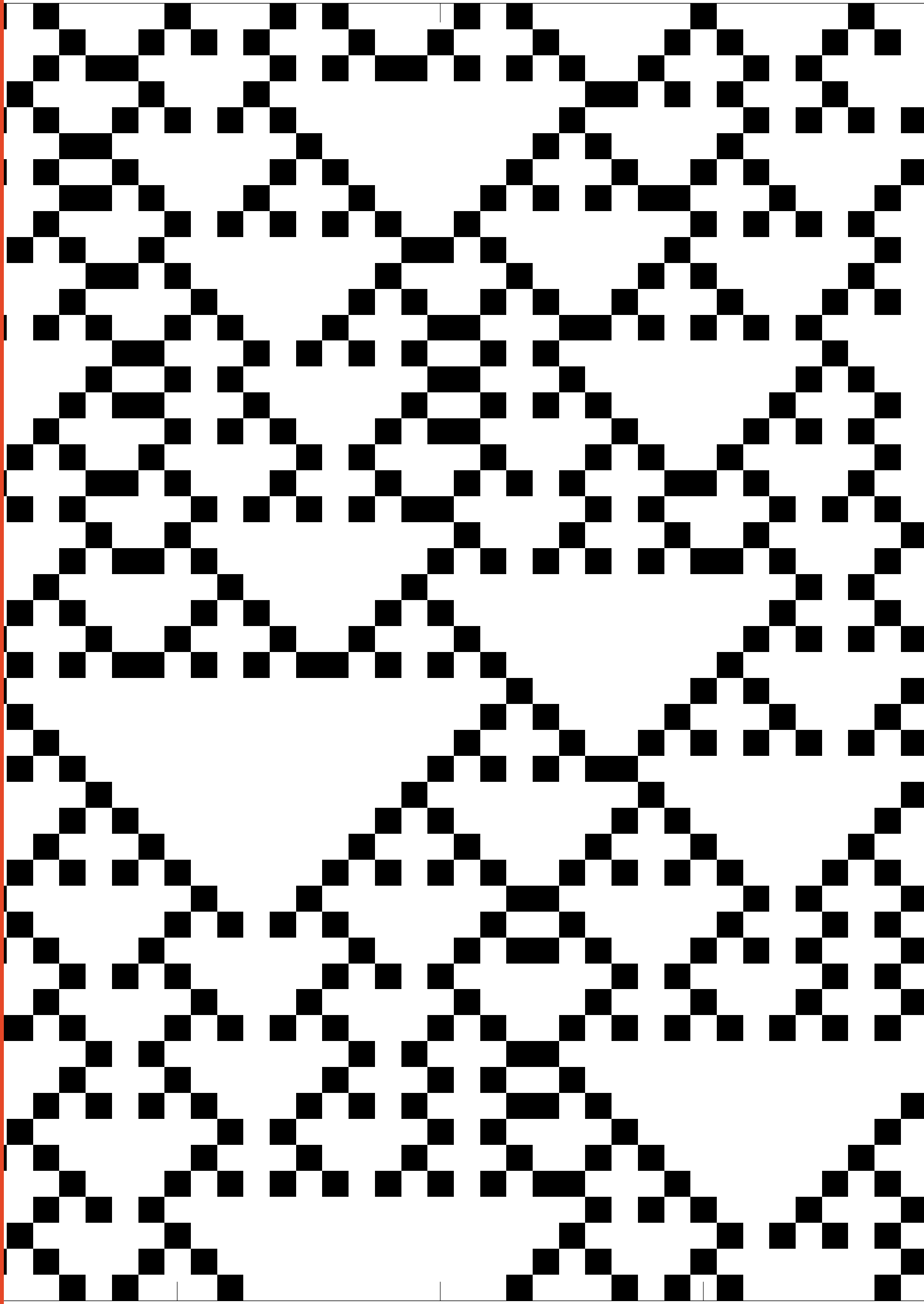
Information dynamics – Part III – Information transfer



Dr. Joseph Lizier

THE UNIVERSITY OF SYDNEY

Information dynamics Part III: session outcomes

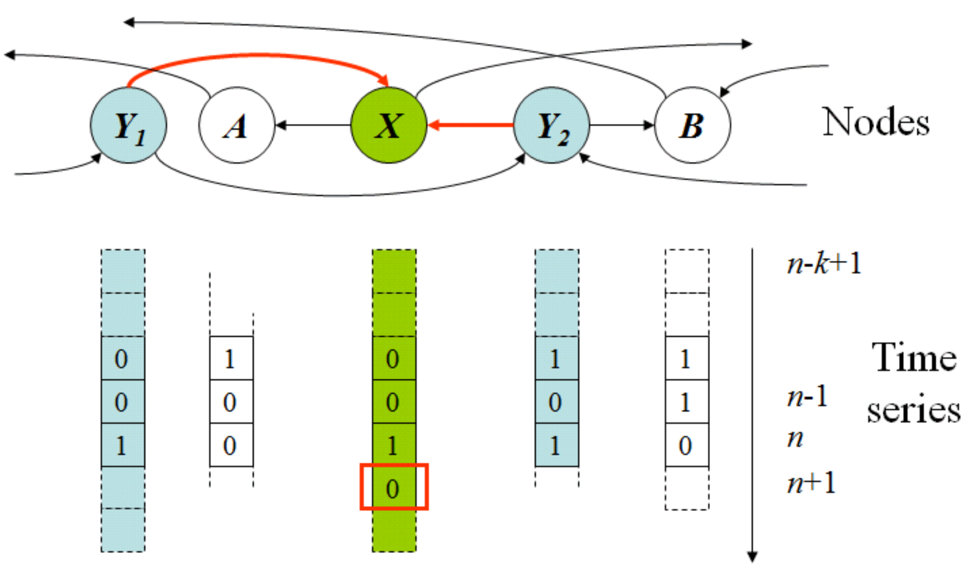
- Understand measures for **information transfer**, and philosophy of how the concept relates to information storage.
  - Transfer entropy as a model for information transfer in a wider perspective of distributed intrinsic computation
- Apply JIDT using AutoAnalyser and extensions of code it produces to analyse information transfer in complex systems data sets.
- Able to apply JIDT to make parameter selections for TE, and evaluate different choices/approaches to analysis

- Primary references (for Info Dynamics II and III sessions):

  - J.T. Lizier, "JIDT: An information-theoretic toolkit for studying the dynamics of complex systems", Frontiers in Robotics and AI, 1:11, 2014; appendix A.2 and A.3
  - J.T. Lizier, "*The local information dynamics of distributed computation in complex systems*", Springer: Berlin/Heidelberg, 2013; chapter 3, 4
  - Bossomaier, Barnett, Harré, Lizier, "An Introduction to Transfer Entropy: Information Flow in Complex Systems", Springer, Cham, 2016; chapter 4 (sections 4.1-4.3); section 5.1

Information dynamics

- Key question: how is the next state of a variable in a complex system computed?
- It is the output of a local computation within the system



Complex system as a multivariate time-series of states

Q: Where does the information in x_{n+1} come from, and how can we measure it?
(Where might we look?)

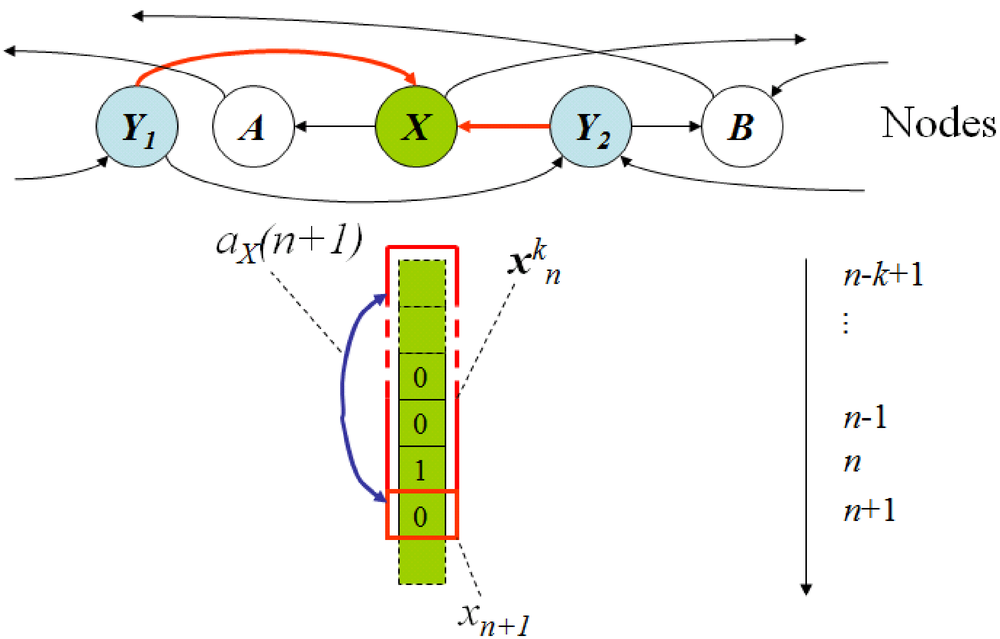
Q: Can we model the information processing in X in terms of:

- how much information was stored?
- how much was transferred, and how was this attributed?

Q: Can we partition them, do they overlap? etc.

Active information storage

– How much information about the next observation X_n of process X can be found in its past state $\boldsymbol{X}_n^{(k)} = \{X_{n-k+1}, \ldots, X_{n-1}, X_n\}$?



Active information storage

$$A_X = \lim_{k \to \infty} I(\boldsymbol{X}_n^{(k)}; X_{n+1})$$

$$A_X(k) = I(\boldsymbol{X}_n^{(k)}; X_{n+1})$$

$$A_X = H(X_{n+1}) - H_{\mu X}$$

$$A_X(k) = \left\langle \log_2 \frac{p(x_{n+1} | \boldsymbol{x}_n^{(k)})}{p(x_{n+1})} \right\rangle$$

$$a_X(n+1, k) = \log_2 \frac{p(x_{n+1} | \boldsymbol{x}_n^{(k)})}{p(x_{n+1})}$$

– AIS: Average information from past state that is in use in predicting the next value

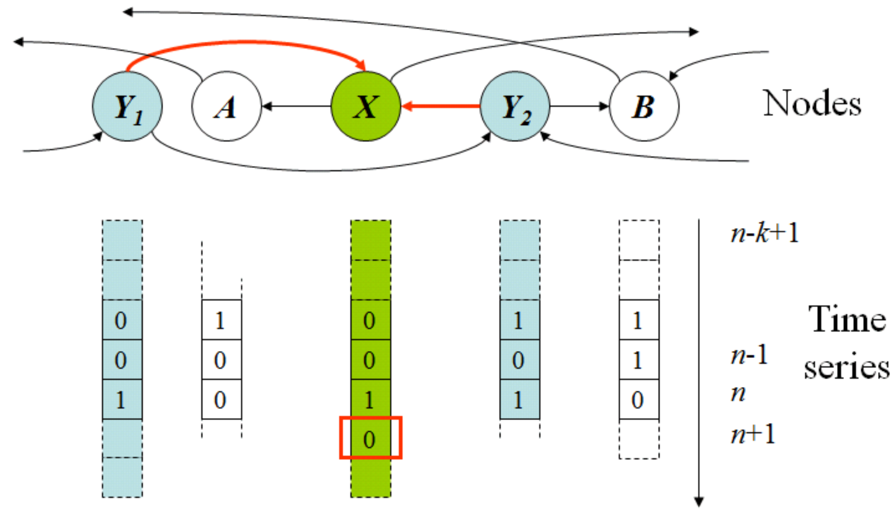
– Local AIS: information from a specific past state in use in predicting specific next value

J.T. Lizier, M. Prokopenko, & A.Y. Zomaya, “Detecting Non-trivial Computation in Complex Dynamics ”, Proc. ECAL, pp. 895-904 (2007).

J.T. Lizier, M. Prokopenko, & A.Y. Zomaya, “*Local measures of information storage in complex distributed computation*”, Information Sciences 208, 39 (2012)

J.T. Lizier, "*JIDT: An information-theoretic toolkit for studying the dynamics of complex systems*", Frontiers in Robotics and AI, 1:11, 2014; appendix A.2 and A.3

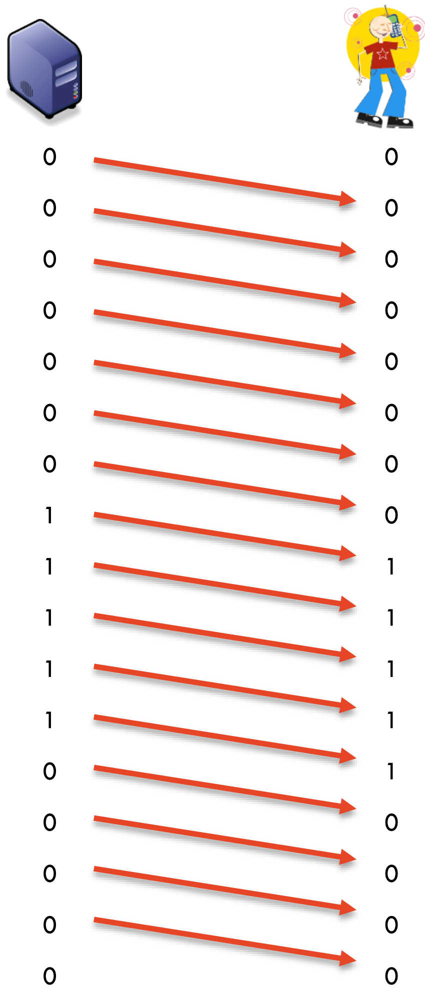
Information transfer



- Recall – we're building a model of the dynamics of the target variable
  - And we've already accounted for the past of the target (Information storage).
- How much information from the past of a source variable helps us predict the next state of the target …
- Or, in modelling the dynamics of the variable, how much information transfer would we include in that model by accounting for the past influence of the source …
- Given that we've already considered the past of the target.

Information transfer – using our intuition

- Simple example: heartbeat messages
  - Target copying simple (Poisson transitioning) messages from source



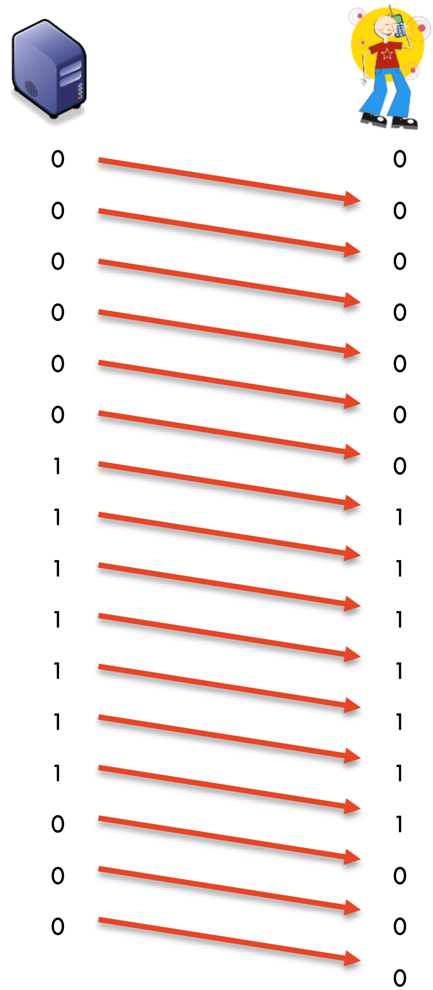
	$s_n = 0$	$s_n = 1$
$s_{n+1} = 0$	$1 - \lambda_1$	λ_0
$s_{n+1} = 1$	λ_1	$1 - \lambda_0$

- $t_{n+1} = s_n$
- Transition rates small (≪0.5), and $\lambda_1 < \lambda_0$ so values are normally 0.
  - E.g. $\lambda_1 = 0.05, \ \lambda_0 = 0.2$

- Try to predict the next values.
- Where did you take the information from?

Information transfer – using our intuition

- Simple example: heartbeat messages
  - Target copying simple (Poisson transitioning) messages from source

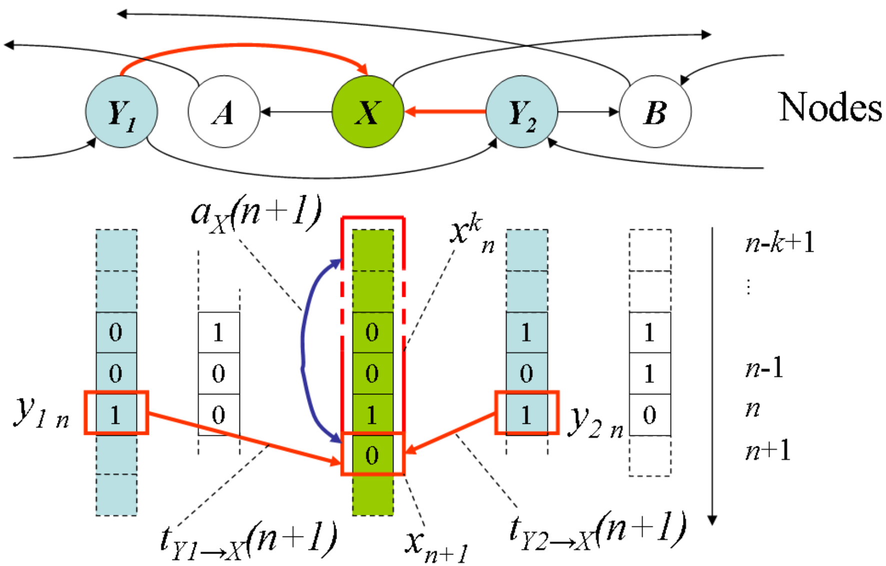


	$s_n = 0$	$s_n = 1$
$s_{n+1} = 0$	$1 - \lambda_1$	λ_0
$s_{n+1} = 1$	λ_1	$1 - \lambda_0$

- $t_{n+1} = s_n$
- Transition rates small (<0.5), and $\lambda_1 < \lambda_0$ so values are normally 0.
  - E.g. $\lambda_1 = 0.05,\ \lambda_0 = 0.2$
- Now: you are (modelling) the destination.
- At each step, model your prediction of the next value.
  - Write down your prediction based on past and how sure you are (scale of 1 to 10).
  - Update your prediction based on source, and evaluate how much source adds (scale 1 to 10)
- Where did source add most/least? Why?

Transfer entropy

- 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 $\boldsymbol{X}_n^{(k)} = \{X_{n-k+1}, \ldots, X_{n-1}, X_n\}$?



Transfer entropy

$$T_{Y\to X} = \lim_{k\to\infty} I\left(Y_n; X_{n+1} \middle| \boldsymbol{X}_n^{(k)}\right)$$

$$T_{Y\to X}(k) = I\left(Y_n; X_{n+1} \middle| \boldsymbol{X}_n^{(k)}\right)$$

$$T_{Y\to X}(k) = \left\langle \log_2 \frac{p\left(x_{n+1} \middle| \boldsymbol{x}_n^{(k)}, y_n\right)}{p\left(x_{n+1} \middle| \boldsymbol{x}_n^{(k)}\right)} \right\rangle$$

$$t_{Y\to X}(k) = \log_2 \frac{p\left(x_{n+1} \middle| \boldsymbol{x}_n^{(k)}, y_n\right)}{p\left(x_{n+1} \middle| \boldsymbol{x}_n^{(k)}\right)}$$

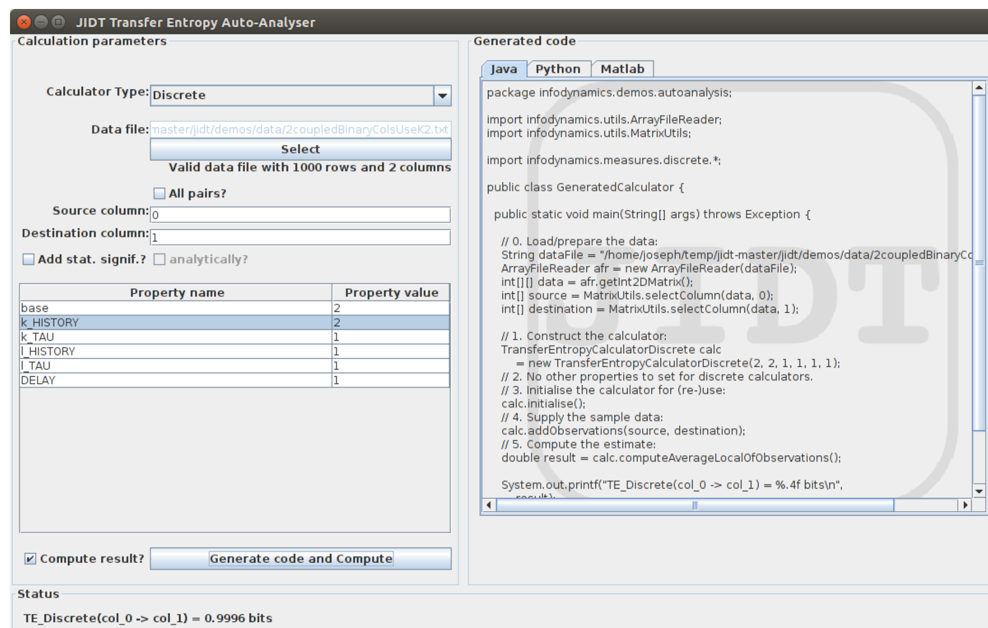
- TE: Average info from source that helps predict next value in context of past.
- Local TE: info from a specific source value to predict specific next value in context of past.

T. Schreiber, "*Measuring Information Transfer*", Physical Review Letters, 85(2), pp. 461-4, 2000.

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.

Transfer entropy in JIDT

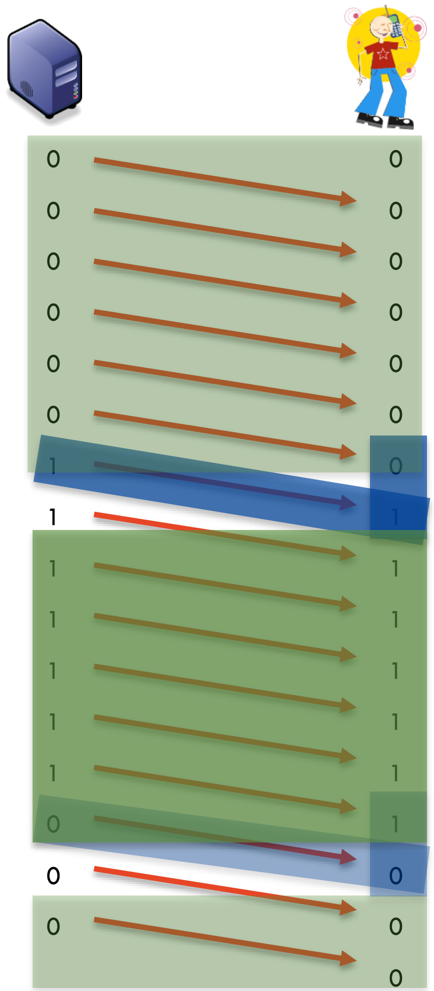
- Start TE AutoAnalyser
- Notice the important embedding parameters (e.g. `k_HISTORY`)
- Has all types of underlying CMI estimators available, same parameters as each &features (e.g. statistical significance, local)
- Using Gaussian model amounts to measuring Granger causality (factor of 2)



L. Barnett, A. B. Barrett, and A. K. Seth, "*Granger causality and transfer entropy are equivalent for Gaussian variables*", Phys. Rev. Lett, 103, 238701, 2009

Transfer entropy on heartbeat messages

– Examine local transfer entropy using JIDT on this example



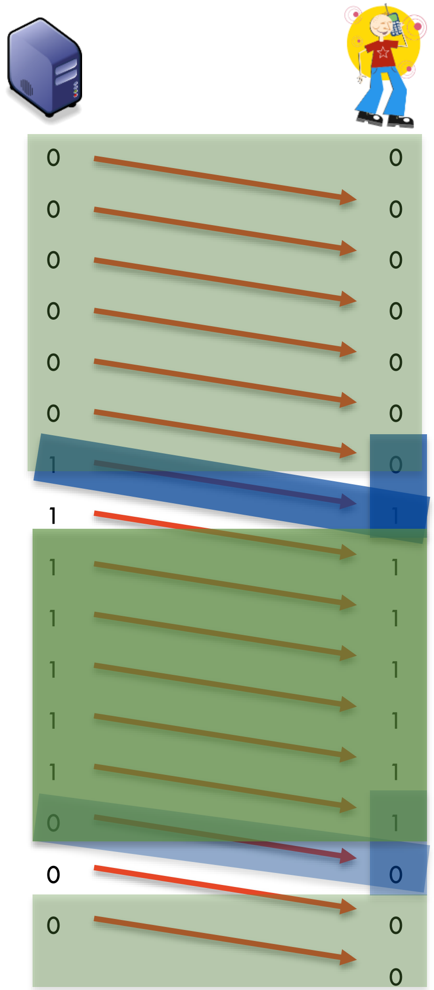
	$s_n = 0$	$s_n = 1$
$s_{n+1} = 0$	$1 - \lambda_1$	λ_0
$s_{n+1} = 1$	λ_1	$1 - \lambda_0$

– See tutorial activity with $\lambda_1 = 0.05$, $\lambda_0 = 0.2, N = 100000$

– Where did source add most?
  – Why?
– Where did source add the least?
  – Why?

Transfer entropy vs (lagged) mutual information

– How does $I\left(Y_n; X_{n+1} \middle| \boldsymbol{X}_n^{(k)}\right)$ contrast to $I(Y_n; X_{n+1})$?



– Both are directed, but conditioning on past makes TE dynamic

– Recall the effect of conditioning in a CMI:

  – Reduces MI in removing redundancies between the source and the target past

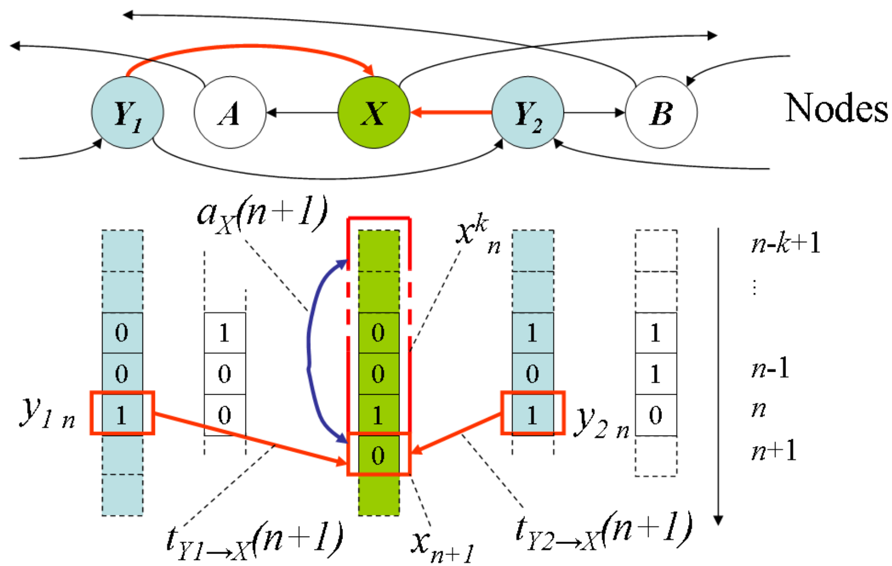
  – Increases MI in including synergies (where the source became more informative in the context of the past)

– *Challenge*: compute local MIs to compare

– These differences directly relate to our intuition about information transfer here!

Role of conditioning on history for TE

1. Provides a contrast to information storage
   - Removes any information storage from being considered as transfer. (removes redundant information in past and source)
   - Transfer is defined in juxtaposition to storage.
   - 2nd step in our modelling process of the computation of the target, after first considering the information that was already present in the target state.

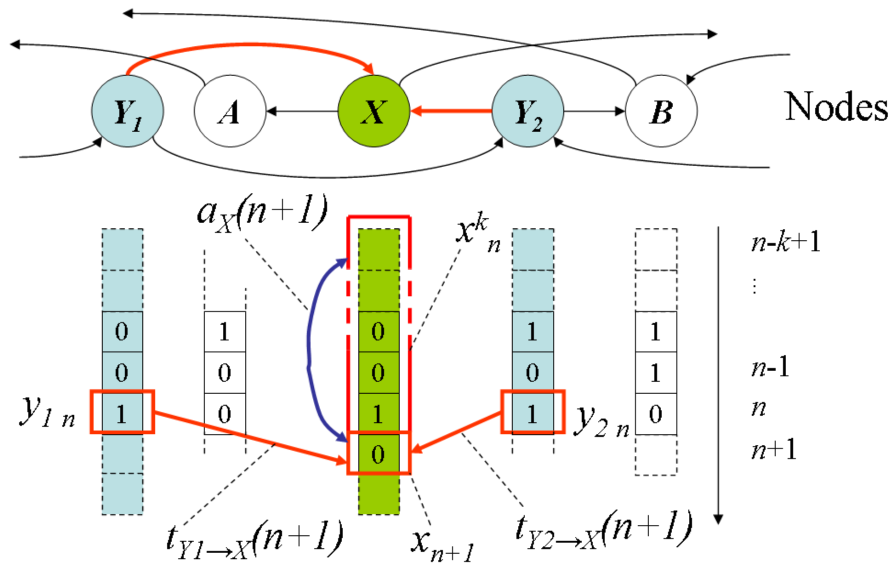


$$
\begin{aligned}
&H(X_{n+1}) \\
&= I\left(\boldsymbol{X}_n^{(k)}; X_{n+1}\right) + I\left(Y_n; X_{n+1} \middle| \boldsymbol{X}_n^{(k)}\right) \\
&+ H\left(X_{n+1} \middle| \boldsymbol{X}_n^{(k)}, Y_n\right)
\end{aligned}
$$

Bossomaier, Barnett, Harré, Lizier, "An Introduction to Transfer Entropy: Information Flow in Complex Systems", Springer, Cham, 2016; section 4.2.2
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 and J.R. Mahoney, "*Moving frames of reference, relativity and invariance in transfer entropy and information dynamics*", Entropy, 15(1), p. 177-197, 2013.

Role of conditioning on history for TE

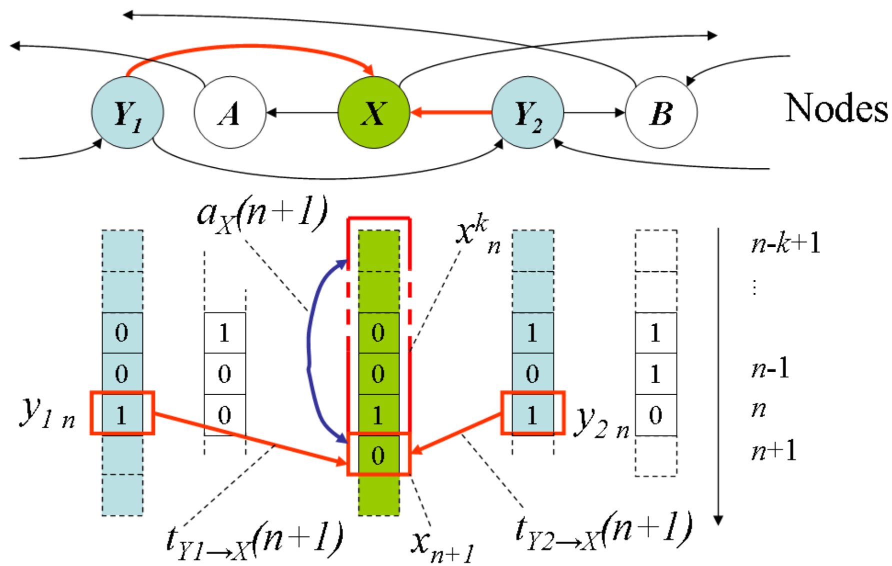
2. TE examines state transitions in the target
   - Recall the interpretation of $\boldsymbol{x}_n^{(k)}$ as a Takens' embedding of the underlying state of the target.
   - We are examining the state transition $\boldsymbol{x}_n^{(k)} \rightarrow x_{n+1}$ (or including redundant information $\boldsymbol{x}_n^{(k)} \rightarrow \boldsymbol{x}_{n+1}^{(k)}$)
   - TE is information provided by source about this dynamic state transition in the target.



T. Schreiber, "*Measuring Information Transfer*", Physical Review Letters, 85(2), pp. 461-4, 2000.
Bossomaier, Barnett, Harré, Lizier, "An Introduction to Transfer Entropy: Information Flow in Complex Systems", Springer, Cham, 2016; section 4.2.2
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 and J.R. Mahoney, "*Moving frames of reference, relativity and invariance in transfer entropy and information dynamics*", Entropy, 15(1), p. 177-197, 2013.

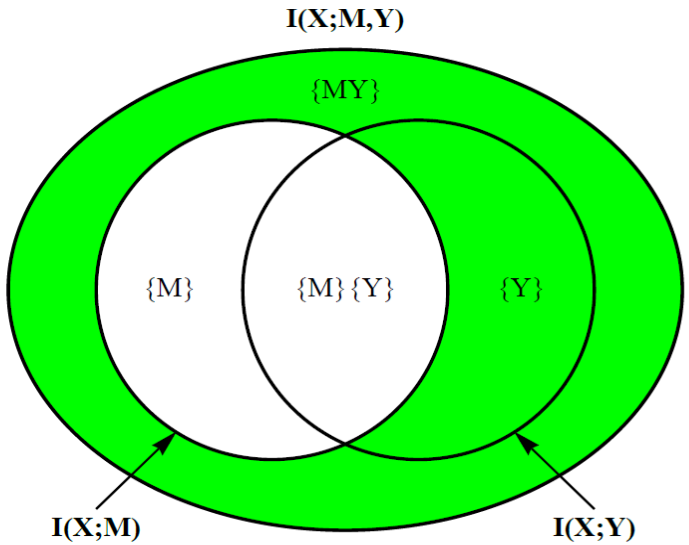
Role of conditioning on history for TE

3. TE examines contribution from source in the context of the target history
   – The target past does not only "condition out" redundant information, but
   – Allows TE to include synergistic (complementary) information from source and target past.
     • "state-dependent" component
   – Think of our modelling process as including the target past $x_n^{(k)}$ as context to consider other sources.



Bossomaier, Barnett, Harré, Lizier, "An Introduction to Transfer Entropy: Information Flow in Complex Systems", Springer, Cham, 2016; section 4.2.2
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 and J.R. Mahoney, "*Moving frames of reference, relativity and invariance in transfer entropy and information dynamics*", Entropy, 15(1), p. 177-197, 2013.
P.L. Williams, R. D. Beer, "*Generalized Measures of Information Transfer*", arXiv:1102.1507, 2011.

Relationship of TE to information storage

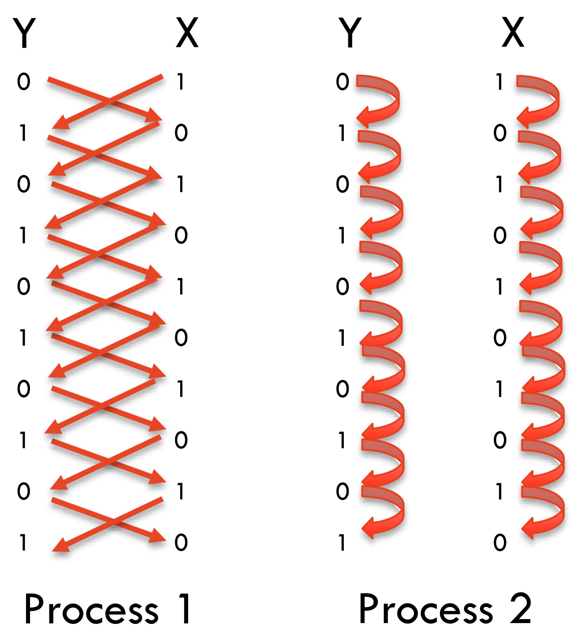


- Partial Information Decomposition helps us to see finer-grained detail inside these components:
  - AIS (white) includes a unique component from the past ({M}), plus a redundant component with transfer source ({M}{Y})
  - Transfer (green) includes a unique component from the source ({Y}), plus a synergistic component with the past of target ({MY})

P.L. Williams, R. D. Beer, "Generalized Measures of Information Transfer", arXiv:1102.1507, 2011.

Differentiating information transfer and causal effect

- Causality is about the effect of interventions (Pearl)
- Information transfer is about modelling dynamics as computation – information processing perspective.
- Heartbeat example: what were the differences?



- Causal attribution is different between processes 1 and 2
- Information processing model attributes info the same way:
  - All information storage
  - No information transfer

J. Pearl, "Causality: Models, Reasoning, and Inference" Cambridge University Press, Cambridge, 2000.
N. Ay, D. Polani, "Information flows in causal networks", Advances in Complex Systems, 11(1), pp. 17-41, 2008.
J.T. Lizier and M. Prokopenko, "Differentiating information transfer and causal effect", European Physical Journal B, vol. 73, no. 4, pp. 605-615, 2010
D. Chicharro, A. Ledberg, "When Two Become One: The Limits of Causality Analysis of Brain Dynamics", PLoS ONE, 7(3), e32466, 2012.

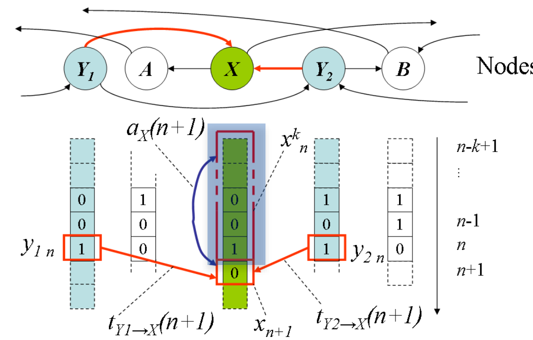
Differentiating information transfer and causal effect

- Information transfer does not directly measure causal effect:
  - Causal sources may serve information storage only
  - Non-causal sources (e.g. hidden common drivers) can have transfer
    - One should restrict analysis to causal sources where possible.
- Information transfer measures emergent computational structure that causality does not, e.g.:
  - Transitions in Heartbeat example
  - Gliders in cellular automata

- Concepts are complementary

J.T. Lizier and M. Prokopenko, "Differentiating information transfer and causal effect", European Physical Journal B, vol. 73, no. 4, pp. 605-615, 2010
Bossomaier, Barnett, Harré, Lizier, "An Introduction to Transfer Entropy: Information Flow in Complex Systems", Springer, Cham, 2016; section 4.2.2.1

Using transfer entropy in practice



1. Set the target embedding
   - Similar concerns as per the active information storage (see Info storage session), and can include an embedding delay:

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

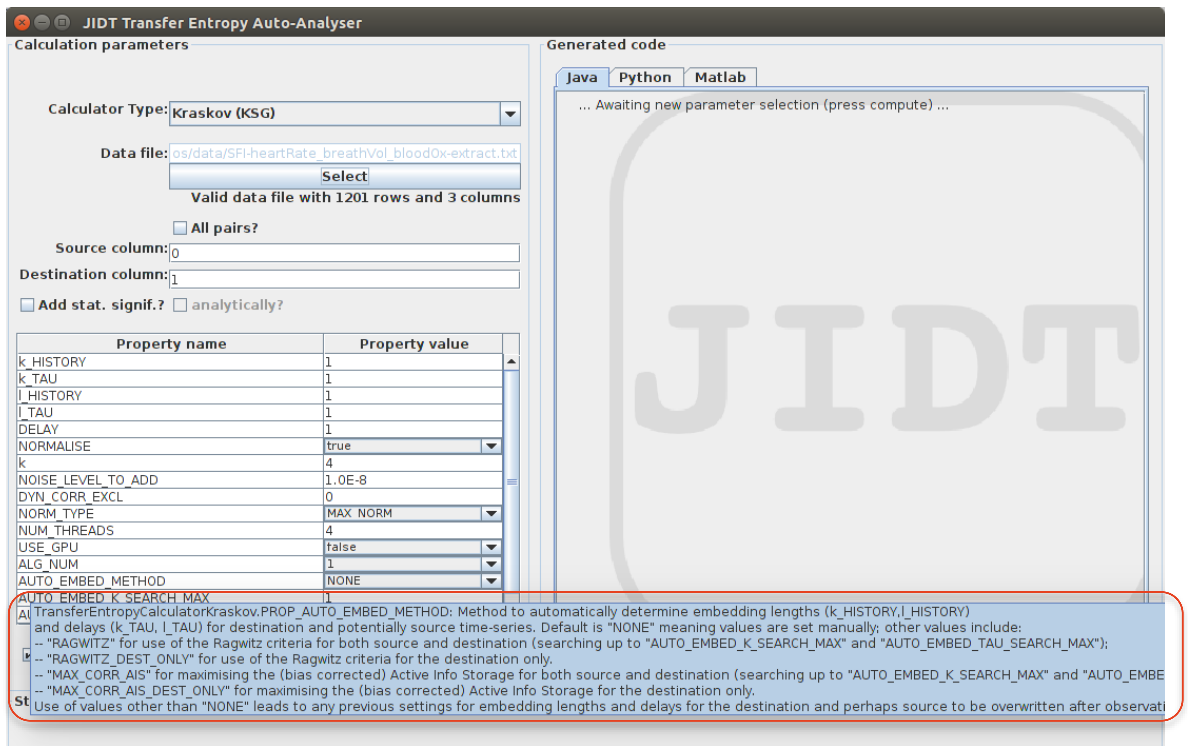
   - Set in the same manner to minimise stored information being counted as transfer, i.e.:
     - Max bias-corrected AIS or Ragwitz criteria or non-uniform.
     - First two options are available in JIDT

   - **Caveats:**
     - Must embed target **first** for proper interpretation as information transfer! (Rather than embedding source-target together)
     - Additional target past values could be missed here which add synergistic information with source.

Bossomaier, Barnett, Harré, Lizier, "An Introduction to Transfer Entropy: Information Flow in Complex Systems", Springer, Cham, 2016; section 4.2
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

TE: setting target embedding parameters in JIDT

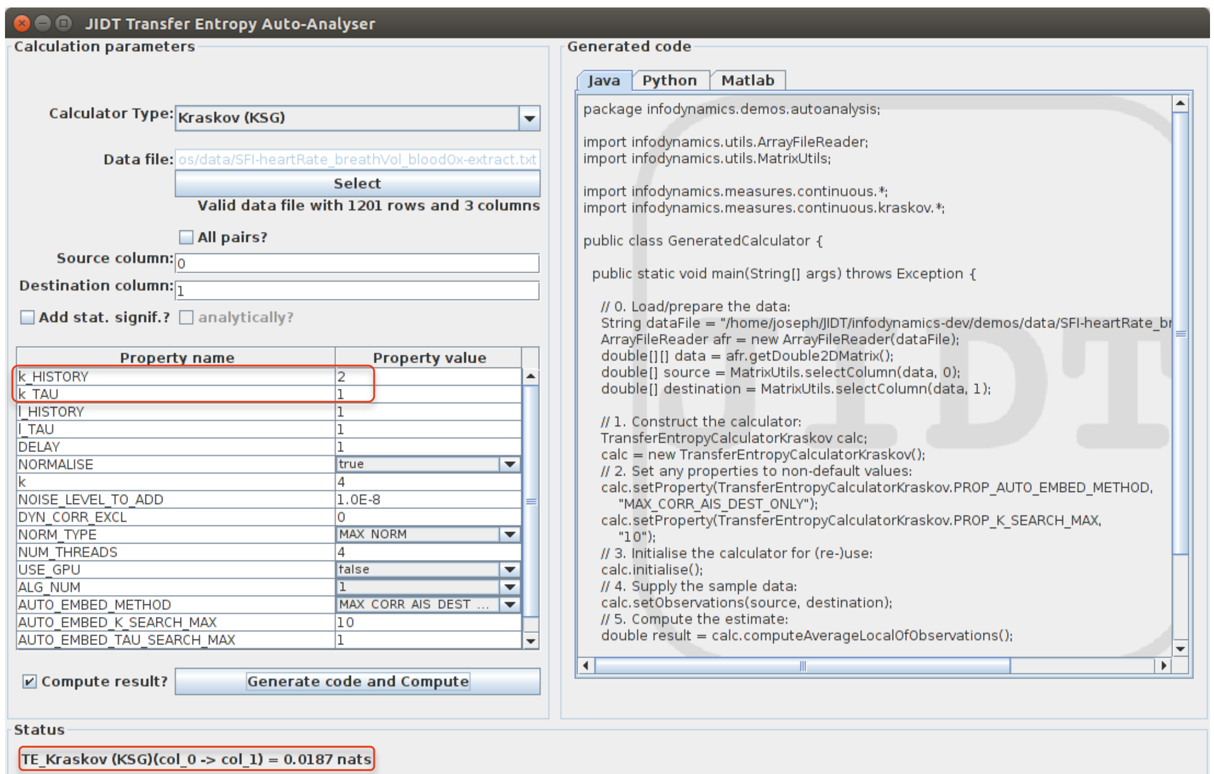
- Select KSG estimator
- `AUTO_EMBED_METHOD` property provides options (hover for info):
  - `NONE`: set `k_HISTORY` and `TAU` manually.
  - `RAGWITZ_DEST_ONLY`:* optimal parameters to minimise prediction error scanned up to `AUTO_EMBED_K_SEARCH_MAX` and `AUTO_EMBED_TAU_SEARCH_MAX`
  - `MAX_CORR_AIS_DEST_ONLY`:* optimal parameters to max. bias-corrected AIS scanned up to `AUTO_EMBED_K_SEARCH_MAX` and `AUTO_EMBED_TAU_SEARCH_MAX`

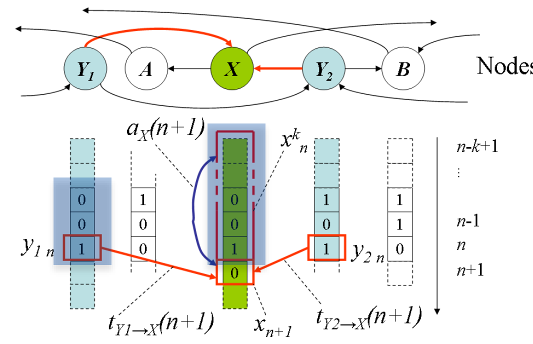


* Note clarification on these parameters re source embedding later!

TE: setting embedding parameters in JIDT

- Select the `SFI-heartRate_breathVol_bloodOx_extract.txt` data set
- Set `AUTO_EMBED_METHOD` to `MAX_CORR_AIS_DEST_ONLY` and `AUTO_EMBED_K_SEARCH_MAX` to 10 and `AUTO_EMBED_TAU_SEARCH_MAX` to 1.
- Click Compute
- The result is returned with optimal parameters shown in `k_HISTORY` and `TAU`. You can retrieve them in code via a `getProperty()` call.





Using transfer entropy in practice

2. Set the source embedding
   - I gave you the simple TE definition earlier!
   - More general including source embedding:

$$T_{Y\to X}(k,l,\tau_X,\tau_Y) = I\left(\boldsymbol{Y}_n^{(l,\tau_Y)}; X_{n+1} \middle| \boldsymbol{X}_n^{(k,\tau_X)}\right)$$

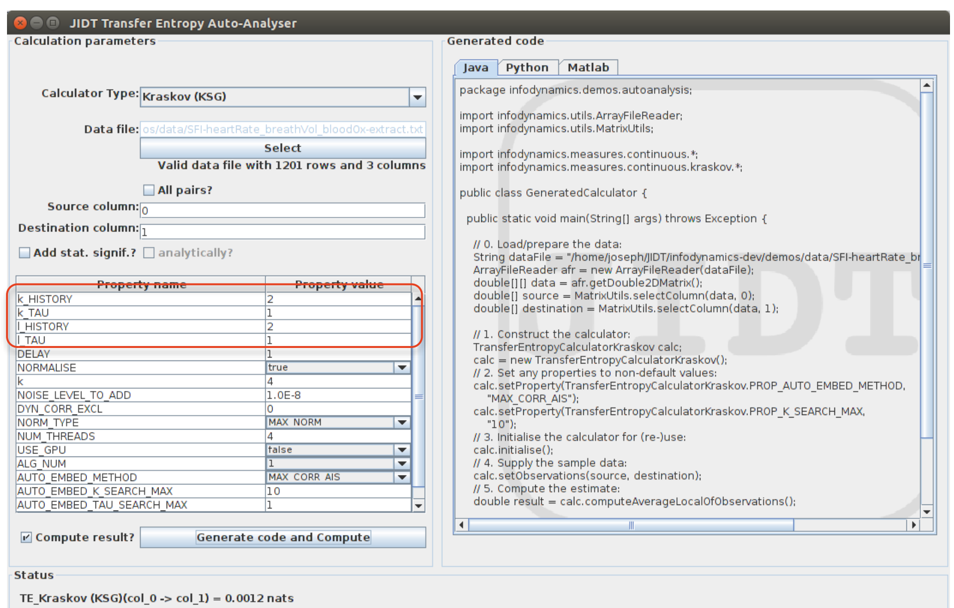
   - When to use source state $\boldsymbol{y}_n^{(k,\tau_Y)}$ directly instead of scalar y_n:
     - If observations y_n masks hidden Markov process
     - If multiple past values of Y are causal to X
     - Otherwise we can build a set of contributors from Y via non-uniform embedding (see next lecture).

Bossomaier, Barnett, Harré, Lizier, "An Introduction to Transfer Entropy: Information Flow in Complex Systems", Springer, Cham, 2016; section 4.2.1.2

J. T. Lizier. "*Measuring the dynamics of information processing on a local scale in time and space*". In M. Wibral, R. Vicente, and J. T. Lizier, editors, *Directed Information Measures in Neuroscience*, pp. 161–193. Springer, Berlin/Heidelberg, 2014.

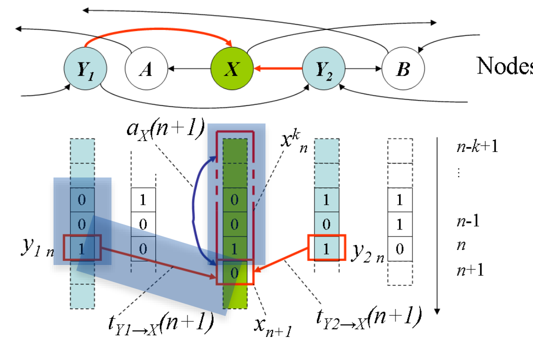
TE: setting source embedding parameters in JIDT

- Select KSG estimator
- `AUTO_EMBED_METHOD` property provides options (hover for info):
  - `NONE`: set target and source embeddings manually.
  - `RAGWITZ`: set target and source embeddings to minimise prediction error
    - `RAGWITZ_DEST_ONLY`: for target only
  - `MAX_CORR_AIS`: set target and source embeddings to max. bias-corrected AIS
    - `MAX_CORR_AIS_DEST_ONLY`: for target only
  - All scan up to `AUTO_EMBED_K_SEARCH_MAX` and `AUTO_EMBED_TAU_SEARCH_MAX`



* Notice the slight change in these options compared to AIS, because we're dealing with source + target

Using transfer entropy in practice



3. Set the source-target delay
   - Ok, I still gave you a simpler TE definition!
   - More general still including source-target delay:

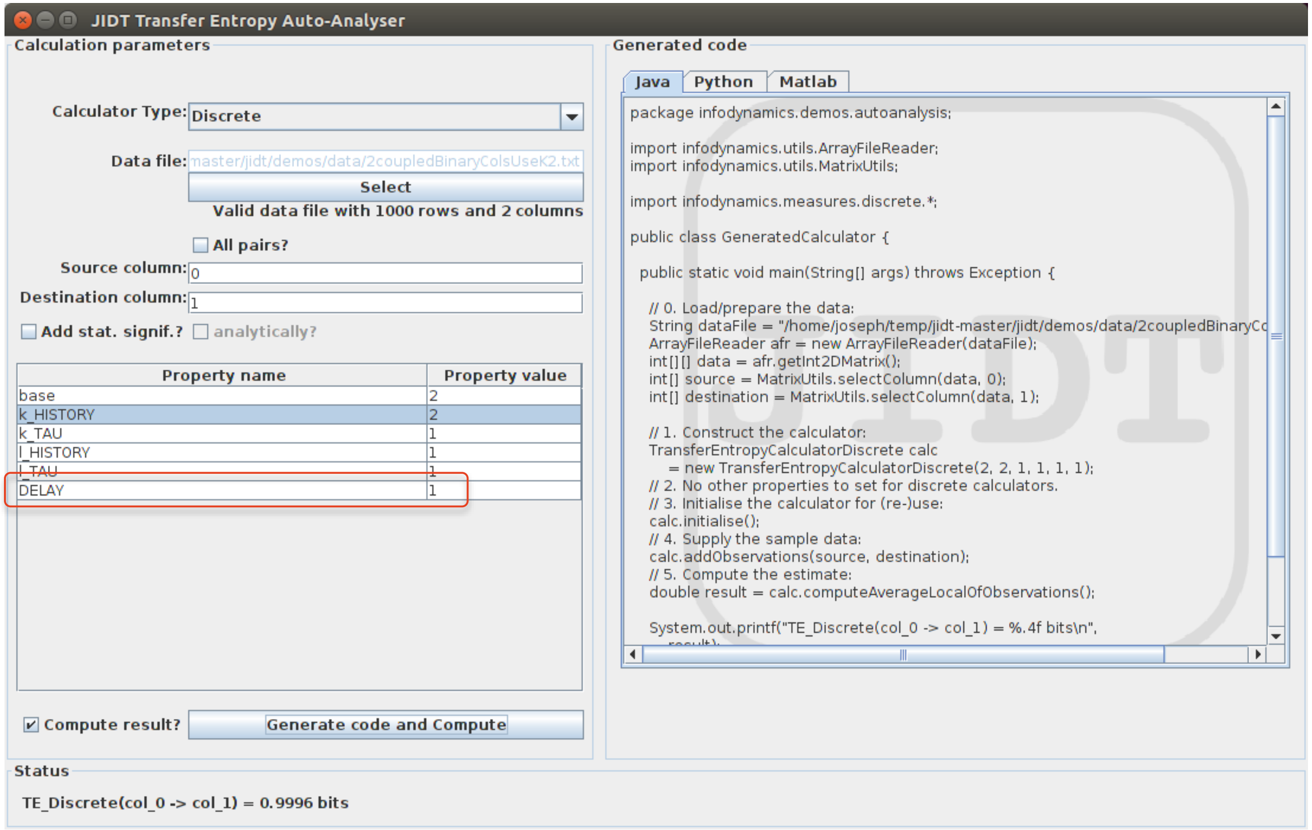
$$T_{Y\to X}(k,l,\tau_X,\tau_Y,u) = I\left(\boldsymbol{Y}_{n+1-u}^{(l,\tau_Y)}; X_{n+1} \middle| \boldsymbol{X}_n^{(k,\tau_X)}\right)$$

   - Set u to maximise $T_{Y\to X}$
     - Proven to match actual causal delay under simple conditions.

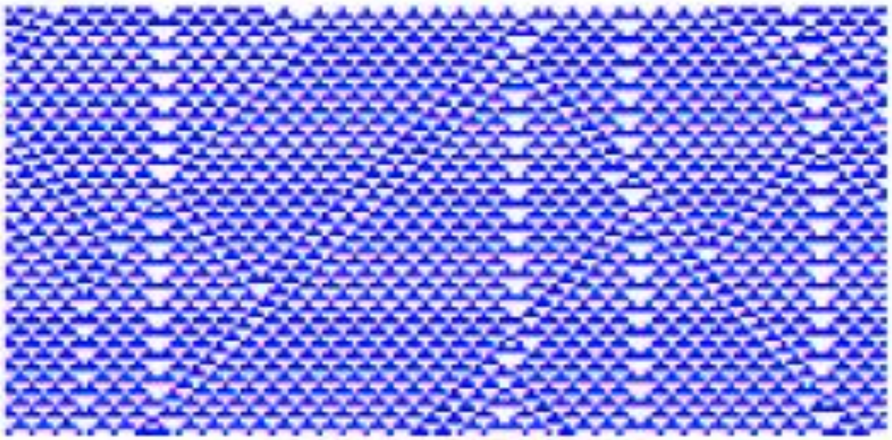
M. Wibral, N. Pampu, V. Priesemann, F. Siebenh¨uhner, H. Seiwert, M. Lindner, J. T. Lizier, and R. Vicente. “*Measuring information-transfer delays*”. *PLoS ONE*, 8(2):e55809+, 2013.
Bossomaier, Barnett, Harré, Lizier, "An Introduction to Transfer Entropy: Information Flow in Complex Systems", Springer, Cham, 2016; section 4.2.4

TE: setting source-target delay in JIDT

– Set `DELAY` property (available on all except kernel estimator).



Example: Computational role of emergent structure in CAs



cells by value

cells by look-up and filtered

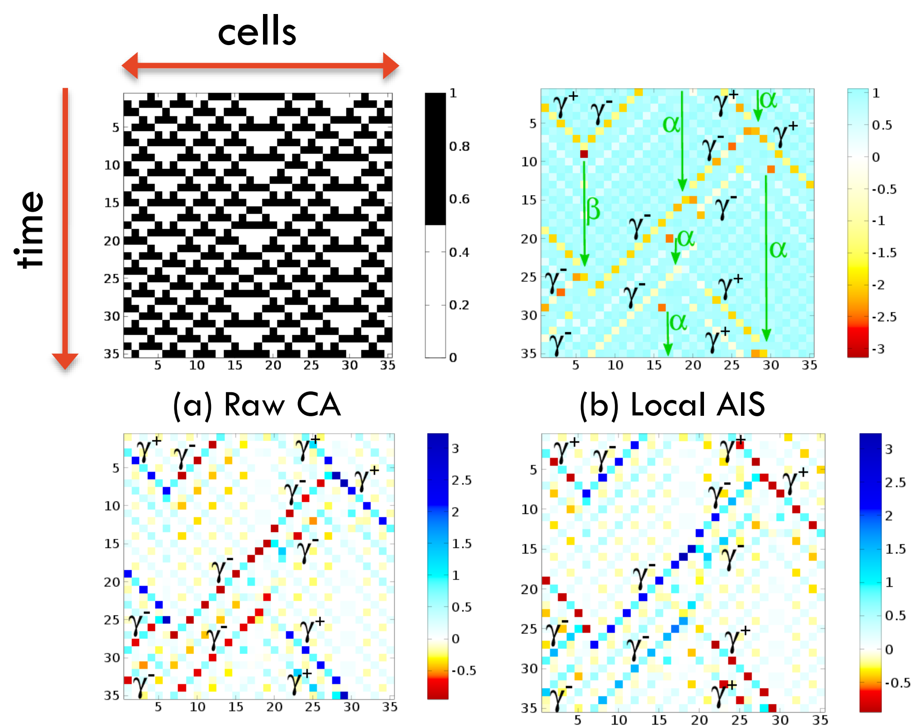
– Emergent structure:
  – Domains, blinkers
  – Particles
    • Gliders, Domain walls
  – Collisions

– Conjectured to represent
  – Information storage
  – Information transfer
    • “
  – Information modification

A. Wuensche, “Classifying cellular automata automatically: Finding gliders, filtering, and relating space-time patterns, attractor basins, and the Z parameter,” Complexity, vol. 4, no. 3, pp. 47–66, 1999. (plus image credit, © Wiley, used with permission)
C. G. Langton, “Computation at the edge of chaos: phase transitions and emergent computation,” Physica D, vol. 42, no. 1-3, pp. 12–37, 1990.
M. Mitchell, J.P. Crutchfield, R. Das, “*Evolving Cellular Automata with Genetic Algorithms: A Review of Recent Work*”, Proc. 1st Int. Conf. Evolutionary Computation and Its Applications (EvCA’96), 1996. (see p. 1/10)

Example: Computational role of emergent structure in CAs

- Go to activity to try out our TE calculator on CA data:
  - We'll use the appropriate embedding length determined for AIS in previous session.
  - We'll compute local TE values and see whether gliders do indeed have strong information transfer values.



(a) Raw CA (b) Local AIS

(c) Local TE right (d) Local TE left

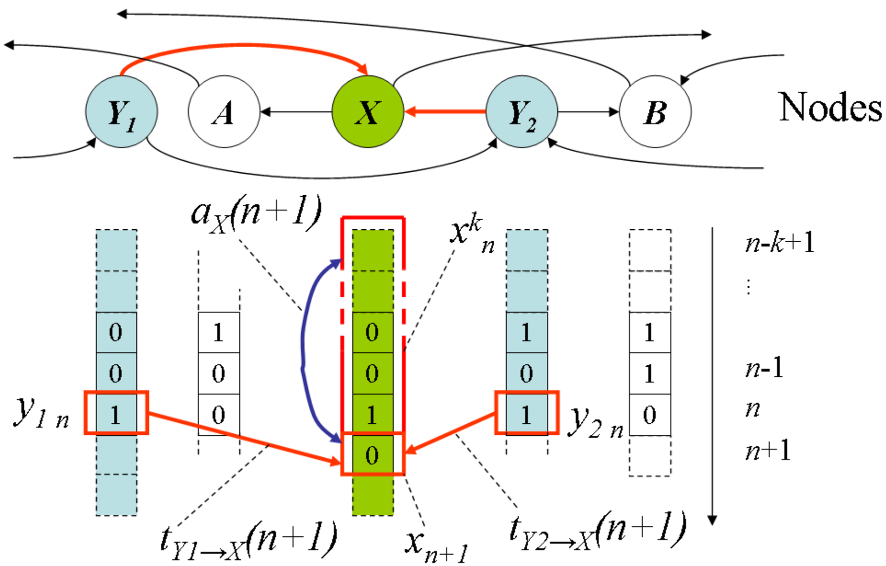
Spoiler alert:
Gliders are dominant transfer entities!

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, "JIDT: An Information-Theoretic toolkit for studying the dynamics of complex systems". Frontiers in Robotics and AI, 1:11, 2014.

Conditional Transfer entropy

– How much info about next observation X_n of process X can be found in observation Y_n of process Y, in context of the past state $\boldsymbol{X}_n^{(k)} = \{X_{n-k+1}, \ldots, X_{n-1}, X_n\}$ and observation Z_n of process Z?



$$T_{Y\to X|Z} = \lim_{k\to\infty} I\left(Y_n ; X_{n+1} \middle| \boldsymbol{X}_n^{(k)}, Z_n\right)$$

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

$$T_{Y\to X|Z}(k) = \left\langle \log_2 \frac{p\left(x_{n+1} \middle| \boldsymbol{x}_n^{(k)}, y_n, z_n\right)}{p\left(x_{n+1} \middle| \boldsymbol{x}_n^{(k)}, z_n\right)} \right\rangle$$

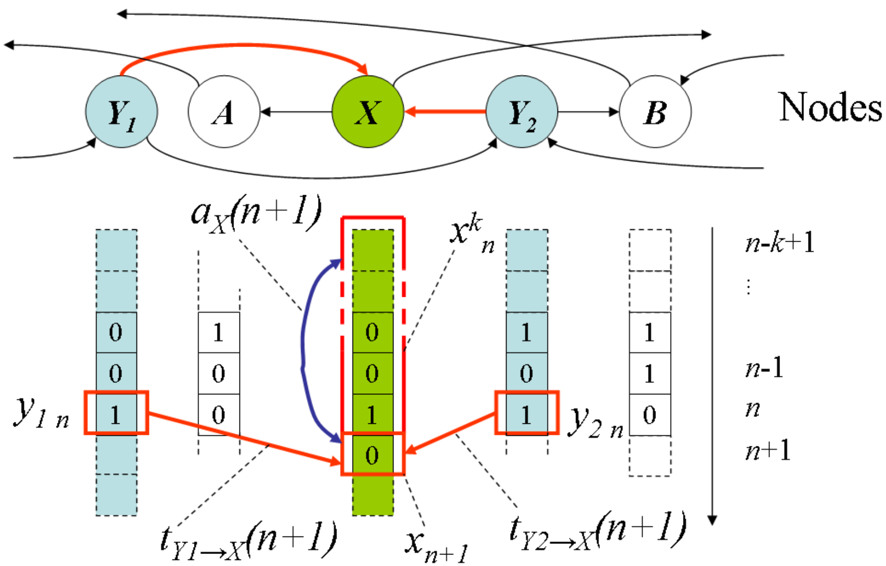
$$t_{Y\to X|Z}(k) = \log_2 \frac{p\left(x_{n+1} \middle| \boldsymbol{x}_n^{(k)}, y_n, z_n\right)}{p\left(x_{n+1} \middle| \boldsymbol{x}_n^{(k)}, z_n\right)}$$

– Can add same source-target delays, conditional-target delays, embeddings, multivariate conditionals etc., all in style of such extensions for TE.

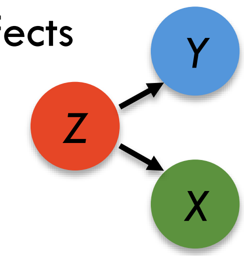
J. T. Lizier, M. Prokopenko, & A. Y. Zomaya. "*Local information transfer as a spatiotemporal filter for complex systems*". Physical Review E, 77(2):026110, 2008.
J. T. Lizier, M. Prokopenko, & A. Y. Zomaya. "*Information modification and particle collisions in distributed computation*", Chaos, 20(3), 037109, 2010.

Contrasting TE and CTE

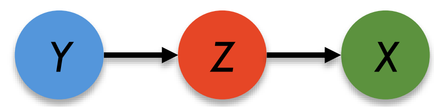
– How does extra conditioning on Z change TE?



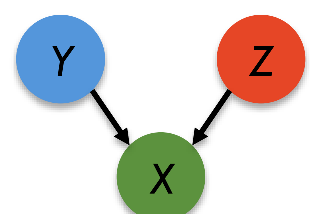
– Removes redundancies between Y and Z, e.g.
  – Common driver effects



  – Pathway effects



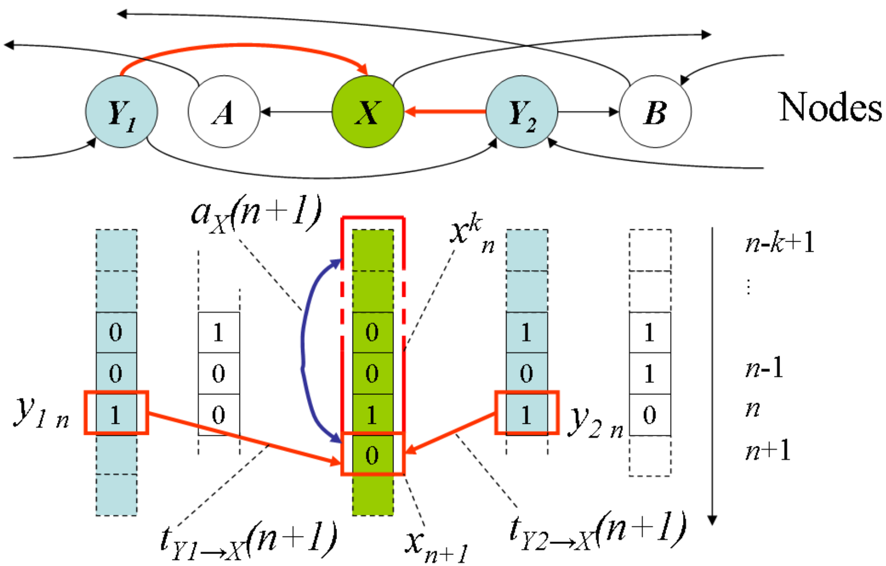
– Includes synergies between Y and Z, e.g.
  – Gated effects



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.

Relating TE and CTE

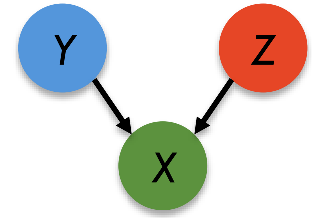
– How does extra conditioning on Z change TE? Do we need them both?



$$T_{Y\to X}(k) = I\left(Y_n; X_{n+1} \middle| \boldsymbol{X}_n^{(k)}\right)$$

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

Information regression



- Our goal in modelling the information processing in X.
- Consider two sources to X. (General case in Lizier 2010):

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

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

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

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

1. Active information storage

2-. Collective transfer entropy

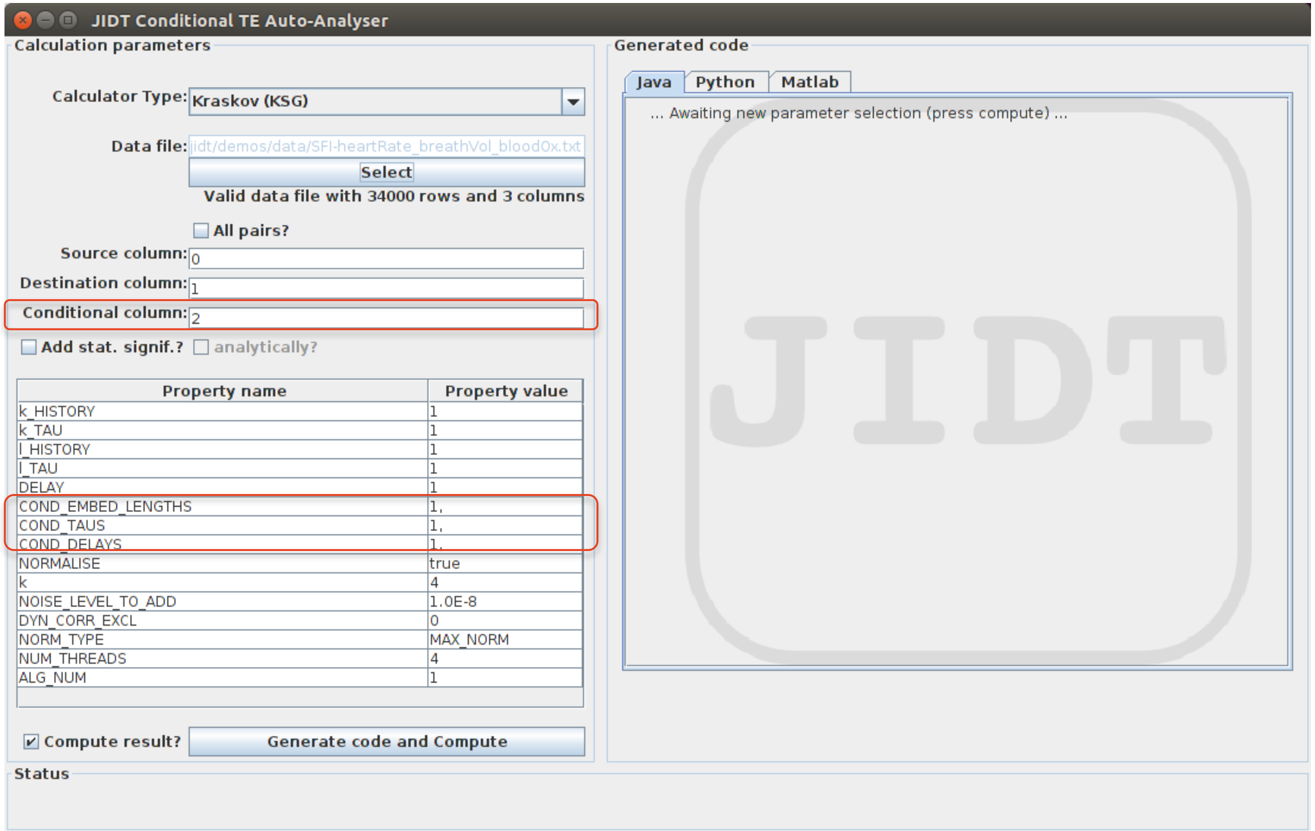
2. Pairwise/apparent transfer entropy

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.

Conditional Transfer entropy in JIDT

- Start CTE AutoAnalyser
- Has all types of underlying CMI estimators available, same parameters as each & features (e.g. statistical significance, local)



Information dynamics Part III: summary

- We've looked at the philosophy behind information storage and transfer and how they are related for analysing information processing in complex systems.
  - Transfer entropy as a model for information transfer in a wider perspective of distributed intrinsic computation
- In particular, we've focussed on how information transfer is characterised.
  - How pairwise and higher order transfer entropies are related.
  - And used JIDT AutoAnalyser and extensions of code to analyse information transfer in complex systems data sets.

- *Coming up*: Inferring effective networks from time-series dynamics.

Questions