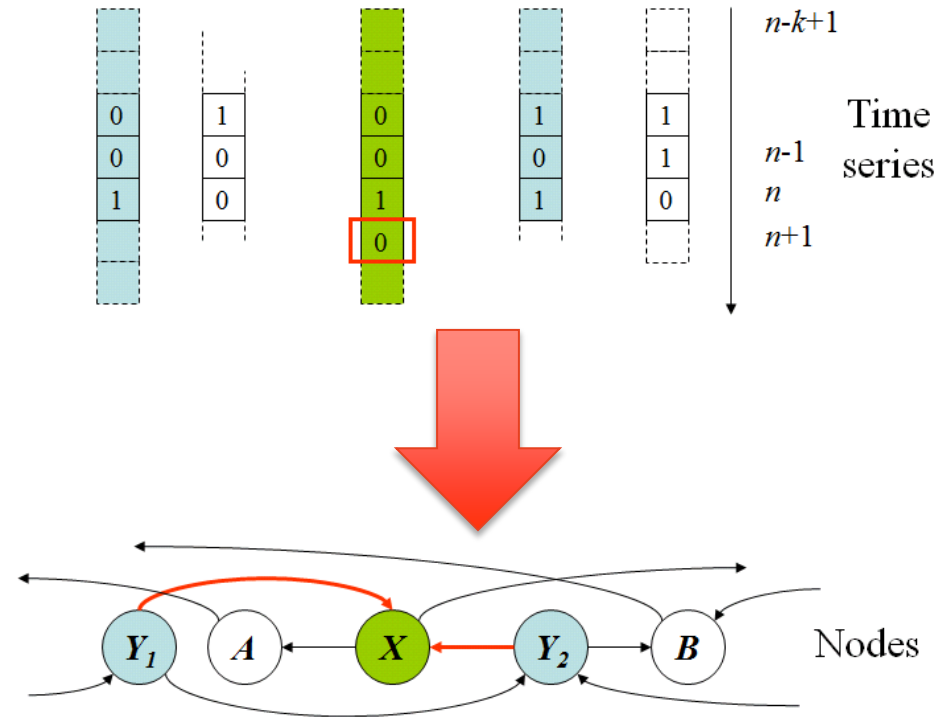


# Inferring network models from multivariate time- series data: *Philosophy, approaches and considerations*

**A/Prof. Joseph Lizier**

School of Computer Science  
Centre for Complex Systems

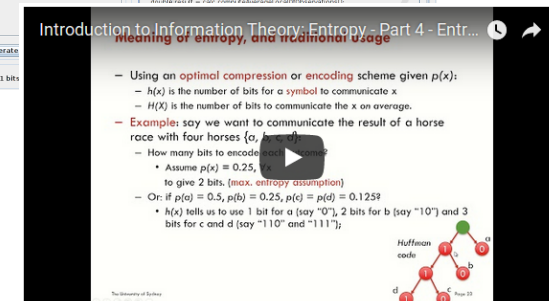
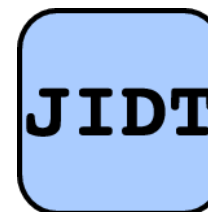
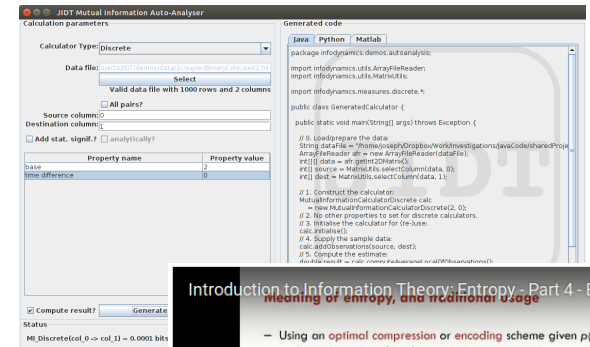


# Inferring network models

## Session outcomes:

1. Understand **philosophy** of different goals in network inference from time-series data
2. Understand **measures** and **approaches** (and how to access them)
3. Understand key **considerations** to incorporate

- Primary references:
- Github:
  - <http://github.com/jlizier/jidt/>
  - <https://github.com/pwollstadt/idxl>
- Short course: <http://bit.ly/jidt-course-alpha>
  - Lecture slides (“course” branch, module 12)
  - Youtube videos
  - Tutorial

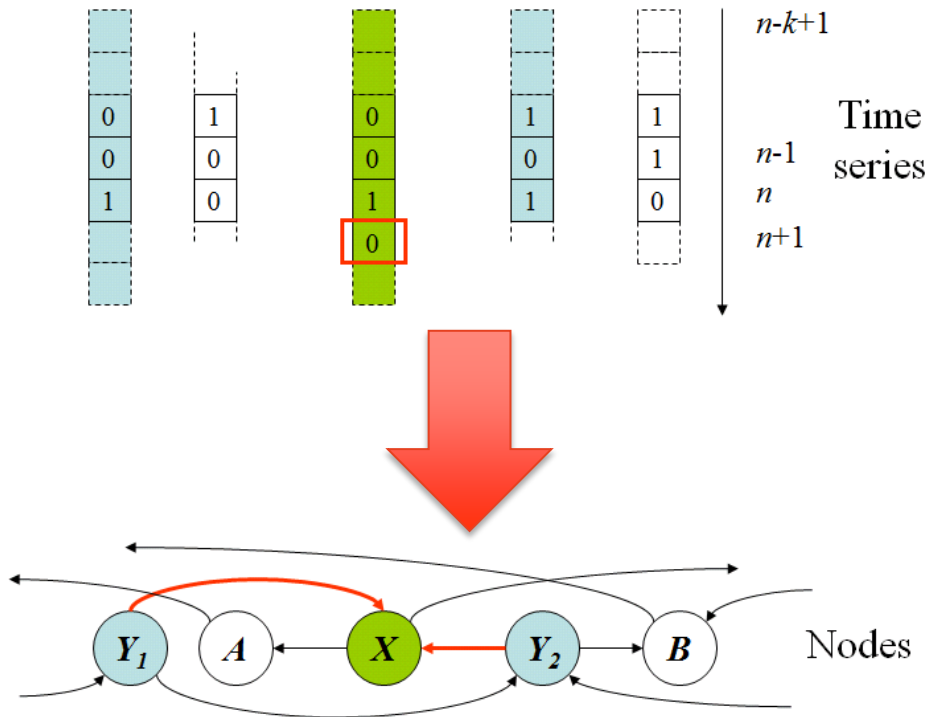


# Scope

- Many, many approaches
  - Cannot cover all
- Focus on:
  - Linear regression models
  - Information-theoretic measures
- Application areas:
  - Neural recordings
  - Financial market data sets
  - Gene regulatory networks
  - Sport analytics
  - ...

# Inferring network models

- Key question: *given only time series for each of a set of variables, how can we build network **model** which represents the **relationships** between these variables?*



Complex system as a **multivariate time-series** of activity of variables.

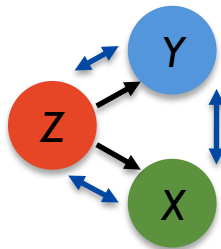
Examples?

Options, philosophically:

1. Functional connectivity
2. Effective connectivity
3. Structural connectivity

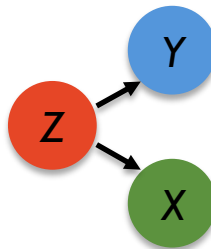
# Functional network inference

- Constructs (**undirected**) networks to represent **pairwise statistical relationships** between nodes
- Usually using a measure of correlation or mutual information (MI) to provide pairwise edge weights
- Trade-off: fast and simple, but does not explain dynamics



# Structural network inference

- Constructs **directed** networks to represent the physical, directed (**causal**) connections
- *Generally* only possible via interventional techniques but not directly from large (observational) multivariate time-series sets alone.
- Does not tell us about time or experimentally modulated changes in how the variables are interacting
- Trade-off: full picture (?), but isn't generally achievable.



O. Sporns. *Networks of the Brain*. MIT Press, Cambridge, Massachusetts, USA, 2011

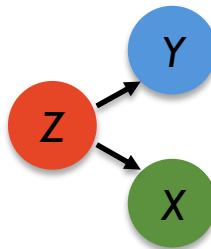
K. J. Friston. Functional and effective connectivity in neuroimaging: A synthesis. *Human Brain Mapping*, 2(1-2):56–78, 1994.

J. Pearl, *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press, 2000.

Bossomaier, Barnett, Harré, Lizier, "An Introduction to Transfer Entropy: Information Flow in Complex Systems", Springer, Cham, 2016; section 7.2

# Effective network inference

- Constructs **directed** networks to provide “*minimal neuronal circuit model*” which can replicate and indeed explain the time series of the nodes
- Not strictly causal, but aims at providing best **generative model** possible from data.
- Hybrid approach, and aligns with dynamical systems/computational views
- Trade-off: multivariate model, data/runtime requirements



O. Sporns. *Networks of the Brain*. MIT Press, Cambridge, Massachusetts, USA, 2011

K. J. Friston. Functional and effective connectivity in neuroimaging: A synthesis. *Human Brain Mapping*, 2(1-2):56–78, 1994.

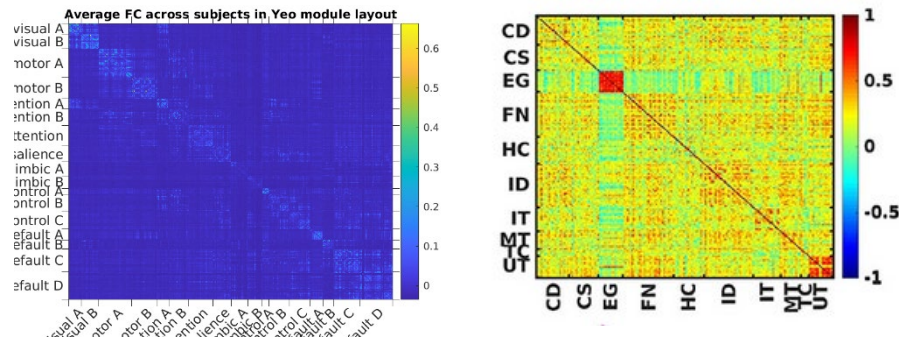
Bossomaier, Barnett, Harré, Lizier, "An Introduction to Transfer Entropy: Information Flow in Complex Systems", Springer, Cham, 2016; section 7.2

# Functional connectivity



# Functional connectivity

- Constructs **undirected** networks to represent **pairwise statistical relationships** between nodes
- Pros:
  - Easy and fast to implement and run (e.g. correlation)
  - Simple to interpret
  - Good at identifying network modules with community detection [1,2]



- Sensitive to changes in the underlying dynamics
  - E.g. changes under conditions such as Alzheimer's disease [3]
  - Even if structure does not change

1. After: B. T. T. Yeo et al., J. Neurophys., 106(3), 1125, (2011)

2. Kukreti et al (2020) *Front. Phys.* 8:323. doi: 10.3389/fphy.2020.00323 (Figure used under CC BY)

3. J. A. Contreras et al., *Neuroimage Clin.*, 22, 101687, (2019)

# Functional connectivity: measures

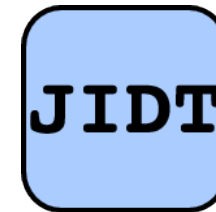
- Correlation (Pearson)

$$r_{xy} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}}$$

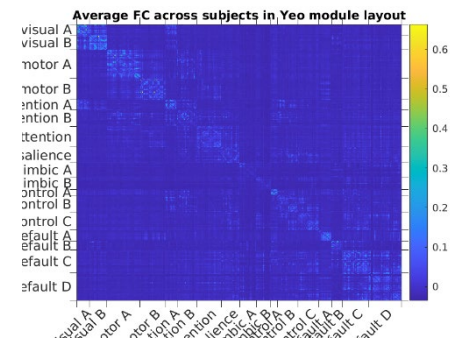
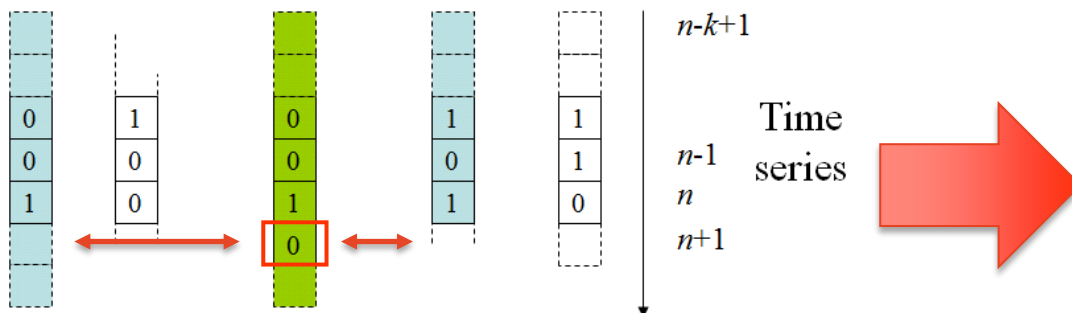
- Spearman correlation

- Mutual information

- $I(X; Y) = \sum_{x,y} p(x, y) \log \frac{p(x|y)}{p(x)}$
- Captures **non-linear** statistical relationships
  - More powerful, but requires more data



<https://github.com/jlazier/jidt>



# Consideration 1: data preprocessing

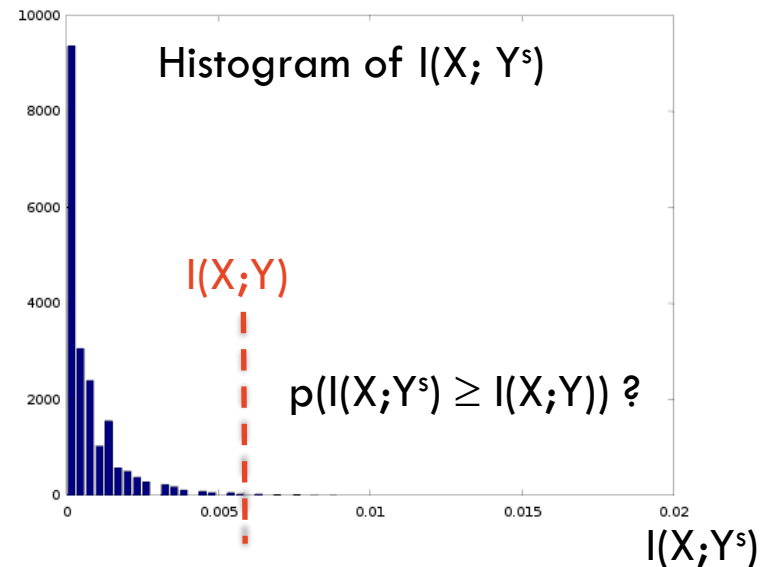
- Important to:
  - remove noise / artifacts
  - stationarise the data
- But it changes the question that you're asking, and may make investigating relationships more difficult.

## Examples:

- Brain fMRI BOLD recordings: detrending, band-pass filtering, global signal regression, deconvolution
- Equity prices in financial markets: log-differences

## Consideration 2: thresholding / statistical significance

- Stay with raw values as edge weights, or:
  - Threshold edges in network (whether remaining weighted or not):
    - By raw value (how to choose?), or
    - Computing statistical significance against a threshold  $\alpha$ 
      - E.g. p-value of a correlation (as a function of N)
  - Bias correct the statistics
- Family-wise error-rate correction
  - Bonferroni correction
    - If making  $N(N-1)$  comparisons, correct  $\alpha \rightarrow \alpha / N(N-1)$



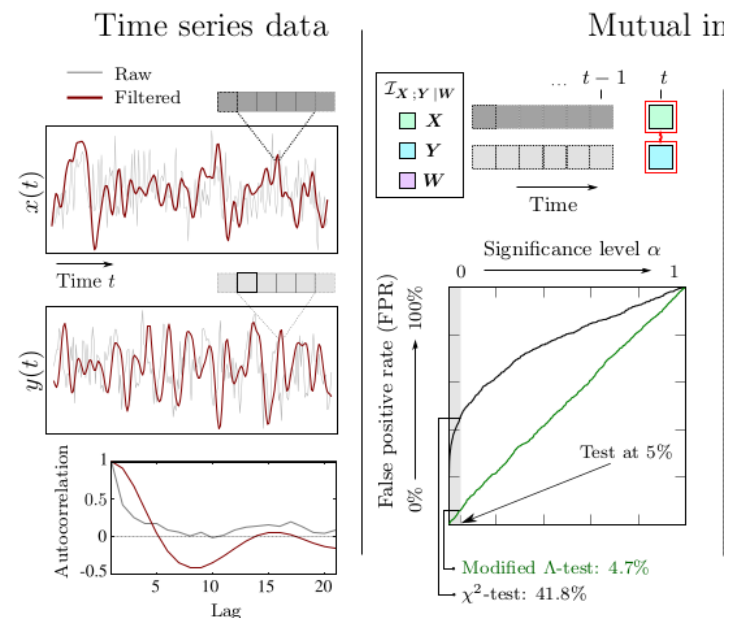
## Consideration 3: autocorrelated data

- Becomes very important for auto-correlated time-series:

- Autocorrelation renders our samples non-independent
- Which increases the variance of our estimates
- Which can make more of them appear significant
- Exacerbated by digital filtering!

- Handle via:

- Correlation: compute the effective number of samples
  - <https://github.com/olivercliff/exact-linear-dependence>
- MI (non-linear estimators): use dynamic correlation exclusion



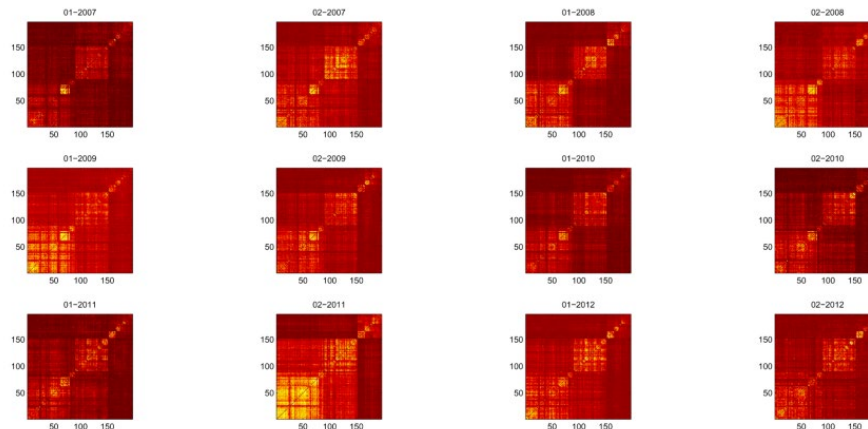
S. Afyouni, S. M. Smith, and T. E. Nichols, "Effective degrees of freedom of the Pearson's correlation coefficient under autocorrelation", *NeuroImage* 199, 609 (2019).

Cliff et al, "Assessing the significance of directed and multivariate measures of linear dependence between time series", *Physical Review Research*, 3, 013145 (2021) doi:10.1103/PhysRevResearch.3.013145

The University of Sydney

## Consideration 4: combining / splitting time series

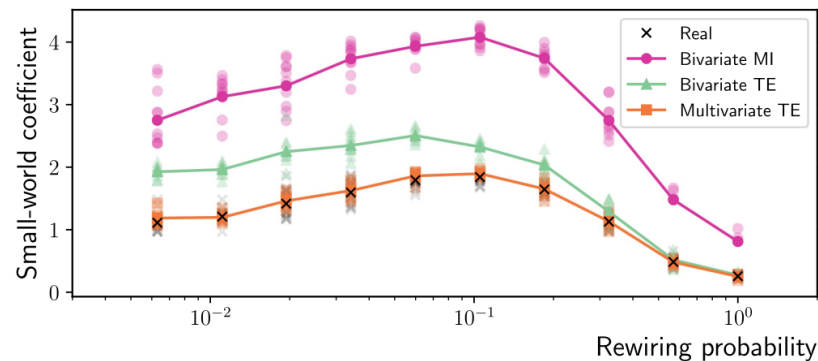
- If you have multiple time series realisations (e.g. trials for different subjects in neuroscience experiments), can/should you:
  - Combine time-series into a single analysis
    - (assumes common statistics across them)
  - Run network analysis separately and combine for group level result
- If you have long and non-stationary time series (e.g. equities over decades), can/should you:
  - Run a single analysis, or partition into different periods?



L. Sandoval, “Structure of a Global Network of Financial Companies Based on Transfer Entropy,” *Entropy*, vol. 16, no. 8, pp. 4443–4482, 2014.  
Used under CC BY 3.0

## Consideration 5: application of network measures?

- Be careful on interpreting network measures on FC networks:
  - Edges are not determined independently, and e.g. clustering is elevated
  - Tend to be a poor reflection of underlying structure and network measures on it
  - E.g. eigenvalues describe persistence of modes in how dynamics are forward propagated; what do eigenvalues of FC matrix mean?
  - Maybe consider network measures (pragmatically) to highlight different structure of statistical relationships in two conditions, but not interpret mechanistically



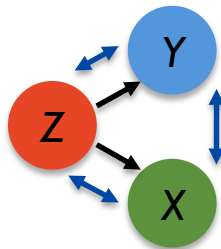
Novelli & Lizier, Inferring network properties from time series using transfer entropy and mutual information: Validation of multivariate versus bivariate approaches, *Network Neuroscience*, 5(2), 373–404. doi:10.1162/netn\_a\_00178

Langford, E., Schwertman, N., & Owens, M. (2001). Is the property of being positively correlated transitive? *The American Statistician*, 55(4), 322–325.

Zalesky, A., Fornito, A., & Bullmore, E. (2012). On the use of correlation as a measure of network connectivity. *NeuroImage*, 60(4), 2096–2106

# Functional connectivity

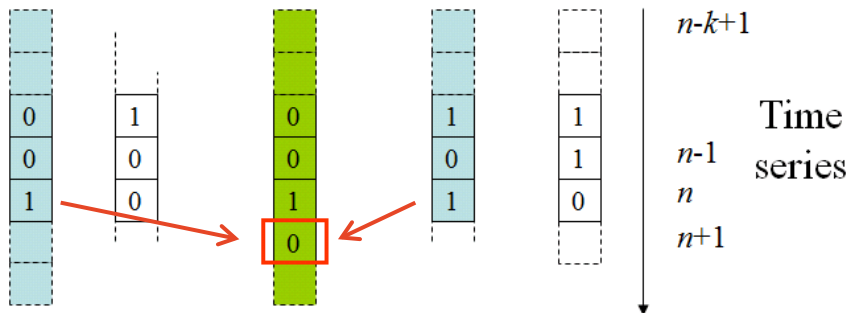
- Constructs **undirected** networks to represent **pairwise statistical relationships** between nodes
- Pros: easy to implement, run and interpret, and sensitive
- Cons:
  - Do not describe directed relationships
  - Do not describe higher-order relationships
  - Do not model/explain how dynamics are generated
  - Need to limit interpretation of network measures





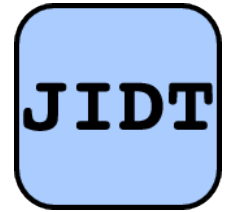
# Directed functional connectivity

- Constructs **directed** networks to represent **pairwise statistical relationships** between nodes
- Pros: easy to implement, run and interpret , and sensitive
  - Describes directed relationships

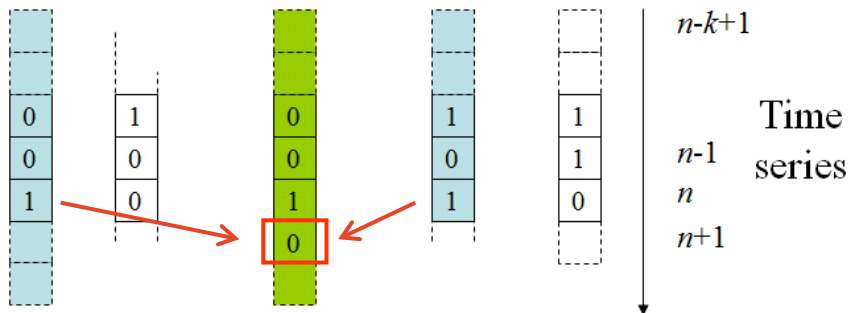


# Directed functional connectivity: measures

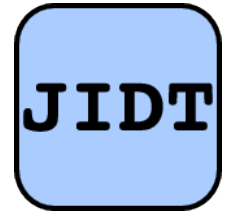
- Lagged correlation
- Lagged mutual information  $I(X(n); Y(n + 1))$



<https://github.com/jlazier/jidt>



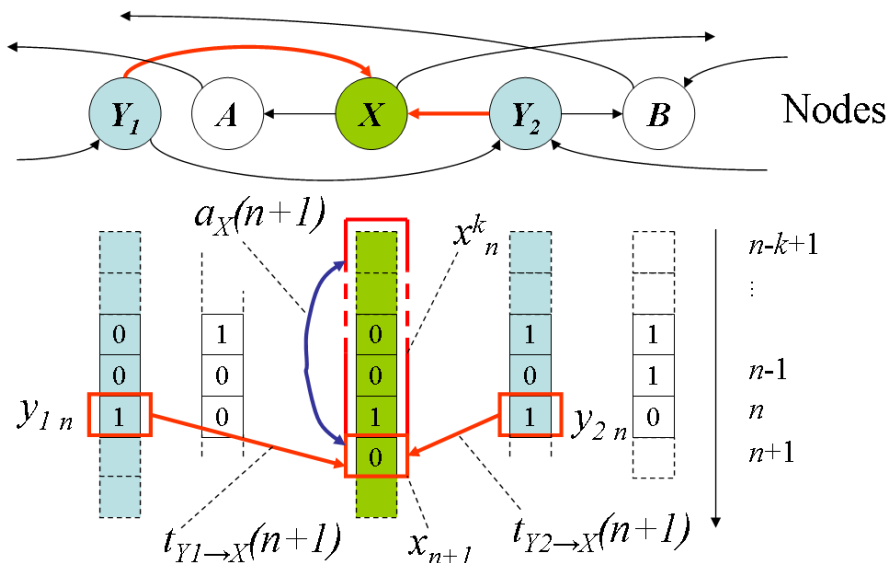
# Directed functional connectivity: measures



Dynamical systems models:

- Granger causality &
- Transfer entropy  $I(Y(n); X(n+1) | X(n)^{(k)})$

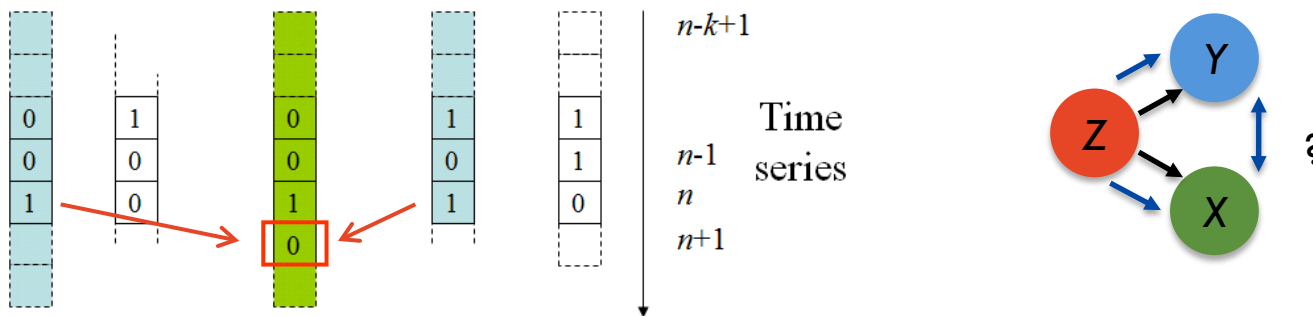
<https://github.com/jlazier/jidt>



- How much information about the next observation  $X_n$  of process  $X$  can be found in observation  $Y_n$  of process  $Y$ , in the context of the past state  $X_n^{(k)} = \{X_{n-k+1}, \dots, X_{n-1}, X_n\}$  ?
- Or, in modelling the dynamics of the target, how much information transfer does the source add to that model (after already including target past)
- Directed and dynamic

# Directed functional connectivity

- Constructs **directed** networks to represent **pairwise statistical relationships** between nodes
- Pros: easy to implement, run and interpret , and sensitive
  - Describes directed relationships
- Cons:
  - Goes part of way to explain how dynamics are generated
  - Better interpretation of network measures
  - Do not describe higher-order relationships

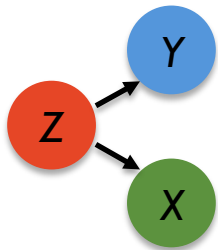


Bossomaier, Barnett, Harré, Lizier, "An Introduction to Transfer Entropy: Information Flow in Complex Systems", Springer, Cham, 2016; section 7.2  
Novelli & Lizier, *Network Neuroscience*, 5(2), 373–404. doi:10.1162/netn\_a\_00178

# Directed functional connectivity

Pairwise approaches, even dynamic ones, do not describe multivariate or higher order relationships

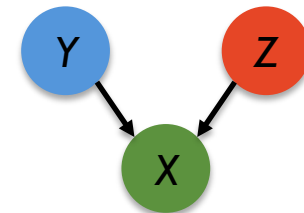
- Because they do not handle **redundancies** or **synergies** between sources.



Common driver



Pathway

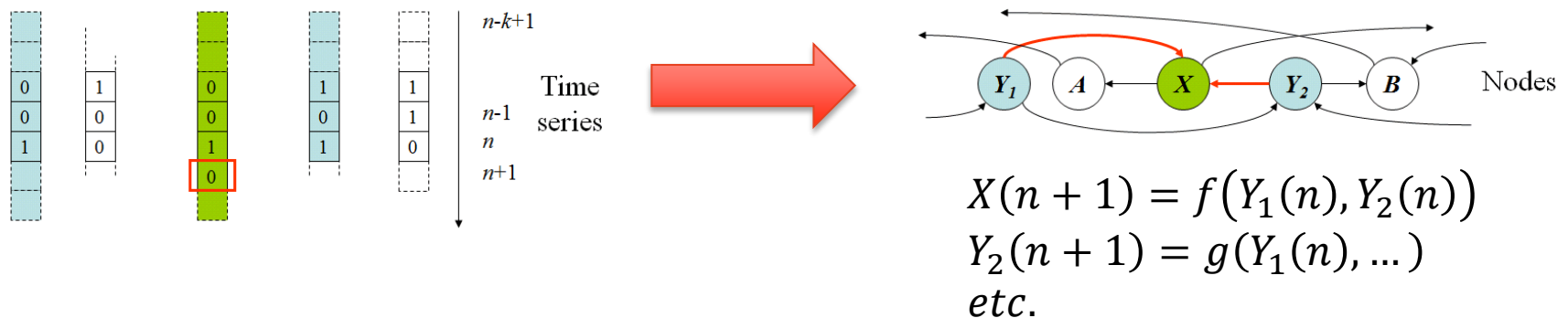


Synergy

# Effective connectivity

# Effective connectivity

- Constructs **directed** networks to provide a “*minimal neuronal circuit model*” [1] which can replicate and indeed explain the time series of the nodes
- Not strictly causal, but aims at providing best **generative model** possible from data.
- Pros:
  - Provides a (directed) generative model that **explains** the dynamics



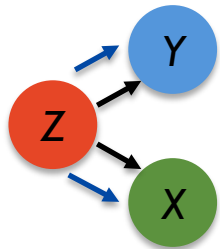
[1] K. J. Friston. Functional and effective connectivity in neuroimaging: A synthesis. *Human Brain Mapping*, 2(1-2):56–78, 1994.

[2] O. Sporns. *Networks of the Brain*. MIT Press, Cambridge, Massachusetts, USA, 2011

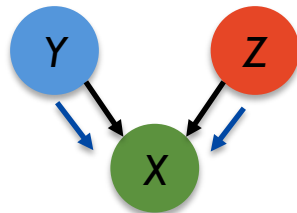
# Effective connectivity

## – Pros:

- Provides a (directed) generative model that **explains** the dynamics
- Inferring the **multivariate set** of parents that provide the best minimal model for dynamics of each target.
  - Higher order network inference!
- Incorporates multivariate effects:
  - Removes redundancies, includes synergies



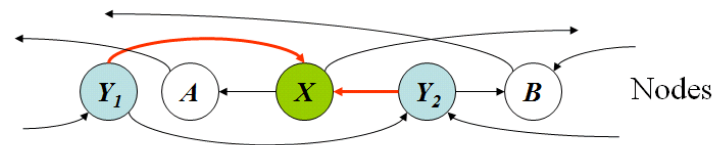
Common driver



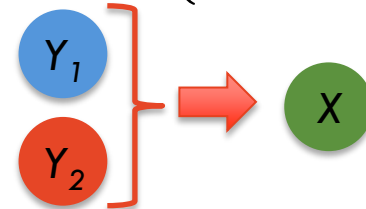
Synergy



Pathway



$$X(n + 1) = f(Y_1(n), Y_2(n))$$





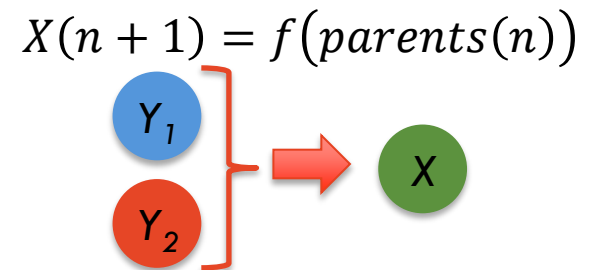
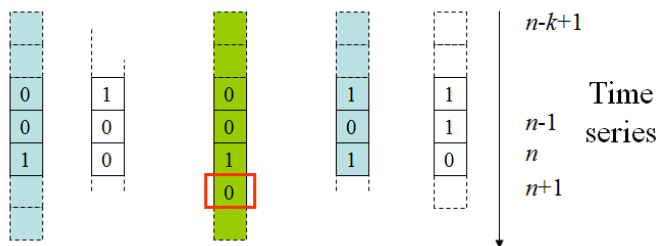
# Effective connectivity

## – Pros:

- Provides a (directed) generative model that **explains** the dynamics
- Inferring the **multivariate set** of parents that provide the best minimal model for dynamics of each target.
  - Higher order network inference!
- Incorporates multivariate effects:
  - Removes redundancies, includes synergies
- Like FC, sensitive to change in dynamics.
- In principle (*more on this later*):
  - Leads to models faithful to underlying structure \*
  - More sensible to apply network measures to EC than FC

# Effective connectivity: approaches (general)

- Inferring the **multivariate set** of parents that provide the best minimal model for dynamics of each target
- Goal, in language of *dynamic Bayesian networks*: given data  $D$ , maximise posterior probability of network model  $G$ :  $p(G | D)$
- Bayesian inversion to consider:
  - $p(G|D) \propto p(D|G)p(G)$
  - $p(G)$  Encodes any constraints / priors on the network (can make it model based)
  - Problem becomes computing likelihood of data given network model
  - For any given target  $X$ , this means maximising  $p()$  given constraints of network model  $p(G)$ , and model of dynamics  $f$  which informs how we evaluate  $p(X(n+1) | parents(n))$



O.M. Cliff et al., An Information Criterion for Inferring Coupling of Distributed Dynamical Systems. *Frontiers in Robotics & AI*, 3, 71, (2016).

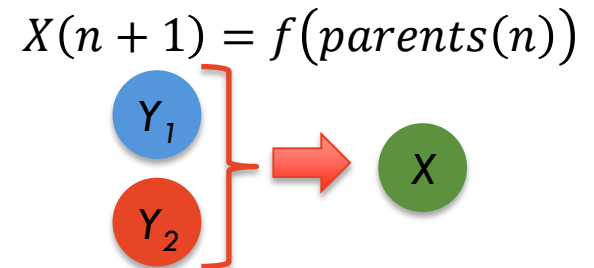
L. Barnett et al., Transfer Entropy as a Log-Likelihood Ratio. *Physical Review Letters*, 109, 138105, (2012).

# Model-based vs model-free approaches

- For any given target  $X$ , this means maximising  $p()$  given constraints of network model  $p(G)$ , and model of dynamics  $f$  which informs how we evaluate  $p(X(n+1) \mid \text{parents}(n))$

- “Model-based” may mean placing constraints on:

- Network structure, e.g.:
  - Biologically feasible brain structures
  - Stochastic block models
- Dynamics, e.g.:
  - Linear dynamics
  - Biologically realistic models



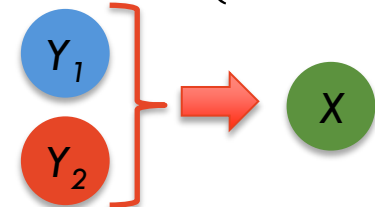
- Model-based approaches are:
  - Faster runtime (particularly linear)
  - More efficient *if* the model is correct
  - Specific to model domain

- Model-free approaches are:
  - Slower runtime
  - Less efficient (more data hungry)
  - Generalisable outside domain and applicable if model unknown

# Model likelihood

- For any given target  $X$ , this means maximising  $p()$  given constraints of network model  $p(G)$ , and model of dynamics  $f$  which informs how we evaluate  $p(X(n+1) \mid \text{parents}(n))$
- Evaluating  $p(G \mid D)$  may be via
  - log-likelihood  $\log p(G \mid D)$ , or
  - prediction error
- Returning to consideration 2 (statistical significance):
  - Score may include penalty for model complexity (number of parents):
    - explicitly (AIC/BIC style) or
    - implicitly (statistical significant test being harder to pass for multivariate models)
  - Should use family-wise error correction if selecting between many models

$$X(n+1) = f(\text{parents}(n))$$



# Approach 1: Least squares regression

- Model for dynamics: Linear vector autoregressive dynamics (VAR) with temporally uncorrelated Gaussian noise:

$$\underline{x}(n+1) = \underline{C}\underline{x}(n) + \underline{r}(n)$$

- Compute lagged covariance matrices from time-series samples:

$$\Omega_\tau = \langle \underline{x}(n+\tau)\underline{x}(n)^T \rangle$$

- Infer [1]:

$$\hat{\underline{C}} = \Omega_1 \Omega_0^{-1}$$

- Very fast; assumes linearity.
- Various extensions, e.g. for optimisation approaches for continuous time dynamics [2,3], or via SVD
- Questions: statistical significance / autocorrelation handling?

- [1] P.-Y. Lai, Reconstructing network topology and coupling strengths in directed networks of discrete-time dynamics, Physical Review E, vol. 95, no. 2, 2017.
- [2] M. Gilson et al., Estimation of Directed Effective Connectivity from fMRI Functional Connectivity Hints at Asymmetries of Cortical Connectome, PLOS CB, vol. 12, no. 3, e1004762, 2016
- [3] R. F. Galán, "On How Network Architecture Determines the Dominant Patterns of Spontaneous Neural Activity, PLOS ONE, vol. 3, no. 5, e2148, 2008

## Approach 2: Dynamic causal modelling (DCM)

- Designed for brain imaging data but general framework may be extended
- Computes log probabilities by combining:
  - Biologically realistic dynamics and mapping to imaging for  $p(D|G)$ , with
  - Biological constraints on connectivity  $p(G)$
- Model comparison focus then:
  - Compares log-likelihoods of models in an exhaustive search
    - Causes an issue for scaling number of nodes but are enhancements, e.g. [3]

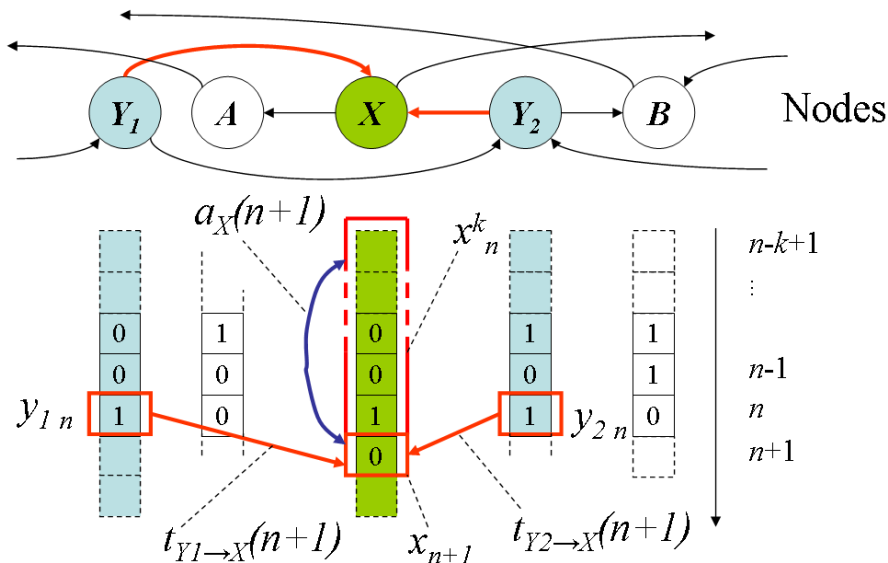
[1] K. J. Friston. Functional and effective connectivity in neuroimaging: A synthesis. *Human Brain Mapping*, 2(1-2):56–78, 1994.

[2] K. J. Friston et al., Dynamic causal modelling, *NeuroImage*, 19(4), 1273, (2003).

[3] A. Razi et al., Large-scale DCMs for resting-state fMRI, *Net. Neurosci.*, 1(3), 222, (2017)

## Approach 3: Multivariate transfer entropy

- Maximise transfer from set of sources  $I(\underline{Y}(n); X(n+1) | X(n)^{(k)})$
- Is a proxy for log probabilities of  $p(D|G)$ 
  - Equivalently Bayesian [3], but
  - Model-free in terms of dynamics and network structure
- Would still have a scaling problem if exhaustive search



- And difficult to conclude on statistical significance of difference between models.

[1] J. T. Lizier and M. Rubinov. Techn Report Preprint 25/2012, Max Planck Institute for Mathematics in the Sciences, 2012.

[2] Bossomaier, Barnett, Harré, Lizier, "An Introduction to Transfer Entropy", Springer, Cham, 2016; section 7.2

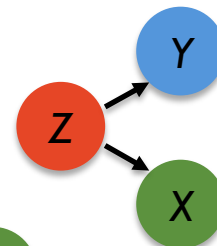
[3] O.M. Cliff et al., An Information Criterion for Inferring Coupling of Distributed Dynamical Systems. Frontiers in Robotics & AI, 3, 71, (2016).

# Conditional Transfer entropy

- $I(Y(n); X(n+1) | X(n)^{(k)}, \underline{Z}(n))$
- Building a model: Marginal benefit (log probability) of adding  $Y$  into the parent set  $\underline{Z}$ .

- Removes **redundancies** between  $Y$  and  $Z$ , e.g.

- Common driver effects

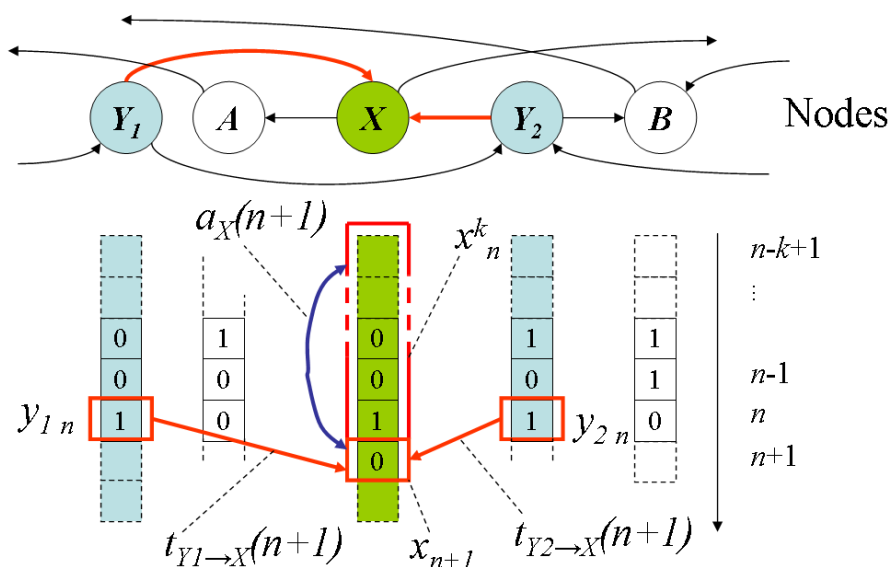
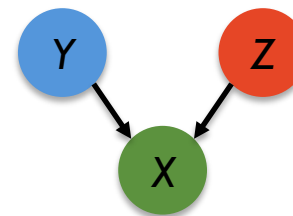


- Pathway effects



- Includes **synergies** between  $Y$  and  $Z$ , e.g.

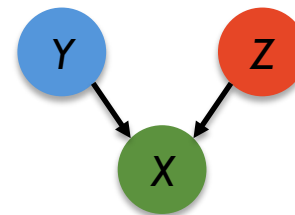
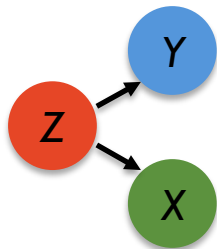
- Gated effects





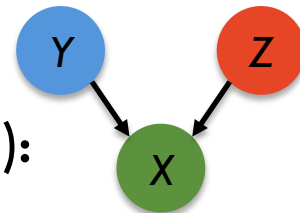
## Approach 3: Multivariate transfer entropy

- Hinted that conditional TE:
  - Computes marginal benefit of adding another parent to the model
  - Address redundancy and synergy
- But what to condition on ...?
  - If we condition on all other variables:
    - We can undersample too easily
    - We can eliminate too many links from our model
- Key: we're going to iteratively build a parent set...



# Approach 3: Multivariate information regression

- **Modelling** the information processing in  $X$ .
- Consider two sources to  $X$ . (General case in Lizier 2010):



Real goal here: to infer this **parent set**  $\{Y, Z\}$  for  $X$

$$H(X_{n+1}) = I(\mathbf{X}_n^{(k)}; X_{n+1}) + I(Y_n, Z_n; X_{n+1} | \mathbf{X}_n^{(k)}) + H(X_{n+1} | \mathbf{X}_n^{(k)}, Y_n, Z_n)$$

$$H(X_{n+1}) = I(\mathbf{X}_n^{(k)}; X_{n+1}) + I(Y_n; X_{n+1} | \mathbf{X}_n^{(k)}) + I(Z_n; X_{n+1} | \mathbf{X}_n^{(k)}, Y_n) + H(X_{n+1} | \mathbf{X}_n^{(k)}, Y_n, Z_n)$$

$$H(X_{n+1}) = I(\mathbf{X}_n^{(k)}; X_{n+1}) + I(Z_n; X_{n+1} | \mathbf{X}_n^{(k)}) + I(Y_n; X_{n+1} | \mathbf{X}_n^{(k)}, Z_n) + H(X_{n+1} | \mathbf{X}_n^{(k)}, Y_n, Z_n)$$

1. Active information storage

2-. Multivariate transfer entropy

2. Pairwise/apparent transfer entropy

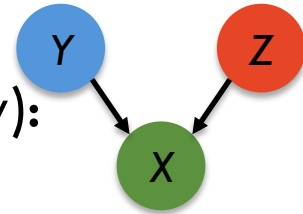
3+. Conditional transfer entropy

J. T. Lizier, M. Prokopenko, & A. Y. Zomaya. "Local information transfer as a spatiotemporal filter for complex systems". Physical Review E, 77(2):026110, 2008.

J. T. Lizier, M. Prokopenko, & A. Y. Zomaya. "Information modification and particle collisions in distributed computation", Chaos, 20(3), 037109, 2010.

## Approach 3: Iterative/greedy information regression

- Maths for two sources to  $X$ . (General case in refs below):



Real goal: to infer this **parent set**  $\{Y, Z\}$  for  $X$

$$H(X_{n+1}) = I(X_n^{(k)}; X_{n+1}) + I(Y_n, Z_n; X_{n+1} | X_n^{(k)}) + H(X_{n+1} | X_n^{(k)}, Y_n, Z_n)$$

- Inferring whole parent set at once is combinatorially difficult.
- Instead, infer parent set one by one in a **greedy** fashion:
  - ➔ 1. Start with empty parent set  $P$
  - ➔ 2. Evaluate TE from each source to target, **conditioning on  $P$** .
  - ➔ 3. Add the source with max conditional TE to  $P$  if **p-value is statistically significant**.
  - ➔ 4. Go back to step 2 if a new parent was added, else terminate.

$$H(X_{n+1}) = I(X_n^{(k)}; X_{n+1}) + I(Y_n; X_{n+1} | X_n^{(k)}) + I(Z_n; X_{n+1} | X_n^{(k)}, Y_n) + H(X_{n+1} | X_n^{(k)}, Y_n, Z_n)$$

$$H(X_{n+1}) = I(X_n^{(k)}; X_{n+1}) + I(Z_n; X_{n+1} | X_n^{(k)}) + H(X_{n+1} | X_n^{(k)}, Z_n)$$

**Synthesis:** L. Novelli et al, "Large-scale directed network inference with multivariate transfer entropy ...," *Network Neuroscience*, 3(3), 827–847, 2019.

J. T. Lizier and M. Rubinov. Technical Report Preprint 25/2012, Max Planck Institute for Mathematics in the Sciences, 2012.

L. Faes, G. Nollo, and A. Porta, "Information-based detection of nonlinear Granger causality ...," *Physical Review E*, 83, 051112, 2011

J. Sun, D. Taylor, E.M. Bollt, Causal network inference by optimal causation entropy. *SIAM Journal on Applied Dynamical Systems*, 14(1), 73–106, 2015

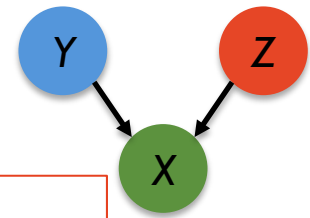
I. Vlachos and D. Kugiumtzis, "Nonuniform state-space reconstruction and coupling detection", *Physical Review E*, 82, 016207, 2010

J. Runge, et al, "Escaping the curse of dimensionality in estimating multivariate transfer entropy. *Physical Review Letters*, 108(25), 258701, 2012

A. Montalto, L. Faes, D. Marinazzo, MuTE: A MATLAB toolbox to compare established and novel estimators of the multivariate TE. *PLoS ONE*, 9(10), e109462, 2014

T. Bossomaier, L. Barnett, M. Harré, J. Lizier, "An Introduction to Transfer Entropy: Information Flow in Complex Systems", Springer, Cham, 2016; section 7.2

## Approach 3: Multivariate transfer entropy



- Instead, infer parents one by one in a **greedy** fashion:
  0. Embed target past
  1. Start with empty parent set  $P$
  2. Evaluate TE from each source to target, conditioning on  $P$ .
  3. Add the source with max cond. TE to  $P$  if p-value is statistically significant.
  4. Go back to step 2 if a new parent was added, else go to step 5.
  5. Prune redundant links in context of final set  $P$ .
  6. Perform statistical test of whole parent set  $P$ .

$$H(X_{n+1}) = I\left(\mathbf{X}_n^{(k)}; X_{n+1}\right) + I\left(Y_n, Z_n; X_{n+1} \mid \mathbf{X}_n^{(k)}\right) + H\left(X_{n+1} \mid \mathbf{X}_n^{(k)}, Y_n, Z_n\right)$$

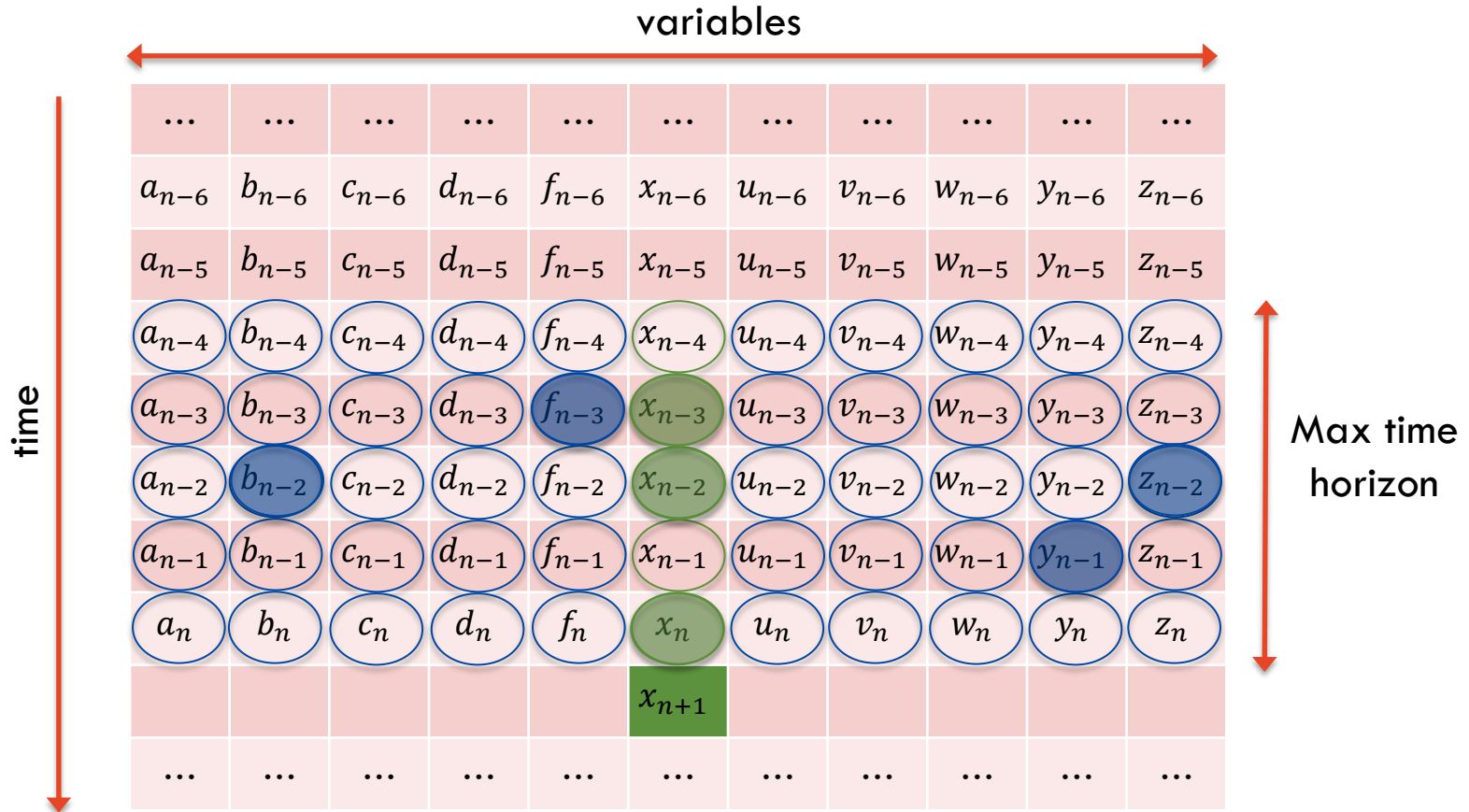
- More **efficient** than brute force search for parent set.
- Handles **redundancies** and **synergies** between parents.
- Could scan for multiple samples from each source (**non-uniform embedding**)
- Statistical tests provide “**automatic brake**” when statistical power of data is exhausted.
- Can add **additional tests**.
- Must control for **multiple comparisons** (e.g. max statistics)
- End result is the **parent set**. Order that nodes were inferred in is no longer relevant.

L. Novelli et al, "Large-scale directed network inference with multivariate transfer entropy and hierarchical statistical testing," *Network Neuroscience*, 3(3), 827–847, 2019.

P. Wollstadt, J.T. Lizier, R. Vicente, C. Finn, M. Martinez-Zarzuela, P. Mediano, L. Novelli and M. Wibral, "IDTxL: The Information Dynamics Toolkit xL: a Python package for the efficient analysis of multivariate information dynamics in networks", *Journal of Open Source Software*, 4(34), 1081, 2019

# Iterative/greedy approaches

- Instead, infer parents one by one in a **greedy** fashion:
  0. Embed target past
  1. Start with empty parent set  $P$
  2. Evaluate TE from each source to target, conditioning on  $P$ .
  3. Add the source with max cond. TE to  $P$  if p-value is statistically significant.
  4. Go back to step 2 if a new parent was added, else go to step 5.
  5. Prune redundant links in context of final set  $P$ .
  6. Perform statistical test of whole parent set  $P$ .



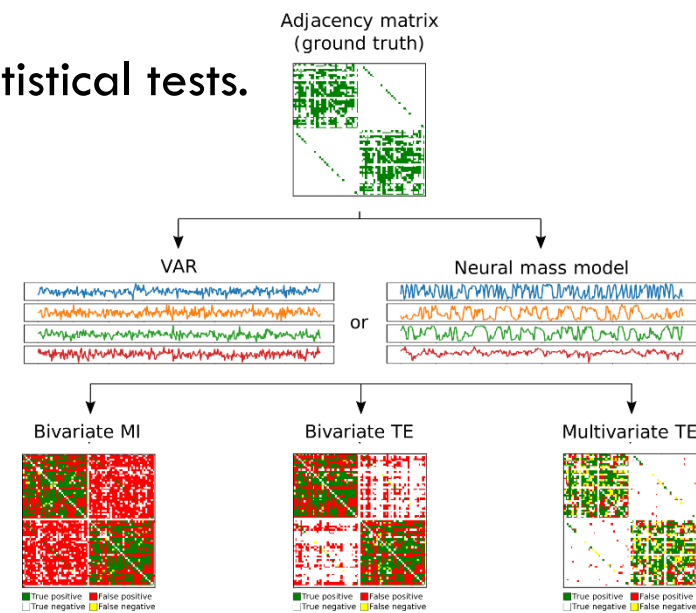
○ Target embedding candidates  
 ● Target embedding selections

○ Source candidates  
 ● Source selections (parent set  $P$ )

# Using iterative/greedy approaches

<https://github.com/pwollstadt/IDTxI>

- **IDTxI** (which uses JIDT as an internal information-theoretic engine) implements the greedy algorithm, including:
  - “Max statistics” for multiple comparison correction in parent selection;
  - Handles TE parameter selection, non-uniform embedding / delay selection of sources;
  - Adds additional steps (pruning step) and statistical tests.
- Validation studies confirm convergence to structure under idealised conditions



P. Wollstadt, et al, "IDTxI: The Information Dynamics Toolkit xl...", *Journal of Open Source Software*, 4(34), 1081, 2019

L. Novelli et al, "Large-scale directed network inference with multivariate transfer ...," *Network Neuroscience*, 3(3), 827–847, 2019.

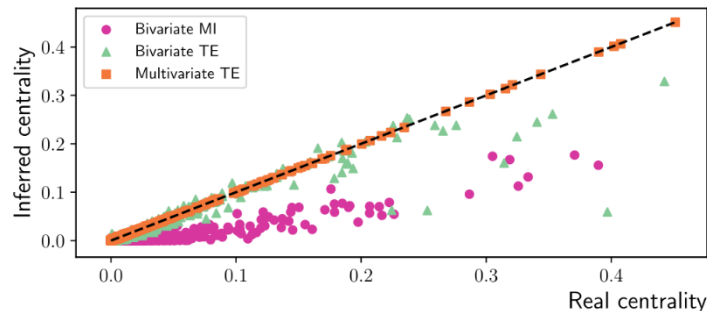
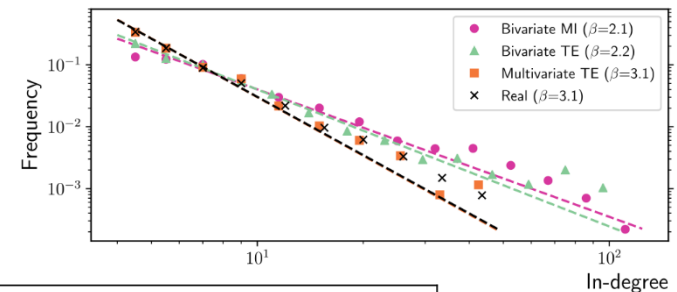
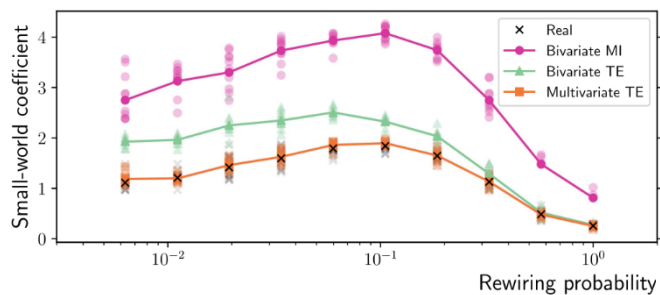
L. Novelli and J. T. Lizier, "Inferring network properties from time series using TE and MI: ...," *Network Neuroscience*, 5(2), 373–404, 2021

D. Shortem et al, "Inferring effective networks of spiking neurons using a continuous-time estimator of transfer entropy", bioRxiv:2024.09.22.614302, 2024

# Effective connectivity

## Pros (continuing from earlier)

- Leads to models faithful to underlying structure:
  - Without hidden nodes
  - With exploration of whole state space
- More sensible to apply network measures to EC than FC



Novelli & Lizier, Inferring network properties from time series using transfer entropy and mutual information: Validation of multivariate versus bivariate approaches, *Network Neuroscience*, 5(2), 373–404. doi:10.1162/netn\_a\_00178

# Effective connectivity

## Cons:

- Data requirements
  - Pairwise methods may be more useful in low data situations?
- Runtime and scaling
- Convergence to underlying structure? (*do we really want this anyway?*)
  - Depends on hidden nodes
  - Data exploring full space?



# Network inference from time-series data: summary

1. Understand philosophy of different goals in network inference from time-series data:
  - Functional, effective and structural connectivity
2. Understand measures and approaches (and how to access them)
  - Model-based and model-free measures
  - Linear and information-theoretic families
  - Pairwise and multivariate (and algorithms serving the latter)
3. Understand key considerations to incorporate
  1. Data pre-processing
  2. Thresholding / statistical significance
  3. Handling autocorrelation
  4. Combining / splitting realisations
  5. Applying network measures

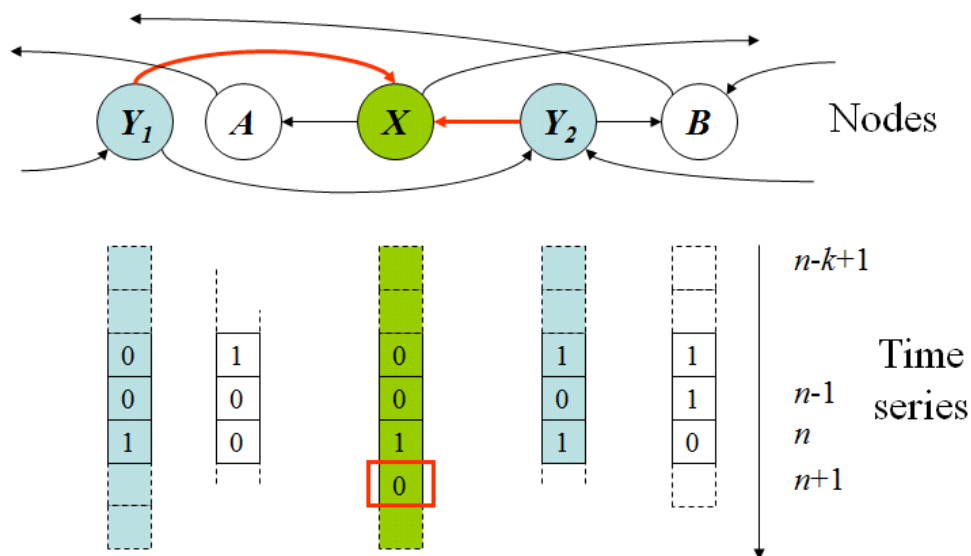
# Questions



THE UNIVERSITY OF  
SYDNEY

# Effective network inference

- Hybrid approach between structural and functional
- Seeks to infer **directed** relationships and “*minimal neuronal circuit model*” which can replicate and indeed explain the time series of the nodes



Not structural (causal) but:

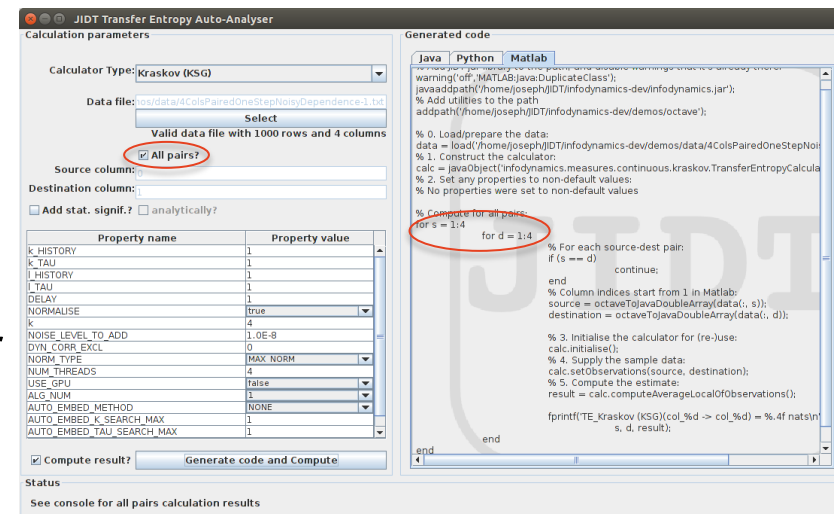
- May be best that can be done with data. (Causality often impossible)
- Reflect dynamic **changes in regime** of the system.
- **Model** the **computation** taking place in the system, revealing **emergent computational structure**.

# TE for directed functional connectivity

1. Measure **pairwise TE** between all pairs of variables in the system;
2. **Threshold** the TE values to select connections for the network

Let's try this in JIDT

- Start TE AutoAnalyser
- Tick “all pairs” checkbox
  - Generates loops for computing TE for all destinations, for all sources
- What is the first issue we encounter?
  - What would we choose for the threshold?



# TE for directed functional connectivity

1. Measure **pairwise TE** between all pairs of variables in the system;
  2. For each source-target pair, obtain the **p-value** for measuring the observed TE under the null distribution; (see TE lecture)
  3. **Threshold the p-values** to select connections for the network.
- More principled approach: threshold is statistically derived, robust, suitable for small data sets.

Let's try this in JIDT

- Start TE AutoAnalyser
- Tick “all pairs” checkbox and “add stat. signif?”
  - Generates loops for computing TE for all destinations, for all sources
  - Provides p-values that can be tested more rigorously

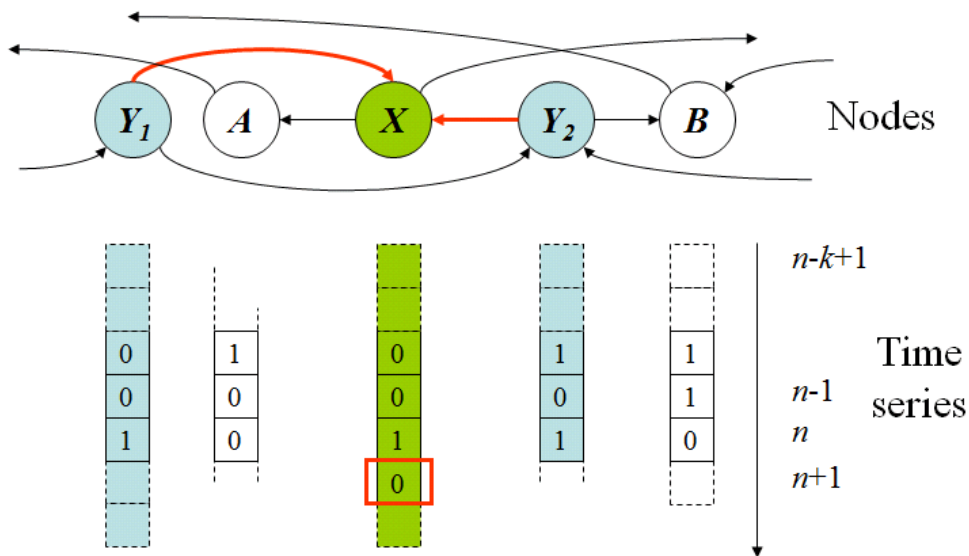


M. Wibral, R. Vicente, and M. Lindner. “Transfer entropy in neuroscience”. In M. Wibral, R. Vicente, and J. T. Lizier, editors, *Directed Information Measures in Neuroscience*, pp. 3–36. Springer, Berlin/Heidelberg, 2014.

Bossomaier, Barnett, Harré, Lizier, “An Introduction to Transfer Entropy: Information Flow in Complex Systems”, Springer, Cham, 2016; section 7.2

# Effective connectivity

- Hybrid approach between structural and functional
- Seeks to infer **directed** relationships and “*minimal neuronal circuit model*” which can replicate and indeed explain the time series of the nodes



Not structural (causal) but:

- May be best that can be done with data. (Causality often impossible)
- Reflect dynamic **changes in regime** of the system.
- **Model** the **computation** taking place in the system, revealing **emergent structure**.

**Transfer entropy is a natural fit!**

O. Sporns. *Networks of the Brain*. MIT Press, Cambridge, Massachusetts, USA, 2011

K. J. Friston. Functional and effective connectivity in neuroimaging: A synthesis. *Human Brain Mapping*, 2(1-2):56–78, 1994.

Bossomaier, Barnett, Harré, Lizier, "An Introduction to Transfer Entropy: Information Flow in Complex Systems", Springer, Cham, 2016; section 7.2  
The University of Sydney