Chaos

Measuring the importance of individual units in producing the collective behavior of a complex network

Cite as: Chaos **31**, 093123 (2021); https://doi.org/10.1063/5.0055051
Submitted: 25 April 2021 . Accepted: 30 August 2021 . Published Online: 23 September 2021

X. San Liang

Measuring the importance of individual units in producing the collective behavior of a complex network

Cite as: Chaos **31**, 093123 (2021); doi: 10.1063/5.0055051
Submitted: 25 April 2021 · Accepted: 30 August 2021 ·
Published Online: 23 September 2021

View Online Export Citation CrossMark

X. San Liang[a)]

AFFILIATIONS

Fudan University, Shanghai 200438, China; Shanghai Qi Zhi Institute (Andrew C. Yao Institute), Shanghai 200232, China; and China Institute for Advanced Study, Central University of Finance and Economics, Beijing 100081, China

[a)]**Author to whom correspondence should be addressed:** sanliang@courant.nyu.edu

ABSTRACT

A quantitative evaluation of the contribution of individual units in producing the collective behavior of a complex network can allow us to understand the potential damage to the structure integrity due to the failure of local nodes. Given a time series for each unit, a natural way to do this is to find the information flowing from the unit of concern to the rest of the network. In this study, we show that this flow can be rigorously derived in the setting of a continuous-time dynamical system. With a linear assumption, a maximum likelihood estimator can be obtained, allowing us to estimate it in an easy way. As expected, this "cumulative information flow" does not equal the sum of the information flows to other individual units, reflecting the collective phenomenon that a group is not the addition of individual members. For the purpose of demonstration and validation, we have examined a network made of Stuart–Landau oscillators. Depending on the topology, the computed information flow may differ. In some situations, the most crucial nodes for the network are not the hubs, i.e., nodes with high degrees; they may have low degrees and, if depressed or attacked, will cause the failure of the entire network. This study can help diagnose neural network problems, control epidemic diseases, trace city traffic bottlenecks, identify the potential cause of power grid failure (e.g., the 2003 great power outage that darkened much of North America), build robust computer networks, and so forth.

Complex networks, for example the brain, power grids, city traffic, financial markets, the Internet, cellular regulatory networks, etc., may cease to function due to the depression or deterioration of certain individual nodes. An example is the 2003 great blackout that darkened much of North America. How to identify the cause(s), i.e., the deteriorated/depressed unit(s), is thence of great importance. Usually, this is studied by preferentially removing a unit and observing the change in behavior of the network. This may be expensive, and even infeasible, for many networks (biological networks, in particular), as breaking a unit means terminating the experiment. This study found that, when time series of measurements for some property of the nodes exist, the problem can be solved by quantitatively assessing, in an easy way, the contribution of each node to the network in terms of cumulant information flow. It is found that the macrostate of a network is not just a simple addition of the individual states, unless all the individual units are independent. It is also found that, in some situations, the most crucial nodes for the network may not be the hubs; suppression of these nodes will shut down the entire network. This study provides an easy approach to measuring the importance of individual units in a complex network, which can help diagnose neural network problems, control epidemic diseases, trace city traffic bottlenecks, identify the potential cause of power grid failure, build robust computer networks, etc.

I. INTRODUCTION

Complex networks provide a framework for the study of many social, biological, and engineering systems, such as the Internet, the brain, power grids, financial trading markets, food webs, and gene regulatory networks, to name a few. A network consists of nodes or vertexes the individual units or organizations contain and links or edges for the interactions among the nodes. For a node, the number of links connected to other nodes is called its degree. By degree distribution, we can have homogeneous and heterogeneous networks. The former class has binomial or Poisson degree distributions, examples including random graphs[1] and some small-world networks,[2] while the latter class is scale free, bearing probability distributions P of degree k following a power law $P(k) \sim k^{-\gamma}$, with an exponent $\gamma \sim 2$–3. Many social,[3] biological,[4] and technological networks[5] have the scale-free property; other topological properties include high clustering coefficient, community and hierarchical structures, and, for directed networks, reciprocity, triad significance profile, etc.[6,7]

A goal of complex network studies is to understand how individuals collaborate to produce the collective behavior. One question to ask is whether the connectivity of a network is robust (see Ref. 8 for a formal definition) to local node failure, deterioration, or functional depression. Of particular interest is whether initially a tiny shock may cascade to disrupt the network on a large scale. This is a field of extensive research. There are a huge number of studies in this regard, among which are Refs. [9–12], to name a few. How to quantify the contribution of a unit to the network as a whole is thence an important issue; it is related to many real world problems, such as power grid failure (e.g., the 2003 massive blackout that darkened much of the North American upper Midwest and Northeast[13]), control of epidemic diseases, identification of bottlenecks in city traffic, etc. Usually, this is studied by observing the connectivity after preferential removal of a unit, which is found to have different effects on the two types of networks. If the removal or attack is random, heterogeneous networks are quite robust as compared to homogeneous networks; if, however, the attack is intentional at some special nodes, then heterogeneous networks could be rather fragile. These special nodes are usually highly connected ones, i.e., hubs, as easily imagined. Recently, Tanaka *et al.*[14] observed that sparsely connected nodes may be more important, which, if functionally depressed, may result in the drastic change in a network structure. That is, the structure integrity or robustness could also be largely influenced by low-degree nodes rather than by hubs. We hence cannot judge the importance of a unit simply by degree. It depends on many different properties of the network topology in question.

As said above, the problem is usually tackled by removing a unit and observing the change in topology of the network of concern. However, in many networks, biological networks in particular, this is often infeasible as breaking a unit means terminating the experiment. On the other hand, we may have time series of measurements. So the whole problem is converted into assessing the importance of a unit from analyzing the signals as observed. Previously, we have rigorously formulated information flow within dynamical systems (e.g., Refs. 15 and 16); it has been widely used for studying the causal relations among dynamical events (see Ref. 17, Sec. 2, and the references therein) and hence is readily available for the study of the interactions among nodes in a network. One may think that the contribution of a given node may be obtained by adding up all the information flows from it to the other nodes. Unfortunately, as we will see in Secs. II A and IV, this is true only when all the nodes are disconnected, i.e., when the nodes do not form a network and hence no collective behaviors emerge. This from one aspect manifests the well-known fact that groups are not simply the addition of their individual members; they could be more or less (some social science examples can be seen in Refs. [18–21]).

In the following, we first present the setting for the problem, and then derive the information flow from an individual unit to the network. Maximum likelihood estimation is made in Sec. III; it yields a formula for easy assessment of the importance of a node from a given time series. As a validation, and also a demonstration of application, Sec. IV presents a network of synchronized Stuart–Landau oscillators which, when a fraction of nodes become deteriorated, may become silent completely. This study is concluded in Sec. V.

II. INFORMATION FLOW FROM A UNIT TO THE ENTIRE NETWORK

Consider a network modeled by an n-dimensional dynamical system,

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}, t) + \mathbf{B}(\mathbf{x}, t)\dot{\mathbf{w}}, \tag{1}$$

where $\mathbf{x}$ is the state variable vector for the n nodes $(x_1, x_2, \ldots, x_n)$, $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{F} = (F_1, \ldots, F_n)$, the differentiable functions of $\mathbf{x}$ and time t that describe the interaction paths (edges/links), $\mathbf{w}$ is a vector of m independent standard Wiener processes, and $\mathbf{B} = (b_{ij})$ an $n \times m$ is the matrix of stochastic perturbation amplitude. Here, we follow the convention in physics not to distinguish a random variable and a deterministic variable. (In probability theory, they are usually distinguished by uppercase and lowercase symbols.) To examine the influence of a unit to the entire network made of the n units, it suffices to consider the component x_1; if not, we can always re-arrange the vector $\mathbf{x}$ to make it so. The whole problem now boils down to finding the information flow from x_1 to $(x_2, x_3, \ldots, x_n)$, which we will be denoting as $\mathbf{x}_{2..n}$ henceforth (i.e., as $\mathbf{x}$ with component 1 removed).

In Ref. 16, the information flow between two individual components x_i and x_j has been rigorously derived from first principles. However, the information flow from one component, here x_1, to a multitude of components, here $\mathbf{x}_{2..n}$, is yet to be implemented. One may conjecture that it is just an addition of all flows from x_1 to all the individual components of $\mathbf{x}_{2..n}$. As we will see below, this is generally not the case, and the nonadditivity is a reflection of the macrostate or collective behavior of a multi-connected network.

We follow the strategy used in Ref. 22 to do the derivation. The information flow is, by the physical argument therein, the amount of entropy transferred from x_1 to $\mathbf{x}_{2..n}$. We hence need to find the evolution of the joint entropy of $\mathbf{x}_{2..n}$ and single out the contribution to this evolution from x_1. This results in the following theorem.

Theorem 2.1 *For the dynamical system (1), if the probability density function (PDF) of $\mathbf{x}$ is compactly supported, then the information flow from x_1 to $(x_2, x_3, \ldots, x_n)$ is*

$$T_{1\to 2..n} = -E\left[\sum_{i=2}^{n} \frac{1}{\rho_{2..n}} \frac{\partial F_i \rho_{2..n}}{\partial x_i}\right] + \frac{1}{2} E\left[\sum_{i=2}^{n}\sum_{j=2}^{n} \frac{1}{\rho_{2..n}} \frac{\partial^2 g_{ij}\rho_{2..n}}{\partial x_i \partial x_j}\right]. \tag{2}$$

The units are nats per unit time. In the equation, $\rho_{2..n}$ is the joint PDF of $(x_2, x_3, \ldots, x_n)$, $g_{ij} = \sum_{k=1}^{m} b_{ik} b_{jk}$, and E signifies mathematical expectation.

Proof. Associated with (1), there is a Fokker–Planck equation governing the evolution of the PDF ρ of $\mathbf{x}$,

$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho F_1}{\partial x_1} + \frac{\partial \rho F_2}{\partial x_2} + \cdots + \frac{\partial \rho F_n}{\partial x_n} = \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{n}\frac{\partial^2 g_{ij}\rho}{\partial x_i \partial x_j}, \tag{3}$$

where $g_{ij} = \sum_{k=1}^{m} b_{ik}b_{jk}$, $i, j = 1, \ldots, n$. This marginal PDF of x_1, $\rho_1(x_1)$, is obtained by integrating out $(x_2, \ldots, x_n)$ in (3). By the assumption of compactness of ρ, the resulting equation becomes

$$\frac{\partial \rho_1}{\partial t} + \frac{\partial}{\partial x_1}\int_{\mathbb{R}^{n-1}} \rho F_1\, d\mathbf{x}_{2..n} = \frac{1}{2}\frac{\partial^2}{\partial x_1^2}\int_{\mathbb{R}^{n-1}} g_{11}\rho\, d\mathbf{x}_{2..n}. \tag{4}$$

For the sake of notational simplicity, here we have written $dx_2 dx_3 \ldots dx_n$ as $d\mathbf{x}_{2..n}$. From this, the evolution of the marginal entropy of x_1, written H_1, can be derived as

$$\frac{dH_1}{dt} = -E\left[F_1 \frac{\partial \log \rho_1}{\partial x_1}\right] - \frac{1}{2}E\left[g_{11}\frac{\partial \log \rho_1}{\partial x_2}\right]. \tag{5}$$

See Ref. 22 for a proof.

To study the impact of x_1 on the rest of the network, we need to consider the evolution of the joint entropy of $(x_2, x_3, \ldots, x_n) = \mathbf{x}_{2..n}$, i.e.,

$$H_{2..n} = -\int_{\mathbb{R}^{n-1}} \rho_{2..n} \log \rho_{2..n}\, d\mathbf{x}_{2..n},$$

where $\rho_{2..n} = \rho_{2..n}(x_2, \ldots, x_n) = \int_{\mathbb{R}} \rho\, dx_1$ is the joint PDF of $(x_2, x_3, \ldots, x_n)$. By integrating out x_1 from Eq. (3), we have

$$\begin{aligned}
&\frac{\partial \rho_{2..n}}{\partial t} + \frac{\partial}{\partial x_2}\int_{\mathbb{R}} \rho F_1\, dx_1 + \cdots + \frac{\partial}{\partial x_n}\int_{\mathbb{R}} \rho F_n\, dx_1 \\
&\quad = \frac{1}{2}\sum_{i=2}^{n}\sum_{j=2}^{n}\frac{\partial^2}{\partial x_i \partial x_j}\int_{\mathbb{R}} g_{ij}\rho\, dx_1.
\end{aligned} \tag{6}$$

Multiply $-(1 + \log \rho_{2..n})$, then integrate over $\mathbb{R}^{n-1}$. The first term is $dH_{2..n}/dt$. By taking advantage of the compactness assumption, the second term on the left-hand side results in

$$\begin{aligned}
&-\int_{\mathbb{R}^{n-1}}\left[(1 + \log \rho_{2..n})\frac{\partial}{\partial x_2}\left(\int_{\mathbb{R}} \rho F_2\, dx_1\right)\right] d\mathbf{x}_{2..n} \\
&= -\int_{\mathbb{R}^{n-1}} \log \rho_{2..n}\frac{\partial}{\partial x_2}\left(\int_{\mathbb{R}} \rho F_2\, dx_1\right) d\mathbf{x}_{2..n} \\
&= \int_{\mathbb{R}^{n-2}}\left\{\left[-\log \rho_{2..n} \cdot \int_{\mathbb{R}} \rho F_2\, dx_1\right]_{-\infty}^{\infty}\right. \\
&\quad \left. + \int_{\mathbb{R}}\left(\int_{\mathbb{R}} \rho F_2\, dx_1\right) \cdot \frac{\partial \log \rho_{2..n}}{\partial x_2}\, dx_2\right\} dx_3 \cdots dx_n \\
&= \int_{\mathbb{R}^n} \rho F_2 \frac{\partial \log \rho_{2..n}}{\partial x_2}\, d\mathbf{x} = E\left[F_2 \frac{\partial \log \rho_{2..n}}{\partial x_2}\right],
\end{aligned}$$

where E signifies mathematical expectation. Likewise, the third term through the nth term is

$$E\left[F_3 \frac{\partial \log \rho_{2..n}}{\partial x_3}\right], \ \ldots, \ E\left[F_n \frac{\partial \log \rho_{2..n}}{\partial x_n}\right].$$

On the right-hand side, the (i, j)th component is

$$\begin{aligned}
&-\int_{\mathbb{R}^{n-1}}\left[(1 + \log \rho_{2..n}) \cdot \frac{1}{2}\frac{\partial^2}{\partial x_i \partial x_j}\int_{\mathbb{R}} g_{ij}\rho\, dx_1\right] d\mathbf{x}_{2..n} \\
&= -\frac{1}{2}\int_{\mathbb{R}^{n-1}} \log \rho_{2..n} \cdot \frac{\partial^2}{\partial x_i \partial x_j}\left(\int_{\mathbb{R}} g_{ij}\rho\, dx_1\right) d\mathbf{x}_{2..n} \\
&= -\frac{1}{2}\int_{\mathbb{R}^{n-2}}\left\{\left[\log \rho_{2..n} \cdot \frac{\partial}{\partial x_j}\left(\int_{\mathbb{R}} g_{ij}\rho\, dx_1\right)\right]_{-\infty}^{\infty}\right. \\
&\quad \left. - \int_{\mathbb{R}} \frac{\partial \log \rho_{2..n}}{\partial x_i} \cdot \frac{\partial \int_{\mathbb{R}} g_{ij}\rho\, dx_1}{\partial x_j}\, dx_i\right\} dx_2, \ldots, dx_{i-1}\, dx_{i+1}, \ldots, dx_n \\
&= \frac{1}{2}\int_{\mathbb{R}^{n-1}} \frac{\partial \log \rho_{2..n}}{\partial x_i} \cdot \frac{\partial \int_{\mathbb{R}} g_{ij}\rho\, dx_1}{\partial x_j}\, d\mathbf{x}_{2..n} \\
&= \frac{1}{2}\int_{\mathbb{R}^{n-2}}\left\{\left[\frac{\partial \log \rho_{2..n}}{\partial x_i}\int_{\mathbb{R}} g_{ij}\rho\, dx_1\right]_{-\infty}^{\infty}\right. \\
&\quad \left. - \int_{\mathbb{R}}\left(\int_{R} g_{ij}\rho\, dx_1\right) \cdot \frac{\partial^2 \log \rho_{2..n}}{\partial x_i \partial x_j}\, dx_j\right\} dx_2, \ldots, dx_{j-1}\, dx_{j+1}, \ldots, dx_n \\
&= -\frac{1}{2}\int_{\mathbb{R}^n} \rho g_{ij}\frac{\partial^2 \log \rho_{2..n}}{\partial x_i \partial x_j}\, d\mathbf{x} = -\frac{1}{2}E\left[g_{ij}\frac{\partial^2 \log \rho_{2..n}}{\partial x_i \partial x_j}\right].
\end{aligned}$$

Putting the above together, we have

$$\frac{dH_{2..n}}{dt} = -\sum_{i=2}^{n} E\left[F_i \frac{\partial \log \rho_{2..n}}{\partial x_i}\right] - \frac{1}{2}\sum_{i=2}^{n}\sum_{j=2}^{n} E\left[g_{ij}\frac{\partial^2 \log \rho_{2..n}}{\partial x_i \partial x_j}\right]. \tag{7}$$

The evolution of $H_{2..n}$ contains two parts, one being the effect of x_1, another being the part with the effect of x_1 excluded. We denote the latter by $dH_{2..n,\not 1}/dt$; it can be found by instantaneously freezing x_1 as a parameter. For this purpose, we examine, on an infinitesimal interval $[t,\ t + \Delta t]$, a system modified from the original (1) by removing its first equation, i.e.,

$$\frac{dx_2}{dt} = F_2(x_1, x_2, \ldots, x_n; t) + \sum_{k=1}^{m} b_{2k}(x_1, x_2, \ldots, x_n; t)\dot{w}_k, \tag{8}$$

$$\frac{dx_3}{dt} = F_3(x_1, x_2, \ldots, x_n; t) + \sum_{k=1}^{m} b_{3k}(x_1, x_2, \ldots, x_n; t)\dot{w}_k, \tag{9}$$

$$\vdots \qquad\qquad \vdots$$

$$\frac{dx_n}{dt} = F_n(x_1, x_2, \ldots, x_n; t) + \sum_{k=1}^{m} b_{nk}(x_1, x_2, \ldots, x_n; t)\dot{w}_k. \tag{10}$$

Note here the F_is and b_{ik}s still have dependence on x_1, but now x_1 appears in the modified system as a parameter. Given the PDF of $\mathbf{x}$ at time t, we need to find the PDF of $\mathbf{x}_{\not 1}$ at time $t + \Delta t$. In Ref. 16, this is fulfilled by first constructing a mapping $\Phi : \mathbb{R}^{n-1} \to \mathbb{R}^{n-1}$, $\mathbf{x}_{\not 1}(t) \mapsto \mathbf{x}_{\not 1}(t + \Delta t)$, then studying the Frobenius–Perron operator of the modified system. Here, we choose an alternative approach. Note that on the interval $[t, t+\Delta t]$, there also exists a Fokker–Planck equation for the modified system,

$$\frac{\partial \rho_{\not 1}}{\partial t} + \frac{\partial F_2 \rho_{\not 1}}{\partial x_2} + \frac{\partial F_3 \rho_{\not 1}}{\partial x_3} + \cdots + \frac{\partial F_n \rho_{\not 1}}{\partial x_n} = \frac{1}{2} \sum_{i=2}^{n} \sum_{j=2}^{n} \frac{\partial^2 g_{ij} \rho_{\not 1}}{\partial x_i \partial x_j}, \quad (11)$$

$$\rho_{\not 1} = \rho_{2..n} \quad \text{at time t.} \quad (12)$$

Here, $g_{ij} = \sum_{k=1}^{m} b_{ik} b_{jk}$ is still as before; $\rho_{\not 1}$ means the joint PDF of $(x_2, \ldots, x_n)$ with x_1 frozen as a parameter. $\rho_{\not 1}$ is somehow similar to the conditional PDF of the former on the latter but not exactly as that. The subscript $\not 1$ signifies that x_1 is removed from the independent variables. Note this is quite different from $\rho_{2..n}$, which has no dependence on x_1 at all, but they are equal at time t.

Divide (11) by $\rho_{\not 1}$ to get

$$\frac{\partial \log \rho_{\not 1}}{\partial t} + \sum_{i=2}^{n} \frac{1}{\rho_{\not 1}} \frac{\partial F_i \rho_{\not 1}}{\partial x_i} = \frac{1}{2\rho_{\not 1}} \sum_{i=2}^{n} \sum_{j=2}^{n} \frac{\partial^2 g_{ij} \rho_{\not 1}}{\partial x_i \partial x_j}.$$

Discretizing and noticing that $\rho_(t) = \rho_{2..n}(t)$, we have

$$\log \rho(\mathbf{x}_{\not 1}; t+\Delta t) = \log \rho_{2..n}(\mathbf{x}_{\not 1}; t) - \Delta t \cdot \sum_{2}^{n} \frac{1}{\rho_{2..n}} \frac{\partial F_i \rho_{2..n}}{\partial x_i} + \frac{\Delta t}{2} \sum_{2}^{n} \sum_{2}^{n} \frac{1}{\rho_{2..n}} \frac{\partial^2 g_{ij} \rho_{2..n}}{\partial x_i \partial x_j} + o(\Delta t).$$

To arrive $dH_{2..n,\not 1}/dt$, we need to find $\log \rho(\mathbf{x}_{\not 1}(t+\Delta t); t+\Delta t)$. Using the Euler–Bernstein approximation,

$$\mathbf{x}_{\not 1}(t+\Delta t) = \mathbf{x}_{\not 1}(t) + \mathbf{F}_{\not 1} \Delta t + \mathbf{B}_{\not 1} \Delta \mathbf{w}, \quad (13)$$

where just like the notation $\mathbf{x}_{\not 1}$,

$$\mathbf{F}_{\not 1} = (F_2, \ldots, F_n)^T,$$

$$\mathbf{B}_{\not 1} = \begin{bmatrix} b_{21} & \cdots & b_{2m} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nm} \end{bmatrix},$$

$$\Delta \mathbf{w} = (\Delta w_1, \ldots, \Delta w_m)^T,$$

and $\Delta w_k \sim N(0, \Delta t)$, we have

$$\begin{aligned}
&\log(\rho_{\not 1}(\mathbf{x}_{\not 1}(t+\Delta t); t+\Delta t) \\
&= \log \rho_{2..n}(\mathbf{x}_{\not 1}(t) + \mathbf{F}_{\not 1}\Delta t + \mathbf{B}_{\not 1}\Delta \mathbf{w}; t) \\
&\quad - \Delta t \cdot \sum_{2}^{n} \frac{1}{\rho_{2..n}} \frac{\partial F_i \rho_{2..n}}{\partial x_i} + \frac{\Delta t}{2} \sum_{2}^{n} \sum_{2}^{n} \frac{1}{\rho_{2..n}} \frac{\partial^2 g_{ij} \rho_{2..n}}{\partial x_i \partial x_j} + o(\Delta t) \\
&= \log \rho_{2..n}(\mathbf{x}_{\not 1}(t)) + \sum_{i=2}^{n} \left[\frac{\partial \log \rho_{2..n}}{\partial x_i} \left(F_i \Delta t + \sum_{k=1}^{m} b_{ik} \Delta w_k \right) \right] \\
&\quad + \frac{1}{2} \cdot \sum_{i=2}^{n} \sum_{j=2}^{n} \left[\frac{\partial^2 \log \rho_{2..n}}{\partial x_i \partial x_j} \left(F_i \Delta t + \sum_{k=1}^{m} b_{ik} \Delta w_k \right) \right. \\
&\quad \left. \cdot \left(F_j \Delta t + \sum_{l=1}^{m} b_{jl} \Delta w_l \right) \right] - \Delta t \cdot \sum_{2}^{n} \frac{1}{\rho_{2..n}} \frac{\partial F_i \rho_{2..n}}{\partial x_i} \\
&\quad + \frac{\Delta t}{2} \sum_{2}^{n} \sum_{2}^{n} \frac{1}{\rho_{2..n}} \frac{\partial^2 g_{ij} \rho_{2..n}}{\partial x_i \partial x_j} + o(\Delta t).
\end{aligned}$$

Take mathematical expectation on both sides. The left-hand side is $-H_{2..n,\not 1}(t+\Delta t)$. By the Corollary III.I of Ref. 16, and noting $E\Delta w_k = 0$, $E\Delta w_k^2 = \Delta t$, and the fact that $\Delta \mathbf{w}$ are independent of $\mathbf{x}_{\not 1}$, we have

$$\begin{aligned}
-H_{2..n,\not 1}(t+\Delta t) &= -H_{2..n}(t) + \Delta t \cdot E \sum_{i=2}^{n} F_i \frac{\partial \log \rho_{2..n}}{\partial x_i} \\
&\quad + \frac{\Delta t}{2} \cdot E \sum_{i=2}^{n} \sum_{j=2}^{n} \sum_{k=1}^{m} \sum_{l=1}^{m} b_{ik} b_{jl} \delta_{kl} \frac{\partial^2 \log \rho_{2..n}}{\partial x_i \partial x_j} \\
&\quad - \Delta t \cdot E \sum_{2}^{n} \frac{1}{\rho_{2..n}} \frac{\partial F_i \rho_{2..n}}{\partial x_i} + \frac{\Delta t}{2} E \sum_{2}^{n} \sum_{2}^{n} \frac{1}{\rho_{2..n}} \frac{\partial^2 g_{ij} \rho_{2..n}}{\partial x_i \partial x_j} + o(\Delta t) \\
&= -H_{2..n}(t) + \Delta t \cdot E \sum_{i=2}^{n} F_i \frac{\partial \log \rho_{2..n}}{\partial x_i} \\
&\quad + \frac{\Delta t}{2} \cdot E \sum_{i=2}^{n} \sum_{j=2}^{n} g_{ij} \frac{\partial^2 \log \rho_{2..n}}{\partial x_i \partial x_j} - \Delta t \cdot E \sum_{2}^{n} \frac{1}{\rho_{2..n}} \frac{\partial F_i \rho_{2..n}}{\partial x_i} \\
&\quad + \frac{\Delta t}{2} E \sum_{2}^{n} \sum_{2}^{n} \frac{1}{\rho_{2..n}} \frac{\partial^2 g_{ij} \rho_{2..n}}{\partial x_i \partial x_j} + o(\Delta t).
\end{aligned}$$

So

$$\begin{aligned}
\frac{dH_{2..n,\not 1}}{dt} &= \lim_{\Delta t \to 0} \frac{H_{2..n,\not 1} - H_{2..n}(t)}{\Delta t} \\
&= -E \sum_{i=2}^{n} \left(F_i \frac{\partial \log \rho_{2..n}}{\partial x_i} - \frac{1}{\rho_{2..n}} \frac{\partial F_i \rho_{2..n}}{\partial x_i} \right) \\
&\quad - \frac{1}{2} E \sum_{i=2}^{n} \sum_{j=2}^{n} \left(g_{ij} \frac{\partial^2 \log \rho_{2..n}}{\partial x_i \partial x_j} + \frac{1}{\rho_{2..n}} \frac{\partial^2 g_{ij} \rho_{2..n}}{\partial x_i \partial x_j} \right).
\end{aligned}$$

Hence, the information flow from x_1 to $\mathbf{x}_{\not 1}$ is

$$\begin{aligned}
T_{1 \to 2..n} &= \frac{dH_{2..n}}{dt} - \frac{dH_{2..n,\not 1}}{dt} \\
&= -E \sum_{i=2}^{n} \left(F_i \frac{\partial \log \rho_{2..n}}{\partial x_i} \right) - \frac{1}{2} E \sum_{i=2}^{n} \sum_{j=2}^{n} \left(g_{ij} \frac{\partial^2 \log \rho_{2..n}}{\partial x_i \partial x_j} \right)
\end{aligned}$$

$$-E\sum_{i=2}^{n}\frac{\partial F_i}{\partial x_i}+\frac{1}{2}E\sum_{i=2}^{n}\sum_{j=2}^{n}\left(g_{ij}\frac{\partial^2\log\rho_{2..n}}{\partial x_i\partial x_j}+\frac{1}{\rho_{2..n}}\frac{\partial^2 g_{ij}\rho_{2..n}}{\partial x_i\partial x_j}\right)$$
$$=-E\left[\sum_{i=2}^{n}\frac{1}{\rho_{2..n}}\frac{\partial F_i\rho_{2..n}}{\partial x_i}\right]+\frac{1}{2}E\left[\sum_{i=2}^{n}\sum_{j=2}^{n}\frac{1}{\rho_{2..n}}\frac{\partial^2 g_{ij}\rho_{2..n}}{\partial x_i\partial x_j}\right].$$

□

There is a nice property regarding noise: when the noise is additive, the stochastic contribution to the information flow vanishes, as stated in the following corollary.

Corollary 2.1 *In (1), if $\mathbf{B}$ does not depend on $\mathbf{x}$, then*

$$T_{1\to 2..n}=-E\left[\sum_{i=2}^{n}\frac{1}{\rho_{2..n}}\frac{\partial F_i\rho_{2..n}}{\partial x_i}\right].$$

Proof. If b_{ij} is independent of $\mathbf{x}$, so is $g_{ij}=\sum_{k=1}^{m}b_{ik}b_{jk}$. Thus,

$$E\sum_i\sum_j\frac{1}{\rho_{2..n}}\frac{\partial^2 g_{ij}\rho_{2..n}}{\partial x_i\partial x_j}=\sum_i\sum_j g_{ij}\int_{R^n}\frac{\partial^2\rho_{2..n}}{\partial x_i\partial x_j}\,d\mathbf{x}$$
$$=\sum_i\sum_j g_{ij}\int_{\mathbb{R}^{n-1}}\frac{\int_{\mathbb{R}}\rho\,dx_1}{\rho_{2..n}}\frac{\partial^2\rho_{2..n}}{\partial x_i\partial x_j}\times dx_2\,dx_3\ldots dx_n$$
$$=\sum_i\sum_j g_{ij}\int_{\mathbb{R}^{n-1}}\frac{\partial^2\rho_{2..n}}{\partial x_i\partial x_j}\,dx_2\,dx_3,\ldots,dx_n,$$

which is zero by the compactness of ρ. □

Formula (2) can be verified with the particular situation in which the rest of the network does not depend on x_1. In this case, x_1 plays no role. Indeed, by following the procedure for the above corollary, one can prove that $T_{1\to 2..n}$ vanishes. So we have

Theorem 2.2 (Principle of nil causality) *If $\mathbf{F}_{\setminus}$ and $\mathbf{B}_{\setminus}$ are independent of x_1, $T_{1\to 2..n}=0$.*

A. Linear systems

Steered by a linear system, a Gaussian process is always Gaussian. In this case, the information flow can be greatly simplified.

Theorem 2.3 *In (1), suppose*

$$F_i=f_i+\sum_{j=1}^{n}a_{ij}x_j,\tag{14}$$

where f_i and a_{ij} are constants, and b_{ij} are also constants. Further suppose that initially $\mathbf{x}$ has a Gaussian distribution, then

$$T_{1\to 2..n}=\sum_{i=2}^{n}\left[\sum_{j=2}^{n}\sigma'_{ij}\left(\sum_{k=1}^{n}a_{ik}\sigma_{kj}\right)-a_{ii}\right].\tag{15}$$

In the equation, σ'_{ij} is the (i,j)th entry of $\begin{bmatrix}1 & 0\\ 0 & \boldsymbol{\Sigma}_{\setminus}^{-1}\end{bmatrix}$, where $\boldsymbol{\Sigma}$ is the covariance matrix, and $\boldsymbol{\Sigma}_{\setminus}$ is $\boldsymbol{\Sigma}$ with the first row and first column deleted.

Proof. In (2), by Corollary 2.1, the stochastic part (second term) can be ignored. Suppose the joint PDF of $\mathbf{x}$ has a form like

$$\rho(x_1,\ldots,x_n)=\frac{1}{\sqrt{(2\pi)^n\det\boldsymbol{\Sigma}}}e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}.\tag{16}$$

Then, it is easy to show

$$\rho_{2..n}(x_2,\ldots,x_n)=\frac{1}{\sqrt{(2\pi)^{n-1}\det\boldsymbol{\Sigma}_{\setminus}}}e^{-\frac{1}{2}(\mathbf{x}_{\setminus}-\boldsymbol{\mu}_{\setminus})^T\boldsymbol{\Sigma}_{\setminus}^{-1}(\mathbf{x}_{\setminus}-\boldsymbol{\mu}_{\setminus})},\tag{17}$$

where $\boldsymbol{\mu}_{\setminus}$ is the vector $\boldsymbol{\mu}$ with the first entry removed. For easy correspondence, we will still count the entries as those as numbered in $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}$. So

$$F_i\frac{\partial\log\rho_{2..n}}{\partial x_i}\left[f_i+\sum_{j=1}^{n}a_{ij}x_j\right]\frac{\partial}{\partial x_i}\left[-\frac{1}{2}(\mathbf{x}_{\setminus}-\boldsymbol{\mu}_{\setminus})^T\boldsymbol{\Sigma}_{\setminus}^{-1}(\mathbf{x}_{\setminus}-\boldsymbol{\mu}_{\setminus})\right]$$
$$=\left(f_i+\sum_{j=1}^{n}a_{ij}x_j\right)\sum_{j=2}^{n}\left(-\frac{\sigma'_{ij}+\sigma'_{ji}}{2}\right)\cdot(x_j-\mu_j).$$

Here, σ'_{ij} is the (i,j)th entry of the matrix $\boldsymbol{\Sigma}_{\setminus}^{-1}$. (Note here the entry indices run from 2 through n, not from 1 through n!) As $\boldsymbol{\Sigma}_{\setminus}$ is symmetric, so is $\boldsymbol{\Sigma}_{\setminus}^{-1}$, and hence $(\sigma'_{ij}+\sigma'_{ji})/2=\sigma'_{ij}$. So

$$-EF_i\frac{\partial\log\rho_{2..n}}{\partial x_i}=0-E\sum_{j=1}^{n}a_{ij}x_j\cdot\sum_{j=2}^{n}(-\sigma'_{ij})\cdot(x_j-\mu_j)$$
$$=E\sum_{k=1}^{n}a_{ik}(x_k-\mu_k)\cdot\sum_{j=2}^{n}\sigma'_{ij}(x_j-\mu_j)$$
$$=\sum_{k=1}^{n}\sum_{j=2}^{n}a_{ik}\sigma'_{ij}E(x_k-\mu_k)(x_j-\mu_j)$$
$$=\sum_{k=1}^{n}\sum_{j=2}^{n}a_{ik}\sigma'_{ij}\sigma_kj.$$

The other term

$$-E\sum_{i=2}^{n}\frac{\partial F_i}{\partial x_i}=-\sum_{i=2}^{n}a_{ii}.$$

Equation (15) follows by summing these two terms together. □

When $n=2$, the above formula can be further simplified. In fact,

$$T_{1\to 2}=a_{21}\sigma'_{22}\cdot\sigma_{12}+a_{22}\sigma'_{22}\cdot\sigma_{22}-a_{22}.$$

In this case, $\sigma'_{22}=1/\sigma_{22}$, so

$$T_{1\to 2}=a_{21}\frac{\sigma_{12}}{\sigma_{22}},$$

a special case of the formula $T_{i\to j}=a_{ji}\sigma_{ij}/\sigma_{jj}$ as obtained before in Ref. 16.

From above one can see that, in general,

$$T_{1\to 2..n} \neq \sum_{j=2}^{n} T_{1\to j}. \tag{18}$$

This must also hold for general nonlinear systems; anyhow, linear systems form a particular subset. That is, the macrostate of a network is not just a simple addition of the individual states. The equality can hold only when the n components are uncorrelated, i.e., when $\mathbf{\Sigma}$ is a diagonal matrix, and hence $\sigma'_{ii} = 1/\sigma_i i$ and $\sigma'_{ij} = 0$ for $i \neq j$. Indeed, in this case, the n components are just independent units; they do not form a network.

B. The impact of $\mathbf{x}_{\!1}$ on x_1

We know that information flow or causality is asymmetric between two entities, that is, the contribution of x_1 to the rest of the network is generally different from that the other way around. For late reference, we here briefly present the result of the information flow from $\mathbf{x}_{\!1}$ to x_1, though it is not needed in this study.

From Ref. 22,

$$\frac{dH_1}{dt} = -E\left[F_1 \frac{\partial \log \rho_1}{\partial x_1}\right] - \frac{1}{2}E\left[g_{11}\frac{\partial^2 \log \rho_1}{\partial x_1^2}\right]. \tag{19}$$

Now if we modify the system on the infinitesimal interval $[t + \Delta t]$ by freezing $(x_2, x_3, \ldots, x_n)$, and follow the above derivation, we finally arrive at the time rate of change of the marginal entropy of x_1 with the effect of $(x_2, x_3, \ldots, x_n)$ excluded is

$$\frac{dH_{1\backslash\!\!\!2..n}}{dt} = E\left(\frac{\partial F_1}{\partial x_1}\right) - \frac{1}{2}E\left(g_{11}\frac{\partial^2 \log \rho_1}{\partial x_1^2}\right) - \frac{1}{2}E\left(\frac{1}{\rho_1}\frac{\partial^2 g_{11}\rho_1}{\partial x_1^2}\right). \tag{20}$$

So the information flow from $\mathbf{x}_{\!1}$ to x_1 is

$$\begin{aligned} T_{2..n\to 1} &= \frac{dH_1}{dt} - \frac{dH_{1\backslash\!\!\!2..n}}{dt} \\ &= -E\left[F_1 \frac{\partial \log \rho_1}{\partial x_1} + \frac{\partial F_1}{\partial x_1}\right] + \frac{1}{2}E\left[\frac{1}{\rho_1}\frac{\partial^2 g_{11}\rho_1}{\partial x_1^2}\right] \\ &= -E\left[\frac{1}{\rho_1}\frac{\partial F_1 \rho_1}{\partial x_1}\right] + \frac{1}{2}E\left[\frac{1}{\rho_1}\frac{\partial^2 g_{11}\rho_1}{\partial x_1^2}\right]. \end{aligned} \tag{21}$$

A seemingly surprising observation is that this is precisely the same in form as that for 2D systems (see Ref. 22), although here the dimensionality can be larger than 2. This does make sense, as we are splitting the system into two subsystems, one with x_1, another with a collection of $n-1$ units. In the meantime, this generally differs in form from those individual information flow formulas for systems with $n > 2$ (see Ref. 16).

III. MAXIMUM LIKELIHOOD ESTIMATION

Given a system like (1), we can rigorously evaluate the information flows among the components. Now suppose, instead of the system, what we have are just n time series with K steps, $K \gg n$, $\{x_1(k)\}, \{x_2(k)\}, \ldots, \{x_n(k)\}$. We can estimate the system from the series and then apply the information flow formula to fulfill the task.

Assume a linear model as shown above, and assume $m = 1$. Following Ref. 15, the maximum likelihood estimator of a_{ij} is equal to the least-square solution of the following over-determined problem:

$$\begin{pmatrix} 1 & x_1(1) & x_2(1) & \cdots & x_n(1) \\ 1 & x_1(2) & x_2(2) & \cdots & x_n(2) \\ 1 & x_1(3) & x_2(3) & \cdots & x_n(3) \\ \vdots & \vdots & \vdots\ddots & \vdots & \\ 1 & x_1(K) & x_2(K) & \cdots & x_n(K) \end{pmatrix} \begin{pmatrix} f_i \\ a_{i1} \\ a_{i2} \\ \vdots \\ a_{in} \end{pmatrix} = \begin{pmatrix} \dot{x}_i(1) \\ \dot{x}_i(2) \\ \dot{x}_i(3) \\ \vdots \\ \dot{x}_i(K) \end{pmatrix},$$

where $\dot{x}_i(k) = (x_i(k+1) - x_i(k))/\Delta t$ (Δt is the time stepsize), for $i = 1, 2, \ldots, n$, $k = 1, \ldots, K$. Use overbar to denote the time mean over the K steps. The above equation is

$$\begin{pmatrix} 1 & \bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_n \\ 0 & x_1(2) - \bar{x}_1 & x_2(2) - \bar{x}_2 & \cdots & x_n(2) - \bar{x}_n \\ 0 & x_1(3) - \bar{x}_1 & x_2(3) - \bar{x}_2 & \cdots & x_n(3) - \bar{x}_n \\ \vdots & \vdots & \vdots\ddots & \vdots & \\ 0 & x_1(K) - \bar{x}_1 & x_2(K) - \bar{x}_2 & \cdots & x_n(K) - \bar{x}_n \end{pmatrix} \begin{pmatrix} f_i \\ a_{i1} \\ a_{i2} \\ \vdots \\ a_{in} \end{pmatrix} = \begin{pmatrix} \bar{\dot{x}}_i \\ \dot{x}_i(2) - \bar{\dot{x}}_i \\ \dot{x}_i(3) - \bar{\dot{x}}_i \\ \vdots \\ \dot{x}_i(K) - \bar{\dot{x}}_i \end{pmatrix}.$$

Denote by $\mathbf{R}$ the matrix

$$\begin{pmatrix} x_1(2) - \bar{x}_1 & x_2(2) - \bar{x}_2 & \cdots & x_n(2) - \bar{x}_n \\ \vdots & \vdots & \vdots\ddots & \vdots \\ x_1(K) - \bar{x}_1 & x_2(K) - \bar{x}_2 & \cdots & x_n(K) - \bar{x}_n \end{pmatrix},$$

$\mathbf{s}$ is the vector $(x_i(2) - \bar{\dot{x}}_i, \ldots, x_i(K) - \bar{\dot{x}}_i)^T$, and $\mathbf{a}_i$ is the row vector $(a_{i1}, \ldots, a_{in})^T$. Then, $\mathbf{R}\mathbf{a}_i = \mathbf{s}$. The least-square solution of $\mathbf{a}_i$, $\hat{\mathbf{a}}_i$, solves

$$\mathbf{R}^T\mathbf{R}\hat{\mathbf{a}}_i = \mathbf{R}^T\mathbf{s}.$$

Note $\mathbf{R}^T\mathbf{R}$ is $K\mathbf{C}$, where $\mathbf{C}$ is the covariance matrix. So

$$\begin{pmatrix} \hat{a}_{i1} \\ \hat{a}_{i2} \\ \vdots \\ \hat{a}_{in} \end{pmatrix} = \mathbf{C}^{-1} \begin{pmatrix} c_{1,di} \\ c_{2,di} \\ \vdots \\ c_{n,di} \end{pmatrix}, \tag{22}$$

where $c_{j,di}$ is the covariance between the series $\{x_j(k)\}$ and $\{(x_i(k+1) - x_i(k))/\Delta t\}$.

So finally, the maximum likelihood estimator (mle) of $T_{1\to 2..n}$ is

$$\hat{T}_{1\to 2..n} = \sum_{i=2}^{n}\left[\sum_{j=2}^{n} c'_{ij}\left(\sum_{k=1}^{n} \hat{a}_{ik}c_{kj}\right) - \hat{a}_{ii}\right], \tag{23}$$

where c'_{ij} is the (i,j)th entry of $\tilde{\mathbf{C}}^{-1}$, and

$$\tilde{\mathbf{C}} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & c_{22} & c_{23} & \cdots & c_{2n} \\ 0 & c_{23} & c_{33} & \cdots & c_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & c_{2n} & c_{3n} & \cdots & c_{nn} \end{pmatrix}. \tag{24}$$

Denoting by $\hat{\mathbf{A}}$ the matrix with entries $(\hat{a}_{ij})$, Eq. (23) can be more succinctly written as

$$\hat{T}_{1\to 2..n} = Tr_{\setminus}\left[\tilde{\mathbf{C}}^{-1}(\hat{\mathbf{A}}\mathbf{C})^T\right] - Tr_{\setminus}\left[\hat{\mathbf{A}}\right]. \tag{25}$$

Here, $Tr_{\setminus}$ means the trace of a matrix with the first term removed. That is, it is defined such that, for matrix $\mathbf{Q}$,

$$Tr_{\setminus}\mathbf{Q} = Tr\mathbf{Q} - Q(1,1).$$

Note that this is made possible by the form of $\tilde{\mathbf{C}}$ (with its special form in 1st row and 1st column); otherwise, the trace of the product of two matrices, say, $\mathbf{P}_{n\times n}\mathbf{Q}_{n\times n} \equiv \mathbf{R}_{n\times n}$, is generally not equal to $Tr[\mathbf{P}(2:n,2:n)\mathbf{Q}(2:n,2:n)] + R(1,1)$.

IV. APPLICATION TO A NETWORK OF COUPLED STUART–LANDAU OSCILLATORS

In this section, we put (23) to application to a network with N nodes, each made of a Stuart–Landau oscillator.[24] This has been used to model many biological networks for phenomena such as circadian rhythms, synchronized neuronal firing, and spatiotemporal activity in the heart and the brain (see Ref. 14 for more examples). Other similar synchronized networks include those made of Rössler oscillators, e.g., Ref. 23. For the purpose of demonstration, here a small number $N = 6$ is chosen. Let the complex state variable of the jth oscillator be z_j. It is defined as (see, e.g., Refs. 14 and 25)

$$\frac{dz_j}{dt} = \left(\alpha_j + i\Omega_j - |z_j|^2\right) z_j + \frac{K}{N}\sum_{k=1}^{N} \Lambda_{jk}(z_k - z_j) + \nu \dot{w}_j,$$
$$j = 1, \ldots, N, \tag{26}$$

where $i = \sqrt{-1}$, Ω_j are the frequencies, α_j are the control parameters, and (Λ) is the adjacency matrix. Here, the coupling coefficient K is chosen to be 1. The notation generally follows that in Ref. 14; the difference lies in an Ω varying oscillator by oscillator, and an additional stochastic term $\nu\dot{w}_j$, where w_j is a standard Wiener process and ν is the stochastic perturbation amplitude. We add some weak stochasticity for convenience (see below). If $K = 0$ and $\nu = 0$, the oscillators are Stuart–Landau oscillators; a positive α_j yields an oscillating state, whereas a negative α_j disables the oscillator. In this study, $K = 1$, $\Omega_j = j/2$, $j = 1, \ldots, N$, are fixed throughout. (We choose these Ω_j simply to make them different. One may feel free to choose other values.) α_j may be 1 or -3, depending on whether z_j is activated or switched off. The adjacency matrix is chosen such that $\Lambda_{2k} = \Lambda_{k2} = 0$, $k = 1, 3, 4$; $\Lambda_{1k} = \Lambda_{k1} = 0$, $k = 4, 6$, and for all other (j,k), $\Lambda_{jk} = 1$. The resulting network is sketched in Fig. 1. As can be seen, z_5 is a highly connected node, or hub, as defined in Sec. I; second to it is z_6. z_1 and z_2 are two sparsely connected nodes.

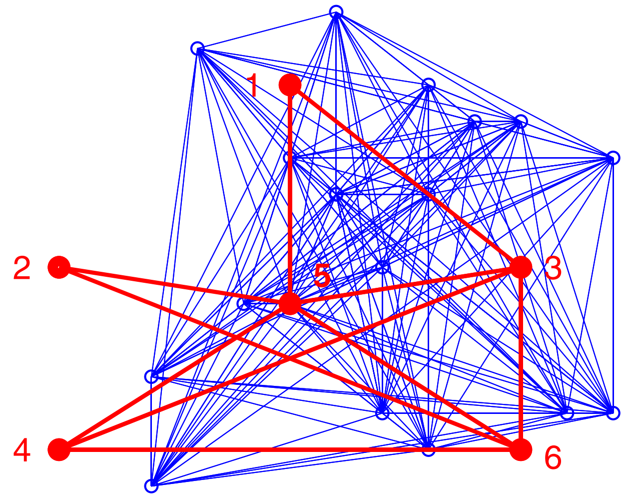

FIG. 1. A schematic of the network of coupled oscillators. For the sake of clarity, in this study, only the six-node (red) subnetwork is considered.

Equation (26) is discretized and solved using the second order Runge–Kutta scheme. The system is initialized with random values, integrated forward with a time stepsize of $\Delta t = 0.1$. Without coupling, the individual oscillators operate on their own, each exhibiting a periodic series with a distinct frequency. Shown in Fig. 2(a) are the active (solid) and inactive (dashed) modes for z_1 when $\nu = 0$. Figure 2(b) displays the corresponding cases when $\nu = 0.1$. We need this slightly perturbed system because, as seen in Fig. 2(a), the trajectories are too regular (periodic), only leaving on the Poincaré plane one point. In other words, they contain no information, making the information flow problem singular. Recently, it is found that this is actually an extreme case,[27] and hence can be handled by perturbing the system slightly with weak stochasticity. (In real systems, noises are ubiquitous.) Figure 2(b) approximates well its deterministic case [Fig. 2(a)] except for some weak ripples superimposed on the curves. So it is reasonable to believe that the addition of the weak perturbation can be used to compute the information flow for the original system.

Figure 3 shows the time series of the six coupled oscillators. In (a), all of them are on. As seen, though the frequencies Ω_j differ, the six oscillators work together to produce completely synchronized oscillations (see Ref. 26 for optimum synchronizations). To assess the importance of a node, a usual practice is to disable it and observe the response. In Figs. 3(b)–3(g), the respective responses when z_1–z_6 are disabled, respectively, are shown. With only one node failure, the network is still alive. But one can see that the impact of z_5 is significantly larger than others, while that from z_1 is by far the least. In Fig. 3(h), when z_5 and z_2 are disabled, then the entire network gradually dies, though in this case $\alpha_1, \alpha_3, \alpha_4$, and α_6 are still positive.

As mentioned in the Introduction, the above assessment by preferential removal of designated node(s) may not be feasible for many networks in nature, neuronal networks in particular. Now use formula (23) to estimate the information flow from the individual oscillators to the network. To begin, note that each z_j actually has

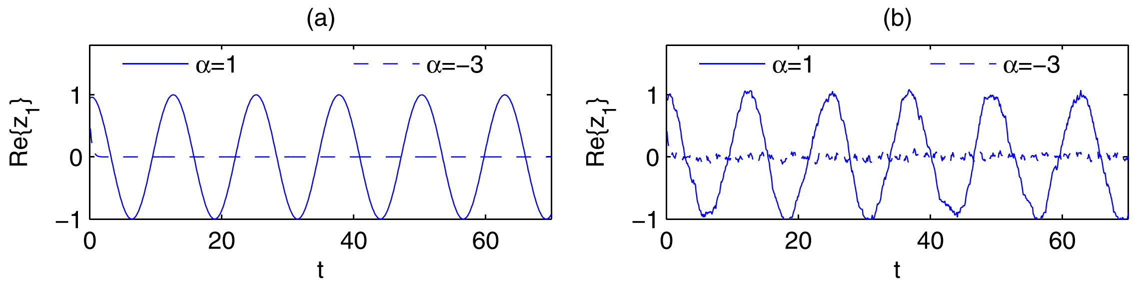

FIG. 2. Time series of a single oscillator z_1 without coupling ($K = 0$). (a) No noise; (b) weak stochasticity applied ($\nu = 0.1$). Only the real parts are drawn.

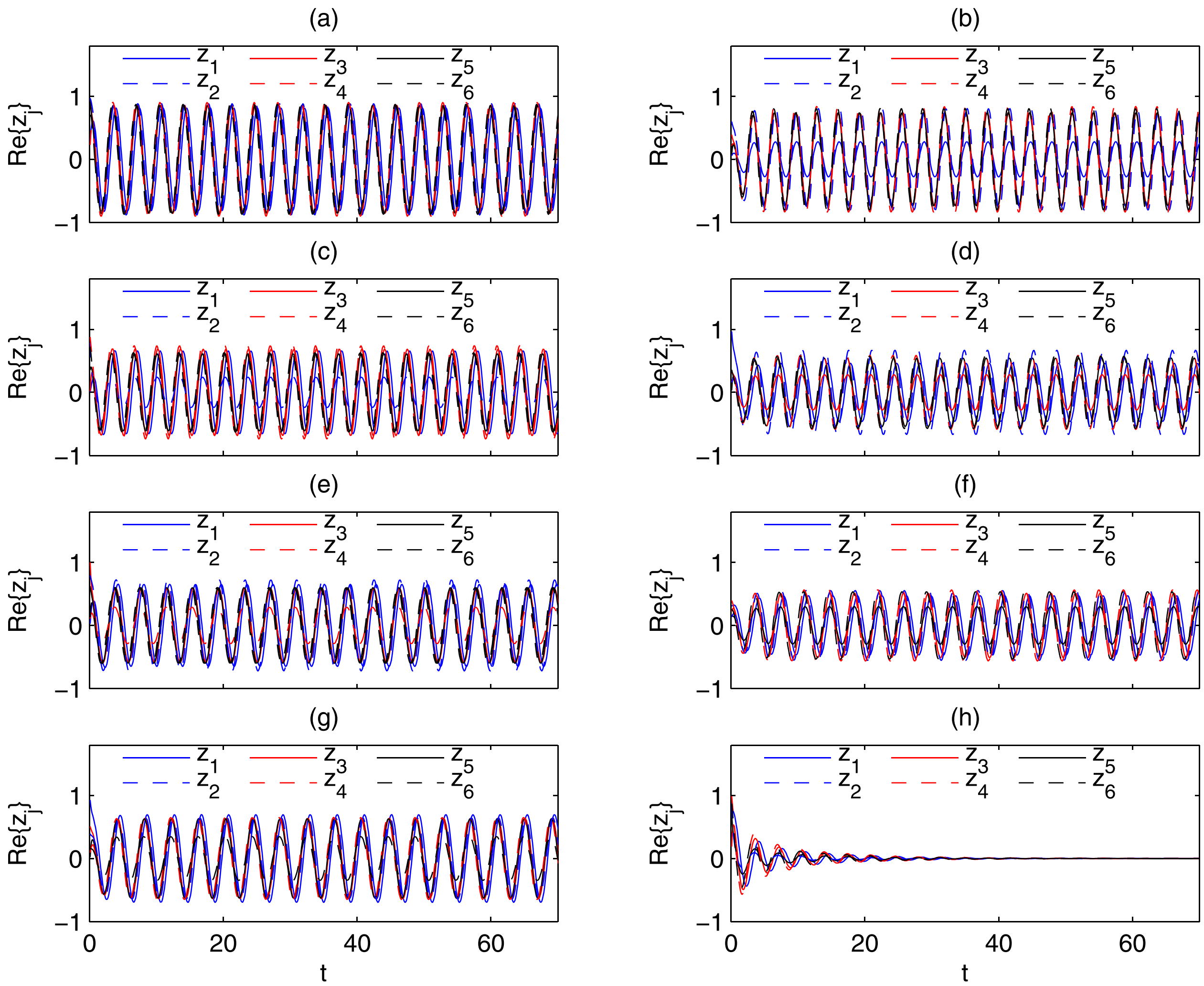

FIG. 3. Time series of the six coupled oscillators z_j (only the real parts of z_j are drawn). (a) All oscillators are active; (b) z_1 inactive (all others are active; same below); (c) z_2 inactive; (d) z_3 inactive; (e) z_4 inactive; (f) z_5 inactive; (g) z_6 inactive; (h) both z_2 and z_5 are inactive.

two components; so they should be taken as two time series. That is, the dynamical system has a dimensionality of $2 \times N$. The remaining computation is straightforward. We generate series with 5000 steps, with the first 100 steps discarded (to ensure stationarity). The computed results are (units in nats per unit time; values may differ slightly due to the random initialization)

$\hat{T}_{1\to network}$	$\hat{T}_{2\to network}$	$\hat{T}_{3\to network}$	$\hat{T}_{4\to network}$	$\hat{T}_{5\to network}$	$\hat{T}_{6\to network}$
0.65	1.00	2.31	2.30	2.96	2.93

By comparison, z_5 and z_6 are most important; second to them are z_3 and z_4. z_1 and z_2 are the least important. The result is just as that as illustrated in Fig. 3. From our common intuition, this makes sense too. As we can check from Fig. 1, z_5 and z_6 are the hubs, whereas z_1 and z_2 are sparsely connected.

As an aside, we have computed the summed pairwise information flows $\sum_j T_{1\to j}$, $\sum_j T_{2\to j}, \ldots, \sum_j T_{6\to j}$, where the summation for a node is over all the indices except that of the node itself. The results are, respectively,

$$1.62,\ 1.51,\ 3.35,\ 2.21,\ 2.97,\ 2.18.$$

They are quite different from the computed $T_{1\to network}, \ldots, T_{6\to network}$ as shown above. Particularly, the result does not correctly assign the largest value to the hub, i.e., node 5. This from one aspect testifies to the validity of (18), i.e., that the macrostate of a network is not just a simple addition of the individual states.

The above network does not have local weights and the links are not directed. If there exist directed links and/or localized weights (e.g., Ref. 26) in the network, hubs need not always be the most crucial units. To see this, let $\Lambda_{52} = 10$ and $\Lambda_{62} = 5$. The computed result is tabulated as follows:

$\hat{T}_{1\to network}$	$\hat{T}_{2\to network}$	$\hat{T}_{3\to network}$	$\hat{T}_{4\to network}$	$\hat{T}_{5\to network}$	$\hat{T}_{6\to network}$
0.35	4.11	2.06	1.95	2.48	0.74

So now the most important node is z_2, though it is sparsely connected! Also, the impact from z_6 has been greatly reduced.

Also, as an aside, we have computed for this case $\sum_j T_{1\to j}$, $\sum_j T_{2\to j}, \ldots, \sum_j T_{6\to j}$, which by computation are, respectively,

$$0.99,\ 6.52,\ 1.74,\ 1.09,\ 0.85,\ 2.20.$$

Again, they are quite different from the computed total information flows as tabulated above.

To see whether the total information flow correctly measures the importance of a node, we do the node deterioration experiments again. Indeed, if z_2 is deteriorated or suppressed, the whole network becomes silent, as shown in Fig. 4(c). The result is hence validated.

V. SUMMARY

A quantitative evaluation of the contribution of individual units in producing the collective behavior of a complex network is important in that it allows us to gain an understanding of which units determine the vulnerability of the network. In this study, we show that a natural measure is the information flow from the unit in concern to the entire network. A formula is derived, and its maximum likelihood estimator is provided. The results are summarized henceforth for easy reference.

For a network modeled with an n-dimensional continuous-time dynamical system,

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}, t) + \mathbf{B}(\mathbf{x}, t)\dot{\mathbf{w}},$$

the information flow from node x_1 to the network $x_2, x_3, \ldots, x_n$ is

$$T_{1\to 2..n} = -E\left[\sum_{i=2}^{n} \frac{1}{\rho_{2..n}} \frac{\partial F_i \rho_{2..n}}{\partial x_i}\right] + \frac{1}{2}E\left[\sum_{i=2}^{n}\sum_{j=2}^{n} \frac{1}{\rho_{2..n}} \frac{\partial^2 g_{ij}\rho_{2..n}}{\partial x_i \partial x_j}\right].$$

When only time series are available, under the assumption of linearity, the maximum likelihood estimator of $T_{1\to 2..n}$ is

$$\hat{T}_{1\to 2..n} = Tr_{\setminus}\left[\tilde{\mathbf{C}}^{-1}(\hat{\mathbf{A}}\mathbf{C})^T\right] - Tr_{\setminus}\left[\hat{\mathbf{A}}\right].$$

In the equation, $Tr_{\setminus}$ means the trace of a matrix with the first term removed, $\mathbf{C} = (c_{ij})$ is the covariance matrix, $\tilde{\mathbf{C}}$ is equal to $\mathbf{C}$ except $\tilde{c}_{1,1} = 1$, $\tilde{c}_{j,1} = \tilde{c}_{1,j} = 0$, $j = 2, 3, \ldots, n$. $\hat{\mathbf{A}} = (\hat{a}_{ij})$ has entries

$$\begin{pmatrix} \hat{a}_{i1} \\ \hat{a}_{i2} \\ \vdots \\ \hat{a}_{in} \end{pmatrix} = \mathbf{C}^{-1} \begin{pmatrix} c_{1,di} \\ c_{2,di} \\ \vdots \\ c_{n,di} \end{pmatrix}, \quad i = 1, 2, \ldots, n,$$

where $c_{j,di}$ is the covariance between the series $\{x_j(k)\}$ and $\{(x_i(k+1) - x_i(k))/\Delta t\}$. Observe that this "cumulative information flow" is not equal to the sum of the information flows to other individual units, reflecting the collective phenomenon that a group is not the addition of the individual members.

The above formula has been put to application to a network consisting of Stuart–Landau oscillators. It is shown that the node with largest information flow is indeed most crucial for the network. Its deterioration or suppression will cause the whole network to cease to function. An observation is depending on the topology, such a node may not be a hub, i.e., the node with high degree; on the contrary, it could be some sparsely connected, low-degree node. This study is expected to be useful in identifying clues to the mystery why initially small shocks at some nodes may trigger a massive, global shutdown of the entire network.

Apart from that considered in the coupled oscillator network example, other node interventions exist. In many biological and technological networks, node failure, attack or deletion, refers to that the node in question is isolated from the network, i.e., the edges to and from it are deleted. In this case, the effect may not appear as a suppression of the oscillation; the oscillation may be still there, but the collective pattern is changed to a pattern without synchronization, a pattern with much higher frequency, etc. The change in pattern is not easy to quantify simply by visual inspection. That is the reason why we chose the above example for validation; anyhow, "dead" and "alive" are two states that make the easiest situation for

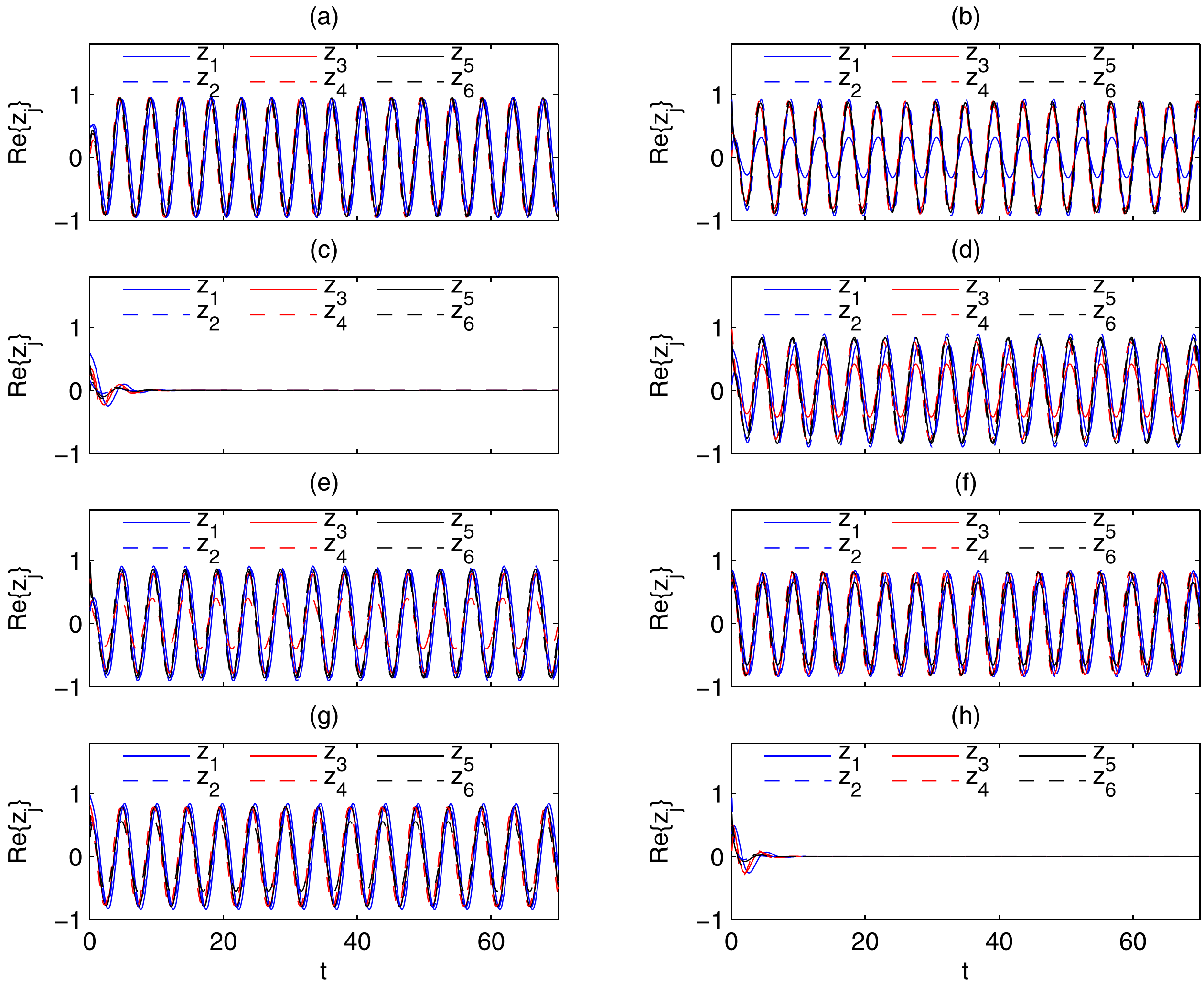

FIG. 4. As in Fig. 3, but with weighted and directed links.

a naked eye to distinguish. Nonetheless, the information flow formalism in this study is generic and hence equally holds for these different situations. This has been testified in many real world applications in previous studies. For information flow and causality and their applications to the diverse real world problems such as global climate change, neuroscience, financial economics, and El Niño, among others, see a brief review in Ref. 17, Sec. 2.

It should be made clear that, if intervention of a node shuts down the network, the information flow from that node must be large. Particularly, if it is the only node that has such an impact, then it must have the largest information flow. However, conversely, intervention of the node with the largest information flow does not guarantee the suppression of the network. In that case, the information flow may not be large enough relative to those from other nodes within the same network. Nonetheless, how large the information flow should be for a node to have such a global impact? This is a rather basic question. It refers to how information flow change by the intervention can be linked to the changes in the network dynamics. This issue, among many others, are to be investigated in future studies.

ACKNOWLEDGMENTS

The paper was written in February 2020 at the kind invitation of Dr. Rio Jie Sun to whom the author owes a debt of gratitude. The suggestions of an anonymous reviewer are appreciated. This study was supported by the National Science Foundation of China (Grant No. 41975064) and the 2015 Jiangsu Program for Innovation Research and Entrepreneurship Groups.

DATA AVAILABILITY

The data and codes that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] P. Erdös and A. Rényi, "On the evolution of random graphs," Publ. Math. Inst. Hung. Acad. Sci. **5**, 17–61 (1960).
[2] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," Nature **393**, 440–442 (1998).
[3] F. Liljeros, C. R. Edling, L. A. N. Amaral, H. E. Stanley, and Y. Aberg, "The web of human sexual contacts," Nature **411**, 907–908 (2001).
[4] R. Albert, "Scale-free networks in cell biology," J. Cell Sci. **118**, 4947–4957 (2005).
[5] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet* (Cambridge University Press, London, 2004).
[6] M. Newman, A.-L. Barabási, and D. J. Watts, *The Structure and Dynamics of Networks* (Princeton University Press, Princeton, NJ, 2006).
[7] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, "Complex networks: Structure and dynamics," Phys. Rep. **424**, 175–308 (2006).
[8] D. L. Alderson and J. C. Doyle, "Can complexity science support the engineering of critical network infrastructures?," in *Proceedings of the 2007 IEEE International Conference on Systems, Man and Cybernetics, Montreal, Canada, October 7–10* (IEEE, 2007).
[9] R. Albert, H. Jeong, and A.-I. Barabási, "Error and attach tolerance of complex networks," Nature **406**, 378–382 (2000).
[10] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, "Network robustness and fragility: Percolation on random graphs," Phys. Rev. Lett. **85**, 5468–5471 (2000).
[11] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, "Breakdown of the Internet under intentional attack," Phys. Rev. Lett. **86**, 3682–3685 (2001).
[12] A. E. Motter and Y.-C. Lai, "Cascade-based attacks on complex networks," Phys. Rev. E **66**, 065102 (2002).
[13] P. Fairley, "The unruly power grid," IEEE Spectrum **41**, 22–27 (2004).
[14] G. Tanaka, K. Morino, and K. Aihara, "Dynamical robustness in complex networks: The crucial role of low-degree nodes," Sci. Rep. **2**, 47 (2012).
[15] X. S. Liang, "Unraveling the cause-effect relation between time series," Phys. Rev. E **90**, 052150 (2014).
[16] X. S. Liang, "Information flow and causality as rigorous notions *ab initio*," Phys. Rev. E **94**, 052201 (2016).
[17] X. S. Liang, "Normalized multivariate time series causality analysis and causal graph reconstruction," Entropy **23**, 679 (2021).
[18] A. Aleta and Y. Moreno, "The dynamics of collective social behavior in a crowd controlled game," EPJ Data Sci. **8**, 425 (2019).
[19] R. F. Baumeister, S. E. Ainsworth, and K. D. Vohs, "Are groups more or less than the sum of their members? The moderating role of individual identification," Behav. Brain Sci. **39**, 11 (2016).
[20] T. W. Malone, R. Laubacher, and C. Dellarocas, "The collective intelligence genome," MIT Sloan Manag. Rev. **51**(3), 21 (2010).
[21] W. Mason and D. J. Watts, "Financial incentives and the performance of crowds," ACM SIGKDD Explor. Newsl. **11**(2), 100–108 (2010).
[22] X. S. Liang, "Information flow within stochastic systems," Phys. Rev. E **78**, 031113 (2008).
[23] M. Paluš, A. Krakovská, J. Jakubík, and M. Chvosteková, "Causality, dynamical systems and the arrow of time," Chaos **28**, 075307 (2018).
[24] S. Strogatz, "Exploring complex networks," Nature **410**, 268–276 (2001).
[25] H. Daido and K. Nakanishi, "Aging transition and universal scaling in oscillator networks," Phys. Rev. Lett. **93**(10), 104101 (2004).
[26] P. S. Skardal, D. Taylor, and J. Sun, "Synchronization of network-coupled oscillators with uncertain dynamics," SIAM J. Appl. Math. **79**, 2409–2433 (2019).
[27] X. S. Liang and X. Yang, "A note on causality versus correlation in an extreme situation," Entropy **23**, 316 (2021).