THE UNIVERSITY OF CHICAGO

# Using Transfer Entropy to Understand Information Flow and Connectivity in Neural Networks

Bowen Zheng
Advised by Brent Doiron and Alexander Strang
Dec 2022

A paper submitted in partial fulfillment of the requirements for the Master of Arts degree in the Master of Arts in Computational Social Science

Faculty Advisor: Brent Doiron
Preceptor: Pedro Alberto Arroyo

**Abstract**

One central goal of psychology is to link human behaviors and cognitive processes with concrete neural network. In social science, there is growing interest in understanding social structures through the lens of social network. Though definition of networks and their dynamics are context dependent, the link between network connectivity and dynamics plays a key role in all fields. In this thesis, we study how transfer entropy can elucidate neural network structures. We use transfer entropy to examine two types of model, a network with firing-rate-based nodes with threshold linear transfer function, and a balanced network consisting of leaky integrate-and-fire type neurons. With a firing rate neural model, the effective connectivity inferred by transfer entropy matches the underlying directional network. For balanced networks, presence of connection between two neurons is not predictive of existence of information flow. However, for a balanced network with cluster or layer structures, the multivariate transfer entropy calculation using samples of each neural cluster accurately captures the direction of the cluster-wise projection. We focused on networks in neuroscience, but the framework can be generalized to understand how social network propagates information.

# 1 Introduction

Psychologists or behavioral neuroscientists aim to find the governing equations explaining human behaviors or brain activities. The neuroscience perspective of cognition and behaviors is essential for understanding the mind. Neural models provide mechanistic models with causal explanations and high predictive powers. More importantly, they give precise predictions of outcomes when the system is perturbed. That is, changes in the neural system correspond to behavioral changes, and therefore a neural mechanic approach to behavioral science is critical for understanding behavioral pathologies. The most famous example, perhaps, is the Hopfield network of memory (Hopfield, 1982). In a Hopfield network, the connectivity weights between neurons store memory. It exemplifies how high-level cognitive processes can be concretely linked to the physical structures of the neurons.

In social science, more researchers are capitalizing on the availability of digital data to construct social networks (Edelmann et al., 2020). These social networks help study concepts like power and influence (Bearden and Mintz, 1987) and social capitals (Borgatti et al., 1998), as well as investing community structures (Brown, 2002). Most social science studies use the network as a static observation (Scott, 2011), but there is a growing interest in incorporating the dynamical processes of the network. For example, Bianchi and Squazzoni (2015) reviewed using agent-based modeling in sociology. They concluded that the dynamic perspective of the network enriches the understanding of the intricate interplay between individual behaviors and social structures. They also called for the continuation of cross-disciplinary integration in the field of network science.

On a high level, all network approaches to model behaviors, cognition, and social processes involve similar elements. A modeler can choose many intertwining components to assemble the network, including types of nodes, interaction rules between nodes, and, most importantly, connectivity. Various fields of research also face similar challenges. Most of the time, the ground truth for connectivity is not observable, and one needs to infer the connectivity from the dynamics of the network.

On the other hand, given the network connectivity, one would also want to answer what quantitative or qualitative information we can predict about the network's dynamics. The node characteristics and interaction rules are field-specific, so I focus on studying the connectivity. In particular, I focus on my study in neuroscience and use it as an example to explore the link between network connectivity and information flow.

Network connectivity can be defined as physical connections like the formation of synapses. Alternatively, the dynamics of the neural activities define functional networks (correlational measures) and effective networks (predictive measures) (Friston, 2011). The interplay between physical network structures and dynamics has attracted many interests. One fruitful line of works assumes the connectivity of the network and shows how the connectivity shapes the statistics of the functional networks(Ocker et al., 2017). By establishing analytical results connecting statistics of neural dynamics with connectivity, this program provides a strong foundation for control theory and network optimization.

On the other hand, in experimental studies one rarely observes the true underlying wiring patterns of the neurons (Ito et al., 2011). Thus, inferring the physical connection based on neural dynamics is equally critical. There are specialized methods and algorithms designed to reconstruct physical connectivity as faithfully as possible. For instance, maximum likelihood estimation based on a generalized linear model assuming concavity has been shown to successfully recover the structure of a balanced network (Zaytsev et al., 2015). However, such methods do not define an effective network that could provide insight into the information transfer within the network. Other studies defined various notions of effective connectivity and observe the correspondence between effective and physical connectivity (Goldenberg and Galván, 2015). Effective connectivity is usually defined with reference to temporal causality following Granger's definition (Liao et al., 2010). Among these methods, transfer entropy has gained some traction in neuroscience because the nonparametric nature of transfer entropy

fits well with complicated interactions between neurons (Vicente et al., 2011; Wibral et al., 2014).

Transfer entropy was first introduced in physics as an information-theoretical measure to quantify directional information flow from one time process to another (Schreiber, 2000). It has since been applied to neuroscience and various fields of social science that involve studying networks. Its usage can be roughly grouped into two kinds. One stream of research uses it as a quantity to explore individual or group difference by highlighting distinct or abnormal patterns of information flows. For instance, in clinical psychology, researchers measure EEG activities of children with ADHD and find that the information flow patterns among children with ADHD show significant differences with the healthy controls (Ekhlasi et al., 2021). Another stream of research, especially in political science and digital social network analysis, use transfer entropy to infer the underlying causal structures of the network. They treat transfer entropy as a measure of influencing power that causally drive downstream agents. For example, Ver Steeg and Galstyan (2012) used transfer entropy to identify and investigate processes governing peer influences on social media data within a network with Twitter data. Using sentiment analysis and transfer entropy calculation, Surano et al. (2022) demonstrated that emotional reactions to Covid-19 lockdown policies in different states are influenced by those of other states and that the direction of influence corresponds to spatial and socioeconomic factors.

Given two time series $X$ and $Y$, the transfer entropy from Y to X is defined as:

$$T_{Y\to X} = \sum p(x_{n+1}, x_n^k, y_n^l) \log\left(\frac{p(x_{n+1} \mid x_n^k, y_n^l)}{p(x_{n+1} \mid x_n^k)}\right) \tag{1}$$

where $x_{n+1}$ is the $X$ at the $n+1$-th time step, and $x_n^k$ is the shorthand notation for $x_n, x_{n-1}, \ldots, x_{n-k+1}$, which is the history of $x_{n+1}$ going $k$-step back. Note that the time series can be multivariate — transfer entropy can measure information transmission from a group of processes to another group. For systems with continuous variables, processes are first approximated by coarse-grained states with resolution $r$. Schreiber showed that for stochastic processes, as $r \to 0$, $T_{Y\to X}$ converges and is independent of the partition. The transfer entropy from process $Y$ to $X$ can also be

expressed as the difference in conditional entropy

$$T_{Y\to X} = H(x_{n+1} \mid x_n^k) - H(x_{n+1} \mid x_n^k, y_n^l)$$

or as the difference in mutual information

$$\begin{aligned} T_t^{n,m}(X,Y) &= I(X_{t+1}; Y_t^m)) \mid X_t^n) \\ &= I(X_{t+1}; Y_t^m, X_t^n) - I(X_{t+1}; X_t^n) \end{aligned} \tag{2}$$

Intuitively, equation 1 establishes that transfer entropy is the reduction in uncertainty of $x_{n+1}$ knowing the history of another process $Y$, conditioning on $X$'s history. Alternatively, it can be thought of as asking how much information $Y$ adds when we know the history of $X$. Note that this quantity is not symmetric by definition. It measures the rate of information flow from one process to another. It can be thought of as a non-parametric extension of the canonical Granger causality assuming a Gaussian linear model (Lindner et al., 2019). Indeed, when the linear and Gaussian assumptions hold, then transfer entropy reduces to Granger causality (Barnett et al., 2009).

Our work differs from previous literature in the following ways. First, previous work emphasized identifying the existence of transfer entropy and used the existence to infer underlying connections (Aru et al., 2015; Avena-Koenigsberger et al., 2018; Vicente et al., 2011). We studied the direct relationship between the topological properties of the network and the resulting transfer entropy and called attention to its relative value. To the extent of our knowledge, this work is also the first to explore information transmission via transfer entropy on balanced networks. Lastly, previous studies have binned spike trains and used continuous estimators for transfer entropy and focused on pairwise neuron activity. We tested binary methods for spike train data directly since binary estimators are unbiased and computationally cheap. We also extended pairwise calculation to multivariate cases.

The paper is organized as follows: First, we estimate the transfer entropy on simple analytical cases. We compare a nonparametric estimator and an estimator with linear and Gaussian assumptions for their convergence rate of the bias and

sample variance. Next, we use transfer entropy to explore information flow in a firing rate-based model (simplified neural mass model). We tested the agreement between effective connectivity constructed with transfer entropy and the true connectivity that governed the system dynamics. We also observe how connection weights, how common and private noise sources affect information transfer, and contrasted transfer entropy with correlation to emphasize their different responses to parameter changes. Lastly, we tested information flow on a balanced network with spiking neurons with leaky-integrate and fire (LIF) dynamics. We implemented different methods for transfer entropy calculation including pairwise bivariate calculation, multivariate calculation for the spike train data, and continuous estimator for the group-level mean firing rate data.

## 2 General Testing of Transfer Entropy Estimators

For the computation of transfer entropy, we implement the transfer entropy estimation algorithm building on an open-source software (Lizier, 2014). To understand the behavior of these statistics and to validate the accuracy of our modified algorithms, we tested various estimators of transfer entropy on simple, analyzable processes. In each case, we calculated the theoretical transfer entropy value and observed the convergence of standard error and bias. For continuous models, there are two estimators for transfer entropy, the Gaussian method, and the KSG estimator developed by Kraskov et al., 2004. The Gaussian method assumes that the variables follow a Gaussian distribution and the interaction between processes is linear. 3 shows that the transfer entropy of linear Gaussian processes reduces to the calculation of the determinant of covariance. If $X, Y, Z$ are Gaussian variables, then the conditional mutual information is given by:

$$\begin{aligned} I(X;Y \mid Z) &= I(X;Y,Z) - I(X;Z) \\ &= \frac{1}{2}\log\left(\frac{\det(\Sigma_X)\det(\Sigma_Y)}{\det(\Sigma_{X,Y,Z})}\right) - \frac{1}{2}\log\left(\frac{\det(\Sigma_X)\det(\Sigma_Z)}{\det(\Sigma_{X,Z})}\right) \\ &= \frac{1}{2}\log\left(\frac{\det(\Sigma_{X,Z})\det(\Sigma_{Y,Z})}{\det(\Sigma_Z)\det(\Sigma_{X,Y,Z})}\right) \end{aligned} \tag{3}$$

, where $\Sigma_X$ denotes the covariance matrix for the Gaussian variable $X$ (in the case where $X$ is a scalar variable, $\Sigma_X$ is the variance).

Different nonparametric methods of estimating transfer entropy vary in how they estimate probability distributions. Given two time series $X$ and $Y$, the kernel density estimation first fixes a kernel function, like box kernel or Gaussian kernel. Then it sums over all kernel functions centered at each sampled data point with appropriate scaling. Note that it uses the same width for the kernel function for the estimation of joint and marginal distribution. The KSG estimator improves traditional kernel density estimation by dynamically adapting the width of the kernel based on sample density. The algorithm first fix a parameter $k$ for choosing nearest neighbors. In estimating the joint probability, for each sample data, the algorithm choose it $k$-nearest neighbors and use the distance to the farthest neighbor as the width for the kernel. The marginal probability calculation inherits the chose of kernel width from the joint space. It has been shown empirically to be more efficient and less sensitive to chosen parameters than the kernel density estimation (Wibral et al., 2014).

The first case we consider is a linear Gaussian model. The process $X$ is simply an independent normal random variable for each $n$; $Y$ is given by $y_n = cx_{n-1} + \sqrt{1-c^2}\mathcal{N}(0,1)$. The one-step time-lag correlation between two processes is $c = 0.5$. Since the process is Gaussian linear, we expect the transfer entropy to reduce to $T_{X\to Y} = -\frac{1}{2}\log(1-c^2)$. This calculation follows from the definition of the transfer entropy. We see that both estimators converge to the theoretical value, as illustrated in Fig 1A.

Next, we test the nonlinear interaction between the two processes. $x_n$ is uniformly distributed on $[-1,1]$ and independent at each time step. $y_n = x_{n-1}^2 + k$, where $k$ also follows a uniform distribution with support $[-1,1]$. We calculated the density distribution and conditional probability analytically and use them to compute the value of the theoretical transfer entropy. Not surprisingly, Granger causality completely fails to capture any interaction (see Fig 1B). Even though this model

is somewhat contrived, it stresses the importance of nonparametric statistics for capturing nonlinear interactions with functions that have some symmetric structures.

The third case is a discrete case. $x_n$ is a Bernoulli variable with probability $p$ of returning 1 and $1-p$ of getting $-1$. $y_n$ takes the value of 0 and 1. The information transfer from $X$ to $Y$ takes the form $y_n = \Theta(x_{n-1} + \mathcal{N}(0,1))$, where $\Theta$ is the step function. The implementation for the discrete case simply uses the empirical distribution for the probability distribution and calculates the TE based on the estimated distribution. Fig 1C shows the case where $p = 0.5$. The binary estimator is unbiased and converges relatively quickly compared to continuous estimators.

The last test case is an autoregressive model of the $10^{th}$ order. The following maps represent the processes:

$$\begin{aligned} x_n &= \sum_{i=1}^{10} a_i x_{n-i} + \mathcal{N}(0,1) \\ y_n &= \sum_{i=1}^{10} a_i y_{n-i} + x_{n-10} + \mathcal{N}(0,1) \end{aligned} \tag{4}$$

, with $a_i$ constant. We derived the exact transfer entropy value theoretically by explicitly analyzing the dynamics of an equivalent linear model. Specifically, we denote $Z_{i\times j}$ as the all-zero matrix and $J_{i\times j}$ the all-one matrix with dimension $i, j$ and $a$ as the vector for constant coefficients.

Let:

$$A = \begin{pmatrix} Z(10,1) & J(10,10) \\ 0 & a' \end{pmatrix}, \quad K = \begin{pmatrix} & Z(10,11) & \\ 0 & 1 & Z(1,9) \end{pmatrix},$$

and:

$$e = \begin{pmatrix} Z(10,1) & 1 & Z(10,1) & 1 \end{pmatrix}^T, \quad E = ee'.$$

The original autoregressive process is equivalent to the following process:

$$\begin{pmatrix} \overrightarrow{x_{n+1}} \\ \overrightarrow{y_{n+1}} \end{pmatrix} = \begin{pmatrix} A & Z(11,11) \\ K & A \end{pmatrix} \begin{pmatrix} \overrightarrow{x_n} \\ \overrightarrow{y_n} \end{pmatrix} + e \tag{5}$$

Then the equivalent process is Gaussian linear with white noise and all variables at each time step follow the Gaussian distribution. Write $G = \begin{pmatrix} A & Z(11,11) \\ K & A \end{pmatrix}$,

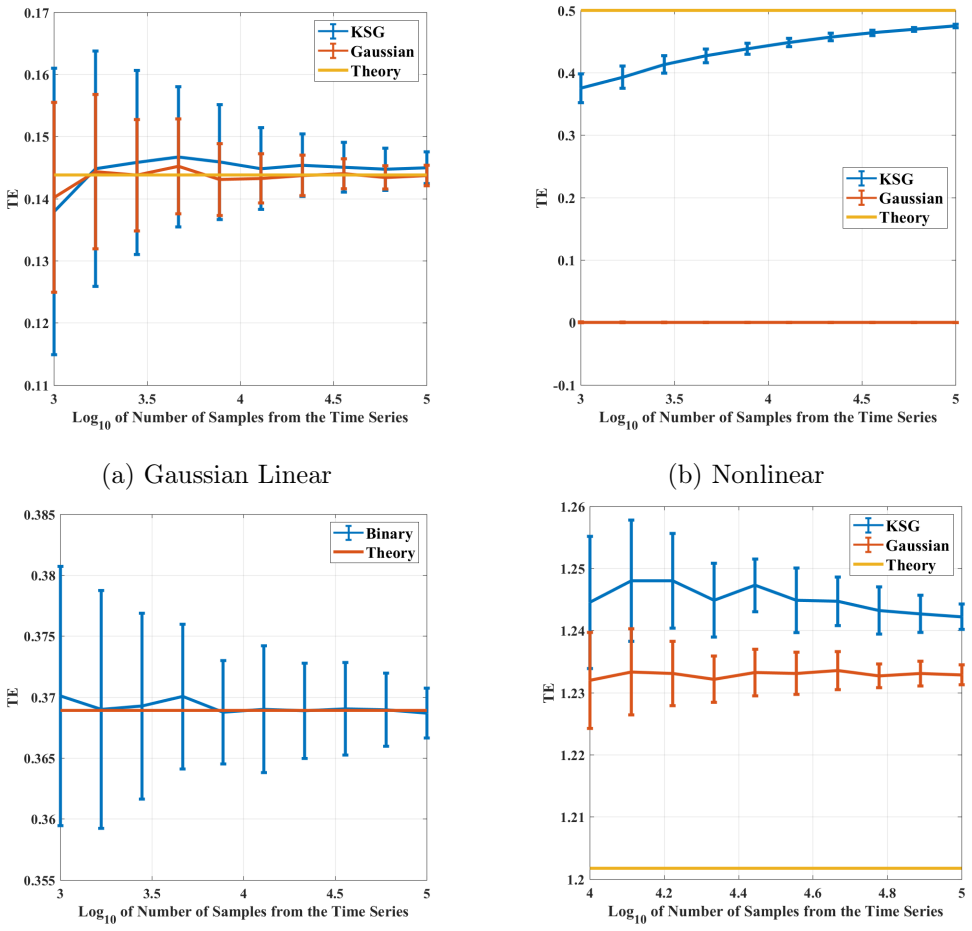
then the covariance of the whole system at time step $n$ is given by $\sum_{i=0}^{n-2} G^i ee'(G')^i$. The transfer entropy calculation reduces to computing the determinant of several sub-matrices from the complete covariance matrix of the whole system.

We use a truncated sum to estimate the covariance matrix. Note that the covariance matrix $\sum_{i=0}^{n-2} G^i ee'(G')^i$ can be computed iteratively. Let $E = ee'$, $H(X) = GXG' + E$, then $\sum_{i=0}^{n-2} G^i ee'(G')^i = H^{(n-1)}(0)$. We justify the truncation by showing that $H(X)$ is a contracting map. Take $X, Y \in \mathbb{R}^{n\times n}$, $|H(X) - H(Y)| = |G(X - Y)G'| \leq |G|^2|X - Y|$. By design, $|G| < 1$, so $|H(X) - H(Y)| < d(X, Y)$. Since the finite matrix space is complete, the contraction map theorem applies. Then, there exists a fixed point $K$, such that $H(K) = GKG' + E$, and the $\sum_{i=0}^{n-2} G^i ee'(G')^i$ converges exponentially to the fixed point of $H(X)$.

Table 1 summarizes the convergence rate of the bias and variance term of the estimator. For each model, we estimate the bias and standard deviation from 100 runs and fir to a power law $b * n^m$ with parameters $b$ and $m$. Cells in Table 1 show the convergent rate $m$ when applicable. Overall, the KSG method has larger variances and the bias term converges to zero slowly. On the other hand, it captures nonlinear interaction better than Granger causality, especially when the interacting function has symmetric structures.

Table 1: Convergence Rate of Various Models

| Model | Convergence Rate of Bias | | | Convergence Rate of SD | | |
|---|---|---|---|---|---|---|
| | KSG | Gaussian | Binary | KSG | Gaussian | Binary |
| Gaussian Linear | -0.15 | unbiased | | -0.49 | -0.51 | |
| Nonlinear | -0.35 | | | -0.48 | | |
| Binary | | | unbiased | | | -0.52 |
| Autoregressive | -0.06 | biased | | -0.55 | -0.52 | |




(a) Gaussian Linear

(b) Nonlinear

(c) Binary

(d) Autoregressive

Figure 1: Plots of Information Flow for Two Nodes

## 3 TE application to Firing Rate-based Models

We start to test information flow in a system with neural network properties. The firing rate model describes how aggregated neural cluster activities behave. The following model is taken from Parmelee et al., 2022. Let $\vec{r}$ be a vector representing the firing rate of each node, and $W$ be the matrix representing a directed graph of neural connectivity. $W_{ij}$ is the weight of the projection from node $j$ to node $i$. $b$ is the external input that drives the firing rate or can be thought of as a parameter controlling the threshold for the threshold linear transfer function. $\mu$ is the noise variable consisting of two sources, common noises, and private noises.

The dynamics of the firing rate-based model are given by:

$$\begin{aligned} \vec{r} &= -\vec{r}dt + [(\vec{b} + W\vec{r})dt + \vec{\mu}dt)]_+ \\ \tau d\vec{\mu} &= -\vec{\mu}dt + k \cdot d\vec{P} + c \cdot dG \end{aligned} \tag{6}$$

where $[\cdot]_+$ is a threshold linear function that preserves the positive input and rectifies negative input to 0, $\vec{P}$ is multivariate Brownian motion that gives private noise to each neuron, and $G$ is a Brownian motion that gives noise common to all neurons. We assume $\tau$ to be transient, which gives the following discretization.

$$\begin{aligned} \vec{r}(n+1) &= \vec{r}(n) + -\vec{r}(n) \triangle t + [(\vec{b} + W\vec{r}(n)) \triangle t + \vec{\mu}(n) \triangle t)]_+ \\ \vec{\mu}(n) \triangle t &= k \cdot \sqrt{\triangle t} \cdot N_P(0, \Sigma) + c \cdot \sqrt{\triangle t} \cdot N_G(0, 1) \end{aligned} \tag{7}$$

where $N_P$ is multivariate Gaussian and $N_G$ is the standard Gaussian. This discretization will be our working model throughout this section. Specific parameters of the model are documented in the online supplementary Matlab live script.

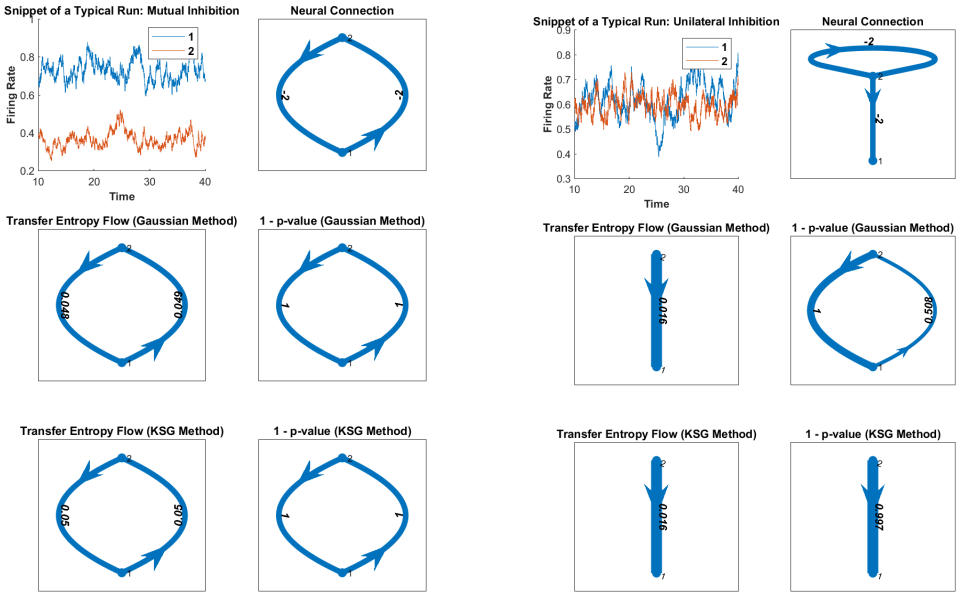
We started with a two-neuron network to confirm that transfer entropy captures the directional information flow and observe how private noises and common noises influence the information flow. Our next step was to test if pairwise transfer entropy estimation could recover the underlying network connectivity on a larger network. For each model, we observe the connection between the weighted adjacency matrix defining the network topology and the matrix generated by calculating the transfer entropy. Additionally, following Lizier, 2014, we implemented statistical testing of significance by calculating a p-value based on permutating the samples to construct a null distribution. For each system, we use 50 trials with 4000 effective time units. We use the Euler method with a time scale of 0.01 to integrate the dynamical equations. The activities of the first 10 time units are taken out for each trial. We sample at a 0.2 effective time interval from each time sequence. For the initial condition, we first run one trial of the system and get the average value of the node activities and pass them as the initial values for the system. We perturb the system with Gaussian noises with zero mean and variance from the initial run such that the system would potentially explore different stable manifolds. We compute a matrix of pairwise transfer entropy,

$TE$, where $TE_{ij}$ is the transfer entropy from node $j$ to $i$, and a $p$-value matrix with the same structure. Both matrices are averaged across the trials.

### 3.1 Two Nodes

For the first test, we consider a simple network consisting of two neurons. In the first example, we consider mutual inhibition, which is often known as the bistable case. Two neural groups compete to fire and one's activity suppresses that of another's. We should expect to see information flowing in both directions. In a second test case, we consider the case where there is only one direction of projection, and we expect to see information flowing only in that direction. The top panel in Fig 2 shows a time-snippet of a typical run of the system and the directed connection graph. The bottom two panels show the transfer entropy calculated with the Gaussian method and the KSG method and graph for simulated $p$-value from permutation data. The $p$-value here represents how unlikely we would observe a nonnegative transfer entropy assuming there is no information transfer along that directed edge. To match the intuition for the graph with other plots, we take $1 - p$.

We see that transfer entropy captures the directional interaction between the nodes. The information is transferred both ways in the mutual inhibition case, and the information flows in the direction of the projection in the unilateral inhibition case.




(a) Mutual Inhibition (b) Unilateral Inhibition

Figure 2: Plots of Information Flow for Two Nodes

## 3.2 Three and Four Nodes

Next, we consider a small network with more neurons. Figure 3 shows the cases with three nodes and Figure 4 for four nodes. We consider four different connections for three and four nodes. The layout of Figures 3 and 4 is the same as Figure 2. Each node also has a self-looping connection such that the aggregated input weights to each node are the same for all neurons. We added this structure because without balancing input to each neuron, the firing rate of some neurons with a more inhibitory drive from other neurons will get suppressed to near zero and fluctuate around zero with noises. Not surprisingly, if a neuron has a firing rate close to zero, it does not transfer much information even when it has dense projection to other neurons.

As the number of nodes increases, the interaction between nodes becomes more complex. We note that even when two nodes are not directly connected, they would potentially transfer information via some connected path. As a result, the effective connectivity based on information transfer between nodes might not accurately reflect the actual wirings. However, we observe that when there is a direct projection, the

transfer entropy along that projection is higher than when the connection is via an indirect path.

While the numerical estimates from the Gaussian and KSG estimators are close, the p-values computed using the Gaussian method (Granger causality) are more sensitive to nonnegative transfer entropy. As a consequence, Granger causality falsely produced projections that are positive in most graphs while the effective connectivity from transfer entropy more faithfully reconstructed the underlying projections. It is most evident in the case with four nodes. The KSG-based estimation aligns with the true connection matrix in each case.

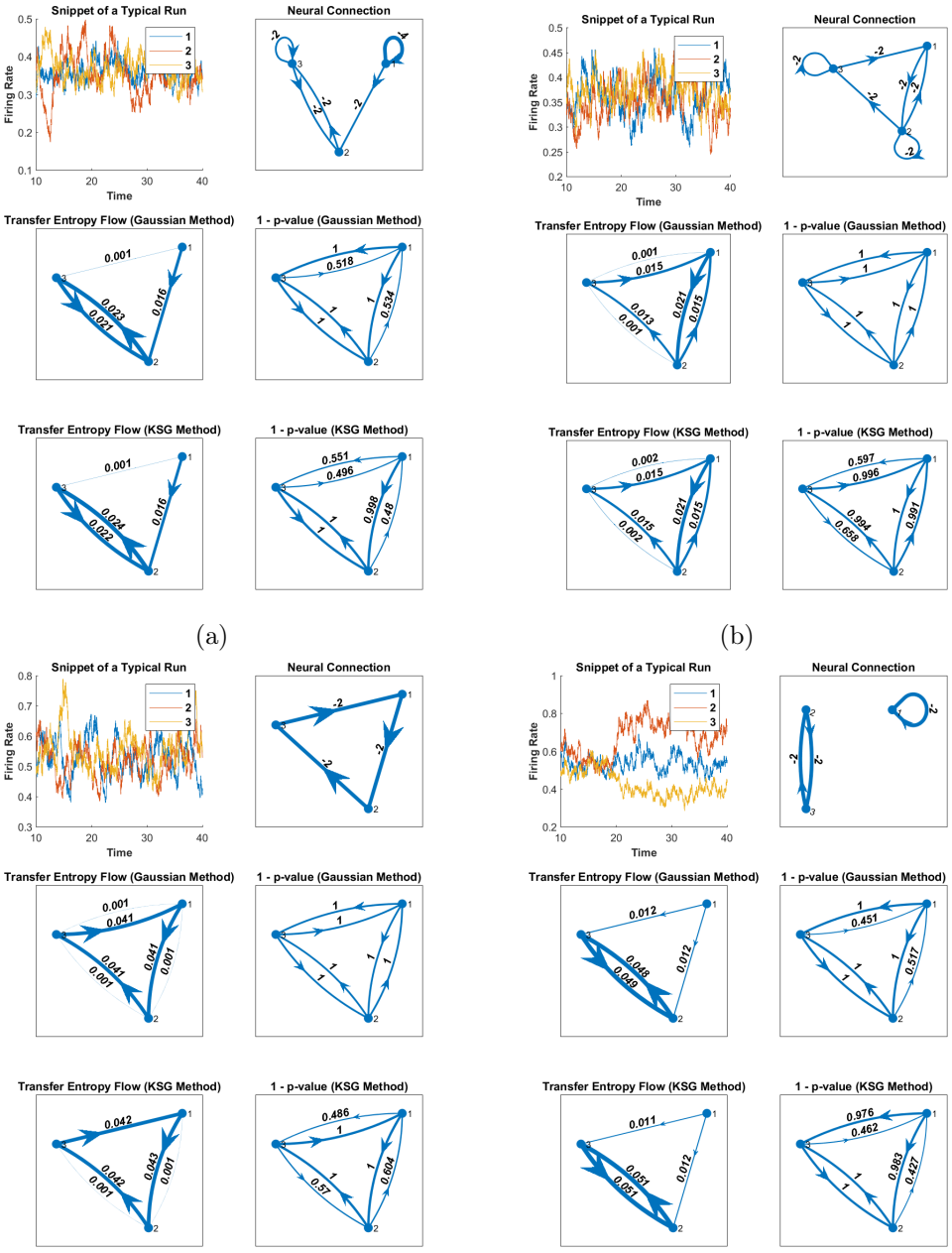



Figure 3: Plots of Information Flows for Three Nodes

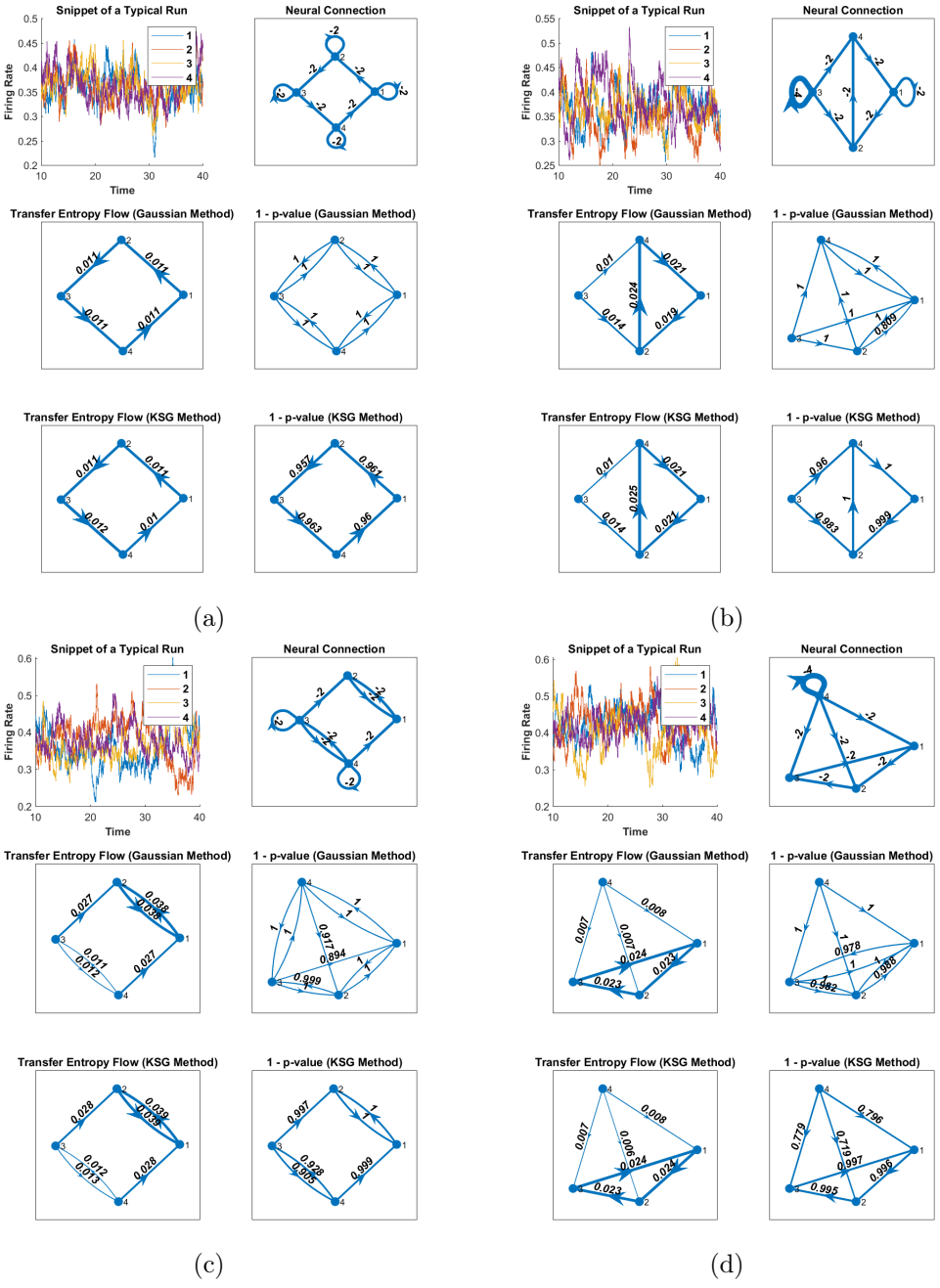



Figure 4: Plots of Information Flows for Four Nodes

We call special attention to the interesting case of (d) for the three-node model. This case illustrates how information transfer on a network would be influenced by common inputs in a subtle way. Even though there is no projection from nodes 1 to 2 and 3, we observe significant non-negative transfer entropy from nodes 1 to 2 and 3. We note that this is not directly caused by all three nodes having common drives.

We see that there is no information transfer from nodes 2 and 3 to one but only in the opposite direction. Why does knowing the history of node 1 informs the present state of node 2 or 3? One possible mechanism is that even though the history of node 1 does not help with inferring the current state of node 3 directly, it gives us some information about node 2 because of the common drive. Since knowing the history of node 2 helps with inferring the state of node 3, the history of node 1 indirectly gives information about node 3. To test this speculation, we use the following simple Gaussian linear model. Covariance matrices can be explicitly written following the same procedure as the autoregressive model in section 2.

$$
\begin{aligned}
x(n+1) &= \alpha x(n) + \beta y(n) + \epsilon_c c(n) + \epsilon_p p_x(n) \\
y(n+1) &= \alpha y(n) + \beta x(n) + \epsilon_c (n) + \epsilon_p p_y(n) \\
z(n+1) &= \alpha z(n) + \epsilon_c c(n) + \epsilon_p p_z(n)
\end{aligned}
$$

where $\alpha, \beta, \epsilon_c$ and $\epsilon_p$ are fixed coefficients, $c(n)$ is the common noise source following normal distribution, and $p_x, p_y$ and , $p_z$ are normal private noise.

We varied the intensity of the common drive $\epsilon_c$ and observed that the transfer entropy from node z to x and y increases as the intensity of the common drive scales up and reaches a maximum before the whole system is swamped by noise and transfer entropy goes down again (see Figure 5).

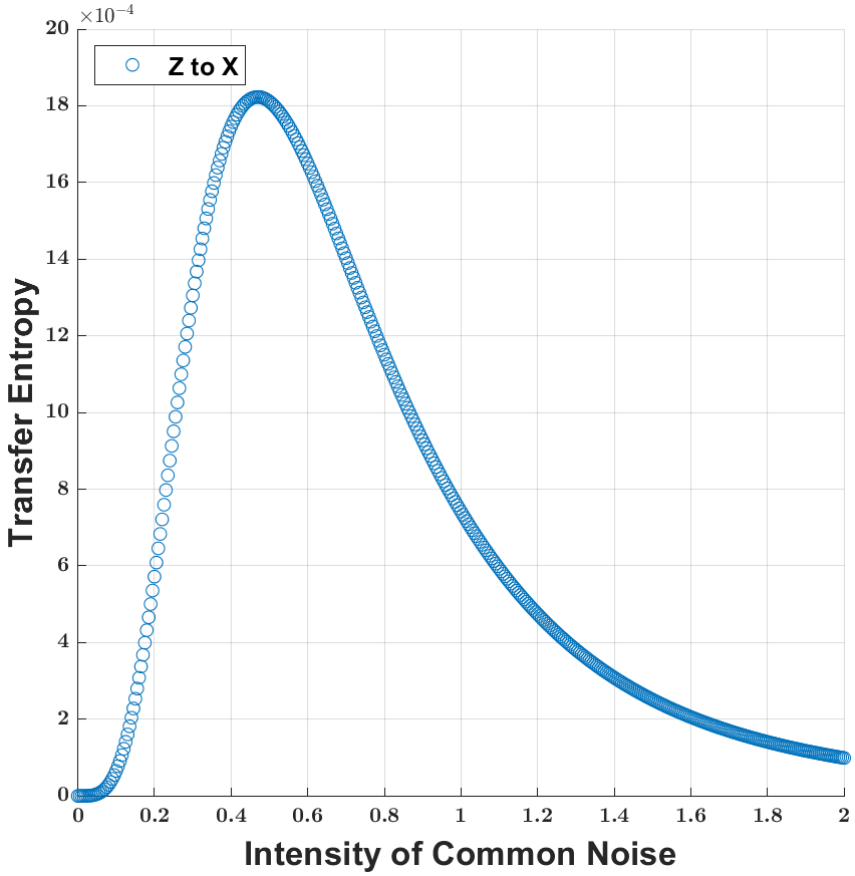



Figure 5: Positive Transfer Entropy in the Absent of Direct Connection

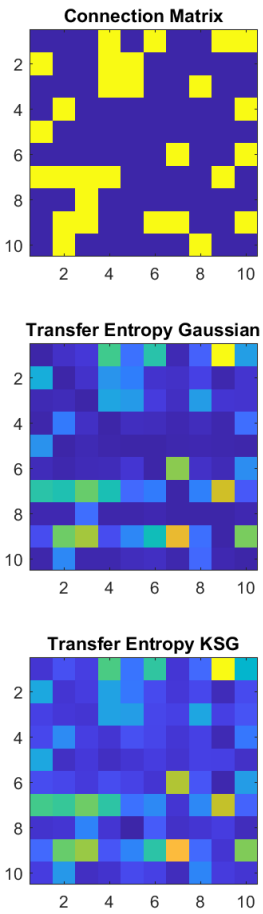
### 3.3 Ten Nodes

As a final exercise in the firing rate model, we push the limit and see if transfer entropy can recover the connectome of a 10-node network. It is computationally heavy to estimate the $p$-value as it involves calculating the transfer entropy for a large number of permuted samples. For the 10-node network, therefore, we calculate the $p$-value graph from one trial. The transfer entropy calculation is averaged across 100 trials to get a more accurate estimation such that we can observe some structures based on the value of the transfer entropy. To visualize the data in Fig. 6, we plot the matrix since the connection graph becomes complicated. There were both false positives and false negatives in the $p$-value graph with a cut-off of 0.1, regardless of the exact method for calculating transfer entropy. However, we still see that when a projection exists in the connection matrix, the transfer entropy is higher. We confirm

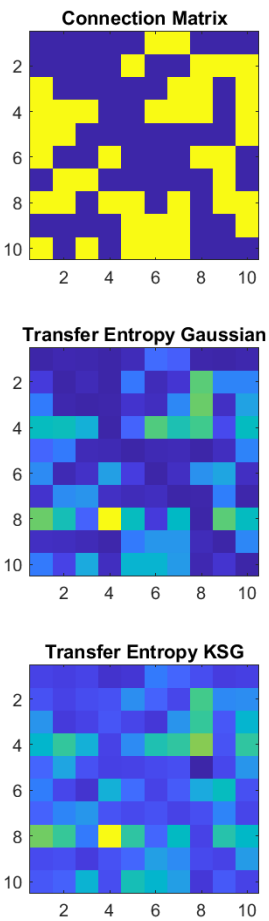
this observation by finding the best-fit parameters for the following minimization problem

$$\text{Minimize: } \|M_{TE} - (\beta_1 W_1 + \beta_2 W_2)\|_F \text{ where } \begin{cases} W_1 = [W - \text{diag}(W)]_+ \\ W_2 = [W^2 - \text{diag}(W^2)]_+ \end{cases}$$

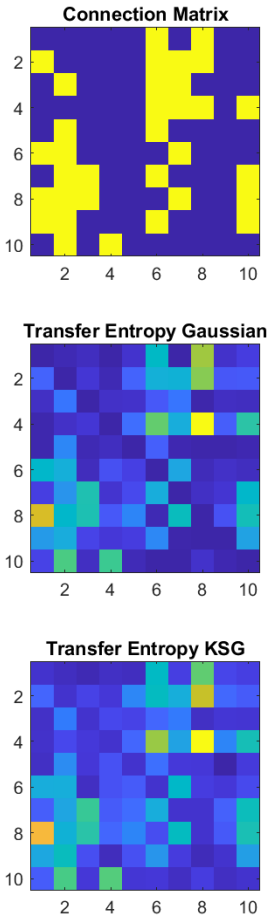
, where $M_{TE}$ is the scaled transfer matrix estimated from data and $\|\cdot\|_F$ is the Frobenius norm of the matrix, Diag(W) is the diagonal matrix with the diagonal term from $W$. $W_1$ denotes whether there is a direct projection between two nodes and $W_2$ denotes whether there is a directed path of length 2 between two nodes. We need to subtract away the diagonal term since the transfer entropy from a process to itself is not defined. The four cases give $\beta_1$ of $1.73, 1.59, 1.80, 0.76$, respectively and $\beta_2$ of $-0.49, -0.79, -0.72, 0$. In the last graph, due to the special structure, $W_1$ and $W_2$ are equivalent. The variance in transfer entropy can not be explained by the existence of projection alone but is positively correlated with direct projection. The path of length 2 disrupts information inflow. This is likely because all projections are inhibitory in the graph and that a path of length 2, therefore, becomes excitatory, masking effects from the length-1 projection.


Connection Matrix
Transfer Entropy Gaussian
Transfer Entropy KSG

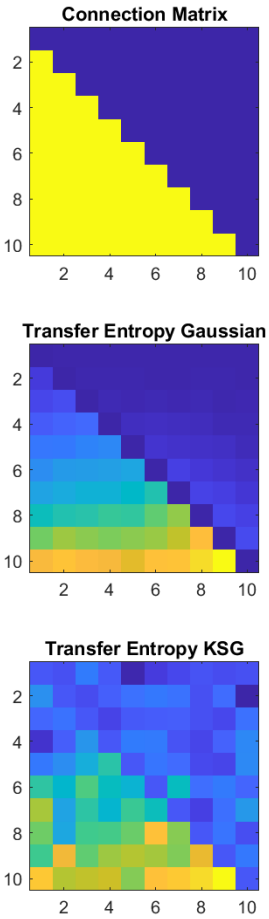
(a)


Connection Matrix
Transfer Entropy Gaussian
Transfer Entropy KSG

(b)


Connection Matrix
Transfer Entropy Gaussian
Transfer Entropy KSG

(c)


Connection Matrix
Transfer Entropy Gaussian
Transfer Entropy KSG

(d)

## 3.4 Effects of Common Noises, Private Noises and Connection Strength on Transfer Entropy

Using the two-nodes unilateral model, we explore how connection weights, common noise, and private noise influence information transfer. The projection is from $Y$ to $X$ and we expect information transfer in this direction but not the other way around. As a comparison, we also calculate the correlation between the two nodes and observe their changes with parameters. The transfer entropy from the projected direction increases as the connection weights between the two neurons increase. The correlation drops down because the influence of common noises goes down in proportion. When private noises feeding into the system increase, we observe that the correlation caused by the common drive gets gradually canceled out and eventually becomes negative, reflecting the inhibition. Meanwhile, the transfer entropy from the $Y$ to $X$ also decreases due to the system becoming noisier. TE and correlation have opposite behaviors when common noises increase. When correlation increases due to co-fluctuation, transfer entropy from $Y$ to $X$ decreases and stays at zero. The opposite behaviors of transfer entropy and correlation suggest that transfer entropy might help neuroscientists tap into network structures not available in correlational analysis.

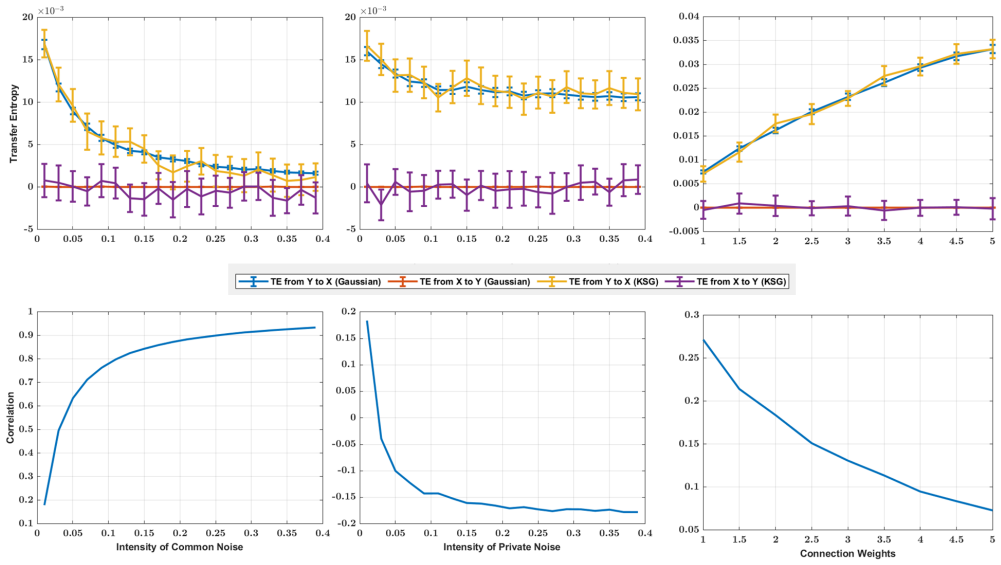
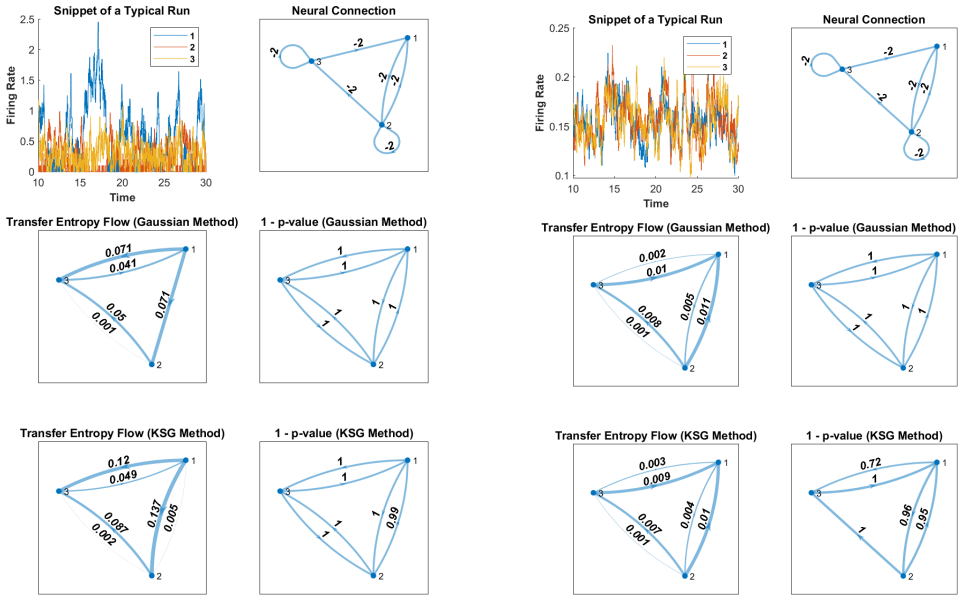



Figure 7: Plots of Influences of Connection Weights and Noises on Transfer Entropy

### 3.5 Nonlinear Transfer Function

In the previous model, the transfer function is nonlinear around 0. We test two other transfer functions with more nonlinear behaviors. The first one is the scaled *sgn* function with scaling to ensure network stability. The second one is a capped absolute value function $f(x) = max(a, |x|)$ with parameter $a$ small enough such that the system does not blow up. KSG and Gaussian methods give more dissimilar estimations compared to the linear transfer function case. Agreeing with Ursino et al., 2020, the structure of the effective connectivity becomes less congruent with the underlying connection. Overall, directed paths still have a much higher information flow than other paths. The difference in estimation is not surprising as the Gaussian estimator assumes all parameters follow a Gaussian distribution and the assumption is evidently false as interaction is not linear. In terms of the identification of projection, both estimators result in false positives.




(a) Sign Transfer Function

(b) Absolute Value Transfer Function

Figure 8: Plots of Influences of Connection Weights and Noises on Transfer Entropy

# 4 TE application to Balanced Network

## 4.1 Introduction to Models and Methods for TE Calculation

Next, we explore information transfer on a neural network with leaky integrate-and-fire neural units. These neural units are organized into a neural network with connections in different parameter regimes. Each neural in the network produces spike trains, a time sequence of binary data points with 1 reflecting a neural spike. The data structure is similar to processed neural recording data. Therefore, it is a suitable exercise to test if transfer entropy can capture information flow on the network and cast light upon wiring structures.

In terms of specific models to test, we use a well-studied simulated neural network, namely the balanced network (van Vreeswijk and Sompolinsky, 1998). The balanced network consists of clusters of excitatory and inhibitory neurons with strong connections between neurons. Despite the strong connections between neurons, all other neurons' net effective input to one neuron averages zero with a large variance.

Therefore, the network does not explode or turn silent, and each neuron behaves as if it receives Gaussian noise and displays a spike train pattern similar to a Poisson process. It is well known that the balanced network internally produces asynchronous variability and that the system has high-dimensional chaotic dynamics. Neurons behave as if they are independent regardless of the underlying wiring. We are interested to see if information flow depicts a similar picture of independently operating neurons.

For a second test case, we add more structures to the balanced network by implementing clusters in the population of excitatory neurons. Specifically, we separate the excitatory neuron population into two groups with denser connections within the cluster and sparse connections between groups. We adjust the ratio of within-cluster and between-cluster connectivity such that the total connectivity across the whole excitatory network remains balanced in proportion to the inhibitory network. The behavior of such a network is described in detail in Litwin-Kumar and Doiron, 2012. Our model is a simplified version assuming transient synaptic dynamics. The clustered network captures firing rate fluctuations present in cortical recording but not in a homogenous balanced network.

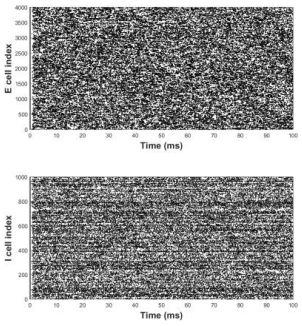
Finally, we consider neural networks with spatial structures. This type of network is built on a balanced network with similar strong excitatory and inhibitory connections. Unlike traditional balanced networks that use Erdős–Rényi graph for connection, a spatial neural network takes into account the fact that spatially close neurons are more likely to form connections. It is also more biologically plausible as it allows the inhibitory neurons to operate at a slower rate. The spatial network has two layers, with layer 1 projecting to layer 2, but not the other way around. The exact method for simulation was described in detail in Huang et al., 2019. This type of network exhibits both high variability and pattern formation. As a consequence of the graph topology, the neural activities of layer 2 neurons are driven by layer 1. Therefore, we expect to see information flow from layer 1 to layer 2. Also, we

expect to see information transfer be higher for neurons that are spatially closer due to higher chances of connectivity.

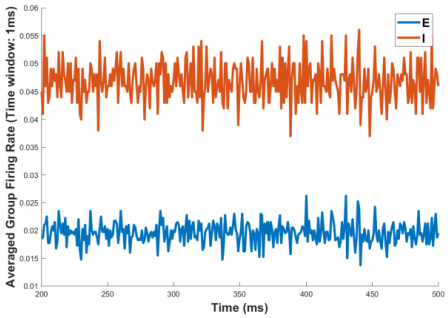
In order to apply transfer entropy to neural recording data, we need to build a pipeline that transforms neural spike train data into information flows. The spike trains are binary and often converted to spike counts data by binning with a given time window in previous studies (Ito et al., 2011). In the following exercise, we try various ways of calculating information flow. We start with the raw neural spiking data and use the plug-in method for binary data by sampling a small subset of neurons from each cluster. Our hope is that this calculation will be sufficient to reveal network structures since the computation for binary data is cheap and the estimator is unbiased with relatively faster converging variance. We also binned the spike trains and computed transfer entropy with the KSG estimator for comparisons.

Next, we subsample large groups of neurons from each cluster and compute the group activities by taking the average of the spike trains. We calculate the transfer entropy between these groups to see if group-level information transfer reveals any structures.

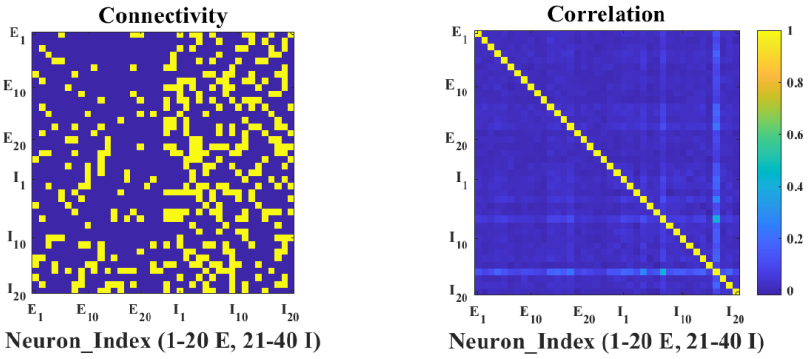
Lastly, we tried multivariate calculation of transfer entropy with the binary estimator or the continuous estimator with the KSG method, using spike train or spike count data respectively. Multivariate methods will not add more insights than simple averaging in cases where all connections are homogenous and without spatial structures. In those cases, each sampled neuron within the cluster contributes equally and their collective information is captured entirely by their summed activities. In the spatial neural net, the contribution from sampled neurons within each region contributes differentially to the target neurons, thus a multivariate nonparametric measure might potentially better capture the information flow through complex mechanisms.



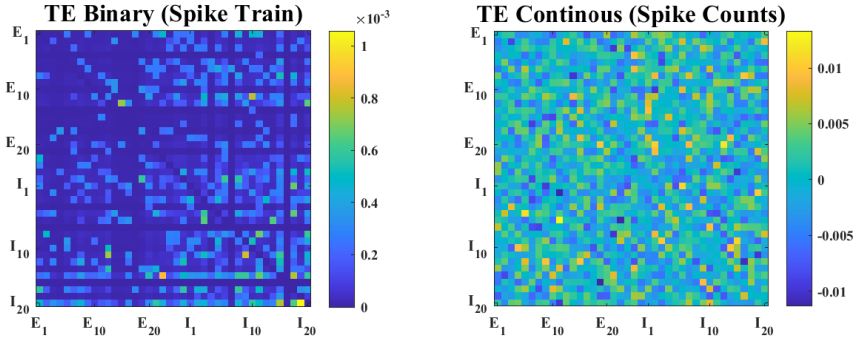

(a) Typical Spike Trains



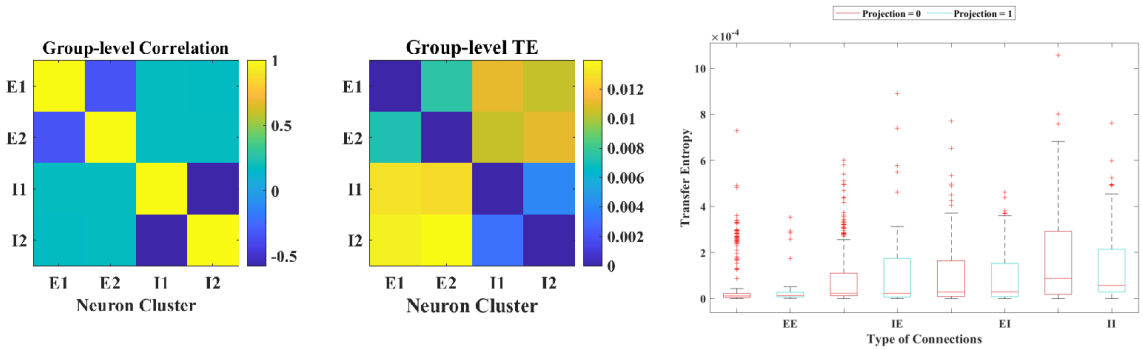

(b) Group Firing Rate




(c) True Connection and Correlation




(d) Pairwise Transfer Entropy




(e) Group-level Transfer Entropy

(f) Distribution of Transfer Entropy

Figure 9: Balanced Network without Extra Structures

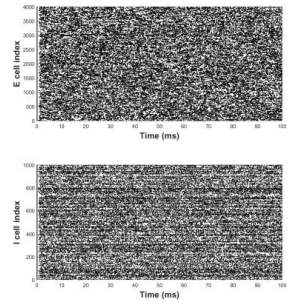
Figures 9 and 10 summaries the results for the network. Fig (a) displays a time-snippet of the spike trains of the sample neurons. Fig (b) displays the group activities for each cluster of neurons. Figure (c) shows the wiring of the sampled neurons and the correlation matrix. Figure (d) shows the transfer entropy calculated based on spike train and spike count data. Figure (e) plots transfer entropy calculation based on group-averaged activities. Two groups are sampled from each cluster. So for the basic balanced network case, there are 4 groups and for the network with two excitatory clusters, there are 6 groups. To see whether there is correspondence between the connection matrix and the pairwise entropy, we also plot the distribution of pairwise entropy with subgroup defined by whether a projection exists in the connection matrix and the neural type at each end of the projection.

In a balanced network without extra structures, the two groups of neurons have stable firing rates, which is characteristic of a balanced network with random connections. The correlation matrix and continuous transfer entropy calculated with spike bins showed little structure. The binary transfer entropy matrix does show more information transfer between the inhibitory and excitatory neurons and within inhibitory neurons, a reflection of denser connections. However, none of the measures were able to recover the true connectomics of the network. This is particularly evident when looking at figure (f). There does not appear to be significant differences between the transfer entropy of directly connected and unconnected pairs. The group-level information transfer shows strong information transfer from inhibitory groups to excitatory groups. There is little information transferred from within the excitatory and inhibitory groups. The group-level transfer entropy matrix shows different characteristics from the correlational matrix.

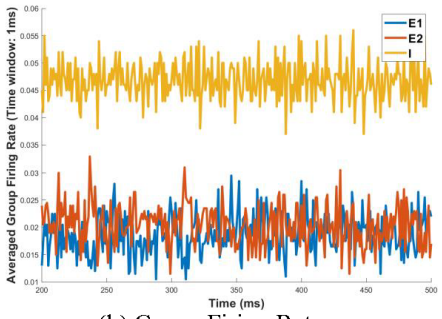
The balanced network with two excitatory clusters showed some interesting patterns. Again, the binary transfer entropy estimator captures the coarse density structures, but none of the measures can match the true connectivity. From the group activity, we see that information transfer is most salient between excitatory and

inhibitory clusters. It appears that between-cluster information transfer is greater than within-cluster information transfer among excitatory neurons.

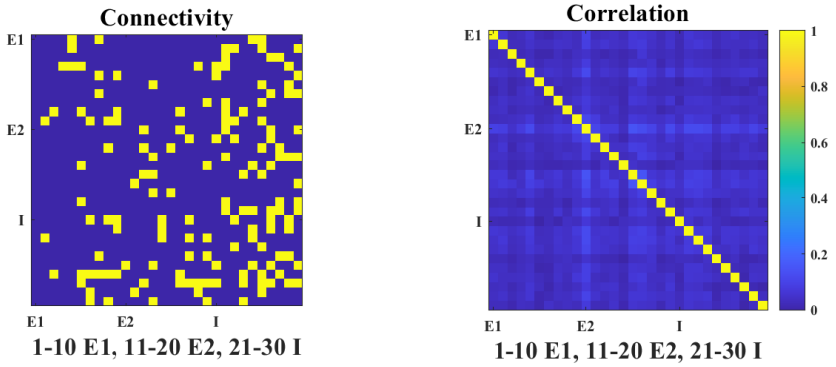
For the balanced network with spatial structures, we see that the transfer entropy estimated using the Gaussian method captures the feedforward projection from neurons in layer 1 to layer 2 of the network. The KSG estimator is too noisy to yield any structure. In part, this is due to the relatively small sample size and the larger variance of the KSG estimator. In fig (b), we use the multivariate transfer entropy calculation to replicate a result documented in Gozel and Doiron, 2022. When the dimensionality of the two layers does not match, communication is hindered, yet downstream activities are still regulated by upstream activities. We want to test if a non-parametric measure of information transfer will exhibit the same pattern. We see that the value of transfer entropy along the forward projecting path is higher compared to the opposite direction. According to our previous tests, transfer entropy tends to be higher along direct paths. The transfer entropy estimated with the Gaussian method falls down as the disparity between the two layers increases. This is expected since transfer entropy estimated with the Gaussian method is still a linear measure. We observe the same pattern when we use another estimator, suggesting that the hindered communication is not an artifact of using a linear measure of information transmission.



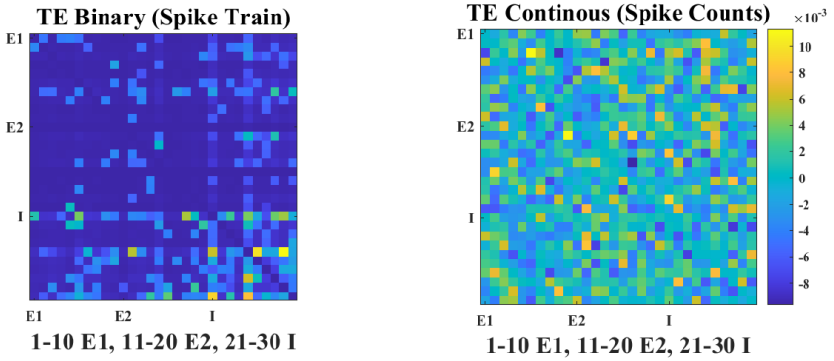

(a) Typical Spike Trains



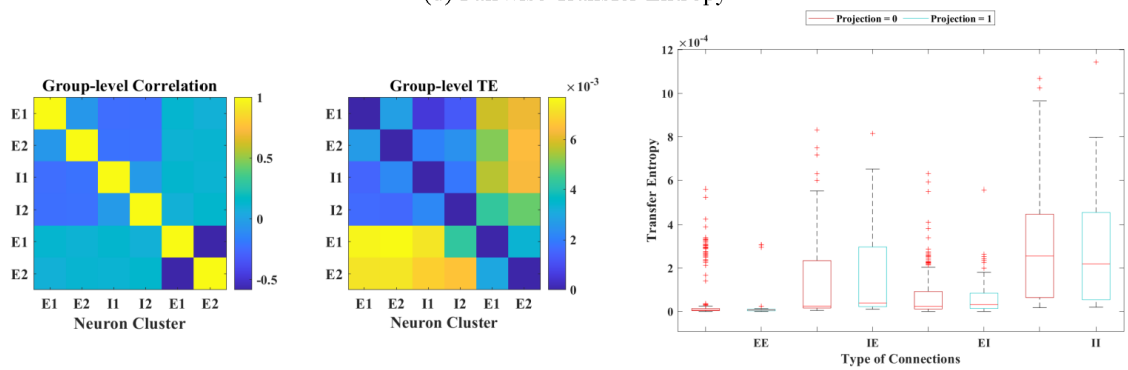

(b) Group Firing Rate




(c) True Connection and Correlation



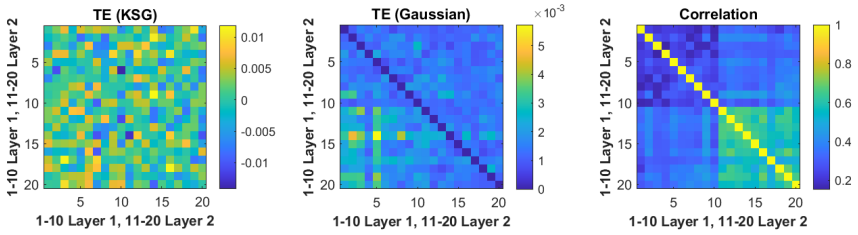

(d) Pairwise Transfer Entropy



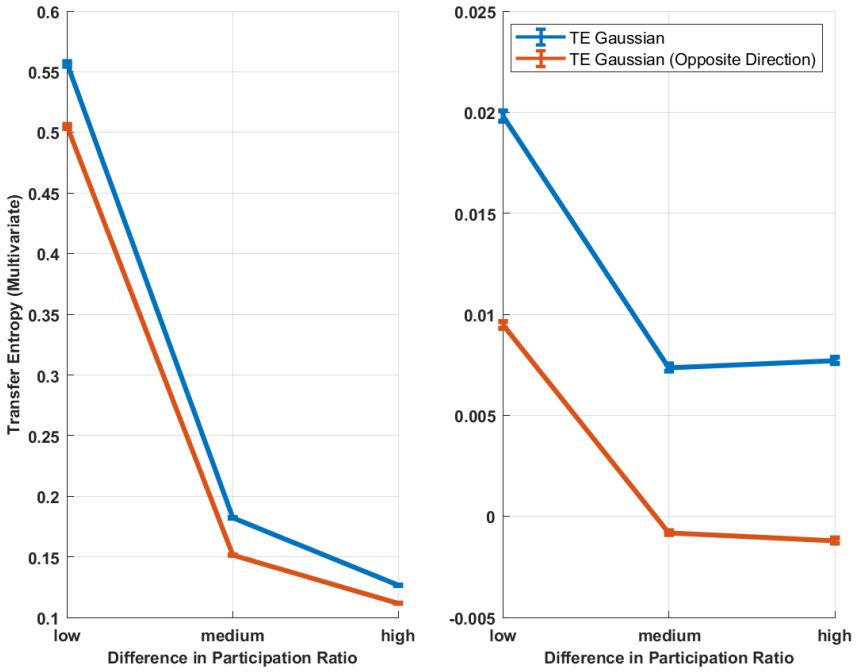

(e) Group-level Transfer Entropy

(f) Distribution of Transfer Entropy

Figure 10: Balanced Network with 2 Excitatory Clusters




(a)




(b)

Figure 11: Spatial Network with 2 Layers

# 5 Discussion

In summary, our experiments demonstrate that the flow of information, as measured by transfer entropy, is intimately related to the network topology. We highlight some general principles that analysts should recognize before applying transfer entropy to neuroscience or social science. We then discuss applications of our current framework in social science with two concrete examples, followed by a discussion of some limitations.

First, while a few papers use a hypothesis testing approach to establish whether or not a direct connection exists between two nodes (Ito et al., 2011; Vicente et al.,

2011), our experiments show that such an approach is less reliable than looking at the relative transfer entropy value. This phenomenon is especially salient if one uses a transfer entropy estimator assuming a Gaussian linear model.

Second, we were able to reconstruct connectivity for firing rate-based models, but not for spiking neural networks. In a balanced network, transfer entropy calculated based on averaged group activities gives more information on the structure. Using binary spike train data, the pairwise transfer entropy calculation matches the overall connection probability but not the specific connectivity. The contribution of each individual neuron to the total inputs for another neuron is minuscule, so it makes sense that the information flow might not be sensitive enough to identify a causal connection. This observation may explain why most articles applying transfer entropy to neuroscience focus on using group-level activities like EEG or fMRI but not spike train data from balanced networks. The implication is similar for social science.

Third, there are many situations where positive information flows exist between two nodes, but there is no causal connection. We expand this discussion in the next section.

Finally, we note that the transfer entropy estimators are often biased when applied to finite time series with self-dependency. This might limit the potential of transfer entropy in studying specific information transmission processes that require an accurate estimation of transfer entropy. Nevertheless, the relative value of transfer entropy can shed light on the network information transmission.

## 5.1 Cautionary Tales for Interpreting Transfer Entropy

Positive transfer entropy does not necessarily implicate a causal structure between two nodes. For instance, consider a simple case where the network is a binary tree with the root transferring information to its subtrees. The left branch has two levels and the right only has one. Thus, both leaves from the left and right subtrees receive information from the root but with different delays. Root variability drives both leaves, but with a lag. Assuming that the left and the right leaves have similar

governing dynamics, the history of the right leave strongly predicts the future of the left leave even though there is no interaction in-between. This simple case illustrates that one should be cautious about the interpretation of structural connection or causality statements when the system of interest has spatial structures or different operating time scales that will cause spurious time causality.

We caution against interpreting information transfers as true causal connection between neurons or neural clusters. First, common inputs or shared noise can cause non-negative information flow between neurons without causal links. We have observed two such mechanisms. One is a time delay in a common input as described above. We observed another case in the firing-based model, where isolated nuerons share a common noise source, and thus predictive information. Even in the simple case of three nodes, non-negative transfer entropy can exist in a pathway with no direct projection. Second, the network topology can confound effective causal connectivity by allowing information to transfer through indirect pathways. Hence, claiming that there is a causal interaction or direct connection based on the existence of transfer entropy can lead to spurious causality or connectivity. In short, transfer entropy measures the predictive information one time series carries about another, not direct influence.

However, whenever there is a direct causal link, transfer entropy is usually present and is higher than transfer entropy due to indirect pathways. For this reason, we recommend that it is safer, despite being more conservative, to only use the transfer entropy with values higher than some threshold as an indication of a causal link.

## 5.2 Application in social science

Social scientists commonly use the network to study social structures and problems. Much work in social network analysis focus on static information of the network like centrality, connectedness, and community detection(Gilbert et al., 2011; Scott, 2011). These static measures give structural information about the underlying social network at a specific time. These descriptions of social network structure need to utilize the rich

information contained in the dynamic processes of network activities. The framework of studying information flow complements current social network analysis approaches by providing a dynamic perspective of how nodes in the network interact. Second, researchers often use linear models like cross-correlation or Granger causality when considering dynamic data. These measures are inappropriate when the interactions between nodes are complicated. Since transfer entropy is a non-parametric measure, it can capture nonlinear interaction, as demonstrated by our analytical cases. Thus, our method adds to the social network methodology by presenting a non-parametric measure that utilizes dynamic information to complement current methods.

For example, researchers defined influencing powers with relative static information like follower counts in the study of influencing power with Twitter data (Grandjean, 2016). We can use the follower network as the structural connectivity of the social network and quantify the information flow on the network by using time series of tweets on specific topics. The information flow network can help identify individuals with high influencing power and study the characteristic of such people. We can also identify communities most insulated from or driven by external information flows.

Information flow is also helpful for the study of academic networks. For example, social scientists used citation networks to study communication efficiency and identify cultural holes (Vilhena et al., 2014). In this paper, the flow of communication was defined with static information of jargon mentioning. We can strengthen the measure of directional influences with transfer entropy that is sensitive to the temporal structures of the data. This dynamic definition of information flow will help validate the current results and might identify different patterns of scientific communication.

## 5.3 Limitations and future research directions

First, note that the experiments in this work are all simulated from models with known parameters. The main parameters are sufficient numbers of time steps back in history (order of the approximated Markov chain) and time delay in information transfer. Neurons can have a complicated dependency on their joint history. Various

types of neurons can also have connections operating at different time scales. In reality, it is impossible to know the exact parameters - one usually has a rough idea of the parameters' regimes or ranges. Das and Fiete, 2020 showed that when there is a mismatch between the assumed models (and parameters) and the true model, connectivity estimation algorithms systematically overestimate connectivity. For future work, this will need to be considered and regularization techniques can be implemented to penalize overestimation.

Another direction worth pursuing is an iterative approach to constructing effective connectivity using conditional transfer entropy (Shahsavari Baboukani et al., 2020). The definition of conditional transfer entropy is the same as transfer entropy while all probability distributions are conditional on given information. One can capitalize on the fact that the existence of connectivity in the network topology often results in elevated information transfer and that spurious effective connection is often a by-product of indirect connection passing through the direct projection. After conditioning out the contribution of the direct pathway, one should expect to see a reduction of false positivity for the identification of network physical connectivity.

Lastly, the results in the current work are mostly computational. Unlike the correlation measure, which already has analytical results connecting network topology and covariance matrix, no such results have yet been demonstrated for transfer entropy. This is because the definition of transfer entropy is less straightforward and involves all orders of statistics. If future studies can connect network topology and information transfer, it will be helpful for understanding neural network functionality based on connectivity information or for designing network connectivity for specific functionality.

## References

Aru, J., Aru, J., Priesemann, V., Wibral, M., Lana, L., Pipa, G., Singer, W., & Vicente, R. (2015). Untangling cross-frequency coupling in neuroscience. *Current opinion in neurobiology*, *31*, 51–61.

Avena-Koenigsberger, A., Misic, B., & Sporns, O. (2018). Communication dynamics in complex brain networks. *Nature reviews neuroscience*, *19*(1), 17–33.

Barnett, L., Barrett, A. B., & Seth, A. K. (2009). Granger causality and transfer entropy are equivalent for gaussian variables. *Physical review letters*, *103*(23), 238701.

Bearden, J., & Mintz, B. (1987). The structure of class cohesion: The corporate network and its dual. *Intercorporate relations: The structural analysis of business*, 187–207.

Borgatti, S. P., Jones, C., & Everett, M. G. (1998). Network measures of social capital. *Connections*, *21*(2), 27–36.

Brown, D. L. (2002). Migration and community: Social networks in a multilevel world. *Rural Sociology*, *67*(1), 1.

Das, A., & Fiete, I. R. (2020). Systematic errors in connectivity inferred from activity in strongly recurrent networks. *Nature Neuroscience*, *23*(10), 1286–1296.

Edelmann, A., Wolff, T., Montagne, D., & Bail, C. A. (2020). Computational social science and sociology. *Annual Review of Sociology*, *46*(1), 61.

Friston, K. J. (2011). Functional and effective connectivity: A review. *Brain connectivity*, *1*(1), 13–36.

Gilbert, F., Simonetto, P., Zaidi, F., Jourdan, F., & Bourqui, R. (2011). Communities and hierarchical structures in dynamic social networks: Analysis and visualization. *Social Network Analysis and Mining*, *1*(2), 83–95.

Goldenberg, D., & Galván, A. (2015). The use of functional and effective connectivity techniques to understand the developing brain. *Developmental cognitive neuroscience*, *12*, 155–164.

Gozel, O., & Doiron, B. (2022). Between-area communication through the lens of within-area neuronal dynamics. *bioRxiv*.

Grandjean, M. (2016). A social network analysis of twitter: Mapping the digital humanities community. *Cogent Arts & Humanities*, *3*(1), 1171458.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, *79*(8), 2554–2558.

Huang, C., Ruff, D. A., Pyle, R., Rosenbaum, R., Cohen, M. R., & Doiron, B. (2019). Circuit models of low-dimensional shared variability in cortical networks. *Neuron*, *101*(2), 337–348.

Ito, S., Hansen, M. E., Heiland, R., Lumsdaine, A., Litke, A. M., & Beggs, J. M. (2011). Extending transfer entropy improves identification of effective connectivity in a spiking cortical network model. *PloS one*, *6*(11), e27431.

Kraskov, A., Stögbauer, H., & Grassberger, P. (2004). Estimating mutual information. *Physical review E*, *69*(6), 066138.

Liao, W., Mantini, D., Zhang, Z., Pan, Z., Ding, J., Gong, Q., Yang, Y., & Chen, H. (2010). Evaluating the effective connectivity of resting state networks using conditional granger causality. *Biological cybernetics*, *102*(1), 57–69.

Lindner, B., Auret, L., Bauer, M., & Groenewald, J. W. (2019). Comparative analysis of granger causality and transfer entropy to present a decision flow for the application of oscillation diagnosis. *Journal of Process Control*, *79*, 72–84.

Litwin-Kumar, A., & Doiron, B. (2012). Slow dynamics and high variability in balanced cortical networks with clustered connections. *Nature neuroscience*, *15*(11), 1498–1505.

Lizier, J. T. (2014). Jidt: An information-theoretic toolkit for studying the dynamics of complex systems. *Frontiers in Robotics and AI*, *1*, 11.

Ocker, G. K., Hu, Y., Buice, M. A., Doiron, B., Josić, K., Rosenbaum, R., & Shea-Brown, E. (2017). From the statistics of connectivity to the statistics of spike times in neuronal networks. *Current opinion in neurobiology*, *46*, 109–119.

Parmelee, C., Moore, S., Morrison, K., & Curto, C. (2022). Core motifs predict dynamic attractors in combinatorial threshold-linear networks. *PloS one*, *17*(3), e0264456.

Schreiber, T. (2000). Measuring information transfer. *Physical review letters*, *85*(2), 461.

Scott, J. (2011). Social network analysis: Developments, advances, and prospects. *Social network analysis and mining*, *1*(1), 21–26.

Shahsavari Baboukani, P., Graversen, C., Alickovic, E., & Østergaard, J. (2020). Estimating conditional transfer entropy in time series using mutual information and nonlinear prediction. *Entropy*, *22*(10), 1124.

Ursino, M., Ricci, G., & Magosso, E. (2020). Transfer entropy as a measure of brain connectivity: A critical analysis with the help of neural mass models. *Frontiers in computational neuroscience*, *14*, 45.

van Vreeswijk, C., & Sompolinsky, H. (1998). Chaotic balanced state in a model of cortical circuits. *Neural computation*, *10*(6), 1321–1371.

Vicente, R., Wibral, M., Lindner, M., & Pipa, G. (2011). Transfer entropy—a model-free measure of effective connectivity for the neurosciences. *Journal of computational neuroscience*, *30*(1), 45–67.

Vilhena, D. A., Foster, J. G., Rosvall, M., West, J. D., Evans, J., & Bergstrom, C. T. (2014). Finding cultural holes: How structure and culture diverge in networks of scholarly communication. *Sociological Science*, *1*, 221.

Wibral, M., Vicente, R., & Lindner, M. (2014). Transfer entropy in neuroscience. In *Directed information measures in neuroscience* (pp. 3–36). Springer.

Zaytsev, Y. V., Morrison, A., & Deger, M. (2015). Reconstruction of recurrent synaptic connectivity of thousands of neurons from simulated spiking activity. *Journal of computational neuroscience*, *39*(1), 77–103.