Early Warning and Diagnosis in Complex High Dimensional Systems

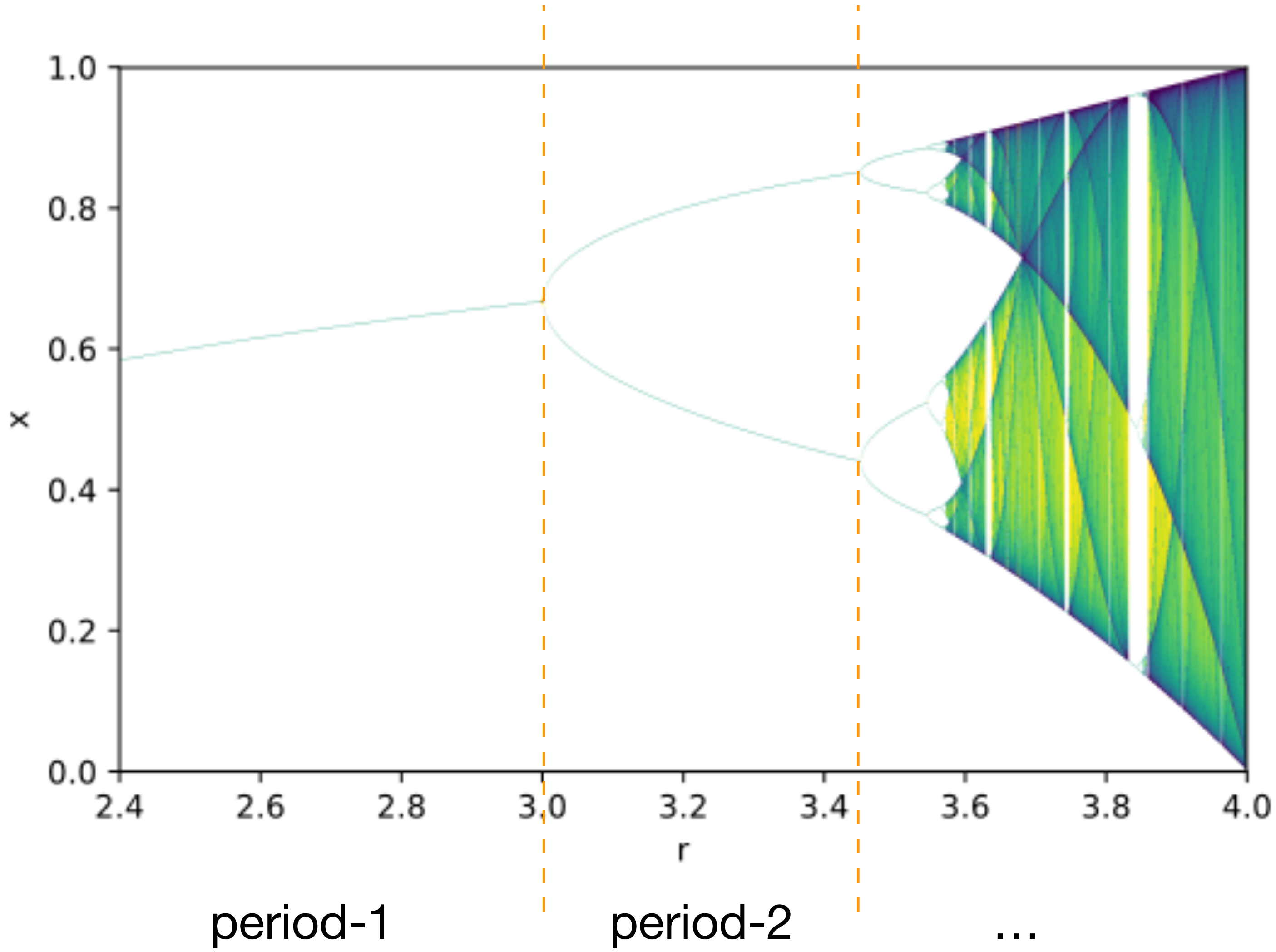
Jie Sun
Department of Mathematics
Clarkson Center for Complex Systems Science (C3S2)
Clarkson University
Potsdam, New York 13699, USA

joint work with Erik Bollt (Clarkson)

Bifurcation, Tipping Point, Etc.

Example: bifurcation diagram of the logistic map

$$x_{t+1} = r x_t (1 - x_t)$$



Bifurcation types

saddle-node

transcritical

pitchfork

Hopf

...

$$\dot{x} = F(x; \mu) \longrightarrow \dot{x} = F(x; \tilde{\mu})$$

Small change of a system's parameter can cause large change of its dynamics

Detecting Bifurcation in Time Series(?)

Vol 461|3 September 2009|doi:10.1038/nature08227 nature

REVIEWS

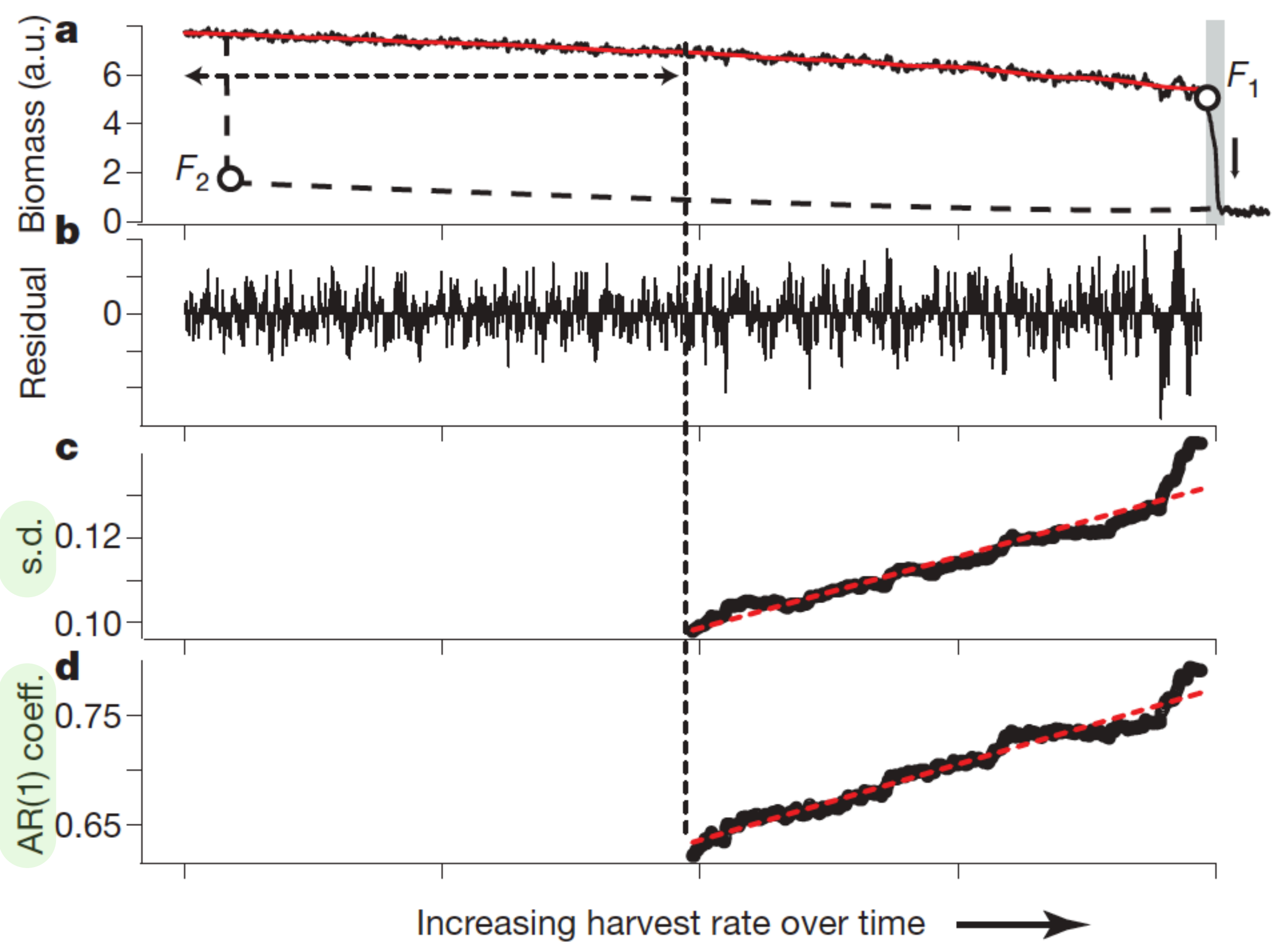
Early-warning signals for critical transitions

Marten Scheffer[1], Jordi Bascompte[2], William A. Brock[3], Victor Brovkin[5], Stephen R. Carpenter[4], Vasilis Dakos[1], Hermann Held[6], Egbert H. van Nes[1], Max Rietkerk[7] & George Sugihara[8]

Idea: track a (scalar) function of the time series
—> Early-Warning Signals (EWS)

scalar time series

EWS



Application of EWS in Detecting Changes

M. Scheffer et al., *Early-warning signals for critical transitions*, Nature **461**, 53-59 (2009)

Examples of EWS: (increased) AR(1), increased variance, etc.

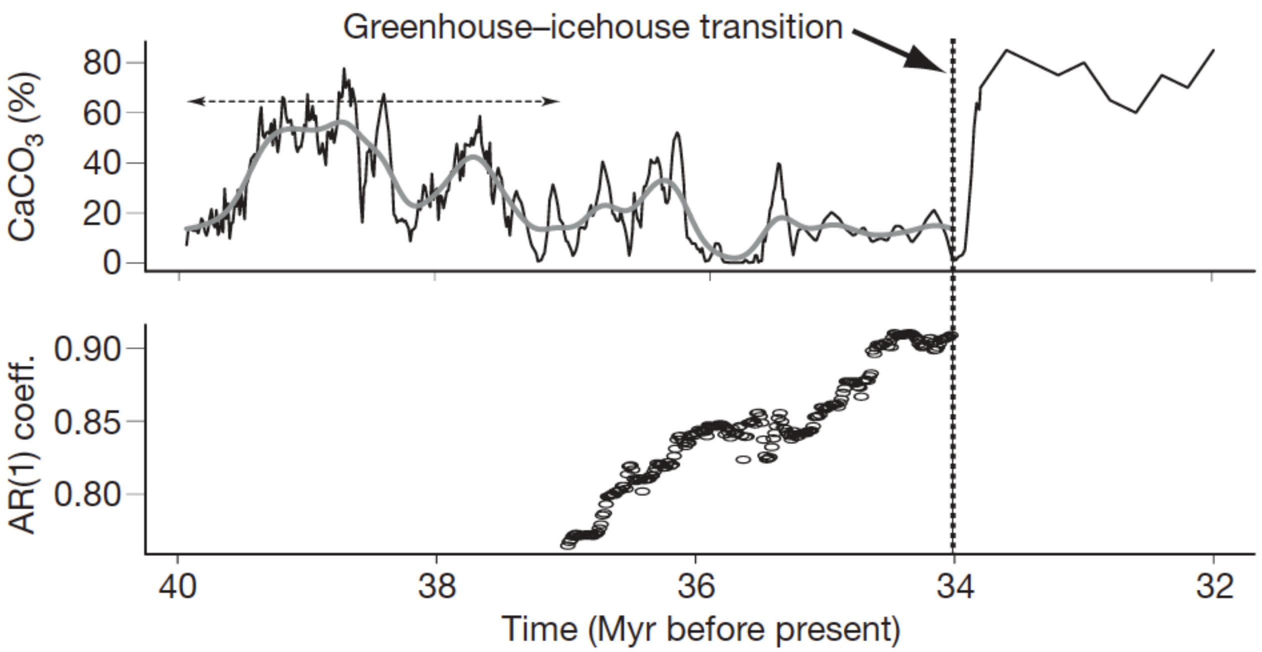


Figure 4 | Critical slowing down indicated by an increase in lag-1 autocorrelation in climate dynamics. We show the period preceding the transition from a greenhouse state to an icehouse state on the Earth 34 Myr ago. The trends in the $CaCO_3$ concentration time series removed by filtering before computing autocorrelation (AR(1) coefficient) are represented by the grey line. The horizontal dashed arrow shows the width of the moving window used to compute the autocorrelation. Modified from ref. 22.

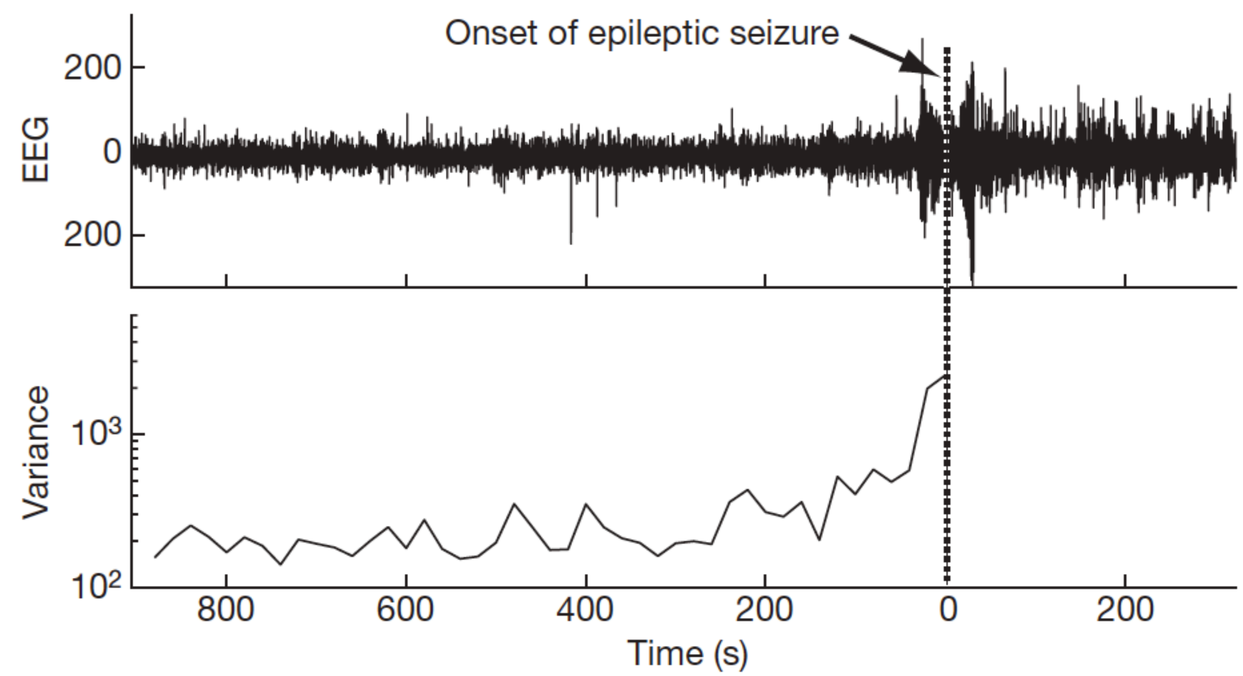
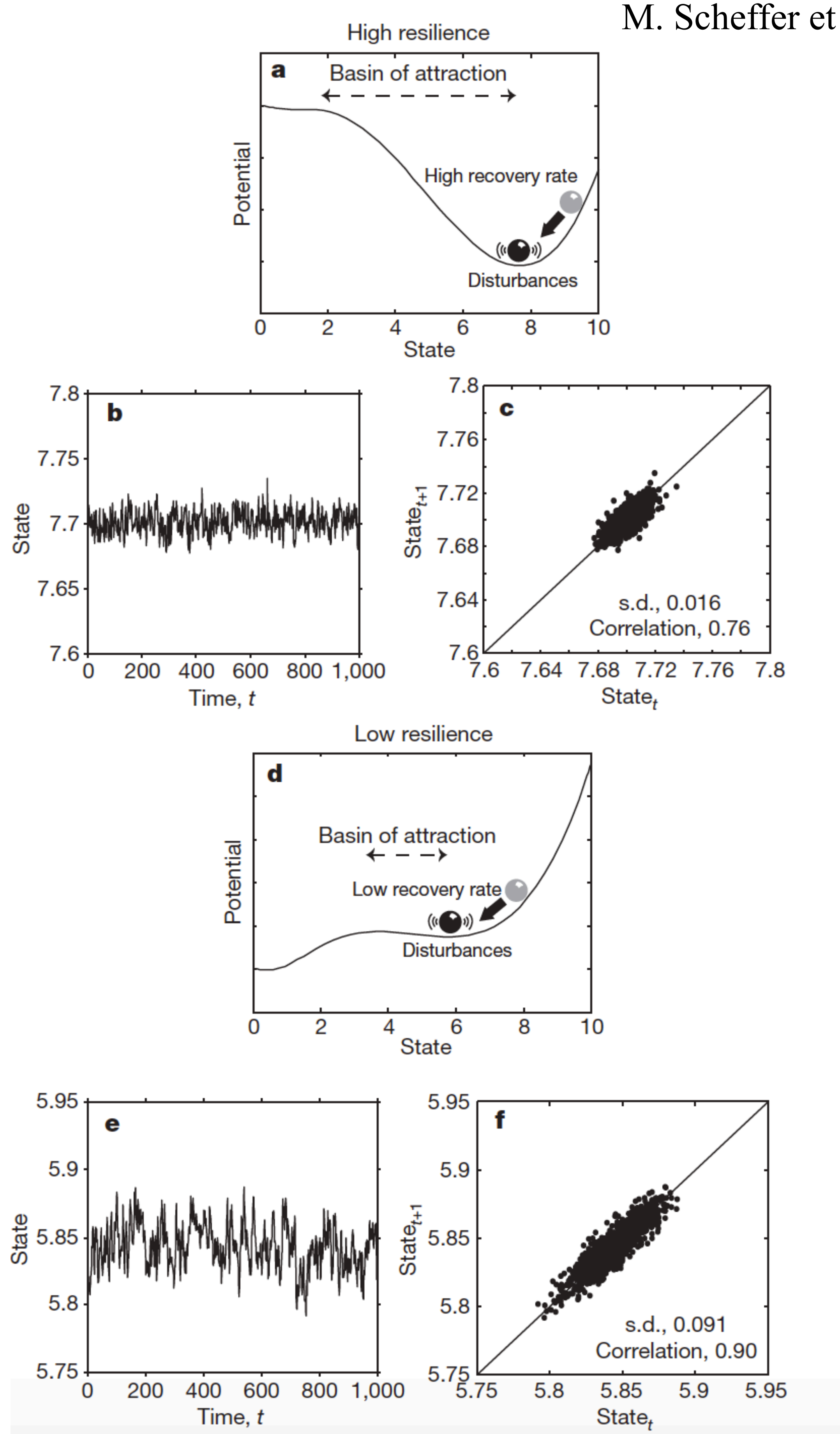


Figure 5 | Subtle changes in brain activity before an epileptic seizure may be used as an early warning signal. The epileptic seizure clinically detected at time 0 is announced minutes earlier in an electroencephalography (EEG) time series by an increase in variance. Adapted by permission from Macmillan Publishers Ltd: Nature Medicine (ref. 3), copyright 2003.

M. Scheffer et al., *Early-warning signals for critical transitions*, Nature **461**, 53-59 (2009)



Box 3 | The relation between critical slowing down, increased autocorrelation and increased variance

Critical slowing down will tend to lead to an increase in the autocorrelation and variance of the fluctuations in a stochastically forced system approaching a bifurcation at a threshold value of a control parameter. The example described here illustrates why this is so. We assume that there is a repeated disturbance of the state variable after each period Δt (that is, additive noise). Between disturbances, the return to equilibrium is approximately exponential with a certain recovery speed, λ. In a simple autoregressive model this can be described as follows:

$$x_{n+1}-\bar{x}=\mathrm{e}^{\lambda\Delta t}(x_n-\bar{x})+\sigma\varepsilon_n$$

$$y_{n+1}=\mathrm{e}^{\lambda\Delta t}y_n+\sigma\varepsilon_n$$

Here y_n is the deviation of the state variable x from the equilibrium, ε_n is a random number from a standard normal distribution and σ is the standard deviation.

If λ and Δt are independent of y_n, this model can also be written as a first-order autoregressive (AR(1)) process:

$$y_{n+1}=\alpha y_n+\sigma\varepsilon_n$$

The autocorrelation $\alpha\equiv \mathrm{e}^{\lambda\Delta t}$ is zero for white noise and close to one for red (autocorrelated) noise. The expectation of an AR(1) process $y_{n+1}=c+\alpha y_n+\sigma\varepsilon_n$ is[18]

$$\mathrm{E}(y_{n+1})=\mathrm{E}(c)+\alpha\mathrm{E}(y_n)+E(\sigma\varepsilon_n)\Rightarrow\mu=c+\alpha\mu+0\Rightarrow\mu=\frac{c}{1-\alpha}$$

For c = 0, the mean equals zero and the variance is found to be

$$\mathrm{Var}(y_{n+1})=\mathrm{E}(y_n^2)-\mu^2=\frac{\sigma^2}{1-\alpha^2}$$

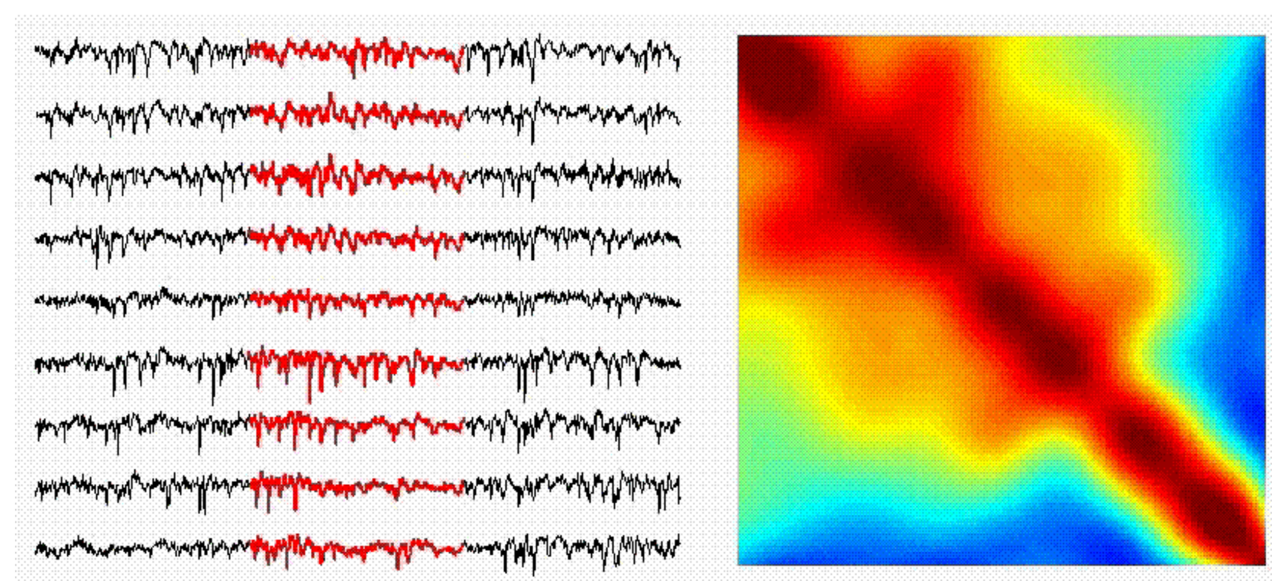
Close to the critical point, the return speed to equilibrium decreases, implying that λ approaches zero and the autocorrelation α tends to one. Thus, the variance tends to infinity. These early-warning signals are the result of critical slowing down near the threshold value of the control parameter.

Flocking of birds

Millennium Bridge

Power grids

Brain Seizure



Networks play a central role in the dynamics and functioning of a system.

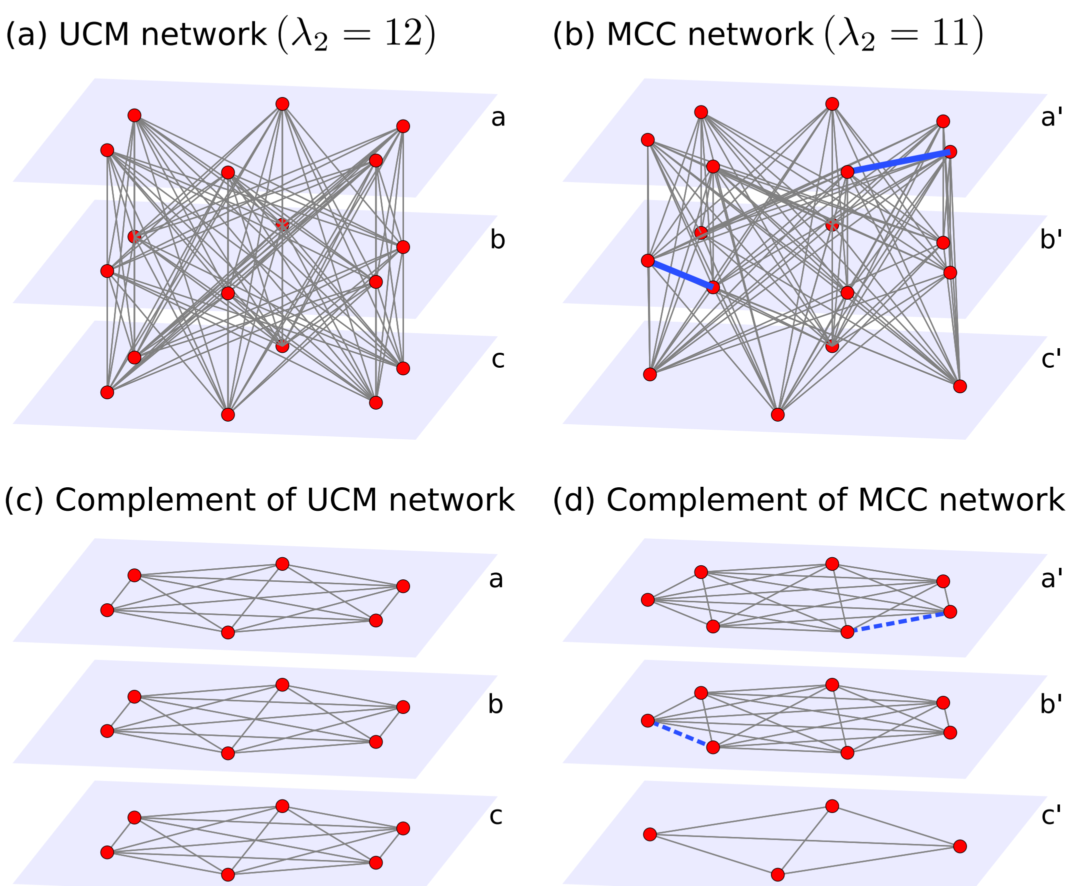
PHYSICAL REVIEW X **7,** 041044 (2017)

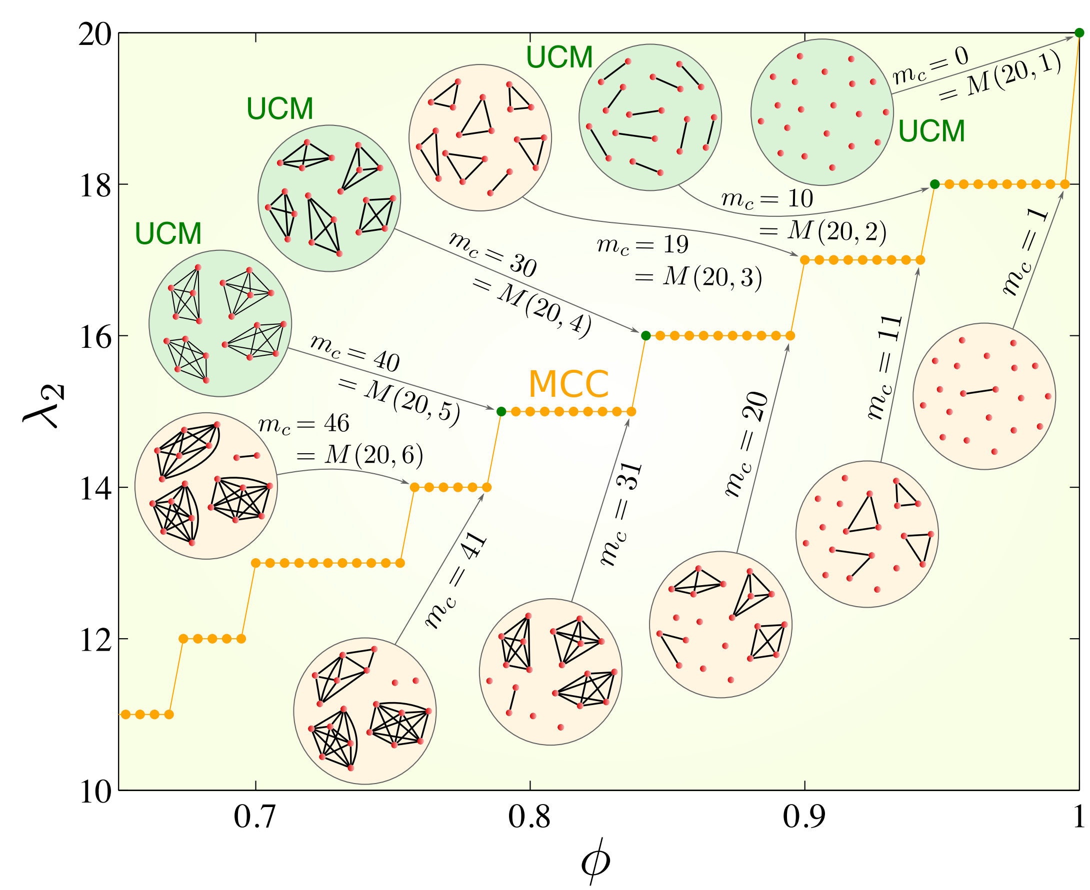
Sensitive Dependence of Optimal Network Dynamics on Network Structure

Takashi Nishikawa,[1] Jie Sun,[2] and Adilson E. Motter[1]

[1] *Department of Physics and Astronomy, Northwestern University, Evanston, Illinois 60208, USA and Northwestern Institute on Complex Systems, Northwestern University, Evanston, Illinois 60208, USA*

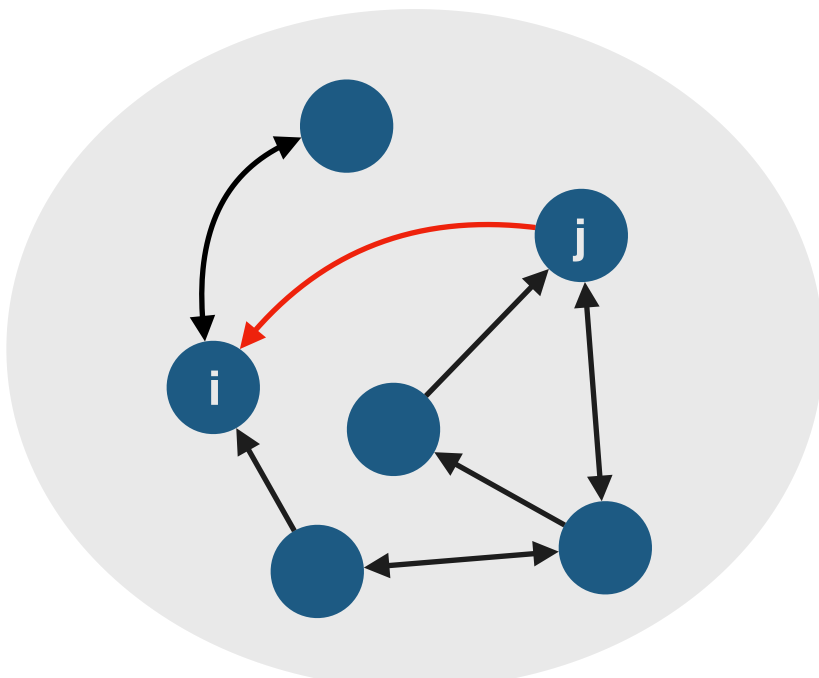
[2] *Department of Mathematics, Clarkson University, Potsdam, New York 13699, USA, Department of Physics, Clarkson University, Potsdam, New York 13699, USA, Department of Computer Science, Clarkson University, Potsdam, New York 13699, USA, and Clarkson Center for Complex Systems Science, Clarkson University, Potsdam, New York 13699, USA*



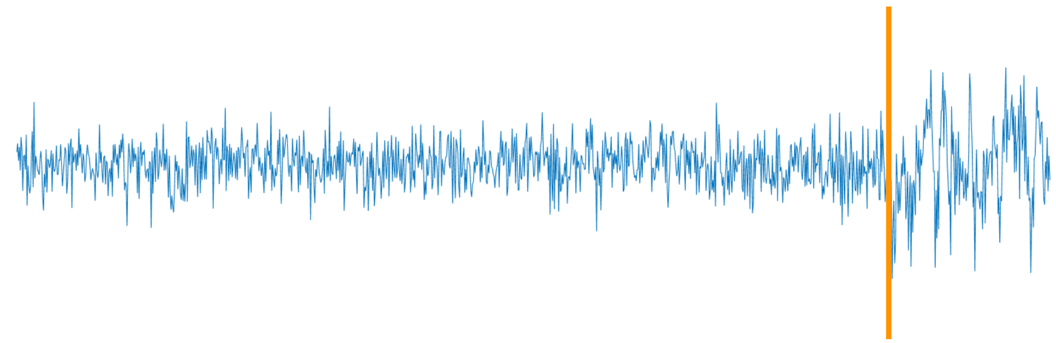


Small change in network structure can lead to large change in network dynamics.

Detecting Changes in a Networked System?



scalar time series —> EWS to detect change



What causes the change?

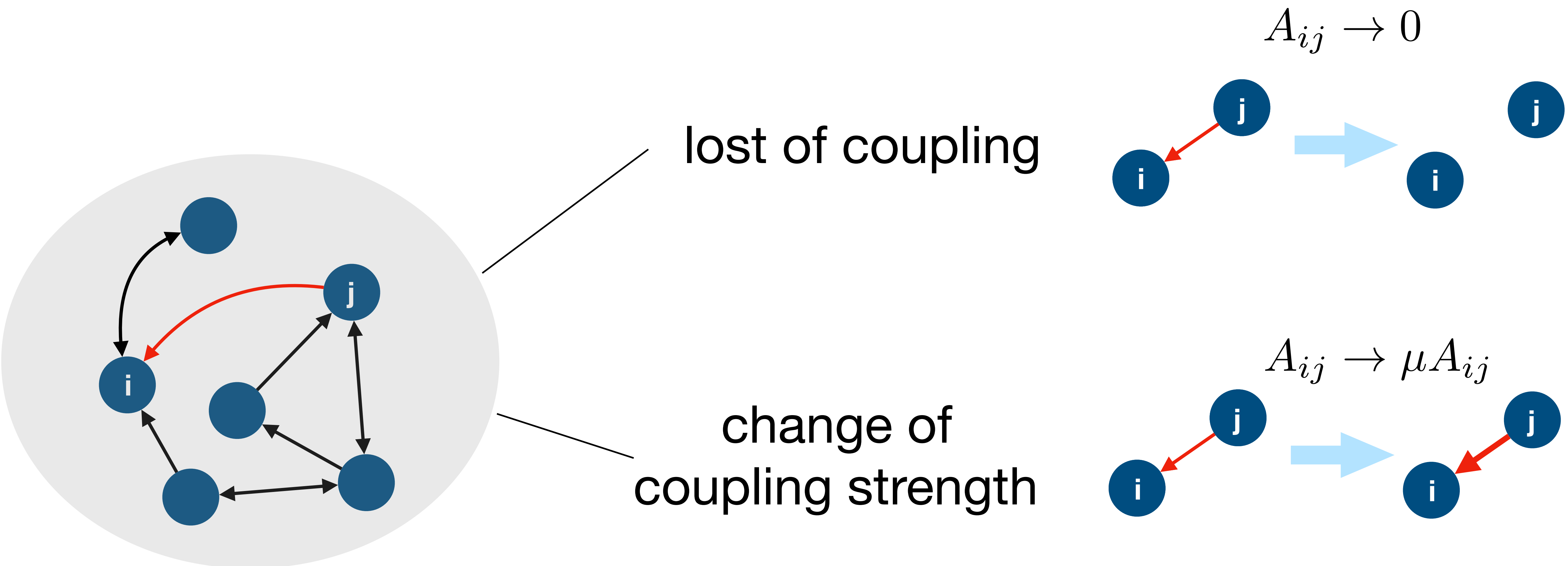
— broken connections
— changing coupling strength
— ...

EWS does not provide information on these questions.

Q: how to detect the coupling changes via analyzing the time series of network dynamics?

Challenges:
— *nonlinearity*
— *high dimensionality*
— *unknown models*

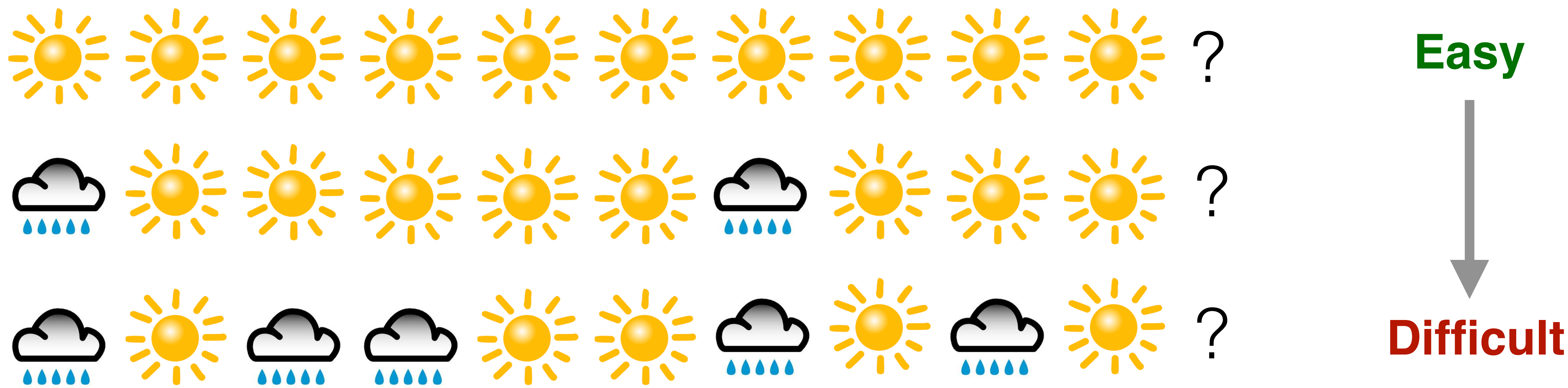
Detecting Changes in a Networked System?



Q: how to detect the coupling changes via analyzing the time series of network dynamics?

Information Theory in a Nutshell

Will it rain the next day? [Yes/no question.]



How difficult is it to predict if it is going to rain the next day?
— Characterization of (un)predicability.

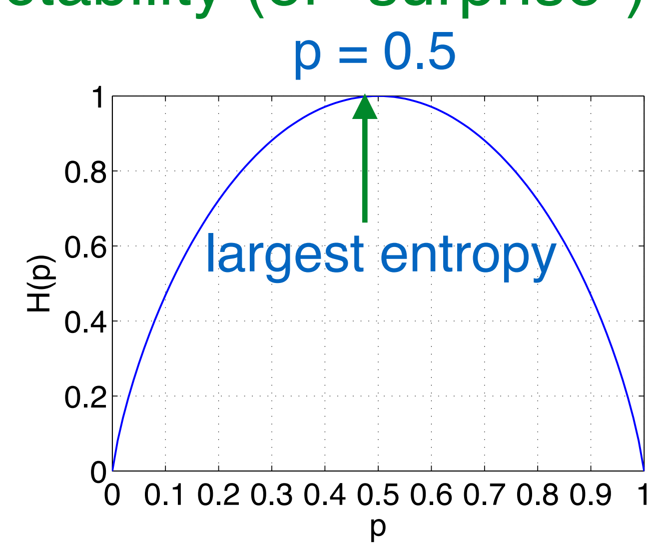
A Mathematical Theory of Communication, Claude E. Shannon (1948).

Entropy $H(X) = -\sum_{x} p(x) \log p(x)$

$$X = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1-p \end{cases}$$

Fair coin is most difficult to predict.

Entropy is a model-free measure of unpredictability (or “surprise”) of a r.v.



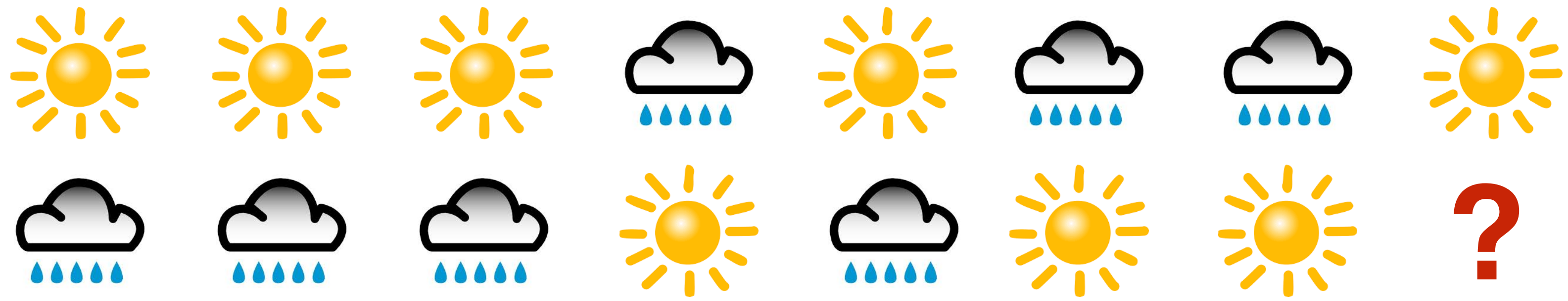
Information Theory in a Nutshell (cont.)

Example: **Will it rain the next day in city B?**

(1)

city A

city B



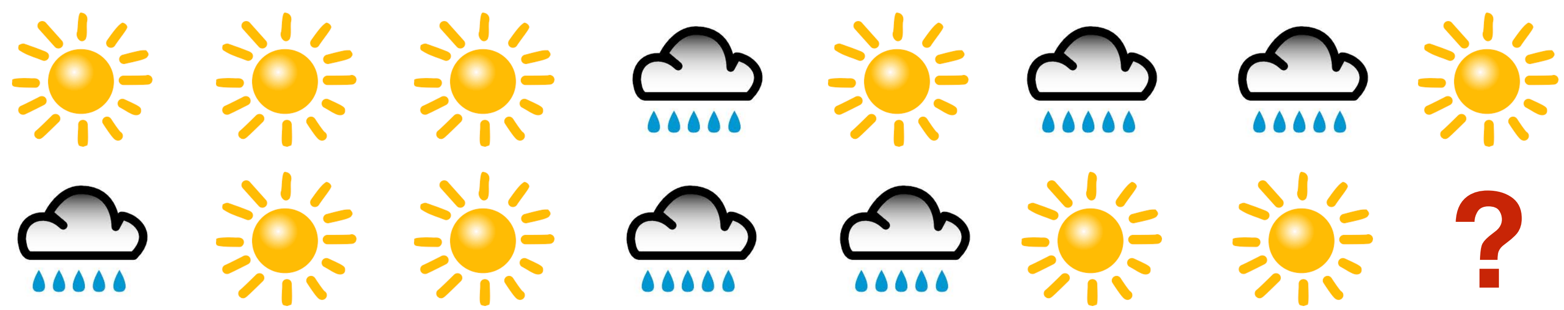
$H(Y|X) \approx 0$

Easy to predict Y once X is known.

(2)

city A

city B

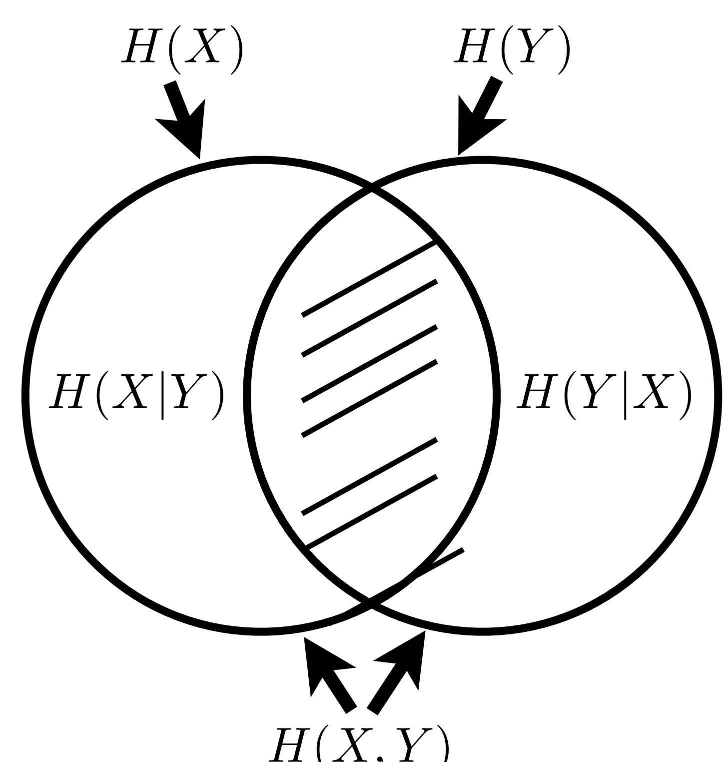


$H(Y|X) \approx H(Y)$

Knowing X does not help to predict Y.

Conditional entropy $H(X|Y) = H(X,Y) - H(Y)$

a measure of “given Y, how much does X remain unpredictable”



Beyond Linear Model: Transfer Entropy

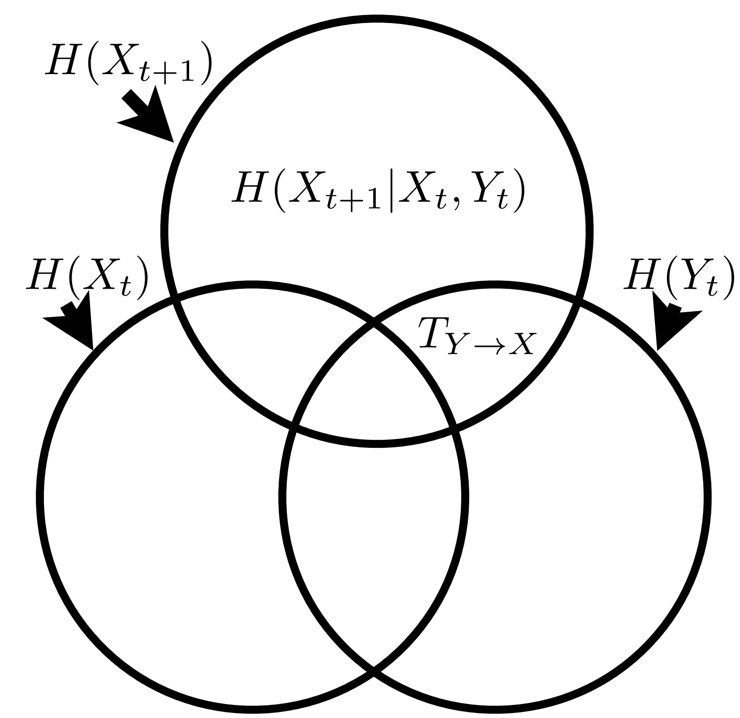
VOLUME 85, NUMBER 2 PHYSICAL REVIEW LETTERS 10 JULY 2000

Measuring Information Transfer

Thomas Schreiber

Idea: apply MI to detect information transfer / causality between two time series

consider two stochastic processes $\{X_t\}$ $\{Y_t\}$



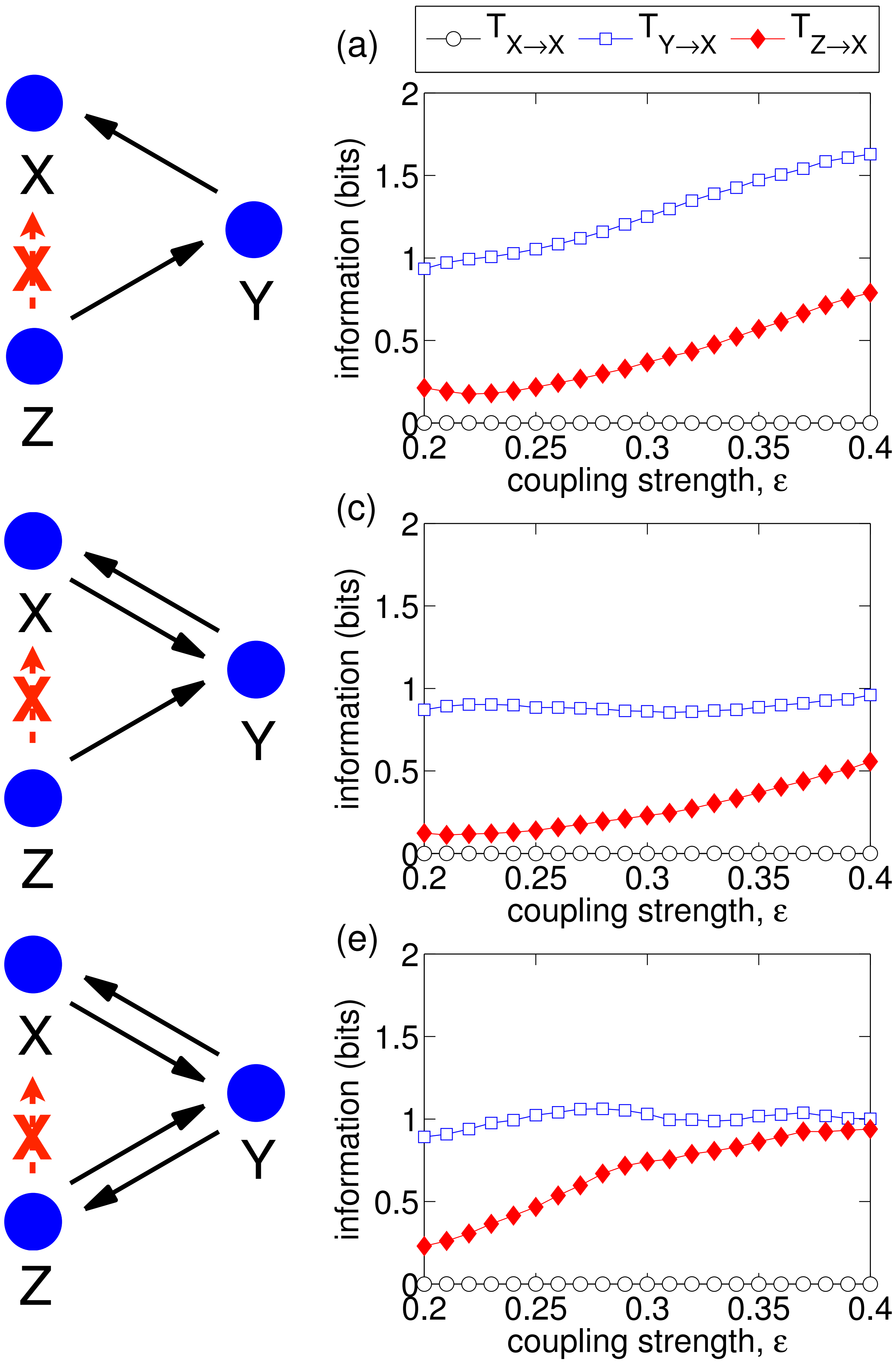
Transfer Entropy (TE)

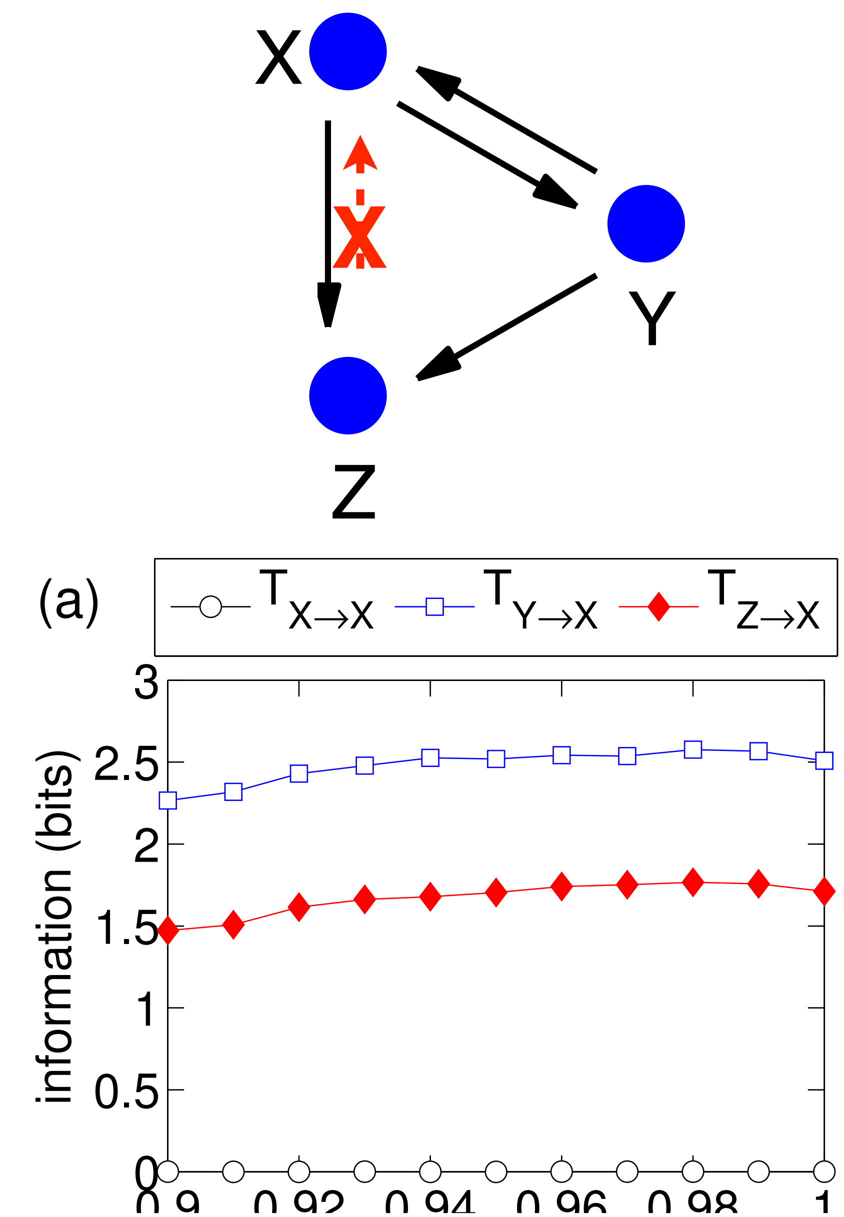
$$T_{Y\to X} = H(X_{t+1}|X_t) - H(X_{t+1}|X_t, Y_t)$$

measures the reduction of **uncertainty** regarding X's future by knowing Y

Systematic Bias in TE-based Network Inference

$$x_{t+1}^{(i)} = f[x_t^{(i)}] + \epsilon \sum_{j \neq i} c_{ij} g[x_t^{(i)}, x_t^{(j)}], \quad i = 1, 2, \ldots, N.$$





JS, E. Bollt, Physica D (2014).

TE-based network inference suffers from systematic bias (not just indirect links or false positives).

Causation Entropy: Measure of Causality in Networks

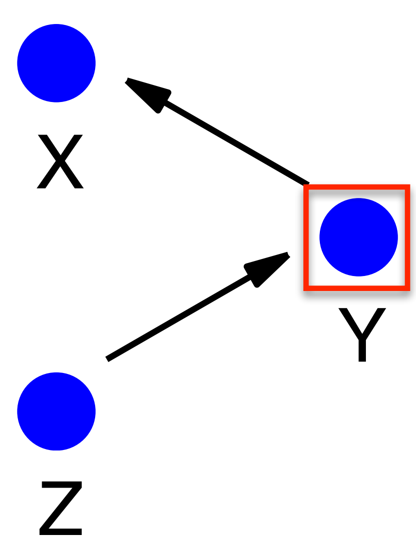
Definition 1 (Causation Entropy). *The causation entropy from process $\mathcal{Q}$ to process $\mathcal{P}$ conditioned on the set of processes $\mathcal{S}$ is defined as*

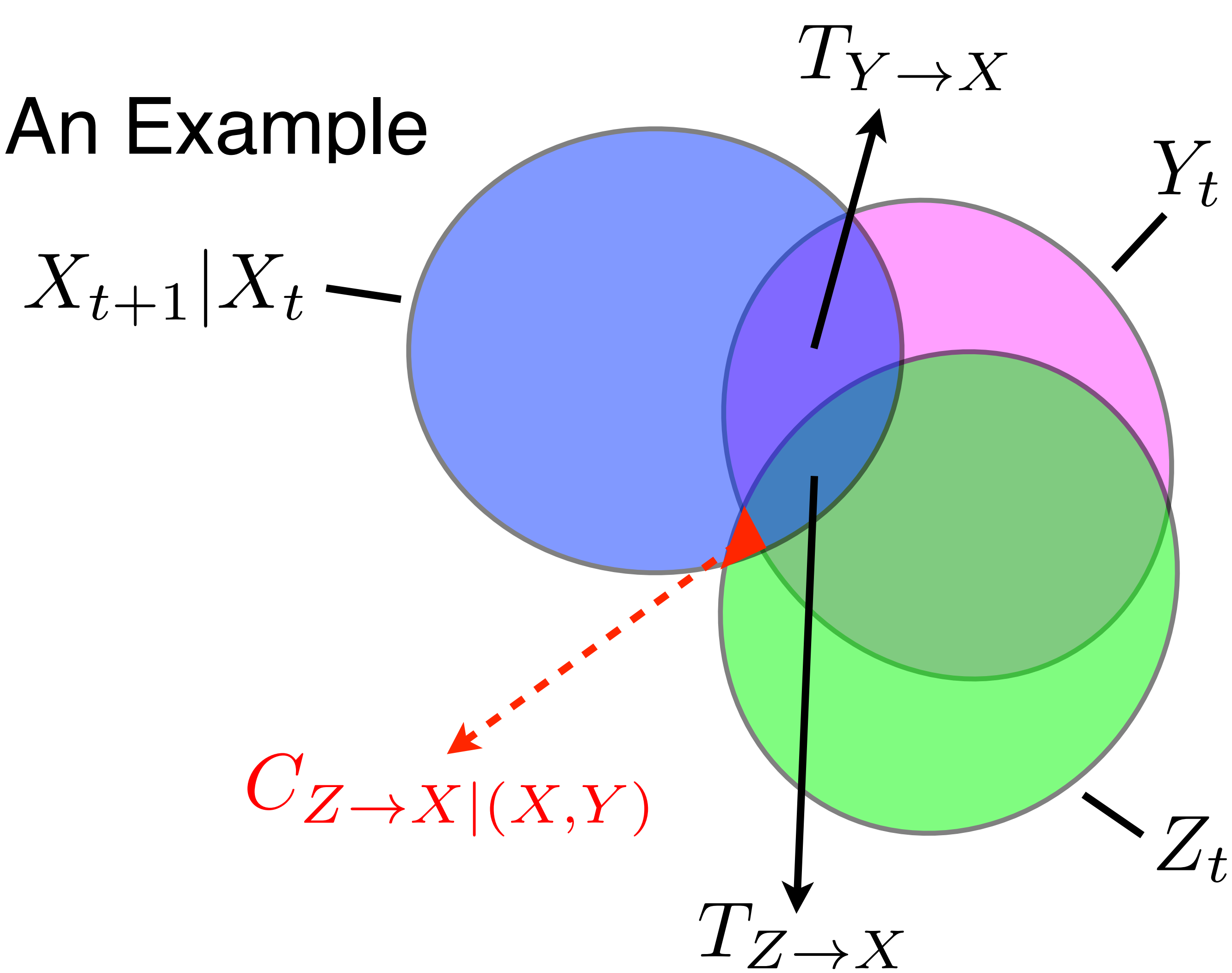
$$C_{\mathcal{Q}\to\mathcal{P}|(\mathcal{S})} = H(\mathcal{P}_{t+1}|\mathcal{S}_t) - H(\mathcal{P}_{t+1}|\mathcal{S}_t, \mathcal{Q}_t).$$

JS, E. Bollt, Physica D (2014).

uncertainty of P's future given S — **uncertainty of P's future given S and Q**

An Example

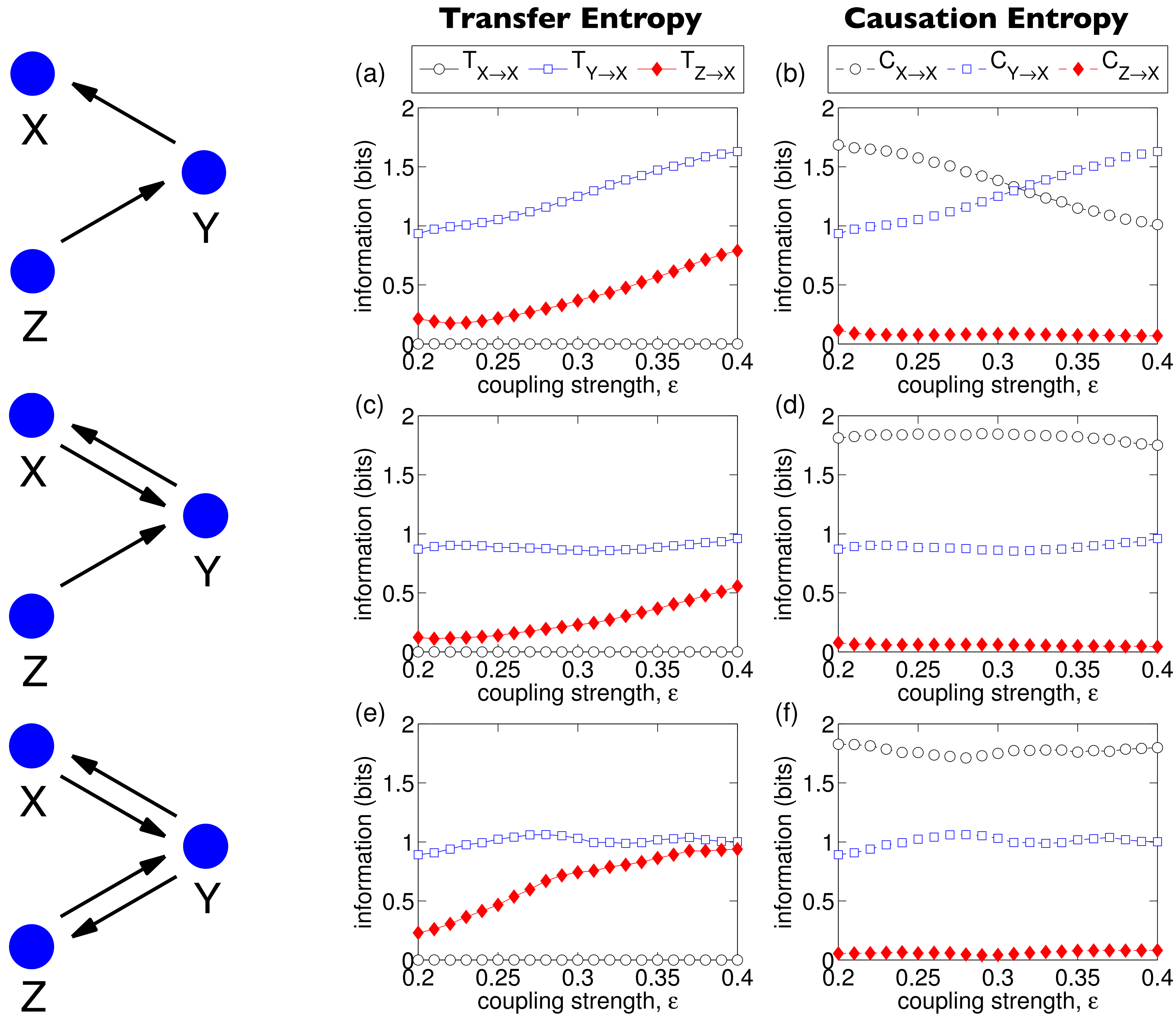




Remarks:

1. CSE itself does not "solve" the causal inference problem.
2. The definition simply emphasizes the fact that cause-and-effect involves all three parts (***cause***, ***effect***, and ***conditioning***).

Transfer Entropy (T) vs. Causation Entropy (C)

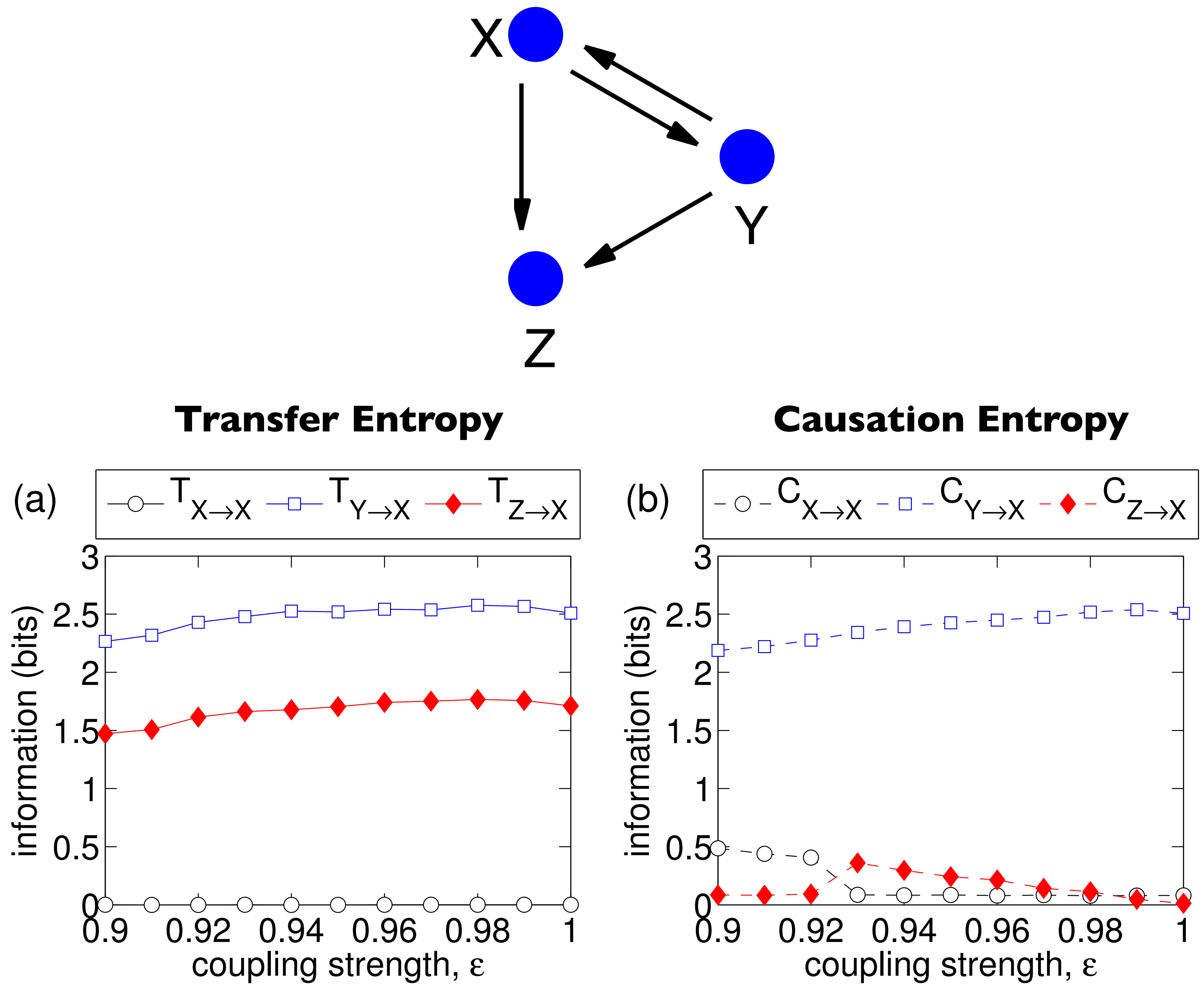


Causation entropy correctly identifies the causal network structure.

Transfer Entropy (T) vs. Causation Entropy (C)

JS and Erik Bollt, "*Causation entropy identifies indirect influences, dominance of neighbors and anticipatory couplings*", Physica D (2014)

JS, Dane Taylor, and Erik Bollt, "*Causal network inference by optimal causation entropy*", SIAM Journal on Applied Dynamical Systems (2015)



Causation entropy correctly identifies the causal network structure.

Optimal Causation Entropy (oCSE) Principle

SIAM J. APPLIED DYNAMICAL SYSTEMS
Vol. 14, No. 1, pp. 73–106

Causal Network Inference by Optimal Causation Entropy*

Jie Sun[†], Dane Taylor[‡], and Erik M. Bollt[†]

⋮

Theorem 2.3 (optimal causation entropy principle for causal network inference). *Suppose that the network stochastic process given by* (2.4) *satisfies the Markov properties in* (2.8). *Let* $I \subset \mathcal{V}$ *be a given set of nodes and* N_I *be the set of* I*'s causal parents, as defined in* (2.3). *It follows that*

(a) *(Direct inference) Node* $j \in N_I$ *iff* $\Leftrightarrow \exists K \supset N_I$ *such that* $C_{j\rightarrow I|(K-\{j\})} > 0 \Leftrightarrow \forall K \subset \mathcal{V}, C_{j\rightarrow I|(K-\{j\})} > 0$.

(b) *(Partial conditioning removal) If there exists* $K \subset \mathcal{V}$ *such that* $C_{j\rightarrow I|(K-\{j\})} = 0$*, then* $j \notin N_I$.

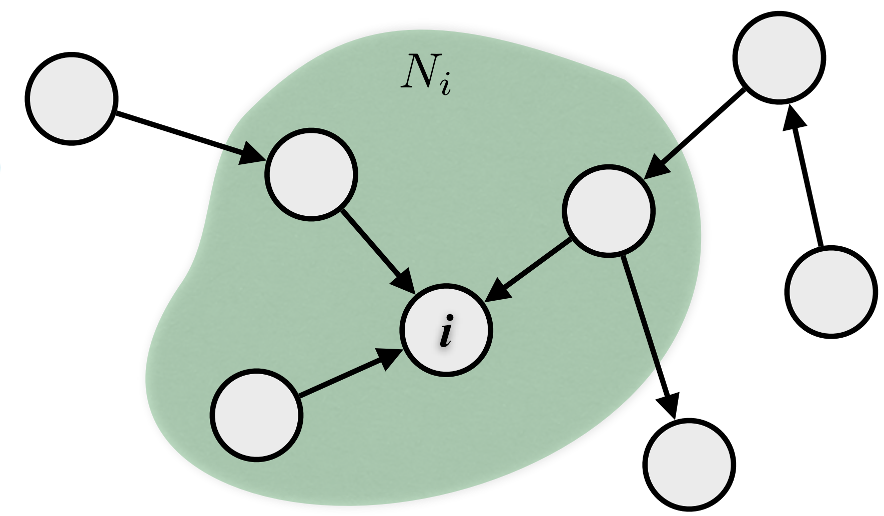
(c) *(Optimal causation entropy principle) The set of causal parents is the minimal set of nodes with maximal causation entropy.*

Define the family of sets with maximal causation entropy as

$$\mathcal{K} = \{K | \forall K' \subset \mathcal{V}, C_{K'\rightarrow I} \leq C_{K\rightarrow I}\}. \quad (2.26)$$

Then the set of causal parents is given by

$$N_I = \cap_{K\in\mathcal{K}} K = \text{argmin}_{K\in\mathcal{K}} K. \quad (2.27)$$



The set of causal "parents" is the *minimal* set of nodes which *maximizes* causation entropy.

The problem of causal network inference is converted into an optimization and estimation problem from given data.

How Entropic Regression Beats the Outliers Problem in Nonlinear System Identification

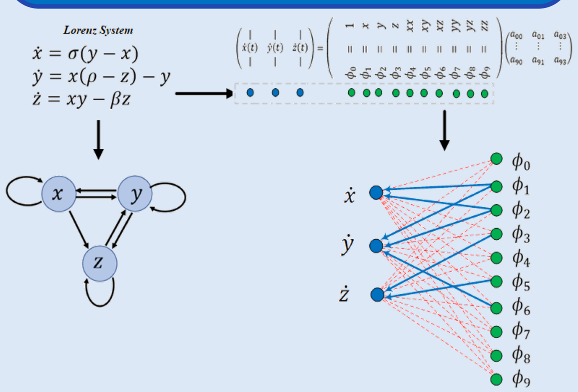
Abd AlRahman AlMomani[1,2], Jie Sun[1,3], Erik Bollt[1,2,3]

[1]*Clarkson Center for Complex Systems Science (C^3S^2),* [2]*Department of ECE,*
[3]*Department of Mathematics, Clarkson University, Potsdam, NY 13699*

Abstract

System identification (SID) is central in science and engineering applications whereby a general model form is assumed, but active terms and parameters must be inferred from observations. Sparse SID has recently become an important approach for SID such as in compressed sensing and Lasso methods. For the current state-of-art methods, it is still challenging to maintain the effectiveness of the methods under realistic scenarios where each observation is subject to non-trivial noise amplitude and sporadically further contaminated by even large noise and outliers. To mitigate such issues of large noise and outliers, we develop an entropic regression approach for nonlinear SID, whereby true model structures are identified based on relevance in reducing information flow uncertainty, not necessarily (just) sparsity. The use of information-theoretic measures as opposed to a metric-based cost function has a unique advantage, thanks to the asymptotic equipartition property of probability distributions, that outliers and other low-occurrence events are conveniently and intrinsically de-emphasized.

System Identification



Suppose that observations $\{\dot{z}(t)\}$ come from a general DE, represented by $\dot{z} = F(z)$, where $z \in \mathbb{R}^N$. Each component function $F_i(z)$ can be represented using a series expansion (for example a power series or a Fourier series), writing generally,

$$\dot{z}_i = F_i(\boldsymbol{z}) = \sum_{k=0}^{\infty} a_{ik}\phi_k(\boldsymbol{z}),$$

In vector form, under a choice of basis and truncation, the nonlinear system identification problem can be recast into the form of a linear inverse problem

$$\boldsymbol{f} = \Phi \boldsymbol{a} + \boldsymbol{\xi},$$

Where $f \in \mathbb{R}^{N\times 1}$ and $\Phi \in \mathbb{R}^{N\times K}$ are given with the goal to estimate $a \in \mathbb{R}^{K\times 1}$.

Entropic Regression

Entropic Regression:

The ER method contains two stages (also see Algorithm 1 for the pseudocode): forward ER and backward ER. In both stages, selection and elimination are based on an entropy criterion and parameters are updated in each iteration

Forward: $\begin{cases} k_i = \arg\max_{k \notin \{k_1,\dots,k_{i-1}\}} I(\Phi R(\Phi, \boldsymbol{f}, \{k\}); \boldsymbol{f} | \boldsymbol{z}_{i-1}), \\ \hat{\boldsymbol{a}} = R(\Phi, \boldsymbol{f}, \{k_1,\dots,k_i\}), \\ \boldsymbol{z}_i = \Phi R(\Phi, \boldsymbol{f}, \{k_1,\dots,k_i\}) \end{cases}$

Termination: $\max_k I(\Phi R(\Phi, \boldsymbol{f}, k); \boldsymbol{f} | \boldsymbol{z}_{i-1}) \approx 0$

Backward: $\begin{cases} \hat{\boldsymbol{a}} = R(\Phi, \boldsymbol{f}, \{k_1,\dots,k_i\}/\{k_i\}), \\ \bar{z}_j = \Phi R(\Phi, \boldsymbol{f}, \{k_1,\dots,k_i\}/\{k_i\}), \end{cases}$

while: $I(\Phi R(\Phi, \boldsymbol{f}, S); \boldsymbol{f} | \bar{z}_j) \approx 0.$

Results

Results are shown for:

(A) Double Well Potential.

$$f(x) = x^4 - x^2.$$

(B) Lorenz System.

$$\begin{cases} \dot{z}_1 = F_1(\boldsymbol{z}) = \sigma(z_2 - z_1), \\ \dot{z}_2 = F_2(\boldsymbol{z}) = z_1(\rho - z_3) - z_2, \\ \dot{z}_3 = F_3(\boldsymbol{z}) = z_1 z_2 - \beta z_3, \end{cases}$$

(C) Kuramoto-Sivashinsky Equations.

$$u_t = -\nu u_{xxxx} - u_{xx} + 2uu_x,$$
$$(t,x) \in [0,\infty) \times (0,L)$$

$$\dot{a}_k = (k^2 - \nu k^4)a_k - k\sum_{m=-\infty}^{\infty} a_m a_{k-m}.$$

Methods

Least Squares (LS): $\min_{a\in\mathbb{R}^K} \|\Phi a - f\|_2$, $\boldsymbol{a} = \Phi^{\dagger}\boldsymbol{f}$

Orthogonal Least Squares (OLS):

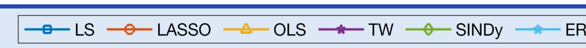
$$\begin{cases} (k_{\ell+1}, a_{k_{\ell+1}}) = \arg\min_{k,c} \|\boldsymbol{r}_\ell - c\phi_k\|_2, \\ \boldsymbol{r}_{\ell+1} = \boldsymbol{r}_\ell - \phi_{k_{\ell+1}} a_{k_{\ell+1}}. \end{cases}$$

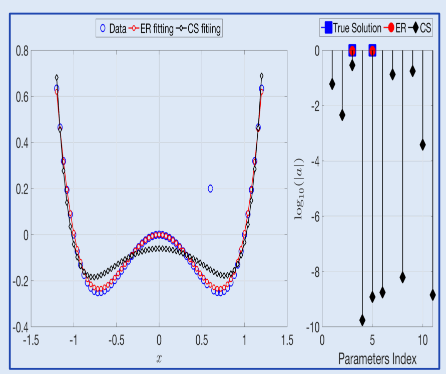
LASSO: $\min_{\boldsymbol{a}\in\mathbb{R}^K} \left(\|\Phi\boldsymbol{a} - \boldsymbol{f}\|_2^2 + \lambda\|\boldsymbol{a}\|_1\right),$

Compressed Sensing (CS): $\begin{cases} \arg\min_{\boldsymbol{a}} \|\boldsymbol{a}\|_1, \\ \text{subject to } \|\Phi\boldsymbol{a} - \boldsymbol{f}\| \le \epsilon, \end{cases}$

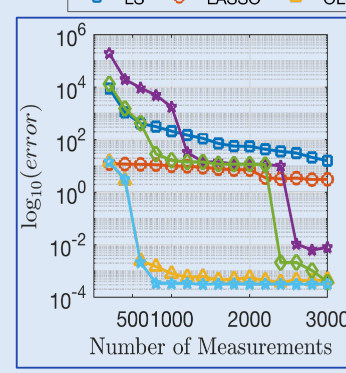
SINDy : Sequential least squares with hard-thresholding.

Tran-Ward (TW): Extend SINDy to the data corruption case, and reconstruct the system assuming that the corrupted data occurs in sparse and isolated time intervals.

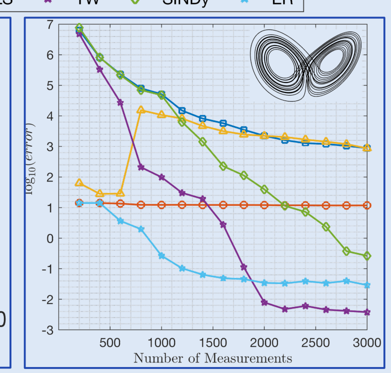




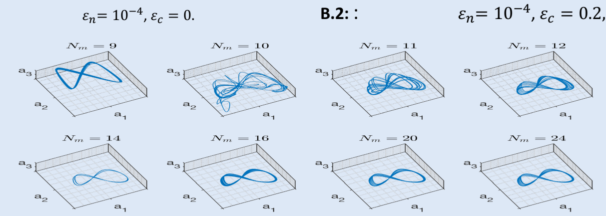
A. *Single Point Outlier.*



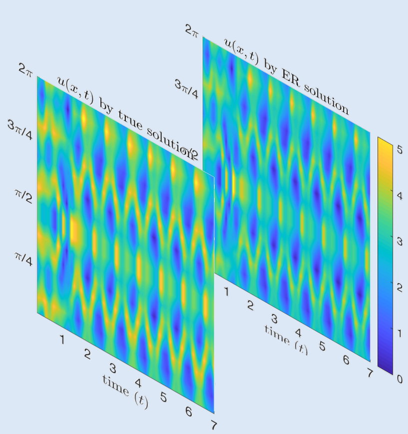
B.1: $\varepsilon_n = 10^{-4}, \varepsilon_c = 0.$



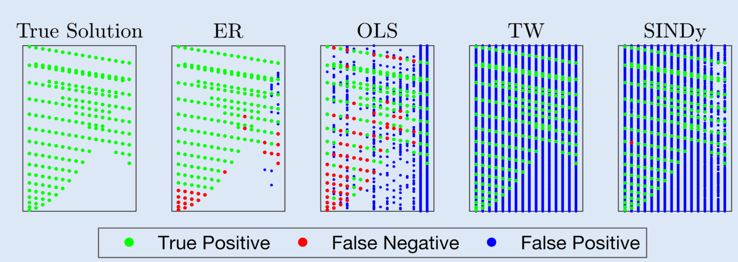
B.2: : $\varepsilon_n = 10^{-4}, \varepsilon_c = 0.2, p = 0.2$



C.2. Effect of number of modes.



C.1. KSE $\boldsymbol{u(x,t)}$ solution, recovered by ER.



C.3. Solution Sparse Structure for different methods.

References

[1] Abd AlRahman R. AlMomani, Jie Sun, and Erik Bollt. How Entropic Regression Beats the Outliers Problem in Nonlinear System Identification. arXiv. 2019.

[2] Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. Proceedings of the National Academy of Sciences, 113(15):3932–3937, 2016.

[3] Emmanuel J. Cand`es, Justin K. Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. Communications on Pure and Applied Mathematics, 59(8):1207–1223, 2006.

[4] Sheng Chen, Stephen A. Billings, and Wan Luo. Orthogonal least squares methods and their application to non-linear system identification. International Journal of Control, 50(5):1873–1896, 1989.

[5] Trevor Hastie, Robert Tibshirani, and Martin Wainwright. Statistical learning with sparsity: The lasso and generalizations. CRC Press, 2015.

[6] Giang Tran and Rachel Ward. Exact recovery of chaotic systems from highly corrupted data. Multiscale Modeling & Simulation, 15:1108–1129, 2017.

Acknowledgment

This work was funded in part by the Simons Foundation Grant No. 318812, the Army Research Office Grant No. W911NF-16-1-0081, the Office of Naval Research Grant No. N00014-15-1-2093, and also DARPA.

Chaos Focus Issue: Causation Inference and Information Flow in Dynamical Systems: Theory and Applications

Chaos **28**, 2018 July Issue
Article no: 075301-075311

Guest editors: Erik Bollt, Jakob Runge, Jie Sun

Complex networks for tracking extreme rainfall during typhoons

U. Ozturk, N. Marwan, O. Korup, H. Saito more...

Detecting directional couplings from multivariate flows by the joint distance distribution

José M. Amigó, and Yoshito Hirata

Transient and equilibrium causal effects in coupled oscillators

Dmitry A. Smirnov

The quoter model: A paradigmatic model of the social flow of written information

James P. Bagrow, and Lewis Mitchell

Detecting causality using symmetry transformations

Subhradeep Roy, and Benjamin Jantzen

Inter-scale information flow as a surrogate for downward causation that maintains spiral waves

Hiroshi Ashikaga, and Ryan G. James

Causality, dynamical systems and the arrow of time

Milan Paluš, Anna Krakovská, Jozef Jakubík, and Martina Chvosteková

Anatomy of leadership in collective behaviour

Joshua Garland, Andrew M. Berdahl, Jie Sun, and Erik M. Bollt

Open or closed? Information flow decided by transfer operators and forecastability quality metric

Erik M. Bollt

Causal network reconstruction from time series: From theoretical assumptions to practical estimation

J. Runge

Causation and information flow with respect to relative entropy

X. San Liang

Analytical Properties of CSE

Theorem 2.2 (basic analytical properties of causation entropy). *Suppose that the network stochastic process given by (2.4) satisfies the Markov assumptions in (2.8). Let $I \subset \mathcal{V}$ be a set of nodes and N_I be its causal parents. Consider two sets of nodes $J \subset \mathcal{V}$ and $K \subset \mathcal{V}$. The following results hold:*

(a) *(Redundancy) If $J \subset K$, then $C_{J \to I|K} = 0$.*
(b) *(No false positive) If $N_I \subset K$, then $C_{J \to I|K} = 0$ for any set of nodes J.*
(c) *(True positive) If $J \subset N_I$ and $J \not\subset K$, then $C_{J \to I|K} > 0$.*
(d) *(Decomposition) $C_{J \to I|K} = C_{(K \cup J) \to I} - C_{K \to I}$.*

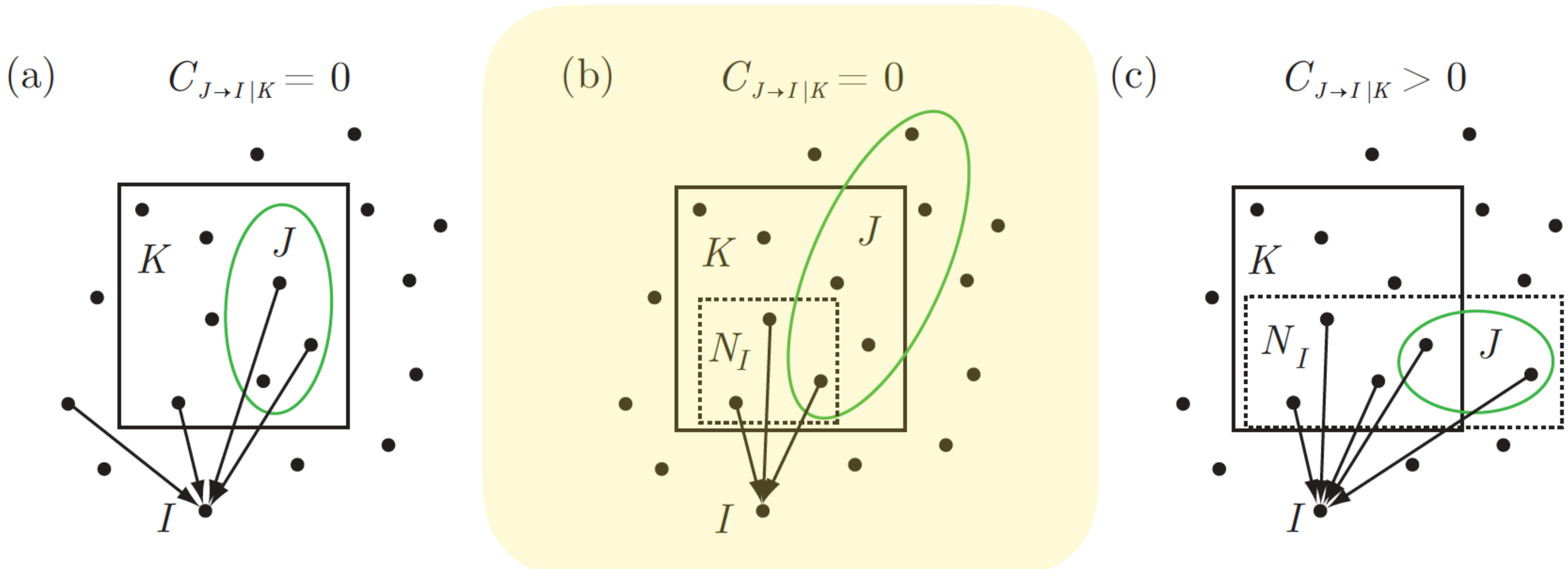
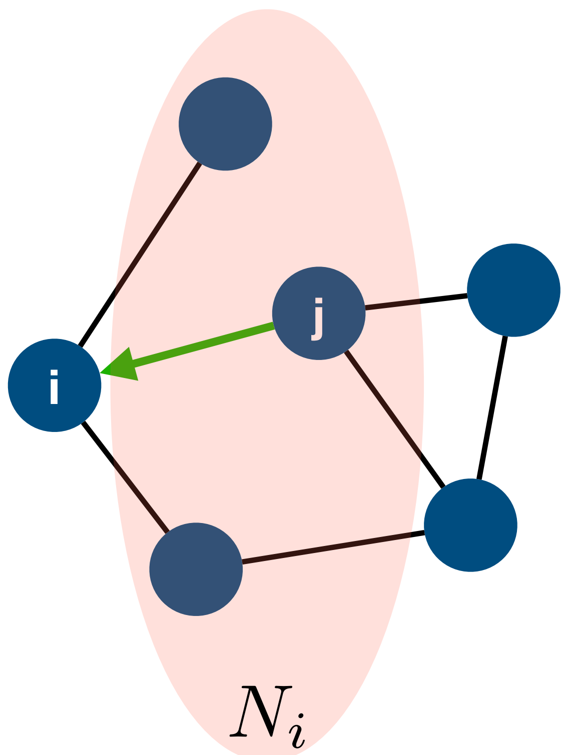


Figure 3. *Basic analytical properties of causation entropy (Theorem 2.2) allowing for the inference of the causal parents N_I of a set of nodes I.* (a) *Redundancy: If J is a subset of the conditioning set K ($J \subset K$), then the causation entropy $C_{J \to I|K} = 0$.* (b) *No false positive: If N_I is already included in the conditioning set K ($N_I \subset K$), then $C_{J \to I|K} = 0$.* (c) *True positive: If a set J contains at least one causal parent of I that does not belong to the conditioning set K, i.e., $(J \subset N_I) \wedge (J \not\subset K)$, then $C_{J \to I|K} > 0$.*

JS, D. Taylor, E. Bollt, *Causal network inference by optimal causation entropy,* SIAM Journal on Applied Dynamical Systems (2015).

Model-free Measures of Pairwise Coupling



Causation Entropy (CSE)

$$C_{j\to i|K_i} = H(X_{t+1}^{(i)}|X_t^{(K_i)}) - H(X_{t+1}^{(i)}|X_t^{(K_i)}, X_t^{(j)})$$

where $K_i = N_i/\{j\}$

The observed "net" influence of j on i given the rest of the nbs of i

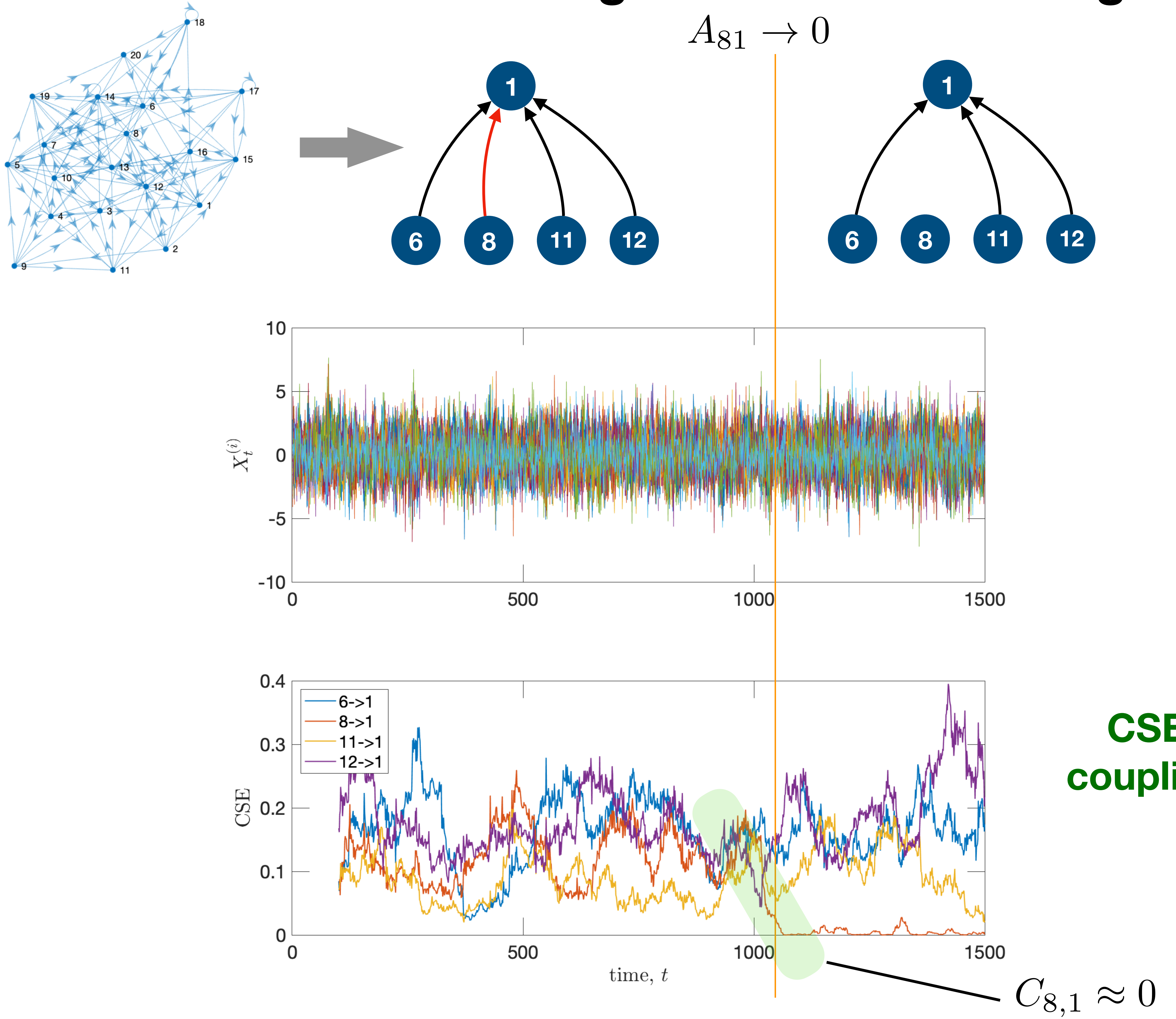
JS and Erik Bollt, "*Causation entropy identifies indirect influences, dominance of neighbors and anticipatory couplings*", Physica D (2014)

JS, Dane Taylor, and Erik Bollt, "*Causal network inference by optimal causation entropy*", SIAM Journal on Applied Dynamical Systems (2015)

Change of link strength $A_{ij} \to A_{ij} + \delta A_{ij} \;\Leftrightarrow\; C_{ij} \to C_{ij} + \delta C_{ij}$

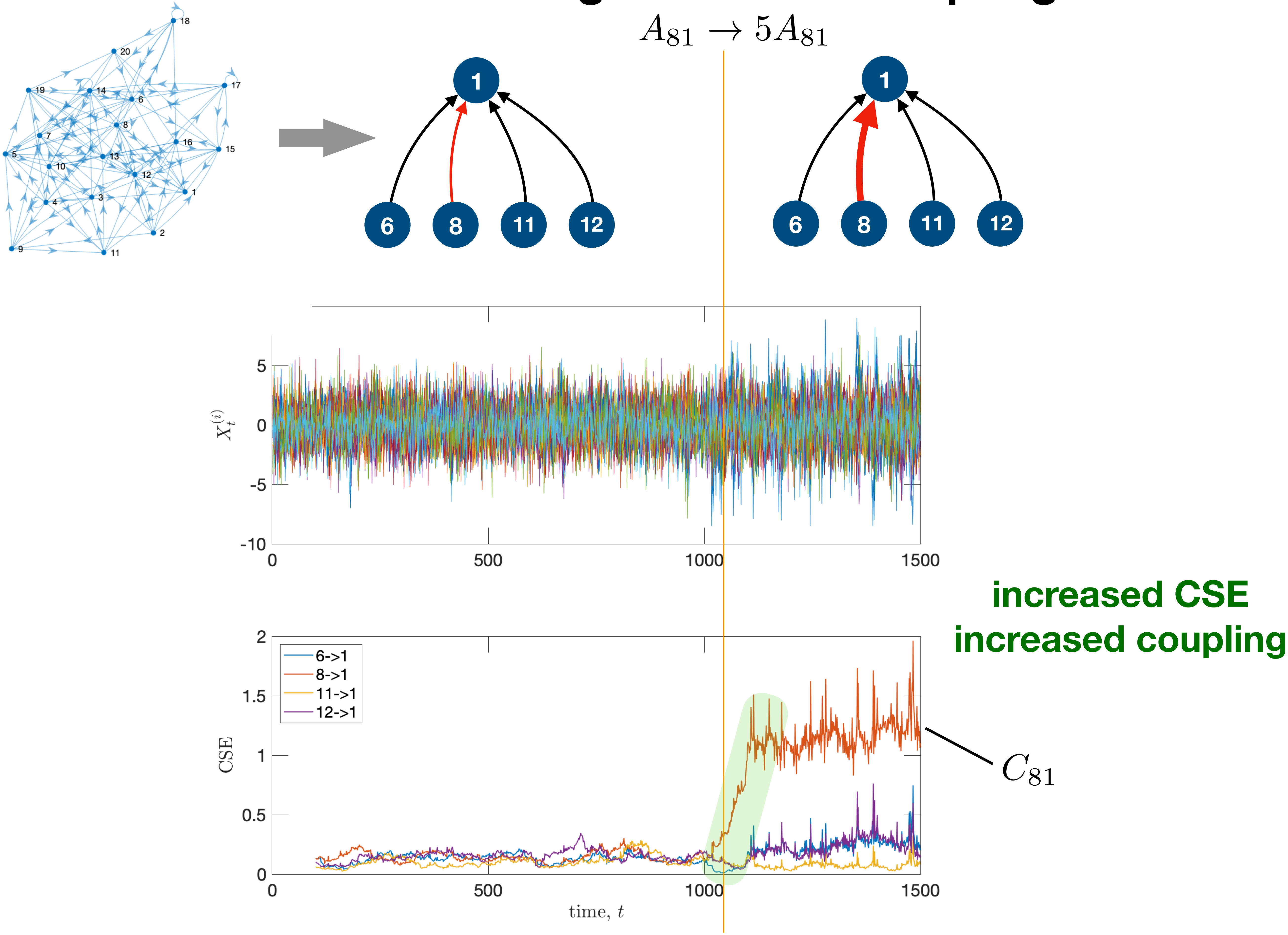
Link removal $A_{ij} \to 0 \Leftrightarrow C_{j\to i|K_i} \to 0$

Results: Detecting a Disconnected Edge

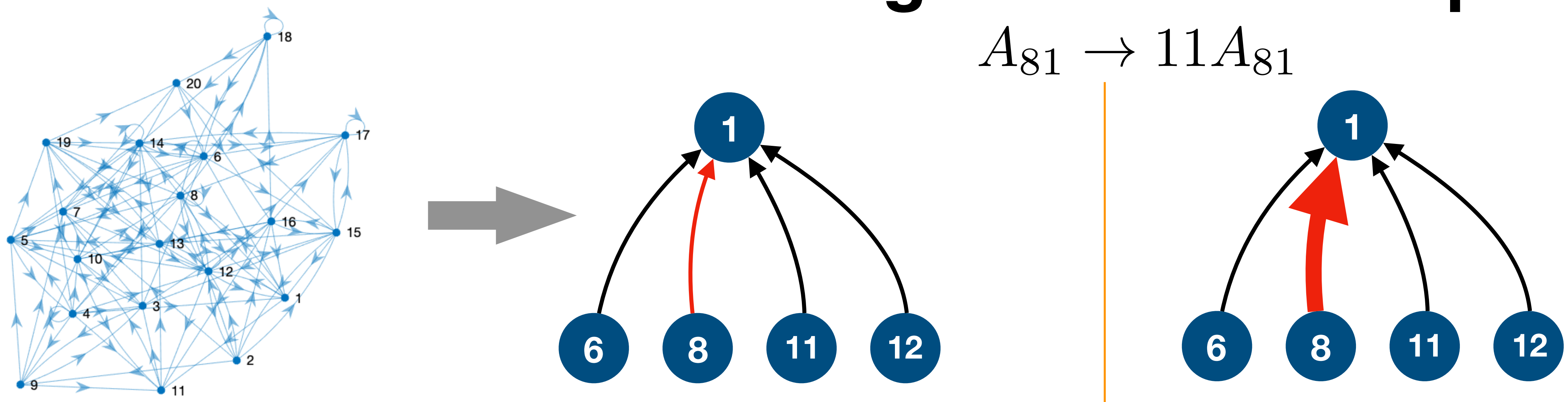


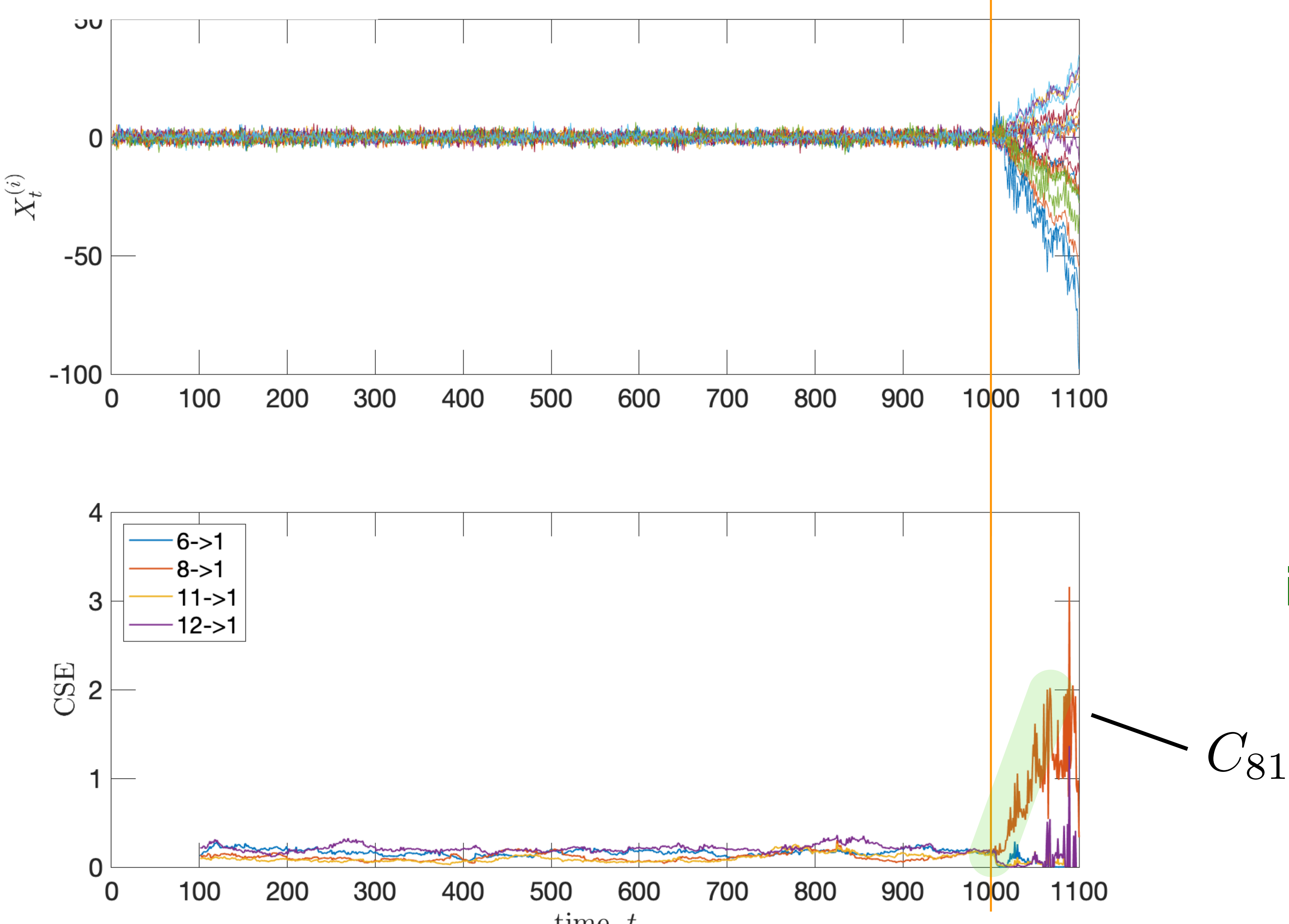
CSE —> 0
coupling —> 0

Results: Detecting Increased Coupling



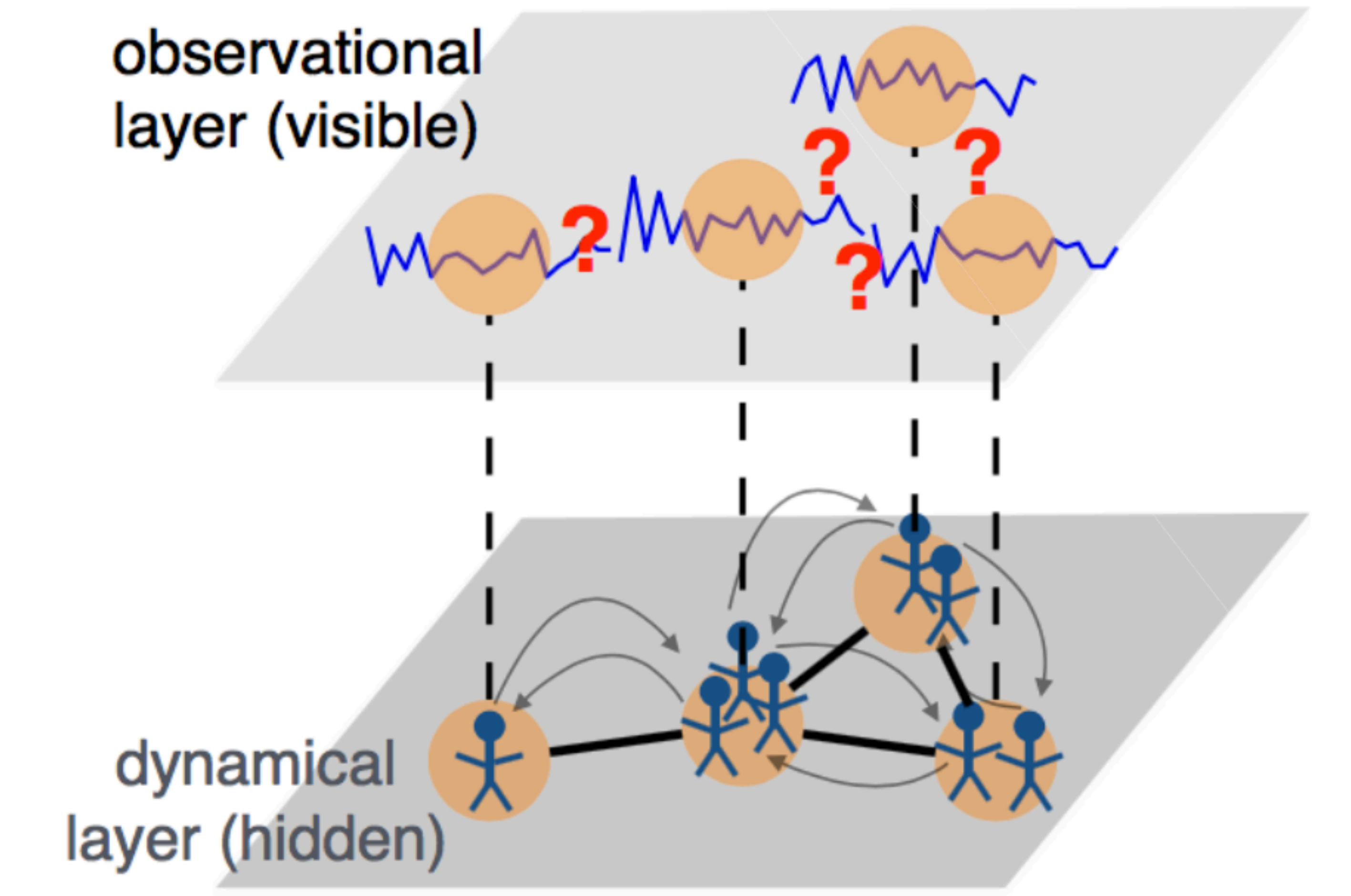
Results: Detecting Increased Coupling



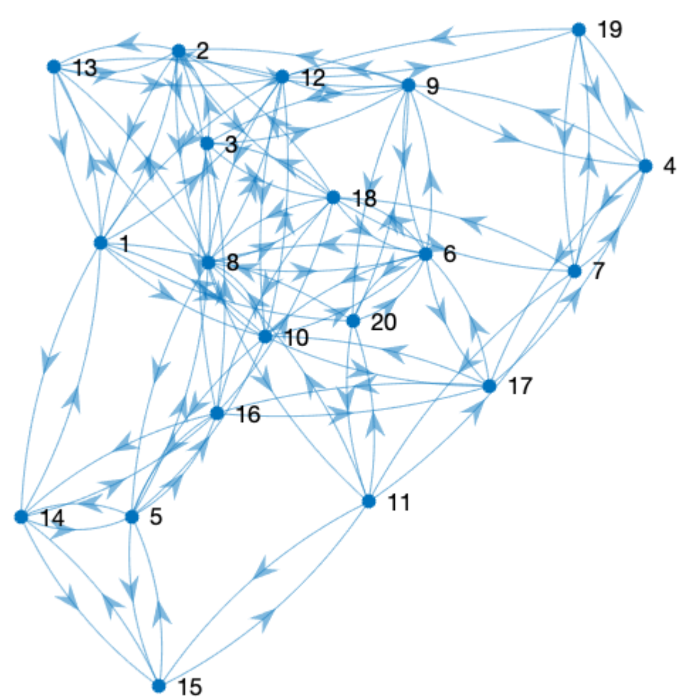


increased CSE
increased coupling

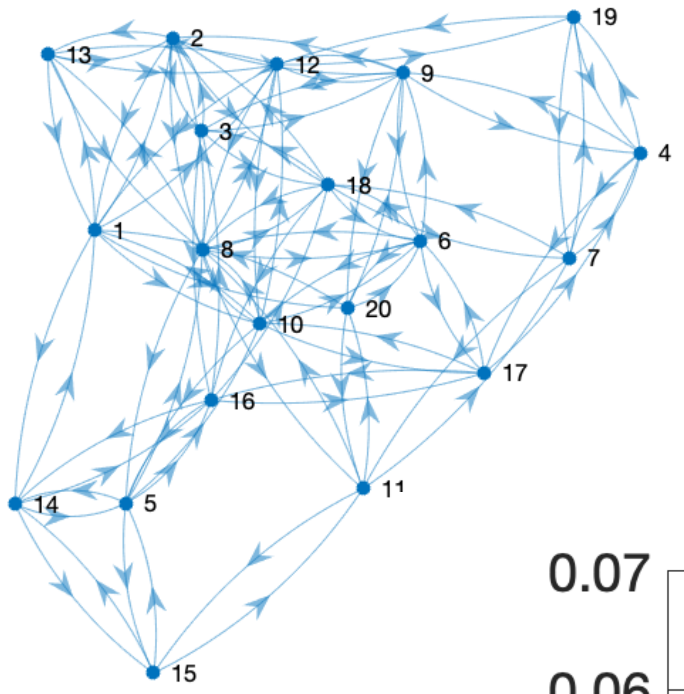
Detection from Random Walk Dynamics on Networks

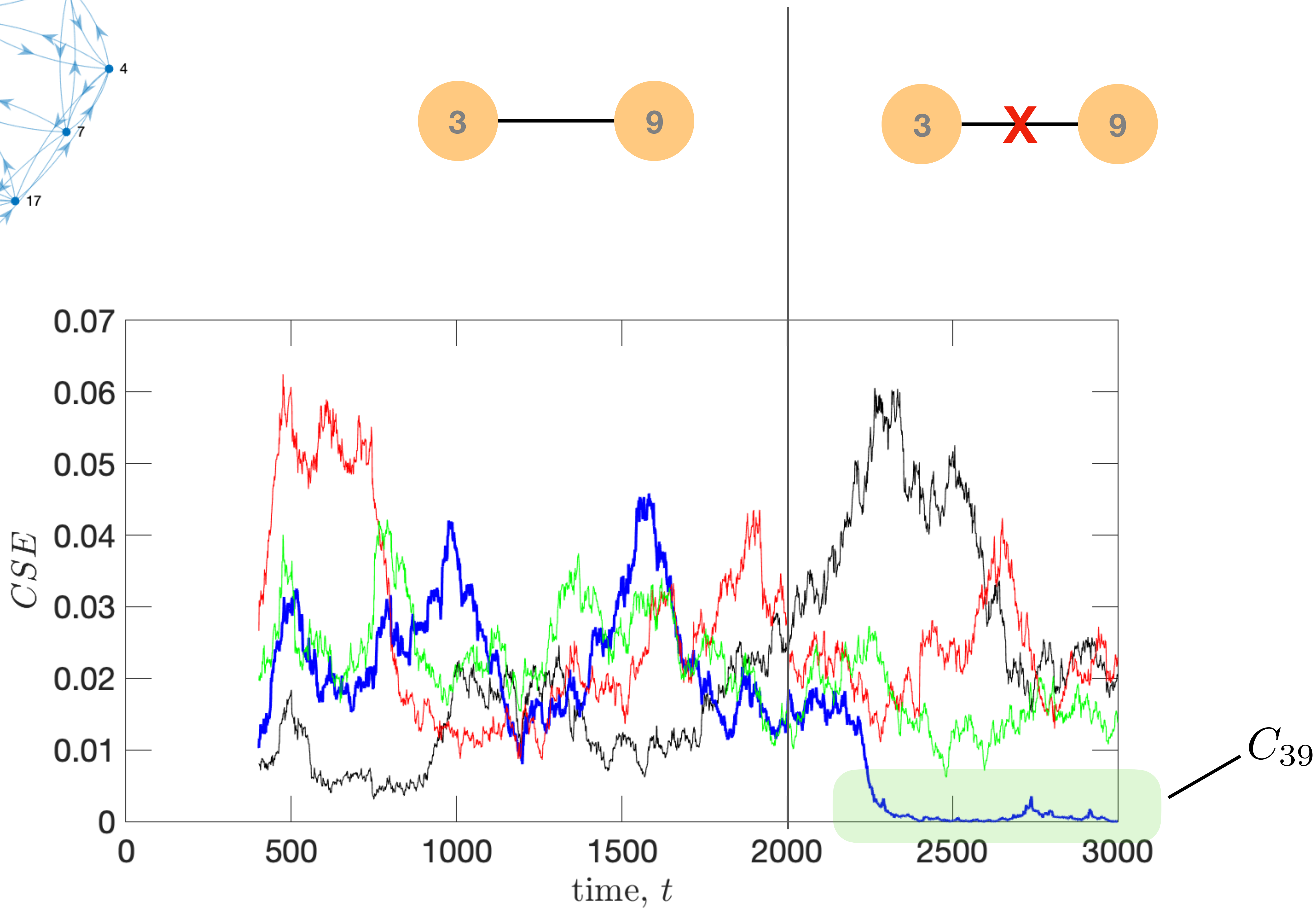


JS, F. Quevedo, E. Bollt, *Data Fusion Reconstruction of Spatially Embedded Complex Networks,* arXiv: 1707.00731.



Q: detecting this edge disconnection?





The use of CSE to track coupling enables detection of:

— edge disconnection (coupling —> 0)

— increase/decrease of information flow