Causation entropy identifies indirect influences, dominance of neighbors and anticipatory couplings

Jie Sun *, Erik M. Bollt

Department of Mathematics, Clarkson University, Potsdam, NY 13699-5815, United States

HIGHLIGHTS

- We demonstrated the success of transfer entropy in detecting information flow in two oscillators.
- We explored the limitations of transfer entropy for causality inference in various scenarios.
- We developed causation entropy for more reliable inference of causality in networks of coupled oscillators.

ARTICLE INFO

ABSTRACT

Inference of causality is central in nonlinear time series analysis and science in general. A popular approach to infer causality between two processes is to measure the information flow between them in terms of transfer entropy. Using dynamics of coupled oscillator networks, we show that although transfer entropy can successfully detect information flow in two processes, it often results in erroneous identification of network connections under the presence of indirect interactions, dominance of neighbors, or anticipatory couplings. Such effects are found to be profound for time-dependent networks. To overcome these limitations, we develop a measure called *causation entropy* and show that its application can lead to reliable identification of true couplings.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

The long-standing puzzle of "what causes what", formally known as the problem of causality inference, is challenging yet central in science [1,2]. Understanding causal relationship between events has important implications in a wide range of areas including as examples social perception [3], epidemiology [4], and econometrics [5]. It is the reliable inference of causality that allows one to untangle complex causal interactions, make predictions, and ultimately design intervention strategies.

Traditional approach of inferring causality between two stochastic processes is to perform the Granger causality test [6]. A main limitation of this test is that it can only provide information about *linear* dependence between two processes and therefore fails to capture intrinsic nonlinearities that are common in real-world systems. To overcome this difficulty, Schreiber developed the concept of *transfer entropy* between two processes [7]. Transfer entropy measures the uncertainty reduction in inferring the future state of a process by learning the (current and past) states of

another process. Being an asymmetric measure by design, transfer entropy is often used to infer the directionality of information flow and further the causality between two processes [8,9]. Recently, it became increasingly popular to use transfer entropy for causality inference in networks of neurons [10,11] and in coupled dynamical systems with parameter mismatches [12], anticipatory couplings [13], and time delays [14]. However, despite the overwhelming number of proposed applications, a clear interpretation of the resulting relationship inferred by transfer entropy is lacking.

In this paper, we study information transfer in the dynamics of small-scale coupled oscillator networks. We show by several examples that causal relationship inferred by transfer entropy is often misleading when the underlying system contains indirect connections, dominance of neighboring dynamics, or anticipatory couplings. To account for these effects, we develop a measure called *causation entropy (CSE)* and show that its appropriate application reveals true coupling structures of the underlying dynamics.

2. Information theory and dynamical systems

In this section we introduce the mathematical tools used in this study, which include elements from both dynamical systems and information theory.

---

* Corresponding author. Tel.: +1 3152682388.
E-mail addresses: sunj@clarkson.edu (J. Sun), bolltem@clarkson.edu
(E.M. Bollt).

2.1. Dynamical system as a stochastic process

Our focus of this paper is on discrete dynamical systems of the form

$$x_{t+1} = f(x_t), \tag{1}$$

where $x_t \in \mathcal{D} \subset \mathbb{R}^m$ is the state variable and $f : \mathcal{D} \to \mathcal{D}$ is the dynamic rule of the system. A trajectory (or orbit) $\{x_t\}$ of Eq. (1) naturally represents a time series. For a continuous dynamical system $\dot{x} = f(x)$, a time series can be obtained by sampling its continuous trajectory at discrete time points. The time points are often chosen to spread uniformly in time or to be the time instances at which the trajectory intersects a given manifold that is transversal to the trajectory, called a *Poincaré section* [15].

A natural bridge between dynamical systems and information theory is the formulation of symbolic dynamics, which requires discretization of the phase space. In particular, a finite *topological partition* $P = \{P_1, \ldots, P_m\}$ of the phase space $\mathcal{D}$ is a collection of pairwise disjoint sets in $\mathcal{D}$ whose union is $\mathcal{D}$ [16]. Defining the associated set of symbols $\Omega = \{1, 2, \ldots, m\}$, one can transform a trajectory $\{x_t\}$ into a *symbolic sequence* $\{s_t\}$, where s_t is defined by [17,18]

$$x_t \in P_i \subset \mathcal{D} \Rightarrow s_t = i \in \Omega. \tag{2}$$

Viewing Ω as the sample space, the symbolic sequence $\{s_t\}$ can be seen as a time series of a stochastic process. Define a probability measure over the partition P, as

$$\mu : P \to \mathbb{R}. \tag{3}$$

If μ is *invariant* under the dynamics, then [19,20]

$$\text{Prob}(s_t = i) = \mu(i), \quad \forall i \in \Omega, \ t \in \mathbb{R}. \tag{4}$$

A partition P is called a *Markov partition* if it gives rise to a stochastic process that is Markovian, i.e., future states of the process depend only on its current state, and not the past states [21,22].

2.2. Information-theoretical measures: entropy, mutual information and transfer entropy

Consider a discrete random variable X whose probability mass function is denoted by $p(x) = \text{Prob}(X = x)$. To quantify the unpredictability of X, one can calculate its *(information) entropy* defined as

$$H(X) = -\sum_x p(x) \log p(x), \tag{5}$$

where by convention, we use "log" to represent "$\log_2$". In general, $H(X)$ approximates the minimal binary description length L of the random variable X, with the following inequality [21]:

$$H(X) \leq L < H(X) + 1. \tag{6}$$

It follows that, among all random variables with c elements, the one with uniform distribution yields the maximum entropy, $\log(c)$.

Consider now two random variables X and Y with joint distribution

$$p(x, y) = \text{Prob}(X = x, Y = y), \tag{7}$$

and conditional distribution

$$p(x|y) = \text{Prob}(X = x|Y = y). \tag{8}$$

The *joint entropy* $H(X, Y)$ and *conditional entropy* $H(X|Y)$ for X and Y are defined, respectively, as

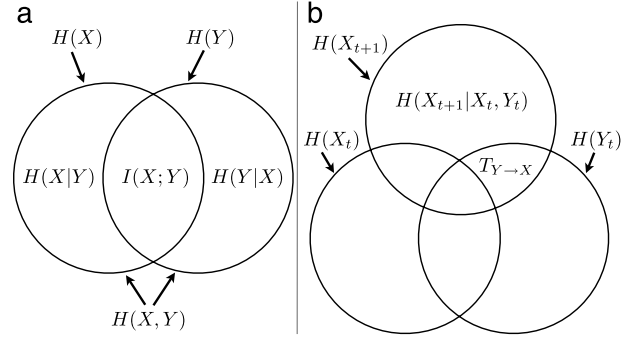$$H(X, Y) = -\sum_{x,y} p(x, y) \log p(x, y), \tag{9}$$

Fig. 1. Venn-like diagrams for information-theoretical measures. (a) Relations between: entropies $H(X)$ and $H(Y)$, joint entropy $H(X, Y)$, conditional entropies $H(X|Y)$ and $H(Y|X)$, and mutual information $I(X; Y)$, of two random variables X and Y. (b) Relations between: transfer entropy $T_{Y \to X}$, entropies of random variables X_{t+1}, X_t, and Y_t, and their joint and conditional entropies. The transfer entropy is the difference between the conditional entropies $H(X_{t+1}|X_t, Y_t)$ and $H(X_{t+1}|X_t)$, which measures the extra information provided by Y_t (in addition to X_t) in the determination of X_{t+1}.

and

$$H(X|Y) = -\sum_y p(y)H(Y|X = x)$$

$$= -\sum_{x,y} p(x, y) \log p(x|y). \tag{10}$$

Similar definition holds for $H(Y|X)$.

It is easy to verify that conditioning reduces entropy, i.e., knowledge of Y will reduce (or at least cannot increase) the uncertainty about X, i.e.,

$$H(X|Y) \leq H(X). \tag{11}$$

Similarly, $H(Y|X) \leq H(Y)$.

The reduction of uncertainty of X (Y) given full information about Y (X) can be measured by the *mutual information* between X and Y, as [21]

$$I(X; Y) = H(X) - H(X|Y) = H(Y) - H(Y|X). \tag{12}$$

The mutual information is symmetric in X and Y, and measures their deviation from independence: if X and Y are fully dependent, then $H(X|Y) = H(Y|X) = 0$ and thus $I(X; Y) = H(X) = H(Y)$; on the other hand, if X and Y are independent, then $H(X|Y) = H(X)$ and $H(Y|X) = H(Y)$ and therefore $I(X; Y) = 0$. In general, we have [21]

$$0 \leq I(X; Y) \leq \min[H(X), H(Y)]. \tag{13}$$

It is convenient to visualize the relationship between entropy, joint entropy, conditional entropy, and mutual information by a Venn-like diagram, as shown in Fig. 1(a).

We now turn to stochastic processes. For a stationary process $\{X_t\}$, its *entropy rate* $H(\mathcal{X})$ can be defined as

$$H(\mathcal{X}) = \lim_{t \to \infty} H(X_t|X_{t-1}, X_{t-2}, \ldots, X_1), \tag{14}$$

which can be thought of as the (asymptotic) growth rate of the joint entropy $H(X_1, X_2, \ldots, X_t)$. If the process is Markovian, then [21]

$$H(\mathcal{X}) = \lim_{t \to \infty} H(X_t|X_{t-1}). \tag{15}$$

For two stochastic processes $\{X_t\}$ and $\{Y_t\}$, the reduction of uncertainty about X_{t+1} due to the information of the past τ_Y states of Y, represented by

$$Y_t^{(\tau_Y)} = (Y_t, Y_{t-1}, \ldots, Y_{t-\tau_Y+1}), \tag{16}$$

in addition to the information of the past τ_X states of X, represented by

$$X_t^{(\tau_X)} = (X_t, X_{t-1}, \ldots, X_{t-\tau_X+1}), \tag{17}$$

is measured by the *transfer entropy* from Y to X, defined as [7]

$$T_{Y \to X} = H(X_{t+1}|X_t^{(\tau_X)}) - H(X_{t+1}|X_t^{(\tau_X)}, Y_t^{(\tau_Y)}). \tag{18}$$

One can similarly define $T_{X \to Y}$, which does not necessarily equal to $T_{Y \to X}$. Note that $T_{Y \to X}$ can also be interpreted as the mutual information between X_{t+1} and $Y_t^{(\tau_Y)}$ conditioned on $X_t^{(\tau_X)}$. In this paper, we focus on the case where

$$\tau_X = \tau_Y = 1, \tag{19}$$

unless specified otherwise. The relationship between transfer entropy, entropy and conditional entropy is illustrated in Fig. 1(b).

3. Measuring information transfer in two coupled oscillators

Coupled oscillator networks are commonly used for modeling the dynamic behavior of complex systems in various areas [23–26]. Here we consider discrete dynamics of coupled oscillator networks, in the form

$$x_{t+1}^{(i)} = f[x_t^{(i)}] + \epsilon \sum_{j \neq i} c_{ij} g[x_t^{(i)}, x_t^{(j)}], \quad i = 1, 2, \ldots, N. \tag{20}$$

Here $x_t^{(i)} \in \mathcal{D} \subset \mathbb{R}^d$ is the state of oscillator i at time t, $f : \mathcal{D} \to \mathcal{D}$ is the dynamics of individual oscillators, $g : \mathcal{D} \times \mathcal{D} \to \mathcal{D}$ is the coupling function, and ϵ is the coupling strength. Term c_{ij} represents the coupling from j to i. In this paper, we use

$$f(x) = ax(1-x) \tag{21}$$

with parameter $a = 4$. The coupling function is chosen to be

$$g(x, y) = f(y) - f(x). \tag{22}$$

The choice of $\epsilon \in [0, 1]$ and normalization condition

$$\sum_{j \neq i} c_{ij} = 1 \tag{23}$$

guarantees that

$$x_t^{(i)} \in \mathcal{D} = [0, 1] \quad \text{for all } i \text{ and } t. \tag{24}$$

We first explore information transfer in two coupled oscillators, with bidirectional and unidirectional couplings, respectively. With a slight abuse of notation, we use X and Y to represent oscillators 1 and 2, respectively. In terms of Eq. (20), the bidirectional coupling corresponds to having $c_{12} = c_{21} = 1$ and unidirectional coupling corresponds to $c_{12} = 1, c_{21} = 0$. Results from numerical simulation are shown in Fig. 2.

One direct observation is that mutual information can be used as a measure of synchrony between two oscillators X and Y. When X and Y are synchronized, their mutual information

$$I(X; Y) = H(X) = H(Y). \tag{25}$$

When they are not synchronized,

$$I(X; Y) < \min[H(X), H(Y)]. \tag{26}$$

We remark that this observation suggests a new and alternative way of measuring generalized synchronization or synchronization among a partial set of nodes in a large-scale network [27,28].

For bidirectionally coupled oscillators, synchrony occurs when the coupling strength [29]

$$\epsilon \in (0.25, 0.75), \tag{27}$$

as shown in Fig. 2(a). The mutual information reaches its maximum for the same range of ϵ. Similarly, synchronization and maximum mutual information both occur when

$$\epsilon \in (0.5, 1], \tag{28}$$

in the case where X and Y are unidirectionally coupled (Fig. 2(c)). Fig. 2(b), (d) show typical time series of the bidirectionally and unidirectionally coupled oscillators with $\epsilon = 0.1$ (unsynchronized trajectories) and $\epsilon = 0.6$ (synchronized trajectories), respectively.

When two oscillators synchronize, the transfer entropy from either one of them to the other becomes zero because no extra information can be gained by learning the past trajectory of the other oscillator (in addition to that from one's own). As a consequence, detection of coupling by transfer entropy (or any other measure) is valid only when the oscillators are not synchronized. Oscillators that are synchronized produce identical trajectories and therefore appear indistinguishable.

When the two oscillators are not synchronized, there is a positive transfer entropy following the directionality of coupling. For bidirectionally coupled oscillators,

$$T_{X \to Y} = T_{Y \to X} > 0 \quad \text{if } \epsilon \in (0, 0.25) \cup (0.75, 1], \tag{29}$$

except for those parameters at which the trajectories of X and Y settle into a periodic orbit (Fig. 2(a)). For unidirectionally coupled oscillators, positive transfer entropy $T_{X \to Y}$ is observed when

$$T_{X \to Y} > T_{Y \to X} = 0 \quad \text{if } \epsilon \in (0, 0.5). \tag{30}$$

This asymmetry of transfer entropy confirms the dominant direction of information flow from X to Y, and not the other way around (Fig. 2(c)).

4. Measuring information transfer in coupled oscillator networks

Having studied the application of transfer entropy in systems of two coupled oscillators, we now turn to networks.

4.1. Effect of indirect influence

First we explore information transfer under the presence of indirect couplings. Consider a directed linear chain

$$Z \to Y \to X, \tag{31}$$

where Z indirectly influences X through Y (Fig. 3(a)). We focus on the dynamics of this three-node network according to Eq. (20), with $\epsilon \in [0.2, 0.4]$, a regime where coupling has a non-negligible effect on the dynamics but not strong enough to result in synchronization.

In Fig. 4(a) we plot values of the transfer entropies $T_{X \to X}$, $T_{Y \to X}$, and $T_{Z \to X}$ ($T_{X \to Y} \approx T_{X \to Z} \approx 0$ are not plotted). By definition, $T_{X \to X} = 0$. The direct influence of Y on X is validated by the positive values of $T_{Y \to X}$. Interestingly, values $T_{Z \to X}$ are also positive, despite the fact that there is no direct coupling from Z to X. Similar results are found for other networks that contain the direct linear chain $Z \to Y \to X$ but without the direct coupling $Z \to X$. See Fig. 3(b)–(c) for the other two networks and Fig. 4(c), (e) for the corresponding results.

One important implication of these results is that the use of transfer entropy for inferring network structure can be inappropriate under the presence of indirect influences. Since indirect couplings are common in many networks, *directed edges that are inferred by measuring transfer entropy can often be "false positive"*.

4.2. Causation entropy

We note that the key reason transfer entropy can fail in identifying indirect couplings from direct ones is that it is a pairwise measure between two processes. For example, the transfer entropy $T_{Z \to X}$ shown in Fig. 4(a), (c), (e) does not account for the fact that the observed information transfer from Z to X is indeed a consequence of the direct information transfer from Z to Y, and then Y to X.
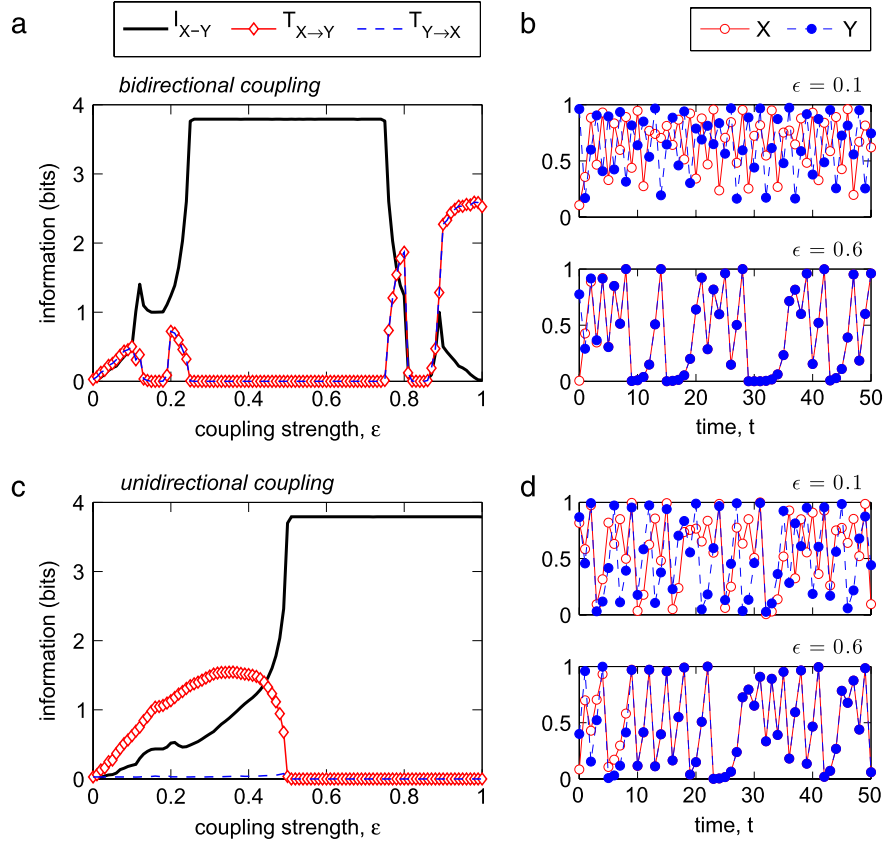
Fig. 2. Measuring information flow in two coupled logistic maps. (a) Dependence of mutual information $I_{X-Y} = I(X; Y)$ and transfer entropies $T_{X \to Y}$ and $T_{Y \to X}$ on coupling strength ϵ for two bidirectionally coupled oscillators. Synchronization occurs when $\epsilon \in (0.25, 0.75)$, which is the same region where the mutual information reaches its maximum. Due to the symmetry of coupling, $T_{X \to Y} = T_{Y \to X}$. (b) Typical time series for two bidirectionally coupled oscillators, with $\epsilon = 0.1$ (top, unsynchronized trajectories) and $\epsilon = 0.6$ (bottom, synchronized trajectories). (c)–(d) Same as (a)–(b), but for two oscillators with unidirectional coupling from oscillator X to oscillator Y. In this case, synchronization appears when $\epsilon \in [0.5, 1)$. For $\epsilon \in (0, 0.5)$, $T_{X \to Y} \gg T_{Y \to X}$, indicating that the dominant direction of information flow between X and Y is from X to Y. In all simulations of the paper, we generate trajectories of length 10^5 and discard the initial 5% segments for all computations. The interval [0, 1] is divided evenly into 2^4 subintervals for the estimation of discrete probabilities. In our simulations, we made the choice of 2^4 based on the balance between the length of the time series and the number of variables in the joint distribution: too few subintervals will only reveal limited information about the true dynamics and on the other hand, too many of them will lead to statistical under-sampling [37,38]. Note that this problem of finding an appropriate number of subintervals for the estimation of entropy is analogous to the problem of finding an appropriate number of bins to construct a histogram, for which no universal "best" solution exists.
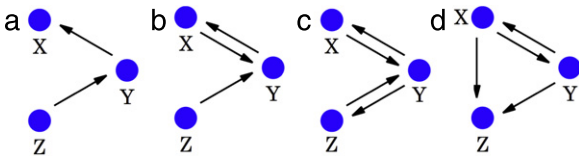

Fig. 3. Small-scale directed binary networks. (a)–(c) Networks with a directed linear chain $Z \to Y \to X$, but no direct coupling $Z \to X$. (d) A network that contains a direct coupling $X \to Z$, but not $Z \to X$.

Here we propose a new measure, which we call *causation entropy*. The causation entropy from Z to X (conditioned on X and Y) is defined as

$$C_{Z \to X|(X,Y)} = H(X_{t+1}|X_t, Y_t) - H(X_{t+1}|X_t, Y_t, Z_t). \qquad (32)$$

Thus, $C_{Z \to X|(X,Y)}$ measures the extra information provided to X by Z *in addition* to the information that is already provided to X by other means.

For an arbitrary set of processes, causation entropy is defined as follows.

Definition 1 (*Causation Entropy*)**.** The causation entropy from process $\mathcal{Q}$ to process $\mathcal{P}$ conditioned on the set of processes $\mathscr{S}$ is defined as

$$C_{\mathcal{Q} \to \mathcal{P}|(\mathscr{S})} = H(\mathcal{P}_{t+1}|\mathscr{S}_t) - H(\mathcal{P}_{t+1}|\mathscr{S}_t, \mathcal{Q}_t). \qquad (33)$$

Causation entropy $C_{\mathcal{Q} \to \mathcal{P}|(\mathscr{S})}$ is a generalization of transfer entropy. In fact, by letting $\mathscr{S} = \mathcal{P}$, we have

$$C_{\mathcal{Q} \to \mathcal{P}|(\mathcal{P})} = T_{\mathcal{Q} \to \mathcal{P}}. \qquad (34)$$

In general, causation entropy $C_{\mathcal{Q} \to \mathcal{P}|(\mathscr{S})}$ measures the reduction in uncertainty in $\mathcal{P}$ due to the extra knowledge of $\mathcal{Q}$ in addition to that of $\mathscr{S}$.

If $\mathscr{S} = \varnothing$, we simply write

$$C_{\mathcal{Q} \to \mathcal{P}} = C_{\mathcal{Q} \to \mathcal{P}|(\varnothing)}. \qquad (35)$$

It follows that

$$C_{\mathcal{Q} \to \mathcal{P}} = H(\mathcal{P}_{t+1}) - H(\mathcal{P}_{t+1}|\mathcal{Q}_t) = I(\mathcal{P}_{t+1}; \mathcal{Q}_t), \qquad (36)$$

which is the mutual information between $\mathcal{P}_{t+1}$ and $\mathcal{Q}_t$. When $\mathscr{S} \neq \varnothing$, causation entropy $C_{\mathcal{Q} \to \mathcal{P}|(\mathscr{S})}$ can be interpreted as the mutual information shared between $\mathcal{P}_{t+1}$ and $\mathcal{Q}_t$ conditioned on $\mathscr{S}_t$.

Fig. 4(b), (d), (f) shows that, for the networks in Fig. 3(a)–(c), both $C_{X \to X}$ and $C_{Y \to X|(X)}$ are positive, as a result of the influence of X on itself (self-dynamics) and the direct influence of Y on X. On the other hand, and by design, the causation entropy $C_{Z \to X|(X,Y)} \approx 0$, in sharp contrast to the positive transfer entropy, $T_{Z \to X} > 0$ (Fig. 4(a), (c), (e)). The reason $C_{Z \to X|(X,Y)}$ is close to zero is that, the information provided by Z (to X) is merely a subset of the information provided by Y. No extra information about X's future state can be gained by learning the current state of Z if those of X and Y are already known.

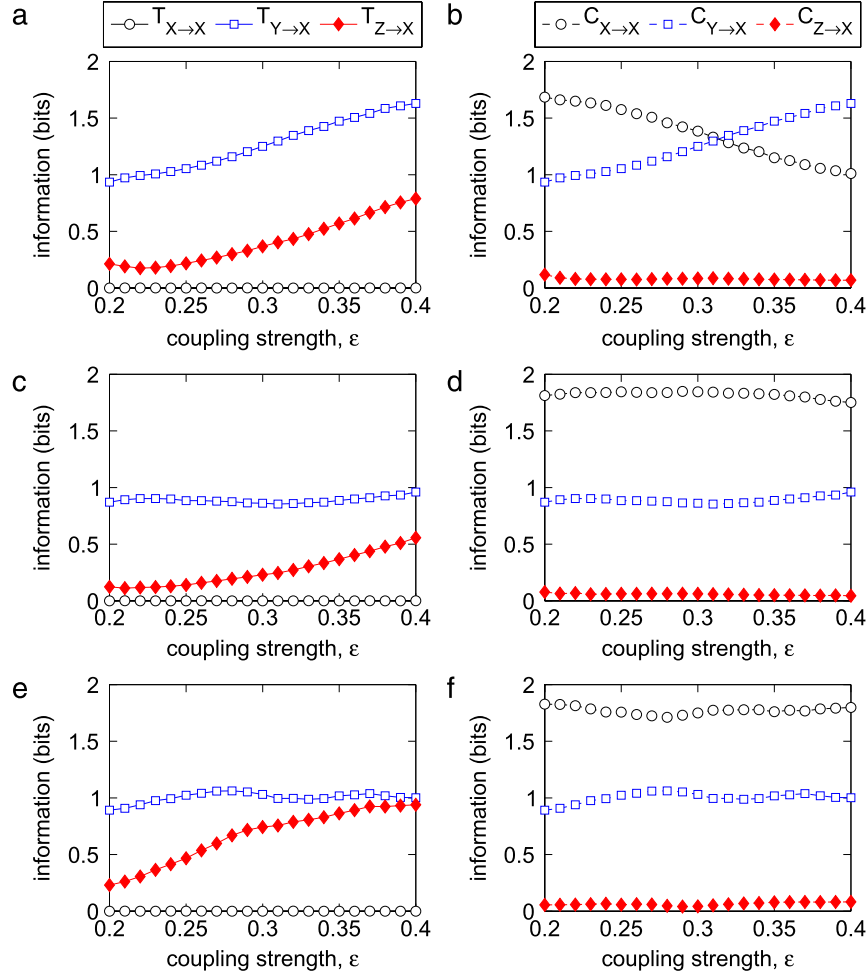**Fig. 4.** Causation entropy and transfer entropy for the identification of indirect coupling. (a)–(b) Transfer entropies $\{T_{X\to X}, T_{Y\to X}, T_{Z\to X}\}$ and causation entropies $\{C_{X\to X}, C_{Y\to X|(X)}, C_{Z\to X|(X,Y)}\}$ for the network shown in Fig. 3(a) with dynamics (20). Note that the transfer entropy $T_{Z\to X}$ is positive despite the absence of direct coupling from Z to X. On the other hand, the causation entropy $C_{Z\to X|(X,Y)} \approx 0$, since information that are being indirectly transferred from Z to X all go through Y. (c)–(d) Same as (a)–(b), for the network in Fig. 3(b). (e)–(f) Same as (a)–(b), for the network in Fig. 3(c).

4.3. Example: dominance of neighbors

Dominance of neighbors refers to a scenario where an oscillator's future state is dominantly determined by the state of its neighboring nodes, rather than by itself. In terms of Eq. (20), this occurs when the coupling strength $\epsilon \approx 1$. We here explore its effect on information transfer. As an example, we consider dynamics by Eq. (20) on the network shown in Fig. 3(d), where node X receives input from Y, but not from Z (even indirectly).

As shown in Fig. 5(a), transfer entropy $T_{Y\to X}$ is positive, due to the direct influence of Y on X. Surprisingly, transfer entropy $T_{Z\to X}$ is also found to be positive, despite the fact that no information flows from Z to X, either directly or indirectly.

The reason positive transfer entropy $T_{Z\to X}$ is found in absence of influence of Z on X is that $T_{Z\to X}$ is taken to be the difference between $H(X_{t+1}|X_t)$ and $H(X_{t+1}|X_t, Z_t)$. Here since X_{t+1} is dominantly determined by Y_t and only depends weakly on X_t, the conditional entropies

$$\begin{cases} H(X_{t+1}|X_t) \approx H(X_{t+1}), \\ H(X_{t+1}|X_t, Z_t) \approx H(X_{t+1}|Z_t). \end{cases} \tag{37}$$

A closer inspection of the network reveals that under the strong coupling regime where the dynamics of an oscillator depends dominantly on its neighbors dynamics, the state of X_{t+1} depends mostly on Y_t (and not X_t). Since Y_t depends mostly on X_{t-1} by

the very same argument, we conclude that the mutual information between X_{t+1} and X_{t-1} is high. Similarly, since Z_t depends mostly on X_{t-1} and Y_{t-1}, there is high mutual information between Z_t and X_{t-1}. Based on this analysis, the mutual information between X_{t-1} and X_t should be low and that between Z_t and X_{t+1} should be non-negligible, which is confirmed in Fig. 5(c)–(d). Although information in the network flows directly from X to Z, without accounting for the dominant factors that determine the value of X_{t+1}, one would indeed infer a directed link from Z to X based on the calculation of the transfer entropy $T_{Z\to X}$.

We note that, because of the dominance of Y on X (as opposed to X on itself), one should indeed measure the causation entropies $C_{Y\to X}$, $C_{X\to X|(Y)}$, and $C_{Z\to X|(X,Y)}$, respectively. Results are shown in Fig. 5(b). The value $C_{Y\to X} > 0$, as expected. The value $C_{X\to X|(Y)} \approx 0$, due to the dominant influence of Y (rather than X itself) on X. The value $C_{Z\to X|(X,Y)} \approx 0$ as well, suggesting the absence of information transfer from Z to X, which is consistent with the structure of the network shown in Fig. 3(d).

4.4. Iterative evaluation of causation entropy in a network of N processes

The determination of causation entropies (i.e., the order Y, X, Z) can in fact be done *a priori*, by first choosing the process $\mathcal{Q}_1 \in \{X, Y, Z\}$ that maximizes the causation entropy $C_{\mathcal{Q}_1\to X}$, and then iter-

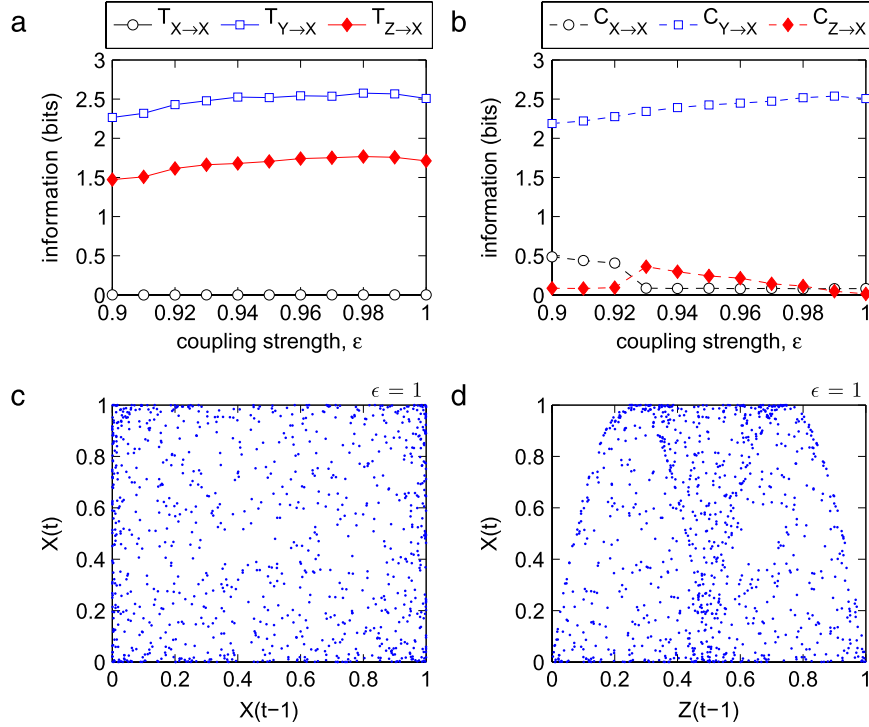**Fig. 5.** Causation entropy versus transfer entropy under the dominance of neighbors. (a) Transfer entropies $\{T_{X\to X}, T_{Y\to X}, T_{Z\to X}\}$. (b) Causation entropies $\{C_{X\to X|(Y)}, C_{Y\to Y}, C_{Z\to X|(X,Y)}\}$ for the network in Fig. 3(d) whose dynamics follow Eq. (20). (c) Scatter plot between X_{t-1} and X_t for $\epsilon = 1$. (d) Scatter plot between Z_{t-1} and X_t for $\epsilon = 1$. In (c) and (d), points are taken from a randomly select trajectory segment (of the full trajectory) of length 1000.

atively select $\mathcal{Q}_k$ as the process that maximizes $C_{\mathcal{Q}_k\to X|(\mathcal{Q}_1,\ldots,\mathcal{Q}_{k-1})}$ (see the following paragraph for details). For the example used in Fig. 5, we found that $\mathcal{Q}_1 = Y$ and $\mathcal{Q}_2 = X$. Therefore, in contrast to transfer entropy, causation entropy can successfully identify the dominance of neighbors and in turn avoid erroneous inference of couplings due to its effect.

For a network of N coupled stochastic processes $\{X_t^{(i)}\}_{i=1}^N$, we propose to identify the set of causal processes of a given process i by iterative maximization of causation entropy. Let $n_0 = i$. We first find process n_1 that satisfies

$$n_1 = \underset{j}{\arg\max}\, C_{X^{(j)}\to X^{(i)}}. \tag{38}$$

Then we iteratively seek for n_k ($k = 2, 3, \ldots$) that satisfies

$$n_k = \underset{j}{\arg\max}\, C_{X^{(j)}\to X^{(i)}|(X^{(n_0)},X^{(n_1)},\ldots,X^{(n_{k-1})})}. \tag{39}$$

We stop the search at step k when

$$C_{X^{(n_k)}\to X^{(i)}} < \theta, \tag{40}$$

where θ is a preselected tolerance value. The processes $n_1, n_2, \ldots, n_{k-1}$ (in the decreasing order of dominance) form the set of causal processes of i.

Note that in theory the value of $C_{X^{(n_k)}\to X^{(i)}}$ will be exactly zero if the dynamics of node n_k does not causal-determine the dynamics of node i. In practice, however, the numerical estimation of $C_{X^{(n_k)}\to X^{(i)}}$ is based on the estimation of probability distributions from finite sample and will at best be close to (but not necessarily equal to) zero for a finite number of data points. A rigorous way of determining whether the numerically computed causation entropy should be identified as zero is to perform a hypothesis test. It can be challenging to do such a test in practice and often times one can instead use a shuffle test to obtain approximate confidence intervals [14].

4.5. Example: anticipatory couplings

Our last example is a unidirectionally coupled dynamical system with anticipatory coupling [13]

$$\begin{cases} x_{t+1} = f(x_t), \\ y_{t+1} = (1-\epsilon)f(y_t) + \epsilon[(1-\alpha)f(x_t) + \alpha f^2(x_t)], \end{cases} \tag{41}$$

where $f(x) = ax(1-x)$ with $a = 4$, parameter $\epsilon \in [0, 1]$ is the coupling strength, and parameter $\alpha \in [0, 1]$ is the strength of anticipatory coupling. Notation f^2 means that the map f is applied twice.

Here we adopt the concept and notation of transfer entropy to define

$$T_{X_t\to Y_{t+1}} = T_{X\to Y} = H(Y_{t+1}|Y_t) - H(Y_{t+1}|Y_t, X_t), \tag{42}$$

and

$$T_{X_{t+1}\to Y_{t+1}} = H(Y_{t+1}|Y_t) - H(Y_{t+1}|Y_t, X_{t+1}). \tag{43}$$

Fig. 6(a) shows that both $T_{X_t\to Y_{t+1}}$ and $T_{X_{t+1}\to Y_{t+1}}$ are positive, with comparable values. Does this suggest that both X_t and X_{t+1} independently influence Y_{t+1}? Standard interpretation (of transfer entropy) would suggest that the answer to this question is yes.

By use of causation entropy, we find that Y_{t+1} is primarily determined by Y_t. The second dominant influence on Y_{t+1} is X_{t+1}, as confirmed by the values of

$$C_{X_{t+1}\to Y_{t+1}|(Y_t)} = H(Y_{t+1}|Y_t) - H(Y_{t+1}|Y_t, X_{t+1}). \tag{44}$$

It turns out that additional information of X_t (beyond Y_t and X_{t+1}) does not contribute to the reduction of uncertainty of Y_{t+1}. This is validated by the causation entropy

$$C_{X_t\to Y_{t+1}|(Y_t,X_{t+1})} = H(Y_{t+1}|Y_t, X_{t+1}) - H(Y_{t+1}|Y_t, X_{t+1}), \tag{45}$$

which remain close to zero, as shown in Fig. 6(b).

Therefore, in contrast to transfer entropy analysis, which would suggest that both X_t and X_{t+1} participate in the determination of Y_{t+1}, causation entropy analysis reveals that information of X_t is indeed completely redundant in inferring Y_{t+1}. In fact, by expressing $f(x_t)$ as x_{t+1} in Eq. (41), it appears that the value of y_{t+1} depends solely on y_t and x_{t+1}, and not on x_t.
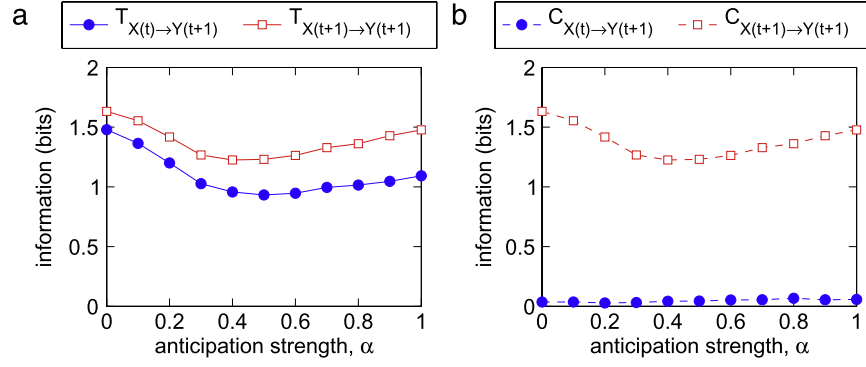
Fig. 6. Causation entropy versus transfer entropy under anticipatory coupling. (a) Transfer entropies $T_{X_t \to Y_{t+1}}$ and $T_{X_{t+1} \to Y_{t+1}}$. (b) Causation entropies $C_{X_{t+1} \to Y_{t+1}|(Y_t)}$ and $C_{X_t \to Y_{t+1}|(Y_t, X_{t+1})}$. Here parameter $\epsilon = 0.3$.
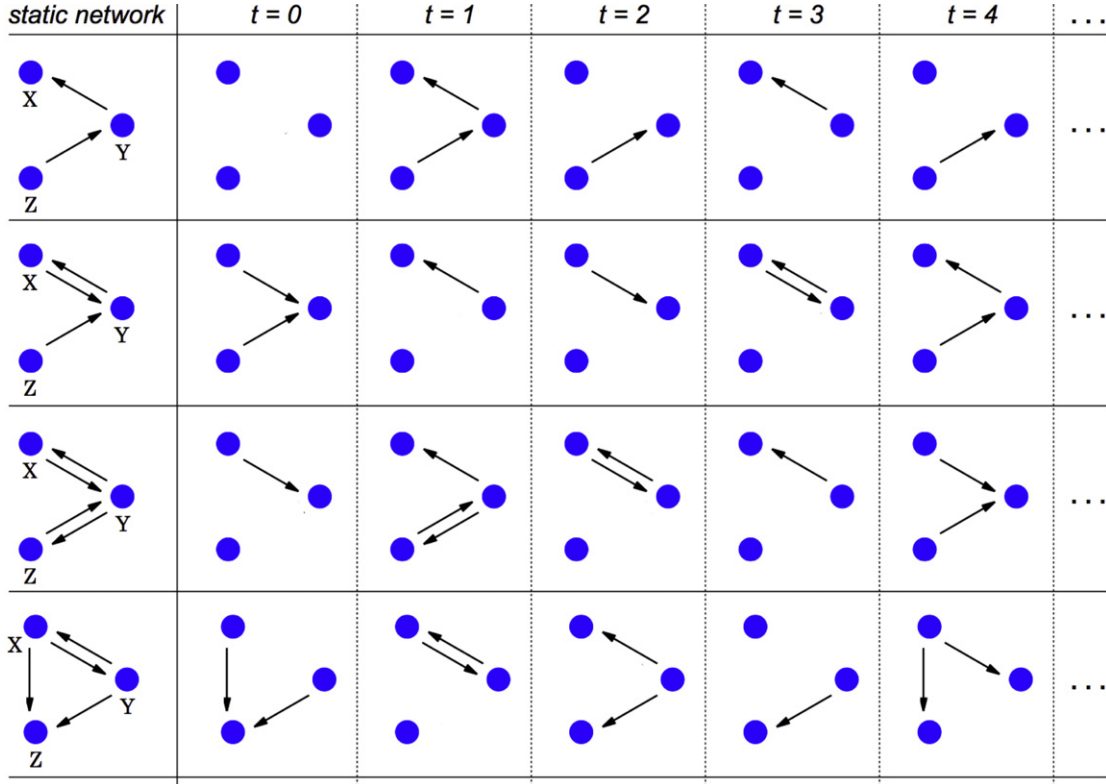

Fig. 7. Examples of time-dependent networks. First (leftmost) column: structure of static networks (the same as those in Fig. 3). Second to the last (rightmost) columns: typical network structures at different times, obtained from keeping each directed edge of the static network independently with probability $p = 0.5$ at each time t.

5. Information transfer in time-dependent networks

The effects of time-dependent structures on network dynamics are often intriguing and pose considerable challenges for analysis. For example, the problem of synchronization stability of coupled oscillators in time-dependent networks has been fully addressed only for a few specific cases [30–34]. Here, our focus is to measure information transfer among oscillators that are coupled through a time-dependent network structure (that is, a network whose edges change in time). In particular, we generalize Eq. (20) to allow for time-dependent interactions in between oscillators, as

$$x_{t+1}^{(i)} = f[x_t^{(i)}] + \epsilon \sum_{j \neq i} c_{ij}(t) g[x_t^{(i)}, x_t^{(j)}], \quad i = 1, 2, \ldots, N. \quad (46)$$

Here all terms in Eq. (46) except for $c_{ij}(t)$ are the same as those in Eq. (20). The term $c_{ij}(t)$ represents the coupling from j to i at time t and explicitly accounts for the time-dependent network structure.

We consider time-dependent networks constructed as follows. Start with a baseline static network whose adjacency matrix is $\bar{C} = [\bar{c}_{ij}]_{n \times n}$. The edges in the network are then allowed to "blink" according to the following rule, to generate a time-dependent network: at each time t,

$$c_{ij}(t) = \begin{cases} \bar{c}_{ij}, & \text{with probability } p; \\ 0, & \text{with probability } 1 - p. \end{cases} \quad (47)$$

Therefore, when the blinking probability $p = 1$, the network is the same as the baseline static network; on the other extreme, when $p = 0$, no edge exists and the network becomes empty (i.e., each oscillator is isolated and does not couple to other oscillators). For the values of p in between 0 and 1, the network structure changes in time in a stochastic fashion (see Fig. 7 for a few illustrative examples).

Our interest lies in the information transfer within such time-dependent networks. Different from its static counterpart, the
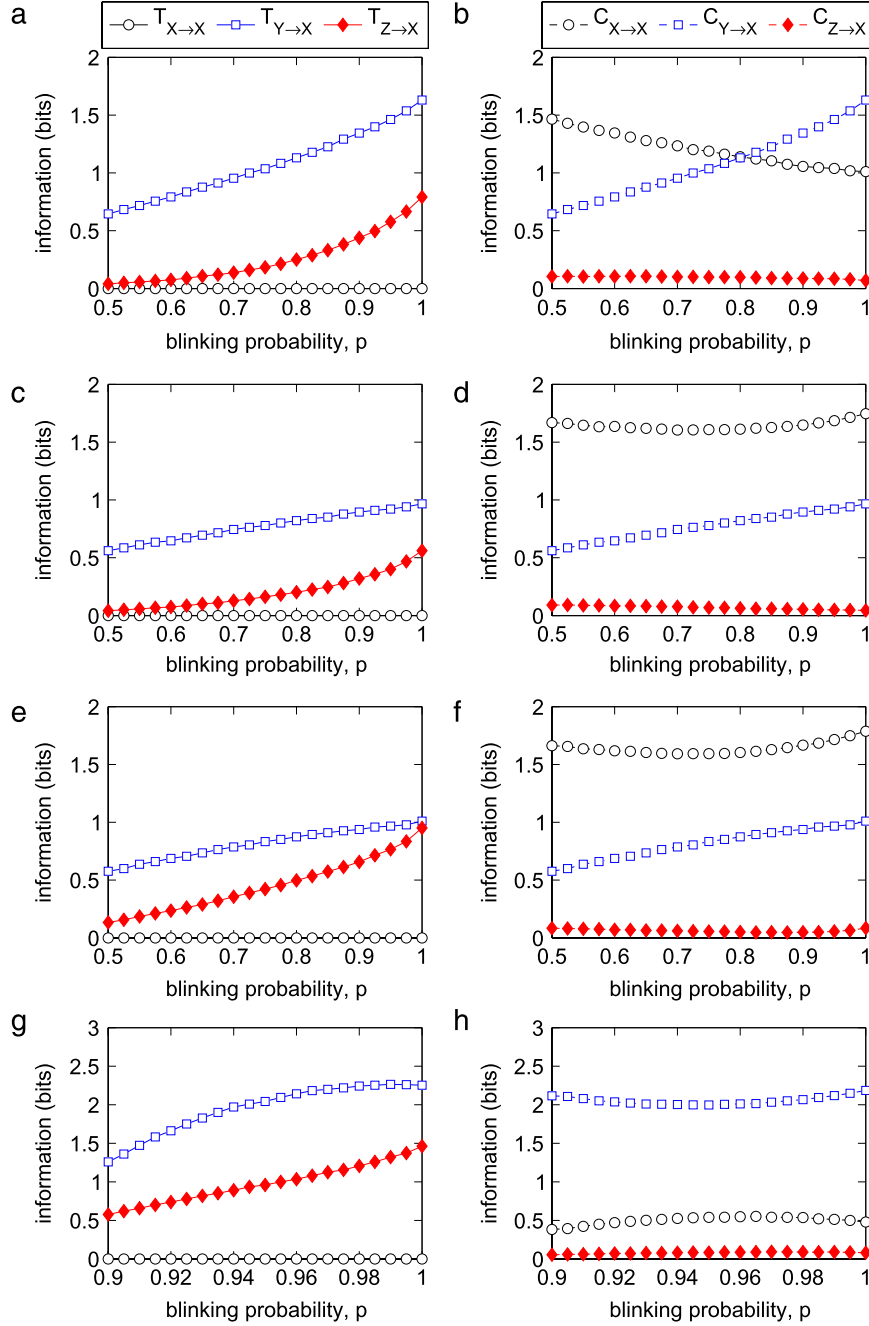
Fig. 8. Causation entropy versus transfer entropy for time-dependent networks. (a)–(b) Transfer entropies $\{T_{X \to X}, T_{Y \to X}, T_{Z \to X}\}$ and causation entropies $\{C_{X \to X}, C_{Y \to X|(X)}, C_{Z \to X|(X,Y)}\}$ for the time-dependent network originates from the network in Fig. 3(a) via Eq. (47) and endowed with dynamics (46), for the fixed coupling strength $\epsilon = 0.4$. (c)–(f) Same as (a)–(b), for the networks in Fig. 3(b) and (c), respectively, and $\epsilon = 0.4$. (g)–(h) Same as (a)–(b), for the network in Fig. 3(d) and $\epsilon = 0.9$.

flow of information in a time-dependent network often cannot be directly obtained from examining the edges $c_{ij}(t)$, because it is possible for a network to be disconnected at all times and yet be able to transfer information from one node to another. Such a scenario has been previously considered in the synchronization of coupled oscillators in time-dependent networks with edges being switched on and off [32] and in moving-neighbor networks whose edges are defined by the local interactions between agents that move in space [31]. In both cases, even though the original static network is connected, the corresponding time-dependent network obtained by blinking the edges might not be (see Fig. 7 for examples).

The connection between these time-dependent networks and the original static network is that the asymptotic temporal average

of each directed edge, $\langle c_{ij} \rangle$, is proportional to the weight of the same edge in the static network:

$$\langle c_{ij} \rangle = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} c_{ij}(t) = p \bar{c}_{ij}. \tag{48}$$

We ran numerical simulation on several time-dependent networks and focus on the information flow measured both by transfer entropy and causation entropy. Fig. 8(a) shows that, for the directed linear chain $Z \to Y \to X$ with fixed coupling strength $\epsilon = 0.4$, when the blinking probability p increases, the transfer entropy $T_{Z \to X}$ becomes increasingly non-negligible, indicating direct information transfer from Z to X from standard interpretation. On the other hand, the causation entropy $C_{Z \to X}$ remains

essentially zero, suggesting that the information transferred from *Z* to *X* is merely a redundancy of the information that are transferred from *Z* to *Y* and *Y* to *X*, respectively, possibly at different times. Fig. 8(b)–(c) show similar comparison between transfer entropy and causation entropy for measuring the information flow in time-dependent networks that originate from the networks shown in Fig. 3(b)–(c) with the fixed coupling strength $\epsilon = 0.4$.

The possible misinterpretation of transfer entropy becomes more evident under the dominance of neighbors scenario, where the coupling strength ϵ is close to 1. As shown in Fig. 8(d), under such scenario, transfer entropy identifies a strong information transfer from *Z* to *X* whereas in the average network of the time-dependent network, it is the exact opposite. Causation entropy, on the other hand, successfully identifies the dominant nodes that influence the dynamics of *X*, namely, its neighbor *Y* and then *X* itself.

6. Discussion and conclusion

Our main message here is that while being an essential problem in science in general, and dynamical systems in particular, the question of what is cause and what is influence in complex system analysis is challenging, not due to the lack of methodology, but rather due to the lack of clear and comprehensible understanding of the applicability of proposed methods, in particular when the underlying system involves complex, possibly time-dependent interactions. The popular concept of transfer entropy has been used lately to serve as a way of inferring causality, without much understanding of its domain of success.

We here explored information flow measured from the dynamics of small-scale coupled oscillators network, attempting to gain insights into the validity of transfer entropy as well as its limitations. For two coupled oscillators, transfer entropy is found to successfully detect the directionality of information flow, even in cases where the couplings are blinking (time-dependent). However, its validity breaks down under the presence of indirect couplings, dominance of neighboring dynamics, anticipatory couplings, or time-dependent couplings, which are common in large-scale complex systems.

To overcome the limitations of transfer entropy, we introduced a new measure of information flow called causation entropy, which is designed to allow inference of causation despite the presence of primary and secondary influences between elements of a larger coupled system. We highlighted the success of our approach with several examples where specifically the transfer entropy cannot distinguish between causation and independence, but causation entropy successfully infers the true causal relationships.

Given the recent advancements in estimating transfer entropy in rather general settings including multivariate time series and infinite time delay [14,35], it is our hope to build on the idea of causation entropy to explore information flow and coupling inference in larger-scale systems, which are important for a wide range of applications across scientific fields. One challenge is that, for large-scale systems, naive binning methods would require an exponential number of data points with respect to the number of variables, in order to reliably calculate entropies (including joint entropy, transfer entropy, and also causation entropy). Nonparametric density estimation methods previously developed for mutual information [8,36] are likely to offer a route toward the reliable estimation of causation entropies in large-scale dynamical systems.

Acknowledgments

References

[1] P.W. Hollanda, Statistics and causal inference, J. Amer. Statist. Assoc. 81 (1986) 945–960.
[2] J. Pearl, Causality: Models, Reasoning and Inference, second ed., Cambridge University Press, Cambridge, UK, 2009.
[3] F. Heider, Social perception and phenomenal causality, Psychol. Rev. 51 (1944) 358–374.
[4] K.J. Rothman, S. Greenland, Causation and causal inference in epidemiology, Am. J. Public Health 95 (2005) S144–S150.
[5] J.J. Heckman, Econometric causality, Int. Stat. Rev. 76 (2008) 1–27.
[6] C.W.J. Granger, Investigating causal relations by econometric models and cross-spectral methods, Econometrica 37 (1969) 424–438.
[7] T. Schreiber, Measuring information transfer, Phys. Rev. Lett. 85 (2000) 461–464.
[8] K. Hlaváčková-Schindlera, M. Paluš, M. Vejmelkab, J. Bhattacharya, Causality detection based on information-theoretic approaches in time series analysis, Phys. Rep. 441 (2007) 1–46.
[9] M. Vejmelka, M. Paluš, Inferring the directionality of coupling with conditional mutual information, Phys. Rev. E 77 (2008) 026214.
[10] C.J. Honey, R. Kötter, M. Breakspear, O. Sporns, Network structure of cerebral cortex shapes functional connectivity on multiple time scales, Proc. Natl. Acad. Sci. USA 104 (2007) 10240–10245.
[11] R. Vicente, M. Wibral, M. Lindner, G. Pipa, Transfer entropy—a model-free measure of effective connectivity for the neurosciences, J. Comput. Neurosci. 30 (2011) 45–67.
[12] E. Bollt, Synchronization as a process of sharing and transferring information, Internat. J. Bifur. Chaos 22 (2012) 1250261.
[13] D.W. Hahs, S.D. Pethel, Distinguishing anticipation from causality: anticipatory bias in the estimation of information flow, Phys. Rev. Lett. 107 (2011) 128701.
[14] J. Runge, J. Heitzig, V. Petoukhov, J. Kurths, Escaping the curse of dimensionality in estimating multivariate transfer entropy, Phys. Rev. Lett. 108 (2012) 258701.
[15] S. Smale, M.W. Hirsch, R.L. Devaney, Differential Equations, Dynamical Systems, and an Introduction to Chaos, second ed., Academic Press, New York, USA, 2004.
[16] J.R. Munkres, Topology, second ed., Prentice Hall, Inc., Upper Saddle River, New Jersey, USA, 2000.
[17] D. Lind, B. Marcus, An Introduction to Symbolic Dynamics and Coding, Cambridge University Press, Cambridge, UK, 1995.
[18] B.P. Kitchens, Symbolic Dynamics: One-Sided, Two-Sided and Countable State Markov Chains, Springer-Verlag, New York, USA, 1998.
[19] J.-P. Eckmann, D. Ruelle, Ergodic theory of chaos and strange attractors, Rev. Modern Phys. 57 (1985) 617–656.
[20] E.M. Bollt, N. Santikissadeekorn, Applied Computational Measurable Dynamics, SIAM, Philadelphia, Pennsylvania, USA, 2013, in press.
[21] T.M. Cover, J.A. Thomas, Elements of Information Theory, second ed., John Wiley & Son, Inc., Hoboken, New Jersey, USA, 2006.
[22] E.M. Bollt, J.D. Skufca, Markov partitions, in: A. Scott (Ed.), Encyclopedia of Nonlinear Science, Routledge, New York, USA, 2005.
[23] L.M. Pecora, T.L. Carroll, Master stability functions for synchronized coupled systems, Phys. Rev. Lett. 80 (1998) 2109–2112.
[24] A. Pikovsky, M. Rosenblum, J. Kurths, Synchronization: A Universal Concept in Nonlinear Sciences, Cambridge University Press, Cambridge, UK, 2003.
[25] A. Arenas, A. Díaz-Guilera, J. Kurths, Y. Moreno, C. Zhou, Phys. Rep. 469 (2008) 93–153.
[26] J. Sun, E.M. Bollt, T. Nishikawa, Master stability functions for coupled nearly identical dynamical systems, Europhys. Lett. 85 (2009) 60011.
[27] J. Sun, E.M. Bollt, T. Nishikawa, Constructing generalized synchronization manifolds by manifold equation, SIAM J. Appl. Dyn. Syst. 8 (2009) 202–221.
[28] D.M. Abrams, S.H. Strogatz, Chimera states for coupled oscillators, Phys. Rev. Lett. 93 (2004) 174102.
[29] J. Juang, Y.-H. Liang, Synchronous chaos in coupled map lattices with general connectivity topology, SIAM J. Appl. Dyn. Syst. 7 (2008) 755–765.
[30] I.V. Belykh, V.N. Belykh, M. Hasler, Blinking model and synchronization in small-world networks with a time-varying coupling, Physica D 195 (2004) 188–206.
[31] M. Porfiri, D.J. Stilwell, E.M. Bollt, J.D. Skufca, Random talk: random walk and synchronizability in a moving neighborhood network, Physica D 224 (2006) 102–113.
[32] D.J. Stilwell, E.M. Bollt, D.G. Roberson, Sufficient conditions for fast switching synchronization in time-varying network topologies, SIAM J. Appl. Dyn. Syst. 5 (2006) 140–156.
[33] F. Sorrentino, E. Ott, Adaptive synchronization of dynamics on evolving complex networks, Phys. Rev. Lett. 100 (2008) 114101.
[34] D. Taylor, E. Ott, J.G. Restrepo, Spontaneous synchronization of coupled oscillator systems with frequency adaptation, Phys. Rev. E 81 (2010) 046214.
[35] M.U. Ahmed, D.P. Mandic, Multivariate multiscale entropy: a tool for complexity analysis of multichannel data, Phys. Rev. E 84 (2011) 061918.
[36] A. Kraskov, H. Stögbauer, P. Grassberger, Estimating mutual information, Phys. Rev. E 69 (2004) 066138.
[37] E.M. Bollt, T. Stanford, Y.-C. Lai, K. Życzkowski, Validity of threshold-crossing analysis of symbolic dynamics from chaotic time series, Phys. Rev. Lett. 85 (2000) 3524.
[38] E.M. Bollt, T. Stanford, Y.-C. Lai, K. Życzkowski, What symbolic dynamics do we get with a misplaced partition? On the validity of threshold crossings analysis of chaotic time-series, Physica D 154 (2001) 259–286.