Efficient reconstruction of directed networks from noisy dynamics using stochastic force inference

Chi-Ho Cheng (鄭智豪) [*] and Pik-Yin Lai (黎璧賢) [†]

Department of Physics, National Changhua University of Education, Changhua 500, Taiwan, Republic of China
and Department of Physics and Center for Complex Systems, National Central University, Chung-Li District, Taoyuan City 320, Taiwan,
Republic of China

We consider coupled network dynamics under uncorrelated noises that fluctuate about the noise-free long-time asymptotic state. Our goal is to reconstruct the directed network only from the time-series data of the dynamics of the nodes. By using the stochastic force inference method with a simple natural choice of linear polynomial basis, we derive a reconstruction scheme of the connection weights and the noise strength of each node. Explicit simulations for directed and undirected random networks with various node dynamics are carried out to demonstrate the good accuracy and high efficiency of the reconstruction scheme. We further consider the case when only a subset of the network and its node dynamics can be observed, and it is demonstrated that the directed weighted connections among the observed nodes can be easily and faithfully reconstructed. In addition, we propose a scheme to infer the number of hidden nodes and their effects on each observed node. The accuracy of these results is illustrated by simulations.

I. INTRODUCTION

The past decades have marked the rise of the big data era [1] with immense increase in the available recording data in a wide range of scientific, engineering, social, and financial areas. This is partly due to the huge increase and reliability in the capacity in the storing devices and partly due to the abundance and convenient availability of these data accessed through the internet. These technologies have enhanced the research activities of data science and complex networks [2–4], especially in the fields of biology and medicine such as systems biology and the associated protein interaction networks [5], metabolism network, food web, and gene regulatory networks. We are in a stage in which the data accumulation rate exceeds tremendously the data to knowledge digestion rate. One of the major ultimate goals in science is to achieve a logical understanding of the phenomenon and hopefully to uncover the basic principles behind and be able to make prediction for the system. Machine learning may be very helpful in summarizing these vast data volumes and classifying them into patterns or structures, but a knowledge-based approach to the analysis of the data can do a much better job in understanding the reasoning behind the data and modeling the system or discovering scientific laws buried in the vast data mines.

The relations between different entities in the data set can often be described in terms of a network consisting of nodes and links. It is a highly nontrivial task to construct the connection topology and link strengths of a network that can faithfully represent the underlying interactions among the network elements. Often the rules for assigning a connection and its strengths between two nodes lack rigorous scientific reasons or are not well justified. Often the connections or their strengths are not known, but the behavior of the nodes can be monitored or measured. One challenging inverse problem is to retrieve the network connection weights or node-node interactions from the passively observed dynamics of the nodes which are often accessible or measurable. In the network reconstruction problems developed in recent years [6–8], with the help of noise, which accounts for external disturbances leading to fluctuating dynamics in a network, it is possible to identify the network connectivity solely from the time-series dynamics of the nodes [9–11]. This noise-bridging approach, among many other methods, is part of a broader research area on the network reconstruction problem [12].

Although it has been demonstrated in simulations that the network reconstruction from noisy node dynamics is rather successful, there are still some problems with current reconstruction methods that must be overcome in practice. For example, the reconstructed connection weight can become a complex number if the quality of the time-series data is not very high. Also some reconstruction methods need to compute the higher-order derivatives of the time-series data and will demand the sampling rate and quality of data to be very high. Another issue in practice is the problem of hidden nodes and/or connections because often not all the nodes in a network can be accessed in reality. Although there was attempt to reconstruct undirected network with hidden nodes [13], the problem of reconstructing a directed network with hidden nodes remains unsolved. On the fundamental side, often the dynamical model of a complex network can be quite arbitrary or artificial; the relationship between the associated fluctuating noisy dynamics and possible physical nonequilibrium systems is still unclear and largely unexplored.

---

[*]phcch@cc.ncue.edu.tw
[†]pylai@phy.ncu.edu.tw

On the other hand, for physical systems such as Brownian dynamics, one aims to uncover the underlying governing equation of motion from the observed time-series data of the trajectories. The recent inverse problem of reconstructing the force field and the noise from an analysis of experimental noisy dynamics is of immense interest because of the advances in experimental recording techniques that lead to the availability of reliable observational data in complex interacting systems of many degrees of freedom. For example, in systems described by the overdamped Langevin equation, recently developed techniques make it possible to efficiently reconstruct the dynamics from observed stochastic trajectories [14–16]. It was recently shown, using communication theory of information capacity, that such trajectories contain information which can be exploited to fit the force field with a linear combination of some appropriate basis functions. Such a stochastic force inference (SFI) approach [16] provides a powerful and data-efficient solution to reconstructing the force field from the fluctuating Brownian noisy dynamics, which could be applied to the noisy network dynamics in the network reconstruction problem here.

In this paper, we implement the SFI on the network reconstruction problem and devise a scheme to reconstruct the connection weights and the noise variances acting on the nodes of a directed network. The validity of our theoretical results will be checked with numerical simulations for known network structures. We also demonstrate that the SFI scheme works well even for reconstructing directed networks with hidden nodes as well as for higher-dimensional intrinsic node dynamics but with only a single component of intrinsic node dynamics observed. Furthermore, we propose a practical scheme to retrieve the number of hidden nodes and the effect of the hidden connections on an individual observed node. The theoretical results are further verified by carrying out explicit simulations on directed and undirected random weighted networks.

II. FLUCTUATING DYNAMICS OF COUPLED NOISY NETWORKS

Consider a network with N nodes whose intrinsic one-dimensional dynamics at time t, $x_i(t)$, is governed by the nonlinear function $f_i(x_i)$. Nodes i and j are connected by directed and weighted links given by the matrix $\mathbf{W}$ and there is no self-connection ($W_{ii} = 0$). In many situations, the weighted matrix takes the form $W_{ij} = g_{ij} A_{ij}$, with the adjacency matrix of elements $A_{ij} = 0$ or 1, and connection weights g_{ij}. We use an overdot to denote time derivative and prime for derivative with respect to the network (spatial) variable x. Assuming the nodes are subjected to temporally uncorrelated noises, the equation of motion of the network dynamics is

$$\dot{x}_i = f_i(x_i) + \sum_{j \neq i}^{N} W_{ij} h(x_i, x_j) + \eta_i(t), \qquad (1)$$

where η_i is zero-mean Gaussian white noise with (spatial) correlation σ_{ij} that acts on the node i,

$$\overline{\eta_i(t)} = 0, \qquad \overline{\eta_i(t)\eta_j(t')} = \sigma_{ij}\delta(t - t'), \qquad (2)$$

where h is the coupling function. The overbar stands for ensemble average over the noise, which can be obtained in practice by a time average over the asymptotic dynamics over an extended period of time, denoted by $\langle \cdots \rangle$. In the presence of noises, the system approaches some asymptotic dynamics and $x_i(t)$ fluctuates around the noise-free solution X_i. We assume that the noisy dynamics fluctuates around some stable noise-free solution, and that f and h can be linearized about the noise-free solution. Denote the deviation from X_i by $\delta x_i \equiv x_i - X_i$; then one has

$$\delta\dot{\vec{x}} = \mathbf{Q}\delta\vec{x} + \vec{\eta}(t), \qquad (3)$$

$$Q_{ij} \equiv W_{ij}\partial_2 h(X_i, X_j) + \left[f_i'(X_i) + \sum_{m=1}^{N} W_{im}\partial_1 h(X_i, X_m) \right]\delta_{ij}, \qquad (4)$$

where $\partial_1 h$ and $\partial_2 h$ denote the partial derivatives with respect to the first and second independent variables in h, respectively. We take h to be of the form $h(x_1, x_2) = h(x_2 - x_1)$ with $h(-z) = -h(z)$ and $h'(0) > 0$; i.e., the coupling tends to synchronize the dynamics of the nodes. For convenience, we will take $h'(0) = 1$ (or absorbed into the definition of W_{ij}) hereafter.

The network reconstruction scheme for networks under white noise using the time-lag correlations is summarized in Appendix A [see Eqs. (4) and (A4) for the reconstruction formula].

III. NETWORK RECONSTRUCTION USING SFI

For the fluctuating dynamics about the stationary noise-free fixed point $\vec{X}$ described by Eq. (3), it is natural to choose an order-1 (linear) polynomial basis, i.e., $n_b = N + 1$ with the set of basis functions $b = \{b_\alpha(\vec{x})\}_{\alpha=1,\ldots,n_b} = \{1, x_1, x_2, \ldots, x_N\}$ for SFI for the force field $\vec{F}(\vec{x}) = \mathbf{Q}(\vec{x} - \vec{X})$. To carry out the SFI from the recorded time-series data of length N_{step}, at times t_k, $k = 1, 2, \ldots, N_{step}$, one computes the average over the time-series data sampled separated with time intervals $\Delta t_k \equiv t_{k+1} - t_k$, denoted by $\langle \cdots \rangle$. To reconstruct the connection and noise matrices, one only needs to compute the following averages or correlation functions:

$$B_{\alpha\beta} \equiv \langle b_\alpha(\vec{x})b_\beta(\vec{x}) \rangle \equiv \frac{1}{N_{step}}\sum_{k=1}^{N_{step}} b_\alpha(\vec{x}(t_k))b_\beta(\vec{x}(t_k)), \quad (5)$$

$$M_{j\beta} \equiv \left\langle \frac{\Delta x_j}{\Delta t} b_\beta \right\rangle \equiv \frac{1}{N_{step}-1}\sum_{k=1}^{N_{step}-1} \frac{\Delta x_j(t_k)}{\Delta t_k} b_\beta(\vec{x}(t_k)), \quad (6)$$

$$\left\langle \frac{\Delta x_i \Delta x_j}{\Delta t} b_\beta \right\rangle \equiv \frac{1}{N_{step}-1}\sum_{k=1}^{N_{step}-1} \frac{\Delta x_i(t_k)\Delta x_j(t_k)}{\Delta t_k} b_\beta(\vec{x}(t_k)), \qquad (7)$$

$$\langle b_\alpha \rangle \equiv \frac{1}{N_{step}}\sum_{k=1}^{N_{step}} b_\alpha(\vec{x}(t_k)), \quad i, j = 1, 2, \ldots, N,$$

$$\alpha, \beta = 1, 2, \ldots, N + 1, \qquad (8)$$

where $\Delta x_j(t_k) \equiv x_j(t_{k+1}) - x_j(t_k)$. Hence from the SFI scheme (see Appendix B), we have the reconstructed **Q** and $\boldsymbol{\sigma}$:

$$Q_{ij} = (\mathbf{MB}^{-1})_{ij+1}, \tag{9}$$

$$\sigma_{ij} = \sum_{\beta,\gamma}^{n_b} \left\langle \frac{\Delta x_i \Delta x_j}{\Delta t} b_\beta \right\rangle B_{\beta\gamma}^{-1} \langle b_\gamma \rangle, \tag{10}$$

$$(\mathbf{Q}\vec{X})_i = -(\mathbf{MB}^{-1})_{i1}. \tag{11}$$

The above SFI network reconstruction formulas will be tested by simulations with time-series data generated by known intrinsic node dynamics with nodes acted on by white noises and connected by a known connection matrix.

A. Reconstruction of undirected and directed network connection weights and noise strengths

To examine the validity of the proposed reconstruction method, we carry out simulations by generating networks with a total of N nodes. Specifically, we generate bidirectional and directed weighted random (denoted by BWR and DWR, respectively) Erdős-Rényi (ER) networks [17,18] with edge connection probability p. The weights of the edges are randomly chosen from a Gaussian distribution of mean μ_W and standard deviation σ_W. The nodes of the entire network are subjected to white noise given by the matrix $\boldsymbol{\sigma}$, and we take σ_{ij} to be diagonal with the elements σ_{ii} uniformly distributed in $[\sigma^2/2, \sigma^2]$; i.e., each node is subjected to a noise of different noise strengths. The intrinsic dynamics of each node is heterogeneous of the logistic form $f(x_i) = r_i x_i(1 - x_i)$. In most cases, we take $h(z) = z$ unless otherwise stated. In the simulations, the values $\mu_W = 2$, $\sigma_W = 1$ or $\mu_W = 10$, $\sigma_W = 2$ are used in the edge weights so that there are positive and negative weights, r_i is randomly chosen in [8,12], and a noise strength $\sigma^2 = 1$ or 0.1 is used. In most cases, we simulate networks of sizes $N = 100$ and 200 for several values of p. The network stochastic dynamics are solved by the Euler method with a time step of 5×10^{-4} for a long time (typically 2×10^3 to 10^4). And for the measurement of various statistical averages they are sampled with a time interval of $\Delta t = 10^{-4}$ to 10^{-3}. It should be noted that the SFI reconstruction involves the inversion of the matrix **B** that may not be possible if it is rank deficient, which may occur if the time series is too short. We tested the minimal length of the time series such that the inversion of the matrix **B** has a problem. We found that, for all the cases we looked into, the minimal length is rather short; for example, in the logistic or FitzHugh-Nagumo (FHN) node dynamics case, the minimal time-series length is less than 0.05. The minimal length for **B** to be noninvertible is much less than the time-series lengths for reasonable reconstruction accuracy, which can be estimated using Eqs. (B13) or (B14).

Figure 1 shows the reconstruction results using SFI against the corresponding actual values for weighted random BWR and DWR networks with connection probability $p = 0.2$ and $N = 100$, with a coarsened sampling interval of $\Delta t = 2\delta t = 10^{-3}$. The SFI reconstructed connection weights between different nodes, $W_{ij} = Q_{ij}(i \neq j)$ as given by Eq. (9), show good agreement with the actual values for both the BWR and DWR networks as shown in Fig. 1(a). The reconstructed diagonal

elements Q_{ii} [which are theoretically given by $r_i - \sum_k^N W_{ik}$ from Eq. (4)] also agree very well with the predicted values [Fig. 1(b)]. The reconstructed noise matrix (diagonal) elements as given by Eq. (10) show perfect agreement with the actual heterogeneous noise variance on the nodes [Fig. 1(c)]. Finally, as a consistency check, the components of the vector $\mathbf{Q}\vec{X}$ are plotted against the prediction from the SFI scheme in Eq. (11) as displayed in Fig. 1(d), showing perfect agreement. The accuracy of the reconstruction can be accessed by evaluating the root-mean-square error of the reconstructed W_{ij}. Figure 2 shows the root-mean-square error, δ_W, as a function of Δ and the length of the time series (τ) for the BWR and DWR networks. δ_W increases with sampling interval Δt for a fixed total time-series data length as shown in Fig. 2(a). On the other hand, δ_W decreases with τ for a fixed Δt, and saturates to a value limited by Δt [Fig. 2(b)]. Notice that δ_W is small compared to the mean of the connection weights ($\mu_W = 10$) if the data length is sufficiently long and the sampling frequency is relatively high.

B. Network reconstruction of higher-dimensional and nonstationary dynamics

We use the FitzHugh-Nagumo model for the two-dimensional node dynamics with the fast and slow variables u and v, respectively. We consider coupled networks of the same type of dynamical variables; i.e., u_i's are coupled among themselves with the network $\mathbf{W}^{(u)}$ and v_i's are coupled with $\mathbf{W}^{(v)}$. The dynamics of the entire system is given by

$$\dot{u}_i = \frac{1}{\epsilon}\left(u_i - \frac{u_i^3}{3} - v_i\right) + \sum_{j \neq i}^N W_{ij}^{(u)} h(u_i, u_j) + \eta_i^{(u)}(t), \tag{12}$$

$$\dot{v}_i = u_i + \alpha_i + \sum_{j \neq i}^N W_{ij}^{(v)} h(v_i, v_j) + \eta_i^{(v)}(t), \tag{13}$$

where ϵ controls the difference in time scales between the fast and slow variable and α_i controls the intrinsic excitability of the ith isolated node: the node is excitable (oscillatory) if $\alpha > 1$ ($\alpha < 1$). $\eta_i^{(u)}(t)$ and $\eta_i^{(v)}(t)$ are the independent zero-mean white noises that act on u_i and v_i, respectively. If both the time series of $u_i(t)$ and $v_i(t)$ are recorded, then one just simply adopts SFI reconstruction for the dynamics in $2N$ phase space and employs the basis $b = \{1, u_1, u_2, \ldots, u_N, v_1, v_2, \ldots, v_N\}$ to reconstruct the networks $\mathbf{W}^{(u)}$ and $\mathbf{W}^{(v)}$ together with the noises as given by Eqs. (9) and (10). However, in many practical situations, often only the dynamics of one component of the intrinsic node dynamics (say, v_i) is recorded and the rest are not (the u_i are "hidden" degrees of freedom). Remarkably, the SFI scheme can still faithfully reconstruct the network and the noise information of the observed degree of freedom. To demonstrate this, we generate BWR and DWR networks with FHN node dynamics with only the time-series data $v_i(t)$ are used for reconstruction. Figure 3 displays the SFI reconstruction results for the network [$W_{ij}^{(v)}$, Fig. 3(a)] and noise strengths [$\sigma_{ii}^{(v)}$, Fig. 3(c)] of the observed degree of freedom, showing faithful reconstruction performance. In addition the diagonal elements Q_{ii} and the prediction (11) for the observed
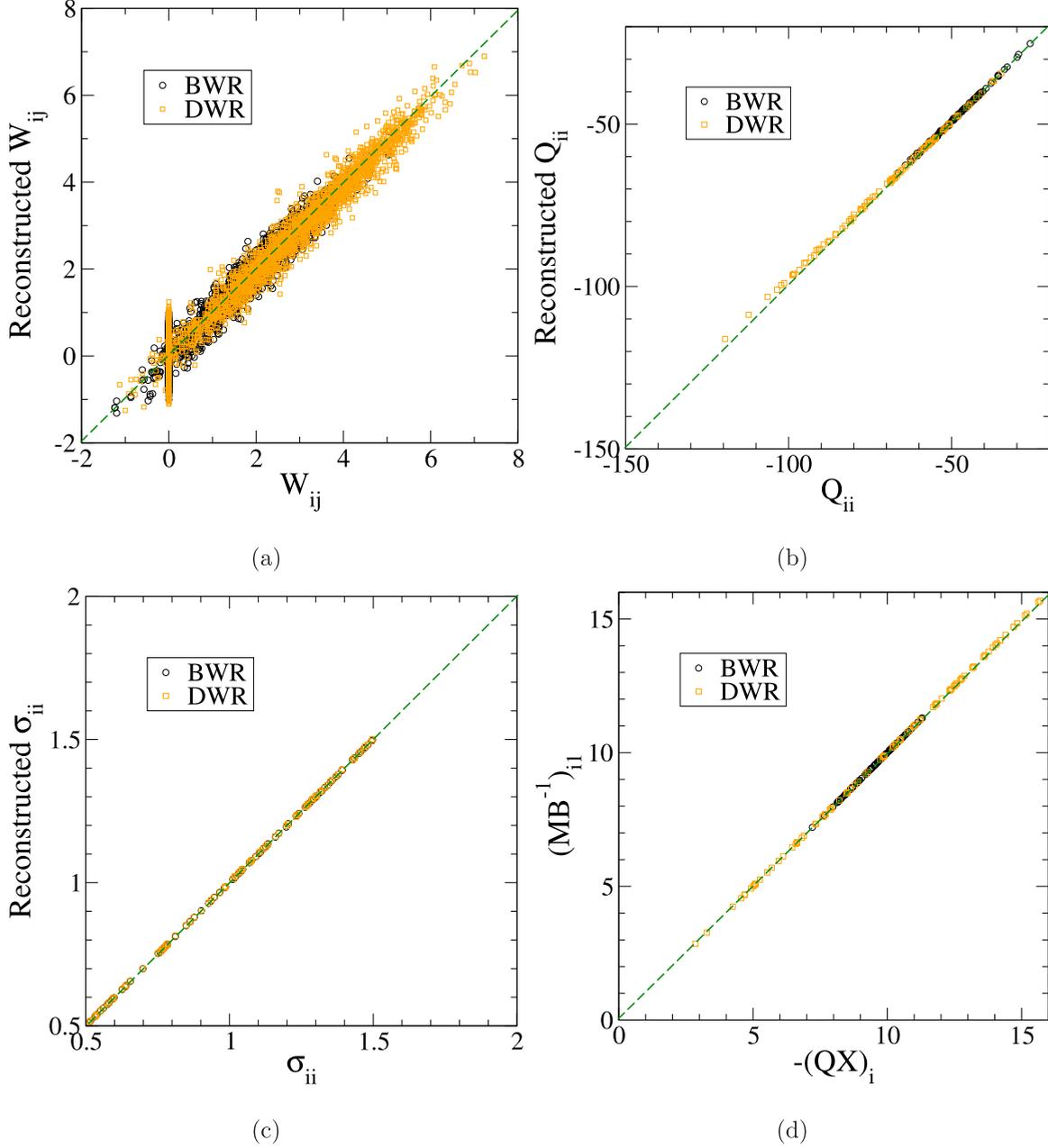
FIG. 1. Network reconstruction of weighted random bidirectional (BWR) and directed (DWR) networks using the SFI method. Sampling interval $\Delta t = 2\delta t = 10^{-3}$ is used. Noisy network dynamics is generated with logistic node dynamics and Gaussian distributed connection weights with $\mu_W = 2$, $\sigma_W = 1$ and noise variance on nodes uniformly distributed in [0.5,1.5]. The total duration of the time-series data is 2000. The dashed straight line marks the $y = x$ line. (a) The reconstructed connection weights between different nodes vs the actual ones. (b) The reconstructed diagonal elements of $\mathbf{Q}$ vs the actual ones. (c) The reconstructed noise variance on each node vs the actual ones. (d) $(\mathbf{MB}^{-1})_{i1}$ plotted against $(\mathbf{Q}\vec{X})_i$, verifying the SFI reconstruction scheme in Eq. (11).

degree of freedom are also well verified. It is worth noting that one needs to have better statistics (longer data length or higher sampling frequency) in order to achieve a similar reconstruction accuracy as the case of no hidden degree of freedom (as in Fig. 1). The success of the reconstruction of the network of the observed degree of freedom regardless of the hidden dynamics can in fact be understood from the theoretical framework of SFI [16]: the fact that some degrees of freedom are hidden is completely equivalent to having a projection basis in which the hidden degrees of freedom are missing and SFI can still

capture the dynamics projected on the observed degrees of freedom with the hidden degrees of freedom being averaged over. Hence the network and noises of the observed degrees of freedom can be accurately reconstructed by SFI.

The two-dimensional and nonlinear intrinsic dynamics for each node allows the possibility for limit cycle periodic dynamics and hence one can explore the SFI reconstruction scheme for noisy network dynamics fluctuating about a nonsteady noise-free state. By setting the parameters $\alpha_i < 1$, the noise-free dynamics undergoes nonlinear periodic
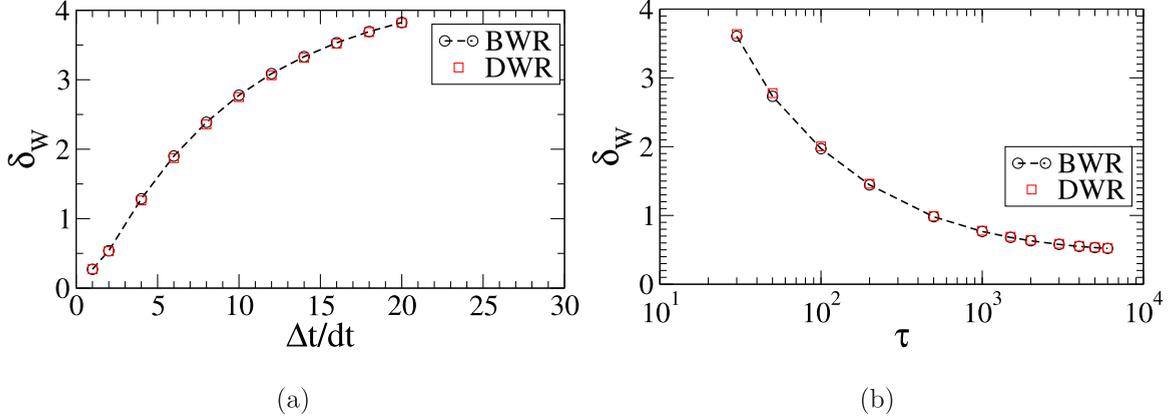
FIG. 2. The root-mean-square error of reconstructed W_{ij} (a) as a function of the normalized Δt for $\tau = 6000$, and (b) as a function of the total data time length τ for $\Delta t = 2\delta t = 10^{-3}$, for weighted random bidirectional (BWR) and directed (DWR) networks of $\mu_W = 10$ and $\sigma_W = 2$, with logistic node dynamics using the SFI method.

oscillations. We carry out SFI reconstruction for the observed dynamics of $v_i(t)$ but with u_i being hidden, similar to the case in Fig. 3. Figure 4 displays the SFI reconstruction results for the network [$W_{ij}^{(v)}$, Fig. 4(a)] and noise strengths [$\sigma_{ii}^{(v)}$, Fig. 4(c)] of the observed degree of freedom, showing faithful reconstruction performance. The diagonal elements Q_{ii} and the prediction (11) for the observed degree of freedom are also well verified. The reason why the SFI network reconstruction works well for the time-dependent noise-free state $\vec{X}(t)$ is presumably because the reconstructed force captures the projected observed dynamics, effectively averaging over $\vec{X}(t)$ which can be thought of also as hidden degrees of freedom.

The fact that SFI can capture the dynamics projected on the observed degrees of freedom suggests that one can accurately reconstruct the noises and network among the observed degrees of freedom, regardless of the network and noises of the hidden variables. To verify this, we generate noisy FHN dynamics with both u_i and v_i coupled by their own (but different) networks, and this time we choose u_i to be the observed variable and v_i are hidden ones. Figure 5 displays the SFI reconstruction results for the observed network [$W_{ij}^{(u)}$, Fig. 5(a)], the associated diagonal elements [$Q_{ii}^{(u)}$, Fig. 5(b)], and the noise strengths [$\sigma_{ii}^{(u)}$, Fig. 5(c)] of the observed nodes, showing accurate reconstructions.

We further look into more complex network dynamics allowed by higher-dimensional intrinsic node dynamics. We consider the node dynamics to be the Rössler dynamics of three-dimensional state variables with nonlinear coupling:

$$\dot{x}_i = -y_i - z_i + \sum_{j \neq i} W_{ij} h(x_i, x_j) + \eta_i(t), \quad (14)$$

$$\dot{y}_i = x_i + a y_i + \sum_{j \neq i} W_{ij} h(y_i, y_j), \quad (15)$$

$$\dot{z}_i = b + z_i(x_i - c) + \sum_{j \neq i} W_{ij} h(z_i, z_j). \quad (16)$$

Here the coupling networks and functions for the three variables are chosen to be the same for simplicity. And the parameters are chosen with $a = b = 0.2$ and $c = 9$ such that the intrinsic dynamics is chaotic without the network coupling.

We take the simple consensus coupling function $h(x_1, x_2) = x_2 - x_1$, and the noise-free synchronized state with the chosen parameter is also chaotic. Only the variables $x_i(t)$ are the observed time series with y_i and z_i being hidden degrees of freedom. The SFI reconstruction results for the BWR and DWR networks are displayed in Fig. 6. The weights of the observed network [Fig. 6(a)], the associated diagonal elements Q_{ii} [Fig. 6(b)], and the noise strengths [Fig. 6(c)] of the observed degree of freedom can be faithful reconstructed.

C. Performance comparison with the time-lag correlation method

To demonstrate the superiority of the SFI reconstruction method, we compare its performance with the time-lag correlation reconstruction method (summarized in Appendix A) in terms of the root-mean-square (rms) errors for various node dynamics, as shown in Table I. For a fair comparison, no preprocessing (such as a smoothing filter) of the time series is carried out, and the same raw time series generated from the same bidirectional random network are used for both methods. The rms errors of the reconstructed network connection weights (δ_W) and that of the reconstructed noise matrix elements (δ_σ) are measured for logistic and various FHN node dynamics of different time-series lengths (τ). For shorter time series, the time-lag correlation method suffered from the unphysical complex constructed weights (denoted by $*$ in the table), whereas the SFI method can still produce reconstruction with decent accuracy. In all cases, the SFI method has a significantly smaller δ_W. The SFI reconstruction of the noise matrix elements is particularly accurate [which can be observed in results in previous sections such as in Fig. 1(c)] whose δ_σ is more than ten times smaller than that of the time-lag correlation method.

IV. NETWORK RECONSTRUCTION WITH HIDDEN NODES

In many realistic situations, it is often that not all nodes and their associated connections can be observed. The behavior (dynamics) of the observed network is strongly influenced by
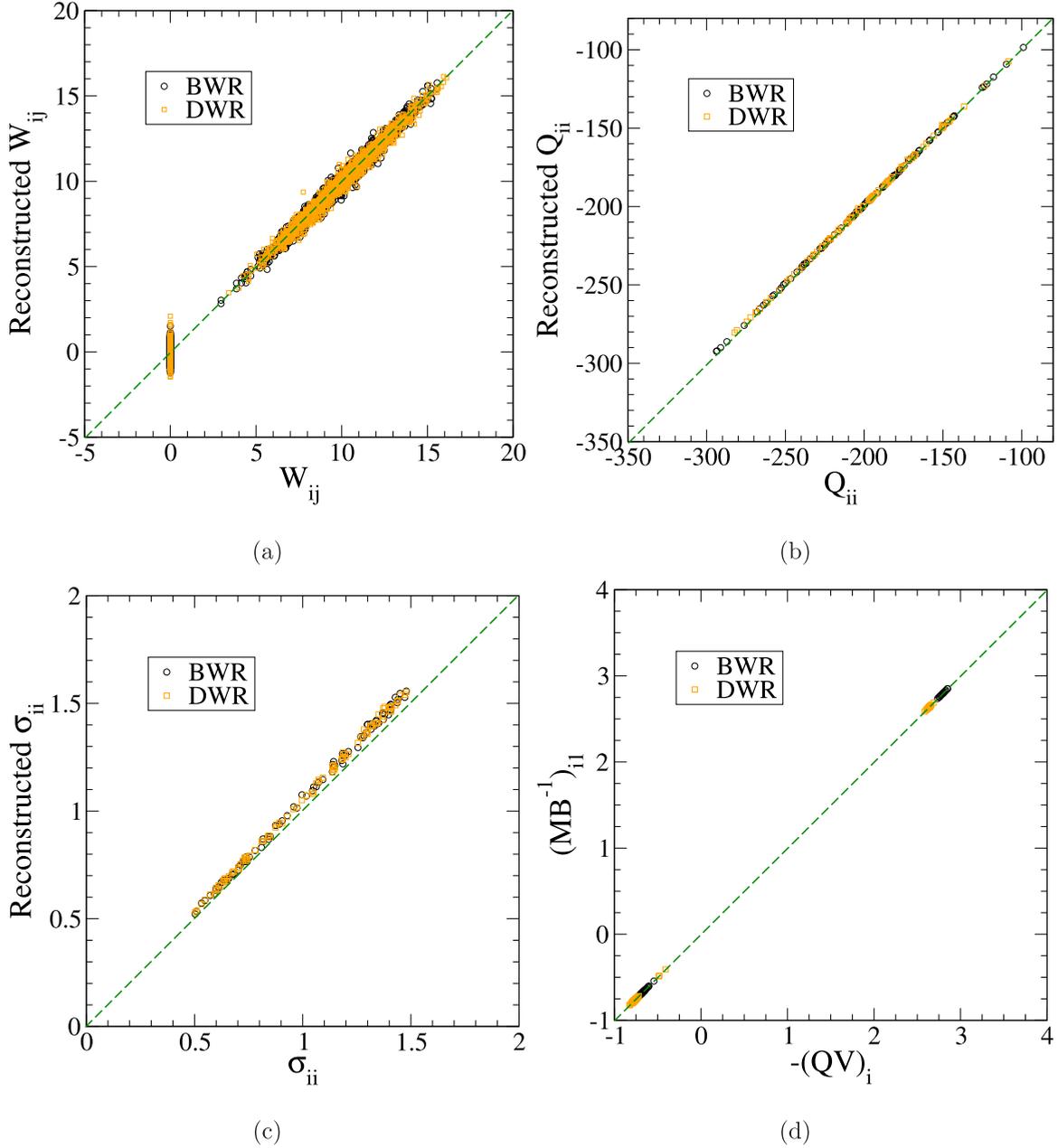
(a)

(b)

(c)

(d)

FIG. 3. Network reconstruction of weighted random bidirectional (BWR) and directed (DWR) networks with FHN dynamics (with $\alpha = 1.05$) using the SFI method. Only v_i are coupled by the network with Gaussian distributed connection weights with $\mu_W = 10$, $\sigma_W = 2$ and $h(x, y) = \tanh(y - x)$. Noise variances on nodes are uniformly distributed in [0.5,1.5]. Only the time-series dynamics $v_i(t)$ are observed. Total time of the series is $\tau = 6000$. $\Delta t = 5 \times 10^{-4}$. (a) The reconstructed connection weights between different nodes vs the actual ones. (b) The reconstructed connection weights between different nodes vs the actual ones. (c) The reconstructed noise variance on each nodes vs the actual ones. (d) $(\mathbf{MB}^{-1})_{i1}$ plotted against $-(\mathbf{Q}\bar{\bar{V}})_i$, verifying the SFI reconstruction scheme in Eq. (11).

the hidden nodes through their links to the observed nodes [19]. Thus reconstructing the network connections among the observed links solely from the dynamics of the observed nodes with the presence of hidden nodes and links is a challenging problem. And so far there is no satisfactory method to achieve the reconstruction of the observed network with hidden nodes for the case of directed networks. It has been recently demonstrated that the effect of the hidden information can be viewed as temporally and spatially correlated noises acting on the observed nodes [19]. Such correlations are due

to the hidden nodes that were interacting with the observed ones. Furthermore, deducing the information on the entire network with only observations from the partial subnetwork of the observed nodes and links is even more challenging, and constitutes the so-called network completion problem [20–22].

The present framework of SFI network reconstruction can be readily extended to the situations where some of the nodes and their connections are not observed, i.e., reconstructing the network of the observed nodes in the presence of hidden nodes
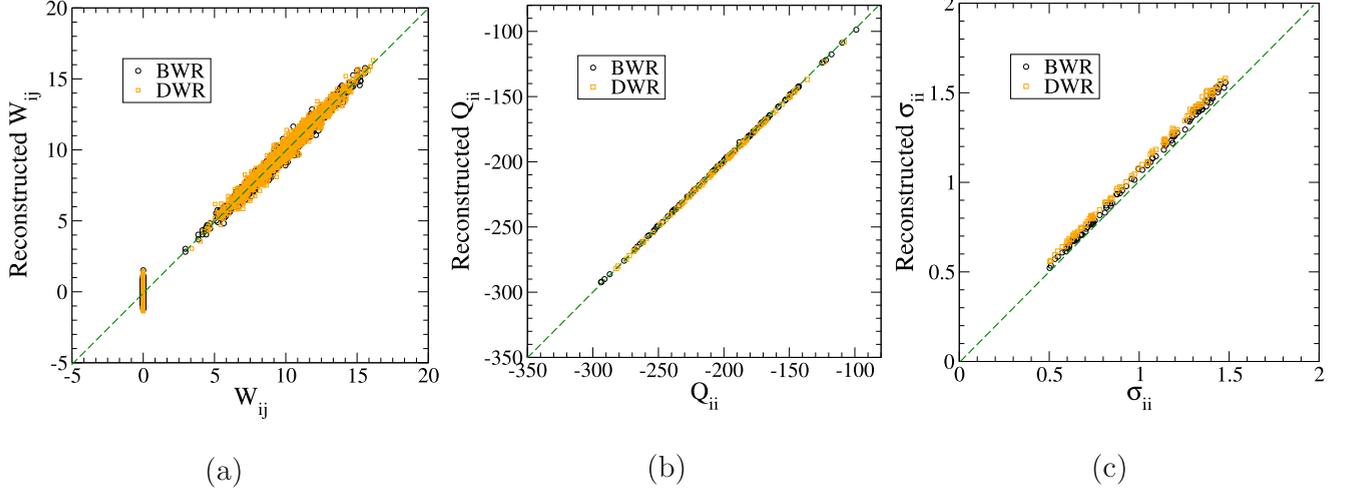
FIG. 4. Network reconstruction of weighted random bidirectional (BWR) and directed (DWR) networks with FHN dynamics (with $\alpha = 0.95$) using the SFI method. Other parameters are the same as in Fig. 3. (a) The reconstructed connection weights between different nodes vs the actual ones. (b) The reconstructed connection weights between different nodes vs the actual ones. (c) The reconstructed noise variance on each node vs the actual ones.

(and their links). Suppose we have a network that consists of a total of N nodes but only the dynamics of $M < N$ nodes can be measured. For simplicity, we consider one-dimensional intrinsic node dynamics $x_i(t)$. By appropriately relabeling the nodes such that the M observed node dynamics are $x_\alpha(t)$, $\alpha = 1, \ldots, M$, we investigate the effect of the unobserved nodes and links on the dynamics of the observed nodes. Hereafter, the observed nodes will be labeled using greek subscript indices. The rest of the unobserved (hidden) node dynamics are $x_m(t)$, $m = M + 1, \ldots, N$.

The major difficulty of reconstructing the observed subnetwork using the time-lag covariance matrix [see Eq. (A4)] is the following: First notice that although the covariance matrix of the observed nodes is the submatrix of the covariance matrix for the entire network ($\mathbf{K}_\tau$), the reconstruction of $\mathbf{Q}$

involves the inverse of the equal-time covariance matrix $\mathbf{K}_0^{-1}$ that mingles the contributions from the hidden connections with the observed ones. Similar to the approach in the previous section, we consider fluctuating dynamics about the noise-free state, and the equation of motion [see Eq. (3)] for the linearized dynamics of the fluctuations of the observed nodes can be written as

$$\dot{\delta x}_\alpha = \sum_{\beta=1}^{M} Q_{\alpha\beta}\delta x_\beta + \sum_{m=M+1}^{N} W_{\alpha m}\delta x_m + \eta_\alpha(t),$$
$$\alpha = 1, 2, \ldots, M. \tag{17}$$

The effect of the extra contributions from the unobserved nodes to x_α has been shown to be similar to a correlated noise [19]. The goal here is to infer the $M \times M$ submatrix $Q_{\alpha\beta}$
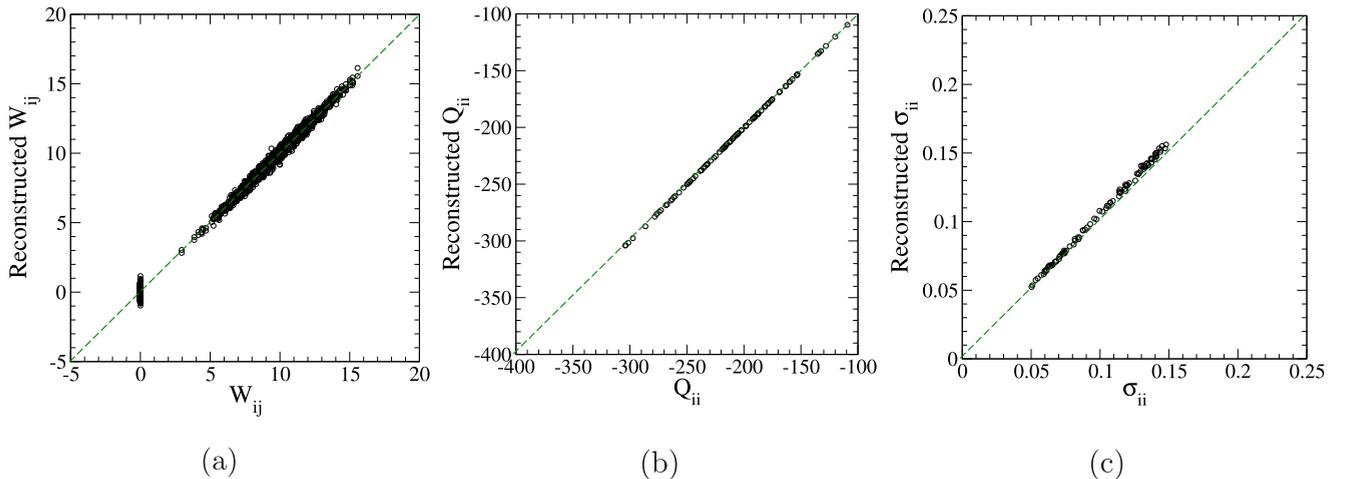

FIG. 5. Network reconstruction of a bidirectional weighted random network (BWR) with FHN dynamics (with $\alpha = 1.05$) using the SFI method. Both u_i and v_i are coupled by their own (different) networks, i.e., $0 \neq \mathbf{W}^{(u)} \neq \mathbf{W}^{(v)} \neq 0$. The noises act on u_i and only the time-series dynamics $u_i(t)$ are observed, and the $v_i(t)$ are hidden. The dashed straight line marks the $y = x$ line. (a) The reconstructed connection weights of the network $W_{ij}^{(u)}$ between different nodes vs the actual ones. (b) The reconstructed diagonal elements of $\mathbf{Q}$ vs the actual ones. (c) The reconstructed noise variance on each node vs the actual ones.
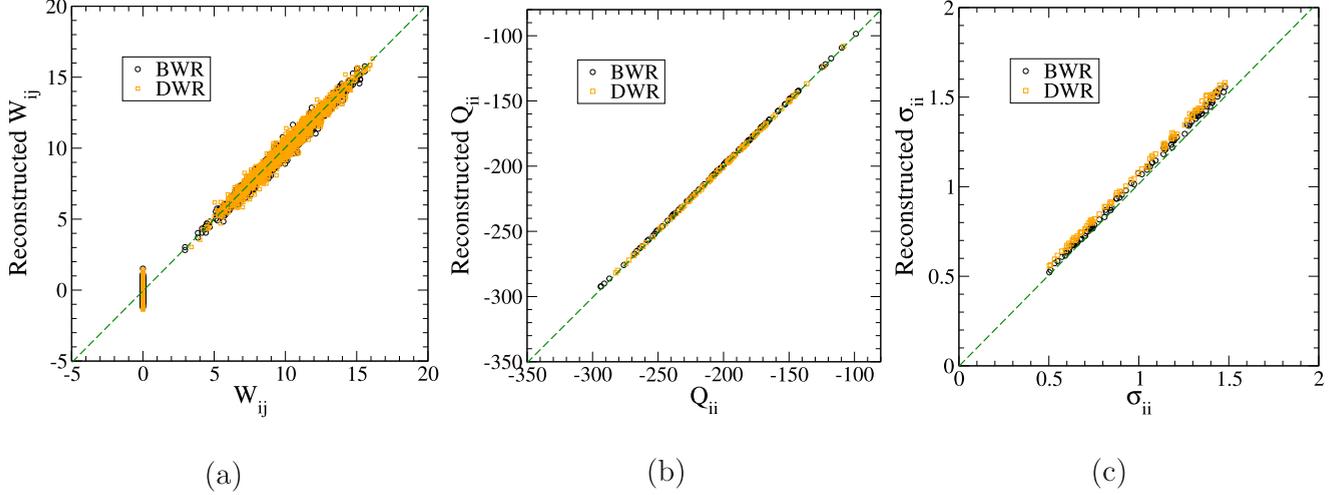
FIG. 6. Network reconstruction of bidirectional (BWR) and directed (DWR) networks with Rössler dynamics using the SFI method. Only the time-series dynamics $x_i(t)$ are observed, and the $y_i(t)$ and $z_i(t)$ are hidden. The dashed straight line marks the $y = x$ line. Rössler dynamics with $a = 0.2$, $b = 0.2$, $c = 9$; $\mu_W = 10$ and $\sigma_W^2 = 2$ for the edge distribution. Both the noise-free dynamics and the noisy dynamics are chaotic attractors. (a) The reconstructed connection weights between different nodes vs the actual ones. (b) The reconstructed diagonal elements of **Q** vs the actual ones. (c) The reconstructed noise variance on each nodes vs the actual ones.

TABLE I. Comparison of the reconstruction performance using the SFI method (denoted by the superscript *SFI*) and time-lag correlation method (denoted by the superscript corr) [see Eqs. (A4) and (A5) in Appendix A] for the weighted random bidirectional networks ($N = 100$ and $p = 0.2$) with logistic and FHN node dynamics for time series of length τ. The time series is sampled with a time interval of $\Delta t = 10^{-3}$. FHN dynamics with only the v components connected by the network are denoted respectively by FHNv (as in Figs. 3 and 4), whereas FHNuv denotes FHN dynamics with both the u and v components separately coupled by their own network, but only the u components of the node dynamics are observed (as in Fig. 5). A time lag of Δt is used in the calculation of the correlation function. The root-mean-square errors of the reconstructed W_{ij} ($i \neq j$) and σ_{ij} are denoted respectively by δ_W and $\delta\sigma$. When the time-series length is not long enough for accurate calculation of the correlation, the time-lag correlation method would result in an unphysical complex reconstruction network matrix and is denoted by an asterisk.

Dynamics	τ	δ_W^{SFI}	δ_σ^{SFI}	δ_W^{corr}	δ_σ^{corr}
Logistic	100	1.97	3.18×10^{-3}	*	*
	1000	0.767	1.01×10^{-3}	*	*
	2000	0.533	7.62×10^{-4}	0.784	1.39×10^{-2}
	5000	0.632	4.88×10^{-4}	0.634	1.21×10^{-2}
FHNv ($\alpha = 1.05$)	100	1.93	3.17×10^{-3}	*	*
	1000	0.757	1.01×10^{-3}	0.942	1.25×10^{-2}
	2000	0.626	7.29×10^{-4}	0.720	1.16×10^{-2}
	5000	0.531	4.93×10^{-4}	0.569	1.13×10^{-2}
FHNv ($\alpha = 0.95$)	100	1.93	3.17×10^{-3}	*	*
	1000	0.759	1.01×10^{-3}	0.951	1.18×10^{-2}
	2000	0.626	7.29×10^{-4}	0.733	1.14×10^{-2}
	5000	0.530	4.93×10^{-4}	0.678	1.12×10^{-2}
FHNuv ($\alpha = 1.05$)	100	2.03	4.77×10^{-3}	*	*
	1000	0.770	1.06×10^{-3}	*	*
	2000	0.636	7.49×10^{-4}	0.967	1.45×10^{-2}
	5000	0.39	4.99×10^{-4}	0.681	1.23×10^{-2}

of the $N \times N$ matrix **Q** of the entire network and the noise covariance of η_α from the time series of the observed nodes $x_\alpha(t)$. Under the SFI framework, the contributions from the fluctuations of the hidden nodes [the $\sum_{m=M+1}^{N} W_{\alpha m} \delta x_m$ term in Eq. (17)] are treated as hidden degrees of freedom and the SFI reconstruction will capture the projection of the dynamics on the observed degrees of freedom.

To investigate the performance of the reconstruction of the connections and the noises of the observed nodes, we generate networks of a total of N nodes and focus on the M observed node dynamics and implement the SFI reconstruction scheme on these observed nodes. Figure 7 shows the reconstructed results against the actual ones for different numbers of observed nodes for BWR and DWR networks with heterogeneous intrinsic logistic node dynamics with consensus coupling, demonstrating that the network between the observed nodes can still be faithfully reconstructed by SFI. The accuracy of the reconstruction depends on the number of hidden nodes, and can be accessed by evaluating the root-mean-square error of the reconstructed W_{ij} and σ_{ij}. Figures 8(a) and 8(b) show respectively the root-mean-square errors of W_{ij} and σ_{ij} as a function of M for the BWR and DWR networks. The reconstruction error increases with the number of hidden nodes as expected. In a similar spirit, the method can be easily applied to the case when the nodes are governed by multicomponent (high-dimensional) intrinsic dynamics and there are both hidden degrees of freedom of the node dynamics as well as hidden nodes. In this case, the network couplings and noise variances of the observed nodes can still be reconstructed by SFI.

Finally, if the information of the dynamics and connections of the observed nodes are available, one can infer the effects of the hidden nodes on the observed one and even deduce the number of hidden nodes in the system. To achieve this goal, one first reconstructs the network connection matrix of
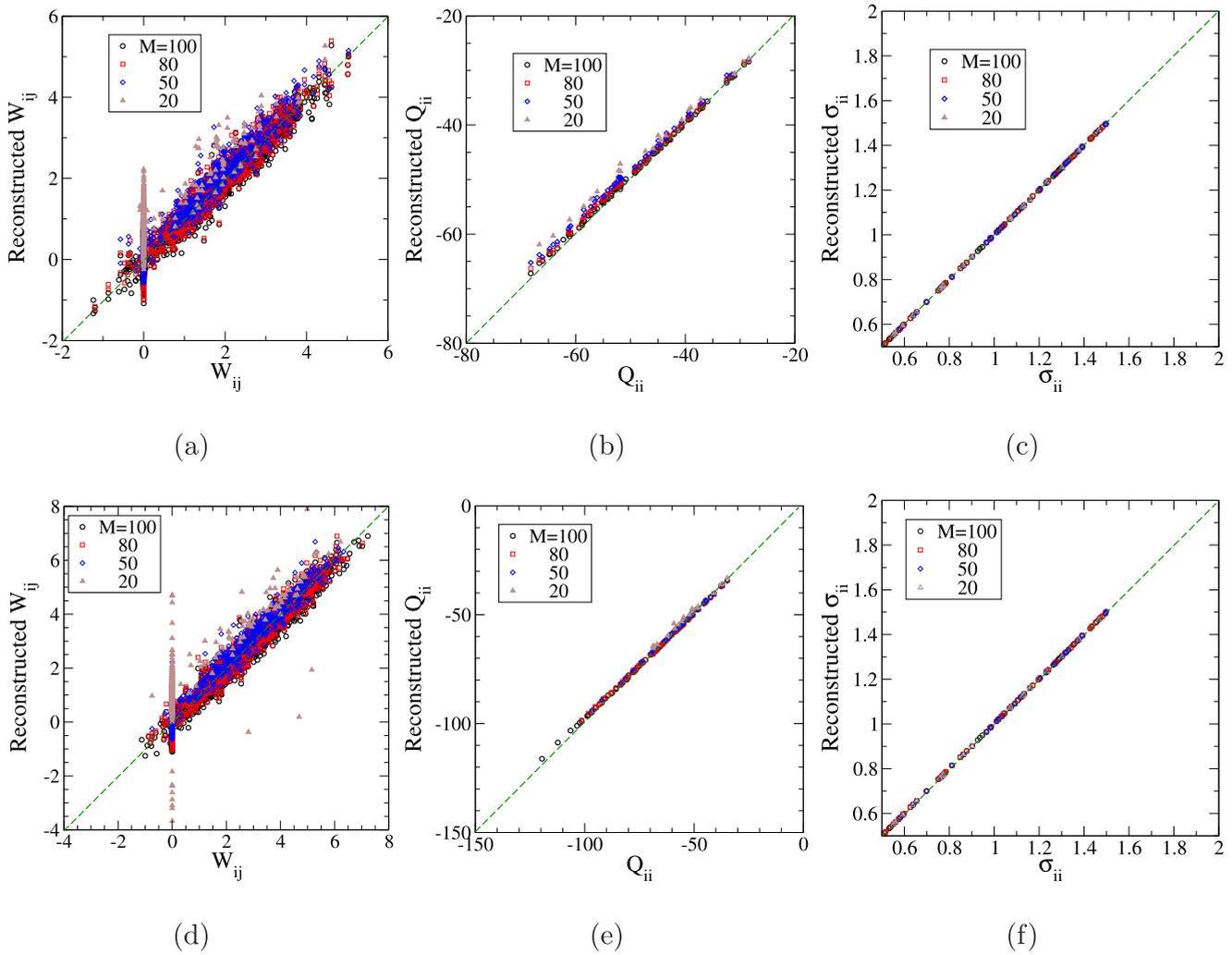
FIG. 7. Network reconstruction with M observed nodes in a network with a total of N nodes with heterogeneous logistic intrinsic node dynamics and consensus coupling $h(x_1, x_2) = x_2 - x_1$ using the SFI method. The dashed straight line marks the $y = x$ line. (a) The reconstructed connection weights between different nodes vs the actual ones for the BWR network. (b) The reconstructed diagonal elements of **Q** vs the actual ones for the BWR network. (c) The reconstructed noise variance on each node vs the actual ones for the BWR network. Panels (d)–(f) are similar to panels (a)–(c) but for the DWR network.
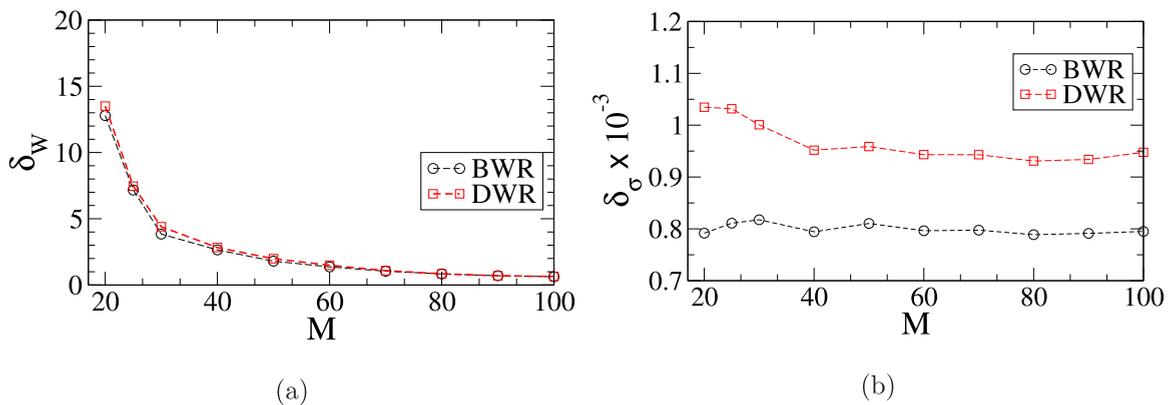

FIG. 8. The root-mean-square error of reconstructed (a) W_{ij} and (b) σ_{ij} as a function of number of observed nodes, M, for weighted random bidirectional (BWR) and directed (DWR) networks with logistic node dynamics using the SFI method for $\tau = 2000$ and $\Delta t = 2\delta t = 10^{-3}$.
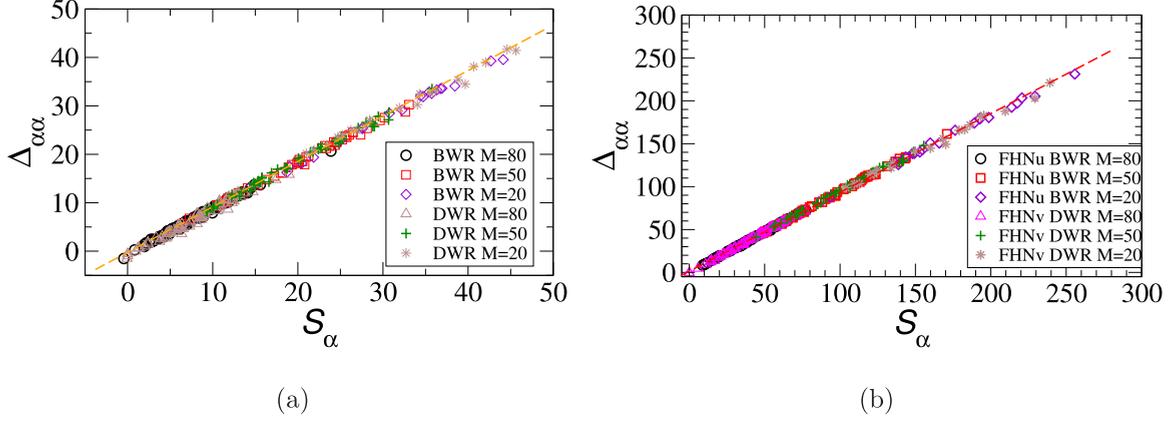
FIG. 9. $\Delta_{\alpha\alpha}$ vs $\mathcal{S}_\alpha$ for BWR and DWR networks with $N = 100$ and $p = 0.2$. The dashed line is a linear fit. (a) Logistic node dynamics with $\mu_W = 2$ and $\sigma_W = 1$. (b) FHN node dynamics with only the u components (FHNu) or only the v components (FHNv) connected by the network. $\mu_W = 10$ and $\sigma_W = 2$.

the observed nodes $\mathbf{Q}^{\text{Recon}}$ by SFI, and compares with the observed connections $\mathbf{Q}^{\text{obs}}$ [defined similar to Eq. (4) for the observed subnetwork] given by

$$Q_{\alpha\beta}^{\text{obs}} \equiv W_{\alpha\beta} + \left[f_\alpha'(X_\alpha) - \sum_{\gamma=1}^{M} W_{\alpha\gamma} \right] \delta_{\alpha\beta}. \qquad (18)$$

Then one can infer the effects of the hidden nodes such as the strength of hidden connections on each observed node and the total number of hidden nodes. Notice that the off-diagonal element of $Q_{\alpha\beta}$ is the same as that of the observed one: $Q_{\alpha\beta} = Q_{\alpha\beta}^{\text{obs}} = W_{\alpha\beta}$ ($\alpha \neq \beta$), but their diagonal element differs by the contributions from the hidden connections, namely, $Q_{\alpha\alpha}^{\text{obs}} - Q_{\alpha\alpha} = \mathcal{S}_\alpha \equiv \sum_{m=M+1}^{N} W_{\alpha m}$, which is the weighted contribution of the hidden connections to the observed node α. As discussed above, $\mathbf{Q}^{\text{Recon}}$ can faithfully find the submatrix of $\mathbf{Q}$ of the whole network. But the diagonal elements of $\mathbf{Q}^{\text{obs}}$ differ from the diagonal elements of $\mathbf{Q}$ since $Q_{\alpha\alpha}^{\text{obs}}$ does not include contributions from the connections from the hidden nodes as revealed in Eq. (18). One expects that, if there are more hidden nodes, the difference between $Q_{\alpha\alpha}^{\text{Recon}}$ and $Q_{\alpha\alpha}^{\text{obs}}$ would be larger. Thus by invoking Eq. (18) one can obtain quantitative information on the hidden nodes by examining

$$\Delta_{\alpha\alpha} \equiv Q_{\alpha\alpha}^{\text{obs}} - Q_{\alpha\alpha}^{\text{Recon}} \simeq \mathcal{S}_\alpha. \qquad (19)$$

The key point is that now one can give an accurate estimate for the effect (weighted sum of the connections) of hidden nodes on the observed node α simply from the value of $\Delta_{\alpha\alpha}$ which can be computed from the observed information. To further investigate the dependence of $\Delta_{\alpha\alpha}$ on the hidden connections, we simulate random networks with one-dimensional logistic and two-dimensional FHN node dynamics under white noise and compute $\Delta_{\alpha\alpha}$. As shown in Fig. 9, $\Delta_{\alpha\alpha}$ displays a rather precise proportional relation with $\mathcal{S}_\alpha$ with a universal proportional constant $\simeq 1$ that appears to be independent of M, N, and the connection probability p of the ER network.

The effect of hidden nodes on the local observed node α can be revealed from Eq. (19), while the information on how

many hidden nodes out there can be obtained from the average of $\Delta_{\alpha\alpha}$ over the observed nodes is defined as

$$[\Delta]_{obs} \equiv \frac{1}{M} \sum_{\alpha=1}^{M} \Delta_{\alpha\alpha} \approx [\mathcal{S}_\alpha]_{obs} \qquad (20)$$

$$\approx \overline{W}(N - M), \qquad (21)$$

where the last equation follows from the assumption of no bias in the choice of hidden nodes.

Furthermore, if there is no bias in the hidden nodes for the underlying network (then $\overline{W}$ can be accurately approximated by sampling over the observed link weights, $[W]_{obs}$), Eq. (21) can be used to estimate the number of hidden nodes from the measurement of the observed network of M nodes to give $N - M \simeq \frac{[\Delta]_{obs}}{[W]_{obs}}$, or the fraction of observed nodes can be estimated as

$$\left(\frac{M}{N} \right)_{\text{est}} \simeq \frac{M}{M + \frac{[\Delta]_{obs}}{[W]_{obs}}}. \qquad (22)$$

Table II displays the estimated observed fraction of nodes for BWR and DWR networks of node dynamics from one dimension to three dimensions with various values of N and M, showing good agreement with the actual values.

V. CONCLUSION AND OUTLOOK

By adopting the SFI method for overdamped Brownian dynamics, we developed the network reconstruction scheme to uncover the connection weights and the noise strengths on the node for a general directed network. The fluctuating noisy dynamics of the nodes allowed a simple natural choice of linear polynomial basis for SFI. The only input is the passive recording of the time-series data of the dynamics of the nodes. The accuracy of the reconstruction is illustrated by explicit generation of undirected and directed random networks of known connection weights and noise variances, and verified by simulations for various intrinsic node dynamics of one to three dimensions. Furthermore, it was demonstrated that even if there is only one observed degree of freedom in the node dynamics and the rest of the degrees of freedom are

TABLE II. Table of the estimated fraction of observed nodes $(\frac{M}{N})_{\text{est}}$ for weighted random bidirectional (BWR) and directed (DWR) networks of logistic, FHN, and Rössler node dynamics. FHN dynamics with only the u components or only the v components connected by the network are denoted respectively by FHNu and FHNv. Total node number $N = 100$ and $p = 0.2$ and number of observed nodes, M. The estimated values $(\frac{M}{N})_{\text{est}}$ are obtained from Eq. (22) and average over five different network realizations.

Dynamics	M	$\frac{M}{N}$	BWR $(\frac{M}{N})_{\text{est}}$	DWR $(\frac{M}{N})_{\text{est}}$
Logistic	20	0.2	0.20 ± 0.03	0.21 ± 0.02
	50	0.5	0.52 ± 0.03	0.51 ± 0.02
	80	0.8	0.82 ± 0.03	0.81 ± 0.02
FHNu	20	0.2	0.20 ± 0.03	0.21 ± 0.03
	50	0.5	0.51 ± 0.02	0.51 ± 0.02
	80	0.8	0.81 ± 0.01	0.80 ± 0.01
FHNv	20	0.2	0.22 ± 0.04	0.20 ± 0.02
	50	0.5	0.52 ± 0.03	0.51 ± 0.02
	80	0.8	0.82 ± 0.02	0.80 ± 0.01
Rössler	20	0.2	0.21 ± 0.03	0.21 ± 0.02
	50	0.5	0.52 ± 0.03	0.51 ± 0.01
	80	0.8	0.81 ± 0.01	0.80 ± 0.01

hidden, the SFI reconstruction for the network and noises of the observed node variables still performs faithfully.

The SFI reconstruction of a network can overcome several problems encountered in noisy network reconstructions. As mentioned, the time-lag correlation method involved taking the principal logarithm of a matrix as given by Eq. (A4) that can lead to the constructed network weights becoming complex numbers, and often some sort of trial and error smoothing filter is needed to preprocess the data to suppress the noise in order to render the reconstructed weights to be real valued. Such a smoothing filter preprocess on the time-series data might lead to uncontrolled effects on the reconstruction of the noise matrix σ_{ij} using the time-correlation method. There is no such problem at all for the SFI reconstruction method due to its straightforward formulation. The other problem in some reconstruction methods is the need to compute higher-order derivatives of the dynamical variables from the time-series data, whose accuracy requires a very high temporal resolution and data sampling rate that can be difficult to achieve, especially in the presence of noise. On the other hand, only first-order time differences are required in the SFI reconstruction method. As for the hidden node or connection problem, an estimated reconstruction scheme for undirected networks was proposed in [13] which involves sophisticated data and correlation matrix manipulations, whereas the principle in the SFI reconstruction naturally takes into account the effects of hidden degrees of freedom in terms of projection onto the observed one and hence can be applied to the hidden node problem for general directed networks in a straightforward manner.

One advantage of using SFI network reconstruction is that the SFI framework provides a convenient evaluation of the quality of the reconstruction from the self-estimate of the error of reconstruction solely from the data (without knowing the true values) using error estimation results (B13) and (B14) in Appendix B. For example, the error of noise variance can

be estimated from Eq. (B14) to be $\delta_\sigma \simeq \sigma_m 2N\sqrt{\Delta t/\tau}$, and with $\tau = 2000$, $\Delta t = 10^{-3}$, $N = 100$, and mean noise variance $\sigma_m = 0.01$ [these are the parameters used in Fig. 8(b)], one gets $\delta_\sigma \simeq 1.4 \times 10^{-3}$ which is of the same order as the measured result from simulation [the $M = 100$ data points in Fig. 8(b)].

One can further improve the reconstruction accuracy by inferring the adjacency matrix using clustering methods to identify node pairs into either connected or disconnected groups (as in [6,7,9,10]). This can be achieved by arranging the reconstructed elements of $W_{ij}(i \neq j)$ in ascending order for fixed i, and employing some cluster algorithm to separate the elements into two clusters, one can identify those nodes j that connect to i. Thus one can identify the zero and nonzero W_{ij} and hence the adjacency matrix. Such a procedure is expected to greatly reduce false connections.

By interpreting the noisy dynamics of a complex network as a Brownian system under stochastic force field, one can broaden the view of network dynamics in more physical terms and calculate the phase space velocity or flow, entropy production rate [see Eq. (B11) in Appendix B], and make connection to stochastic thermodynamics whose concepts and well-developed techniques can be helpful in network dynamics research and vice versa.

In this work, we assumed the dynamics fluctuates about some stable attractor and the system can be linearized about some noise-free state. However, in some situations the intrinsic dynamics can be highly nonlinear or the coupling function is not linearizable: for example, if $h(x_i, x_j) = (x_j - x_i)^3$ the leading-order network coupling term about the synchronized state is nonlinear. In this highly nonlinear situation, network reconstruction based on SFI can still be carried out by employing higher-order basis functions. And the SFI framework also provides a criterion for choosing the order of the basis functions in terms of the information rate [as defined in Eq. (B12) in Appendix B] to avoid overfitting [16] to ensure good reconstruction quality.

The present work focused on network dynamics subjected to temporally uncorrelated additive white noise; it can also be easily generalized to multiplicative white noises in which the noise strength on the node depends on the dynamical state of the node. A diffusive drift will result from the multiplicative noises and SFI can handle the situation equally well [16]. Furthermore, the network reconstruction problem is more challenging if the noises acting on the nodes are temporally correlated, and more sophisticated schemes involving longer time-lag covariances are required which also suffered from the problem of a complex reconstructed connection matrix [11]. In this case, the SFI network reconstruction scheme would be more straightforward since the Ornstein-Uhlenbeck process for the correlated noises can be introduced as an extra dynamical variable subjected to white noise. And this extra variable can be viewed as a hidden degree of freedom as discussed in Sec. III B; then the SFI network reconstruction can be implemented relatively easily.

On the other hand, in some situations where the nature or the source of the noises is unknown, or the time-series data are prone to an unknown source of errors and may even be unreliable, then one may employ an empirical approach using the Bayesian method [23] to reconstruct the underlying network

structure with optimal likelihood. These methods require the input of data and network models and have been demonstrated [24–26] to work well even with scarce data. By comparing the reconstructed network structures using the SFI or time-lag correlation methods with that of Bayesian network inference, one can gain some insight on the nature of the noises or errors in the network dynamics.

Finally, since the noisy node dynamics considered in this work is quite general, there is a broad area of possible applications. In particular, our method should be most relevant to infer network structures resulted from complex physical interactions in which the nodes have some sort of direct physical interaction or direct influence with other nodes. Some possible applications involve networks or interacting systems in which the dynamical time-series data have been measured experimentally. Some examples include functional magnetic resonance imaging (fMRI) time-series measurement [27] for brain function networks, multielectrode array recordings [28] for *in vitro* neuronal networks, gene expression time-series data for gene regulatory networks [29,30], membrane receptors or proteins interaction from diffusion dynamics on living cells [31,32], intercellular communication networks in cells [33], or bacteria colonies [34].

APPENDIX A: SUMMARY FOR THE TIME-LAG CORRELATION METHOD OF NETWORK RECONSTRUCTION FOR COUPLED NOISY NETWORK

Consider a network with N nodes whose intrinsic one-dimensional node dynamics $x_i(t)$ is described by Eq. (1) and the corresponding fluctuating dynamics governed by Eqs. (3) and (4). Assuming the matrix $\mathbf{Q}$ in Eq. (4) is time independent (which is true if the asymptotic noise-free solution is time independent, i.e., the system fluctuates around a stable stationary solution), then the information of the network connections can be retrieved by measuring the long-time limit of the time-lag correlations of the node dynamics [9–11]:

$$\mathbf{K}_t \equiv \lim_{t' \to \infty} \langle [\vec{x}(t' + t) - \langle \vec{x}(t' + t) \rangle][\vec{x}(t') - \langle \vec{x}(t') \rangle]^\mathsf{T} \rangle. \quad \text{(A1)}$$

In practice, the noise-free solution can be approximated by $X_i \simeq \langle x_i \rangle$. Network connections contained in the matrix $\mathbf{Q}$ can be retrieved by measuring the long-time limit of the time-lag correlations of the node dynamics defined in Eq. (A1). From [9], the reconstruction formula for a directed network is

$$\mathbf{K}_t = e^{t\mathbf{Q}} \mathbf{K}_0. \quad \text{(A2)}$$

For the special case of undirected network, i.e., symmetric $\mathbf{Q}$ and uniform uncorrelated noises, $\sigma = \sigma^2 \mathbf{I}$, one has [6,7]

$$\mathbf{Q} = -\frac{\sigma^2}{2} \mathbf{K}_0^{-1}. \quad \text{(A3)}$$

However, the properties of the white noise are usually not known and need to be reconstructed also. Therefore, in general, for directed or undirected networks under white noises,

the matrix $\mathbf{Q}$ can be constructed from the measurement of the time-lag covariance matrices via

$$\mathbf{Q} = \frac{1}{t} \ln(\mathbf{K}_t \mathbf{K}_0^{-1}), \quad \text{(A4)}$$

and the noise matrix is then reconstructed from the Lyapunov equation

$$\sigma = -\mathbf{Q}\mathbf{K}_0 - \mathbf{K}_0 \mathbf{Q}^\mathsf{T}. \quad \text{(A5)}$$

For the situations that the noise-free solution is not stationary, such as periodic dynamics (limit cycle) or even chaotic, it has been demonstrated numerically [7,9] that the above reconstruction scheme can still work rather well.

The above covariance-relation-based method [Eq. (A4)] involves a calculation of the principal matrix logarithm, which is very sensitive to the noise in the data. For example, when applied directly to the multielectrode array data for neuropotential time-series recordings, an unphysical complex $\mathbf{Q}$ is obtained. A similar complex matrix was also found when this covariance relation was used directly to estimate directed connectivity for fMRI measurements [27]. In some cases, such a complex matrix problem can be avoided if a low-pass filter (such as a moving average filter) is first applied to the data to reduce the random noises [28].

APPENDIX B: SUMMARY FOR STOCHASTIC FORCE INFERENCE OF OVERDAMPED BROWNIAN DYNAMICS

Consider an overdamped stochastic system with N_d-dimensional phase-space coordinates x_i governed by Brownian dynamics under the force field $\vec{F}(\vec{x})$ (with the mobility tensor absorbed in its definition) and obeying the following Langevin equation:

$$\dot{\vec{x}} = \vec{F}(\vec{x}) + \vec{\eta}(t), \quad \text{(B1)}$$

$$\overline{\vec{\eta}(t)\vec{\eta}^\mathsf{T}(t')} = 2\mathbf{D}\delta(t - t'), \qquad \overline{\vec{\eta}(t)} = 0, \quad \text{(B2)}$$

where $\mathbf{D}$ is the $d \times d$ diffusion tensor. In general the zero-mean white noise $\vec{\eta}$ can be multiplicative with a spatially dependent diffusion field $\mathbf{D}(\vec{x})$, in which an extra drift term $\nabla \cdot \mathbf{D}$ should be added to the right-hand side of Eq. (B1) when it is written in Itô form. Based on the communication-theory notion of capacity, Frishman and Ronceray showed that these stochastic trajectories contain a limited amount of information, and SFI exploits such information to fit the force field accurately with a linear combination of some appropriately chosen basis functions [16]. By choosing a suitable projection basis $b_\alpha(\vec{x})$ ($\alpha = 1, \ldots, n_b$, a set of n_b fitting functions), the method of stochastic force inference takes the discrete time-series data of $\vec{x}(t_k)$ as input, and returns with the inferred force field $\vec{F}(\vec{x})$, phase-space velocity field $\vec{v}(\vec{x})$, and diffusion field $\mathbf{D}(\vec{x})$. The SFI method is described briefly as follows; the detailed information-theoretic foundation, derivations, and examples in various physical systems can be found in [16].

Here we simply give the calculation steps with the stochastic trajectories as the time-series data together with a chosen projection basis as input. First construct the $n_b \times n_b$ matrix

B and the $N_d \times n_b$ matrix **M** from the average of the basis vectors evaluated from the time-series data:

$$B_{\alpha\beta} = \langle b_\alpha(\vec{x}_i) b_\beta(\vec{x}_i) \rangle \equiv \frac{1}{N_{\text{step}}} \sum_{k=1}^{N_{\text{step}}} b_\alpha(\vec{x}(t_k)) b_\beta(\vec{x}(t_k)), \quad \text{(B3)}$$

$$M_{j\beta} = \langle \frac{\Delta x_j}{\Delta t} b_\beta \rangle \equiv \frac{1}{N_{\text{step}}} \sum_{i=k}^{N_{\text{step}}} \frac{\Delta x_j(t_k)}{\Delta t_k} b_\beta(\vec{x}(t_k)). \quad \text{(B4)}$$

Then the force field can be inferred as

$$\vec{F}(\vec{x}) \simeq \mathbf{M}\mathbf{B}^{-1} \vec{b}(\vec{x}). \quad \text{(B5)}$$

And the diffusion tensor field can be inferred as

$$D_{ij}(\vec{x}) \simeq \sum_{\beta\gamma}^{n_b} \langle \frac{\Delta x_i \Delta x_j}{2\Delta t} b_\beta \rangle B_{\beta\gamma}^{-1} b_\gamma(\vec{x}), \quad \text{(B6)}$$

$$\langle \frac{\Delta x_i \Delta x_j}{2\Delta t} b_\beta \rangle \equiv \frac{1}{N_{\text{step}}} \sum_{k=1}^{N_{\text{step}}} \frac{\Delta x_i(t_k) \Delta x_j(t_k)}{2\Delta t_k} b_\beta(\vec{x}(t_k)). \quad \text{(B7)}$$

For the purpose of network reconstruction considered in this paper, **D** is taken to be spatially independent. For the case of a spatially independent diffusion tensor, Eq. (B6) is given by the average over the time-series samples:

$$D_{ij} = \sum_{\beta\gamma}^{n_b} \langle \frac{\Delta x_i \Delta x_j}{2\Delta t} b_\beta \rangle B_{\beta\gamma}^{-1} \langle b_\gamma \rangle,$$

$$\langle b_\gamma \rangle \equiv \frac{1}{N_{\text{step}}} \sum_{k=1}^{N_{\text{step}}} b_\gamma(\vec{x}(t_k)). \quad \text{(B8)}$$

The phase-space velocity field can be inferred as

$$\vec{v}(\vec{x}) \simeq \mathbf{M}^S \mathbf{B}^{-1} \vec{b}(\vec{x}), \quad \text{(B9)}$$

$$M_{j\beta}^S \equiv \frac{1}{N_{\text{step}}} \sum_{i=k}^{N_{\text{step}}} \frac{\Delta x_j(t_k)}{\Delta t_k} b_\beta \left(\frac{\vec{x}(t_{k+1}) + \vec{x}(t_k)}{2} \right). \quad \text{(B10)}$$

The average entropy production rate and information rate along the trajectory are given by

$$\dot{S} = \frac{\Delta S}{\tau} \simeq \frac{1}{N_{\text{step}}} \sum_{k=1}^{N_{\text{step}}} \vec{v}^\top(\vec{x}_k) \mathbf{D}^{-1}(\vec{x}_k) \vec{v}(\vec{x}_k), \quad \text{(B11)}$$

$$\frac{I_b}{\tau} \simeq \frac{1}{4N_{\text{step}}} \sum_{k=1}^{N_{\text{step}}} \vec{F}^\top(\vec{x}_k) \mathbf{D}^{-1}(\vec{x}_k) \vec{F}(\vec{x}_k). \quad \text{(B12)}$$

The SFI method can also provide a self-consistent estimate of the errors of the inferred force and diffusion fields from the time-series data as follows [16]:

$$\left(\frac{\delta F}{F} \right)^2 = \frac{N_d n_b}{2I_b}, \quad \text{(B13)}$$

$$\left(\frac{\delta D}{D} \right)^2 = \frac{4N_d n_b \Delta t}{\tau}. \quad \text{(B14)}$$

Note that the above relative errors can be obtained solely from the data without the knowledge of the corresponding actual values, which is essential in practice and also can provide valuable information on the inference quality.

[1] Nature, Special issue on Big Data 455, No. 7209 (2008).

[2] S. H. Strogatz, Exploring complex networks, Nature (London) **410**, 268 (2001).

[3] M. E. J. Newman, The structure and function of complex networks, SIAM Rev. **45**, 167 (2003).

[4] M. E. J. Newman, *Networks: An Introduction* (Oxford University Press, Oxford, UK, 2010).

[5] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási, The large-scale organization of metabolic networks, Nature (London) **407**, 651 (2000).

[6] E. S. C. Ching, P.-Y. Lai, and C. Y. Leung, Extracting connectivity from dynamics of networks with uniform bidirectional coupling, Phys. Rev. E **88**, 042817 (2013).

[7] E. S. C. Ching, P.-Y. Lai, and C. Y. Leung, Reconstructing weighted networks from dynamics, Phys. Rev. E **91**, 030801(R) (2015).

[8] Z. Zhang, Z. Zheng, H. Niu, Y. Mi, S. Wu, and G. Hu, Solving the inverse problem of noise-driven dynamic networks, Phys. Rev. E **91**, 012814 (2015).

[9] E. S. C. Ching and H. C. Tam, Reconstructing links in directed networks from noisy dynamics, Phys. Rev. E **95**, 010301(R) (2017).

[10] P.-Y. Lai, Reconstructing network topology and coupling strengths in directed networks of discrete-time dynamics, Phys. Rev. E **95**, 022311 (2017).

[11] H. C. Tam, E. S. C. Ching, and Pik-Yin Lai, Reconstructing networks from dynamics with correlated noise, Physica A **502**, 106 (2018).

[12] F. Goetze and P.-Y. Lai, Reconstructing positive and negative couplings in Ising spin networks by sorted local transfer entropy, Phys. Rev. E **100**, 012121 (2019).

[13] E. S. C. Ching and P. H. Tam, Effects of hidden nodes on the reconstruction of bidirectional networks, Phys. Rev. E **98**, 062318 (2018).

[14] L. Pérez García, J. Donlucas Pérez, G. Volpe, A. V. Arzola, and G. Volpe, High-performance reconstruction of microscopic force fields from Brownian trajectories, Nat. Commun. **9**, 5166 (2018).

[15] M. Beheiry, M. Dahan, and J. B. Masson, InferenceMAP: Mapping of single-molecule dynamics with Bayesian inference, Nat. Methods **12**, 594 (2015).

[16] A. Frishman and P. Ronceray, Learning Force Fields from Stochastic Trajectories, Phys. Rev. X **10**, 021009 (2020).

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

[18] B. Bollobás, *Random Graphs* (Academic, London, 1985).

[19] B. Gemao and P.-Y. Lai, Effects of hidden nodes on noisy network dynamics, Phys. Rev. E **103**, 062302 (2021).

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

[21] H. Haehne, J. Casadiego, J. Peinke, and M. Timme, Detecting Hidden Units and Network Size from Perceptible Dynamics, Phys. Rev. Lett. **122**, 158301 (2019).

[22] R. Shi, W. Jiang, and S. Wan, Detecting network structures from measurable data produced by dynamics with hidden variables, Chaos **30**, 013138 (2020).

[23] C. T. Butts, Network inference, error, and informant (in)accuracy: A Bayesian approach, Soc. Networks **25**, 103 (2003).

[24] M. E. J. Newman, Network structure from rich but noisy data, Nat. Phys. **14**, 542 (2018).

[25] T. P. Peixoto, Network Reconstruction and Community Detection from Dynamics, Phys. Rev. Lett. **123**, 128301 (2019).

[26] J.-G. Young, G. T. Cantwell, and M. E. J. Newman, Bayesian inference of network structure from unreliable data, J. Complex Networks **8**, cnaa046 (2021).

[27] M. Gilson, R. Moreno-Bote, A. Ponce-Alvarez, P. Ritter, and G. Deco, Estimation of directed effective connectivity from fMRI functional connectivity hints at asymmetries of cortical connectome, PLoS Comput. Biol. **12**, e1004762 (2016).

[28] C. Sun, K. C. Lin, C. Y. Yeung, E. S. C. Ching, Y.-T. Huang, P.-Y. Lai, and C. K. Chan, Revealing directed effective connec-

tivity of cortical neuronal networks from measurements, Phys. Rev. E **105**, 044406 (2022).

[29] V. A. Huynh-Thu and P. Geurts, dynGENIE3: Dynamical GE-NIE3 for the inference of gene networks from time series expression data, Sci. Rep. **8**, 3384 (2018).

[30] Z. Bar-Joseph, A. Gitter, and I. Simon, Studying and modelling dynamic biological processes using time-series gene expression data, Nat. Rev. Genet. **13**, 552 (2012).

[31] W. He, H. Song, Y. Su, L. Geng, B. J. Ackerson, H. B. Peng, and P. Tong, Dynamic heterogeneity and non-Gaussian statistics for acetylcholine receptors on live cell membrane, Nat. Commun. **7**, 11701 (2016).

[32] Y. Shen, C. Luo, Y. Wen, W. He, P. Huang, H.-Y. Chen, P.-Y. Lai, and P. Tong, Directed motion of membrane proteins under an entropy-driven potential field generated by anchored proteins, Phys. Rev. Research **3**, 043195 (2021).

[33] D. B. Brückner, N. Arlt, A. Fink, P. Ronceray, J. O. Rädler, and C. P. Broedersz, Learning the dynamics of cell-cell interactions in confined cell migration, Proc. Natl. Acad. Sci. USA **118**, e2016602118 (2021).

[34] J. d'Alessandro, A. Solon, Y. Hayakawa, C. Anjard, F. Detcheverry, J.-P. Rieu, and C. Rivière, Contact enhancement of locomotion in spreading cell colonies, Nat. Phys. **13**, 999 (2017).