http://www.aimspress.com/journal/mine

Mathematics in Engineering, XYZ(x): xxx–xxx
DOI:10.3934/mine.2018.x.xxx
Received: XYZ
Accepted: XYZ
Published: XYZ

Type of article

Topological features determining the error in the inference of networks using transfer entropy

Roy H. Goodman[1,*] **and Maurizio Porfiri**[2,*]

[1] Department of Mathematical Sciences, New Jersey Institute of Technology, Newark, New Jersey 07102, USA

[2] Department of Mechanical and Aerospace Engineering & Department of Biomedical Engineering, New York University, Tandon School of Engineering, Brooklyn, New York 11201, USA

* **Correspondence:** goodman@njit.edu; Tel: +1-973-642-4261; Fax: +1-973-596-5591, mporfiri@nyu.edu; Tel: +1-646-997-3681; Fax +1-646-997-3532.

Abstract: The problem of inferring interactions from observations of individual behavior in networked dynamical systems is ubiquitous in science and engineering. From brain circuits to financial networks, there is a dire need for robust methodologies that can unveil network structures from individual time series. Originally formulated to identify asymmetries in pairs of coupled dynamical systems, transfer entropy has been proposed as a model-free, computationally-inexpensive framework for network inference. While previous studies have cataloged a library of pathological instances in which transfer entropy-based network reconstruction can fail, we presently lack analytical results that can help quantify the accuracy of the identification and pinpoint scenarios where false inferences results are more likely to be registered. Here, we present a detailed analytical study of a Boolean network model of policy diffusion. Through perturbation theory, we establish a closed-form expression for the transfer entropy between any pair of nodes in the network up to the third order in an expansion parameter that is associated with the spontaneous activity of the nodes. While for slowly-varying dynamics, transfer entropy is successful in capturing the weight of any link, for faster dynamics, the error in the inference is controlled by local topological features of the node pair. Specifically, the error in the inference of a weight between two nodes depends on the mismatch between their weighted in-degrees that serves as a common uncertainty bath upon which we must tackle the inference problem. Interestingly, an equivalent result is discovered when numerically studying a network of coupled chaotic tent maps, suggesting that heterogeneity in the in-degree is a critical factor that can undermine the success of transfer entropy-based network inference.

Keywords: Boolean network, data-driven, discrete systems, information theory, Markov chain, perturbation theory, policy diffusion

1. Introduction

Networks arise in all areas of science and engineering whenever we study the interactions between discrete actors [4]. To understand the functioning of the brain, it is critical to know which regions are sending signals to which other regions [8]. In ecology, the goal may be to understand the structure of the food web, that is to identify which organisms depend on which others for sustenance [10]. In social science or epidemiology, the question may be to identify links between people along which information or disease might spread [13]. Similarly, financial regulators often wish to understand the interdependence of financial institutions that are enmeshed by a web of investments and loans, in order to determine how a shock might ripple through the community [25].

The network topology is not known explicitly in most of these real-world examples, thereby calling for data-driven computational methods to identify links from time-series collected at each node. In this process of network inference, one seeks to determine which actors influence the dynamics of which other actors, from the joint time evolution of their states. Despite considerable progress in network inference, the precise quantification of the errors associated with the implementation of these methods remain elusive. This is the motivating question for the present study.

In mathematical terms, a network is modeled by a graph $\Gamma = (\mathcal{V}, \mathcal{E})$, a collection $\mathcal{V}$ of vertices (or nodes), some subset of which are connected pairwise by a set $\mathcal{E}$ of edges (or links) [11]. In particular, we consider weighted directed networks, in which each edge interpreted as beginning at a node j and end ending at a node i, and to which there is associated a scalar weight W_{ij}. With respect to network inference, given two nodes i and j, one seeks to infer if node j influences the behavior of node i and, in addition, the strength W_{ij} of that influence. In particular, the interest is in *direct influence*, that is, if node i depends on node j and node j on node k, then node k influences node i, but perhaps only indirectly, and it is important to distinguish this from the *direct* influence of j on i.

In an authentic model-free vein, information theory provides a mathematical framework for drawing such inferences based solely on time-series [5]. Information theory grew out of the effort to understand the transmission of signals along noisy communication channels and has strong connections both with statistical physics and with optimal strategies for gambling and investment [9]. The information in an event is associated with the uncertainty in its outcome, with unlikely events possessing more information. Information may flow among the units of a network, and can be used to reconstruct its topology. For example, Squartini and colleagues [25] have recently reviewed information-theoretic results for network inference, focusing on financial networks, and Wibral and colleagues [29] have presented an overview of the state-of-the-art of information-theoretic methods in neuroscience.

Introduced almost twenty years ago by Schreiber [24] to quantify asymmetries in coupled dynamical systems, transfer entropy is emerging as an approach of choice to guide the process of network inference from time-series. A thorough survey of its mathematical properties and many applications is provided in [6]. Given a pair of time-series, transfer entropy measures the reduction in uncertainty in predicting the future state of one of the time-series from its present value, compared with predicting its future state from its present value and that of the second series. In the context of network inference, one predicts that a link is present from node j to node i given a value of transfer entropy from node j to node i that is statistically different from zero.

While network inference based on transfer entropy is intuitive and computationally inexpensive, it is not free of technical limitations that may beget false positive and negative results. Most of these lim-

itations are rooted in the pairwise definition of transfer entropy that does not contemplate the existence of indirect influence from any other node in the network. Extending the transfer entropy definition to multivariate interactions could address some of these limitations, but it would require the estimation of high dimensional probability mass functions that could be computationally unfeasible beyond simple motifs [6]. While Runge [23] an Sun and colleagues [27] pinpoint several pathological instances in which transfer entropy-based network reconstruction can fail, we currently lack analytical results that can help assess the accuracy of the inference and determine scenarios where false results should be expected. Filling this gap in knowledge is the chief aim of this study.

Specifically, we continue the mathematically-principled treatment of a random Boolean network (RBN) model proposed in [19] as a minimalistic representation of policy diffusion. The model is a simplification of the more general setup presented in [2, 12] to describe enactment and changes to alcohol-related policies in the 50 state of the United States of America from 1080. In the RBN, each node is assigned a Boolean variable, whose probability of activation at a given time-step depends on both the state of its neighboring nodes and on its own internal dynamics, each of which is specified by weighting parameters. The model possesses a small parameter that allows for the use of perturbation methods to compute closed-form equations approximating the relations between system parameters and the transfer entropy, allowing the reconstruction of the former from the latter. In [18], the analysis was completed and extended to the case in which the parameters vary with a prescribed temporal period τ, while here we assume, for simplicity, that it is time-independent.

The model provides an interesting test case for a number of reasons. Foremost is that its explicit and linear mathematical formulation allows for closed-form analysis that is not possible for real-world datasets, nor for most nonlinear mathematical models. Second is that because it is defined in terms predetermined parameters, the ground truth is known and deviations from that ground truth can be exactly measured. It is exceptionally simple and fast to evaluate, allowing us to compute long time-series that would be unavailable in most experimental settings. Finally, because the state space of the model is binary, computation of the transfer entropy is especially fast and inexpensive.

The RBN model consists of N nodes with state $X^i, i = 1 \dots N$, each of which can take the two states zero or one. The system's state at a given time depends on its previous state according to the linear transition law

$$\Pr\left[X^i(t+1) = 1 | X^1(t) = x^1, \dots, X^N(t) = x^N\right] = \Theta\left[1 + \sum_{j=1}^{N} W_{ij} x^j\right]. \tag{1.1}$$

The coefficient Θ represents the probability of spontaneous activation of each node, while the term W_{ij} represents the nonnegative weighted influence on node i from its neighbor j. We assume for simplicity that the network contains no self-connected edges, that is,

$$W_{ii} \equiv 0. \tag{1.2}$$

In the context of policy diffusion, Θ measures the tendency of a legal unit to spontaneously enact or change a policy in the absence of any interaction with other legal units. With reference to real policy-making in the United States of America, adopting a time-step of a month would yield small values of Θ for policies that have a high start-up cost and unknown, long-term, benefits as well as for policies that attend rare, specific problems that could have significant opposition [2, 12]. The RBN

shares similarities with synapsis models in theoretical neuroscience, where influence between neurons is represented through excitatory networks [22]. This is the case of slowly-varying dynamics on which we focus, thereby enabling a perturbation argument that treats Θ as a small parameter.

The primary analytical result of [18] is a closed-form expression relating the transfer entropy from j to i with the weight W_{ij}, namely,

$$\mathrm{TE}^{j\to i} = \Theta^2 G^{(2)}(W_{ij}) + O\left(\Theta^3\right). \tag{1.3}$$

where $G^{(2)}$ is the one-to-one map

$$G^{(2)}(x) = -x + (1+x)\log(1+x). \tag{1.4}$$

This expression was derived earlier in [19] where W_{ij} was assumed to take only the binary values zero or one. In [18], it is proposed that truncating this expression and solving for W_{ij} should be an effective way to estimate the network weights. This is shown via a handful of examples to largely reproduce the network structure of several small time-dependent networks and one larger random network. These examples include networks with as few as two and as many as 100 nodes.

In the present work, we compute the next-order, $O\left(\Theta^3\right)$, term in this approximation, providing an estimate of the error associated with indirect influence from other nodes in the network. Approximation (1.3) is local in the sense that the weight corresponding to a link (i, j) is determined entirely by the transfer entropy from j to i. The correction, by contrast, depends on the difference between the total weighted inputs incident on each of the nodes i and j, revealing the effect of the more global network structure on the transfer entropy.

This paper is organized as follows. In Sec. 2 we introduce our motivating example, whose surprising behavior we wish to explain. In Sec. 3, we summarize additional mathematical background to this problem. In Sec. 4, we calculate the next order term in the expansion (1.3) to understand the errors made in the example calculation. Then in Sec. 5 we return to this example, which is illuminated by our calculation. To offer evidence in favor of the generality of the analysis, we numerically study a network of coupled chaotic maps in Sec. 6. Finally, we present our main conclusions in Sec. 7.

2. A surprising example

We begin by describing the network at the center of the example. The Barabási-Albert (BA) network was introduced in 1999 to model the network structure of the internet, although related models date back at least to Yule [1, 30]. Such a network is constructed by an algorithm exhibiting preferential attachment that proceeds as follows. Begin with a small 'seed' graph, which we take to be the complete network with m_0 nodes. Then at step $k+1 = m_0+1, \ldots N$, choose m existing nodes at random according to the weights

$$p_j = \frac{d_j}{\sum_{i=1}^{k} d_i},$$

where d_i is the in-degree of node i (the number of links pointing toward i), and create links between node $k+1$ and the m chosen nodes. By construction, the network have no self-directed edges connecting a node to itself (assuming that the seed network satisfies this property.)

The standard BA network is undirected; Prettejohn et al. describe a directed network variation [20]. In this model, the links added at each step are directed edges from node $k+1$ to node j. If that were

the entire model all links would point toward the seed graph, and none would point away from it. In the modified version, once such a directed edge is chosen, then with the same probability p_j another directed edge is created in the opposite direction.

After constructing such a network Γ we introduce a random weight W_{ij}, chosen from the uniform distribution on $[0, 1]$, to each link from a node j to a node i. The elements of the weight matrix W in (1.1) represent the influence of j on i. It is visualized in Fig. 1. The out-degree of each node, that is, the number of nonzero entries in each row is fairly homogeneous: 40 of the 50 nodes have out-degree equal to five, while the in-degree (number of nonzero entries per column) is much more heterogeneous, with 15 different values ranging from 0 to 20.

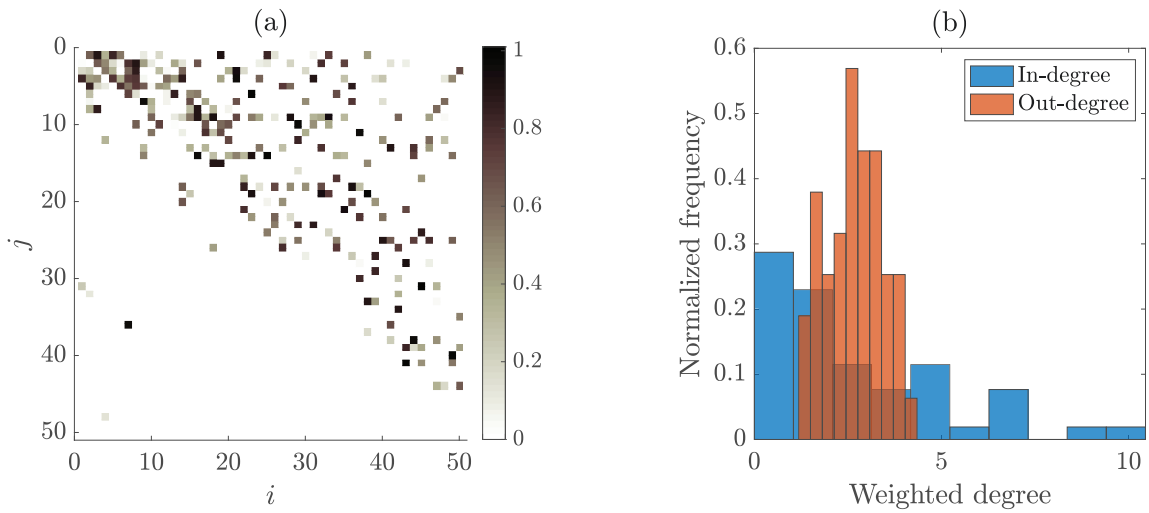

Figure 1. (a) Visualization of the matrix W with larger entries shaded darker. **(b)** Histograms of the in- and out-degree of Γ, measuring the distribution of the weighted in- and out-degrees in the network. The weighted in-degree of a node is defined as the corresponding row-sum of W, and, similarly, the out-degree is the column-sum. This histogram is normalized to have unit area, as are all others in the paper. The weighted in-degree and out-degree both have mean 2.7. The in-degree has variance 6.7, while the out-degree has variance 0.58.

In the computational example we apply the result of [18] to reconstruct two distinct but related networks. The first is described above and the second is its transpose Γ^{T}, obtained by reversing the orientation of each directed link. The weight matrix for Γ^{T} is simply the transposed weight matrix W^{T}. The in-degree of a node i of Γ^{T} is equal to the out-degree of the same node in Γ, and vice-versa.

To analyze the model, we run 100 time-series each of 10^5 steps. For each of the 2,450 pairs (i, j) (excluding the terms with $i = j$ since there are no self-loops by construction), we compute the transfer entropy $\mathrm{TE}^{j\to i}$ via a simple plug-in estimation [16] where we count the occurrences of the possible states that define the interaction between the node pair. Then, using the leading-order term in (1.3), an approximate value of the corresponding weight. We utilize $\Theta = 0.05$ so that the probability on the right-hand side of (1.1) is less than one for all i. We average the computed value of the transfer entropy over the 100 realizations. If the $\mathcal{O}\left(\Theta^3\right)$ error term in (1.3) is negligible, then the inference of the influence of j on i in Γ should be identical to the influence of i on j in Γ^{T}.

In Fig. 2, we plot the inferred weights as a function of their exact values for both these matrices,

here showing only the 262 nonzero entries. When the weight is small, the computation significantly overestimates it in both cases. For larger weights, there is an increased amount of variation in the predicted values, but it is distributed in markedly different ways for the two matrices. For the network Γ, the variation appears to be distributed evenly above and below the exact value. For the network Γ^{T}, the error in the inference is nearly always positive, and appears to be significantly smaller. This is confirmed in the right panel of the figure, which shows histograms of the errors, normalized to have unit area. The histogram for the network Γ has mean close to zero and is considerably broader than the histogram for Γ^{T}, which has a positive mean value.

Since most of the entries of the weight matrix W are zero, a critical measure of success in the network reconstruction is minimizing false positives, that is, links that are inferred between disconnected nodes. The overestimation of weights lower than 0.1 evident in Fig. 2 suggests that some false positives are unavoidable. Of the 262 largest inferred weights for both Γ and Γ^{T}, 11 entries for W and 13 for W^{T} would be classified as false positives. Note that this computation was performed on the computation based on the transfer entropy averaged over 100 realizations. Significantly more false positives were found if the calculation was performed using a single realization.

Clearly, the calculated errors in the weights in the calculation for Γ are larger than for the similar calculation using Γ^{T}, despite the fact that equation (1.3) predicts the same leading-order accuracy. The analysis of this paper aims to determine what topological features of the two networks conspire to create this difference and to correct it.

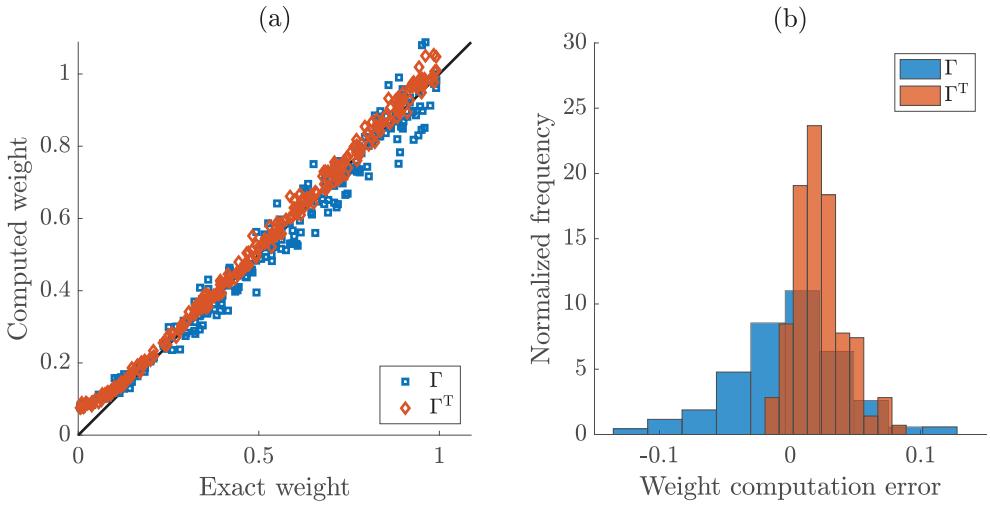

Figure 2. **(a)** Computed values of the nonzero matrix entries vs. the exact values for both Γ and Γ^{T}. **(b)** Histogram of the error in the computed approximations for the nonzero entries of the two matrices. For Γ the inference error has mean -7.8×10^{-4} and variance 1.8×10^{-3}. For Γ^{T} the inference error has mean 2.3×10^{-2} and variance 4.8×10^{-4}.

3. Mathematical background

3.1. Markov chains

We assume familiarity with the basics of Markov chains and use this section mainly to set notation [7]. We consider a discrete-time finite-state Markov chain $Z(t)$, $t \in \mathbb{N}$ evolving in a sample space $\mathcal{Z}$ whose generic element is denoted as z_i, $i = 1 \dots |\mathcal{Z}|$ where $|\mathcal{Z}|$ is the cardinality of the set. Lower case letters will be used to denote realizations of random variables. The transition matrix is a matrix $P \in \mathbb{R}_+^{|\mathcal{Z}|\times|\mathcal{Z}|}$, whose columns each sum to one. The entry P_{ij} represents the probability that the the Markov chain will transition from state z_i to state z_j at any given step,

$$P_{ij} = \Pr[Z(t+1) = z_j | Z(t) = z_i].$$

Considered as a row vector, the distribution $\nu(t)$ evolves according to the following recursion,

$$\nu(t+1) = \nu(t)P, \tag{3.1}$$

for $t \in \mathbb{N}$ and initial distribution $\nu(0) = \nu_0$.

The solution to this recursion is a probability distribution for all times t, such that its entries are non-negative and sum to one. Given mild assumptions on the transition matrix [7] P the recursion converges exponentially to a unique limiting distribution denoted π as $t \to \infty$, called the stationary distribution. This is given by the left eigenvector with unit eigenvalue,

$$\pi = \pi P, \quad \sum_{i=1}^{N} \pi_i = 1. \tag{3.2}$$

3.2. Elements of information theory

Our goal is to reconstruct the network Γ from a time-series of finite duration generated by the dynamics of (1.3). Since this time series is finite, it is impossible to determine with certainty the matrix used to create it. The object, then, is to find the network that is consistent with the data while requiring the fewest unsupported assumptions. Loosely speaking, we can say that it should not be too surprising [9]. The idea of surprise is that unlikely events convey more information, so a "surprise function" should assign a large value to unlikely events and a small value to expected events.

The information associated with the event that a random variable X drawn from a sample space $\mathcal{X}$ takes the value x is

$$I(x) = -\log \Pr(X = x).$$

Here, we make use of natural logarithms, so that information is measured in *nats*. The (Shannon) entropy of the random variable X is then the expectation of the information

$$H(X) = \mathrm{E}[I(X)] = -\sum_{x\in\mathcal{X}} \Pr(X = x) \log \Pr(X = x), \tag{3.3}$$

which quantifies the amount of uncertainty in X, where E denotes expectation. The entropy is the unique functional over probability distributions that satisfies the Shannon-Khinchin axioms [14], making it the best choice for measuring the uncertainty of a distribution. Since the entropy is the expectation of $I(X)$, it is generically non-negative.

Given another random variable Y, the notions of joint and conditional entropies are similarly defined by

$$H(X,Y) = -\sum_{x\in\mathcal{X},y\in\mathcal{Y}} \Pr(X=x,Y=y)\log\Pr(X=x,Y=y), \tag{3.4a}$$

$$H(X|Y) = -\sum_{x\in\mathcal{X},y\in\mathcal{Y}} \Pr(X=x,Y=y)\log\Pr(X=x|Y=y), \tag{3.4b}$$

where $\mathcal{Y}$ is the sample space of Y.

These notions can be readily extended to stochastic processes, such that given two stationary processes X and Y, transfer entropy from Y to X is defined as [24]

$$\begin{aligned}\mathrm{TE}^{Y\to X} &= H(X(t+1)|X(t)) - H(X(t+1)|X(t),Y(t)) \\ &= \sum_{\substack{x_+\in\mathcal{X}\\ x\in\mathcal{X}\\ y\in\mathcal{Y}}} \left\{\Pr\left[X(t+1)=x_+, X(t)=x, Y(t)=y\right]\times\log\frac{\Pr\left[X(t+1)=x_+|X(t)=x,Y(t)=y\right]}{\Pr\left[X(t+1)=x_+|X(t)=x\right]}\right\}.\end{aligned} \tag{3.5}$$

Transfer entropy measures the reduction in the uncertainty of predicting $X(t+1)$ from both $X(t)$ and $Y(t)$ relative to predicting it from $X(t)$ alone. As a simple consequence of its definition, transfer entropy is non-negative.

We acknowledge that building on this definition, a variety of amelioration could be undertaken. For example, one may consider delayed interactions between the two processes, such that $Y(t)$ in (3.5) should be replaced with $Y(t-\delta)$ with δ being a suitable time-delay [28]. Also, one may attempt at a symbolic treatment of the time-series [26], or pursue an analysis in terms of recurrence plots [21]. In this work, we focus on the classical definition in (3.5) toward gathering analytical insight into the role of the topology on the accuracy of transfer entropy-based network inference.

4. Analysis

We briefly present the computation of the next-order correction in the transfer entropy formula (1.3). This necessarily builds on the calculation in [18], the details of which are summarized below. As in most perturbative calculations [15], the number and complexity of terms greatly increases as the order of the calculation increases. The details of this calculation were assisted and verified using Mathematica.

4.1. The stationary distribution

System (1.1) can be reformulated as a Markov chain, with $\mathcal{Z} = 2^N$ states $z = [X^1,\dots,X^N] \in \{0,1\}^N$, such that each state is a binary vector whose entries take the value one or zero depending on whether that node is active. Letting $Z(t) = [X^1(t),\dots,X^N(t)]$, then the transition probability in (1.1) can be written as

$$\Pr\left[X^i(t+1)=1|Z(t)=z\right] = \Theta\left[1+e_i^{\mathrm{T}}Wz\right], \tag{4.1}$$

where the unit vector e_i is column i of the identity matrix. For both the possible realizations of $X^i(t+1)$, this can be written as

$$\Pr\left[X^i(t+1)=x_+^i|Z(t)=z\right] = \left(1-x_+^i\right) + \Theta\left(2x_+^i-1\right)\left[1+e_i^{\mathrm{T}}Wz\right]. \tag{4.2}$$

Taking a product of such terms yields the transition matrix of this Markov chain

$$P_{ij}(t) = \Pr\left[Z(t+1) = z_j | Z(t) = z_i\right] = \prod_{k=1}^{N} \left\{\left(1 - e_k^{\mathrm{T}} z_j\right) + \Theta\left(2e_k^{\mathrm{T}} z_j - 1\right)\left[1 + e_k^{\mathrm{T}} W z_i\right]\right\} \tag{4.3}$$

The transition matrix P can be expanded in powers of Θ as follows:

$$P = P^{(0)} + \Theta P^{(1)} + \Theta^2 P^{(2)} + \Theta^3 P^{(3)} + O\left(\Theta^4\right).$$

The first three terms are given in [18]. Specifically, the zeroth order term is

$$P_{ij}^{(0)} = \prod_{k=1}^{N} \left[\left(1 - e_k^{\mathrm{T}} z_j\right)\right] = \begin{cases} 1 & \|z_j\| = 0, \\ 0 & \|z_j\| > 0. \end{cases} \tag{4.4a}$$

Here, the norm of a binary vector is taken to be the number of nonzero entries, namely,

$$\|z\| = \sum_{i=1}^{N} X^i.$$

The first order term is

$$P_{ij}^{(1)} = \sum_{r=1}^{N} \left\{\left(2e_r^{\mathrm{T}} z_j - 1\right)\left[1 + e_r^{\mathrm{T}} W z_i\right] \prod_{\substack{k=1 \\ k \neq r}}^{N} \left(1 - e_k^{\mathrm{T}} z_j\right)\right\} = \begin{cases} -\left[N + 1_N^{\mathrm{T}} W z_i\right] & \|z_j\| = 0, \\ \left[1 + z_j^{\mathrm{T}} W z_i\right] & \|z_j\| = 1, \\ 0 & \|z_j\| > 1, \end{cases} \tag{4.4b}$$

where 1_N denotes a column vector of N ones. Finally, the second order term is

$$\begin{aligned} P_{ij}^{(2)} &= \sum_{\substack{r,s=1 \\ r>s}}^{N} \left\{\left(2e_r^{\mathrm{T}} z_j - 1\right)\left[1 + e_r^{\mathrm{T}} W z_i\right]\left(2e_s^{\mathrm{T}} z_j - 1\right)\left[1 + e_s^{\mathrm{T}} W z_i\right] \prod_{\substack{k=1 \\ k \neq r,s}}^{N} \left(1 - e_k^{\mathrm{T}} z_j\right)\right\} \\ &= \begin{cases} \sum_{\substack{r,s=1 \\ r>s}}^{N} \left\{\left[1 + e_r^{\mathrm{T}} W z_i\right]\left[1 + e_s^{\mathrm{T}} W z_i\right]\right\} & \|z_j\| = 0, \\ -\left[1 + z_j^{\mathrm{T}} W z_i\right]\left[N - 1 + \left(1_N^{\mathrm{T}} - z_j^{\mathrm{T}}\right) W z_i\right] & \|z_j\| = 1, \\ \left\{1 + z_j^{\mathrm{T}} W z_i + \left[e_{\mathcal{I}_1(z)}^{\mathrm{T}} W z_i\right]\left[e_{\mathcal{I}_2(z_j)}^{\mathrm{T}} W z_i\right]\right\} & \|z_j\| = 2. \\ 0 & \|z_j\| > 2, \end{cases} \end{aligned} \tag{4.4c}$$

where $\mathcal{I}_1(z_j)$ and $\mathcal{I}_2(z_j)$ are used to identify the two entries of z_j that are different from zero for the case $\|z_j\| = 2$.

In a similar manner, we compute $P^{(3)}$. The calculation below will only need the values of $P^{(3)}$ for which $\|z_i\| = 0$. Accordingly, that is all we report here

$$P_{ij}^{(3)}\Big|_{z_i=0} = \begin{cases} -\binom{N}{3}, & \|z_j\| = 0 \\ \binom{N-1}{2}, & \|z_j\| = 1 \\ -(N-2), & \|z_j\| = 2 \\ 1, & \|z_j\| = 3 \\ 0, & \|z_j\| > 3. \end{cases} \tag{4.5}$$

We remark that $P_{ij}^{(n)} \neq 0$ only in the case that both $\|z_i\| \leq n$ and $\left\|z_j\right\| \leq n$. There are of course $\binom{N}{n}$ vectors z with $\|z\| = n$, so that if $n \ll N$ then the approximate matrix, computed for all transitions with $\|z\| \leq n$, is low rank. Given a time- series of system (1.1), this provides a way to check whether Θ has been chosen small enough that the proposed approximation is valid.

Next, we expand the stationary distribution to the third order, similar to the transition matrix,

$$\pi = \sum_{j=0}^{3} \Theta^j \pi^{(j)} + O\left(\Theta^4\right), \tag{4.6}$$

By replacing (4.6) into (3.2) and grouping terms of the same power in Θ, we determine the following chain of relationships:

$$\pi^{(0)}\left(I - P^{(0)}\right) = 0, \qquad \sum_{j=1}^{N} \pi_j^{(0)} = 1, \tag{4.7a}$$

$$\pi^{(1)}\left(I - P^{(0)}\right) = \pi^{(0)} P^{(1)}, \qquad \sum_{j=1}^{N} \pi_j^{(1)} = 0, \tag{4.7b}$$

$$\pi^{(2)}\left(I - P^{(0)}\right) = \pi^{(0)} P^{(2)} + \pi^{(1)} P^{(1)}, \qquad \sum_{j=1}^{N} \pi_j^{(2)} = 0, \tag{4.7c}$$

$$\pi^{(3)}\left(I - P^{(0)}\right) = \pi^{(0)} P^{(3)} + \pi^{(1)} P^{(2)} + \pi^{(2)} P^{(1)}, \qquad \sum_{j=1}^{N} \pi_j^{(3)} = 0. \tag{4.7d}$$

The left column of equations contains the expansion of the stationary equation, while the right column enforces the condition that π is a probability distribution. The matrix $\left(I - P^{(0)}\right)$ in each equation is singular, with a one-dimensional null space. The first equation (4.7a) is solved by its null vector. The other equations are all solvable, as long as the vectors on their right-hand sides sum to zero, and they all do.

The first three equations in (4.7) for the zeroth, first, and second order terms in the expansion are solved in [18] and are given by

$$\pi_i^{(0)} = \begin{cases} 1 & \|z_i\| = 0, \\ 0 & \|z_i\| > 0, \end{cases} \tag{4.8a}$$

$$\pi_i^{(1)} = \begin{cases} -N & \|z_i\| = 0, \\ 1 & \|z_i\| = 1, \\ 0 & \|z_i\| > 1, \end{cases} \tag{4.8b}$$

and

$$\pi_i^{(2)} = \begin{cases} \binom{N}{2} - 1_N^{\mathrm{T}} W 1_N & \|z_i\| = 0, \\ -(N-1) + z_i^{\mathrm{T}} W 1_N & \|z_i\| = 1, \\ 1 & \|z_i\| = 2, \\ 0 & \|z_i\| > 2, \end{cases} \tag{4.8c}$$

To extend the computation to the next order, we replace (4.4) and (4.8) into (4.7d). By solving the equation and imposing the constraint that the perturbation is zero-sum, we determine

$$\pi^{(3)} = \begin{cases} -\binom{N}{3} + (N-1)1_N^{\mathrm{T}} W 1_N + \frac{1}{2} 1_N^{\mathrm{T}} \left(W W^{\mathrm{T}} - W \circ W \right) 1_N - 1_N^{\mathrm{T}} W^2 1_N & \left\| z_j \right\| = 0, \\ \binom{N-1}{2} - (N-2) z_i^{\mathrm{T}} W 1_N - 1_N^{\mathrm{T}} W 1_N & \\ \qquad + \sum_{k=1}^{N} \left(\left(z_i^{\mathrm{T}} W e_k \right)^2 + (z_i^{\mathrm{T}} W e_k) \left(e_k^{\mathrm{T}} W 1_N - 1_N^{\mathrm{T}} W e_k \right) \right) & \|z_i\| = 1, \\ -(N-2) + z_i^{\mathrm{T}} W 1_N + \sum_{k=1}^{N} (e_{\mathcal{I}_1(z_i)} W e_k)(e_{\mathcal{I}_2(z_i)} W e_k) & \|z_i\| = 2, \\ 1 & \|z_i\| = 3, \\ 0 & \|z_i\| > 3, \end{cases} \tag{4.9}$$

where the Hadamard product [3], or element-wise product, $(A \circ B)_{ij} = A_{ij} B_{ij}$, is in the $\left\| z_j \right\| = 0$ term.

Marginalizing (4.6) via (4.8) and (4.9), we compute the stationary periodic distribution of each node

$$\Pr(X^i(t) = x^i) = (1 - x^i) + (2x^i - 1) \left(\Theta + \Theta^2 e_i^{\mathrm{T}} W 1_N + \Theta^3 \sum_{k=1}^{N} (e_k^{\mathrm{T}} W 1_N)(e_i^{\mathrm{T}} W e_k) \right) + \mathcal{O}\left(\Theta^4\right). \tag{4.10}$$

This distribution describes the probability that a node is active or non-active at a given time. Should the expansion be truncated at the second order in Θ, only the row sum of W at node i (quantifying the weighted in-degree of node i) will enter the probability distribution in (4.10). These nodes are those that influence node i with respect to system (1.1). Extending the analysis to the third power in Θ brings forward a more complex dependence of node i on the other nodes in the network, in the form of a weighted sum of all the elements of the matrix W (including those that are influenced by node i).

Similarly, marginalizing (4.6) with respect to any nodes but node pair ij, we determine the following joint distribution for ij:

$$\begin{aligned} \Pr\left(X^i(t) = x^i, X^j(t) = x^j \right) = \Pr\left(X^i(t) = x^i \right) \Pr\left(X^j(t) = x^j \right) \\ + (2x^i - 1)(2x^j - 1) e_i^{\mathrm{T}} W W^{\mathrm{T}} e_j \Theta^3 + \mathcal{O}\left(\Theta^4\right). \end{aligned} \tag{4.11}$$

While a second order expansion in Θ would yield that node i and j are independent, retaining the third order power yields a different claim. Specifically, (4.11) cannot be obtained from (4.10) if we are interested in the quantification of the third order term in Θ.

4.2. Transfer entropy

We now develop the two leading-order terms in the expansion of the transfer entropy, using the definition (3.5) and the expansion (4.6) developed in the last section. Without loss of generality we consider the transfer entropy from node 2 to node 1

$$\mathrm{TE}^{2\to 1} = \sum_{x_+^1, x^1, x^2} \Pr\left[X^1(t+1) = x_+^1, X^1(t) = x^1, X^2(t) = x^2 \right] \log \frac{\Pr\left[X^1(t+1) = x_+^1 | X^1(t) = x^1, X^2(t) = x^2 \right]}{\Pr\left[X^1(t+1) = x_+^1 | X^1(t) = x^1 \right]}, \tag{4.12}$$

where the probabilities are evaluated using the stationary distributions derived in the previous section.

We can compute the conditional probability in the denominator of the logarithm term using the transition probability (4.2) and properly marginalizing the stationary distribution (4.6), as follows:

$$\Pr\left[X^1(t+1) = x^1_+ | X^1(t) = x^1\right] = \left(1 - x^1_+\right) + \left(2x^1_+ - 1\right)\left(\Theta + \Theta^2 d_1 + \Theta^3\left(e_1^{\mathrm{T}} W^2 1_N + x^1 e_1^{\mathrm{T}} W^2 W^{\mathrm{T}} e_1\right)\right) + \mathcal{O}\left(\Theta^4\right) \quad (4.13)$$

where

$$d_j = \sum_{k=1}^{N} W_{jk}. \quad (4.14)$$

is the weighted in-degree of node j. In agreement with one's expectation, for small values of Θ, such that an affine expansion would hold, the probability of a transition does not depend on the network topology. Retaining the second order power in Θ, introduces a dependence on all the nodes that influence node 1, through d_1. Increasing further the order of the expansion, we observe a richer influence from the entire network, expressed through quadratic and cubic terms in the elements of W.

A similar calculation yields the term in the numerator of (4.12). By using (4.11) and (4.6), we establish

$$\Pr\left[X^1(t+1) = x^1_+ | X^1(t) = x^1, X^2(t) = x^2\right] = \left(1 - x^1_+\right) + \left(2x^1_+ - 1\right)\left[\Theta\left[1 + W_{12}x^2\right] + \Theta^2(d_1 - W_{12}) + \Theta^3\left(\tilde{w}^1_2 W 1_N + \tilde{w}^1_2 W W^{\mathrm{T}}\left(x^1 e_1 + x^2 e_2\right) + x^1 x^2 \sum_{k=3}^{N}\left(W_{1k}\sum_{j=1}^{N} W_{1j}W_{2j}W_{kj}\right)\right)\right] + \mathcal{O}\left(\Theta^4\right) \quad (4.15)$$

where $\tilde{w}^1_2 = (e_1^{\mathrm{T}} W - W_{12} e_2^{\mathrm{T}})$ identifies the first row of W with its second element set to zero. Similar to (4.13), for small values of Θ, the probability of node 1 to transition to a give state conditional to its present state and the present styate of node 2 depends only on W_{12}. Reaching to the second power in Θ brings about the effect of all the other nodes that influence node 1 (excluding node 2). The third power in Θ depicts a further degree of interaction, wherein the entire network contributes to the probability of transition (4.15), involving both cubic and quartic terms in terms of the elements of W.

The joint probability $\Pr\left[X^1(t+1) = x^1_+, X^1(t) = x^1, X^2(t) = x^2\right]$ in (4.12) is calculated as the product of (4.11) and (4.15). From this joint probability and the conditional probabilities in (4.13) and (4.15), we can calculate an expansion for transfer entropy. Switching from labeling the indices from $(1, 2)$ to (i, j) with $i \neq j$, we establish

$$\mathrm{TE}^{j\to i} = \Theta^2 G^{(2)}(W_{ij}) + \Theta^3 G^{(3)}_{ij}(W) + \mathcal{O}\left(\Theta^4\right) \quad (4.16)$$

where the scalar function $G^{(2)}$ is defined in (1.4) and the matrix function $G^{(3)}$ is defined entry-wise as

$$G^{(3)}_{ij}(W) = W_{ij}\left(W_{ij} - d_i - d_j\right) + \log\left(1 + W_{ij}\right)\left(d_i - W_{ij} + (1 + W_{ij})d_j\right). \quad (4.17)$$

In agreement with the discussion on the effect of the network network topology on local joint and transition probabilities in (4.13) and (4.15), the asymptotic expansion for transfer entropy in (4.16) suggests that increasing the order of the expansion in $\mathrm{TE}^{j\to i}$ hampers the possibility of mapping transfer entropy one-to-one with the corresponding weights. Surprisingly, the term in Θ^3 depends on the influence of j on i along with the in-degree of both i and j, which encapsulate the overall influence of any other network node on the pair.

4.3. Solving for the weights

In the process of network inference, we should compute all the entries of W from transfer entropy values estimated from time-series. To accomplish this goal, we must solve the system of equations (4.16) with $i, j = 1, \ldots, N$, $i \neq j$, for W. For convenience, we scale transfer entropy by Θ^2, yielding the following matrix-valued equation for W

$$T_{ij} = G^{(2)}(W_{ij}) + \Theta G^{(3)}_{ij}(W) + \mathcal{O}\left(\Theta^2\right), \quad i, j = 1, \ldots, N, \; i \neq j \tag{4.18}$$

where $T_{ij} = \Theta^{-2}\mathrm{TE}^{j\to 1}$.

Formally, we can solve (4.18) for W as a series in Θ, as follows:

$$W = W^{(0)} + \Theta W^{(1)} + \mathcal{O}\left(\Theta^2\right)$$

and find the two equations

$$\begin{aligned} \mathcal{O}(1): &\quad T_{ij} = G^{(2)}\left(W^{(0)}_{ij}\right), \\ \mathcal{O}(\Theta): &\quad 0 = \left[G^{(2)}\right]'\left(W^{(0)}_{ij}\right) W^{(1)}_{ij} + G^{(3)}_{ij}\left(W^{(0)}\right). \end{aligned}$$

These have formal solutions

$$W^{(0)}_{ij} = \left[G^{(2)}\right]^{-1}(T_{ij}) \tag{4.19a}$$

$$W^{(1)}_{ij} = -\frac{G^{(3)}_{ij}\left(W^{(0)}\right)}{\left[G^{(2)}\right]'\left(W^{(0)}_{ij}\right)} \tag{4.19b}$$

Ultimately, we compute any element of W by following these steps. *(i)* For every pair $i \neq j$, we solve (4.19a) to compute value $W^{(0)}_{ij}$ and we assemble the matrix $W^{(0)}$ (with zeros on the diagonal). *(ii)* For every pair $i \neq j$, we compute the correction $W^{(1)}_{ij}$ by using the entire matrix $W^{(0)}$.

To shed light on the topological underpinnings of the correction $W^{(1)}$ that relates $\mathrm{TE}^{j\to i}$ to weights beyond that between i and j, we can perform a further perturbation analysis. Specifically, we can estimate the correction term in (4.19b) in the case that the entries in W are small using Taylor series. As shown in [18], $G^{(2)}(W_{ij}) \sim \frac{W^2_{ij}}{2}$. Similarly, expanding (4.17) under the assumption that $W_{ij} \ll 1$ but the other entries are $\mathcal{O}(1)$, we find

$$G^{(3)}_{ij}(W) \sim \frac{W^2_{ij}}{2}\left(d_j - d_i + W_{ij}\right) + \mathcal{O}\left(\|W\|^3\right). \tag{4.20}$$

Therefore the correction (4.19b) approximately satisfies

$$W^{(1)}_{ij} \sim \frac{W^{(0)}_{ij}}{2}\left(d^{(0)}_j - d^{(0)}_i + W^{(0)}_{ij}\right) + \mathcal{O}\left(\left\|W^{(0)}\right\|^2\right), \tag{4.21}$$

where $d^{(0)}_i$ is the in-degree of node i with respect to matrix $W^{(0)}$, which, based on (4.19a) encapsulates the overall information transfer to node i from its neighbors. The same expansion holds, with a different remainder, if $W_{ij} \ll 1$ but the other weights in (4.17) are $\mathcal{O}(1)$.

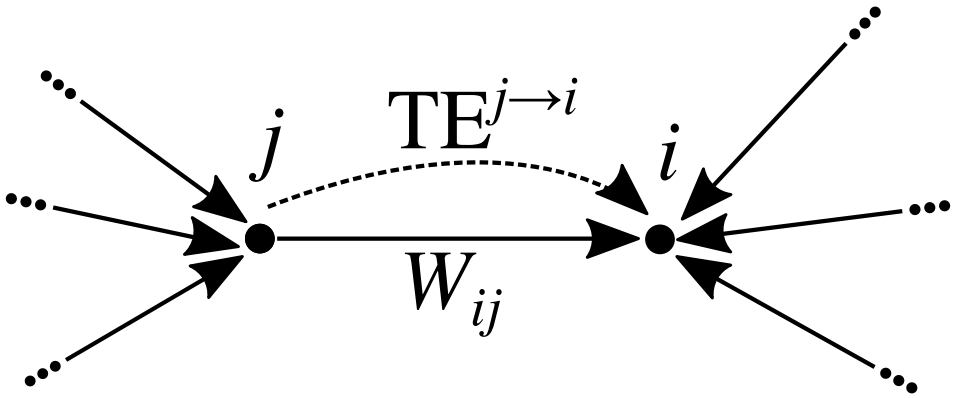

Figure 3. Schematic of a graph, showing a link from node i to node j with weight W_{ij}, as well as the additional links pointing to these nodes from other parts of the graph. Node j has in-degree 3 and out-degree 1. Node i has in-degree 4 and out-degree 0. Whereas only the edge labeled W_{ij} contributes $\mathrm{TE}^{j\to i}$ at leading order. The correction depends on all the pictured edges.

Overall, (4.21) indicates that $W_{ij}^{(1)}$ depends on all the transfer entropy values associated with incoming information to nodes i and j. Note that $d_i^{(0)}$ and $d_j^{(0)}$ have opposite signs in the term (4.21). Specifically, the correction in the inference of the weight between nodes i and j can be split into two terms $W_{ij}^{\mathrm{link}} = \frac{1}{2}\left(W_{ij}^{(0)}\right)^2$ and $W_{ij}^{\mathrm{topology}} = \frac{1}{2}W_{ij}^{(0)}(d_j^{(0)} - d_i^{(0)})$. The first of these, W_{ij}^{link}, represents a correction that depends only on the computed transfer entropy for the pair ij. The second, by contrast, depends on the mismatch between their weighted in-degrees from $W^{(0)}$ that serves as a common uncertainty bath upon which we must tackle the inference problem. The computation of $W_{ij}^{(0)}$ treats the pair ij in isolation, while the correction term $W_{ij}^{\mathrm{topology}}$ accounts for the information which both nodes i and j receive from the rest of the network; see Fig. 3. We attribute the different behaviors of the simulations reported in Sec. 2 to this observation, as we relate in the next Section.

We further comment that to this order in the expansion, equation (4.17) yields that if $W_{ij} = 0$ then $T_{ij} = 0$. Since transfer entropy $\mathrm{TE}^{j\to i}$ computed by time series is never exactly zero, this suggests that the third order approximation will do no better than the second-order approximation at discriminating between $W_{ij} = 0$ and W_{ij} positive but small. However, it can contribute to the accuracy of the identification of nonzero weights, in terms of both the mean and variance of the prediction.

5. Continued numerical example

If transfer entropy $\mathrm{TE}^{j\to i}$ were a function solely of the weight W_{ij} and not on further topological properties of the graph, then the inferred weights for a network Γ and its transpose, computed to leading order, should be identical. That they were not found to be identical in Sec. 2 motivated the computation, via perturbation expansion, of the next order correction terms, detailed in Sec. 4. Equation (4.21) provides an intuitive explanation for the differing behaviors. The correction term $W_{ij}^{\mathrm{topology}}$ is proportional to the mismatch between the weighted in-degrees of nodes i and j, with an additional factor of W_{ij}. Fig. 1(b) shows that the weighted in-degrees for Γ vary more widely than the weighted in-degrees of Γ^{T} (which equal the weighted out-degrees of Γ). Therefore the range in their mismatch

is larger, giving rise to the wider variation in the inferred weights for Γ than for Γ^{T}.

We now apply the correction (4.19b) to the numerical simulations described in Sec. 2. In Fig. 4 we compare the second and third order approximations for the computation with the networks Γ and Γ^{T}. Subfigures (a) and (b) show the second and third order approximations to the weights for the network Γ. With respect to Γ, the variance of the error is reduced in the improved approximation, while the mean error, which was already close to zero, stays about the same. For the network Γ^{T}, as shown in Subfigures (c) and (d), the mean error is reduced while the error variance is improved only slightly. After the correction is applied, the inferred weights for both the networks Γ and Γ^{T} are computed to about the same accuracy.

Clearly, the correction term accounts for the marked difference between the inferred weights seen in the initial numerical experiment. When the correction term is included in the computation, then transfer entropy, already demonstrated to be a useful tool for network reconstruction, only improves. Past studies have noted that $\mathrm{TE}^{j\to i}$ is in itself insufficient for reconstructing W_{ij} since it fails to condition on the dynamics of the remaining nodes in the network [27]. The procedure outlined in Sec. 4.3 provides a method to nonlinearly combine the transfer entropies of all node-pairs in order to effectively approximate such a conditioning.

6. Additional numerical example

The analysis performed in this paper applies only to the RBN model (1.1). It is tempting, however, to draw from it a more general conclusion about transfer entropy. Namely, consider models in which the coupling between the nodes of the network is weak. To leading-order, transfer entropy between two nodes should depend only on the strength of the coupling between them, but this is subject to a correction like that given in (4.20), dependent on the difference in total weighted in-degree of the two nodes forming the pair.

In this section, we explore this hypothesis with a numerical example. We consider a coupled system of chaotic tent maps subject to additive noise. The evolution of the ith state is given by the following:

$$x^i(t+1) = F\left(x^i(t)\right) + \Theta \sum_{j=1}^{N} W_{ij}\left(F\left(x^j(t)\right) - F\left(x^i(t)\right)\right) + \sigma n^i(t). \tag{6.1}$$

Here, $n^i(t)$ is additive noise drawn from a normal distribution, with small multiplier σ and the nonlinear function F, known as the tent map, is

$$F(x) = \begin{cases} 2x & 0 \le x < \frac{1}{2}, \\ 2 - 2x & \frac{1}{2} \le x \le 1, \\ 0 & x < 0 \text{ or } x > 1. \end{cases} \tag{6.2}$$

In the absence of noise—i.e., when $\sigma = 0$—solutions to (6.2) may synchronize. A condition for the existence of stable synchronous solutions for this system is given by [**?**] based on the earlier work of [17]. The condition is derived as follows. Define an $N \times N$ matrix L by

$$L_{ij} = \begin{cases} W_{ij}, & i \neq j, \\ -\sum_{k\neq i} W_{ik}, & i = j. \end{cases} \tag{6.3}$$

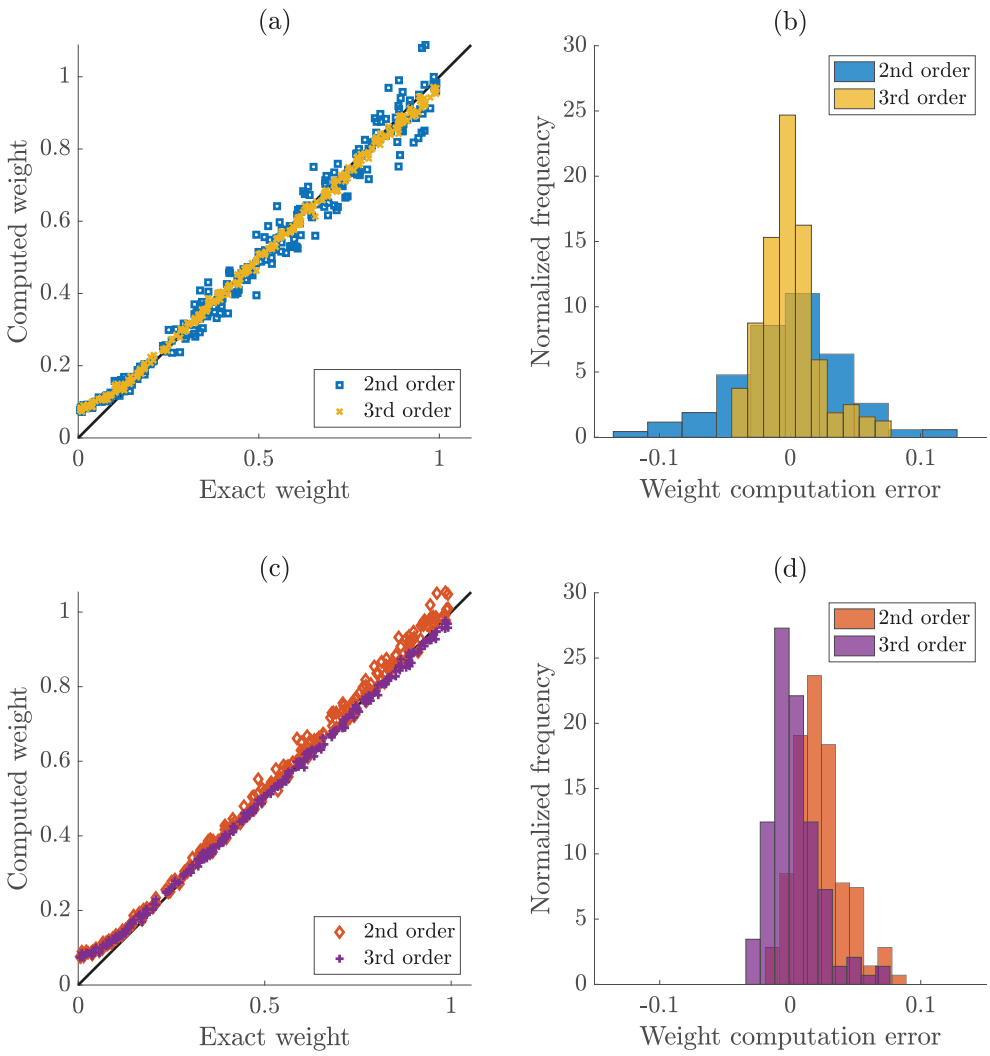

Figure 4. **(a)** Computed values of the nonzero matrix entries vs. the exact values W using second and third order approximations. **(b)** Histogram of the error in the two computed approximations for the weight matrix W. Blue points and histogram repeated from Fig. 2. **(c)** and **(d)**. As in (a) and (b) but for the matrix W^{T} with red points and histogram repeated from Fig. 2. For Γ the inference error, including the correction has mean 2.9×10^{-4} and variance 4.8×10^{-4}. For W^{T} the inference error has mean 3.1×10^{-4} and variance 3.5×10^{-4}. These compare favorably with the values reported in the caption for Fig. 2.

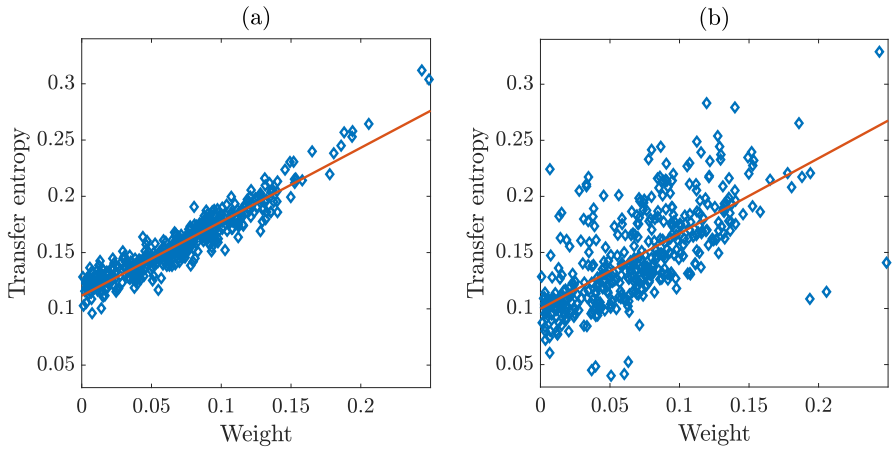

Figure 5. The pairwise transfer entropy computed as a function of the nonzero weights for **(a)** the network Γ and **(b)** the network Γ^{T}, together with best-fit linear functions. The root-mean-square deviation from the best-fit line is .0096 for Γ, and .033 for Γ^{T}.

Then stable solutions exist if, for all nonzero eigenvalues λ of W,

$$|1 + \Theta\lambda| < \frac{1}{2}. \tag{6.4}$$

We were unable to find any examples of networks of the type used in Section 2 for which synchronization was possible for both the weighted network Γ and its transpose Γ^{T}. Instead we construct the following simple family of weighted networks that may lead to synchronization. Let Γ be a directed Erdős-Rényi network. Todo: cite? In this type of network the nodes are connected randomly such that a directed link exists from j to $i \neq j$, with a fixed probability $p < 1$. Then for each node i, choose the weights W_{ij} such that $d_i = \sum_j W_{ij} = 1$.

Thus the term $W_{ij}^{\mathrm{topology}}$ is identically zero for Γ but nonzero for the network Γ^{T}. We expect from (4.21) that the computed transfer entropy for (6.2) using Γ^{T} will be have larger variance than the computed transfer entropy for Γ. Fig. 5 shows the results of this computation. In this computation $N = 30$ and $p = \frac{1}{2}$.

Therefore, we conclude that the wider variation in in-degree for Γ than for Γ^{T} is responsible for the difference in the behaviors. This supports the claims analytically derived for the RBN apply to a wider class of coupled dynamical systems. The main message of this work is that the success of pairwise inference through transfer entropy depends on the topological features of the network to be discovered. Networks with homogeneous in-degree can be more accurately inferred than those with heterogeneous in-degree. The out-degree plays a surprisingly marginal role on the quality of the inference.

7. Conclusion

Transfer entropy is emerging as a promising approach to carry out network reconstruction from the time-series of the network nodes. In this vein, one utilizes the value of transfer entropy between a pair of network nodes to decide whether a link exists between them and, potentially, predict its weight. This process is based on the premise that transfer entropy maps one-to-one with the weight of the link

between two nodes, which is in general difficult to guarantee. In fact, the definition of transfer entropy between two nodes deliberately excludes the effect of any other node in the network, thereby neglecting indirect influence between the two nodes that could be supported by the entire network.

In this work, we seek to offer a precise quantification of the topological features that challenge transfer-entropy network reconstruction in the context of a simple Boolean network model, inspired by policy diffusion. Through perturbation analysis of the high-dimensional Markov chain associated with the model evolution on the entire network, we establish a closed-form expression for the stationary distribution. This is, in turn, utilized to derive our main result, consisting of a third-order accurate form of transfer entropy between any two nodes in the network.

Our closed-form result indicates that for sufficiently slow dynamics, transfer entropy between two nodes maps one-to-one the corresponding weight of the link between them. As the perturbation parameter increases to encompass dynamics that could evolve at a faster scale, the one-to-one map breaks down, in favor of a more complex relationship between topology and transfer entropy. Specifically, we discover that inferring the weight between two nodes calls for examining transfer entropy from any other node in the network to the node pair under scrutiny. The higher the mismatch between the two nodes in terms of their total incoming transfer entropy, the less accurate is the one-to-one mapping between transfer entropy and influence.

This finding is particularly important when studying heterogeneous networks, such as scale-free models where wide variations in the incoming information flow should be expected as we seek to discover links around potential hubs. To detail this aspect, we have examined two instances of a weighted, directed scale-free network: one in which heterogeneity manifests in terms of out-degree and, the other, constructed through simple matrix transposition that displays heterogeneity in terms of the in-degree distribution. In the former case, transfer entropy between each pair of nodes offers an unbiased measure of the mutual influence between the nodes, such that one might estimate the corresponding weight from transfer entropy reading with an uncertainty of zero mean. In the latter case, we observe a systematic bias of transfer entropy-based inference, such that the uncertainty in the estimation has a non-zero mean, although it is characterized by a narrower variance. Accounting for the high-order correction derived in this paper, we successfully address network inference in both cases, attaining an unbiased, tight reconstruction of every weight in the network from transfer entropy measures.

Although numerical results on coupled chaotic tent maps seem to align with our analytical predictions, the generality of our results remains an open question for future research. Specifically, our future work should seek to establish a general framework for error quantification in transfer entropy-based network inference, beyond the case of Boolean network model examined herein. While it is unlikely that one could determine a closed-form expression of transfer entropy in terms of the network topology as proposed in this paper, it may be tenable to establish conservative bounds on information transfer associated with indirect influence between nodes. Another area of future inquiry encompasses the extension of the analytical framework to encompass time-varying dynamics, leading to a non-stationary Markov process, and self-loops in the network, which would likely exacerbate the role of heterogeneity on transfer entropy.

Acknowledgments

This work was undertaken while RG was on sabbatical at the NYU Tandon School of Engineering. He thanks them for their hospitality. MP acknowledges the support of the National Science Foundation under Grant nos. CMMI 1433670, CMMI 1561134, and CBET 1547864.

References

1. R. Albert and A.-L. Barabási, Statistical mechanics of complex networks, *Rev. Mod. Phys.*, **74** (2002), 47–97.
2. R. P. Anderson, G. Jimenez, J. Y. Bae, D. Silver, J. Macinko and M. Porfiri, Understanding policy diffusion in the us: an information-theoretical approach to unveil connectivity structures in slowly evolving complex systems, *SIAM J. Appl. Dyn. Syst*, **15** (2016), 1384–1409.
3. D. S. Bernstein, *Scalar, vector, and matrix mathematics: Theory, facts, and formulas–revised and expanded edition*, Princeton University Press, 2018.
4. S. Boccaletti, V. Latora, Y. Moreno, M. Chavez and D.-U. Hwang, Complex networks: Structure and dynamics, *Phys. Rep.*, **424** (2006), 175–308.
5. E. M. Bollt, J. Sun and J. Runge, Introduction to focus issue: Causation inference and information flow in dynamical systems: Theory and applications, *Chaos*, **28** (2018), 075201.
6. T. Bossomaier, L. Barnett, M. Harré and J. Lizier, *An introduction to transfer entropy: Information flow in complex systems*, Springer International Publishing, 2016.
7. P. Brémaud, *Markov chains: Gibbs fields, Monte Carlo simulation, and queues*, vol. 31, Springer Science & Business Media, 2013.
8. E. Bullmore and O. Sporns, Complex brain networks: graph theoretical analysis of structural and functional systems, *Nat. Rev. Neurosci.*, **10** (2009), 186.
9. T. Cover and J. Thomas, *Elements of information theory*, Wiley, 2012.
10. J. A. Dunne, R. J. Williams and N. D. Martinez, Food-web structure and network theory: the role of connectance and size, *Proc. Natl. Acad. Sci. U.S.A.*, **99** (2002), 12917–12922.
11. C. Godsil and G. F. Royle, *Algebraic graph theory*, vol. 207, Springer Science & Business Media, 2013.
12. C. Grabow, J. Macinko, D. Silver and M. Porfiri, Detecting causality in policy diffusion processes, *Chaos*, **26** (2016), 083113.
13. M. J. Keeling and K. T. Eames, Networks and epidemic models, *J. R. Soc. Interface*, **2** (2005), 295–307.
14. A. Khinchin, *Mathematical foundations of information theory*, Dover Books on Mathematics, Dover Publications, 1957.
15. A. H. Nayfeh, *Introduction to perturbation techniques*, John Wiley & Sons, 2011.
16. L. Paninski, Estimation of entropy and mutual information, *Neural Comput.*, **15** (2003), 1191–1253.

17. L. M. Pecora and T. L. Carroll, Synchronization in chaotic systems, *Physical review letters*, **64** (1990), 821.

18. M. Porfiri and M. Ruiz Marín, Inference of time-varying networks through transfer entropy, the case of a Boolean network model, *Chaos*, **28** (2018), 103123.

19. M. Porfiri and M. Ruiz Marín, Information flow in a model of policy diffusion: An analytical study, *IEEE Trans. Network Sci. Eng.*, **5** (2018), 42–54.

20. B. J. Prettejohn, M. J. Berryman and M. D. McDonnell, Methods for generating complex networks with selected structural properties for simulations: A review and tutorial for neuroscientists, *Front. Comput. Neurosci.*, **5**.

21. A. M. Ramos, A. Builes-Jaramillo, G. Poveda, B. Goswami, E. E. Macau, J. Kurths and N. Marwan, Recurrence measure of conditional dependence and applications, *Phys. Rev. E*, **95** (2017), 052206.

22. A. Roxin, V. Hakim and N. Brunel, The statistics of repeating patterns of cortical activity can be reproduced by a model network of stochastic binary neurons, *J. Neurosci.*, **28** (2008), 10734–10745.

23. J. Runge, Causal network reconstruction from time series: From theoretical assumptions to practical estimation, *Chaos*, **28** (2018), 075310.

24. T. Schreiber, Measuring information transfer, *Phys. Rev. Lett.*, **85** (2000), 461–464.

25. T. Squartini, G. Caldarelli, G. Cimini, A. Gabrielli and D. Garlaschelli, Reconstruction methods for networks: The case of economic and financial systems, *Phys. Rep.*, **757** (2018), 1–47.

26. M. Staniek and K. Lehnertz, Symbolic transfer entropy, *Phys. Rev. Lett.*, **100** (2008), 158101.

27. J. Sun, D. Taylor and E. M. Bollt, Causal network inference by optimal causation entropy, *SIAM J. Appl. Dyn. Syst*, **14** (2015), 73–106.

28. M. Wibral, N. Pampu, V. Priesemann, F. Siebenhühner, H. Seiwert, M. Lindner, J. T. Lizier and R. Vicente, Measuring information-transfer delays, *PloS One*, **8** (2013), e55809.

29. M. Wibral, R. Vicente and J. T. Lizier, *Directed information measures in neuroscience*, Springer, 2014.

30. G. U. Yule, II.—A mathematical theory of evolution, based on the conclusions of Dr. JC Willis, FRS, *Philos. Trans. R. Soc. London, Ser. B*, **213** (1925), 21–87.

AIMS Press

© 2019 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0)