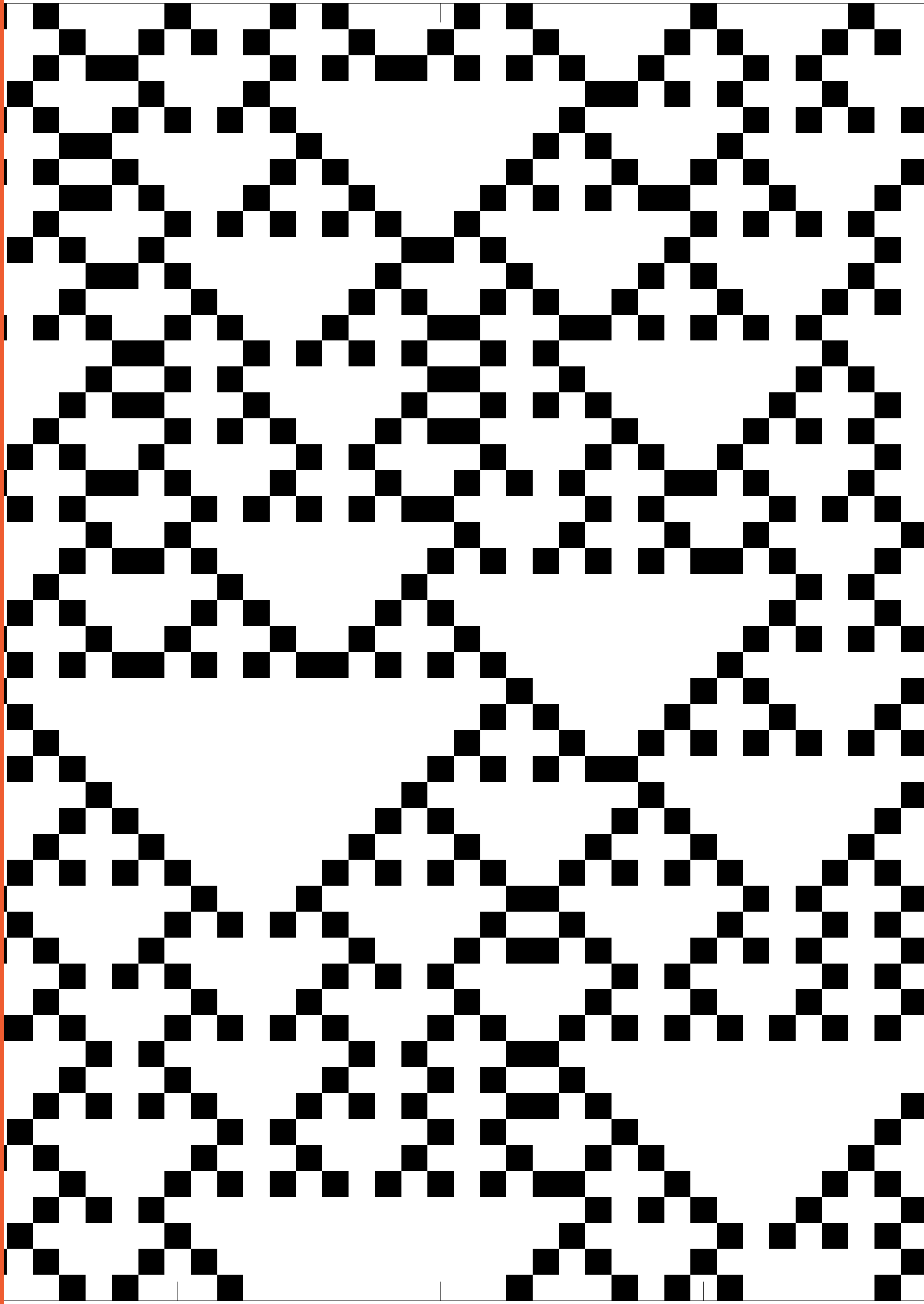
Differentiating information transfer and causal effect.

Or: why you should be interested in more than causality

Dr. Joseph Lizier

THE UNIVERSITY OF SYDNEY



Differentiating information transfer and causal effect

– J.T. Lizier and M. Prokopenko, "*Differentiating information transfer and causal effect*", European Physical Journal B, vol. 73, no. 4, pp. 605-615, 2010. doi: 10.1140/epjb/e2010-00034-5

Differentiating information transfer and causal effect

J.T. Lizier[1,2,a] and M. Prokopenko[1,3]

[1] CSIRO Information and Communications Technology Centre, Locked Bag 17, North Ryde, 1670 NSW, Australia
[2] School of Information Technologies, The University of Sydney, 2006 NSW, Australia
[3] Max Planck Institute for Mathematics in the Sciences, Inselstraße 22-26, 04103 Leipzig, Germany

Received 19 January 2009 / Received in final form 4 December 2009
Published online 27 January 2010 – © EDP Sciences, Società Italiana di Fisica, Springer-Verlag 2010

Abstract. The concepts of information transfer and causal effect have received much recent attention, yet often the two are not appropriately distinguished and certain measures have been suggested to be suitable for both. We discuss two existing measures, transfer entropy and information flow, which can be used separately to quantify information transfer and causal information flow respectively. We apply these measures to cellular automata on a local scale in space and time, in order to explicitly contrast them and emphasize the differences between information transfer and causality. We also describe the manner in which the measures are complementary, including the conditions under which they in fact converge. We show that causal information flow is a primary tool to describe the causal structure of a system, while information transfer can then be used to describe the emergent computation on that causal structure.

Concepts of information transfer and causal effect

- Discussions around "information transfer" appear to subsume two differing concepts
  - **Predictive** or computational information transfer
    - *"if I know the state of the source, how much does that help to predict the state of the destination?"*
  - **Causal** effect
    - *"if I change the state of the source, to what extent does that cause or alter the state of the destination?"*
- Several questions persist regarding these concepts:
  - Is information transfer akin to causal effect?
  - If not, what is the difference?
  - When examining the "effect" of one variable on another (e.g. between brain regions), should one seek to measure information transfer or causal effect?

Concepts of information transfer and causal effect

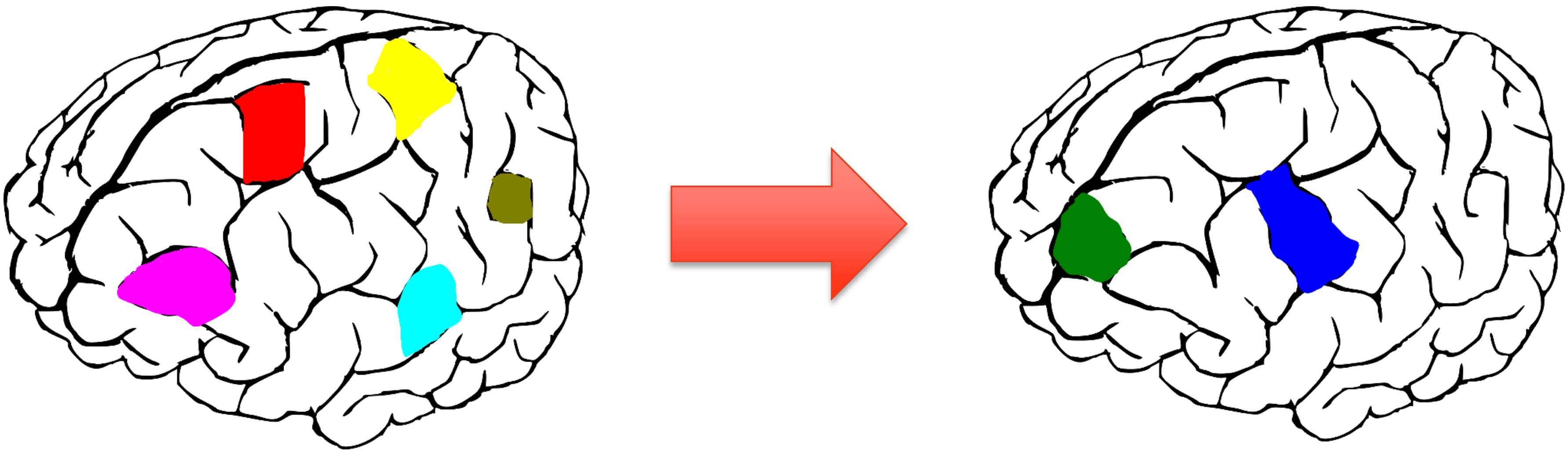
- Discussions around "information transfer" appear to subsume two differing concepts
  - **Predictive** or computational information transfer
    - *"if I know the state of the source, how much does that help to predict the state of the destination?"*
  - **Causal** effect
    - *"if I change the state of the source, to what extent does that cause or alter the state of the destination?"*
- Issues:
  - Both concepts have been associated with "information transfer" by various authors
  - Predictive transfer has been used to infer causality and directly equated with it.
  - This keeps me awake at night!

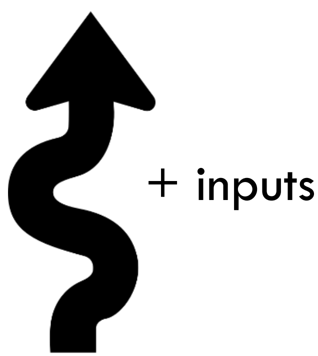
Overview

Concept	Causal effect	Information transfer
Measure	Causal information flow	Transfer entropy

- **Thesis:** concepts are quite distinct
  - **predictive transfer** (or the computational perspective) is more closely aligned with the popularly understood notion of information transfer,
  - while **causal information flow** should be considered as a separate fundamental notion.
- Clarify, and contrast and emphasise the differences
- Show their complementarity and where they converge
- **Message:**
  1. causal information flow is a primary tool to establish system structure,
  2. transfer entropy reveals information transfer in emergent computation.

Predictive information transfer: bigger picture

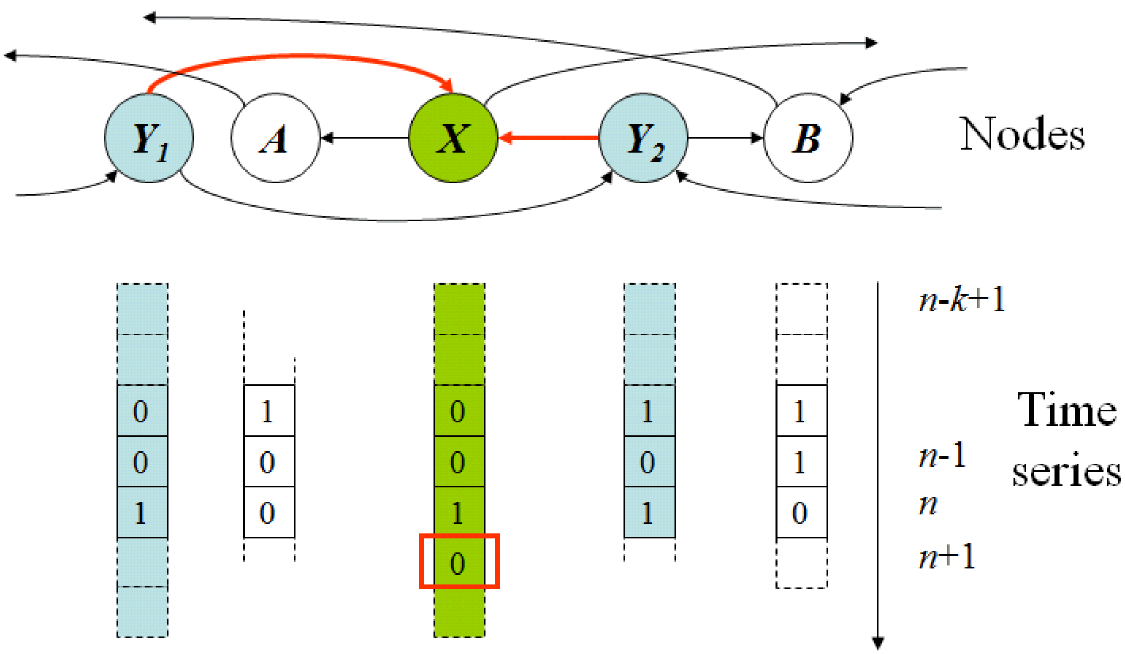




– Q: how can we quantify the unfolding intrinsic computation in terms of the interaction between variables in the complex system?

Predictive information transfer: bigger picture

– Key question: how is the next state of a variable in a complex system **computed**?



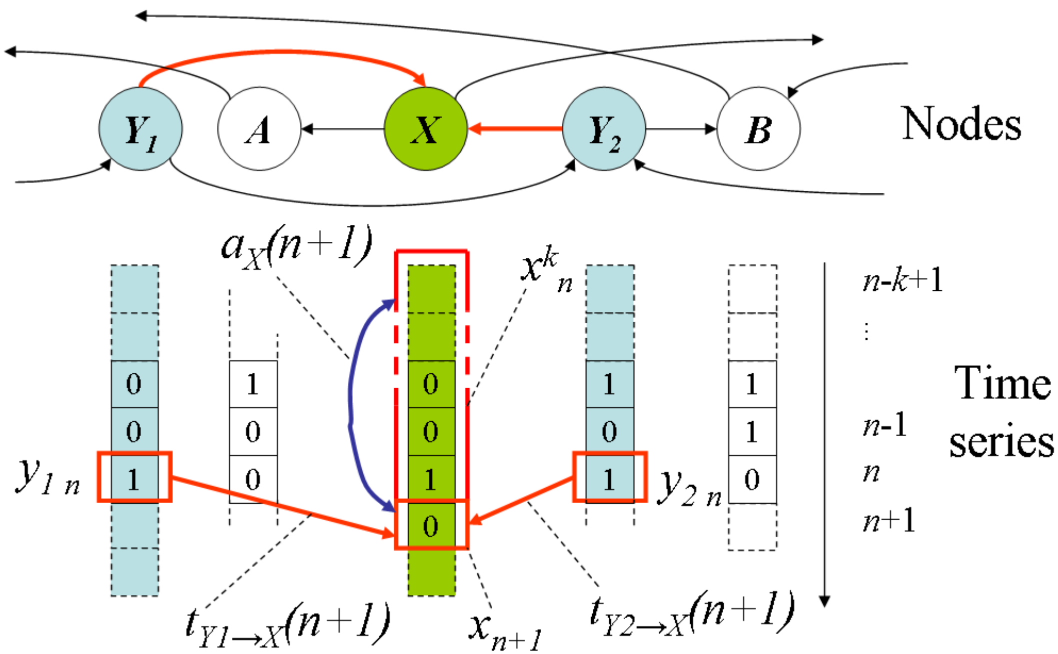
Q: Where does the information in x n+1 come from, and how can we measure it?

Q: How much was stored, how much was transferred, can we partition them or do they overlap?

Complex system as a multivariate time-series of states

Information dynamics

– Studies computation of the next state of a target variable in terms of information storage, transfer and modification: (Lizier et al., 2008, 2010, 2012)



The measures examine:

- State updates of a target variable;
- Dynamics of the measures in space and time.

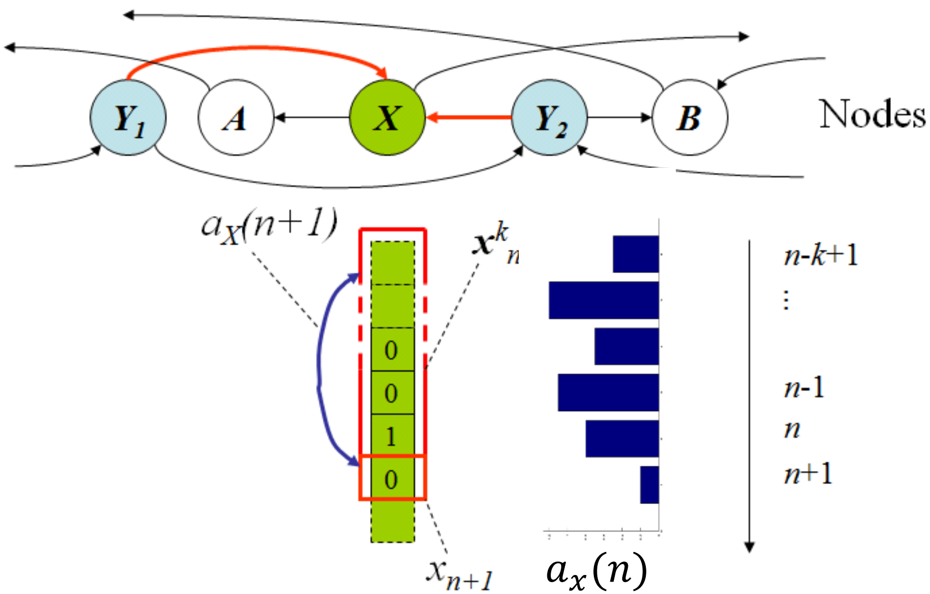
Active information storage (Lizier et al., 2012)

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

Active information storage:

$$A_X = I\left(\boldsymbol{X_n^{(k)}}; X_{n+1}\right)$$
$$= \left\langle \log \frac{p(x_{n+1}|\boldsymbol{x_n^{(k)}})}{p(x_{n+1})} \right\rangle$$

Average information from past state that is in use in predicting the next value.



Local active information storage:

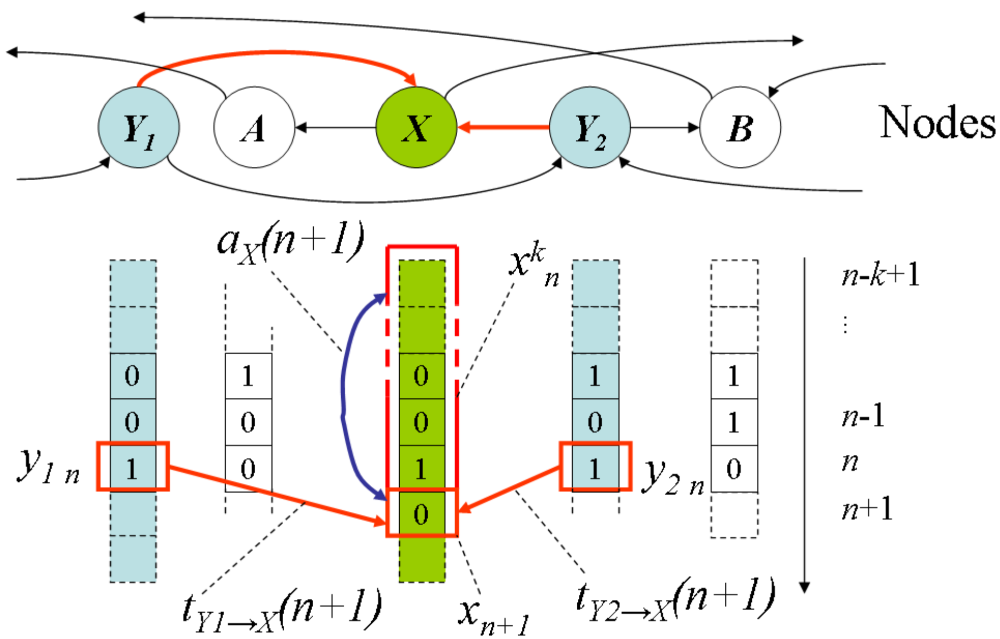
$$a_x(n) = i\left(\boldsymbol{x_n^{(k)}}; x_{n+1}\right)$$
$$= \log \frac{p(x_{n+1}|\boldsymbol{x_n^{(k)}})}{p(x_{n+1})}$$

Information from a specific past state that is in use in predicting the specific next value.

$$A_X = a_x(n)$$

Information transfer

– How much information about the state transition $\boldsymbol{X}_{\boldsymbol{n}}^{(\boldsymbol{k})} \rightarrow X_{n+1}$ of X can be found in the past state $\boldsymbol{Y}_{\boldsymbol{n}}^{(\boldsymbol{l})}$ of a source process Y?



Transfer entropy (Schreiber, 2000):

$$T_{Y\rightarrow X} = I\left(\boldsymbol{Y}_{\boldsymbol{n}}^{(\boldsymbol{l})}; X_{n+1} | \boldsymbol{X}_{\boldsymbol{n}}^{(\boldsymbol{k})}\right)$$

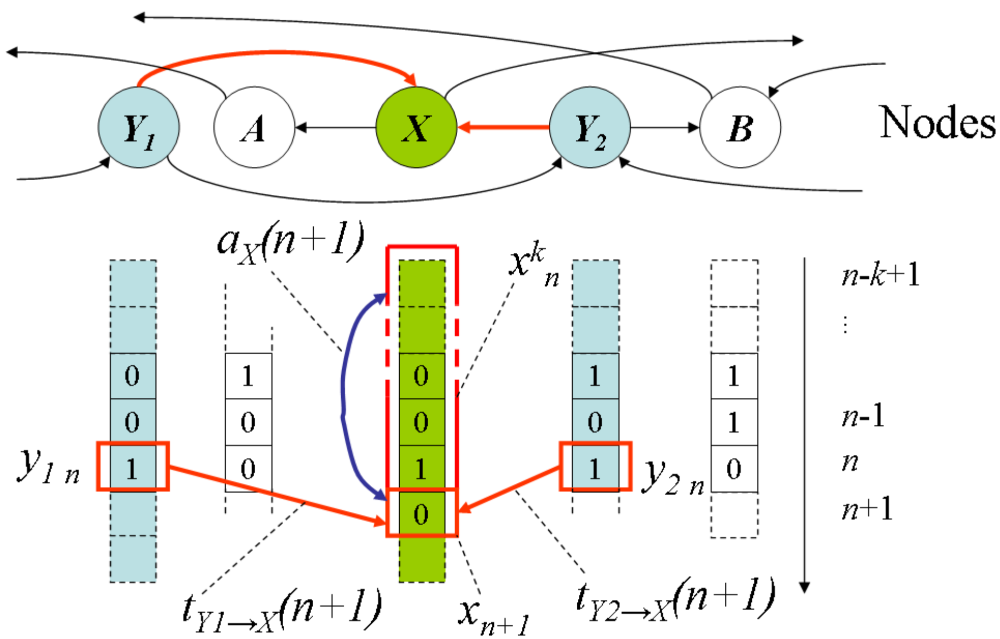
$$= \left\langle \log \frac{p(x_{n+1} | \boldsymbol{x}_{\boldsymbol{n}}^{(\boldsymbol{k})}, \boldsymbol{y}_{\boldsymbol{n}}^{(\boldsymbol{l})})}{p(x_{n+1} | \boldsymbol{x}_{\boldsymbol{n}}^{(\boldsymbol{k})})} \right\rangle$$

Average info from source that helps predict next value in context of past.

Storage and transfer are complementary:
$H_X = A_X + T_{Y\rightarrow X}$ + higher order terms

Information transfer

– How much information about the state transition $\boldsymbol{X_n^{(k)}} \rightarrow X_{n+1}$ of X can be found in the past state $\boldsymbol{Y_n^{(l)}}$ of a source process Y?



Transfer entropy (Schreiber, 2000):

$$T_{Y\rightarrow X} = I\left(\boldsymbol{Y_n^{(l)}}; X_{n+1} | \boldsymbol{X_n^{(k)}}\right)$$

$$= \left\langle \log \frac{p(x_{n+1} | \boldsymbol{x_n^{(k)}}, \boldsymbol{y_n^{(l)}})}{p(x_{n+1} | \boldsymbol{x_n^{(k)}})} \right\rangle$$

Average info from source that helps predict next value in context of past.

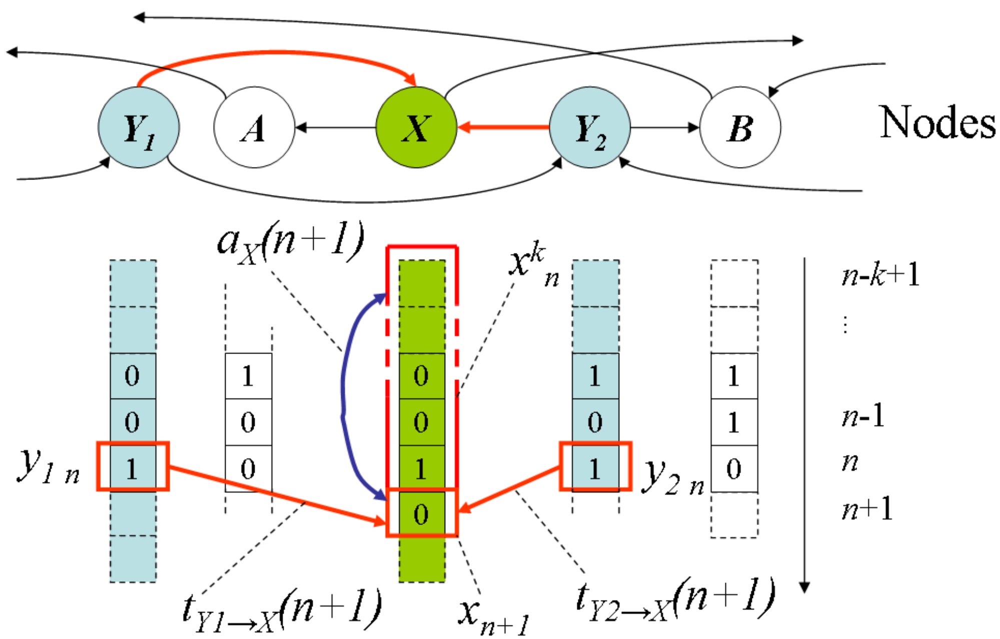
Local transfer entropy (Lizier et al., 2008):

$$t_{Y\rightarrow X} = \log \frac{p(x_{n+1} | \boldsymbol{x_n^{(k)}}, \boldsymbol{y_n^{(l)}})}{p(x_{n+1} | \boldsymbol{x_n^{(k)}})}$$

Information from a specific observation about the specific next value

Conditional transfer entropy

- What about accounting for other sources?
  - Conditioning removes redundancies and includes synergies



Conditional TE (Lizier et al., 2008, 2010):

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

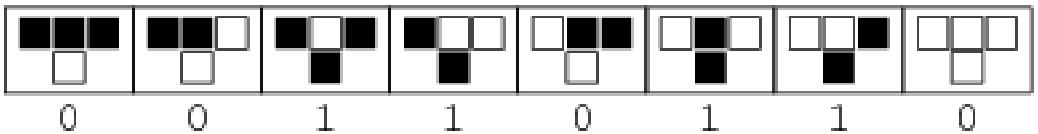
$$= \left\langle \log \frac{p(x_{n+1}|\boldsymbol{x}_{\boldsymbol{n}}^{(\boldsymbol{k})}, \boldsymbol{y}_{\boldsymbol{n}}^{(\boldsymbol{l})}, z_n)}{p(x_{n+1}|\boldsymbol{x}_{\boldsymbol{n}}^{(\boldsymbol{k})}, z_n)} \right\rangle$$

Average info from source that helps predict next value in context of past and conditional.

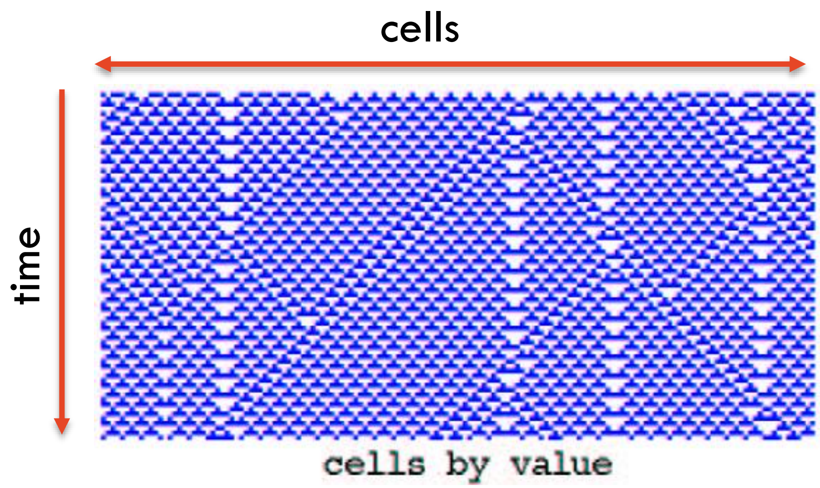
Extend information regression:

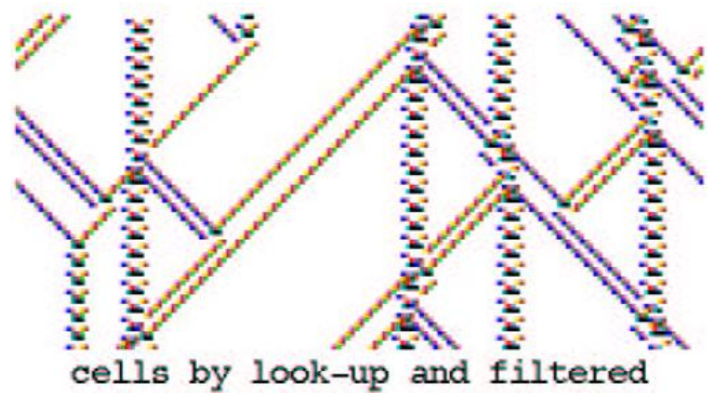
$$H_X = A_X + T_{Y\to X} + T_{Z\to X|Y} + \cdots$$

Cellular automata



CAs: simple dynamical systems; known causal structures and rules





(Wuensche, 1999)

Emergent structure:

- Domain, blinkers
- Particles
  - Gliders, domain walls
- Particle collisions

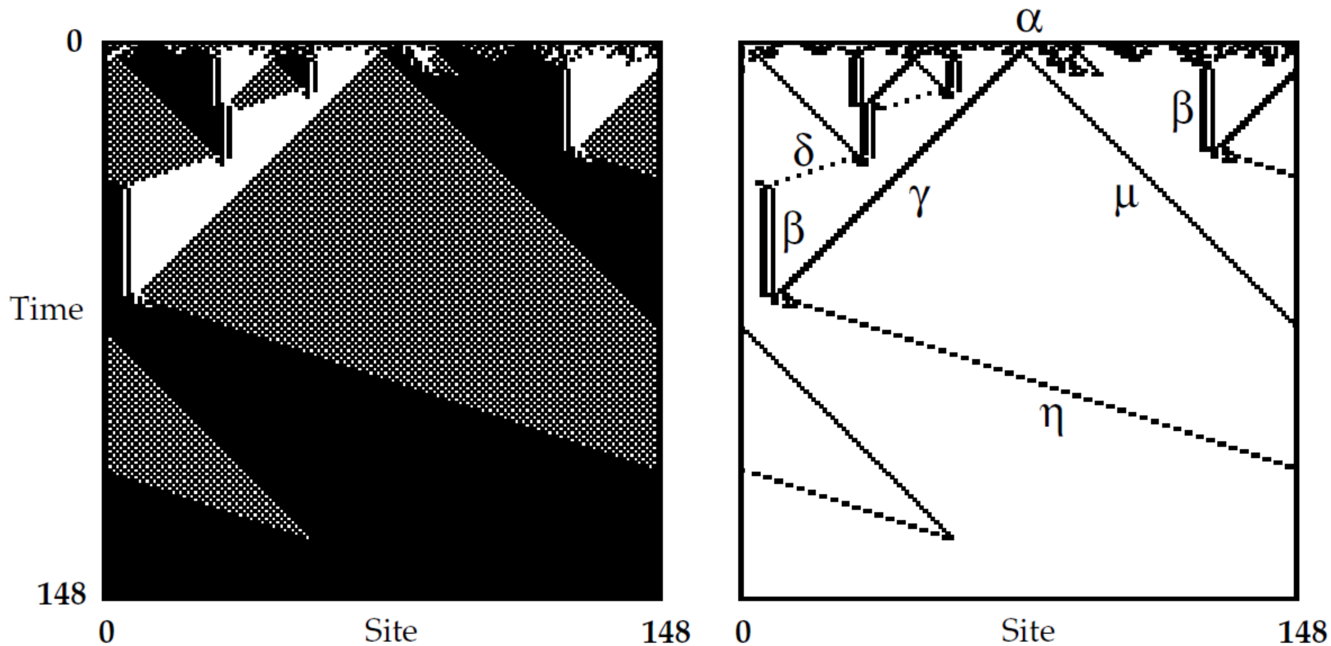
Conjectured to represent:

- Information storage
- Information transfer
  - “
- Information modification

It’s easy to identify which components store, transfer and modify information in a PC – it’s not so easy in complex systems.

(Langton, 1990)

Cellular automata

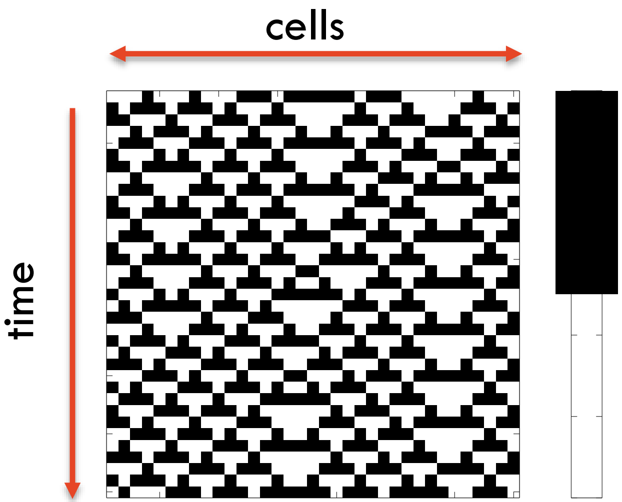


Mitchell et al. (1994, 1996) used GAs to evolve CAs to solve specific computational tasks.

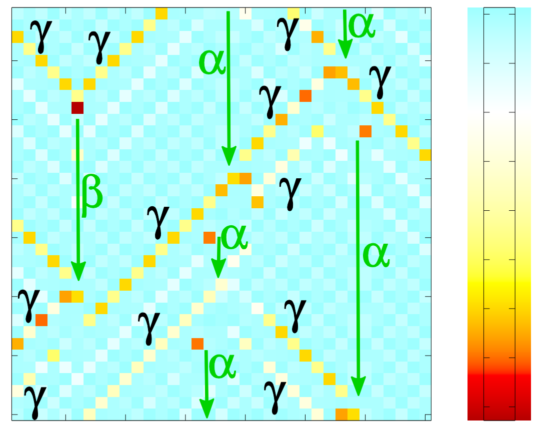
In attempting the density classification task (above), the CA uses:

- domains and blinkers β to store information;
- gliders γ,η to transfer information;
- glider collisions e.g. $\gamma + \beta \rightarrow \eta$ to modify/process information

Application to cellular automata (rule 54)



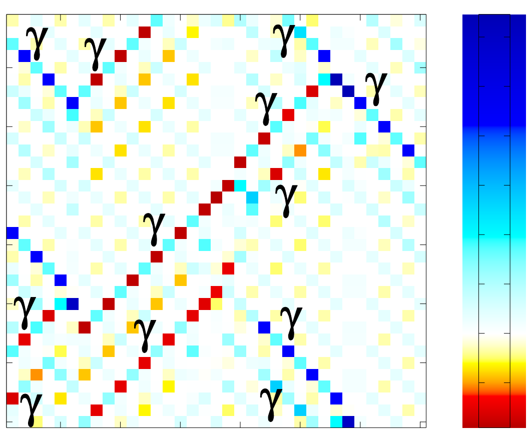
(a) Raw CA

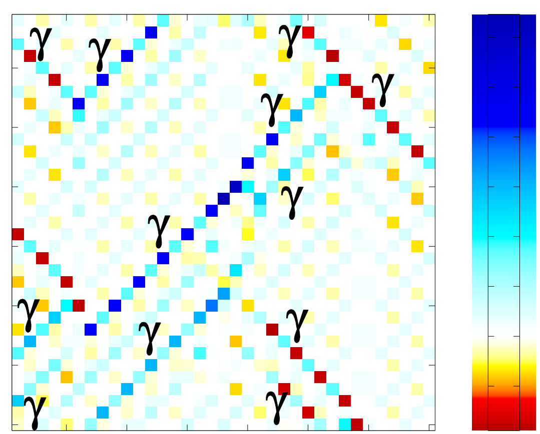

(b) AIS

Informative storage during regular patterns (domains and blinkers);

Misinformative storage at gliders, with change in phase or pattern of activity

JIDT Lizier et al. (2008-14)

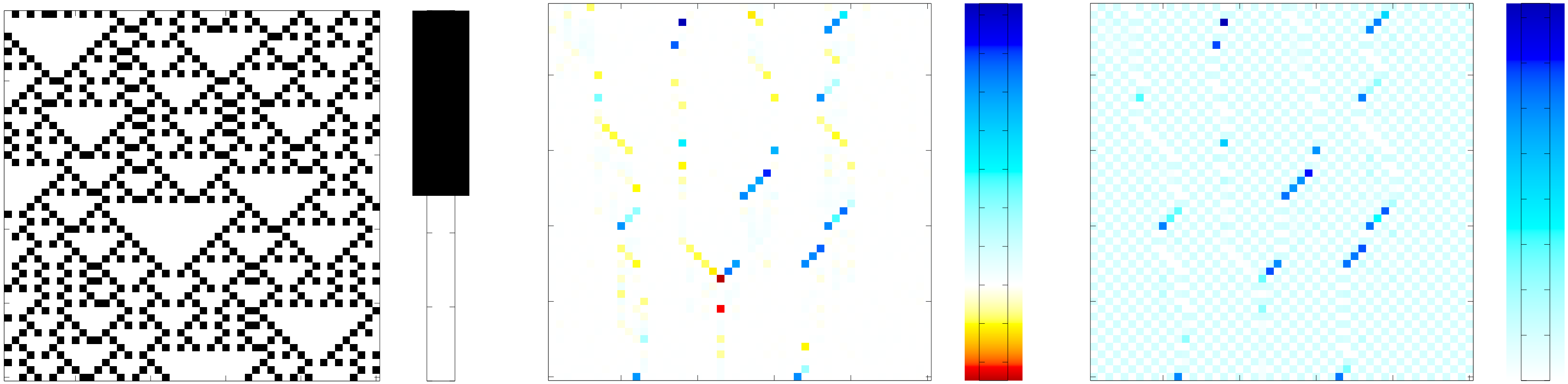

(c) TE (to right)


(d) TE (to left)

Gliders are the dominant information transfer entities.

Misinformative transfer in opposite direction

Application to cellular automata (rule 18)



(a) Raw CA (b) TE left (c) Conditional TE left

– Conditional TE accounts for other sources – removes redundancies and includes synergies – can be very different

JIDT Lizier et al. (2008-14)

Causal information flow

– Requires interventions or perturbations (Pearl, 2000)

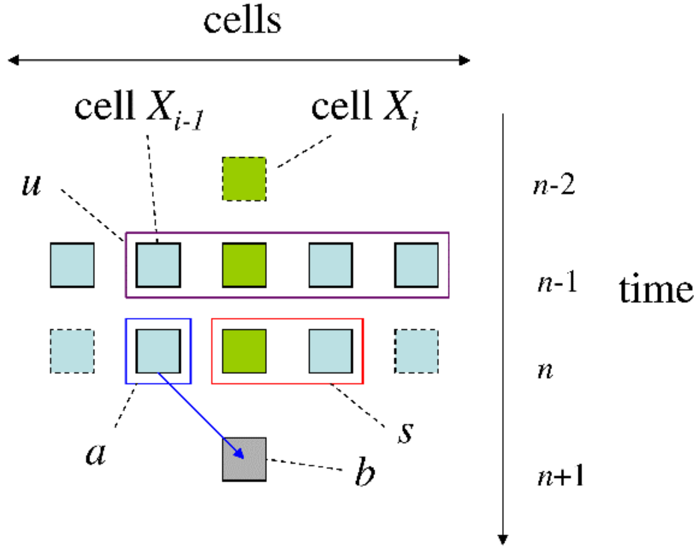
– Interventional probabilities: $p(a|\hat{s})$
  – Distribution of a given that we have intervened to impose $\hat{s}$
  – E.g. Consider a and s with a common (sole) cause:
    - $p(a|s) \neq p(a)$ in general, but
    - $p(a|\hat{s}) = p(a)$
  – Can measure from known system dynamics, results of intervention, or observations using e.g. back-door adjustment

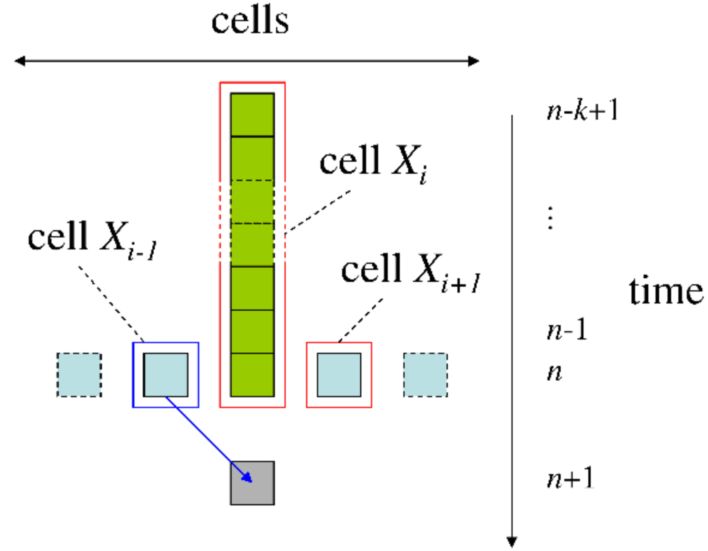
Causal information flow (Ay and Polani, 2008)

– Measures deviation of target B from causal independence on the source A, imposing conditionals S.

$$I_p\left(A \to B \middle| \hat{S}\right) = \sum_{s} p(s) \sum_{a} p(a|\hat{s}) \sum\nolimits_{b} p(b|\hat{a}, \hat{s}) \ \log \frac{p(b|\hat{s}, \hat{a})}{\sum_{a'} p(a|\hat{s}) p(b|\widehat{a'}, \hat{s})}$$

– Contrast to conditional transfer entropy, which measures deviation of target from stochastic independence on the source, given conditionals





Local causal information flow (Lizier and Prokopenko, 2010)

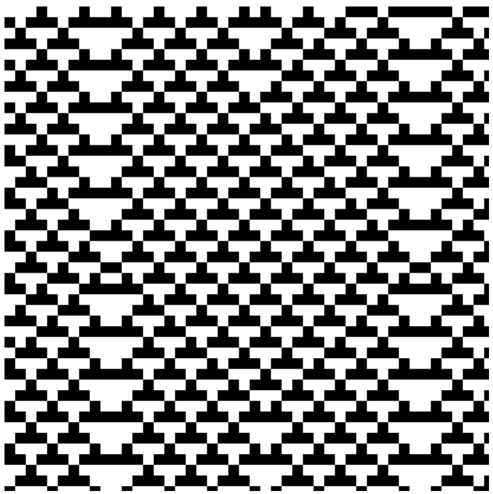
– Measures, for a given sample (*b*,*a*,*s*), deviation of target *b* from causal independence on the source *a*, imposing conditionals *s*.

$$f(a \to b|\hat{s}) = \log \frac{p(b|\hat{s},\hat{a})}{\sum_{a'} p(a|\hat{s})p(b|\widehat{a'},\hat{s})}$$

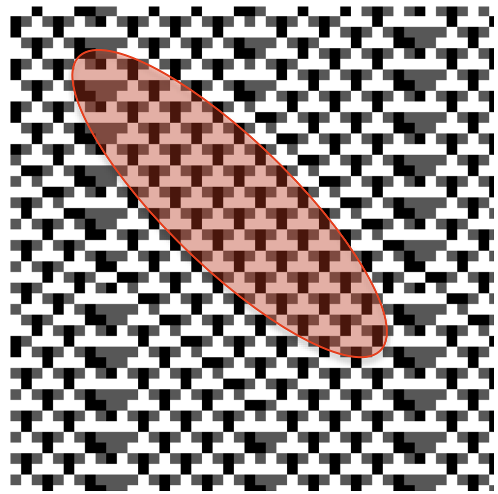
– Contrast to conditional transfer entropy which measures deviation from stochastic independence for a given sample.

– Is an attribution of the causal information flow in a given sample, however *f* does not average to I_p

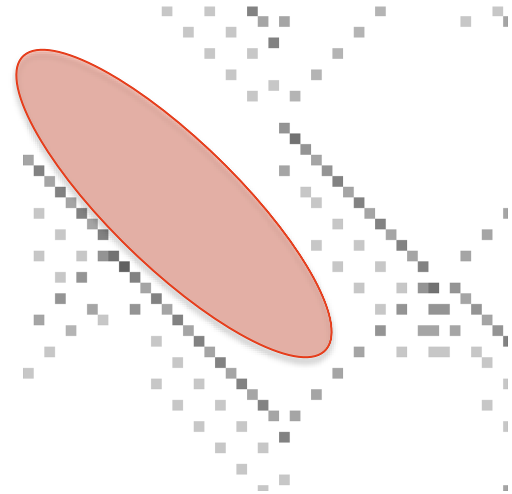
1. Causal connections can embody information storage rather than transfer



(a) Raw CA



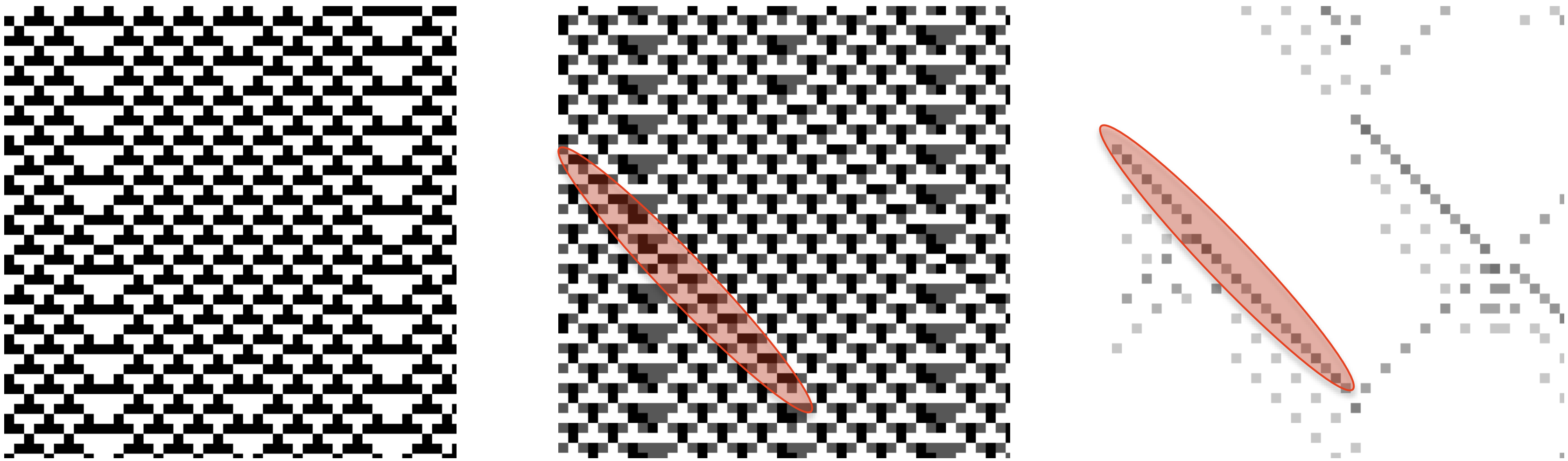
(b) Flow right



(c) Conditional TE right

- Regular periodic background domain and blinkers have:
  - Low transfer entropy, high information storage
  - Non-distinct causal information flow
- Results are correct from different perspectives:
  - Backgrounds/blinkers are same configurations as in gliders, so just as causal
  - ➢ Causal connections can support information storage as well as / instead of transfer
  - ➢ Transfer entropy does not detect all causal interactions that flow does

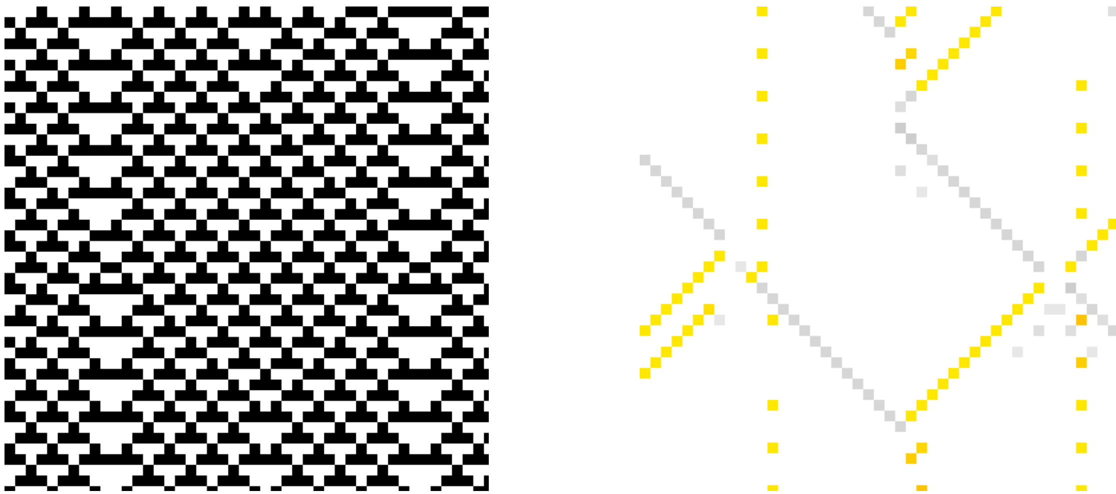
2. Emergent transfer structures are not strongly causal



(a) Raw CA (b) Flow right (c) Conditional TE right

- Gliders structures have:
  - High transfer entropy, low information storage
  - Non-distinct causal information flow
- Results are correct from different perspectives:
  - Gliders are same configurations as in backgrounds, so just as causal
  - ➢ Information flow does not distinguish emergent computational structure
  - ➢ Context of past gives transfer entropy the computational perspective

3. Transfer entropy is not causal

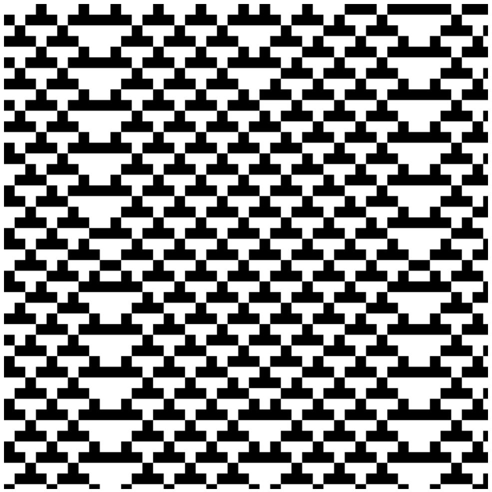


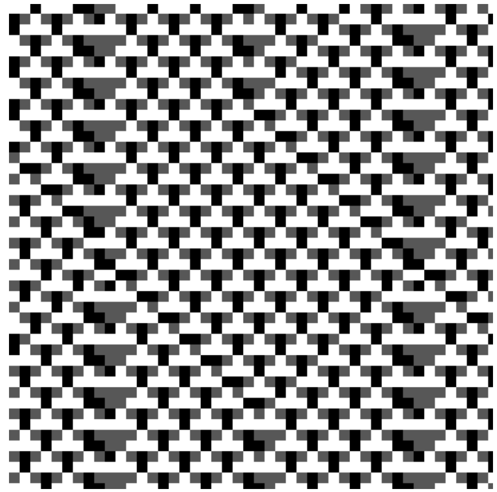
(a) Raw CA (b) TE 2 cells to right

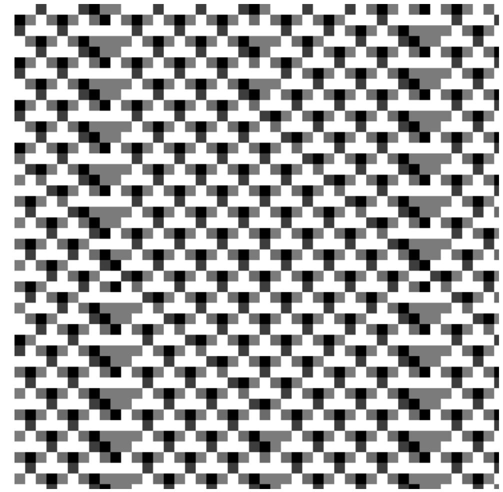
– Transfer entropy is still high for gliders for 2 steps to right / time step (outside "light cone")
  – It is a predictive measure, not causal

➢ To be interpreted as information transfer, TE should be measured from causal sources

4. Complete transfer entropy to infer causal information flow


(a) Raw CA


(b) Flow right


(c) Conditional TE right, k=1

– Convergence occurs (as ~here) when:
  – TE parameter *k* is set within causal structure
  – All {*a*,*s*} combinations are observed (required for averages only)
  – $p(a|\hat{s})=p(a|s)$ (close to being met here)

➢ Where one cannot intervene or use inference such as back-door, conditional TEs offer an alternative
  – E.g. Inferring effective networks in neural data; see Bossomaier et al. (2016) sec. 7.2

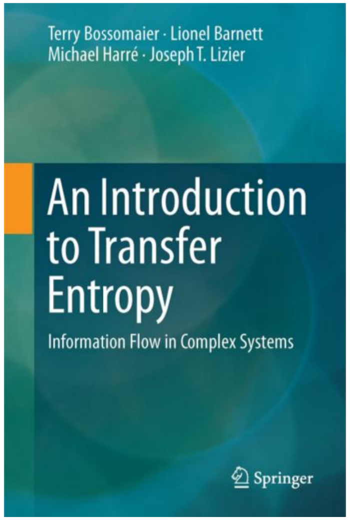
Conclusion

- Concepts of causal information flow and information transfer are equal citizens.
- They are captured by causal information flow and transfer entropy respectively.
- They are complementary as follows:
  - Causal information flow is a primary tool to establish system structure,
  - Transfer entropy reveals information transfer in emergent computation, along with information storage and modification.

Advertisements!

- Java Information Dynamics Toolkit (JIDT) – http://jlizier.github.io/jidt/

- *"An Introduction to Transfer Entropy: Information Flow in Complex Systems"*, Terry Bossomaier, Lionel Barnett, Michael Harré and Joseph T. Lizier, Springer, 2016.



- Master of Complex Systems at Usyd – starting in 2017!

Questions

References

– N. Ay and D. Polani. Information Flows in Causal Networks. Advances in Complex Systems, 11(1):17–41, 2008.
– T. Bossomaier, L. Barnett, M. Harré, J. T. Lizier, An Introduction to Transfer Entropy: Information Flow in Complex Systems, Springer, Cham, 2016.
– R. G. James, N. Barnett, and J. P. Crutchfield. Information flows? a critique of transfer entropies. Physical Review Letters, 116:238701+, 2016
– C. G. Langton. Computation at the Edge of Chaos. Physica D 42, 12-37, 1990.
– J. T. Lizier, M. Prokopenko, and A. Y. Zomaya. Local information transfer as a spatiotemporal filter for complex systems. Physical Review E, 77(2):026110+, 2008.

References

– J. T. Lizier, M. Prokopenko, and A. Y. Zomaya. Information modification and particle collisions in distributed computation. Chaos, 20(3):037109+, 2010
– J.T. Lizier and M. Prokopenko, Differentiating information transfer and causal effect. European Physical Journal B, 73(4), pp. 605-615, 2010
– J. T. Lizier, M. Prokopenko, and A. Y. Zomaya. Local measures of information storage in complex distributed computation. Information Sciences, 208:39–54, 2012.
– J.T. Lizier. JIDT: An information-theoretic toolkit for studying the dynamics of complex systems. Frontiers in Robotics & AI, 1:11, 2014.

References

- M. Mitchell, J. P. Crutchfield, and P. T. Hraber. Evolving Cellular Automata to Perform Computations: Mechanisms and Impediments. Physica D, 75:361–391, 1994.
- M. Mitchell, J. P. Crutchfield, and R. Das. Evolving Cellular Automata with Genetic Algorithms: A Review of Recent Work. In E. D. Goodman, W. Punch, and V. Uskov, editors, Proceedings of the First International Conference on Evolutionary Computation and Its Applications, Moscow, Russia, 1996. Russian Academy of Sciences.
- J. Pearl. Causality: Models, Reasoning, and Inference. Cambridge University Press, Cambridge, 2000.
- T. Schreiber. Measuring Information Transfer. Physical Review Letters, 85(2): 461–464, 2000.

References

– A. Wuensche. Classifying cellular automata automatically: Finding gliders, filtering, and relating space-time patterns, attractor basins, and the Z parameter. Complexity, 4(3):47–66, 1999