A Derivation

Here we present a derivation showing that the transfer entropy equals the difference between two types of mutual information:

$$TE(X \to Y) = \sum p\left(y_{t+1}, y_t^{(k)}, x_t^{(l)}\right) \log \frac{p\left(y_{t+1} \mid y_t^{(k)}, x_t^{(l)}\right)}{p\left(y_{t+1} \mid y_t^{(k)}\right)}. \tag{1}$$

Applying the conditional Bayes formula $p\left(y \mid x\right) = \frac{p(y,x)}{p(x)}$ on the numerator and denominator in the log term of equation 1:

$$TE(X \to Y) = \sum p\left(y_{t+1}, y_t^{(k)}, x_t^{(l)}\right) \log \frac{\frac{p\left(y_{t+1}, y_t^{(k)}, x_t^{(l)}\right)}{p\left(y_t^{(k)}, x_t^{(l)}\right)}}{\frac{p\left(y_{t+1}, y_t^{(k)}\right)}{p\left(y_t^{(k)}\right)}}. \tag{2}$$

Adding the marginal distribution of time series $\mathbf{Y}$ to the numerator and denominator simultaneously:

$$TE(X \to Y) = \sum p\left(y_{t+1}, y_t^{(k)}, x_t^{(l)}\right) \log \frac{\frac{p\left(y_{t+1}, y_t^{(k)}, x_t^{(l)}\right)}{p(y_r) p\left(y_t^{(k)}, x_t^{(l)}\right)}}{\frac{p\left(y_{t+1}, y_t^{(k)}\right)}{p(y_r) p\left(y_t^{(k)}\right)}} \tag{3}$$

$$= \sum p\left(y_{t+1}, y_t^{(k)}, x_t^{(l)}\right) \log \frac{p\left(y_{t+1}, y_t^{(k)}, x_t^{(l)}\right)}{p\left(y_r\right) p\left(y_t^{(k)}, x_t^{(l)}\right)}$$

$$- \sum p\left(y_{t+1}, y_t^{(k)}\right) \log \frac{p\left(y_{t+1}, y_t^{(k)}\right)}{p\left(y_r\right) p\left(y_t^{(k)}\right)} \tag{4}$$

$$= MI\left(Y_{t+1}, Y_t^{(k)}, X_t^{(l)}\right) - MI\left(Y_{t+1}, Y_t^{(k)}\right). \tag{5}$$

In these expressions, y_r is sampled from time series $\mathbf{Y}$ randomly each time step and independently of the time step t.

B Transfer Entropy Neural Estimator

B.1 Consistency

Definition. A neural estimator $\widehat{S}(X,Y)_n$ which uses n samples from the data distribution to estimate a statistic $S(X,Y)$ on variables X, Y is *strongly consistent* if for any $\epsilon > 0$, there exists a positive integer N and a choice of neural network such that:

$$\forall n \geq N, \quad \mid S(X,Y) - S(X,Y)_n \mid \leq \epsilon, \text{almost everywhere (a.e.)} \tag{6}$$

The Mutual Information Neural Estimator (MINE) depends on a choice of a neural network and the number of samples n from the data distribution [S1]. Let f_θ be the family of functions parameterized by the neural network with parameters $\theta \in \Theta$. MINE is defined as:

$$\widehat{MI}(X,Y)_n = \sup_{\theta \in \Theta} E_{P_{XY}^{(n)}} [f_\theta] - \log\left(E_{P_X^{(n)} \otimes P_Y^{(n)}} \left[e^{f_\theta}\right]\right). \tag{7}$$

17 **Theorem 1** [S1]**.** MINE is strongly consistent.

18 The Transfer Entropy Neural Estimator (TENE) consists of two independent MINE and depends on
19 choice of neural network and sample number n. TENE is defined as:

$$\widehat{TE}(X \to Y)_n = \widehat{MI}\left(Y_{t+1}, \left(Y_t^{(k)}, X_t^{(l)}\right)\right)_n - \widehat{MI}\left(Y_{t+1}, Y_t^{(k)}\right)_n. \tag{8}$$

20 We use $MI^{[1]}$, $MI^{[2]}$, $\widehat{MI}_n^{[1]}$, $\widehat{MI}_n^{[2]}$ as abbreviations of $MI\left(Y_{t+1}, \left(Y_t^{(k)}, X_t^{(l)}\right)\right)$ and
21 $MI\left(Y_{t+1}, Y_t^{(k)}\right)$, $\widehat{MI}\left(Y_{t+1}, \left(Y_t^{(k)}, X_t^{(l)}\right)\right)_n$ and $\widehat{MI}\left(Y_{t+1}, Y_t^{(k)}\right)_n$ respectively.

22 We will prove the following:
23 **Theorem 2.** TENE is strongly consistent.
24 **Proof.** Let $\epsilon > 0$. By Theorem 1, we can choose neural networks and integers N_1, N_2 and such that

$$\forall n \geq N_1, \quad \left|MI^{[1]} - \widehat{MI}_n^{[1]}\right| \leq \epsilon/2, \text{a.e.} \tag{9}$$

$$\forall n \geq N_2, \quad \left|MI^{[2]} - \widehat{MI}_n^{[2]}\right| \leq \epsilon/2, \text{a.e.} \tag{10}$$

25 Letting $N = \max\{N_1, N_2\}$, for $n \geq N$ and for some neural network we have, a.e.,

$$\forall n \geq N, \left|TE(X \to Y) - \widehat{TE}(X \to Y)_n\right| = \left|(MI^{[1]} - MI^{[2]}) - (\widehat{MI}_n^{[1]} - \widehat{MI}_n^{[2]})\right| \tag{11}$$

$$= \left|(MI^{[1]} - \widehat{MI}_n^{[1]}) - (MI^{[2]} - \widehat{MI}_n^{[2]})\right| \tag{12}$$

$$\leq \left|(MI^{[1]} - \widehat{MI}_n^{[1]})\right| + \left|(MI^{[2]} - \widehat{MI}_n^{[2]})\right| \tag{13}$$

$$\leq \epsilon/2 + \epsilon/2 = \epsilon. \tag{14}$$

26 The proof is complete.

B.2 Variation of bias vary with dimension and noise

28 We examine the performance of TENE for the considered class of neural networks on linear dynamic
29 system, consisting of variables X and Y defined as:

$$x_{t+1} = \alpha x_t + \varepsilon_x \tag{15}$$
$$y_{t+1} = \beta y_t + g_c x_t + \varepsilon_y \tag{16}$$

30 We set $\alpha = \beta = 0.5$ and $\varepsilon_x = \varepsilon_x \sim N(0, \sigma^2)$. The true values of transfer entropy $TE(X \to Y)$
31 in this simple coupled system can be determined analytically [S2]. We can increase the dimension
32 of the system by considering multiple independent copies of variables X and Y, in which case the
33 mutual information and transfer entropy scale linearly with the dimension of the system. For each
34 considered dimension, standard deviation σ, and coupling strength g_c in an interval from -0.4 to
35 0.4, we generate a time series of length 50,000. We also consider an alternative non-parametric
36 estimator of mutual information, the Kraskov estimator [S3] with $k = 5$ nearest neighbours. In Fig 1
37 we compare the results of MINE with the analytic formula and the Kraskov estimator. MINE shows
38 marked improvement over the Kraskov estimator, especially when variables are high-dimensional.
39 Comparing Fig 1(a,b) or (c,d) shows that the amplitude of the driving Gaussian noise has little
40 influence on estimates. Interestingly, as coupling strength g_c grows small, i.e., as X and Y become
41 more independent, the Kraskov estimator can suggest a negative value of the mutual information,
42 i.e., we estimate that $MI_n(Y_{t+1}, Y_t, X_t) < MI_n(Y_{t+1}, Y_t)$. We deduce that irrelevant information
43 about the nearly independent variable X_t interferes with the estimation of the mutual information by
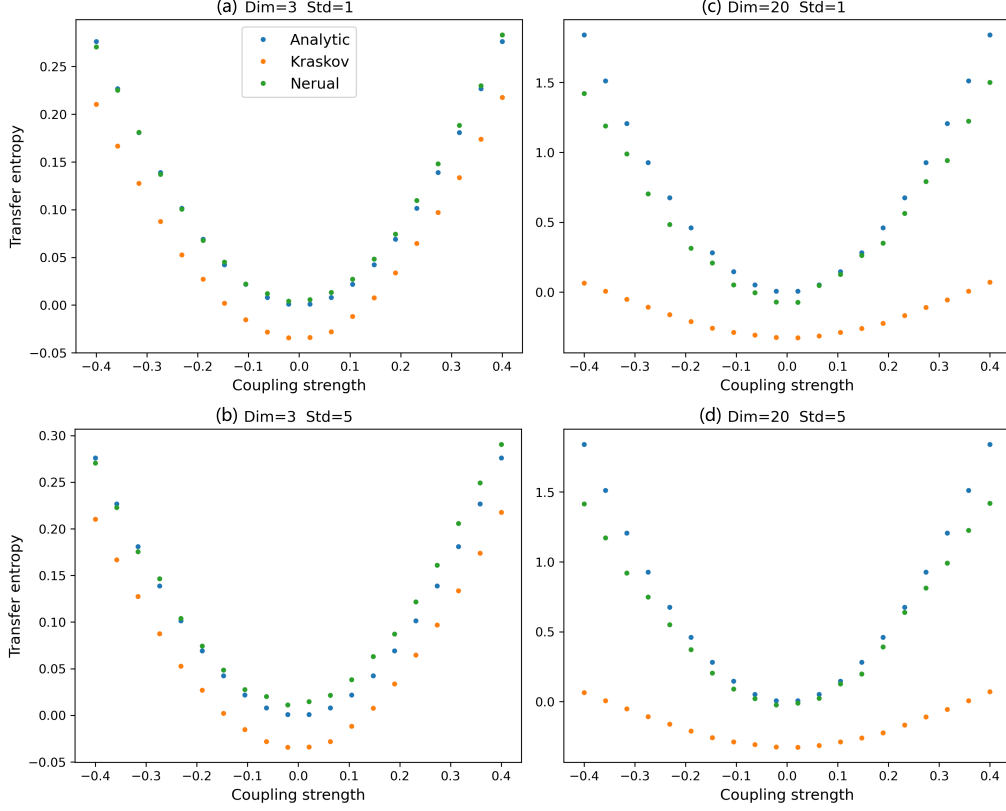44 the Kraskov estimator.

Figure 1: True and estimated transfer entropy versus coupling strength g_c. The dimension and standard deviation (std) σ of system noise is indicated in the titles of subplots. The units of transfer entropy are bits.

C Algorithm

Details on the implementation of coupling attention mechanism are provided in Algorithm 1.

Algorithm 1 Coupling Relationship Inference

Input: Small samples e.g., $\{(x_i, y_i)\}_{i=1}^{L}$ as train set S, total number of epochs E
1: $\theta, \phi, \alpha \leftarrow$ initialize network parameters
2: **Stage 1:** iteration until epoch exceeds E
3:     Assign coupling attention coefficients $\{a_i\}_{i=1}^{L}$
4:     Compute $\mathcal{L}_1, \mathcal{L}_2$ on S
5:     $\theta, \phi \leftarrow \theta + \nabla_\theta \mathcal{L}_1, \phi + \nabla_\phi \mathcal{L}_2$
6:     Redo lines 3-4
7:     $\alpha \leftarrow \alpha + \nabla_\alpha (\mathcal{L}_1 - \mathcal{L}_2)$
8:     $\eta \leftarrow$ initialize network parameters
9:     Compute $\mathcal{L}_3$ on S', e.g., $\{(x_i, y_i) * a_i\}_{i=1}^{L}$
10:     $\eta \leftarrow \eta - \nabla_\eta \mathcal{L}_3$
11: Record optimal parameter η^* during **Stage 1**
12: **Stage 2:** iteration until $\mathcal{L}_3$ convergence
13:     Compute $\mathcal{L}_3$ on S'
14:     $\eta^* \leftarrow \eta^* - \nabla_{\eta^*} \mathcal{L}_3$

D Model Neuronal Dynamics

Here we give detailed information about five neuronal dynamics applied to modeling membrane potential and relevant quantities in biological connectomes. We input to each causal discovery algorithm the coordinate corresponding to the neuron membrane voltage potential, because this variable is most likely to be experimentally accessible.

D.1 Hindmarsh-Rose dynamics

The spikes of activity in neurons are considered an important part of the brain's information processing [S4, S5]. Hindmarsh and Rose [S6] (HR) proposed a phenomenological neuron model that is a simplification of the Hodgkin-Huxley model [S7]. The HR model is described by

$$\dot{p} = q - ap^3 + bp^2 - n + I_{\text{ext}}$$
$$\dot{q} = c - dp^2 - q$$
$$\dot{n} = r\left[s\left(p - p_0\right) - n\right]$$

where $p(t)$ is the action potential of the membrane, $q(t)$ is related to the fast current and $n(t)$ is associated with the slow current. Presynaptic neurons with an action potential p_j coupled by chemical synapses to neurons i modifying its action potential p_i according to

$$\dot{p}_i = q_i - ap_i^3 + bp_i^2 - n + I_{\text{ext}} + \Gamma$$
$$\Gamma = g_c\left(V_{\text{syn}} - p_i\right)\sum_{j=1}^{N}\frac{B_{ij}}{1 + \exp(-\lambda\left(p_j - \Theta_{\text{syn}}\right))}$$

where $i, j = 1, \ldots, N$, N is the number of neurons, g_c is the chemical coupling strength and B_{ij} describes neurons' chemical connections. The chemical synapse function is modeled by the above sigmoidal function, with $\Theta_{\text{syn}} = 1.0$. We use parameters $a = 1, b = 3, c = 1, u = 5, s = 4, r = 0.005, p_0 = -1.60$, coupling strength $g_c = 0.1, V_{\text{syn}} = 2, \lambda = 10$, and external current $I_{\text{ext}} = 3.24$, for which HR neurons exhibits a chaotic burst behavior.

D.2 Izhikevich dynamics

Izhikevich dynamics reproduce spiking and bursting behavior of known types of cortical neurons, and combine the biological plausibility of Hodgkin–Huxley-type dynamics and the computational efficiency of integrate-and-fire neurons [S8]. The equations governing Izhikevich spike dynamics are:

$$\dot{v} = 0.04v^2 + 5v + 140 - u + I + g_c\sum B_{ij}u_j$$
$$\dot{u} = a(bv - u)$$

with the auxiliary after-spike resetting

$$\text{if } v \geq +30\text{mV}, \quad \text{then }\left\{\begin{array}{l} v \leftarrow c \\ u \leftarrow u + d \end{array}\right. .$$

Here, variable v represents the membrane potential of the neuron and u represents a membrane recovery variable, which accounts for the activation of K^+ ionic currents and inactivation of Na^+ ionic currents, and it provides negative feedback to v. Here, we use the parameters $a = 0.2, b = 2, c = -56, d = -16, I = -99$. After the spike reaches its apex ($+30\text{mV}$), the membrane voltage and the recovery variable are reset. If v skips over 30, then it first is reset to 30, and then to c so that all spikes have equal magnitudes.

D.3 Rulkov dynamics

The Rulkov model is a map-based neuron model with a surprising abundance of features, such as periodic and chaotic spiking, and bursting. The Rulkov map is an abstract mathematical model, although it shares some specific features with others neuron models closer to experimental observations. We use synthetic time series where each neuron is simulated using the Rulkov

4

model [S9], which has two variables, u and w, evolving at different timescales as described by $\boldsymbol{x}(t+1) = (u(t+1), v(t+1)) = \boldsymbol{F}(\boldsymbol{x}(t)) = (F_1(u(t), w(t)), F_2(u(t), w(t)))$, with

$$F_1(u, w) = \frac{\beta}{1 + u^2} + w + \Gamma(u) \quad \text{and} \quad F_2(u, w) = w - \nu u - \sigma.$$

The two variables reflect the two important time scales of a neuron model. The variable u represents the fast dynamics of the system and usually models the membrane voltage of the neuron, whereas w is the slow variable and represents the variations of the ionic recovery currents. Different combinations of parameters σ and β give rise to different dynamical states of the neuron, such as resting, tonic spiking, and chaotic bursts. As for the coupling, we consider chemical synaptic coupling, that is, $\boldsymbol{H}(\boldsymbol{x}_i, \boldsymbol{x}_j) = (h(u_i, u_j), 0)$ with $h(u_i, u_j) = (u_i - V_s)\Gamma(u_j)$, where

$$\Gamma(u_j) = \frac{1}{1 + \exp\{\lambda(u_j - \Theta_s)\}}$$

and electrical synaptic coupling, $\boldsymbol{H}(\boldsymbol{x}_i, \boldsymbol{x}_j) = (h(u_i, u_j), 0)$, with $h(u_i, u_j) = u_j - u_i$. In the chemical coupling, V_s is a parameter called the reverse potential. Here, we use the parameters with $\beta = 4.4, \sigma = \nu = 0.001,, V_s = 20, \Theta_s = -0.25$ and $\lambda = 10$.

D.4 FitzHugh-Nagumo dynamics

A FitzHugh-Nagumo neuron comprises a two-dimensional system of smooth ODEs, so cannot exhibit autonomous chaotic dynamics and bursting. Adding noise allows for stochastic bursting [S10]. The equations governing the FitzHugh-Nagumo neuronal network dynamics are

$$\dot{v} = a + bv + cv^2 + dv^3 - u + \Gamma$$
$$\dot{u} = \varepsilon(ev - u)$$

with the coupling term

$$\Gamma(v_i) = -g_c \sum_{j=1}^{N} B_{ij}(v_j - v_i).$$

The FitzHugh-Nagumo dynamics capture the firing behaviors of neurons with two components. The first component v represents the membrane potential, which contains self- and interaction dynamics, and the second component u represents a recovery variable. To simulate the shape of each spike, the time step in the model must be relatively small, e.g., $\tau = 0.25$ ms. Here, we use the parameters $a = 0.28, b = 1, c = 0, d = -1, \varepsilon = 0.04, e = 12.5$. Moreover, the parameters in the FitzHugh–Nagumo model can be tuned so that the model describes spiking dynamics of many resonator neurons.

D.5 Time series generation

To obtain the time series from above neural dynamics, we use Runge-Kutta method with variable-step to solve the ordinary differential equation of Hindmarsh-Rose and FitzHugh-Nagumo with sample interval $\tau = 0.1$. Izhikevich dynamics are solved by the Euler formula with time step $h = 0.05$. For the Rulkov map we consider a unit sample interval.

E   Real Connectomes Information

E.1   Cat connectome

The cat connectivity dataset comprises a description of cortical connections in the cat brain [S11], a connectivity set resulting from a comprehensive literature search of anatomical tracing studies in the cat cortex. Detailed information on the delineated regions, including information on the used parcellation scheme, abbreviations and possible overlap with other parcellation schemes, as well as information on the physiological characteristics of these regions, is given in the appendix of the original study Ref. [S11]. The connectivity dataset incorporates data of one hemisphere, including 65 regions and 1139 interregional macroscopic axonal projections [S12].

E.2 Macaque connectome

The macaque connectivity data set used in this study comprises anatomical data from 410 tract tracing studies collated in the online neuroinformatics data base CoCoMac (http://cocomac.org), first analyzed and made publicly available in Ref. [S13]. In the present study they focused primarily on an analysis of the connectivity among regions of the cerebral cortex. The cortical connection matrix was extracted from the primary connection data by removing all subcortical (thalamus, basal ganglia, brainstem) regions. In addition, regions that did not maintain at least one incoming and one outgoing connection were also removed to ensure that the network was strongly connected. The remaining connection data set used in this study consisted of 242 regions and 4090 directed projections represented in binary format (connection present = 1, connection absent = 0) [S14].

E.3 Mouse connectome

The Allen Mouse Brain Connectivity Atlas uses enhanced green fluorescent protein (EGFP)-expressing adeno-associated viral vectors to trace axonal projections from defined regions and cell types, and high-throughput serial two-photon tomography to image the EGFP-labelled axons throughout the brain. This systematic and standardized approach allows spatial registration of individual experiments into a common three dimensional (3D) reference space, resulting in a whole-brain connectivity matrix. The Allen Mouse Brain Connectivity Atlas is a freely available, foundational resource for structural and functional investigations into the neural circuits that support behavioural and cognitive processes in health and disease [S15].

E.4 C.elegans connectome

All the chemical and gap junction synapses, the connectome, in the posterior nervous system of the C.elegans adult male are identified by serial section electron microscopy [S16]. The feasibility of comprehensive synapse-level nervous system reconstruction by this method was a primary reason for the initial selection of C. elegans as an experimental model. They developed a PC-based software platform to facilitate assembly of a connectome from electron micrographic images. The connectome is of a single adult animal and was produced from a series of 5000 serial thin sections of 70 to 90 nm encompassing the posterior half of the body.

E.5 Rat connectome

The Rat Connectome used in this work represents connection patterns between distinct gray matter regions in a rat brain, and was presented as an example of the Brain Architecture Knowledge Management System (BAMS) [S17], where it is called the rat macroconnectome. The rat (macro)connectome was inferred by combining data from neuroscientific literature as well as connectivity reports inserted into BAMS [S17].

E.6 Drosophila connectome

Electron microscopy was used to collect the connectome of a Drosophila brain [S18]. The brain was cut into very thin slices and photographed with an electron microscope. A three-dimensional map of the neurons and connections in the brain was then reconstructed from these images using machine learning algorithms. By matching reconstructed neurons to examples from light microscopy, they assigned neurons to cell types and assembled a connectome of the repeating module of the medulla.

Information about the above datasets is summarized in Table 1. We access all the brain connectomes from NeuroData's Graph DataBase from `https://neurodata.io/project/connectomes/`.

6

Table 1: Statistical information of six real networks: dataset name, type of network, number of nodes, number of edges, mean degree $\langle k \rangle$, and data acquisition method.

Dataset	Region	#Nodes	#Edges	Mean degree	Sensor
Cat	Cerebral Cortex	65	1139	17.5	Tract tracing studies
Macaque	Cerebral Cortex	242	4090	16.9	Tract tracing studies
Mouse	Cerebral Cortex	195	214	1.1	Electron microscopy
C. elegans	Neural	272	4451	16.4	Electron Microscopy
Rat	Gray matter	503	30088	59.8	Neuroanatomical experiments
Drosophila	Medulla	1781	33641	18.9	Electron Microscopy

F Additional Experiments

Performance on real coupled networks. Results are provided in Table 2345 as the supplement of main text Table 2.

Table 2: Performance comparison on Macaque connectome. Each point contains the mean and standard deviation of AUROC.

	HR	IZH	RULKOV	FHN
GRANGER	0.50±0.01	0.48±0.02	0.54±0.02	0.54±0.01
TE KRASKOV	0.62±0.02	0.50±0.02	0.60±0.02	0.45±0.02
CCM	0.44±0.02	0.51±0.01	0.55±0.01	0.56±0.01
LATENT CCM	0.47±0.01	0.51±0.01	0.53±0.02	0.53±0.01
PCMCI	0.47±0.01	0.50±0.02	0.50±0.01	0.52±0.01
PCMCI+	0.47±0.02	0.51±0.02	0.52±0.02	0.52±0.01
CLASSIFIER	0.72±0.04	0.57±0.04	0.74±0.03	0.60±0.04
TENE	0.53±0.07	0.47±0.02	0.71±0.03	0.51±0.05
ATEN	**0.90**±0.1	**0.63**±0.14	**0.79**±0.10	**0.84**±0.13

Table 3: Performance comparison on Mouse connectome.

	HR	IZH	RULKOV	FHN
GRANGER	0.62±0.03	0.54+0.03	0.89+0.01	—
TE KRASKOV	0.50±0.02	0.54±0.03	0.86±0.03	—
CCM	0.55±0.05	0.47±0.04	0.58±0.01	—
LATENT CCM	0.47±0.01	0.51±0.01	0.53±0.02	—
PCMCI	0.50±0.02	0.50±0.02	0.77±0.03	—
PCMCI+	0.53±0.03	0.50±0.05	0.79±0.02	—
CLASSIFIER	**0.78**±0.07	**0.64**±0.05	**0.99**±0.01	—
TENE	0.53±0.04	0.51±0.03	0.80±0.08	—
ATEN	0.72±0.12	0.63±0.20	0.90±0.09	—

Table 4: Performance comparison on Cat connectome.

	HR	IZH	RULKOV	FHN
GRANGER	0.50±0.01	0.59±0.02	0.65±0.01	0.53±0.01
TE KRASKOV	0.62±0.02	0.50±0.02	0.60±0.02	0.45±0.02
CCM	0.75±0.01	0.51±0.01	0.57±0.01	0.68±0.02
LATENT CCM	0.69±0.02	0.48±0.03	0.51±0.01	0.64±0.02
PCMCI	0.53±0.01	0.49±0.01	0.66±0.01	0.53±0.01
PCMCI+	0.57±0.01	0.52±0.02	0.63±0.02	0.56±0.01
CLASSIFIER	0.72±0.04	0.57±0.04	0.74±0.03	0.60±0.04
TENE	0.53±0.07	0.47±0.02	0.71±0.03	0.51±0.05
ATEN	**0.84**±0.13	**0.62**±0.10	**0.89**±0.07	**0.75**±0.13

Table 5: Performance comparison on Rat connectome.

	HR	IZH	RULKOV	FHN
GRANGER	0.50±0.01	0.50±0.01	0.46±0.01	0.52±0.02
TE KRASKOV	0.56±0.01	0.56±0.01	0.43±0.02	0.51±0.02
CCM	0.63±0.02	0.54±0.05	0.50±0.01	0.52±0.03
LATENT CCM	0.63±0.02	0.55±0.01	0.53±0.02	0.51±0.03
PCMCI	0.51±0.01	0.53±0.01	0.46±0.01	0.49±0.02
PCMCI+	0.51±0.02	0.55±0.01	0.44±0.02	0.49±0.01
CLASSIFIER	0.86±0.04	0.78±0.07	0.82±0.03	0.77±0.06
TENE	0.89±0.03	0.32±0.10	0.86±0.03	0.70±0.06
ATEN	**0.94**±0.06	**0.93**±0.03	**0.90**±0.06	**0.85**±0.11

We observe that our method usually substantially improves reconstruction performance on real coupled networks, exception for the Mouse connectome with very low mean degree (see Table. 1).

References

S1. Belghazi, M. I. *et al.* Mutual information neural estimation. In *International Conference on Machine Learning*, 531–540 (PMLR, 2018).

S2. Kaiser, A. & Schreiber, T. Information transfer in continuous processes. *Physica D: Nonlinear Phenomena* **166**, 43–62 (2002).

S3. Kraskov, A., Stögbauer, H. & Grassberger, P. Estimating mutual information. *Physical review E* **69**, 066138 (2004).

S4. Borges, F. *et al.* Inference of topology and the nature of synapses, and the flow of information in neuronal networks. *Physical Review E* **97**, 022303 (2018).

S5. Rabinovich, M. I., Varona, P., Selverston, A. I. & Abarbanel, H. D. Dynamical principles in neuroscience. *Reviews of Modern Physics* **78**, 1213 (2006).

S6. Hindmarsh, J. L. & Rose, R. A model of neuronal bursting using three coupled first order differential equations. *Proceedings of the Royal Society of London. Series B. Biological Sciences* **221**, 87–102 (1984).

S7. Hodgkin, A. L. & Huxley, A. F. Propagation of electrical signals along giant nerve fibres. *Proceedings of the Royal Society of London. Series B-Biological Sciences* **140**, 177–183 (1952).

S8. Izhikevich, E. M. Simple model of spiking neurons. *IEEE Transactions on Neural Networks* **14**, 1569–1572 (2003).

S9. Eroglu, D., Tanzi, M., van Strien, S. & Pereira, T. Revealing dynamics, communities, and criticality from data. *Physical Review X* **10**, 021047 (2020).

S10. FitzHugh, R. Impulses and physiological states in theoretical models of nerve membrane. *Biophysical Journal* **1**, 445–466 (1961).

S11. Scannell, J. W., Blakemore, C. & Young, M. P. Analysis of connectivity in the cat cerebral cortex. *Journal of Neuroscience* **15**, 1463–1483 (1995).

S12. de Reus, M. A. & van den Heuvel, M. P. Rich club organization and intermodule communication in the cat connectome. *Journal of Neuroscience* **33**, 12929–12939 (2013).

S13. Modha, D. S. & Singh, R. Network architecture of the long-distance pathways in the macaque brain. *Proceedings of the National Academy of Sciences* **107**, 13485–13490 (2010).

S14. Harriger, L., Van Den Heuvel, M. P. & Sporns, O. Rich club organization of macaque cerebral cortex and its role in network communication. *PLoS ONE* **7**, e46497 (2012).

S15. Oh, S. W. *et al.* A mesoscale connectome of the mouse brain. *Nature* **508**, 207–214 (2014).

S16. Jarrell, T. A. *et al.* The connectome of a decision-making neural network. *Science* **337**, 437–444 (2012).

S17. Bota, M., Dong, H.-W. & Swanson, L. W. Combining collation and annotation efforts toward completion of the rat and mouse connectomes in bams. *Frontiers in Neuroinformatics* **6**, 2 (2012).

S18. Takemura, S.-y. *et al.* A visual motion detection circuit suggested by drosophila connectomics. *Nature* **500**, 175–181 (2013).