the need of any clone system. The basic idea came from the fact that, if the local dynamics and the coupling functions are available, then the only unknown parameters are the coupling strengths associated with the adjacency matrix. With data collected at different times, a set of equations for the coupling strengths can be obtained, which can be solved by using the standard Euclidean  $L_2$ -norm minimization. Numerical tests showed that this reconstruction method is effective for both transient and attracting dynamics [20].

### 4.3 Network reconstruction via phase-space linearization

For nonlinear dynamical networks, there was a method [10] based on chaotic time-series analysis, where time series were assumed to be available from some or all nodes of the network. The nodal

dynamics were assumed to be described by autonomous systems with a few coupling terms. That is, the network is sparse. The delay-coordinate embedding method [43, 59, 60] was then used to reconstruct the phase space of the underlying networked dynamical system. The principal idea was to estimate the Jacobian matrix of the underlying dynamics, which governs the evolutions of infinitesimal vectors in the tangent space along a typical trajectory of the system. Mathematically, the entries of the Jacobian matrix are mutual partial derivatives of the dynamical variables on different nodes in the network. A statistically significant entry in the matrix implies a connection between the two nodes specified by the row and the column indices of that entry. Because of the mathematical nature of the Jacobian matrix, i.e., it is meaningful only for infinitesimal tangent vectors, linearization of the dynamics in the neighborhoods of the reconstructed phase-space points is needed, for which constrained optimization techniques [112–118] were found to be effective [10].

The approach [10] was based on using  $L_1$ -minimization in the phase space of a networked system to reconstruct the topology without knowledge of the self-dynamics of the nodes and without using any external perturbation to the state of nodes. In particular, one data point  $\mathbf{x}_*^t$  is chosen from the phase space at time  $t$  and from the neighborhood of  $N$  nearest data points from the time series. For node  $i$  (the  $i$ th coordinate), the dynamics at  $t$  around  $\mathbf{x}_*^t$  can be expressed by using a Taylor series expansion, yielding

$$\mathbf{x}_i^{t+1} = \mathbf{F}_i(\mathbf{x}_*^t) + D\mathbf{F}_i(\mathbf{x}_*^t) \cdot (\mathbf{x}_i^t - \mathbf{x}_*^t) + O(\|\mathbf{x}_i^t - \mathbf{x}_*^t\|^2), \quad (84)$$

where the higher-order terms can be omitted when implementing the reconstruction. A key parameter affecting the reconstruction accuracy is size of the neighborhood in the phase space. If the size is relatively large, more data points will be available but the equations will have larger errors. If the size is small, the amount of available data will be reduced. There is thus a trade-off for choosing the neighborhood size. Selecting a different time  $t$ , a set of equations in the form of Eq. (84) can be established for reconstruction. The direct neighbors of node  $i$  (the  $i$ th column of the adjacency matrix  $\mathcal{A}$ ) can be recovered by estimating the coefficients for  $\mathbf{F}_i$  in the local neighborhoods in the phase space. The coefficients can be solved from the set of equations by using a standard  $L_1$ -minimization procedure. After the coefficients have been estimated, it is necessary to set a threshold to discern the neighbors of node  $i$ , where true and false positive rates can be used to quantify the reconstruction performance. Discrete dynamical systems were used to demonstrate the working of the reconstruction method [10].

#### 4.4 Reconstruction of oscillator networks based on noise induced dynamical correlation

The effect of noise on the dynamics of nonlinear and complex system has been intensively investigated in the field of nonlinear science. For example, the interplay between nonlinearity and stochasticity can lead to interesting phenomenon such as stochastic resonance [284–289], where a suitable amount of noise can counter-intuitively optimize the characteristics of the system output such as the signal-to-noise ratio. Previous studies also established the remarkable phenomenon of noise-induced frequency [290] or coherence resonance [291–293]. Specifically, when a nonlinear oscillator is under stochastic driving, a dominant Fourier frequency in its oscillations can emerge, resulting in a signal that can be much more temporarily regular than that without noise [290–293]. A closely related phenomenon is noise-induced collective oscillation or stochastic resonance in



the absence of an external periodic driving in excitable dynamical systems [294]. In complex networks, there was a study of the effect of noise on the fluctuation of nodal states about the synchronization manifold [295], with a number of scaling properties uncovered.

The rich interplay between nonlinear dynamics and stochastic fluctuations pointed at the possibility that noise may be exploited for network reconstruction, leading to the development of a number of methods [13, 15, 27]. In general, these methods do not assume any *a priori* knowledge about the nodal dynamics and there is no need to impose any external perturbation. Under the condition that the influence of noise on the evolution of infinitesimal tangent vectors in the phase space of the underlying networked dynamical system is dominant, it can be argued [15, 27] that the dynamical correlation matrix that can be computed readily from the available nodal time series approximates the network adjacency matrix, fully unveiling the network topology.

In a general sense, noise is beneficial for network reconstruction. Suppose that all nodes in an oscillator network are in a synchronous state. Without external perturbation, the coupled oscillators behave as a single oscillator so that the effective interactions among the oscillators vanish, rendering impossible to extract the interaction pattern from measurements. However, noise can induce desynchronization so that the time series would contain information about the underlying interaction patterns.

It was demonstrated [15, 27] that noise can bridge dynamics and the network topology in that nodal interactions can be inferred from the noise-induced correlations. Consider  $N$  *nonidentical* oscillators, each of which satisfies  $\dot{\mathbf{x}}_i = \mathbf{F}_i(\mathbf{x}_i)$  in the absence of coupling, where  $\mathbf{x}_i$  denotes the  $d$ -dimensional state variable of the  $i$ th oscillator. Under noise, the dynamics of the whole coupled-oscillator system can be expressed as:

$$\dot{\mathbf{x}}_i = \mathbf{F}_i(\mathbf{x}_i) - c \sum_{j=1}^N L_{ij} \mathbf{H}(\mathbf{x}_j) + \eta_i, \quad (85)$$

where  $c$  is the coupling strength,  $\mathbf{H}: \mathbb{R}^d \rightarrow \mathbb{R}^d$  denotes the coupling function of oscillators,  $\eta_i$  is the noise term,  $L_{ij} = -1$  if  $j$  connects to  $i$  (otherwise 0) for  $i \neq j$  and  $L_{ii} = -\sum_{j=1, j \neq i}^N L_{ij}$ . Due to nonidentical oscillators and noise, an invariant synchronization manifold does not exist. Let  $\bar{\mathbf{x}}_i$  be the counterpart of  $\mathbf{x}_i$  in the absence of noise, and assume a small perturbation  $\xi_i$ , we can write  $\mathbf{x}_i = \bar{\mathbf{x}}_i + \xi_i$ . Substituting this into Eq. (85), we obtain:

$$\dot{\xi} = [D\hat{\mathcal{F}}(\bar{\mathbf{x}}) - c\hat{\mathcal{L}} \otimes D\hat{\mathcal{H}}(\bar{\mathbf{x}})] \cdot \xi + \eta, \quad (86)$$

where  $\xi = [\xi_1, \xi_2, \dots, \xi_N]^T$  denotes the deviation vector,  $\eta = [\eta_1, \eta_2, \dots, \eta_N]^T$  is the noise vector,  $\hat{\mathcal{L}}$  names the Laplacian matrix of elements  $L_{ij}$  ( $i, j = 1, \dots, N$ ),

$$D\hat{\mathcal{F}}(\bar{\mathbf{x}}) = \text{diag}[D\hat{\mathcal{F}}_1(\bar{\mathbf{x}}_1), D\hat{\mathcal{F}}_2(\bar{\mathbf{x}}_2), \dots, D\hat{\mathcal{F}}_N(\bar{\mathbf{x}}_N)]$$

and  $D\hat{\mathcal{F}}_i$  are  $d \times d$  Jacobian matrices of  $\mathbf{F}_i$ ,  $\otimes$  denotes direct product, and  $D\hat{\mathcal{H}}$  is the Jacobian matrix of the coupling function  $\mathbf{H}$ .

Let  $\hat{\mathcal{C}}$  denote the dynamical correlation matrix of oscillators  $\langle \xi \xi^T \rangle$ , wherein  $C_{ij} = \langle \xi_i \xi_j \rangle$  and  $\langle \cdot \rangle$  is the time average. One obtains

$$0 = \langle d(\xi \xi^T)/dt \rangle = -\hat{\mathcal{A}} \cdot \hat{\mathcal{C}} - \hat{\mathcal{C}} \cdot \hat{\mathcal{A}}^T + \langle \eta \xi^T \rangle + \langle \xi \eta^T \rangle, \quad (87)$$

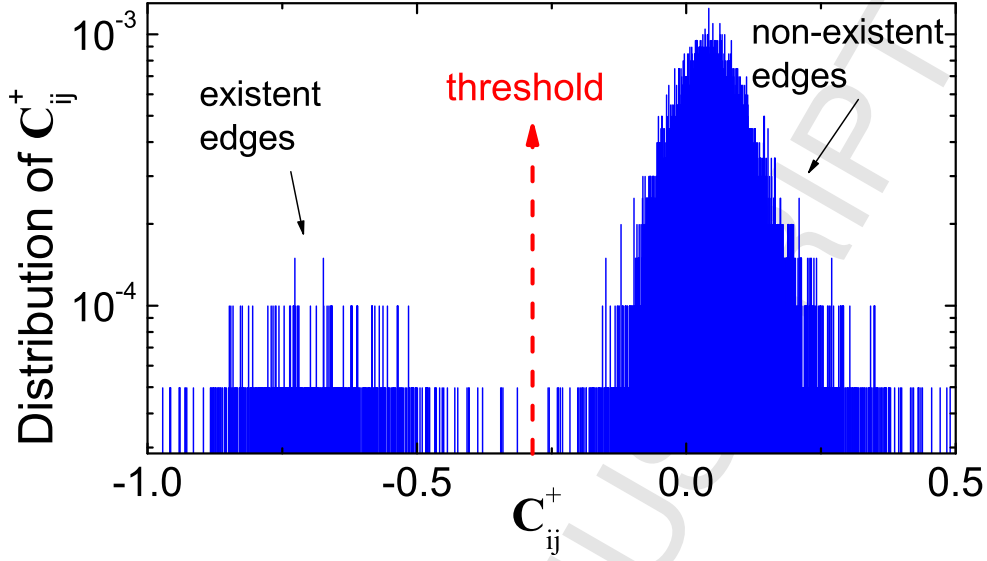


Figure 35: **Reconstruction of an oscillator network based on noise-induced dynamical correlation.** Illustration of the distribution of the element values of the pseudo inverse matrix of the dynamical correlation matrix  $\hat{C}$ . The theoretical threshold is marked by red dashed arrow.

where  $\hat{\mathcal{A}} = -D\hat{\mathcal{F}}(\bar{\mathbf{x}}) + c\hat{\mathcal{L}} \otimes D\hat{\mathcal{H}}(\bar{\mathbf{x}})$ .  $\xi(t)$  can be solved from Eq. (86), yielding the expression of  $\langle \eta \xi^T \rangle$  and  $\langle \xi \eta^T \rangle$ :

$$\langle \xi \eta^T \rangle = \langle \eta \xi^T \rangle = \hat{\mathcal{D}}/2.$$

As a result, Eq. (87) can be simplified to:

$$\hat{\mathcal{A}} \cdot \hat{C} + \hat{C} \cdot \hat{\mathcal{A}}^T = \hat{\mathcal{D}}. \quad (88)$$

The general solution of  $\hat{C}$  can be written as

$$\text{vec}(\hat{C}) = \text{vec}(\hat{\mathcal{D}})/(\hat{I} \otimes \hat{\mathcal{A}} + \hat{\mathcal{A}} \otimes \hat{I}),$$

where  $\text{vec}(\hat{X})$  is a vector containing all columns of matrix  $\hat{X}$ .

For one-dimensional state variable and linear coupling, with Gaussian white noise  $\hat{\mathcal{D}} = \sigma^2 \hat{I}$ , and negligible intrinsic dynamics  $D\hat{\mathcal{F}}$ , Eq. (88) can be simplified to

$$\hat{\mathcal{L}} \cdot \hat{C} + \hat{C} \cdot \hat{\mathcal{L}}^T = \sigma^2 \hat{I}/c.$$

For undirected networks, the solution of  $\hat{C}$  becomes

$$\hat{C} = \frac{\sigma^2}{2c} \hat{\mathcal{L}}^\dagger, \quad (89)$$

where  $\hat{\mathcal{L}}^\dagger$  denotes the pseudo inverse of matrix  $\mathcal{L}$ . Since  $\hat{C}$  can be calculated from time series, and  $\sigma$  and  $c$  are constant, setting a threshold in the element values of matrix  $\hat{C}$  for identifying real nonzero elements can lead to full reconstruction of all the connections (Fig. 35). A key issue is how to determine the threshold.

In Refs. [15, 27], it was argued that the threshold can be set based on the diagonal elements (autocorrelation) in  $\hat{C}$ :

$$C_{ii} \simeq \frac{\sigma^2}{2ck_i} \left( 1 + \frac{1}{\langle k \rangle} \right). \quad (90)$$

The formula of  $C_{ii}$  obtained from a second-order approximation is consistent with the finding of noise-induced algebraic scaling law in Ref. [13]. Specifically, from Eq. (90), let

$$S \equiv \sum_{i=1}^N 1/C_{ii} = 2cl^2/[\sigma^2(N+l)],$$

where  $l = \sum_{i=1}^N k_i = N\langle k \rangle$  is twice the total number of links. The integral part of  $l$  can be identified via

$$l = (S\sigma^2 + \sqrt{S^2\sigma^4 + 8cNS\sigma^2})/4c.$$

The threshold  $C_M^+$  (or  $[\sigma^2/(2c)]C_M^+$ ) is chosen such that  $\sum_{m=1}^M \Phi(C_m^+) = l$ , where  $\Phi(C_m^+)$  is the unnormalized distribution of  $C_m^+$ . Then the connection matrix  $\mathcal{L}$  can be obtained (see Fig. 35). Figure 36 shows the scaling property of autocorrelation versus the node degree  $k$  for different nodal dynamics and different types of networks, as predicted by Eq. (90). The numerical results are in good agreement with the theoretical prediction.

The reconstruction method by virtue of noise was improved [296], based on reconstruction formula (89). It was argued that, to use the formula, the coupling strength  $c$  and the noise variance  $\sigma$  should be known *a priori*, a condition that may be difficult to meet in realistic situations. A simple method to eliminate both  $c$  and  $\sigma$  without requiring any *a priori* parameters was then introduced [296]. Specifically, note that

$$\begin{aligned} \frac{\sigma^2}{2c} C_{ii}^+ &= k_i, \\ \frac{\sigma^2}{2c} C_{ij}^+ &= -A_{ij}, \quad i \neq j, \end{aligned} \quad (91)$$

where  $\sigma$  and  $c$  can be simultaneously eliminated by the ratio  $r_{ij}$  of the diagonal and off-diagonal elements of  $\hat{C}^+$  in Eq. (91):

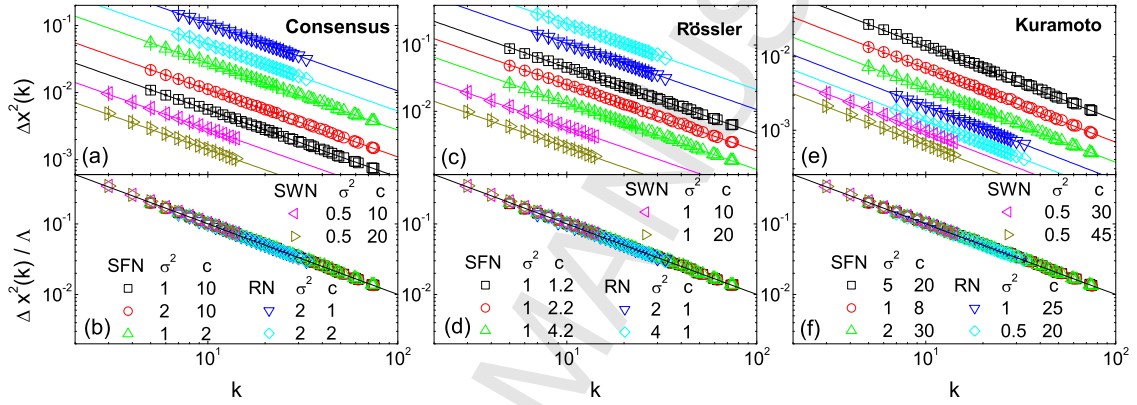
$$r_{ij} = \begin{cases} 0, & i \text{ and } j \text{ are disconnected,} \\ -\frac{1}{k_i}, & i \text{ and } j \text{ are connected.} \end{cases} \quad (92)$$

A threshold is still necessary to fully separate links and zero elements in the adjacency matrix. A heuristic method was developed [296] for determining the threshold. While based on the same principle, the improved reconstruction method appears indeed more practical.

An alternative method to reconstruct the network topology of a dynamical system contaminated by white noise was developed [297] through measurement of both the nodal state  $\mathbf{x}$  and its velocity  $\dot{\mathbf{x}}$ , with the following formula:

$$\hat{\mathcal{A}} = \hat{\mathcal{B}} \cdot \hat{C}^{-1}, \quad (93)$$

where  $\hat{C} = \langle \mathbf{x}\mathbf{x}^T \rangle$  is the state-state correlation matrix, and  $\hat{\mathcal{B}} = \langle \dot{\mathbf{x}}\mathbf{x}^T \rangle$  is the velocity-state correlation matrix. An interesting feature is that, by measuring the velocity, the white noise variance is



**Figure 36: Scaling of noise-induced fluctuations with nodal degree.** Average fluctuation  $\Delta x^2(k)$  as a function of the nodal degree  $k$  for different values of noise variance  $\sigma^2$  and coupling strength  $c$  for scale-free, random and small-world networks and three types of nodal dynamical processes [(a) consensus dynamics, (c) Rössler dynamics, and (e) Kuramoto dynamics]. (b,d,f) Rescaled quantity  $\Delta x^2(k)/\Lambda$ , where  $\Lambda = \sigma^2(1 + 1/\langle k \rangle)/(2c)$ , versus  $k$  for consensus, Rössler and Kuramoto dynamics, respectively. Data points are from a single network configuration and  $\Delta x^2(k)$  is obtained by averaging over all nodes of degree  $k$  with error bars. The parallel lines in (a), (c) and (e) are theoretical predictions, and the lines in (b), (d) and (f) are the function  $1/k$ . Network size is 500. For the scale-free network, the lowest degree is  $k_{\min} = 5$ . For the random network, the connection probability among nodes is 0.03. For the small-world network, the average degree is 8 and the rewiring probability is 0.1. The natural frequency  $\omega_i$  in the Kuramoto model is chosen independently from a prescribed probability distribution  $g(\omega) = 3/4(1 - \omega^2)$  for  $|\omega| \leq 1$  and  $g(\omega) = 0$  otherwise. From Ref. [15] with permission.

not required to be known and the formula applies (in principle) to any strength of noise. However, measuring the instantaneous velocity of the state variables can be difficult and the errors would affect the reconstruction performance dramatically. There is in fact a trade-off between the generality of the reconstruction method and the measurement accuracy. Nevertheless, the work [297] is theoretically valuable, and it was validated by using a linear system and a nonlinear cell cycle dynamics model. In the nonlinear model, there are both active and inhibitive links, corresponding to positive and negative elements in the adjacency matrix. It was demonstrated that both classes of links can be successfully reconstructed [297].

## 4.5 Reverse engineering of complex systems

**Automated reverse engineering of nonlinear dynamical systems.** Reconstructing nonlinear biological and chemical systems is of great importance. An automated reverse engineering approach was developed to solve the problem by using partitioning, automated probing and snipping [151]. Firstly, partitioning allows the algorithm to characterize the variables in a nonlinear system separately by decoupling the interdependent variables. It was argued that Bayesian networks cannot model mutual dependencies among the variables, a common pattern in biological and other regulation networks with feedback. It was also articulated that the partitioning procedure is able to reveal the underlying structure of the system to a higher degree as compared with alternative methods. Secondly, instead of passively accepting data to model a system, automated probing uses automatically synthesized models to create a test criterion to eliminate unsuitable models. Because the test cannot be analytically derived, candidate tests are optimized to maximize the agreement between data and model predictions. Thirdly, snipping automatically simplifies and optimizes models. For all models obtained by automated probing, one calculates their prediction errors against system data, perturbs the existent models randomly, evaluate the newly created models, and compare the performance of the modified models with that of the corresponding original models. In this way, models are constantly evolving, with those with better performance replacing the inferior ones. This process tends to yield more accurate and simpler models. The process is akin to a genetic algorithm, but in the former, models evolve and are constantly improved, whereas in the latter, some adaptive functions evolve.

Four synthetic systems and two physical systems were used [151] to test the automated reverse engineering approach. The method was demonstrated to be robust against noise and scalable to interdependent and nonlinear systems with many variables. The limitations of the method were discussed as well. Especially, the method is restricted to systems in which all variables are observable. In addition, discrete time series as input to the method can lead to inconsistency with the synthesized models. Without including any fuzzy effect, the process may not yield explicit models. Potential solutions to these problems were suggested [151].

**Constructing minimal models for complex system dynamics.** In Ref. [298], the authors introduced a method of reverse engineering to recover a class of complex networked system described by ordinary differential equations of the form

$$\frac{dx_i}{dt} = M_0(x_i(t)) + \sum_{j=1}^N A_{ij} M_1(x_i(t)) M_2(x_j(t)), \quad (94)$$

where the adjacency matrix  $\mathcal{A}$  defines the interacting components,  $M_0(x_i(t))$  describes the self-dynamics of each node,  $M_1(x_i(t))M_2(x_j(t))$  characterizes the interaction between nodes  $i$  and  $j$ . Equation (94) was argued to be suitable for many social, technological and social systems.

The system dynamics is uniquely characterized by three independent functions and the aim is to construct the model in the form

$$\mathbf{m} = (M_0(x), M_1(x), M_2(x)), \quad (95)$$

corresponding to a point in the model space  $\mathbb{M}$ . The goal was to develop a general method to infer the subspace  $\mathbb{M}(\mathcal{X})$  by relying on minimal *a priori* knowledge of the structure of  $M_0(x)$ ,  $M_1(x)$ , and  $M_2(x)$ . The key lies in using the system's response to external perturbations, a common technique used in biological experiments such as genetic perturbation, which is feasible for technological and social networked systems as well. The link between the observed system response and the leading terms of  $\mathbf{m}$  was established, enabling the formulation of  $\mathcal{X}$  into a dynamical equation. Contrary to traditional reverse engineering, the approach [298] gives the boundaries of  $\mathbb{M}(\mathcal{X})$  rather than a specific model  $\mathbf{m}$ .

To infer  $\mathbf{m}$ , the three functions were expressed [298] in terms of a Hahn series [299]

$$\begin{aligned} M_0(x) &= \sum_{n=0}^{\infty} A_n (x_0 - x)^{\Pi_0(n)} \\ M_1(x) &= \sum_{n=0}^{\infty} B_n (x_0 - x)^{\Pi_1(n)} \\ M_2(x) &= \sum_{n=0}^{\infty} C_n (x_0 - x)^{\Pi_2(n)}. \end{aligned} \quad (96)$$

The challenge was to discern the coefficients  $A_n$ ,  $B_n$  and  $C_n$  for uncovering the functional form (96), which can be overcome by exploiting the system's response to external perturbation.

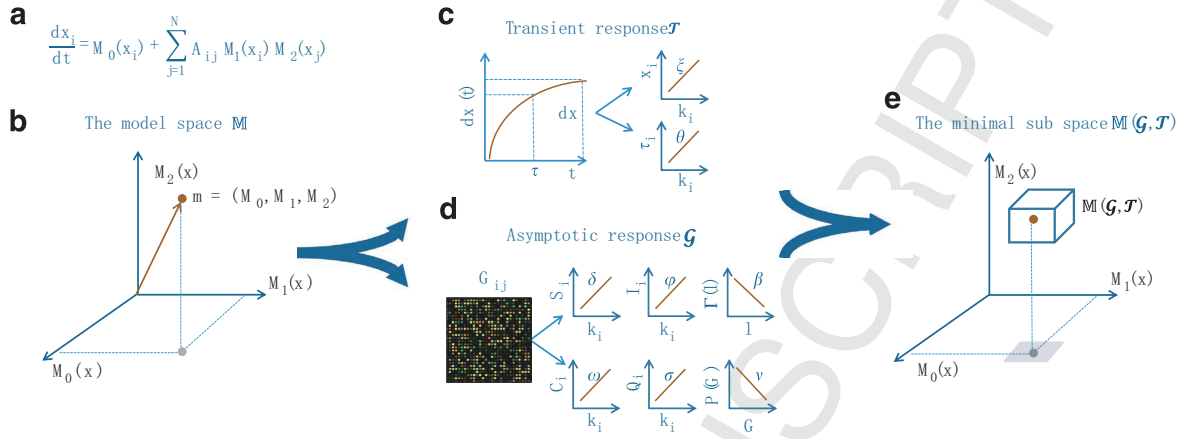
Under perturbation, the temporal dynamics is characterized by the time-dependent relaxation from the original steady state,  $x_i$ , to the perturbed steady-state  $x_i(t \rightarrow \infty) = x_i + dx_i$ . System (94) can then be linearized about the steady state. After relaxation, the system's new state is characterized by a response matrix. Based on the permanent perturbation,  $\mathbf{m}$  can be estimated via

$$\begin{aligned} M_0(x) &\sim \frac{1}{(R(x))^2} \int \left[ (R(x))^{2+\theta} + O((R(x))^{\phi+(2+\theta)}) \right] dx \\ M_1(x) &\sim M_0(x)R(x) \\ M_2(x) &\sim \begin{cases} (R(x))^{\delta+1-\phi} + O((R(x))^{\phi+(\delta+1-\phi)}), & \beta = 0 \\ y_0 - (R(x))^\beta + O((R(x))^{\phi+\beta}), & \beta > 0 \end{cases} \end{aligned} \quad (97)$$

where

$$R(x) \sim \begin{cases} x^{-\frac{1}{\xi}} + O(x^{\phi(-\frac{1}{\xi})}), & \delta = 0 \\ (x_0 - x)^{\frac{1}{\delta}} + O((x_0 - x)^{\phi(\frac{1}{\delta})}), & \delta > 0 \end{cases} \quad (98)$$

and  $x_0$  and  $y_0$  are arbitrary constants. A set of parameters need to be determined, which exhibit certain scaling properties. Specifically, the relaxation time scales with  $\theta$ , i.e.,  $\tau_i \sim k_i^\theta$ , the steady-state activity either scales with  $\xi$  as  $x_i \sim k_i^\xi$  or has the relation  $x_i \sim k_i^\xi$ , node  $i$ 's impact on neighbors obeys  $I_i \sim k_i^\phi$ ; and  $i$ 's stability against perturbation in its vicinity leads to  $S_i \sim k_i^\delta$  (see Fig. 37).



**Figure 37: Constructing a minimal model for complex dynamics.** (a) The unknown nonlinear equation of a dynamical system, which is to be reconstructed from empirical observations. (b) The dynamic model  $\mathbf{m}$  is composed of three functions:  $M_0(x)$ ,  $M_1(x)$  and  $M_2(x)$ , corresponding to a point in the model space,  $\mathbb{M}$ . (c,d) The system's response to perturbations is used to extract a set of observable functions directly linked to  $\mathbf{m}$ . (c) The transient response is characterized by the steady-state  $\xi_i$  and the relaxation time  $\tau_i$ , from which the exponents  $x$  and  $y$  can be extracted. (d) The asymptotic response for measuring six additional functions, providing  $\delta$ ,  $\phi$ ,  $\beta$ ,  $\omega$ ,  $\sigma$  and  $v$ . (e) The subspace in the model space. From Ref. [298] with permission.

The determination of the parameters gives only the leading terms of  $M_0(x)$ ,  $M_1(x)$  and  $M_2(x)$ . Additional terms can be used to better model complex systems. The reconstructed model is called a minimal model because it captures the essential features of the underlying mechanism without any additional constraint. The subspace given by the reconstruction method is robust to parameter selection. Models subject to the subspace can produce consistent data with observations [298].

## 5 Inference approaches to reconstruction of biological networks

High-throughput technologies such as microarrays and RNA sequencing produce a large amount of experimental data, making genome-scale inference of transcriptional gene regulation possible. The reconstruction of biological interactions among genes is of paramount importance to biological sciences at different levels. A variety of approaches aiming to reconstruct gene co-expression networks or regulation networks have been developed.

In general, there are three types of experimental gene data: gene co-expression data, gene knockout data, and transcriptional factors. Gene interaction networks can be classified into two categories: (1) co-expression networks, in which the nodes represent genes and the edges represent the degree of similarity in the expression profiles of the genes, and (2) transcription-regulatory networks, in which the nodes represent either transcription factors or target genes, and edges characterize the causal regulatory relationships.

Reconstructing neuronal networks and brain functional networks from observable data has also been an active area of research. Typical experimental data include blood oxygen level dependent (BOLD) signals, electroencephalography (EEG) and stereoelectroencephalography (SEEG) data, magnetoencephalography (MEG) data, spike data, calcium image data, etc. To uncover the interaction structure among the neurons or distinct brain domains from the signals, a number of reconstruction approaches have been developed and utilized in neuroscience.

### 5.1 Correlation based methods

Correlation based methods are widely used for inferring the associations between two variables.

#### 5.1.1 Value based methods

*Pearson's correlation coefficient* measuring the strength of the linear relationship between two random variables is widely used in many fields [300]. The correlation is defined as

$$\text{corr} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n-1)S_x S_y}, \quad (99)$$

where  $\bar{x}$  and  $\bar{y}$  are the sample means and  $S_x$  and  $S_y$  are the standard deviations of  $x$  and  $y$ , respectively. The Pearson's correlation assumes that data is normally distributed and is sensitive to outliers.

*Distance Covariance* (dCov) provides a nonparametric test to examine the statistical dependence of two variables [301]. For some given pairs of measurement  $(x_i, y_i), i = 1, 2, \dots, n$  for variables  $X$  and  $Y$ , let  $\mathcal{A}$  denote the pairwise Euclidean distance matrix of  $X$  with  $a_{ij} = |x_i - x_j|$  and  $\mathcal{B}$  be the corresponding matrix for  $Y$ , where  $|\cdot|$  denote the Euclidean norm. Define the doubly centered distance matrix  $\mathcal{A}^c$  whose elements are given by  $a_{ij}^c = a_{ij} - \bar{a}_{i\cdot} - \bar{a}_{\cdot j} + \bar{a}_{\cdot\cdot}$ , where  $\bar{a}_{i\cdot}$  is  $i^{\text{th}}$  row mean,  $\bar{a}_{\cdot j}$  is the  $j^{\text{th}}$  column mean, and  $\bar{a}_{\cdot\cdot}$  is the grand mean of  $\mathcal{A}$ . Centered distance matrix  $\mathcal{B}^c$  can be defined similarly. The squared sample distance covariance is defined to be the arithmetic average of the products of  $\mathcal{A}^c$  and  $\mathcal{B}^c$ :

$$\text{dcov}^2(X, Y) = \frac{1}{n^2} \sum_{ij} a_{ij}^c b_{ij}^c. \quad (100)$$



The *Theil-Sen estimator*, proposed in Refs. [302, 303], is defined as

$$\hat{\beta}_1 = \text{median}\{m_{ij} = \frac{y_i - y_j}{x_i - x_j} : x_i \neq x_j, 1 \leq i \leq j \leq n\}, \quad (101)$$

where the median can characterize the relationship between the two variables. This estimate is robust, unbiased, and less sensitive to outliers.

*Partial correlation and information theory* (PCIT) [304, 305] extracts all possible interaction triangles and applies Data Processing Inequality (DPI) to filter indirect interactions using partial correlation. The partial correlation coefficient  $c_{ij}^p$  between two genes  $i$  and  $j$  within an interaction triangle  $(i, j, k)$  is defined as

$$c_{ij}^p = \frac{\text{corr}(x_i, x_j) - \text{corr}(x_i, x_k)\text{corr}(x_j, x_k)}{\sqrt{(1 - \text{corr}(x_i, x_k))^2(1 - \text{corr}(x_j, x_k))^2}}, \quad (102)$$

where  $\text{corr}(\cdot, \cdot)$  is Pearson's correlation coefficient.

### 5.1.2 Rank based methods

Compared with value based correlation measures, rank based correlation measures are more robust and insensitive to outliers. The *Spearman and Kendall measures* are the most commonly used. In particular, Spearman's correlation is simply the Pearson's correlation coefficient incorporated with ranked expression [306], and Kendall's  $\tau$  coefficient [307] is defined as

$$\tau(x_i, x_j) = \frac{\text{con}(x_i^r, x_j^r) - \text{dis}(x_i^r, x_j^r)}{\frac{1}{2}n(n-1)}, \quad (103)$$

where  $x_i^r$  and  $x_j^r$  are the ranked expression profiles of genes  $i$  and  $j$ ,  $\text{con}(\cdot, \cdot)$  and  $\text{dis}(\cdot, \cdot)$  represent the numbers of concordant and discordant pairs, respectively.

*Inner Composition Alignment* (ICA) [19] was proposed to infer directed networks from short time series by extending the Kendall's  $\tau$  measure. Given time series  $x^l$  and  $x^k$  of length  $n$  from subsystems  $l$  and  $k$  over the same time intervals, let  $\pi^l$  be the permutation that arranges  $x^l$  in a nondecreasing order. The series  $g^{k,l} = x^k(\pi^l)$  is the reordering of the time series  $x^k$  with respect to  $\pi^k$ . The ICA is formulated as

$$\tau^{l \rightarrow k} = 1 - \frac{\sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} w_{ij} \Theta[(g_{j+1}^{k,l} - g_i^{k,l})(g_i^{k,l} - g_j^{k,l})]}{\frac{1}{2}(n-1)(n-2)}, \quad (104)$$

where  $w_{ij}$  denotes the weight between points  $i$  and  $j$ , and  $\Theta[x]$  is the Heaviside step function. Compared with Kendall's  $\tau$  measure, ICA can infer direct interactions and eliminate indirect interaction by using the partial version of ICA.

*Hoeffding's D coefficient* is a rank-based nonparametric measure of association [308]. The statistic coefficient  $D$  is defined as

$$D = \frac{(n-2)(n-3)D_1 + D_2 - 2(n-2)D_3}{n(n-1)(n-2)(n-3)(n-4)}, \quad (105)$$

where

$$D_1 = \sum_{i=1}^n Q_i(Q_i - 1), \quad (106)$$

$$D_2 = \sum_{i=1}^n (R_i - 1)(R_i - 2)(S_i - 1)(S_i - 2), \quad (107)$$

$$D_3 = \sum_{i=1}^n (R_i - 2)(S_i - 2)Q_i, \quad (108)$$

$R_i$  is the rank of  $X_i$ ,  $S_i$  is the rank of  $Y_i$ , and the bivariate rank  $Q_i$  is the number of both  $X$  and  $Y$  values less than the  $i^{\text{th}}$  point.

It was pointed out [309] that real gene interactions may change as the intrinsic cellular state varies or may exist only under a specific condition. That is, for a long time series of the co-expression data, perhaps only part of the data are meaningful for revealing interactions. In this regard, two new co-expression measures based on the matching patterns of local expression ranks were proposed. Specifically, when dealing with time-course data, the measure  $W_1$  is defined as

$$W_1 = \sum_{i=1}^{n-k+1} I(\phi(x_i, \dots, x_{i+k-1}) = \phi(y_i, \dots, y_{i+k-1})) + I(\phi(x_i, \dots, x_{i+k-1}) = \phi(-y_i, \dots, -y_{i+k-1})), \quad (109)$$

where  $I(\cdot)$  is an indicator function, and  $\phi$  is the rank function that returns the indices of the elements after they have been sorted in an increasing order, and  $W_1$  counts the number of continuous subsequences of length  $k$  with matching and reverse rank patterns. For non-time-series data, where the order is not meaningful, a more general measure,  $W_2$  can be defined:

$$W_2 = \sum_{1 \leq i_1 < \dots < i_k \leq n} I(\phi(x_{i_1}, \dots, x_{i_k}) = \phi(y_{i_1}, \dots, y_{i_k})) + I(\phi(x_{i_1}, \dots, x_{i_k}) = \phi(-y_{i_1}, \dots, -y_{i_k})). \quad (110)$$

## 5.2 Causality based measures

*Wiener-Granger Causality* (WGC), pioneered by Wiener [310] and Granger [119], is a commonly used measure to infer the causal influence between two variables, e.g., the expression time series of genes and the spiking neural time series. The basic idea of WGC is straightforward. Given two time series  $\{X_t\}$  and  $\{Y_t\}$ , if using the history of both  $X$  and  $Y$  is more successful to predict  $X_{t+1}$  than exclusively using the history of  $X$ ,  $Y$  is said to be G-cause  $X$ . The idea of WGC is similar to that of the transfer entropy, but WGC uses some correlation measure instead of mutual information. WGC is widely used in measuring the functional connectivity among subdomains of brain based on EEG, MEG, SEEG data [311]. Fundamentally, Granger test is a linear method operated on the hypothesis that the underlying system can be described as a multivariate stochastic process. Thus, in principle, there is no guarantee that the method would be effective for nonlinear systems, in spite of efforts to extend the methodology to strongly coupled systems [312–314].

In the traditional Granger framework, measurement noise is generally detrimental in the sense that, as its amplitude is increased the value of the detected causal influence measure decreases monotonically, leading to spurious detection outcomes [315]. The transfer entropy framework

is applicable [316] to both linear and nonlinear systems, but often the required data amount is prohibitively large. In the special case of Gaussian dynamical variables, the two methods, one of the autoregressive nature (Granger test) and another based on information theoretic concepts (transfer entropy, to be discussed below), are in fact equivalent to each other [317]. An alternative information theoretic measure, the causation entropy, was also proposed [318–320].

*Convergent cross mapping framework* (CCM) is based on delay coordinate embedding, the paradigm of nonlinear time series analysis [43, 44, 60, 321]. The CCM method can deal with both linear and nonlinear systems with small data sets, and it has been applied to data from different contexts, such as EEG data [322], FMRI [323], fishery data [324], economic data [325], and cerebral auto-regulation data [326]. It was also found that properly applied noise can enhance the CCM performance in inferring causal relations [327].

The nonlinear dynamics based CCM method was proposed [328] to detect and quantify causal influence between a pair of dynamical variables through the corresponding time series. The starting point is to reconstruct a phase space, for each variable, based on the delay-coordinate embedding method [43]. Specifically, for time series  $x(t)$ , the reconstructed vector is  $X(t) = [x(t), x(t - \tau), \dots, x(t - (E_x - 1)\tau)]$ , where  $\tau$  is the delay time and  $E_x$  is the embedding dimension. For variable  $y$ , a similar vector can be constructed in the  $E_y$  dimensional space. Let  $\mathbf{M}_X$  and  $\mathbf{M}_Y$  denote the attractor manifolds in the  $E_x$ - and  $E_y$ -dimensional space, respectively. If  $x$  and  $y$  are dynamically coupled, there is a mapping relation between  $\mathbf{M}_X$  and  $\mathbf{M}_Y$ . The CCM method measures how well the local neighborhoods in  $\mathbf{M}_X$  correspond to those in  $\mathbf{M}_Y$ . In particular, the cross-mapping estimate of a given  $Y(t)$ , denoted as  $\hat{Y}(t)|\mathbf{M}_X$ , is based on a simplex projection [93, 329] that is essentially a *nearest-neighbor algorithm* involving  $E + 1$  nearest neighbors of  $X(t)$  in  $\mathbf{M}_X$ . (Note that  $E + 1$  is the minimum number of points required for a bounding simplex in the  $E$ -dimensional space.) The time indices of the  $E + 1$  nearest neighbors are denoted as  $t_1, t_2, \dots, t_{E+1}$  in the order of distances to  $X(t)$  from the nearest to the farthest, i.e., point  $X(t_1)$  is the nearest-neighbor point of  $X(t)$  in  $\mathbf{M}_X$ . These time indices are used to identify the points (putative neighborhoods) in  $\mathbf{M}_Y$ , namely, to find the points at the corresponding instants:  $Y(t_1), Y(t_2), \dots$ , and  $Y(t_{E+1})$ , which are used to estimate  $\hat{Y}(t)$  through the weighted average

$$\hat{Y}(t)|\mathbf{M}_X = \sum_{i=1}^{E+1} w_i(t) \cdot Y(t_i), \quad (111)$$

where

$$w_i(t) = \mu_i(t) / \sum_j \mu_j(t) \quad (112)$$

is the weight of the vector  $Y(t_i)$ ,

$$\mu_i(t) = \exp\{-d[X(t), X(t_i)]/d[X(t), X(t_1)]\}, \quad (113)$$

and  $d[X(t), X(t_i)]$  is the Euclidean distance between the two vector points  $X(t)$  and  $X(t_i)$  in  $\mathbf{M}_X$ . An estimated time series  $\hat{y}(t)$  can then be obtained from  $\hat{Y}(t)|\mathbf{M}_X$ . Likewise, the cross mapping from  $Y$  to  $X$  can be defined analogously so that the time series of  $x(t)$  can be predicted from the cross-mapping estimate  $\hat{X}(t)|\mathbf{M}_Y$ .

The correlation coefficient between the original time series  $y(t)$  and the predicted time series  $\hat{y}(t)$  from  $\mathbf{M}_X$ , denoted as  $\rho_{Y|\mathbf{M}_X}$ , is a measure of CCM causal influence from  $y$  to  $x$ . Larger value

of  $\rho_{Y|M_X}$  implies that  $y$  is a stronger cause of  $x$ , while  $\rho_{Y|M_X} \leq 0$  indicates that  $y$  has no influence on  $x$ . The relative strength of causal influence can be defined as  $R = \rho_{X|M_Y} - \rho_{Y|M_X}$ , which is a quantitative measure of the casual relationship between  $x$  and  $y$ . A positive value of  $R$  indicates that  $x$  is the CCM cause of  $y$ .

### 5.3 Information-theoretic based methods

#### 5.3.1 Mutual information

*Mutual Information* (MI) [330] is used widely to quantify the pairwise mutual dependency between two variables. For two variables  $X$  and  $Y$ , the mutual information is defined in term of

$$\mathbf{MI}(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \left( \frac{p(x,y)}{p(x)p(y)} \right), \quad (114)$$

where  $p(x,y)$  is the joint probability distribution functions of  $X$  and  $Y$ , and  $p(x)$  and  $p(y)$  are the marginal probability distribution function of  $X$  and  $Y$ , respectively. Note that MI is nonnegative and symmetric:  $\mathbf{MI}(X;Y) = \mathbf{MI}(Y;X)$ . The larger the value of MI, the more highly the two variable are correlated. The key to applying the mutual information for quantifying associations in continuous data is to estimate the probability distribution from the data. A difficult and open issue is how to obtain unbiased mutual information value.

There are four commonly used MI-based network reconstruction methods: relevance network approach (RELNET) [331, 332], context likelihood of relatedness (CLR) [333], maximum relevance/minimum redundancy feature selection (MRNET) [334, 335], and the algorithm for accurate reconstruction of cellular networks (ARACNE) [336, 337].

In RELNET, a threshold  $\tau$  is used to distinguish actual links from null connections. However, this method has a limitation, i.e., the indirect relationship may lead to high mutual information values and failures to distinguish direct from indirect relationships.

The CLR method integrates z-score to measure the significance of the calculated MI. For node  $i$ , the mean  $\mu_i$  and the standard deviation  $\delta_i$  of the empirical distribution of the mutual information  $\mathbf{MI}(x_i; x_k)$ , ( $k = 1, \dots, n$ ) are calculated. The z-score  $z_{ij}$  of  $\mathbf{MI}(x_i; x_j)$  is defined as

$$z_{ij} = \max \left( 0, \frac{\mathbf{MI}(x_i; x_j) - \mu_i}{\delta_i} \right). \quad (115)$$

The value of  $z_{ji}$  can be defined analogously. A combined score, which quantifies the relatedness between  $i$  and  $j$ , is expressed as  $\hat{z}_{ij} = \sqrt{z_{ij}^2 + z_{ji}^2}$ .

The MRNET method can be used to infer a network by repeating the maximum relevance, minimum redundancy (MRMR) [338, 339] feature selection method for all variables. The key to the MRMR algorithm lies in selecting a set of variables with a high value of the mutual information with the target variable (maximum relevance) but meanwhile the selected variables are mutually maximally independent (minimum redundancy). The aim of this method is to associate direct interactions with high rank values, but associate indirect interactions with low rank values.

The MRMR algorithm is essentially a greedy algorithm. Firstly, the MRMR selects the variable  $x_i$  with the highest MI value regarding the target  $y$ . Next, given a set of selected variables, a new

variable  $x_k$  is chosen to maximize the value of  $s_k^y = u_k^y - r_k^y$ , where  $u_k^y$  is the relevance term and  $r_k^y$  is the redundancy term. More precisely, one has

$$u_k^y = MI(x_k; y), \quad (116)$$

$$r_k^y = \frac{1}{|S|} \sum_{x_l \in S} MI(x_k; x_l). \quad (117)$$

By setting each gene to be the target, one can calculate all the values of  $s_{x_j}^{x_i}$ , and the relatedness value between any pair of  $x_i$  and  $x_j$  is the maximum of  $(s_{x_j}^{x_i}, s_{x_i}^{x_j})$ . A modification of MRNET can be found in Ref. [335].

ARACNE is the extension of RELNET. In the ARACNE, Data Processing Inequality (DPI) [340], a well-known information theoretic property, is used to overcome the limitation in the RELNET. Specifically, the DPI stipulates that if variables  $x_i$  and  $x_j$  interact only through a third variable  $x_k$ , one has

$$MI(x_i; x_j) \leq \min(MI(x_i, x_k), MI(x_j, x_k)), \quad (118)$$

which is used to eliminate indirect interactions. ARACNE starts from a network in which the value of MI for each edge is larger than  $\tau$ . The weakest edge in each triplet, for example the edge between  $i$  and  $j$ , is regarded as an indirect interaction and is removed if

$$MI(x_i; x_j) \leq \min(MI(x_i, x_k), MI(x_j, x_k)) - \epsilon, \quad (119)$$

where  $\epsilon$  is a tolerance parameter.

### 5.3.2 Maximal information coefficient

*Maximal Information Coefficient* (MIC) is a recently proposed information based method, belonging to the larger class of maximal information-based, non-parametric exploration statistics [341]. The basic idea of MIC is that, if a relationship exists between two variables, a grid can be drawn on the scatterplot of the two variables to partition the data so that the relationship between the two variables in the two-dimensional diagram can be uncovered.

The MIC of a given pair of data  $D$  is calculated, as follows. First, one partitions the  $x$ -values of  $D$  into  $x$  bins and the  $y$ -values into  $y$  bins according to a uniform partition grid  $G(x, y)$ . One then calculates the mutual information with respect to this partition, which is denoted as  $I(D|G)$ . For fixed  $D$  with a different grid  $G$ , even with the same  $x$  bins and  $y$  bins, the value of the mutual information will be different. A characteristic matrix  $\mathcal{M}(D)$  can be defined as

$$\mathcal{M}(D_{x,y}) = \frac{I^*(D, x, y)}{\log \min(x, y)}, \quad (120)$$

where  $I^*(D, x, y) = \max\{I(D|G)\}$  is the maximum over all grids  $G$  with  $x$  and  $y$  bins. The value of MIC of the variable pair  $D$  is given by

$$\text{MIC}(D) = \max\{\mathcal{M}(D)_{x,y}\}. \quad (121)$$

There were claims [342, 343] that the MIC does not outperform MI, suggesting that the mutual information may be a more natural and practical way to quantify statistical associations.

### 5.3.3 Maximum entropy

One way to model the spiking activity of a population of neurons is through the classic Ising model [344, 345], in terms of the following maximum entropy distribution:

$$P(\sigma_1, \sigma_2, \dots, \sigma_N) = \frac{1}{Z} \exp \left[ \sum_i h_i \sigma_i + \frac{1}{2} \sum_{i \neq j} J_{ij} \sigma_i \sigma_j \right], \quad (122)$$

where  $\sigma_i = \pm 1$  is the spiking activity of cell  $i$  and the normalized factor is the partition function  $Z$  in statistical physics. The effective coupling strengths  $J_{ij}$  are then chosen so that the averages  $(\langle \sigma_i \rangle, \langle \sigma_i \sigma_j \rangle)$  in this distribution agree with the results from numerical or actual experiments.

## 5.4 Bayesian network

The Bayesian networks (BNs) [346, 347] are a commonly used probabilistic graphical model, represented as a direct acyclic graph (DAG), in which nodes represents a set of random variables and the direct links signify their conditional dependencies. For example, in the case of inferring a gene regulatory network, a Bayesian network could represent the probabilistic relationships between transcription factors and their target genes. The BN method is associated with, however, high computational cost (especially for large networks). Unlike the alternative pairwise association estimation methods discussed above, the BN method requires that all possible DAGs be exhaustively or heuristically searched, scored, and kept either as a best-scoring network or a network constructed by averaging the scores of all the networks. In order to optimize the posterior probabilities in BN, a variety of heuristic search algorithms were developed, e.g., simulated annealing, max-min parent and children algorithm [348], Markov blanket algorithm [349], Markov boundary induction algorithm [350]. The optimization algorithms integrated into BN notwithstanding, the method is practically applicable but only to biological networks of relatively small size.

## 5.5 Regression and resampling

Interring gene regulatory networks can be treated as a feature selection problem. The expression level of a target gene can be predicted by its direct regulate transcription factors. It is often assumed that the links among the genes are sparse. For example, the lasso [351] is a widely used regression method for network inference, and it is a shrinkage and selection method for linear regression. Specifically, it minimizes the usual sum of squared errors, with a bound on the  $L_1$  norm of the coefficients, which is used to regularize the model. Given the expression data of genes, in a steady state experiment, the feature selection can be formulated as

$$\min \frac{1}{2n} \|y - Xw\|_2^2 + \lambda \|w\|_1, \quad (123)$$

where  $\lambda$  is an adjustable parameter for controlling the sparsity of the coefficients. A direct use of Lasso to infer networks has two shortcomings: (a) it is known to be an unstable procedure in terms of the selected features, and (b) it does not provide confidence scores for the selected features. As a result, a stability selection procedure is often integrated into Lasso [352]. One

first resamples the data into several sub-data based on bootstrap, then applies the Lasso to solve these sub-data regression problems, and aggregates the final score for each feature to select more confident features.

Beside the steady-state data, time series data can also be tackled using the Lasso in the sense that the current expression level of the transcription factors is the predictors of the change in the expression of the target gene. A group Lasso method using both steady-state and time series data was proposed [353], in which the pair coefficients of a single transcription factor across both steady-state and time-series data are either both zero or both non-zero.

The gene regulation can also be modeled with nonlinear models, e.g., polynomial regression models [354, 355] and sigmoid functions [356].

## 5.6 Supervised and semi-supervised methods

Supervised and semi-supervised methods treat reconstruction as a classification problem. A number of methods were developed to accomplish this goal, with some works showing that the performance of the supervised methods is better than that of the unsupervised methods [357]. Generally, prior to applying a supervised method, two types of training data sets are required: the expression profile of each gene and a list of prior information about whether the known transcriptional factors and genes are regulated.

*Support Vector Machine* (SVM) is a prevailing supervised classification method that has been successfully used in inferring gene regulatory networks. The basic idea of SVM is to find an apparent gap that divide the data points of the separate categories as widely as possible. By integrating the kernel function, SVM can be extended to nonlinear classification. *Gene Network Inference with Ensemble of Trees* (GENIE), similar to MRNET, identifies the best subset of regulator genes with random forest and extra-trees for regression and feature selection instead of MI and MRMR. The method of *Supervised inference of regulatory networks* (SIRENE) [358] uses SVM as a classifier to learn the decision boundary given the training dataset, and then generates the labels (whether links exist or do not) for the prediction dataset. Through a semi-supervised method, the unlabeled data are also included into the training data [359].

## 5.7 Transfer and joint entropies

*Transfer Entropy* (TE), as an information theoretic method, is often used for reconstructing neuron networks [360–362]. Compared with the mutual information, TE can be employed to reveal the causal relationships from the historical time series of two variables. Given the time series  $\{x_t\}$  and  $\{y_t\}$  for  $x$  and  $y$  respectively, TE can be expressed as:

$$TE_{y \rightarrow x} = H(x_t | x_{t-1} : x_{t-m}) - H(x_t | y_{t-1} : y_{t-l}, x_{t-1} : x_{t-m}), \quad (124)$$

where  $H(x)$  is the Shannon entropy of  $x$ , and  $l, m$  are the length of the historical time series. In the case of inferring neuronal networks,  $x_t$  can be the number of spikes in a specific time window, and  $l$  and  $m$  are often set to be 1. *Joint Entropy* (JE) [363]. In calculating the MI or correlations for neural networks, the number of spikes in each time window is used, but the temporal patterns

between spikes are ignored. For JE, the cross-inter-spike-intervals (cISI), defined as  $\text{cISI} = t_y - t_x$  between two neurons, is used. The JE is calculated based on cISI as

$$\text{JE}(X, Y) = H(\text{cISI}), \quad (125)$$

where  $H$  is the entropy of cISI. In Ref [363], it was shown that TE and JE perform better than other methods. In Ref. [364], the authors demonstrated that, by integrating high-order historical time series ( $l > 1$  and  $m > 1$ ) and multiple time delays, TE can be improved markedly.

## 5.8 Distinguishing between direct and indirect interactions

Based on a motif analysis, it was found [365] that three types of generic and systematic errors exist for inferring networks: (a) fan-out error, (b) fan-in error, and (c) cascade error. The cascade error originates from the tendency for incorrectly predicted “shortcuts” to cascades, also known as an error to misinterpret indirect links as direct links. For example, if nodes 1 and 2, and nodes 2 and 3 in the true (direct) network are strongly dependent upon each other, then high correlations will also be visible between nodes 1 and 3 in the observed (direct and indirect) network. Such errors are always present in the pairwise correlation, mutual information or other similarity metrics between a pair of nodes.

It was proposed [366] that a deconvolution method can be used to distinguish direct dependencies in networks. In particular, the weights of an observed network  $\mathcal{G}_{obs}$  to be inferred by the correlation or mutual information can be modeled as a sum of both direct weights  $\mathcal{G}_{dir}$  in the real network and indirect weights  $\mathcal{G}_{dir}^2, \mathcal{G}_{dir}^3$ , etc. due to indirect paths:

$$\mathcal{G}_{obs} = \mathcal{G}_{dir} + \mathcal{G}_{dir}^2 + \mathcal{G}_{dir}^3 + \cdots = \mathcal{G}_{dir} \cdot (I - \mathcal{G}_{dir})^{-1}. \quad (126)$$

The direct network can then be obtained as

$$\mathcal{G}_{dir} = \mathcal{G}_{obs} \cdot (I + \mathcal{G}_{obs})^{-1} \quad (127)$$

Through a matrix similarity transformation, one obtains  $\mathcal{G}_{obs} = \mathcal{U} \cdot \mathcal{S}_{obs} \cdot \mathcal{U}^{-1}$ , where  $\mathcal{S}_{obs}$  is a diagonal matrix whose elements are  $\lambda_{obs}$ , which holds similarly for  $\mathcal{G}_{dir}$ :  $\mathcal{G}_{dir} = \mathcal{U} \cdot \mathcal{S}_{dir} \cdot \mathcal{U}^{-1}$ , and  $\lambda_{dir} = \lambda_{obs} / (1 + \lambda_{obs})$ . The direct network can then be inferred.

The problem of extracting local response from global response to perturbation was first addressed using a modular response analysis (MRA) [367]. Since then many works have appeared [368–370]. A local response coefficient  $r_{ij}$  is defined as

$$r_{ij} = \lim_{\Delta x_j} \left( \frac{\partial \ln x_i}{\partial \ln x_j} \right), i \neq j, \quad (128)$$

which quantifies the sensitivity of module  $i$  to the change in module  $j$  with the states of other modules unchanged. However, in a real situation, an external intervention to perturb a parameter  $p_j$  intrinsic to module  $j$  can cause a local change in  $x_j$  and then propagates to the whole system. After the network has relaxed into a new steady state, the global response coefficient, an often measured quantity in practice, can be obtained as  $R_{ip_j} = d \ln x_i / d p_j$ . The relation between the local and global responses can be established.



The idea of extracting local from global responses using statistical similarity measure based methods and a global silence method were articulated and developed [371]. Given the observed global response matrix  $\mathcal{G}$ , measured by the correlation or the mutual information, the local response matrix  $\mathcal{S}$  can be obtained as

$$\mathcal{S} = [\mathcal{G} - I + \mathcal{D} \cdot (\mathcal{G} - I) \cdot \mathcal{G}] \cdot \mathcal{G}^{-1}, \quad (129)$$

where  $I$  is the identity matrix and the function  $D(\mathcal{M})$  sets the off-diagonal elements of matrix  $\mathcal{M}$  to be zero. Through this method, the indirect links become silenced so that the direct and indirect interactions can be distinguished.

## 6 Discussions and future perspectives

Nearly two decades of intense research in complex networks have resulted in a large body of knowledge about network structures and their effects on various dynamical processes on networks. The types of processes that have been investigated include synchronization [372–389], virus spreading [263–270], traffic flow [390–397], and cascading failures [172, 398–408]. A typical approach in the field is to implement a particular dynamical process of interest on networks whose connecting topologies are completely specified. Often, real-world networks such as the Internet, the power grids, transportation networks, and various biological and social networks are used as examples to demonstrate the relevance of the dynamical phenomena found from model networks. While this line of research is necessary for discovering and understanding various fundamental phenomena in complex networks, the “inverse” or “reverse engineering” problem of predicting network structure and dynamics from data is also extremely important. The basic hypothesis underlying the inverse problem is that the detailed structure of the network and the nodal dynamics are totally unknown, but only a limited set of signals or time series measured from the network is available. The question is whether the intrinsic structure of the network and the dynamical processes can be inferred solely from the set of measured time series. Compared with the “direct” network dynamics problem, there has been less effort in the inverse problem. Data-based reverse engineering of complex networks, with great application potential, is important not only for advancing network science and engineering at a fundamental level but also for meeting the need to address an array of applications where large-scale, complex networks arise. In this Review, we attempt to provide a review, as comprehensive as possible, of the recent advances in the network inverse problem.

A focus of this Review is on exploiting compressive sensing for addressing various inverse problems in complex networks (Secs. 2 and 3). The basic principle of compressive sensing is to reconstruct a signal from sparse observations, which are obtained through linear projections of the original signal on a few predetermined vectors. Since the requirements for observations can be relaxed considerably as compared with those associated with conventional signal reconstruction schemes, compressive sensing has received much recent attention and it is becoming a powerful technique to obtain high-fidelity signal for applications where sufficient observations are not available.

The key idea rendering possible exploiting compressive sensing for reverse engineering of complex network structures and dynamics lies in the fact that any nonlinear system equation or coupling function can be approximated by a power-series expansion. The task of prediction becomes then that of estimating all the coefficients in the power-series representations of the vector fields governing the nodal dynamics and the interactions. Since the underlying vector fields are unknown, the power series can contain high-order terms. The number of coefficients to be estimated can therefore be quite large. Conventional wisdom would count this as a difficult problem since a large amount of data is required and the computations involved can be extremely demanding. However, compressive sensing is ideally suited for this task. In this Review we have described in detail how compressive sensing can be used for uncovering the full topology and nodal dynamical processes of complex, nonlinear oscillator networks, for revealing the interaction patterns in social networks hosting evolutionary game dynamics, for detecting hidden nodes, for identifying chaotic elements in neuronal networks, for reconstructing complex geospatial networks and nodal positioning, and for mapping out spreading dynamics on complex networks based on binary data.

In addition to the compressive sensing based paradigm, we have also reviewed a number of alternative methods for reconstruction of complex networks. These include methods based on response to external driving signals, system clone (synchronization), phase-space linearization, noise induced dynamical correlation, and reverse engineering of complex systems. Representative methodologies to reconstruct biological networks have also been briefly reviewed: methods based on correlation, causality, information-theoretic measures, Bayesian inference, regression and resampling, supervision and semi-supervision, transfer and joint entropies, and distinguishing between direct and indirect interactions.

Data-based reconstruction of complex networks belongs to the broad field of nonlinear and complex systems identification, prediction, and control. There are many outstanding issues and problems. In the following we list several open problems which, in our opinion, are at the forefront of this area of research.

## 6.1 Localization of diffusion sources in complex networks

Dynamical processes taking place in complex networks are ubiquitous in nature and in engineering systems, examples of which include disease or epidemic spreading in the human society [409, 410], virus invasion in computer networks [268, 411], and rumor propagation in online social networks [412]. Once an epidemic emerges, it is often of great interest to be able to identify its source within the network accurately and quickly so that proper control strategies can be devised to contain or even to eliminate the spreading process. In general, various types of spreading dynamics can be regarded as diffusion processes in complex networks, and it is of fundamental interest to be able to locate the *sources of diffusion*. A straightforward, brute-force search for the sources requires accessibility of global information about the dynamical states of the network. However, for large networks, a practical challenge is that our ability to obtain and process global information can often be quite limited, making brute-force search impractical with undesired or even disastrous consequences. For example, the standard breadth-first search algorithm for finding the shortest paths, when implemented in online social networks, can induce information explosion even for a small number of searching steps [413]. Recently, in order to locate the source of the outbreak of Ebola virus in Africa, five medical practitioners lost their lives [414]. All these call for the development of efficient methodologies to locate diffusion sources based only on limited, practically available information *without the need of acquiring global information about the dynamical states of the entire network*.

There have been pioneering efforts in addressing the source localization problem in complex networks, such as those based on the maximum likelihood estimation [415, 416], belief propagation [417], the phenomena of hidden geometry of contagion [418], particle diffusion and coloration [419], and inverse spreading [420]. In addition, some approaches have been developed for identifying super spreaders that promote spreading processes stemming from sources [421–423]. In spite of these efforts, achieving accurate source localization from a small amount of observation remains challenging. A systematic framework dealing with the localization of diffusion sources for arbitrary network structures and interaction strength has been missing.

A potential approach [424] to addressing the problem of network source localization is to investigate the fundamental issue of *locatability*: given a complex network and limited (sparse) observation, are diffusion sources locatable? Answering this question can then lead to a potential

solution to the practical and challenging problem of actual localization: given a network, can a minimum set of nodes be identified which produce sufficient observation so that sources at arbitrary locations in the network can actually be located? A two-step solution strategy was recently suggested [424]. Firstly, one develops a minimum output analysis to identify the minimum number of messenger/sensor nodes, denoted as  $N_m$ , to fully locate any number of sources in an efficient way. The ratio of  $N_m$  to the network size  $N$ ,  $n_m \equiv N_m/N$ , thus characterizes the source locatability of the network in the sense that networks requiring smaller values of  $n_m$  are deemed to have a stronger locatability of sources. The minimum output analysis can be carried out by taking advantage of the dual relation between the controllability theory [425] and the canonical observability theory [426]. Secondly, given  $N_m$  messenger nodes, one can formulate the source localization problem as a sparse signal reconstruction problem, which can be solved by using compressive sensing. The basic properties of compressive sensing allow one to accurately locate sources from a small amount of measurement from the messenger nodes, much less than that required by the conventional observability theory. Testing these ideas using variety of model and real-world networks, the authors found that the connection density and degree distribution play a significant role in source locatability, and sources in a homogeneous and denser network are more readily to be located. A striking and counterintuitive finding [424] was that, for an undirected network with one connected component and random link weights, a single messenger node is sufficient to locate any number of sources.

Theoretically, the combination of the minimum output analysis (derived from the controllability and observability theories for complex networks) and the compressive sensing based localization method constitutes a general framework for locating diffusion sources in complex networks. It represents a powerful paradigm to quantify the source locatability of a network exactly and to actually locate the sources efficiently and accurately. Because of the compressive sensing based methodology, the framework is robust against noise, paving the way to practical implementation in a noisy environment. The framework also provides significant insights into the open problem of developing source localization methods for time variant complex networks hosting nonlinear diffusion processes.

## 6.2 Data based reconstruction of complex networks with binary-state dynamics

Complex networked systems consisting of units with binary-state dynamics are common in nature, technology, and society [427]. 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 [428], cooperation or defection in networks hosting evolutionary game dynamics [206], being susceptible or infected in epidemic spreading on social and technological networks [429], two competing opinions in social communities [430], 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 [431], cooperation among selfish individuals [432], oscillations in biological systems [433], power blackout [434], financial crisis [435], and phase transitions in natural systems [436]. A variety of models were introduced to gain insights into binary-state dynamics on complex networks [42], such as the voter models for competition of

two opinions [437], stochastic propagation models for epidemic spreading [438], models of rumor diffusion and adoption of new technologies [439], cascading failure models [440], Ising spin models for ferromagnetic phase transition [441], and evolutionary games for cooperation and altruism [442]. A general theoretical approach to dealing with networks hosting binary state dynamics was developed [443] based on the pair approximation and the master equations, providing a good understanding of the effect of the network structure on the emergent phenomena.

Compressive sensing can be exploited to 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. Of particular relevance to this problem 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 [37] demonstrated that the propagation network structure can be reconstructed and the sources of spreading can be detected by exploiting compressive sensing (Sec. 3.6). However, for binary state network dynamics, a general reconstruction framework has been missing. 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, difficulties may 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 methods that are robust against internal and external random effects.

A potential approach [444] to developing a general and robust framework to reconstruct complex networks based solely on the binary states of the nodes without any knowledge about the switching functions is based on the idea of linearizing the switching functions from binary data. This allows one to convert the network reconstruction into a sparse signal reconstruction problem for local structures associated with each node. Exploiting the natural sparsity of complex networks, one can employ the lasso [445], 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. The underlying mechanism that justifies the linearization procedure by conducting tests using a number of linear, nonlinear and piecewise binary-state dynamics can be established on a large number of model and real complex networks [444]. Universally high reconstruction accuracy was found even for small data amount with noise. Because of its high accuracy, efficiency and robustness against noise and missing data, the framework is promising as a general solution to the inverse problem of network reconstruction from binary-state time series. The data-based linearization method can be useful for dealing with general nonlinear systems with a wide range of applications.

### 6.3 Universal structural estimator and dynamics approximator for complex networks

Is it possible to develop a universal and completely data-driven framework for extracting network topology and identifying the dynamical processes, without the need to know *a priori* the specific types of network dynamics? An answer to this question would be of significant value not only to complexity science and engineering but also to modern data science where the goal is to unearth the hidden structural information and to predict the future evolution of the system. A partial solution to this problem emerged recently, where the concept of *universal structural estimator and dynamics approximator* for complex networked systems was proposed [446], and it was demonstrated that such a framework or “machine” can indeed be developed for a large number of distinct types of network dynamical processes. While it remains an open issue to develop a rigorous mathematical framework for the universal machine, the preliminary work [446] can be regarded as an initial attempt towards the development of a universal framework for network reconstruction and dynamics prediction.

The key principle is the following. In spite of the dramatic difference in the types of dynamics in terms of, e.g., the interaction setting and state updating rules, there are two common features shared by many dynamical processes on complex networks: (1) they are stochastic, first-order Markovian processes, i.e., only the current states of the systems determine their immediate future; and (2) the nodal interactions are local. The two features are characteristic of a *Markov network* (or a *Markov random field*) [447, 448]. In particular, a Markov network is an undirected and weighted probabilistic graphical model that is effective at determining the complex probabilistic interdependencies in situations where the directionality in the interaction between connected nodes cannot be naturally assigned, in contrast to the directed Bayesian networks [447, 448]. A Markov network has two types of parameters: a node bias parameter that controls its preference of state choice, and a weight parameter characterizing the interaction strength of each undirected link. The joint probability distribution of the state variables  $\mathbf{X} = (x_1, x_2, \dots, x_N)^T$  is given by  $P(\mathbf{X}) = \prod_C \phi(\mathbf{X}_C) / \sum_{\mathbf{X}} \prod_C \phi(\mathbf{X}_C)$ , where  $\phi(\mathbf{X}_C)$  is the potential function for a well-connected network clique  $C$ , and the summation in the denominator is over all possible system state  $\mathbf{X}$ . If this joint probability distribution is available, literally all conditional probability interdependencies can be obtained. The way to define a clique and to determine its potential function plays a key role in the Markov network’s representation power of modeling the interdependencies within the system. In Ref. [446], the possibility was pursued of modeling the conditional probability interdependence of a variety of dynamical processes on complex networked systems via a binary Ising Markov network with its potential function in the form of the Boltzmann factor,  $\exp(-E)$ , where  $E$  is the energy determined by the local states and their interactions along with the network parameters (the link weights and node biases) in a *log-linear* form [449]. This is effectively a sparse Boltzmann machine [449] adopted to complex network topologies without hidden units. (Note that hidden units can play a critical role in ordinary Boltzmann machines [449]). A temporal evolutionary mechanism was introduced [446] as a persistent sampling process for such a machine based on the conditional probabilities obtained via the joint probability, and generate a Markov chain of persistently sampled state configurations to form state transition time series for each node. The model was named a *sparse dynamical Boltzmann machine* (SDBM) [446].

In dynamical processes on complex networks, such as epidemic spreading or evolutionary

game dynamics, the state of each node at the next time step is determined by the probability conditioned on the current states of its neighbors (and its own state in some cases). There is freedom to manipulate the conditional probabilities that dictate the system behavior in the immediate future through change in the parameters such as weights and biases. A basic question is then, for an SDBM, is it possible to properly choose these parameters so that the conditional probabilities produced are identical or nearly identical to those of a typical dynamical process with each given observed system state configuration? If the answer is affirmative, the SDBM can serve as a dynamics approximator of the original system, and the approximated conditional probabilities possess predictive power for the system state at the next time step. When such an SDBM is found for many types of dynamical process on complex networks, it effectively serves as a universal dynamics approximator. Moreover, if the detailed statistical properties of the state configurations can be reproduced in the long time limit, that is, if the time series generated by this SDBM are statistically identical or nearly identical to those from the original system, the SDBM is a generative model of the observed data (in the language of machine learning), which is potentially capable of long term prediction.

When an approximator exists for each type of dynamics on a complex network, the topology of the SDBM is nothing but that of the original network, providing a simultaneous solution to the problem of network structure reconstruction. Previous works on the inverse static or kinetic Ising problems led to methods of reconstruction for Ising dynamics through maximization of the data likelihood (or pseudo-likelihood) function via the gradient descent approaches [450–461]. Instead of adopting these approaches, compressive sensing can be exploited. By incorporating a K-means clustering algorithm into the sparse solution obtained from compressive sensing, it was demonstrated [446] that nearly perfect reconstruction of the complex network topology can be achieved. In particular, using 14 different types of dynamical processes on complex networks, it was found that, if the time series data generated by these dynamical processes are assumed to be from its equivalent SDBMs, the universal reconstruction framework is capable of recovering the underlining network structure of each original dynamics with almost zero error. This represents solid and concrete evidence that SDBM is capable of serving as a universal structural estimator for complex networks. In addition to being able to precisely reconstruct the network topologies, the SDBM also allows the link weights and the nodal biases to be calculated with high accuracy. In fact, the the universal SDBM is fully automated and does not require any subjective parameter choice [446].

## 6.4 Controlling nonlinear and complex dynamical networks

An ultimate goal of systems identification and prediction is to control. The coupling between nonlinear dynamics and complex network structure presents tremendous challenges to our ability to formulate effective control methodologies. In spite of the rapid development of network science and engineering toward understanding, analyzing and predicting the dynamics of large complex network systems in the past fifteen years, the problem of controlling nonlinear dynamical networks has remained largely open.

The field of controlling chaos in nonlinear dynamical systems has been active for more than two decades since the seminal work of Ott, Grebogi, and Yorke [462]. The basic idea was that chaos, while signifying random or irregular behavior, possesses an intrinsically sensitive dependence on

initial conditions that can be exploited for controlling the system using only small perturbation. This feature, in combination with the fact that a chaotic system possesses an infinite set of unstable periodic orbits, each leading to different system performance, implies that a chaotic system can be stabilized about some desired state with optimal performance. Controlling chaos has since been studied extensively and examples of successful experimental implementation abound in physical, chemical, biological, and engineering systems [463]. The vast literature on controlling chaos, however, has been mostly limited to low-dimensional systems, systems that possess one or a very few unstable directions (i.e., one or a very few positive Lyapunov exponents [464]). Complex networks with nonlinear dynamics are generally high dimensional, rendering inapplicable existing methodologies of chaos control.

In the past several years, a framework for determining the *linear* controllability of network based on the traditional control and graph theories emerged [425, 465–479], leading to a quantitative understanding of the effect of network structure on its controllability. In particular, a structural controllability framework was proposed [468], revealing that the ability to steer a complex network toward any desired state, as measured by the minimum number of driver nodes, is determined by the set of maximum matching, which is the maximum set of links that do not share starting or ending nodes. A main result was that the number of driver nodes required for full control is determined by the network's degree distribution [468]. The framework was established for weighted and directed networks. An alternative framework, the exact controllability framework, was subsequently formulated [425], which was based on the principle of maximum multiplicity to identify the minimum set of driver nodes required to achieve full control of networks with arbitrary structures and link weight distributions. Generally, a limitation of such rigorous mathematical frameworks of controllability is that the nodal dynamical processes must be assumed to be *linear*. For nonlinear dynamical networks, to establish a mathematical controllability framework similar to those based on the classic Kalman's rank condition [480–482] for linear networks is an unrealistically broad objective. Traditionally, controllability for nonlinear control can be formulated based on Lie brackets [483], but to extend the abstract framework to complex networks may be difficult. A recent work extended the linear controllability and observability theory to nonlinear networks with symmetry [477]. In spite of the previous works, at the present there is no known general framework for controlling nonlinear dynamics on complex networks.

Due to the high dimensionality of nonlinear dynamical networks and the rich variety of behaviors that they can exhibit, it may be prohibitively difficult to develop a control framework that is universally applicable to different kinds of network dynamics. In particular, the classic definition of linear controllability, i.e., a network system is controllable if it can be driven from an arbitrary initial state to an arbitrary final state in finite time, is generally not applicable to nonlinear dynamical networks. Instead, controlling collective dynamical behaviors may be more pertinent and realistic [246–248, 484]. Our viewpoint is that, for nonlinear dynamical networks, control strategies may need to be specific and system-dependent. Recently, a control framework for systems exhibiting multistability was formulated [485]. A defining characteristic of such systems is that, for a realistic parameter setting, there are multiple coexisting attractors in the phase space [156, 486–490]. The goal is to drive the system from one attractor to another using physically meaningful, temporary and finite parameter perturbation, assuming that the system is likely to evolve into an undesired state (attractor) or the system is already in such a state and one wishes to implement control to bring the system out of the undesired state and steer it into a desired one. It should be noted that



dynamical systems with multistability are ubiquitous in the real world ranging from biological and ecological to physical systems [186, 491–496].

In biology, nonlinear dynamical networks with multiple attractors have been employed to understand fundamental phenomena such as cancer mechanisms [497], cell fate differentiation [498–501], and cell cycle control [502–504]. For example, Boolean network models were used to study gene evolution [505], attractor number variation with asynchronous stochastic updating [506], gene expression in the state space [498], and organism system growth rate improvement [507]. Another approach is to abstract key regulation genetic networks [508, 509] (or motifs) from all associated interactions, and to employ synthetic biology to modify, control and finally understand the biological mechanisms within these complicated systems [499, 503]. An earlier application of this approach led to a good understanding of the ubiquitous phenomenon of bistability in biological systems [510], where there are typically limit cycle attractors and, during cell cycle control, noise can trigger a differentiation process by driving the system from a limit circle to another steady state attractor [499]. In general, there are two candidate mechanisms for transition or switching between different attractors [500]: through signals transmitted within cells and through noise, which were demonstrated recently using synthetic genetic circuits [511, 512]. More recently, a detailed numerical study was carried out of how signal-induced bifurcations in a tri-stable genetic circuit can lead to transitions among different cell types [501].

The control and controllability framework for nonlinear dynamical networks with multistability can be formulated [485] based on the concept of *attractor networks* [513]. An attractor network is defined in the phase space of the underlying nonlinear system, in which each node represents an attractor and a directed edge from one node to another indicates that the system can be driven from the former to the latter using experimentally feasible, temporary, and finite parameter changes. A well connected attractor network implies a strong feasibility that the system can be controlled to reach a desired attractor. The connectivity of the attractor network can then be used to characterize the controllability of the nonlinear dynamical network. More specifically, for a given pair of attractors, the relative weight of the shortest path is the number of accessible control parameters whose adjustments can lead to the attractor transition as specified by the path. Gene regulatory networks (GRNs) were used [485] to demonstrate the *practicality* of the control framework, which includes low-dimensional, experimentally realizable synthetic gene circuits and a realistic T-cell cancer network of 60 nodes. A finding was that noise can facilitate control by reducing the required amplitude of the control signal. In fact, the development of the nonlinear control framework [485] was based entirely on physical considerations, rendering feasible experimental validation.

The framework can be adopted to controlling nonlinear dynamical networks other than the GRNs. For example, for the Northern European power grid network recently studied by Menck et al. [514], a rewiring method was proposed and demonstrated to be able to enhance the system stability through the addition of extra transmission lines. For a power grid network, the synchronous states are desired while other states, e.g., limit cycles, detrimental. Treating the link density (or number) as a tunable parameter, the minimum transfer capacity required for extra lines to realize the control can be estimated through our method. Another example is Boolean networks with discrete dynamics, for which a perturbation method was proposed based on modifying the update rules to rescue the system from the undesired states [515]. An attractor network can be constructed based on perturbation to multiple parameters to drive the system out of the undesired, damaged states toward a normal (desired) state. For biological systems, an epigenetic state network (ESN)

approach was proposed [516] to analyze the transitions among different phenotypic processes. In an ESN, nodes represent attractors and edges represent pathways between a pair of attractors. By construction, different parameter values would result in a different ESN. This should be contrasted to an attractor network, in which nodes are attractors but edges are directed and represent controllable paths (through parameter perturbation) to drive the system from one attractor to another.

## Acknowledgement

We thank Dr. R. Yang (formerly at ASU), Dr. R.-Q. Su (formerly at ASU), and Mr. Zhesi Shen for their contributions to a number of original papers on which this Review is partly based. This work was supported by ARO under Grant No. W911NF-14-1-0504. W.-X. Wang was also supported by NSFC under Grants No. 61573064 and No. 61074116, as well as by the Fundamental Research Funds for the Central Universities, Beijing Nova Programme.

## References

- [1] Gruen, S., Diesmann, M. & Aertsen, A. Unitary events in multiple single neuron spiking activity. i. detection and significance. *Neu. Comp.* **14**, 43–80 (2002).
- [2] Güttig, R., Aertsen, A. & Rotter, S. Statistical significance of coincident spikes: count-based versus rate-based statistics. *Neu. Comp.* **14**, 121–153 (2002).
- [3] Gardner, T. S., di Bernardo, D., Lorenz, D. & Collins, J. J. Inferring genetic networks and identifying compound mode of action via expression profiling. *Science* **301**, 102–105 (2003).
- [4] Pipa, G. & Grün, S. Non-parametric significance estimation of joint-spike events by shuffling and resampling. *Neurocomputing* **52**, 31–37 (2003).
- [5] Brovelli, A. *et al.* Beta oscillations in a large-scale sensorimotor cortical network: directional influences revealed by Granger causality. *Proc. Nat. Acad. Sci. USA* **101**, 9849–9854 (2004).
- [6] Yu, D., Righero, M. & Kocarev, L. Estimating topology of networks. *Phys. Rev. Lett.* **97**, 188701 (2006).
- [7] Bongard, J. & Lipson, H. Automated reverse engineering of nonlinear dynamical systems. *Proc. Natl. Acad. Sci. (USA)* **104**, 9943–9948 (2007).
- [8] Timme, M. Revealing network connectivity from response dynamics. *Phys. Rev. Lett.* **98**, 224101 (2007).
- [9] Tang, W. K.-S., Yu, M. & Kocarev, L. Identification and monitoring of biological neural network. In *Circuits and Systems, 2007. ISCAS 2007. IEEE Inter. Symp.*, 2646–2649 (IEEE, 2007).
- [10] Napoletani, D. & Sauer, T. D. Reconstructing the topology of sparsely connected dynamical networks. *Phys. Rev. E* **77**, 026103 (2008).
- [11] Sontag, E. Network reconstruction based on steady-state data. *Essays Biochem.* **45**, 161–176 (2008).
- [12] Clauset, A., Moore, C. & Newman, M. E. J. Hierarchical structure and the prediction of missing links in networks. *Nature (London)* **453**, 98–101 (2008).
- [13] Wang, W.-X., Chen, Q.-F., Huang, L., Lai, Y.-C. & Harrison, M. Scaling of noisy fluctuations in complex networks and applications to network prediction. *Phys. Rev. E* **80**, 016116 (2009).
- [14] Donges, J., Zou, Y., Marwan, N. & Kurths, J. The backbone of the climate network. *EPL (Europhys. Lett.)* **87**, 48007 (2009).
- [15] Ren, J., Wang, W.-X., Li, B. & Lai, Y.-C. Noise bridges dynamical correlation and topology in coupled oscillator networks. *Phys. Rev. Lett.* **104**, 058701 (2010).

- [16] Chan, J., Holmes, A. & Rabadan, R. Network analysis of global influenza spread. *PLoS Comput Biol* **6**, e1001005 (2010).
- [17] Yuan, Y., Stan, G.-B., Warnick, S. & Goncalves, J. Robust dynamical network reconstruction. In *Decision and Control (CDC), 2010 49th IEEE Conference on*, 810–815 (IEEE, 2010).
- [18] Levnajić, Z. & Pikovsky, A. Network reconstruction from random phase resetting. *Phys. Rev. Lett.* **107**, 034101 (2011).
- [19] Hempel, S., Koseska, A., Kurths, J. & Nikoloski, Z. Inner composition alignment for inferring directed networks from short time series. *Phys. Rev. Lett.* **107**, 054101 (2011).
- [20] Shandilya, S. G. & Timme, M. Inferring network topology from complex dynamics. *New J. Phys.* **13**, 013004 (2011).
- [21] Yu, D. & Parlitz, U. Inferring network connectivity by delayed feedback control. *PLOS ONE* **6**, e24333 (2011).
- [22] Wang, W.-X., Lai, Y.-C., Grebogi, C. & Ye, J.-P. Network reconstruction based on evolutionary-game data via compressive sensing. *Phys. Rev. X* **1**, 021021 (2011).
- [23] Wang, W.-X., Yang, R., Lai, Y.-C., Kovanis, V. & Grebogi, C. Predicting catastrophes in nonlinear dynamical systems by compressive sensing. *Phys. Rev. Lett.* **106**, 154101 (2011).
- [24] Wang, W.-X., Yang, R., Lai, Y.-C., Kovanis, V. & Harrison, M. A. F. Time-series-based prediction of complex oscillator networks via compressive sensing. *EPL (Europhys. Lett.)* **94**, 48006 (2011).
- [25] Yang, R., Lai, Y.-C. & Grebogi, C. Forecasting the future: is it possible for time-varying nonlinear dynamical systems? *Chaos* **22**, 033119 (2012).
- [26] Pan, W., Yuan, Y. & Stan, G.-B. Reconstruction of arbitrary biochemical reaction networks: a compressive sensing approach. In *Decision and Control (CDC), 2012 IEEE 51st Annu. Conf.*, 2334–2339 (IEEE, 2012).
- [27] Wang, W.-X., Ren, J., Lai, Y.-C. & Li, B. Reverse engineering of complex dynamical networks in the presence of time-delayed interactions based on noisy time series. *Chaos* **22**, 033131 (2012).
- [28] Berry, T., Hamilton, F., Peixoto, N. & Sauer, T. Detecting connectivity changes in neuronal networks. *J. Neurosci. Meth.* **209**, 388–397 (2012).
- [29] Stetter, O., Battaglia, D., Soriano, J. & Geisel, T. Model-free reconstruction of excitatory neuronal connectivity from calcium imaging signals. *PLoS Comp. Biol.* **8**, e1002653 (2012).
- [30] Su, R.-Q., Ni, X., Wang, W.-X. & Lai, Y.-C. Forecasting synchronizability of complex networks from data. *Phys. Rev. E* **85**, 056220 (2012).
- [31] Su, R.-Q., Wang, W.-X. & Lai, Y.-C. Detecting hidden nodes in complex networks from time series. *Phys. Rev. E* **85**, 065201 (2012).
- [32] Hamilton, F., Berry, T., Peixoto, N. & Sauer, T. Real-time tracking of neuronal network structure using data assimilation. *Phys. Rev. E* **88**, 052715 (2013).
- [33] Zhou, D., Xiao, Y., Zhang, Y., Xu, Z. & Cai, D. Causal and structural connectivity of pulse-coupled nonlinear networks. *Phys. Rev. Lett.* **111**, 054102 (2013).
- [34] Timme, M. & Casadiego, J. Revealing networks from dynamics: an introduction. *J. Phys. A. Math. Theo.* **47**, 343001 (2014).
- [35] Su, R.-Q., Lai, Y.-C. & Wang, X. Identifying chaotic fitzhugh-nagumo neurons using compressive sensing. *Entropy* **16**, 3889–3902 (2014).

- [36] Su, R.-Q., Lai, Y.-C., Wang, X. & Do, Y.-H. Uncovering hidden nodes in complex networks in the presence of noise. *Sci. Rep.* **4**, 3944 (2014).
- [37] Shen, Z., Wang, W.-X., Fan, Y., Di, Z. & Lai, Y.-C. Reconstructing propagation networks with natural diversity and identifying hidden sources. *Nat. Commun.* **5**, 4323 (2014).
- [38] Su, R.-Q., Wang, W.-W., Wang, X. & Lai, Y.-C. Data based reconstruction of complex geospatial networks, nodal positioning, and detection of hidden node. *R. Soc. Open Sci.* **1** (2015).
- [39] Albert, R. & Barabási, A.-L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97 (2002).
- [40] Newman, M. E. J. The structure and function of complex networks. *SIAM Rev.* **45**, 167–256 (2003).
- [41] Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. & Hwang, D.-U. Complex networks: Structure and dynamics. *Phys. Rep.* **424**, 175–308 (2006).
- [42] Newman, M. E. J. *Networks: An Introduction* (Oxford University Press, 2010).
- [43] Takens, F. Detecting strange attractors in fluid turbulence. In Rand, D. & Young, L. S. (eds.) *Dynamical Systems and Turbulence*, vol. 898 of *Lecture Notes in Mathematics*, 366–381 (Springer-Verlag, Berlin, 1981).
- [44] Kantz, H. & Schreiber, T. *Nonlinear Time Series Analysis* (Cambridge University Press, Cambridge, UK, 1997), first edn.
- [45] Hegger, R., Kantz, H. & Schreiber, R. *TISEAN* (<http://www.mpi-pks-dresden.mpg.de/tisean/TISEAN3.01/index.html>, Dresden, 2007), e-book edn.
- [46] Packard, N. H., Crutchfield, J. P., Farmer, J. D. & Shaw, R. S. Geometry from a time series. *Phys. Rev. Lett.* **45**, 712–716 (1980).
- [47] Theiler, J. Spurious dimension from correlation algorithms applied to limited time series data. *Phys. Rev. A* **34**, 2427–2432 (1986).
- [48] Liebert, W. & Schuster, H. G. Proper choice of the time-delay for the analysis of chaotic time-series. *Phys. Lett. A* **142**, 107–111 (1989).
- [49] Liebert, W., Pawelzik, K. & Schuster, H. G. Optimal embeddings of chaotic attractors from topological considerations. *Europhys. Lett.* **14**, 521–526 (1991).
- [50] Buzug, T. & Pfister, G. Optimal delay time and embedding dimension for delay-time coordinates by analysis of the global static and local dynamic behavior of strange attractors. *Phys. Rev. A* **45**, 7073–7084 (1992).
- [51] Kember, G. & Fowler, A. C. A correlation-function for choosing time delays in-phase portrait reconstructions. *Phys. Lett. A* **179**, 72–80 (1993).
- [52] Rosenstein, M. T., Collins, J. J. & Luca, C. J. D. Reconstruction expansion as a geometry-based framework for choosing proper delay times. *Physica D* **73**, 82–98 (1994).
- [53] Lai, Y.-C., Lerner, D. & Hayden, R. An upper bound for the proper delay time in chaotic time series analysis. *Phys. Lett. A* **218**, 30–34 (1996).
- [54] Lai, Y.-C. & Lerner, D. Effective scaling regime for computing the correlation dimension in chaotic time series analysis. *Physica D* **115**, 1–18 (1998).
- [55] Pecora, L. M., Carroll, T. L. & Heagy, J. F. Statistics for mathematical properties of maps between time series embeddings. *Phys. Rev. E* **52**, 3420–3439 (1995).
- [56] Pecora, L. M. & Carroll, T. L. Discontinuous and nondifferentiable functions and dimension increase induced by filtering chaotic data. *Chaos* **6**, 432–439 (1996).

- [57] Pecora, L. M., Carroll, T. L. & Heagy, J. F. Statistics for continuity and differentiability: an application to attractor reconstruction from time series. *Fields Inst. Commun.* **11**, 49–62 (1997).
- [58] Goodridge, C. L., Pecora, L. M., Carroll, T. L. & Rachford, F. J. Detecting functional relationships between simultaneous time series. *Phys. Rev. E* **64**, 026221 (2001).
- [59] Pecora, L. M., Moniz, L., Nichols, J. & Carroll, T. A unified approach to attractor reconstruction. *Chaos* **17**, 013110 (2007).
- [60] Sauer, T. D., Yorke, J. A. & Casdagli, M. Embedology. *J. Stat. Phys.* **65**, 579–616 (1991).
- [61] János, J., Flepp, L. & Tél, T. Exploring transient chaos in an NMR-laser experiment. *Phys. Rev. Lett.* **73**, 529–532 (1994).
- [62] János, J. & Tél, T. Time-series analysis of transient chaos. *Phys. Rev. E* **49**, 2756–2763 (1994).
- [63] Dhamala, M., Lai, Y.-C. & Kostelich, E. J. Detecting unstable periodic orbits from transient chaotic time series. *Phys. Rev. E* **61**, 6485–6489 (2000).
- [64] Dhamala, M., Lai, Y.-C. & Kostelich, E. J. Analysis of transient chaotic time series. *Phys. Rev. E* **64**, 056207 (2001).
- [65] Triandaf, I., Bollt, E. & Schwartz, I. B. Approximating stable and unstable manifolds in experiments. *Phys. Rev. E* **67**, 037201 (2003).
- [66] Taylor, S. R. & Campbell, S. A. Approximating chaotic saddles in delay differential equations. *Phys. Rev. E* **75**, 046215 (2007).
- [67] Lathrop, D. P. & Kostelich, E. J. Characterization of an experimental strange attractor by periodic orbits. *Phys. Rev. A* **40**, 4028–4031 (1989).
- [68] Badii, R. *et al.* Progress in the analysis of experimental chaos through periodic orbits. *Rev. Mod. Phys.* **66**, 1389–1415 (1994).
- [69] Pierson, D. & Moss, F. Detecting periodic unstable points in noisy chaotic and limit-cycle attractors with applications to biology. *Phys. Rev. Lett.* **75**, 2124–2127 (1995).
- [70] Pei, X. & Moss, F. Characterization of low-dimensional dynamics in the crayfish caudal photoreceptor. *Nature (London)* **379**, 618–621 (1996).
- [71] So, P. *et al.* Detecting unstable periodic orbits in chaotic experimental data. *Phys. Rev. Lett.* **76**, 4705–4708 (1996).
- [72] Allie, S. & Mees, A. Finding periodic points from short time series. *Phys. Rev. E* **56**, 346–350 (1997).
- [73] Grassberger, P. & Procaccia, I. Measuring the strangeness of strange attractors. *Physica D* **9**, 189–208 (1983).
- [74] Grassberger, P. Do climatic attractors exist? *Nature (London)* **323**, 609–612 (1986).
- [75] Procaccia, I. Complex or just complicated? *Nature (London)* **333**, 498–499 (1988).
- [76] Osborne, A. & Provenzale, A. Finite correlation dimension for stochastic systems with power-law spectra. *Physica D* **35**, 357–381 (1989).
- [77] Lorenz, E. Dimension of weather and climate attractors. *Nature (London)* **353**, 241–244 (1991).
- [78] Ding, M., Grebogi, C., Ott, E., Sauer, T. D. & Yorke, J. A. Plateau onset for correlation dimension: when does it occur? *Phys. Rev. Lett.* **70**, 3872–3875 (1993).
- [79] Wolf, A., Swift, J. B., Swinney, H. L. & Vastano, J. A. Determining Lyapunov exponents from a time series. *Physica D* **16**, 285–317 (1985).

- [80] Sano, M. & Sawada, Y. Measurement of the Lyapunov spectrum from a chaotic time series. *Phys. Rev. Lett.* **55**, 1082–1085 (1985).
- [81] Eckmann, J. P. & Ruelle, D. Ergodic theory of chaos and strange attractors. *Rev. Mod. Phys.* **57**, 617–656 (1985).
- [82] Eckmann, J. P., Kamphorst, S. O., Ruelle, D. & Ciliberto, S. Liapunov exponents from time series. *Phys. Rev. A* **34**, 4971–4979 (1986).
- [83] Brown, R., Bryant, P. & Abarbanel, H. D. I. Computing the Lyapunov spectrum of a dynamical system from an observed time series. *Phys. Rev. A* **43**, 2787–2806 (1991).
- [84] Sauer, T. D., Tempkin, J. A. & Yorke, J. A. Spurious Lyapunov exponents in attractor reconstruction. *Phys. Rev. Lett.* **81**, 4341–4344 (1998).
- [85] Sauer, T. D. & Yorke, J. A. Reconstructing the Jacobian from data with observational noise. *Phys. Rev. Lett.* **83**, 1331–1334 (1999).
- [86] Casdagli, M. Nonlinear prediction of chaotic time series. *Physica D* **35**, 335–356 (1989).
- [87] Sugihara, G. *et al.* Distinguishing error from chaos in ecological time series. *Phil. Trans. Roy. Soc. London B* **330**, 235–251 (1990).
- [88] Kurths, J. & Ruzmaikin, A. A. On forecasting the sunspot numbers. *Solar Phys.* **126**, 407–410 (1990).
- [89] Grassberger, P. & Schreiber, T. Nonlinear time sequence analysis. *Int. J. Bif. Chaos* **1**, 521–547 (1990).
- [90] Tsonis, A. A. & Elsner, J. B. Nonlinear prediction as a way of distinguishing chaos from random fractal sequences. *Nature (London)* **358**, 217–220 (1992).
- [91] Longtin, A. Nonlinear forecasting of spike trains from sensory neurons. *Int. J. Bif. Chaos* **3**, 651–661 (1993).
- [92] Murray, D. B. Forecasting a chaotic time series using an improved metric for embedding space. *Physica D* **68**, 318–325 (1993).
- [93] Sugihara, G. Nonlinear forecasting for the classification of natural time series. *Philos. T. Roy. Soc. A.* **348**, 477–495 (1994).
- [94] Finkenstädt, B. & Kuhsier, P. Forecasting nonlinear economic time series: A simple test to accompany the nearest neighbor approach. *Empiri. Econ.* **20**, 243–263 (1995).
- [95] Schiff, S. J., So, P., Chang, T., Burke, R. E. & Sauer, T. Detecting dynamical interdependence and generalized synchrony through mutual prediction in a neural ensemble. *Phys. Rev. E* **54**, 6708–6724 (1996).
- [96] Hegger, R., Kantz, H. & Schreiber, T. Practical implementation of nonlinear time series methods: The tisean package. *Chaos* **9**, 413–435 (1999).
- [97] Sello, S. Solar cycle forecasting: a nonlinear dynamics approach. *Astron. Astrophys.* **377**, 312–320 (2001).
- [98] Matsumoto, T., Nakajima, Y., Saito, M., Sugi, J. & Hamagishi, H. Reconstructions and predictions of nonlinear dynamical systems: a hierarchical bayesian approach. *IEEE Trans. Signal Proc.* **49**, 2138–2155 (2001).
- [99] Smith, L. A. What might we learn from climate forecasts? *Proc. Nat. Acad. Sci. (USA)* **19**, 2487–2492 (2002).
- [100] Judd, K. Nonlinear state estimation, indistinguishable states, and the extended kalman filter. *Physica D* **183**, 273–281 (2003).
- [101] Farmer, J. D. & Sidorowich, J. J. Predicting chaotic time series. *Phys. Rev. Lett.* **59**, 845–848 (1987).

- [102] Gouesbet, G. Reconstruction of standard and inverse vector fields equivalent to a Rössler system. *Phys. Rev. A* **44**, 6264–6280 (1991).
- [103] Baake, E., Baake, M., Bock, H. G. & Briggs, K. M. Fitting ordinary differential equations to chaotic data. *Phys. Rev. A* **45**, 5524–5529 (1992).
- [104] Sauer, T. Reconstruction of dynamical systems from interspike intervals. *Phys. Rev. Lett.* **72**, 3811–3814 (1994).
- [105] Sauer, T. D. Reconstruction of shared nonlinear dynamics in a network. *Phys. Rev. Lett.* **93**, 198701 (2004).
- [106] Parlitz, U. Estimating model parameters from time series by autosynchronization. *Phys. Rev. Lett.* **76**, 1232 (1996).
- [107] Szpiro, G. G. Forecasting chaotic time series with genetic algorithms. *Phys. Rev. E* **55**, 2557–2568 (1997).
- [108] Tao, C., Zhang, Y. & Jiang, J. J. Estimating system parameters from chaotic time series with synchronization optimized by a genetic algorithm. *Phys. Rev. E* **76**, 016209 (2007).
- [109] Hegger, R., Kantz, H., Matassini, L. & Schreiber, T. Coping with nonstationarity by overembedding. *Phys. Rev. Lett.* **84**, 4092–4095 (2000).
- [110] Kuramoto, Y. *Chemical Oscillations, Waves and Turbulence* (Springer-Verlag, Berlin, 1984), first edn.
- [111] Strogatz, S. H. From Kuramoto to Crawford: exploring the onset of synchronization in populations of coupled oscillators. *Physica D* **143**, 1–20 (2000).
- [112] Chen, S. S., Donoho, D. L. & Saunders, M. A. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.* **20**, 33–61 (1998).
- [113] Candès, E., Romberg, J. & Tao, T. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Info. Theory* **52**, 489–509 (2006).
- [114] Candès, E., Romberg, J. & Tao, T. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.* **59**, 1207–1223 (2006).
- [115] Candès, E. Compressive sampling. In *Proceedings of the International Congress of Mathematicians*, vol. 3, 1433–1452 (Madrid, Spain, 2006).
- [116] Donoho, D. Compressed sensing. *IEEE Trans. Info. Theory* **52**, 1289–1306 (2006).
- [117] Baraniuk, R. G. Compressed sensing. *IEEE Signal Process. Mag.* **24**, 118–121 (2007).
- [118] Candès, E. & Wakin, M. An introduction to compressive sampling. *IEEE Signal Process. Mag.* **25**, 21–30 (2008).
- [119] Granger, C. W. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37**, 424–438 (1969).
- [120] Sommerlade, L. *et al.* Inference of Granger causal time-dependent influences in noisy multivariate time series. *J. Neurosci. Methods* **203**, 173–185 (2012).
- [121] Ramb, R. *et al.* The impact of latent confounders in directed network analysis in neuroscience. *Phil. Trans. Roy. Soc. A* **371**, 20110612 (2013).
- [122] Gao, L. *et al.* Granger causal time-dependent source connectivity in the somatosensory network. *Sci. Rep.* **5**, 10399 (2015).
- [123] Schelter, B. *et al.* Overarching framework for data-based modelling. *EPL* **105**, 30004 (2014).



- [124] Makarov, V. A., Panetsos, F. & Feo, O. A method for determining neural connectivity and inferring the underlying network dynamics using extracellular spike recordings. *J. Neurosci. Meth.* **144**, 265–279 (2005).
- [125] Yao, C. & Bollt, E. M. Modeling and nonlinear parameter estimation with Kronecker product representation for coupled oscillators and spatiotemporal systems. *Physica D* **227**, 78–99 (2007).
- [126] Davidson, E. H. *et al.* A genomic regulatory network for development. *Science* **295**, 1669–1678 (2002).
- [127] Oates, A. C., Gorfinkiel, N., González-Gaitán, M. & Heisenberg, C.-P. Quantitative approaches in developmental biology. *Nat. Rev. Gene.* **10**, 517–530 (2009).
- [128] Bonneau, R. *et al.* A predictive model for transcriptional control of physiology in a free living cell. *Cell* **131**, 1354–1365 (2007).
- [129] Gardner, T. S., di Bernardo, D., Lorenz, D. & Collins, J. J. Inferring genetic networks and identifying compound mode of action via expression profiling. *Science* **301**, 102–105 (2003).
- [130] Faith, J. J. *et al.* Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS Biol.* **5**, e8 (2007).
- [131] Rice, J. J., Tu, Y. & Stolovitzky, G. Reconstructing biological networks using conditional correlation analysis. *Bioinfo.* **21**, 765–773 (2005).
- [132] Perkins, T. J., Jaeger, J., Reinitz, J. & Glass, L. Reverse engineering the gap gene network of *Drosophila melanogaster*. *PLoS Comp. Biol.* **2**, e51 (2006).
- [133] Bonneau, R. *et al.* The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets *de novo*. *Geno. Biol.* **7**, R36 (2006).
- [134] Margolin, A. *et al.* ARACNE: An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinfo.* **7**, S7 (2006).
- [135] Bonneau, R. Learning biological networks: from modules to dynamics. *Nat. Chem. Biol.* **4**, 658–664 (2008).
- [136] Turing, A. M. The chemical basis of morphogenesis. *Phil. Trans. Royal Soc. London Ser. B Biol. Sci.* **237**, 37–72 (1951).
- [137] BC, G. & SA, K. Spatial harmonics and pattern specification in early *Drosophila* development. part i. bifurcation sequences and gene expression. *J. Theo. Biol.* **144**, 303–319 (1990).
- [138] BC, G. & SA, K. Spatial harmonics and pattern specification in early *Drosophila* development. part ii. the four colour wheels model. *J. Theo. Biol.* **144**, 321–345 (1990).
- [139] Kauffman, S. A. Homeostasis and differentiation in random genetic control networks. *Nature (London)* **224**, 177–178 (1969).
- [140] Kauffman, S. A. Metabolic stability and epigenesis in randomly constructed genetic nets. *J. Theo. Biol.* **22**, 437–467 (1969).
- [141] Shmulevich, I. & Kauffman, S. A. Activities and sensitivities in boolean network models. *Phys. Rev. Lett.* **93**, 048701 (2004).
- [142] Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Nat. Acad. Sci. (USA)* **95**, 14863–14868 (1998).
- [143] Butte, A. J., Tamayo, P., Slonim, D., Golub, T. R. & Kohane, I. S. Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks. *Proc. Nat. Acad. Sci. (USA)* **97**, 12182–12186 (2000).

- [144] Friedman, N., Linial, M., Nachman, I. & Pe'er, D. Using bayesian networks to analyze expression data. *J. Comp. Biol.* **7**, 601–620 (2000).
- [145] Friedman, N. Inferring cellular networks using probabilistic graphical models. *Science* **303**, 799–805 (2004).
- [146] Segal, E. *et al.* Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat. Gene.* **34**, 166–176 (2003).
- [147] Chickering, D. M. Learning Bayesian networks is NP-complete. In Fisher, D. & Lenz, H. (eds.) *Learning from Data: Artificial Intelligence and Statistics V*, 121–130 (Springer-Verlag, 1996).
- [148] Chickering, D. M., Heckerman, D. & Meek, C. Large-sample learning of Bayesian networks is NP-hard. *J. Machine Learn. Res.* **5**, 1287–1330 (2004).
- [149] Gustafsson, M., Hornquist, M. & Lombardi, A. Constructing and analyzing a large-scale gene-to-gene regulatory network-lasso-constrained inference and biological validation. *IEEE/ACM Trans. Comp. Biol. Bioinfo.* **2**, 254–261 (2005).
- [150] Shabani, A. *et al.* Efficient measurement of quantum dynamics via compressive sensing. *Phys. Rev. Lett.* **106**, 100401 (2011).
- [151] Bongard, J. & Lipson, H. Automated reverse engineering of nonlinear dynamical systems. *Proc. Nat. Acad. Sci. (USA)* **104**, 9943–9948 (2007).
- [152] Ye, J.-J. Optimality conditions for optimization problems with complementarity constraints. *SIAM J. Optim.* **9**, 374–387 (1999).
- [153] Hénon, M. A two-dimensional mapping with a strange attractor. *Commun. Math. Phys.* **50**, 69–77 (1976).
- [154] Grebogi, C., Ott, E. & Yorke, J. A. Crises, sudden changes in chaotic attractors and chaotic transients. *Physica D* **7**, 181–200 (1983).
- [155] Ott, E. *Chaos in Dynamical Systems* (Cambridge University Press, Cambridge, UK, 2002), second edn.
- [156] Lai, Y.-C. & Tél, T. *Transient Chaos - Complex Dynamics on Finite Time Scales* (Springer, New York, 2011), first edn.
- [157] Chirikov, B. V. & Izraelev, F. M. Some numerical experiments with a nonlinear mapping: Stochastic component. *Colloques. Int. du CNRS* **229**, 409–428 (1973).
- [158] Chirikov, B. V. A universal instability of many-dimensional oscillator systems. *Phys. Rep.* **52**, 263–379 (1979).
- [159] Lorenz, E. N. Deterministic nonperiodic flow. *J. Atmos. Sci.* **20**, 130–141 (1963).
- [160] Rössler, O. E. Equation for continuous chaos. *Phys. Lett. A* **57**, 397–398 (1976).
- [161] Stewart, H. B., Ueda, Y., Grebogi, C. & Yorke, J. A. Double crises in two-parameter dynamical systems. *Phys. Rev. Lett.* **75**, 2478–2481 (1995).
- [162] Bonatto, C. & Gallas, J. A. C. Periodicity hub and nested spirals in the phase diagram of a simple resistive circuit. *Phys. Rev. Lett.* **101**, 054101 (2008).
- [163] Stoop, R., Benner, P. & Uwate, Y. Real-world existence and origins of the spiral organization of shrimp-shaped domains. *Phys. Rev. Lett.* **105**, 074102 (2010).
- [164] Gladwell, M. *The Tipping Point: How Little Thinkgs Can Make a Big Difference* (Little, Brown and Company, USA, 2000).
- [165] Scheffer, M. *et al.* Early-warning signals for critical transitions. *Nature (London)* **461**, 53–59 (2009).

- [166] Rahmstorf, S. Ocean circulation and climate during the past 120,000 years. *Nature (London)* **419**, 207–214 (2002).
- [167] Drake, J. M. & Griffen, B. D. Early warning signals of extinction in deteriorating environments. *Nature (London)* **467**, 456–459 (2010).
- [168] Dai, L., Vorselen, D., Korolev, K. S. & Gore, J. Generic indicators for loss of resilience before a tipping point leading to population collapse. *Science* **336**, 1175–1177 (2012).
- [169] Tylianakis, J. M. & Cux, C. Tipping points in ecological networks. *Trends. Plant. Sci.* **19**, 281–283 (2014).
- [170] Lever, J. J., Nes, E. H., Scheffer, M. & Bascompte, J. The sudden collapse of pollinator communities. *Ecol. lett.* **17**, 350–359 (2014).
- [171] Scheffer, M. *Ecology of Shallow Lakes* (Springer Science & Business Media, 2004).
- [172] Buldyrev, S. V., Parshani, R., Paul, G., Stanley, H. E. & Havlin, S. Catastrophic cascade of failures in interdependent networks. *Nature (London)* **464**, 1025–1028 (2010).
- [173] Gao, J.-X., Buldyrev, S. V., Stanley, H. E. & Havlin, S. Networks formed from interdependent networks. *Nat. Phys.* **8**, 40–48 (2012).
- [174] Scheffer, M. Foreseeing tipping points. *Nature (London)* **467**, 411–412 (2010).
- [175] Linkov, I. *et al.* Measurable resilience for actionable policy. *Environ. Sci. Tech.* **47**, 10108–10110 (2013).
- [176] Park, J., Seager, T. P., Rao, R. S. C., Convertino, M. & Kinkov, I. Integrating risk and resilience approaches to catastrophe management in engineering systems. *Risk Ana.* **33**, 356–367 (2013).
- [177] Park, J., Seager, T. P. & Rao, P. S. C. Understanding resilience of complex value chain networks. In Linggren, A., Maon, F., Vanhamme, J. & Sen, S. (eds.) *Sustainable Value Chain Management: Analyzing, Designing, Implementing, and Monitoring for Social and Environmental Responsibility*, 329–352 (Ashgate Publishing, Surrey UK, 2013).
- [178] Dakos, V., Cartenter, S. R., van Nes, E. H. & Scheffer, M. Resilience indicators: prospects and limitations for early warnings of regime shifts. *Philos. Trans. B* **370**, 20130263 (2015).
- [179] Gao, J.-X., Barzel, B. & Barabasi, A.-L. Universal resilience patterns in complex networks. *Nature (London)* **530**, 307–312 (2016).
- [180] Goldstein, H., Poole, C. & Safko, J. *Classical Mechanics* (Addison Wesley, San Francisco, 2002), third edn.
- [181] Waddington, C. H. *The Strategy of the Genes* (Allen & Unwin, London, 1957).
- [182] Huang, S. Reprogramming cell fates: reconciling rarity with robustness. *Bioessays* **31**, 546–560 (2009).
- [183] MacArthur, B. D., Maayan, A. & Lemischka, I. R. Systems biology of stem cell fate and cellular reprogramming. *Nat. Rev. Mol. Cell Biol.* **10**, 672–681 (2009).
- [184] Wang, J., Xu, L. & Wang, E. K. Potential landscape and flux framework of nonequilibrium networks: Robustness, dissipation, and coherence of biochemical oscillations. *Proc. Natl. Acad. Sci. (USA)* **105**, 12271–12276 (2008).
- [185] Wang, J., Xu, L., Wang, E. K. & Huang, S. The potential landscape of genetic circuits imposes the arrow of time in stem cell differentiation. *Biophys. J.* **99**, 29–39 (2010).
- [186] Wang, J., Zhang, K., Xu, L. & Wang, E. K. Quantifying the waddington landscape and biological paths for development and differentiation. *Proc. Natl. Acad. Sci. (USA)* **108**, 8257–8262 (2011).

- [187] Zhang, F., Xu, L., Zhang, K., Wang, E. K. & Wang, J. The potential and flux landscape theory of evolution. *J. Chem. Phys.* **137**, 065102 (2012).
- [188] Garham, R. & Tél, T. Existence of a potential for dissipative dynamical systems. *Phys. Rev. Lett.* **52**, 9–12 (1984).
- [189] Graham, R., Hamm, A. & Tél, T. Nonequilibrium potentials for dynamical systems with fractal attractors and repellers. *Phys. Rev. Lett.* **66**, 3089–3092 (1991).
- [190] Tél, T. & Lai, Y.-C. Quasipotential approach to critical scaling in noise-induced chaos. *Phys. Rev. E* **81**, 056208 (2010).
- [191] Gartner, J. & Wheellock, C. Electric vehicles on the grid. *Research Report, PikeResearch LLC* (2009).
- [192] Mak, H.-I., Rong, Y. & Shen, Z.-J. M. Infrastructure planning for electric vehicles with battery swapping. *Manage. Sci.* **59**, 1557–1575 (2013).
- [193] Holmes, P. J. & Rand, D. A. The bifurcations of duffing's equation: An application of catastrophe theory. *J. Sound Vibr.* **44**, 237–253 (1976).
- [194] Sommerlade, L. *et al.* Optimized spectral estimation for nonlinear synchronizing systems. *Phys. Rev. E* **89**, 032912 (2014).
- [195] Strogatz, S. H. Exploring complex networks. *Nature (London)* **410**, 268–276 (2001).
- [196] Erdős, P. & Rényi, A. On random graphs i. *Publ. Math. Debrecen* **6**, 290–291 (1959).
- [197] Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
- [198] Lusseau, D. *et al.* The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations. *Behav. Ecol. Sociobiol.* **54**, 396–405 (2003).
- [199] Zachary, W. W. An information flow model for conflict and fission in small groups. *J. Anthropol. Res.* **33**, 452–473 (1977).
- [200] Milo, R. *et al.* Superfamilies of evolved and designed networks. *Science* **303**, 1538–1542 (2004).
- [201] Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature (London)* **393**, 440–442 (1998).
- [202] Smith, J. M. *Evolution and the Theory of Games* (Cambridge University Press, Cambridge, UK, 1982).
- [203] Weibull, J. W. *Evolutionary Game Theory* (MIT Press, Cambridge, MA, 1997).
- [204] Hofbauer, J. & Sigmund, K. *Evolutionary Games and Population Dynamics* (Cambridge University Press, Cambridge, UK, 1998).
- [205] Nowak, M. A. *Evolutionary Dynamics: Exploring the Equations of Life* (Harvard University Press, Cambridge, MA, 2006).
- [206] Szabó, G. & Fath, G. Evolutionary games on graphs. *Phys. Rep.* **446**, 97–216 (2007).
- [207] Frean, M. & Abraham, E. R. Rock–scissors–paper and the survival of the weakest. *P. Roy. Soc. B-Biol. Sci.* **268**, 1323–1327 (2001).
- [208] Traulsen, A., Claussen, J. C. & Hauert, C. Coevolutionary dynamics: from finite to infinite populations. *Phys. Rev. Lett.* **95**, 238701 (2005).
- [209] Reichenbach, T., Mobilia, M. & Frey, E. Mobility promotes and jeopardizes biodiversity in rockpaper scissors games. *Nature (London)* **448**, 1046–1049 (2007).
- [210] Reichenbach, T., Mobilia, M. & Frey, E. Noise and correlations in a spatial population model with cyclic competition. *Phys. Rev. Lett.* **99**, 238105 (2007).

- [211] Peltomäki, M. & Alava, M. Three-and four-state rock-paper-scissors games with diffusion. *Phys. Rev. E* **78**, 031906 (2008).
- [212] Szabó, G. & Szolnoki, A. Phase transitions induced by variation of invasion rates in spatial cyclic predator-prey models with four or six species. *Phys. Rev. E* **77**, 011906 (2008).
- [213] Berr, M., Reichenbach, T., Schottenloher, M. & Frey, E. Zero-one survival behavior of cyclically competing species. *Phys. Rev. Lett.* **102**, 048102 (2009).
- [214] Shi, H., Wang, W.-X., Yang, R. & Lai, Y.-C. Basins of attraction for species extinction and coexistence in spatial rock-paper-scissors games. *Phys. Rev. E* **81**, 030901 (2010).
- [215] Ni, X., Yang, R., Wang, W.-X., Lai, Y.-C. & Grebogi, C. Basins of coexistence and extinction in spatially extended ecosystems of cyclically competing species. *Chaos* **20**, 045116 (2010).
- [216] Yang, R., Wang, W.-X., Lai, Y.-C. & Grebogi, C. Role of intraspecific competition in the coexistence of mobile populations in spatially extended ecosystems. *Chaos* **20**, 023113 (2010).
- [217] He, Q., Mobilia, M. & Täuber, U. C. Spatial rock-paper-scissors models with inhomogeneous reaction rates. *Phys. Rev. E* **82**, 051909 (2010).
- [218] Wang, W.-X., Lai, Y.-C. & Grebogi, C. Effect of epidemic spreading on species coexistence in spatial rock-paper-scissors games. *Phys. Rev. E* **81**, 046113 (2010).
- [219] Frey, E. Evolutionary game theory: Theoretical concepts and applications to microbial communities. *Physica A* **389**, 4265–4298 (2010).
- [220] Ni, X., Wang, W.-X., Lai, Y.-C. & Grebogi, C. Cyclic competition of mobile species on continuous space: Pattern formation and coexistence. *Phys. Rev. E* **82**, 066211 (2010).
- [221] Venkat, S. & Pleimling, M. Mobility and asymmetry effects in one-dimensional rock-paper-scissors games. *Phys. Rev. E* **81**, 021917 (2010).
- [222] Wang, W.-X., Ni, X., Lai, Y.-C. & Grebogi, C. Pattern formation, synchronization, and outbreak of biodiversity in cyclically competing games. *Phys. Rev. E* **83**, 011917 (2011).
- [223] Juul, J., Sneppen, K. & Mathiesen, J. Clonal selection prevents tragedy of the commons when neighbors compete in a rock-paper-scissors game. *Phys. Rev. E* **85**, 061924 (2012).
- [224] Lamouroux, D., Eule, S., Geisel, T. & Nagler, J. Discriminating the effects of spatial extent and population size in cyclic competition among species. *Phys. Rev. E* **86**, 021911 (2012).
- [225] Avelino, P., Bazeia, D., Losano, L., Menezes, J. & Oliveira, B. Junctions and spiral patterns in generalized rock-paper-scissors models. *Phys. Rev. E* **86**, 036112 (2012).
- [226] Avelino, P., Bazeia, D., Losano, L. & Menezes, J. Von Neumann's and related scaling laws in rock-paper-scissors-type games. *Phys. Rev. E* **86**, 031119 (2012).
- [227] Jiang, L.-L., Wang, W.-X., Lai, Y.-C. & Ni, X. Multi-armed spirals and multi-pairs antispirals in spatial rockpaper-scissors games. *Phys. Lett. A* **376**, 2292–2297 (2012).
- [228] Juul, J., Sneppen, K. & Mathiesen, J. Labyrinthine clustering in a spatial rock-paper-scissors ecosystem. *Phys. Rev. E* **87**, 042702 (2013).
- [229] Knebel, J., Krüger, T., Weber, M. F. & Frey, E. Coexistence and Survival in Conservative Lotka-Volterra Networks. *Phys. Rev. Lett.* **110**, 168106 (2013).
- [230] Hua, D.-Y., Dai, L.-C. & Lin, C. Four- and three-state rock-paper-scissors games with long-range selection. *Europhys. Lett.* **101**, 38004 (2013).
- [231] Park, J., Do, Y., Huang, Z.-G. & Lai, Y.-C. Persistent coexistence of cyclically competing species in spatially extended ecosystems. *Chaos* **23**, 023128 (2013).

- [232] Kang, Y., Pan, Q., Wang, X. & He, M. A golden point rule in rock-paper-scissors-lizard-spock game. *Physica A* **392**, 2652–2659 (2013).
- [233] Dobrinevski, A., Alava, M., Reichenbach, T. & Frey, E. Mobility-dependent selection of competing strategy associations. *Phys. Rev. E* **89**, 012721 (2014).
- [234] Zheng, H.-Y. *et al.* Mesoscopic interactions and species coexistence in evolutionary game dynamics of cyclic competitions. *Sci. Rep.* **4**, 7486 (2014).
- [235] Nowak, M. A. & May, R. M. Evolutionary games and spatial chaos. *Nature (London)* **359**, 826–829 (1992).
- [236] Hauert, C. & Doebeli, M. Spatial structure often inhibits the evolution of cooperation in the snowdrift game. *Nature (London)* **428**, 643–646 (2004).
- [237] Szabó, G. & Tóke, C. Evolutionary prisoner’s dilemma game on a square lattice. *Phys. Rev. E* **58**, 69–73 (1998).
- [238] Santos, F. C., Santos, M. D. & Pacheco, J. M. Social diversity promotes the emergence of cooperation in public goods games. *Nature (London)* **454**, 213–216 (2008).
- [239] Szolnoki, A., Perc, M. & Danku, Z. Towards effective payoffs in the prisoners dilemma game on scale-free networks. *Physica A* **387**, 2075–2082 (2008).
- [240] Santos, F. C., Pacheco, J. M. & Lenaerts, T. Evolutionary dynamics of social dilemmas in structured heterogeneous populations. *Proc. Natl. Acad. Sci. (USA)* **103**, 3490–3494 (2006).
- [241] Gardner, T. S., di Bernardo, D., Lorenz, D. & Collins, J. J. Inferring genetic networks and identifying compound mode of action via expression profiling. *Science* **301**, 102–105 (2003).
- [242] Bansal, M., Belcastro, V., Ambesi-Impiombato, A. & di Bernardo, D. How to infer gene networks from expression profiles. *Mol. Syst. Biol.* **3**, 78 (2007).
- [243] Geier, F., Timmer, J. & Fleck, C. Reconstructing gene-regulatory networks from time series, knock-out data, and prior knowledge. *BMC Syst. Biol.* **1**, 11 (2007).
- [244] Hecker, M., Lambeck, S., Toepferb, S., van Someren, E. & Guthke, R. Gene regulatory network inference: Data integration in dynamic models - a review. *BioSyst.* **96**, 86–103 (2009).
- [245] Gardiner, C. W. *Handbook of Stochastic Methods* (Springer, Berlin, 1985).
- [246] Wang, X. F. & Chen, G. Pinning control of scale-free dynamical networks. *Physica A* **310**, 521–531 (2002).
- [247] Li, X., Wang, X. F. & Chen, G. Pinning a complex dynamical network to its equilibrium. *IEEE Trans. Circ. Syst. I* **51**, 2074–2087 (2004).
- [248] Sorrentino, F., di Bernardo, M., Garofalo, F. & Chen, G. Controllability of complex networks via pinning. *Phys. Rev. E* **75**, 046103 (2007).
- [249] Yu, W., Chen, G. & Lü, J. On pinning synchronization of complex dynamical networks. *Automatica* **45**, 429–435 (2009).
- [250] Hodgkin, A. L. & Huxley, A. F. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physio.* **17**, 500–544 (1952).
- [251] FitzHugh, R. Impulses and physiological states in theoretical models of nerve membrane. *Biophys. J.* **1**, 445–466 (1961).
- [252] Nagumo, J., Arimoto, S. & Yoshizawa, S. An active pulse transmission line simulating nerve axon. *Proc. IRE* **50**, 2061–2070 (1962).
- [253] Wu, X. Synchronization-based topology identification of weighted general complex dynamical networks with time-varying coupling delay. *Physica A* **387**, 997–1008 (2008).

- [254] Husmeier, D. Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic bayesian networks. *Bioinformatics* **19**, 2271–2282 (2003).
- [255] Sayed, A. H., Tarighat, A. & Khajehnouri, N. Network-based wireless location: challenges faced in developing techniques for accurate wireless location information. *IEEE Signal Process. Mag.* **22**, 24–40 (2005).
- [256] Moon, I.-C. & Carley, K. Modeling and simulating terrorist networks in social and geospatial dimensions. *IEEE Intell. Syst.* **22**, 40–49 (2007).
- [257] Li, C. & Chen, G. Synchronization in general complex dynamical networks with coupling delays. *Physica A* **343**, 263–278 (2004).
- [258] Dhamala, M., Jirsa, V. & Ding, M. Enhancement of neural synchrony by time delay. *Phys. Rev. Lett.* **92**, 074104 (2004).
- [259] Gross, T., Rudolf, L., Levin, S. & Dieckmann, U. Generalized models reveal stabilizing factors in food webs. *Science* **325**, 747–750 (2009).
- [260] Shang, Y. & Ruml, W. Improved mds-based localization. In *INFOCOM 2004. Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies*, vol. 4, 2640–2651 (IEEE, 2004).
- [261] Shang, Y., Ruml, W., Zhang, Y. & Fromherz, M. P. Localization from mere connectivity. In *Proceedings of the 4th ACM International Symposium on Mobile ad hoc Networking and Computing*, 201–212 (ACM, 2003).
- [262] Bellen, A. & Zennaro, M. *Numerical Methods for Delay Differential Equations* (Oxford University Press, 2013).
- [263] Pastor-Satorras, R. & Vespignani, A. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* **86**, 3200–3203 (2001).
- [264] Eames, K. T. D. & Keeling, M. J. Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases. *Proc. Natl. Acad. Sci. (USA)* **99**, 13330–13335 (2002).
- [265] Watts, D. J., Muhamad, R., Medina, D. C. & Dodds, P. S. Multiscale, resurgent epidemics in a hierarchical metapopulation model. *Proc. Natl. Acad. Sci. (USA)* **102**, 11157–11162 (2005).
- [266] Colizza, V., Barrat, A., Barthélemy, M. & Vespignani, A. The role of the airline transportation network in the prediction and predictability of global epidemics. *Proc. Natl. Acad. Sci. (USA)* **103**, 2015–2020 (2006).
- [267] Gómez-Gardeñes, J., Latora, V., Moreno, Y. & Profumo, E. Spreading of sexually transmitted diseases in heterosexual populations. *Proc. Natl. Acad. Sci. (USA)* **105**, 1399–1404 (2008).
- [268] Wang, P., González, M. C., Hidalgo, C. A. & Barabási, A.-L. Understanding the spreading patterns of mobile phone viruses. *Science* **324**, 1071–1076 (2009).
- [269] Merler, S. & Ajelli, M. The role of population heterogeneity and human mobility in the spread of pandemic influenza. *Proc. Roy. Soc. B Bio. Sci.* **277**, 557–565 (2010).
- [270] Balcan, D. & Vespignani, A. Phase transitions in contagion processes mediated by recurrent mobility patterns. *Nat. Phys.* **7**, 581–586 (2011).
- [271] Boguná, M., Pastor-Satorras, R. & Vespignani, A. Absence of epidemic threshold in scale-free networks with degree correlations. *Phys. Rev. Lett.* **90**, 028701 (2003).

- [272] Parshani, R., Carmi, S. & Havlin, S. Epidemic threshold for the susceptible-infectious-susceptible model on random networks. *Phys. Rev. Lett.* **104**, 258701 (2010).
- [273] Castellano, C. & Pastor-Satorras, R. Thresholds for epidemic spreading in networks. *Phys. Rev. Lett.* **105**, 218701 (2010).
- [274] Gleeson, J. P. High-accuracy approximation of binary-state dynamics on networks. *Phys. Rev. Lett.* **107**, 068701 (2011).
- [275] Gomez, R. M., Leskovec, J. & Krause, A. Inferring networks of diffusion and influence. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1019–1028 (ACM, 2010).
- [276] Pinto, P. C., Thiran, P. & Vetterli, M. Locating the source of diffusion in large-scale networks. *Phys. Rev. Lett.* **109**, 068702 (2012).
- [277] Myers, S. A. & Leskovec, J. On the convexity of latent social network inference. *Adv. Neu. Info. Proc. Sys.* **23**, 1741–1749 (2010).
- [278] Castellano, C. & Pastor-Satorras, R. Non-mean-field behavior of the contact process on scale-free networks. *Phys. Rev. Lett.* **96**, 038701 (2006).
- [279] Volz, E. & Meyers, L. A. Epidemic thresholds in dynamic contact networks. *J. Roy. Soc. Interface* **6**, 233–241 (2009).
- [280] Cohen, R., Havlin, S. & ben Avraham, D. Efficient immunization strategies for computer networks and populations. *Phys. Rev. Lett.* **91**, 247901 (2003).
- [281] Forster, G. A. & Gilligan, C. A. Optimizing the control of disease infestations at the landscape scale. *Proc. Natl. Acad. Sci. (USA)* **104**, 4984–4989 (2007).
- [282] Klepac, P., Laxminarayan, R. & Grenfell, B. T. Synthesizing epidemiological and economic optima for control of immunizing infections. *Proc. Natl. Acad. Sci. (USA)* **108**, 14366–14370 (2011).
- [283] Kleczkowski, A., Oleś, K., Gudowska-Nowak, E. & Gilligan, C. A. Searching for the most cost-effective strategy for controlling epidemics spreading on regular and small-world networks. *J. R. Soc. Interface* **9**, 158–169 (2012).
- [284] Benzi, R., Sutera, A. & Vulpiani, A. The mechanism of stochastic resonance. *J. Phys. A* **14**, L453–L457 (1981).
- [285] Benzi, R., Parisi, G., Sutera, A. & Vulpiani, A. A theory of stochastic resonance in climatic-change. *J. Appl. Math.* **43**, 565–578 (1983).
- [286] McNamara, B. & Wiesenfeld, K. Theory of stochastic resonance. *Phys. Rev. A* **39**, 4854–4869 (1989).
- [287] Moss, F., Pierson, D. & O’Gorman, D. Stochastic resonance - tutorial and update. *Int. J. Bif. Chaos* **4**, 1383–1397 (1994).
- [288] Gailey, P. C., Neiman, A., Collins, J. J. & Moss, F. Stochastic resonance in ensembles of nondynamical elements: The role of internal noise. *Phys. Rev. Lett.* **79**, 4701–4704 (1997).
- [289] Gammaitoni, L., Hänggi, P., Jung, P. & Marchesoni, F. Stochastic resonance. *Rev. Mod. Phys.* **70**, 223–287 (1998).
- [290] Sigeti, D. & Horsthemke, W. Pseudo-regular oscillations induced by external noise. *J. Stat. Phys.* **54**, 1217 (1989).
- [291] Pikovsky, A. S. & Kurths, J. Coherence resonance in a noise-driven excitable system. *Phys. Rev. Lett.* **78**, 775–778 (1997).
- [292] Liu, Z. & Lai, Y.-C. Coherence resonance in coupled chaotic oscillators. *Phys. Rev. Lett.* **86**, 4737–4740 (2001).



- [293] Lai, Y.-C. & Liu, Z. Noise-enhanced temporal regularity in coupled chaotic oscillators. *Phys. Rev. E* **64**, 066202 (2001).
- [294] Hu, G., Ditzinger, T., Ning, C. Z. & Haken, H. Stochastic resonance without external periodic force. *Phys. Rev. Lett.* **71**, 807 (1993).
- [295] Zhou, C. & Kurths, J. Hierarchical synchronization in complex networks with heterogeneous degrees. *Chaos* **16**, 015104 (2006).
- [296] Ching, E. S., Lai, P.-Y. & Leung, C. Extracting connectivity from dynamics of networks with uniform bidirectional coupling. *Phys. Rev. E* **88**, 042817 (2013).
- [297] Zhang, Z. *et al.* Solving the inverse problem of noise-driven dynamic networks. *Phys. Rev. E* **91**, 012814 (2015).
- [298] Barzel, B., Liu, Y.-Y. & Barabási, A.-L. Constructing minimal models for complex system dynamics. *Nat. Commun.* **6** (2015).
- [299] Hahn, H. Über die nichtarchimedischen größensysteme. In *Hans Hahn Gesammelte Abhandlungen Band 1/Hans Hahn Collected Works Volume 1*, 445–499 (Springer, 1995).
- [300] Pearson, K. Note on regression and inheritance in the case of two parents. *Proc. Royal Soc. (London)* **58**, 240–242 (1895).
- [301] Székely, G. J., Rizzo, M. L. *et al.* Brownian distance covariance. *Ann. Appl. Stat.* **3**, 1236–1265 (2009).
- [302] Wilcox, R. A note on the theil-sen regression estimator when the regressor is random and the error term is heteroscedastic. *Biomet. J.* **40**, 261–268 (1998).
- [303] Peng, H., Wang, S. & Wang, X. Consistency and asymptotic distribution of the theil-sen estimator. *J. Stat. Plan. Inference* **138**, 1836–1850 (2008).
- [304] de la Fuente, A., Bing, N., Hoeschele, I. & Mendes, P. Discovery of meaningful associations in genomic data using partial correlation coefficients. *Bioinfo.* **20**, 3565–3574 (2004).
- [305] Reverter, A. & Chan, E. K. F. Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks. *Bioinfo.* **24**, 2491–2497 (2008).
- [306] Nie, J. *et al.* Tf-cluster: A pipeline for identifying functionally coordinated transcription factors via network decomposition of the shared coexpression connectivity matrix (sccm). *BMC Sys. Bio.* **5**, 1 (2011).
- [307] Kendall, M. G. A new measure of rank correlation. *Biometrika* **30**, 81–93 (1938).
- [308] Fujita, A. *et al.* Comparing pearson, spearman and hoeffding’s d measure for gene expression association analysis. *J. Bioinfo. Comp. Bio.* **7**, 663–684 (2009).
- [309] Wang, Y. R., Waterman, M. S. & Huang, H. Gene coexpression measures in large heterogeneous samples using count statistics. *Proc. Natl. Acad. Sci. (USA)* **111**, 16371–16376 (2014).
- [310] Wiener, N. The theory of prediction. *Mod. Math. Eng.* **1**, 125–139 (1956).
- [311] Bressler, S. L. & Seth, A. K. Wiener-Granger causality: a well established methodology. *Neuroimage* **58**, 323–329 (2011).
- [312] Chen, Y., Rangarajan, G., Feng, J. & Ding, M. Analyzing multiple nonlinear time series with extended Granger causality. *Phys. Lett. A* **324**, 26–35 (2004).
- [313] Ancona, N., Marinazzo, D. & Stramaglia, S. Radial basis function approach to nonlinear Granger causality of time series. *Phys. Rev. E* **70**, 056221 (2004).
- [314] Marinazzo, D., Pellicoro, M. & Stramaglia, S. Kernel method for nonlinear Granger causality. *Phys. Rev. Lett.* **100**, 144103 (2008).

- [315] Nalatore, H., Sasikumar, N. & Rangarajan, G. Effect of measurement noise on Granger causality. *Phys. Rev. E* **90**, 062127 (2014).
- [316] Schreiber, T. Measuring information transfer. *Phys. Rev. Lett.* **85**, 461–464 (2000).
- [317] Barnett, L., Barrett, A. B. & Seth, A. K. Granger causality and transfer entropy are equivalent for Gaussian variables. *Phys. Rev. Lett.* **103**, 238701 (2009).
- [318] Sun, J., Cafaro, C. & Bollt, E. M. Identifying coupling structure in complex systems through the optimal causation entropy principle. *Entropy* **16**, 3416–3433 (2014).
- [319] Cafaro, C., Lord, W. M., Sun, J. & Bollt, E. M. Causation entropy from symbolic representations of dynamical systems. *Chaos* **25**, 043106 (2015).
- [320] Sun, J., Taylor, D. & Bollt, E. M. Causal network inference by optimal causation entropy. *SIAM J. Dyn. Syst.* **14**, 73–106 (2015).
- [321] Deyle, E. R. & Sugihara, G. Generalized theorems for nonlinear state space reconstruction. *PLoS One* **6**, e18295 (2011).
- [322] McBride, J. C. *Dynamic Complexity and Causality Analysis of Scalp EEG for Detection of Cognitive Deficits*. Ph.D. thesis, University of Tennessee, Knoxville (2014).
- [323] Wismüller, A., Wang, X.-X., DSouza, A. M. & Nagarajan, N. B. A framework for exploring non-linear functional connectivity and causality in the human brain: mutual connectivity analysis (mca) of resting-state functional mri with convergent cross-mapping and non-metric clustering. *arXiv:1407.3809* (2014).
- [324] Harford, W. *et al.* Can climate explain temporal trends in king mackerel (*scomberomorus cavalla*) catch-per-unit-effort and landings? Tech. Rep., SEDAR (2014). SEDAR38-AW-04. SEDAR, North Charleston, SC.
- [325] Huffaker, R. & Fearn, A. Empirically testing for dynamic causality between promotions and sales beer promotions and sales in england. *Proc. Food. Syst. Dyn* 270–274 (2014).
- [326] Heskamp, L., Meel-van den Abeelen, A. S., Lagro, J. & Claassen, J. A. Convergent cross mapping: a promising technique for cerebral autoregulation estimation. *IJCNMH* **1**(Suppl. 1), S20 (2014).
- [327] Jiang, J.-J., Huang, Z.-G., Huang, L., Liu, H. & Lai, Y.-C. Direct dynamical influence is more detectable with noise. *Sci. Rep.* (2016).
- [328] Sugihara, G. *et al.* Detecting causality in complex ecosystems. *Science* **338**, 496–500 (2012).
- [329] Sugihara, G. & May, R. M. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature (London)* **344**, 734–741 (1990).
- [330] MacKay, D. J. *Information Theory, Inference and Learning Algorithms* (Cambridge university press, 2003).
- [331] Butte, A. J., Tamayo, P., Slonim, D., Golub, T. R. & Kohane, I. S. Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks. *Proc. Natl. Acad. Sci. (USA)* **97**, 12182–12186 (2000).
- [332] Butte, A. J. & Kohane, I. S. Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. In *Pac. Symp. Biocomput.*, vol. 5, 418–429 (World Scientific, 2000).
- [333] Faith, J. J. *et al.* Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. *PLoS Biol.* **5**, 54–66 (2007).
- [334] Meyer, P. E., Kontos, K., Lafitte, F. & Bontempi, G. Information-theoretic inference of large transcriptional regulatory networks. *EURASIP J. Bioinfo. Sys. Bio.* **2007**, 79879 (2007).

- [335] Meyer, P., Marbach, D., Roy, S. & Kellis, M. Information-theoretic inference of gene networks using backward elimination. In *BIOCOMP*, 700–705 (2010).
- [336] Margolin, A. *et al.* Aracne: An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinfo.* **7**, S7 (2006). NIPS Workshop on New Problems and Methods in Computational Biology, Whistler, CANADA, DEC 18, 2004.
- [337] Zoppoli, P., Morganella, S. & Ceccarelli, M. Timedelay-aracne: Reverse engineering of gene networks from time-course data by an information theoretic approach. *BMC Bioinfo.* **11**, 154 (2010).
- [338] Tourassi, G. D., Frederick, E. D., Markey, M. K. & Floyd Jr, C. E. Application of the mutual information criterion for feature selection in computer-aided diagnosis. *Med. Phys.* **28**, 2394–2402 (2001).
- [339] Ding, C. & Peng, H. Minimum redundancy feature selection from microarray gene expression data. *J. Bioinfo. Comp. Bio.* **3**, 185–205 (2005).
- [340] Cover, T. M. & Thomas, J. A. *Elements of Information Theory* (John Wiley & Sons, 2012).
- [341] Reshef, D. N. *et al.* Detecting novel associations in large data sets. *Science* **334**, 1518–1524 (2011).
- [342] Kinney, J. B. & Atwal, G. S. Equitability, mutual information, and the maximal information coefficient. *Proc. Natl. Acad. Sci. (USA)* **111**, 3354–3359 (2014).
- [343] Simon, N. & Tibshirani, R. Comment on “detecting novel associations in large data sets” by reshef et al, science dec 16, 2011. *arXiv preprint arXiv:1401.7645* (2014).
- [344] Cocco, S., Leibler, S. & Monasson, R. Neuronal couplings between retinal ganglion cells inferred by efficient inverse statistical physics methods. *Proc. Natl. Acad. Sci. (USA)* **106**, 14058–14062 (2009).
- [345] Schneidman, E., Berry, M., Segev, R. & Bialek, W. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature (London)* **440**, 1007–1012 (2006).
- [346] Yeung, K. Y., Bumgarner, R. E. & Raftery, A. E. Bayesian model averaging: development of an improved multi-class, gene selection and classification tool for microarray data. *Bioinfo.* **21**, 2394–2402 (2005).
- [347] Ferrazzi, F., Sebastiani, P., Ramoni, M. F. & Bellazzi, R. Bayesian approaches to reverse engineer cellular systems: a simulation study on nonlinear Gaussian networks. *BMC Bioinfo.* **8**, S2 (2007). 10th Annual International Conference on Research in Computational Molecular Biology, Venice, ITALY, APR 02-05, 2006.
- [348] Tsamardinos, I., Aliferis, C. F. & Statnikov, A. Time and sample efficient discovery of markov blankets and direct causal relations. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 673–678 (ACM, 2003).
- [349] Aliferis, C. F., Statnikov, A., Tsamardinos, I., Mani, S. & Koutsoukos, X. D. Local causal and markov blanket induction for causal discovery and feature selection for classification part i: Algorithms and empirical evaluation. *J. Mach. Learn. Res.* **11**, 171–234 (2010).
- [350] Statnikov, A. & Aliferis, C. F. Analysis and computational dissection of molecular signature multiplicity. *PLoS Comp. Bio.* **6**, e1000790 (2010).
- [351] Tibshirani, R. Regression shrinkage and selection via the lasso. *J. Royal Stat. Soc. Ser. B (Method.)* **58**, 267–288 (1996).
- [352] Haury, A.-C., Mordelet, F., Vera-Licona, P. & Vert, J.-P. Tigress: Trustful inference of gene regulation using stability selection. *BMC Sys. Bio.* **6**, 145 (2012).

- [353] Yuan, M. & Lin, Y. Model selection and estimation in regression with grouped variables. *J. Royal Stat. Soc. Ser. B (Stat. Method.)* **68**, 49–67 (2006).
- [354] Li, H. & Zhan, M. Unraveling transcriptional regulatory programs by integrative analysis of microarray and transcription factor binding data. *Bioinfo.* **24**, 1874–1880 (2008).
- [355] Song, L., Langfelder, P. & Horvath, S. Comparison of co-expression measures: mutual information, correlation, and model based indices. *BMC Bioinfo.* **13**, 328 (2012).
- [356] Maetschke, S. R., Madhamshettiwar, P. B., Davis, M. J. & Ragan, M. A. Supervised, semi-supervised and unsupervised inference of gene regulatory networks. *Brief. Bioinfo.* **15**, 195–211 (2014).
- [357] Huynh-Thu, V. A., Irrthum, A., Wehenkel, L. & Geurts, P. Inferring regulatory networks from expression data using tree-based methods. *PLOS ONE* **5**, e12776 (2010).
- [358] Mordelet, F. & Vert, J.-P. Sirene: supervised inference of regulatory networks. *Bioinfo.* **24**, I76–I82 (2008). Joint Meeting of the 7th European Conference on Computational Biology/5th Meeting of the Bioinformatics-Italian-Society, Cagliari, ITALY, SEP 22–26, 2008.
- [359] Cerulo, L., Elkan, C. & Ceccarelli, M. Learning gene regulatory networks from only positive and unlabeled data. *BMC Bioinfo.* **11**, 228 (2010).
- [360] Lungarella, M. & Sporns, O. Mapping information flow in sensorimotor networks. *PLoS Comp. Bio.* **2**, e144 (2006).
- [361] Buehlmann, A. & Deco, G. Optimal information transfer in the cortex through synchronization. *PLoS Comput. Biol.* **6**, e1000934 (2010).
- [362] Honey, C. J., Kötter, R., Breakspear, M. & Sporns, O. Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *Proc. Natl. Acad. Sci. (USA)* **104**, 10240–10245 (2007).
- [363] Garofalo, M., Nieuws, T., Massobrio, P. & Martinoia, S. Evaluation of the performance of information theory-based methods and cross-correlation to estimate the functional connectivity in cortical networks. *PLOS ONE* **4**, e6482 (2009).
- [364] Ito, S. *et al.* Extending transfer entropy improves identification of effective connectivity in a spiking cortical network model. *PLOS ONE* **6**, e27431 (2011).
- [365] Marbach, D. *et al.* Revealing strengths and weaknesses of methods for gene network inference. *Proc. Natl. Acad. Sci. (USA)* **107**, 6286–6291 (2010).
- [366] Feizi, S., Marbach, D., Medard, M. & Kellis, M. Network deconvolution as a general method to distinguish direct dependencies in networks. *Nat. Biotechnol.* **31**, 726–733 (2013).
- [367] Kholodenko, B. *et al.* Untangling the wires: A strategy to trace functional interactions in signaling and gene networks. *Proc. Natl. Acad. Sci. (USA)* **99**, 12841–12846 (2002).
- [368] Stark, J., Callard, R. & Hubank, M. From the top down: towards a predictive biology of signalling networks. *Trend. Biotech.* **21**, 290–293 (2003).
- [369] Kholodenko, B. N. Untangling the signalling wires. *Nat. Cell Bio.* **9**, 247–249 (2007).
- [370] Kholodenko, B., Yaffe, M. B. & Kolch, W. Computational approaches for analyzing information flow in biological networks. *Sci. Signal.* 2002961 (2012).
- [371] Barzel, B. & Barabasi, A.-L. Network link prediction by global silencing of indirect correlations. *Nat. Biotechnol.* **31**, 720–725 (2013).
- [372] Lago-Fernandez, L. F., Huerta, R., Corbacho, F. & Sigüenza, J. A. Fast response and temporal coherent oscillations in small-world networks. *Phys. Rev. Lett.* **84**, 2758–2761 (2000).

- [373] Gade, P. M. & Hu, C.-K. Synchronous chaos in coupled map lattices with small-world interactions. *Phys. Rev. E* **62**, 6409–6413 (2000).
- [374] Jost, J. & Joy, M. P. Spectral properties and synchronization in coupled map lattices. *Phys. Rev. E* **65**, 016201 (2001).
- [375] Barahona, M. & Pecora, L. M. Synchronization in small-world systems. *Phys. Rev. Lett.* **89**, 054101 (2002).
- [376] Wang, X. F. & Chen, G. Synchronization in small-world dynamical networks. *Int. J. Bifurcation Chaos Appl. Sci.* **12**, 187–192 (2002).
- [377] Wang, X. F. & Chen, G. Making a continuous-time minimum-phase system chaotic by using time-delay feedback. *IEEE Trans. on Circ. Sys., Part I* **48**, 641–645 (2001).
- [378] Hong, H., Choi, M. Y. & Kim, B. J. Synchronous chaos in coupled map lattices with small-world interactions. *Phys. Rev. E* **65**, 026139 (2002).
- [379] Barahona, M. & Pecora, L. M. Heterogeneity in oscillator networks: Are smaller worlds easier to synchronize? *Phys. Rev. Lett.* **91**, 014101 (2003).
- [380] Belykh, V., Belykh, I. & Hasler, M. Connection graph stability method for synchronized coupled chaotic systems. *Physica D* **195**, 159–187 (2004).
- [381] Belykh, I., Hasler, M., Lauret, M. & Nijmeijer, H. Synchronization and graph topology. *Int. J. Bif. Chaos* **15**, 3423–3433 (2005).
- [382] Chavez, M., Hwang, D.-U., Amann, A., Hentschel, H. G. E. & Boccaletti, S. Synchronization is enhanced in weighted complex networks. *Phys. Rev. Lett.* **94**, 218701 (2005).
- [383] Donetti, L., Hurtado, P. I. & noz, M. A. M. Entangled networks, synchronization, and optimal network topology. *Phys. Rev. Lett.* **95**, 188701 (2005).
- [384] Zhou, C. & Kurths, J. Dynamical weights and enhanced synchronization in adaptive complex networks. *Phys. Rev. Lett.* **96**, 164102 (2006).
- [385] Zhou, C. & Kurths, J. Hierarchical synchronization in complex networks with heterogeneous degrees. *Chaos* **16**, 015104 (2006).
- [386] Park, K., Lai, Y.-C., Gupte, S. & Kim, J.-W. Synchronization in complex networks with a modular structure. *Chaos* **16**, 015105 (2006).
- [387] Huang, L., Park, K., Lai, Y.-C., Yang, L. & Yang, K. Abnormal synchronization in complex clustered networks. *Phys. Rev. Lett.* **97**, 164101 (2006).
- [388] Wang, X. G., Huang, L., Lai, Y.-C., & Lai, C.-H. Optimization of synchronization in gradient clustered networks. *Phys. Rev. E* **76**, 056113 (2007).
- [389] Guan, S.-G., Wang, X.-G., Lai, Y.-C. & Lai, C. H. Transition to global synchronization in clustered networks. *Phys. Rev. E* **77**, 046211 (2008).
- [390] Arenas, A., Díaz-Guilera, A. & Guimerà, R. Communication in networks with hierarchical branching. *Phys. Rev. Lett.* **86**, 3196–3199 (2001).
- [391] Echenique, P., Gómez-Gardeñes, J. & Moreno, Y. Improved routing strategies for internet traffic delivery. *Phys. Rev. E* **70**, 056105 (2004).
- [392] Zhao, L., Lai, Y.-C., Park, K. & Ye, N. Onset of traffic congestion in complex networks. *Phys. Rev. E* **71**, 026125 (2005).
- [393] Wang, W.-X., Wang, B.-H., Yin, C.-Y., Xie, Y.-B. & Zhou, T. Traffic dynamics based on local routing protocol on a scale-free network. *Phys. Rev. E* **73**, 026111 (2006).
- [394] Meloni, S., Gómez-Gardeñes, J., Latora, V. & Moreno, Y. Scaling breakdown in flow fluctuations on complex networks. *Phys. Rev. Lett.* **100**, 208701 (2008).

- [395] Tang, M. & Zhou, T. Efficient routing strategies in scale-free networks with limited bandwidth. *Phys. Rev. E* **84**, 026116 (2011).
- [396] Yang, H.-X., Wang, W.-X., Xie, Y.-B., Lai, Y.-C. & Wang, B.-H. Transportation dynamics on networks of mobile agents. *Phys. Rev. E* **83**, 016102 (2011).
- [397] Morris, R. G. & Barthelemy, M. Transport on coupled spatial networks. *Phys. Rev. Lett.* **109**, 128703 (2012).
- [398] Motter, A. E. & Lai, Y.-C. Cascade-based attacks on complex networks. *Phys. Rev. E* **66**, 065102(R) (2002).
- [399] Zhao, L., Park, K. & Lai, Y.-C. Attack vulnerability of scale-free networks due to cascading breakdown. *Phys. Rev. E* **70**, 035101(R) (2004).
- [400] Zhao, L., Park, K., Lai, Y.-C. & Ye, N. Tolerance of scale-free networks against attack-induced cascades. *Phys. Rev. E* **72**, 025104(R) (2005).
- [401] Galstyan, A. & Cohen, P. Cascading dynamics in modular networks. *Phys. Rev. E* **75**, 036109 (2007).
- [402] Gleeson, J. P. Cascades on correlated and modular random networks. *Phys. Rev. E* **77**, 046117 (2008).
- [403] Huang, L., Lai, Y.-C. & Chen, G. Understanding and preventing cascading breakdown in complex clustered networks. *Phys. Rev. E* **78**, 036116 (2008).
- [404] Yang, R., Wang, W.-X., Lai, Y.-C. & Chen, G. Optimal weighting scheme for suppressing cascades and traffic congestion in complex networks. *Phys. Rev. E* **79**, 026112 (2009).
- [405] Wang, W.-X., Yang, R. & Lai, Y.-C. Cascade of elimination and emergence of pure cooperation in coevolutionary games on networks. *Phys. Rev. E* **81**, 035102(R) (2010).
- [406] Parshani, R., Buldyrev, S. V. & Havlin, S. Interdependent networks: reducing the coupling strength leads to a change from a first to second order percolation transition. *Phys. Rev. Lett.* **105**, 048701 (2010).
- [407] Parshani, R., Buldyrev, S. V. & Havlin, S. Critical effect of dependency groups on the function of networks. *Proc. Natl. Acad. Sci. (USA)* **108**, 10071010 (2011).
- [408] Liu, R.-R., Wang, W.-X., Lai, Y.-C. & Wang, B.-H. Cascading dynamics on random networks: Crossover in phase transition. *Phys. Rev. E* **85**, 026110 (2012).
- [409] Neumann, G., Noda, T. & Kawaoka, Y. Emergence and pandemic potential of swine-origin h1n1 influenza virus. *Nature (London)* **459**, 931–939 (2009).
- [410] Hvistendahl, M., Normile, D. & Cohen, J. Despite large research effort, H7N9 continues to baffle. *Science* **340**, 414–415 (2013).
- [411] Berger, N., Borgs, C., Chayes, J. T. & Saberi, A. On the spread of viruses on the internet. In *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, 301–310 (Society for Industrial and Applied Mathematics, 2005).
- [412] Liu, D. & Chen, X. Rumor propagation in online social networks like twitter—a simulation study. In *Multimedia Information Networking and Security (MINES), 2011 Third Inter. Conf.*, 278–282 (IEEE, 2011).
- [413] Broder, A. *et al.* Graph structure in the web. *Comput. Netw.* **33**, 309–320 (2000).
- [414] Gire, S. K. *et al.* Genomic surveillance elucidates ebola virus origin and transmission during the 2014 outbreak. *Science* **345**, 1369–1372 (2014).
- [415] Shah, D. & Zaman, T. Detecting sources of computer viruses in networks: theory and experiment. In *ACM SIGMETRICS Performance Evaluation Review*, vol. 38, 203–214 (ACM, 2010).

- [416] Luo, W., Tay, W. P. & Leng, M. Identifying infection sources and regions in large networks. *IEEE Trans. Signal Proces.* **61**, 2850–2865 (2013).
- [417] Altarelli, F., Braunstein, A., Dall’Asta, L., Lage-Castellanos, A. & Zecchina, R. Bayesian inference of epidemics on networks via belief propagation. *Phys. Rev. Lett.* **112**, 118701 (2014).
- [418] Brockmann, D. & Helbing, D. The hidden geometry of complex, network-driven contagion phenomena. *Science* **342**, 1337–1342 (2013).
- [419] Zhu, K. & Ying, L. Information source detection in the sir model: A sample path based approach. In *Information Theory and Applications Workshop (ITA), 2013*, 1–9 (IEEE, 2013).
- [420] Shen, Z. *et al.* Locating the source of spreading in complex networks. *arXiv preprint arXiv:1501.06133* (2015).
- [421] Kitsak, M. *et al.* Identification of influential spreaders in complex networks. *Nat. Phys.* **6**, 888–893 (2010).
- [422] Pei, S., Muchnik, L., Andrade Jr, J. S., Zheng, Z. & Makse, H. A. Searching for super-spreaders of information in real-world social media. *Sci. Rep.* **4** (2014).
- [423] Morone, F. & Makse, H. A. Influence maximization in complex networks through optimal percolation. *Nature (London)* (2015).
- [424] Hu, Z.-L., Han, X., Lai, Y.-C. & Wang, W.-X. Optimal localization of diffusion sources in complex networks. *Preprint* (2016).
- [425] Yuan, Z., Zhao, C., Di, Z., Wang, W.-X. & Lai, Y.-C. Exact controllability of complex networks. *Nat. Commun.* **4**, 2447 (2013).
- [426] Kalman, R. On the general theory of control systems. *IRE Trans. Automat. Contr.* **4**, 110–110 (1959).
- [427] Barrat, A., Barthélemy, M. & Vespignani, A. *Dynamical Processes on Complex Networks* (Cambridge University Press, 2008).
- [428] Kumar, A., Rotter, S. & Aertsen, A. Spiking activity propagation in neuronal networks: reconciling different perspectives on neural coding. *Nat. Rev. Neuro.* **11**, 615–627 (2010).
- [429] Pastor-Satorras, R., Castellano, C., Van Mieghem, P. & Vespignani, A. Epidemic processes in complex networks. *Rev. Mod. Phys.* **87**, 925–979 (2015).
- [430] Shao, J., Havlin, S. & Stanley, H. E. Dynamic opinion model and invasion percolation. *Phys. Rev. Lett.* **103**, 018701 (2009).
- [431] Granell, C., Gómez, S. & Arenas, A. Dynamical interplay between awareness and epidemic spreading in multiplex networks. *Phys. Rev. Lett.* **111**, 128701 (2013).
- [432] Santos, F. C. & Pacheco, J. M. Scale-free networks provide a unifying framework for the emergence of cooperation. *Phys. Rev. Lett.* **95**, 098104 (2005).
- [433] Koseska, A., Volkov, E. & Kurths, J. Oscillation quenching mechanisms: Amplitude vs. oscillation death. *Phys. Rep.* **531**, 173–199 (2013).
- [434] Buldyrev, S. V., Parshani, R., Paul, G., Stanley, H. E. & Havlin, S. Catastrophic cascade of failures in interdependent networks. *Nature (London)* **464**, 1025–1028 (2010).
- [435] Galbiati, M., Delpini, D. & Battiston, S. The power to control. *Nat. Phys.* **9**, 126–128 (2013).
- [436] Balcan, D. & Vespignani, A. Phase transitions in contagion processes mediated by recurrent mobility patterns. *Nat. Phys.* **7**, 581–586 (2011).
- [437] Sood, V. & Redner, S. Voter model on heterogeneous graphs. *Phys. Rev. Lett.* **94**, 178701 (2005).

- [438] Pastor-Satorras, R. & Vespignani, A. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* **86**, 3200 (2001).
- [439] Castellano, C., Fortunato, S. & Loreto, V. Statistical physics of social dynamics. *Rev. Mod. Phys.* **81**, 591 (2009).
- [440] Bashan, A., Berezin, Y., Buldyrev, S. V. & Havlin, S. The extreme vulnerability of interdependent spatially embedded networks. *Nat. Phys.* **9**, 667–672 (2013).
- [441] Krapivsky, P. L., Redner, S. & Ben-Naim, E. *A Kinetic View of Statistical Physics* (Cambridge University Press, 2010).
- [442] Santos, F. C., Santos, M. D. & Pacheco, J. M. Social diversity promotes the emergence of cooperation in public goods games. *Nature (London)* **454**, 213–216 (2008).
- [443] Gleeson, J. P. Binary-state dynamics on complex networks: pair approximation and beyond. *Phys. Rev. X* **3**, 021004 (2013).
- [444] Li, J.-W., Wang, W.-X., Lai, Y.-C. & Grebogi, C. Data based reconstruction of complex networks with binary-state dynamics. *Preprint* (2016).
- [445] Hastie, T., Tibshirani, R., Friedman, J. & Franklin, J. The elements of statistical learning: data mining, inference and prediction. *Math. Intell.* **27**, 83–85 (2005).
- [446] Chen, Y.-Z. & Lai, Y.-C. Universal structural estimator and dynamics approximator for complex networks. *Preprint* (2016).
- [447] Bishop, C. M. *Pattern Recognition and Machine Learning* (springer, 2006).
- [448] Russell, S. & Norvig, P. *Artificial Intelligence: A Modern Approach (3rd Edition)* (Prentice Hall, 2009), 3 edn.
- [449] Ackley, D. H., Hinton, G. E. & Sejnowski, T. J. A learning algorithm for Boltzmann machines. *Cogn. Sci.* **9**, 147–169 (2014).
- [450] Tanaka, T. Mean-field theory of Boltzmann machine learning. *Phys. Rev. E* **58**, 2302 (1998).
- [451] Cocco, S. & Monasson, R. Adaptive cluster expansion for inferring Boltzmann machines with noisy data. *Phys. Rev. Lett.* **106**, 090601 (2011).
- [452] Aurell, E. & Ekeberg, M. Inverse Ising inference using all the data. *Phys. Rev. Lett.* **108**, 090201 (2012).
- [453] Nguyen, H. C. & Berg, J. Mean-field theory for the inverse Ising problem at low temperatures. *Phys. Rev. Lett.* **109**, 050602 (2012).
- [454] Ricci-Tersenghi, F. The bethe approximation for solving the inverse Ising problem: a comparison with other inference methods. *J. Stat. Mech.* **2012**, 08015 (2012).
- [455] Roudi, Y. & Hertz, J. Mean field theory for nonequilibrium network reconstruction. *Phys. Rev. Lett.* **106**, 048702 (2011).
- [456] Méard, M. & Sakellariou, J. Exact mean-field inference in asymmetric kinetic Ising systems. *J. Stat. Mech.* **2011**, 08015 (2011).
- [457] Sohl-Dickstein, J., Battaglino, P. B. & DeWeese, M. R. New method for parameter estimation in probabilistic models: Minimum probability flow. *Phys. Rev. Lett.* **107**, 220601 (2011).
- [458] Zhang, P. Inference of kinetic Ising model on sparse graphs. *J. Stat. Phys.* **148**, 502–512 (2012).
- [459] Zeng, H.-L., Alava, M., Aurell, E., Hertz, J. & Roudi, Y. Maximum likelihood reconstruction for Ising models with asynchronous updates. *Phys. Rev. Lett.* **110**, 210601 (2013).
- [460] Barnett, L., Lizier, J. T., Harré, M., Seth, A. K. & Bossomaier, T. Information flow in a kinetic Ising model peaks in the disordered phase. *Phys. Rev. Lett.* **111**, 177203 (2013).



- [461] Decelle, A. & Ricci-Tersenghi, F. Pseudolikelihood decimation algorithm improving the inference of the interaction network in a general class of Ising models. *Phys. Rev. Lett.* **112**, 070603 (2014).
- [462] Ott, E., Grebogi, C. & Yorke, J. A. Controlling chaos. *Phys. Rev. Lett.* **64**, 1196–1199 (1990).
- [463] Boccaletti, S., Grebogi, C., Lai, Y.-C., Mancini, H. & Maza, D. Control of chaos: theory and applications. *Phys. Rep.* **329**, 103–197 (2000).
- [464] Grebogi, C. & Lai, Y.-C. Controlling chaotic dynamical systems. *Syst. Control Lett.* **31**, 307–312 (1997).
- [465] Lombardi, A. & Hörnquist, M. Controllability analysis of networks. *Phys. Rev. E* **75**, 056110 (2007).
- [466] Liu, B., Chu, T., Wang, L. & Xie, G. Controllability of a leader-follower dynamic network with switching topology. *IEEE Trans. Automat. Contr.* **53**, 1009–1013 (2008).
- [467] Rahmani, A., Ji, M., Mesbahi, M. & Egerstedt, M. Controllability of multi-agent systems from a graph-theoretic perspective. *SIAM J. Contr. Optim.* **48**, 162–186 (2009).
- [468] Liu, Y.-Y., Slotine, J.-J. & Barabási, A.-L. Controllability of complex networks. *Nature (London)* **473**, 167–173 (2011).
- [469] Wang, W.-X., Ni, X., Lai, Y.-C. & Grebogi, C. Optimizing controllability of complex networks by small structural perturbations. *Phys. Rev. E* **85**, 026115 (2011).
- [470] Nacher, J. C. & Akutsu, T. Dominating scale-free networks with variable scaling exponent: heterogeneous networks are not difficult to control. *New J. Phys.* **14**, 073005 (2012).
- [471] Yan, G., Ren, J., Lai, Y.-C., Lai, C.-H. & Li, B. Controlling complex networks: How much energy is needed? *Phys. Rev. Lett.* **108**, 218703 (2012).
- [472] Nepusz, T. & Vicsek, T. Controlling edge dynamics in complex networks. *Nat. Phys.* **8**, 568–573 (2012).
- [473] Liu, Y.-Y., Slotine, J.-J. & Barabási, A.-L. Observability of complex systems. *Proc. Natl. Acad. Sci. (USA)* **110**, 2460–2465 (2013).
- [474] Menichetti, G., Dall’Asta, L. & Bianconi, G. Network controllability is determined by the density of low in-degree and out-degree nodes. *Phys. Rev. Lett.* **113**, 078701 (2014).
- [475] Ruths, J. & Ruths, D. Control profiles of complex networks. *Science* **343**, 1373–1376 (2014).
- [476] Wuchty, S. Controllability in protein interaction networks. *Proc. Natl. Acad. Sci. (USA)* **111**, 7156–7160 (2014).
- [477] Whalen, A. J., Brennan, S. N., Sauer, T. D. & Schiff, S. J. Observability and controllability of nonlinear networks: The role of symmetry. *Phys. Rev. X* **5**, 011005 (2015).
- [478] Yan, G. *et al.* Spectrum of controlling and observing complex networks. *Nat. Phys.* **11**, 779–786 (2015).
- [479] Chen, Y.-Z., Wang, L.-Z., Wang, W.-X. & Lai, Y.-C. The paradox of controlling complex networks: control inputs versus energy requirement. *arXiv:1509.03196v1* (2015).
- [480] Kalman, R. E. Mathematical description of linear dynamical systems. *J. Soc. Indus. Appl. Math. Ser. A* **1**, 152–192 (1963).
- [481] Lin, C.-T. Structural controllability. *IEEE Trans. Automat. Contr.* **19**, 201–208 (1974).
- [482] Luenberger, D. G. *Introduction to Dynamical Systems: Theory, Models, and Applications* (John Wiley & Sons, Inc, New Jersey, 1999), first edn.
- [483] Slotine, J.-J. E. & Li, W. *Applied Nonlinear Control* (Prentice-Hall, New Jersey, 1991).

- [484] Chen, Y.-Z., Huang, Z.-G. & Lai, Y.-C. Controlling extreme events on complex networks. *Sci. Rep.* **4**, 6121 (2014).
- [485] Wang, L.-Z. *et al.* Control and controllability of nonlinear dynamical networks: a geometrical approach. *arXiv:1509.07038* (2015).
- [486] Grebogi, C., McDonald, S. W., Ott, E. & Yorke, J. A. Final state sensitivity: an obstruction to predictability. *Phys. Lett. A* **99**, 415–418 (1983).
- [487] McDonald, S. W., Grebogi, C., Ott, E. & Yorke, J. A. Fractal basin boundaries. *Physica D* **17**, 125–153 (1985).
- [488] Feudel, U. & Grebogi, C. Multistability and the control of complexity. *Chaos* **7**, 597–604 (1997).
- [489] Feudel, U. & Grebogi, C. Why are chaotic attractors rare in multistable systems? *Phys. Rev. Lett.* **91**, 134102 (2003).
- [490] Ni, X., Ying, L., Lai, Y.-C., Do, Y.-H. & Grebogi, C. Complex dynamics in nanosystems. *Physical Review E* **87**, 052911 (2013).
- [491] May, R. M. Thresholds and breakpoints in ecosystems with a multiplicity of stable states. *Nature (London)* **269**, 471–477 (1977).
- [492] Alley, R. B. *et al.* Abrupt climate change. *Science* **299**, 2005–2010 (2003).
- [493] Chase, J. M. Experimental evidence for alternative stable equilibria in a benthic pond food web. *Ecol. Lett.* **6**, 733–741 (2003).
- [494] Schröder, A., Persson, L. & De Roos, A. M. Direct experimental evidence for alternative stable states: a review. *Oikos* **110**, 3–19 (2005).
- [495] Badzey, R. L. & Mohanty, P. Coherent signal amplification in bistable nanomechanical oscillators by stochastic resonance. *Nature (London)* **437**, 995–998 (2005).
- [496] Huang, S., Guo, Y.-P., May, G. & Enver, T. Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Developmental Biol.* **305**, 695–713 (2007).
- [497] Huang, S. Genetic and non-genetic instability in tumor progression: link between the fitness landscape and the epigenetic landscape of cancer cells. *Cancer Metastasis Rev.* **32**, 423–448 (2013).
- [498] Huang, S., Eichler, G., Bar-Yam, Y. & Ingber, D. E. Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Phys. Rev. Lett.* **94**, 128701 (2005).
- [499] Süel, G. M., Garcia-Ojalvo, J., Liberman, L. M. & Elowitz, M. B. An excitable gene regulatory circuit induces transient cellular differentiation. *Nature (London)* **440**, 545–550 (2006).
- [500] Furusawa, C. & Kaneko, K. A dynamical-systems view of stem cell biology. *Science* **338**, 215–217 (2012).
- [501] Li, X., Zhang, K. & Wang, J. Exploring the mechanisms of differentiation, dedifferentiation, reprogramming and transdifferentiation. *PLOS ONE* **9**, e105216 (2014).
- [502] Battogtokh, D. & Tyson, J. J. Bifurcation analysis of a model of the budding yeast cell cycle. *Chaos* **14**, 653–661 (2004).
- [503] Yao, G., Tan, C., West, M., Nevins, J. & You, L. Origin of bistability underlying mammalian cell cycle entry. *Mol. Syst. Biol.* **7** (2011).
- [504] Radmaneshfar, E. *et al.* From start to finish: the influence of osmotic stress on the cell cycle. *PLOS ONE* **8**, e68067 (2013).
- [505] Kauffman, S., Peterson, C., Samuelsson, B. & Troein, C. Genetic networks with canalizing boolean rules are always stable. *Proc. Natl. Acad. Sci. (USA)* **101**, 17102–17107 (2004).

- [506] Greil, F. & Drossel, B. Dynamics of critical Kauffman networks under asynchronous stochastic update. *Phys. Rev. Lett.* **95**, 048701 (2005).
- [507] Motter, A., Gulbahce, N., Almaas, E. & Barabási, A.-L. Predicting synthetic rescues in metabolic networks. *Mol. Sys. Biol.* **4** (2008).
- [508] Ma, W., Trusina, A., El-Samad, H., Lim, W. A. & Tang, C. Defining network topologies that can achieve biochemical adaptation. *Cell* **138**, 760–773 (2009).
- [509] Faucon, P. C. *et al.* Gene networks of fully connected triads with complete auto-activation enable multistability and stepwise stochastic transitions. *PLOS ONE* **9**, e102873 (2014).
- [510] Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in *Escherichia coli*. *Nature (London)* **403**, 339–342 (2000).
- [511] Wu, M. *et al.* Engineering of regulated stochastic cell fate determination. *Proc. Natl. Acad. Sci. (USA)* **110**, 10610–10615 (2013).
- [512] Wu, F., Menn, D. & Wang, X. Quorum-sensing crosstalk-driven synthetic circuits: From unimodality to trimodality. *Chem. Biol.* **21**, 1629–1638 (2014).
- [513] Lai, Y.-C. Controlling complex, nonlinear dynamical networks. *Nat. Sci. Rev.* **1**, 339–341 (2014).
- [514] Menck, P. J., Heitzig, J., Kurths, J. & Schellnhuber, H. J. How dead ends undermine power grid stability. *Nat. Comm.* **5** (2014).
- [515] Campbell, C. & Albert, R. Stabilization of perturbed boolean network attractors through compensatory interactions. *BMC Syst. Biol.* **8**, 53 (2014).
- [516] Wang, P. *et al.* Epigenetic state network approach for describing cell phenotypic transitions. *Interface Focus* **4**, 20130068 (2014).