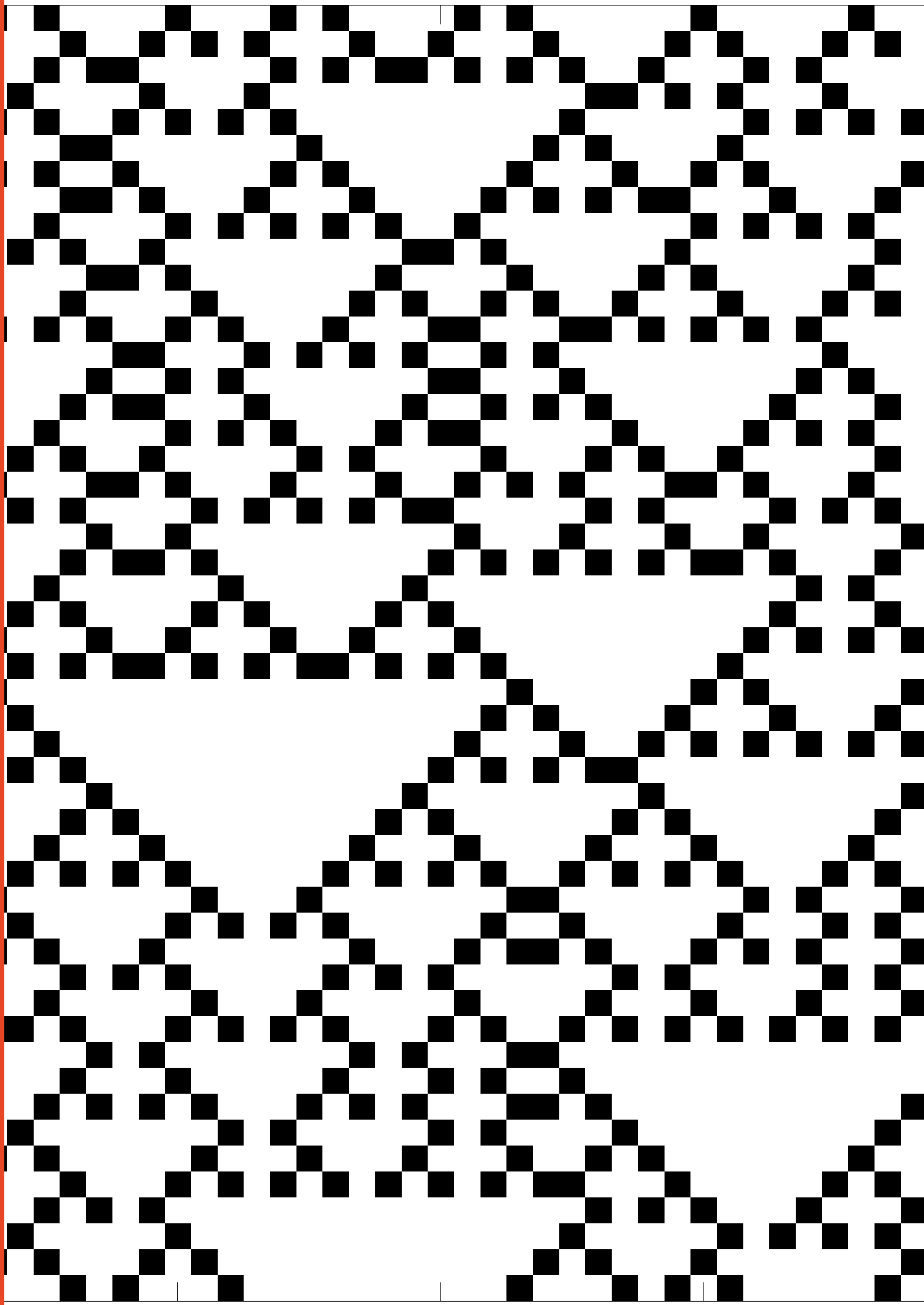
Effective network inference

Dr. Joseph Lizier

THE UNIVERSITY OF SYDNEY



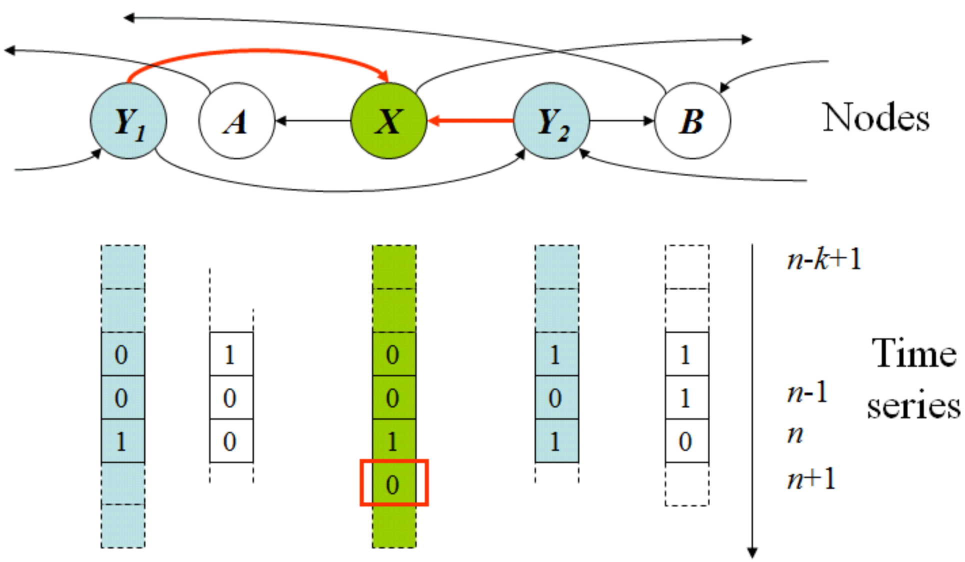
Effective network inference: session outcomes

- Understand different options for network inference from time-series data, as well as complexities and subtleties involved in effective network inference.
- Able to use JIDT for simple effective network inference (pairwise)
- Understand advantages of more advanced techniques, e.g. those implemented in the IDTxl toolkit

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

Network inference

– 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?*



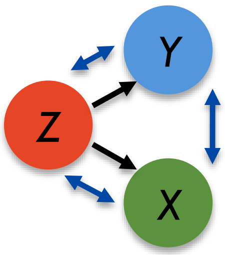
Complex system as a multivariate time-series of states

Options:

1. Functional networks
2. Effective networks
3. Structural networks

Functional network inference

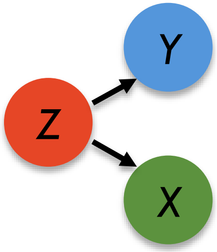
- Constructs undirected networks to represent relationships between nodes
- Usually using a measure of correlation or MI
- Provides no explanation for how the relationship manifests



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.

Structural network inference

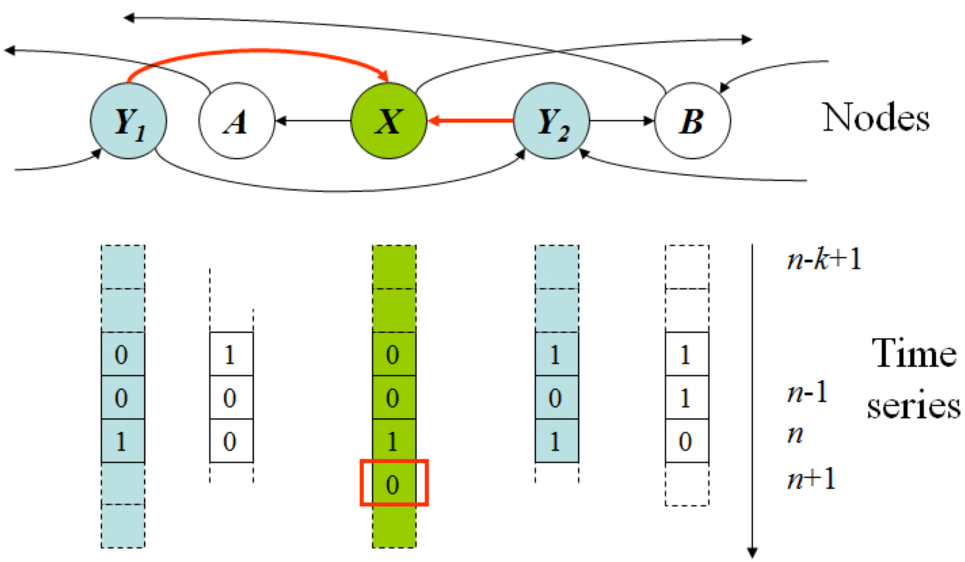
- Constructs directed networks to represent the physical, directed (causal) connections
- *Generally* only possible via interventional techniques but not directly from large (observational) multivariate time-series sets alone.
- Does not tell us about time or experimentally modulated changes in how the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



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

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:

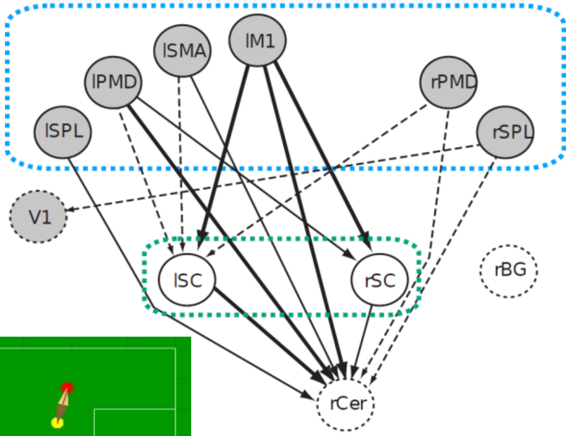
- May be best that can be done with data. (Causality often impossible)
- Reflect dynamic changes in regime of the system.
- **Model** the computation taking place in the system, revealing emergent computational structure.

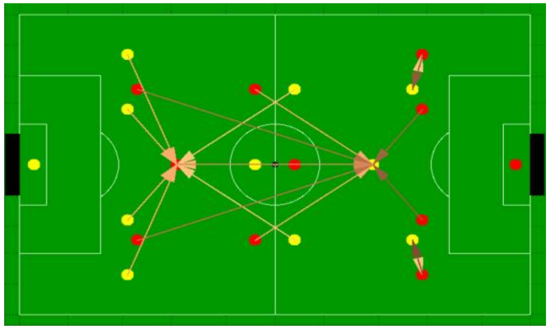
O. Sporns. *Networks of the Brain.* MIT Press, Cambridge, Massachusetts, USA, 2011
K. J. Friston. Functional and effective connectivity in neuroimaging: A synthesis. *Human Brain Mapping*, 2(1-2):56–78, 1994.
Bossomaier, Barnett, Harré, Lizier, "An Introduction to Transfer Entropy: Information Flow in Complex Systems", Springer, Cham, 2016; section 7.2

What can we use effective network inference for?

- Model/infer underlying information network for the system
- Examine how network changes with condition
- Examine how network changes over time

- Application areas:
  - Neuroimaging
  - Financial market data sets
  - Gene regulatory networks
  - Social media analysis
  - Sport analytics
  - …

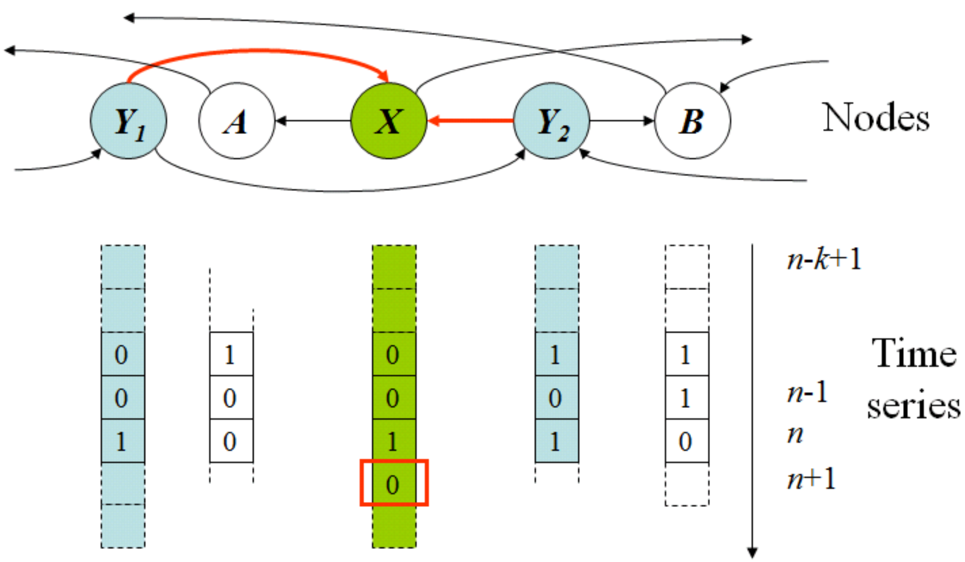




J. T. Lizier, J. Heinzle, A. Horstmann, J.-D. Haynes, and M. Prokopenko.. *Journal of Computational Neuroscience*, 30(1):85–107, 2011
O. M. Cliff, J. T. Lizier, X. R. Wang, P. Wang, O. Obst, and M. Prokopenko. In S. Behnke, M. Veloso, A. Visser, and R. Xiong, editors, *RoboCup 2013: Robot World Cup XVII*, volume 8371 of *Lecture Notes in Computer Science*, pages 1–12. Springer, Berlin/Heidelberg, 2014

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 structure.

Transfer entropy is a natural fit!

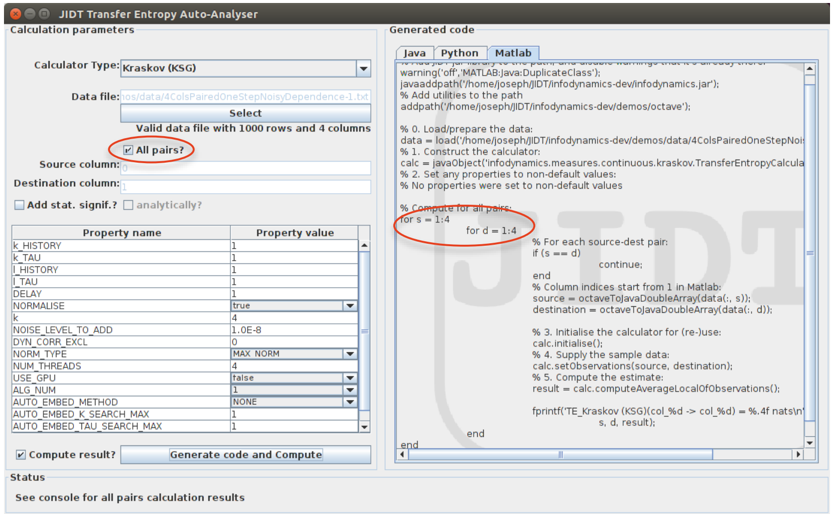
O. Sporns. *Networks of the Brain.* MIT Press, Cambridge, Massachusetts, USA, 2011
K. J. Friston. Functional and effective connectivity in neuroimaging: A synthesis. *Human Brain Mapping*, 2(1-2):56–78, 1994.
Bossomaier, Barnett, Harré, Lizier, "An Introduction to Transfer Entropy: Information Flow in Complex Systems", Springer, Cham, 2016; section 7.2

TE for effective network inference: basic approach

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?

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

TE for effective network inference: standard approach

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

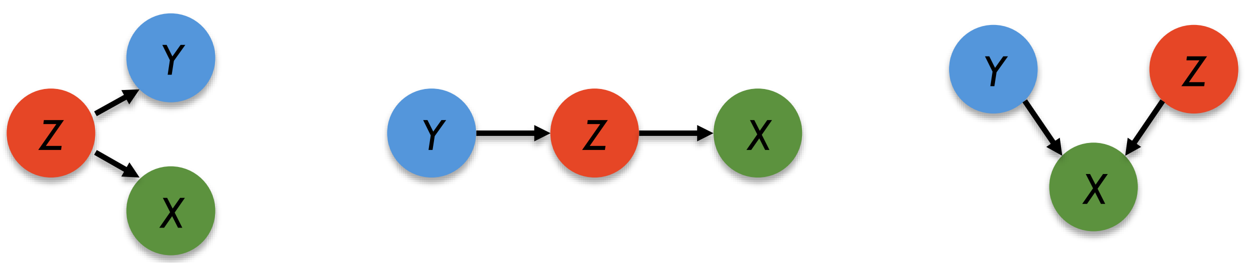
– Start TE AutoAnalyser
– Tick "all pairs" checkbox and "add stat. signif?"
  – Generates loops for computing TE for all destinations, for all sources
  – Provides p-values that can be tested more rigorously

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

Notes on standard approach

1. Need to correct for multiple comparisons using either:
   - family-wise error rates
     - e.g. Bonferroni correction → drop threshold α by a factor of M, where M is the number of tests we make. Normally M=G.(G-1) for network of size G, which gives a very small threshold!
   - false discovery rates.
2. Need to properly set TE parameters, else can get false positives/negatives:
   - Set embeddings;
   - Set source-target delay – e.g. scan over a range
3. Does not handle redundancies or synergies between sources.



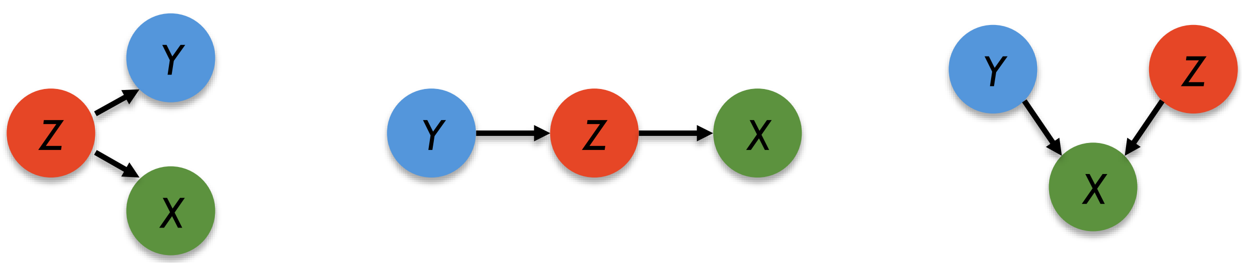
M. Wibral, R. Vicente, and M. Lindner. "*Transfer entropy in neuroscience*". In M. Wibral, R. Vicente, and J. T. Lizier, editors, *Directed Information Measures in Neuroscience*, pp. 3–36. Springer, Berlin/Heidelberg, 2014.

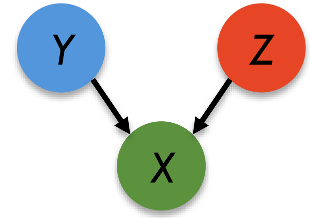
Bossomaier, Barnett, Harré, Lizier, "An Introduction to Transfer Entropy: Information Flow in Complex Systems", Springer, Cham, 2016; section 7.2

Multivariate effective network inference with TE

- What could we do to address redundancy and synergy?
- Recall that conditional TE does this

- 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





Recall: Information regression/model

– 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

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

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

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

1. Active information storage

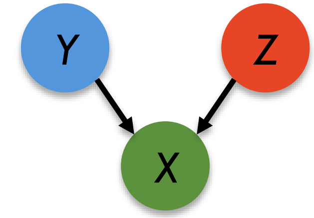
2-. Collective transfer entropy

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.

Iterative/greedy approaches



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

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

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

– Inferring whole parent set at once is combinatorially difficult.

– Instead, infer parent set one by one in a greedy fashion:

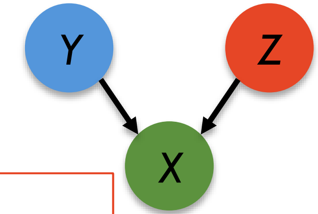
1. Start with empty parent set *P*
2. Evaluate TE from each source to target, conditioning on *P*.
3. Add the source with max conditional TE to *P* if p-value is statistically significant.
4. Go back to step 2 if a new parent was added, else terminate.

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

$$H(X_{n+1}) = I\left(\boldsymbol{X}_n^{(k)}; X_{n+1}\right) + I\left(Z_n; X_{n+1} \middle| \boldsymbol{X}_n^{(k)}\right) + H\left(X_{n+1} \middle| \boldsymbol{X}_n^{(k)}, Z_n\right)$$

J. T. Lizier and M. Rubinov. "*Multivariate construction of effective computational networks from observational data*". Technical Report Preprint 25/2012, Max Planck Institute for Mathematics in the Sciences, 2012.

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



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*.

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

- 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.

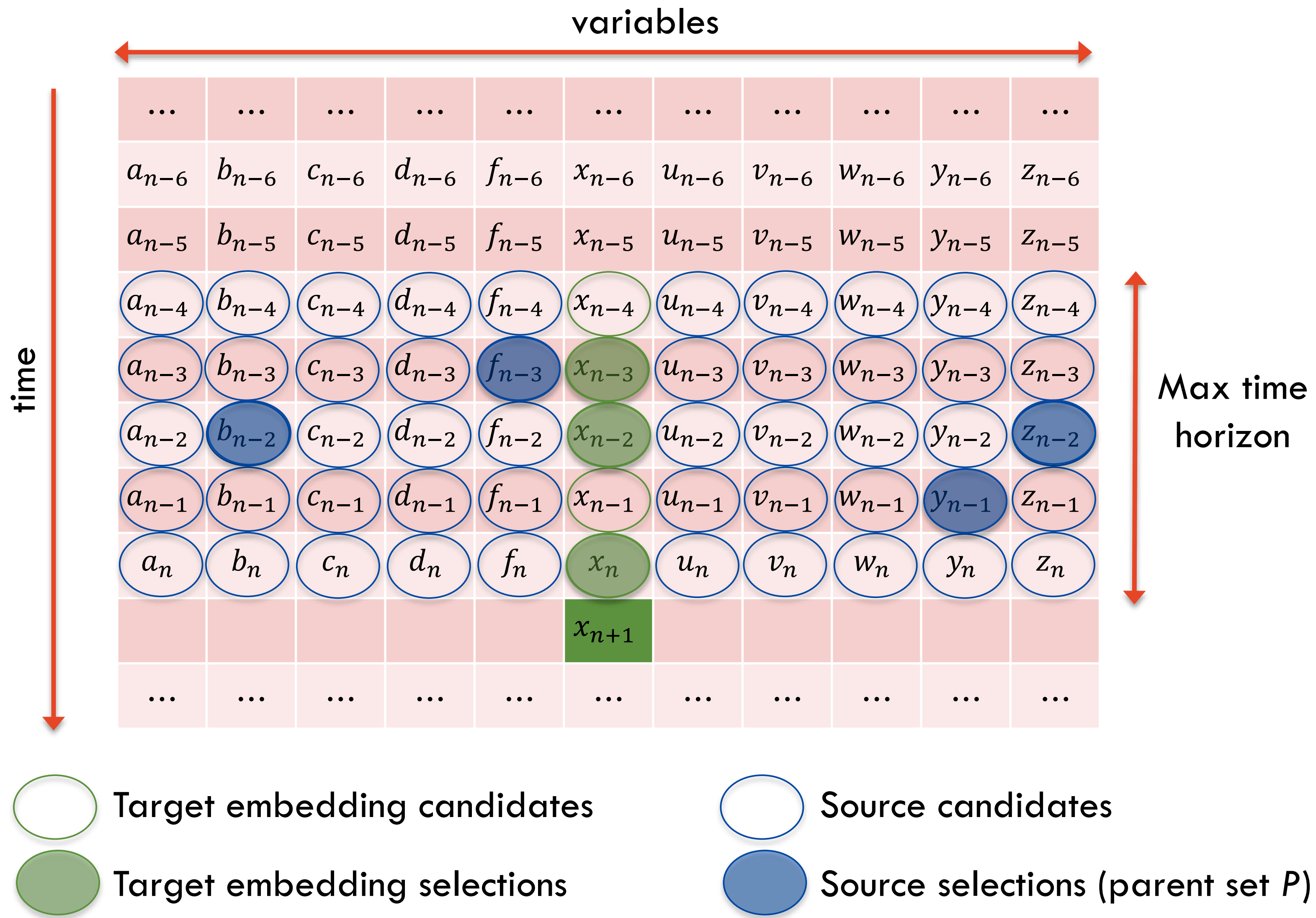
J. T. Lizier and M. Rubinov. “*Multivariate construction of effective computational networks from observational data*”. Technical Report Preprint 25/2012, Max Planck Institute for Mathematics in the Sciences, 2012.

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

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", arXiv:1807.10459, 2018

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*.



Using iterative/greedy approaches

- IDTxl (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.

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", arXiv:1807.10459, 2018
https://github.com/pwollstadt/IDTxl

Effective network inference: summary

- We’ve looked at different options for network inference from time-series data, complexities and subtleties involved in effective network inference.
- Able to use JIDT for simple effective network inference (pairwise)
- Understand advantages of more advanced techniques in the IDTxl toolkit

- *Coming up*: Summary

Questions