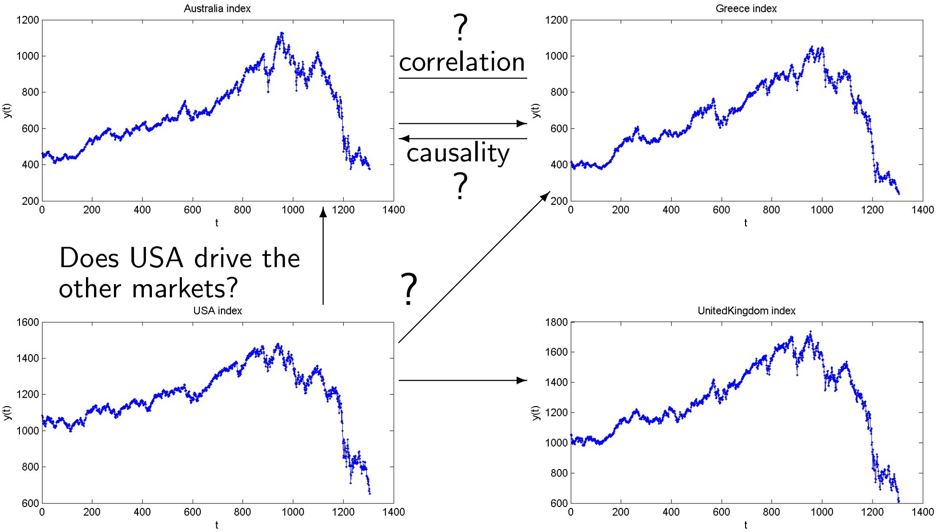
Connectivity networks and applications

Dimitris Kugiumtzis

e-mail: dkugiu@gen.auth.gr http:\\users.auth.gr\dkugiu

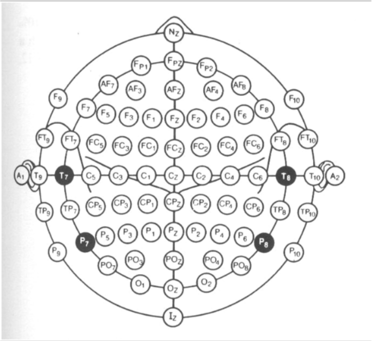
November 13, 2019

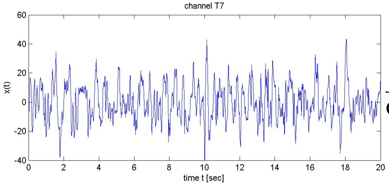
Financial World Markets



Electroencephalogram (EEG)

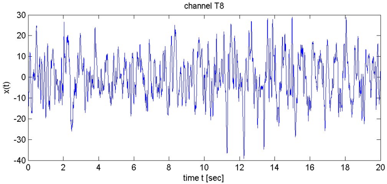
http://en.wikipedia.org/wiki/File:EEG_cap.jpg

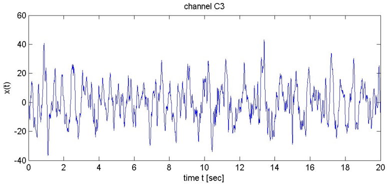


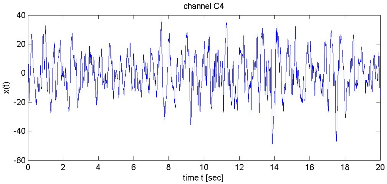


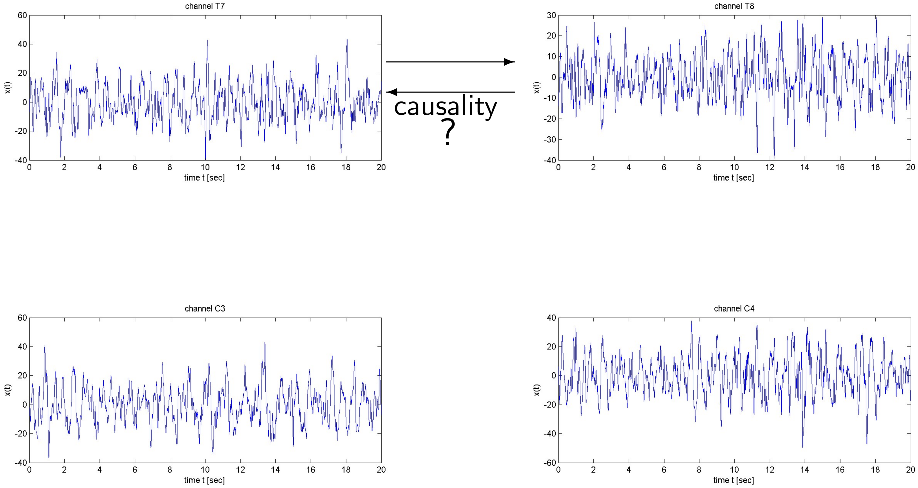
correlation

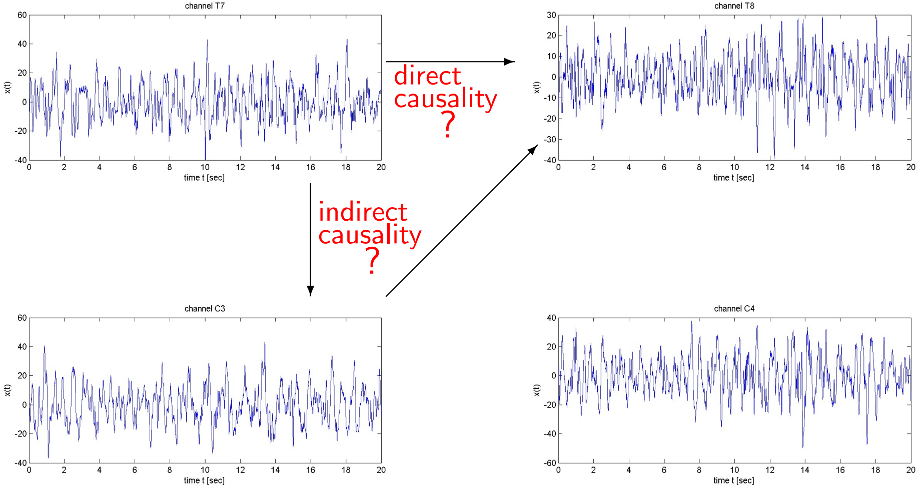
?

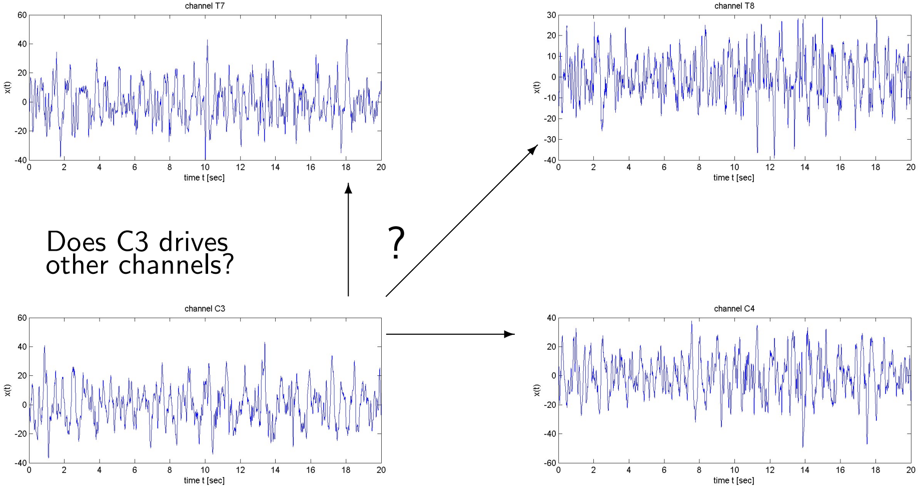






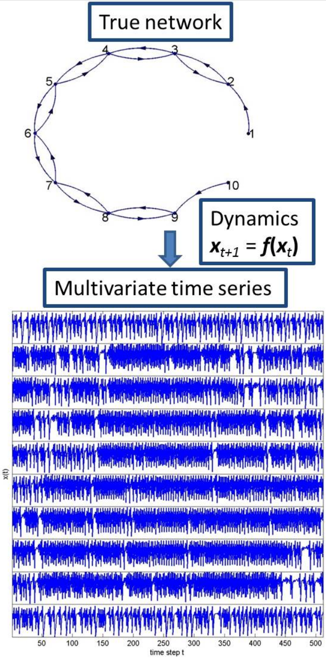

channel T7
x(t)
time t [sec]
causality
?
channel T8
x(t)
time t [sec]
channel C3
x(t)
time t [sec]
channel C4
x(t)
time t [sec]


channel T7
channel T8
channel C3
channel C4
direct causality ?
indirect causality ?
x(t)
time t [sec]

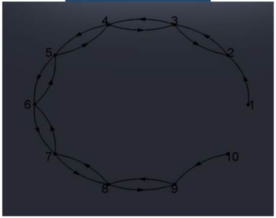


channel T7
channel T8
channel C3
channel C4
x(t)
time t [sec]
Does C3 drives
other channels?
?

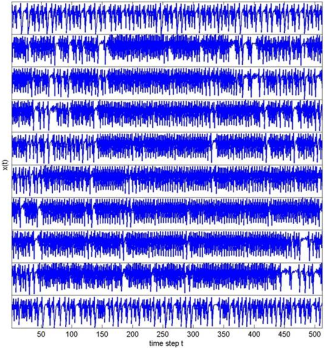
Complex networks from multivariate time series



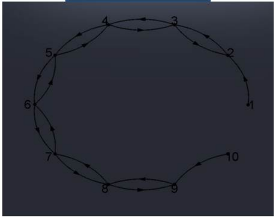
Complex networks from multivariate time series



Multivariate time series

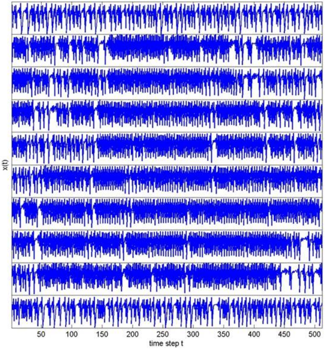


[redacted]



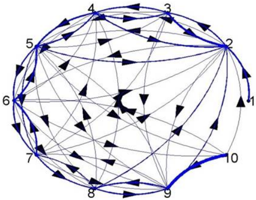
Complex networks from multivariate time series

Multivariate time series

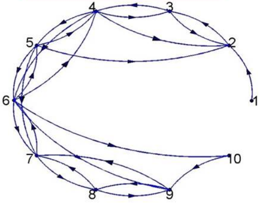


Estimated network

weighted connections



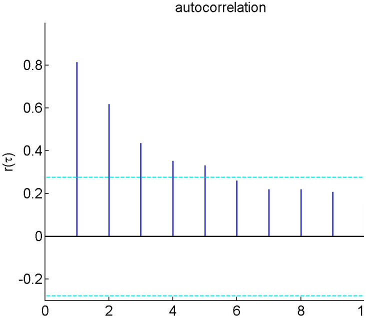
binary connections

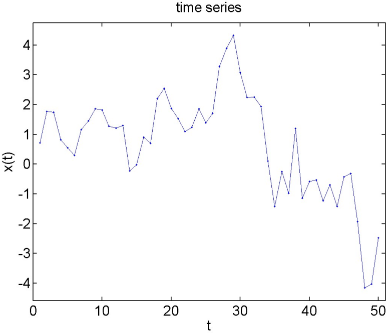


(auto)correlation $r(X_t; X_{t-\tau})$

Are X_t and X_{t-1} linearly correlated? $r(X_t; X_{t-1}) \neq 0$?

Are X_t and X_{t-2} linearly correlated? $r(X_t; X_{t-2}) \neq 0$?

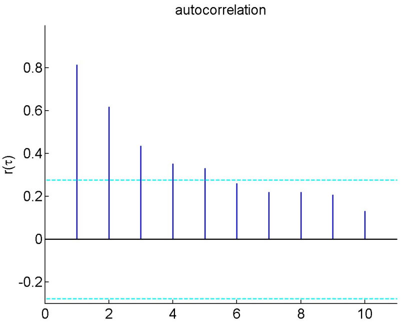




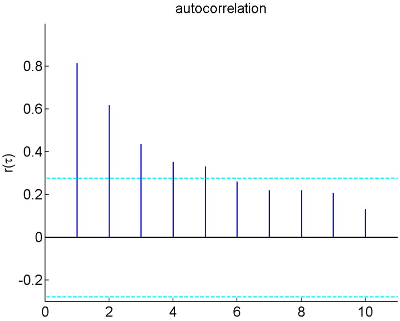
(auto)correlation $r(X_t; X_{t-\tau})$

Are X_t and X_{t-1} linearly correlated? $r(X_t; X_{t-1}) \neq 0$? Yes

Are X_t and X_{t-2} linearly correlated? $r(X_t; X_{t-2}) \neq 0$? Yes



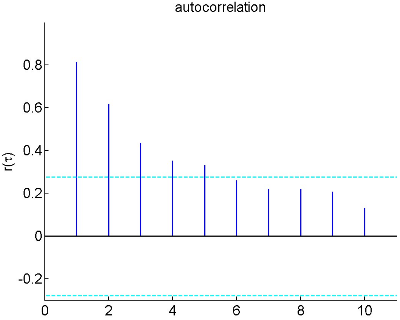
Are X_t and X_{t-2} directly linearly correlated?



Are X_t and X_{t-2} directly linearly correlated?

Are X_t and X_{t-2} linearly correlated given X_{t-1}?

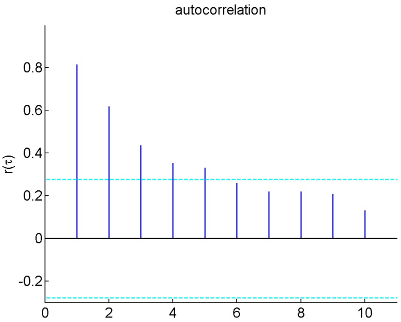
$$r(X_t; X_{t-2}|X_{t-1}) \neq 0?$$

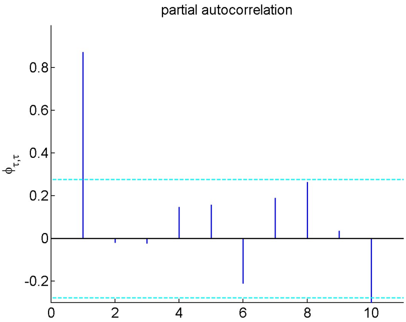


Are X_t and X_{t-2} directly linearly correlated?

Are X_t and X_{t-2} linearly correlated given X_{t-1}?

$$r(X_t; X_{t-2}|X_{t-1}) \neq 0? \qquad \text{No}$$





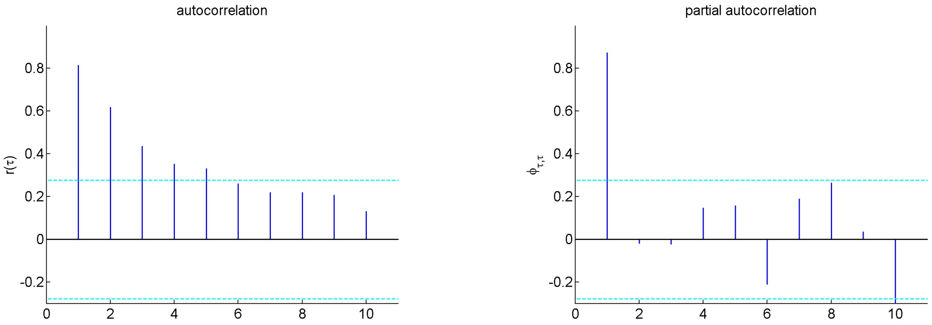
Are X_t and X_{t-2} directly linearly correlated?

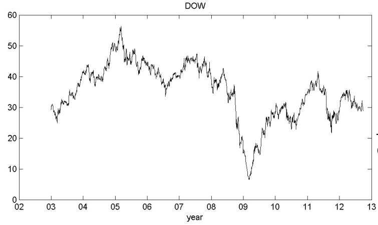
Are X_t and X_{t-2} linearly correlated given X_{t-1}?

$$r(X_t; X_{t-2}|X_{t-1}) \neq 0? \qquad \text{No}$$

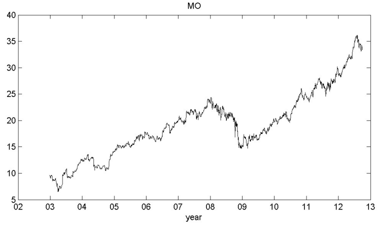
Are X_t and X_{t-2} linearly or/and nonlinearly correlated given X_{t-1}?

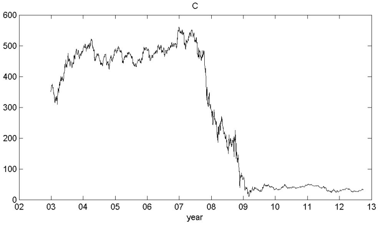
$$I(X_t; X_{t-2}|X_{t-1}) \neq 0?$$

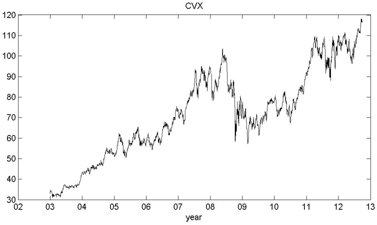




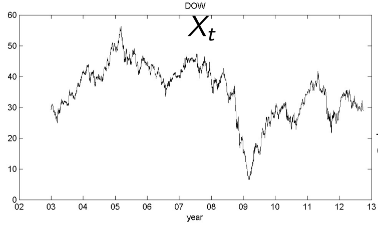
correlation
?

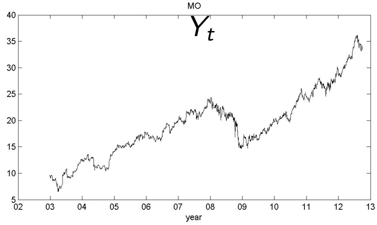






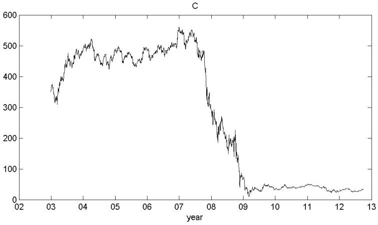
$r(X_t; Y_t)$

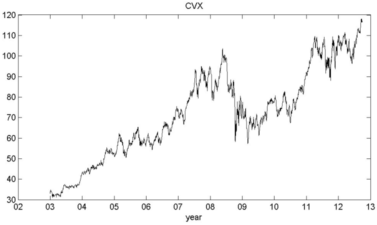




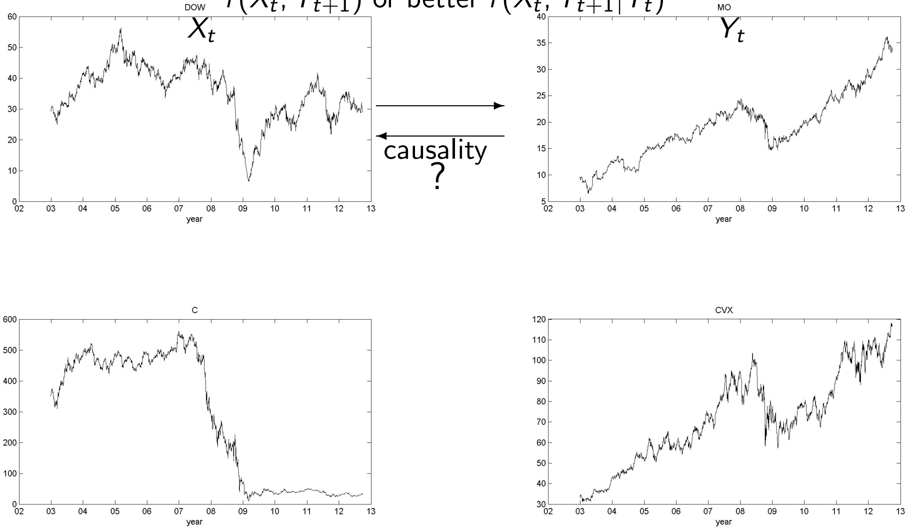
$\overline{\text{correlation}}$

?

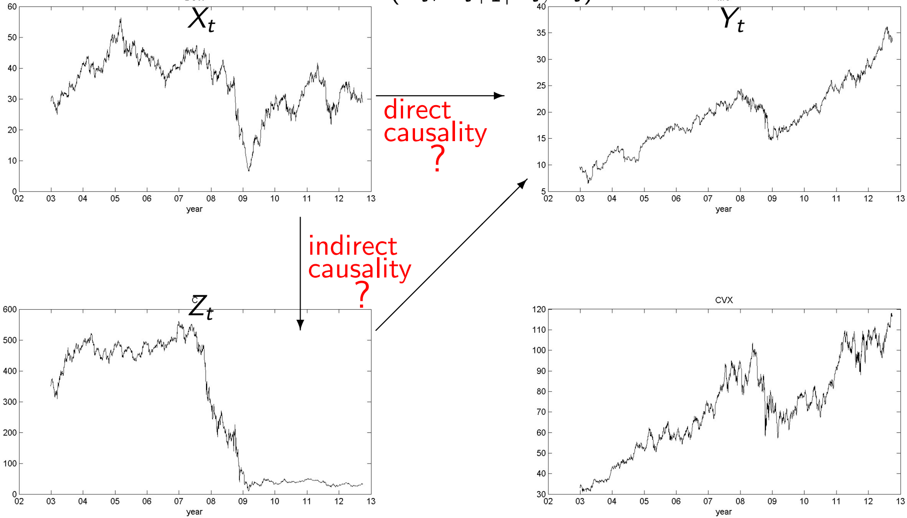


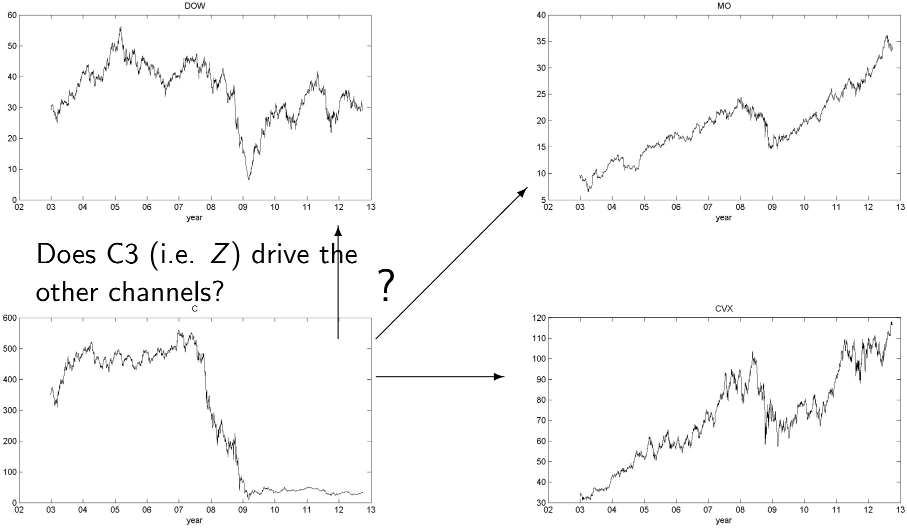


$r(X_t; Y_{t+1})$ or better $r(X_t; Y_{t+1} | Y_t)$



$$r(X_t; Y_{t+1} | Y_t, Z_t)$$


DOW
X_t
year
MO
Y_t
year
direct causality ?
indirect causality ?
C
Z_t
year
CVX
year



Does C3 (i.e. Z) drive the other channels?

Correlation measures

	$X \sim Y$	$X \sim Y \mid Z$
Linear	- Cross-Correlation	- Partial Correlation
	- Coherence	- Partial Coherence
Nonlinear	- Phase Synchronization	?
	- Cross Mutual Information	?

Granger Causality measures

	$X \rightarrow Y$	$X \rightarrow Y \mid Z$
Linear	- Granger Causality Index (GCI)	- Conditional (Partial) GCI (CGCI)
	- Directed Coherence (DC)	- Partial DC (PDC)
	- Directed Transfer Function (DTF)	- direct DTF (dDTF)
Nonlinear	- Directionality Index	?
	- Mean Conditional Recurrence	?
	- Transfer entropy (TE)	- Partial TE (PTE)
	- Mutual Information from Mixed Embedding (MIME)	- Partial MIME (PMIME)

Correlation measures

Bivariate time series $\{x_t, y_t\}_{t=1}^{n}$

Linear correlation measures:

Estimate of cross-covariance

$$c_{XY}(\tau) = \hat{\gamma}_{XY}(\tau) = \frac{1}{n-\tau}\sum_{t=1}^{n-\tau}(x_t - \bar{x})(y_{t+\tau} - \bar{y})$$

$\bar{x}$ and $\bar{y}$ are sample means.

Estimate of **cross-correlation**:

$$r_{XY}(\tau) = r(X_t, Y_{t+\tau}) = \hat{\rho}_{XY}(\tau) = \frac{c_{XY}(\tau)}{c_{XY}(0)} = \frac{c_{XY}(\tau)}{s_X s_Y}$$

s_X and s_Y are sample standard deviations.

- $|r_{XY}(\tau)| \leq 1$
- $r_{XY}(\tau) = r_{YX}(-\tau)$ but $r_{XY}(\tau) \neq r_{XY}(-\tau)$

Noninear correlation measures:

Entropy: information from each sample of X (assume proper discretization of X)

$$H(X) = \sum_{x} p_X(x) \log p_X(x)$$

Mutual information: information for Y knowing X and vice versa

$$I(X,Y) = H(X)+H(Y)-H(X,Y) = \sum_{x,y} p_{XY}(x,y) \log \frac{p_{XY}(x,y)}{p_X(x)p_Y(y)}$$

For $X \to X_t$ and $Y \to Y_{t+\tau}$,

cross-delayed mutual information:

$$I_{XY}(\tau) = I(X_t, Y_{t+\tau}) = \sum_{x_t, y_{t+\tau}} p_{X_t Y_{t+\tau}}(x_t, y_{t+\tau}) \log \frac{p_{X_t Y_{t+\tau}}(x_t, y_{t+\tau})}{p_{X_t}(x_t) p_{Y_{t+\tau}}(y_{t+\tau})}$$

To compute $I_{XY}(\tau)$ make a partition of $\{x_t\}_{t=1}^n$, a partition of $\{y_t\}_{t=1}^n$ and compute probabilities for each cell from the relative frequency.

$r_{XY}(0) \neq 0$:

$\Longrightarrow$ (linear) correlation of x_t and y_t

$\Longrightarrow$ systems X and Y are correlated, $X \sim Y$

$r_{XY}(0) \neq 0$:

$\Longrightarrow$ (linear) correlation of x_t and y_t

$\Longrightarrow$ systems X and Y are correlated, $X \sim Y$

$r_{XY}(\tau) \neq 0$:

$\Longrightarrow$ (linear) correlation of x_t and $y_{t+\tau}$

$\Longrightarrow$ X effects the future of Y

$\Longrightarrow$ $X \to Y$

$r_{XY}(-\tau) \neq 0 \quad \Longrightarrow \quad Y \to X$

$r_{XY}(0) \neq 0$:

$\Longrightarrow$ (linear) correlation of x_t and y_t

$\Longrightarrow$ systems X and Y are correlated, $X \sim Y$

$r_{XY}(\tau) \neq 0$:

$\Longrightarrow$ (linear) correlation of x_t and $y_{t+\tau}$

$\Longrightarrow$ X effects the future of Y

$\Longrightarrow$ $X \to Y$

$r_{XY}(-\tau) \neq 0 \quad \Longrightarrow \quad Y \to X$

Thus $r_{XY}(\tau) = r(X_t, Y_{t+\tau})$ and $I_{XY}(\tau) = I(X_t, Y_{t+\tau})$ indicate the direction of interaction.

$r_{XY}(0) \neq 0$:

$\Longrightarrow$ (linear) correlation of x_t and y_t

$\Longrightarrow$ systems X and Y are correlated, $X \sim Y$

$r_{XY}(\tau) \neq 0$:

$\Longrightarrow$ (linear) correlation of x_t and $y_{t+\tau}$

$\Longrightarrow$ X effects the future of Y

$\Longrightarrow$ $X \rightarrow Y$

$r_{XY}(-\tau) \neq 0 \quad \Longrightarrow \quad Y \rightarrow X$

Thus $r_{XY}(\tau) = r(X_t, Y_{t+\tau})$ and $I_{XY}(\tau) = I(X_t, Y_{t+\tau})$ indicate the direction of interaction.

Can they also be used as causality measures?

$r_{XY}(0) \neq 0$:

$\Longrightarrow$ (linear) correlation of x_t and y_t

$\Longrightarrow$ systems X and Y are correlated, $X \sim Y$

$r_{XY}(\tau) \neq 0$:

$\Longrightarrow$ (linear) correlation of x_t and $y_{t+\tau}$

$\Longrightarrow$ X effects the future of Y

$\Longrightarrow$ $X \rightarrow Y$

$r_{XY}(-\tau) \neq 0 \quad \Longrightarrow \quad Y \rightarrow X$

Thus $r_{XY}(\tau) = r(X_t, Y_{t+\tau})$ and $I_{XY}(\tau) = I(X_t, Y_{t+\tau})$ indicate the direction of interaction.

Can they also be used as causality measures?

How can we approximate: $r(X_t; Y_{t+1}|Y_t)$, $r(X_t; Y_{t+1}|Y_t)$?

$r_{XY}(0) \neq 0$:

$\Longrightarrow$ (linear) correlation of x_t and y_t

$\Longrightarrow$ systems X and Y are correlated, $X \sim Y$

$r_{XY}(\tau) \neq 0$:

$\Longrightarrow$ (linear) correlation of x_t and $y_{t+\tau}$

$\Longrightarrow$ X effects the future of Y

$\Longrightarrow$ $X \rightarrow Y$

$r_{XY}(-\tau) \neq 0 \quad \Longrightarrow \quad Y \rightarrow X$

Thus $r_{XY}(\tau) = r(X_t, Y_{t+\tau})$ and $I_{XY}(\tau) = I(X_t, Y_{t+\tau})$ indicate the direction of interaction.

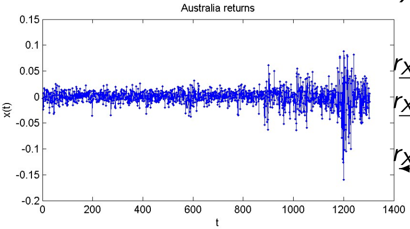
Can they also be used as causality measures?

How can we approximate: $r(X_t; Y_{t+1}|Y_t)$, $r(X_t; Y_{t+1}|Y_t)$?

...or even $r(X_t; Y_{t+1}|Y_t, Z_t)$, $r(X_t; Y_{t+1}|Y_t, Z_t)$?

Example: Returns for USA, UnitedKingdom, Greece and Australia.

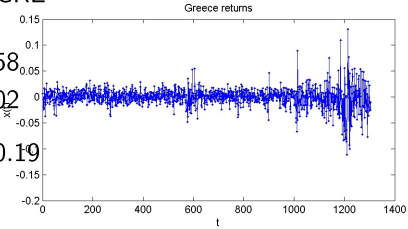
X:AUS, Y:GRE



$r_{XY}(0) = 0.58$

$r_{XY}(1) = 0.02$ →

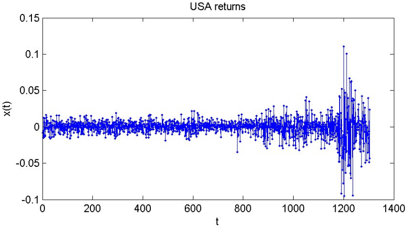
$r_{XY}(-1) = 0.19$ ←

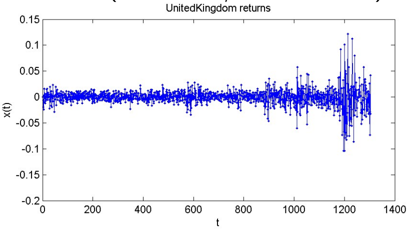


returns:

$x_t = \log(y_t) - \log(y_{t-1})$

Is the measure significant?
Can I draw a link? (directed / undirected)





Significance test for a correlation / causality measure q,
$H_0: q = 0 \qquad H_1: q \neq 0$

Randomization test

1. Generate M resampled (surrogate) time series, each by shifting the original observations with a random time step w:
original time series: $\{x_t\} = \{x_1, x_2, \ldots, x_n\}$
i-th surrogate time series:
$\{x_t^{*i}\} = \{x_{w+1}, x_{w+2}, \ldots, x_n, x_1, \ldots, x_{w-1}, x_w\}$
2. Compute the statistic q on the original pair, q_0, and on the M surrogate pairs, $q_1, \ldots, q_M$,
e.g. $q_0 \equiv r_{XY}(\tau) = \text{Corr}(x_t, y_{t+\tau})$ and $q_i \equiv \text{Corr}(x_t^{*i}, y_{t+\tau}^{*i})$
3. If q_0 is at the tails of the empirical null distribution formed by $q_1, \ldots, q_M$, reject H_0.
We use rank ordering: for a two-sided test, the p-value of the test is [Yu and Huang, 2001]

$$\begin{array}{lll} 2\frac{r_{q_0}-0.326}{M+1+0.348} & \text{if} & r_{q_0} < \frac{M+1}{2} \\ 2(1-\frac{r_{q_0}-0.326}{M+1+0.348}) & \text{if} & r_{q_0} \geq \frac{M+1}{2} \end{array}$$

Example: Returns for USA, UnitedKingdom, Greece and Australia.
Correlation matrix for delay 1, $r_{XY}(1)$

$$R(1) = \begin{bmatrix} & 0.382 & 0.333 & 0.596 \\ 0.049 & & 0.039 & 0.303 \\ 0.096 & 0.001 & & 0.190 \\ 0.031 & -0.001 & -0.021 & \end{bmatrix}$$

Randomization significance test for $r_{XY}(1)$ ($M = 1000$)

Matrix of p-values   Adjacency matrix

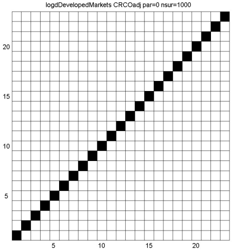
$$P(R(1)) = \begin{bmatrix} & 0.0013 & 0.0013 & 0.0033 \\ 0.0732 & & 0.1991 & 0.0013 \\ 0.0073 & 0.8901 & & 0.0033 \\ 0.2450 & 0.9760 & 0.4028 & \end{bmatrix} A = \begin{bmatrix} & 1 & 1 & 1 \\ 0 & & 0 & 1 \\ 1 & 0 & & 1 \\ 0 & 0 & 0 & \end{bmatrix}$$

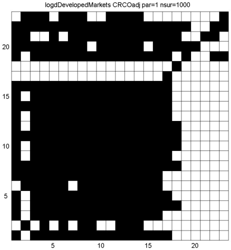
For significance level, say $\alpha = 0.05$, there may be $p < \alpha$ more often than it should be due to multiple testing.
Correction with e.g. False Discovery Rate

Network for World Financial Markets

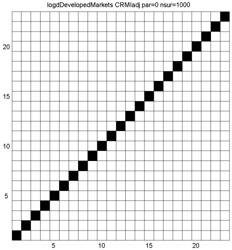
index	*market*
1	Austria
2	Belgium
3	Denmark
4	Finland
5	France
6	Germany
7	Greece
8	Ireland
9	Italy
10	Netherlands
11	Norway
12	Portugal
13	Spain
14	Sweden
15	Switzerland
16	UnitedKingdom
17	USA
18	Canada
19	Australia
20	HongKong
21	Japan
22	NewZealand
23	Singapore

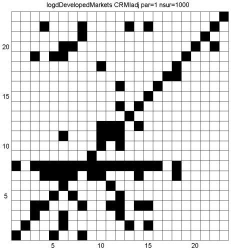
$r_{XY}(0)$ Adjacency matrix $r_{XY}(1)$



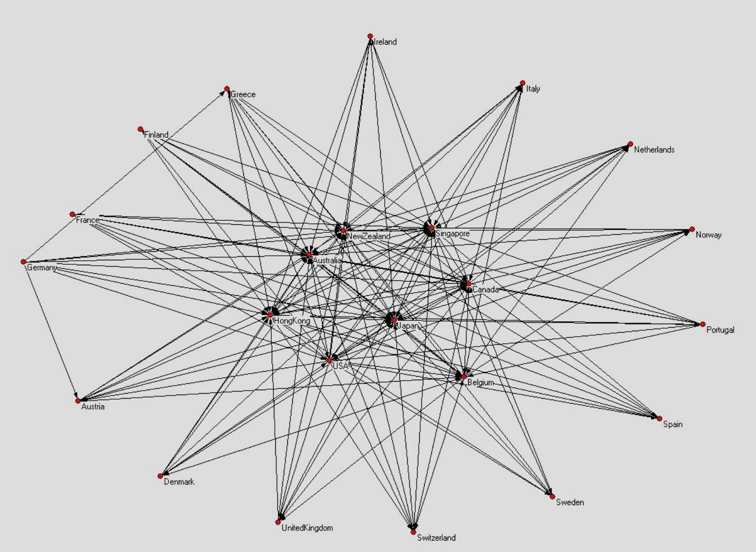


$I_{XY}(0)$ Adjacency matrix $I_{XY}(1)$





Correlation network, nodes: 23 financial markets, directed links: $r_{XY}(1)$



$r_{XY}(0) \neq 0$:

$\Longrightarrow$ (linear) correlation of x_t and y_t

$\Longrightarrow$ systems X and Y are correlated, $X \sim Y$

$r_{XY}(0) \neq 0$:

$\Longrightarrow$ (linear) correlation of x_t and y_t

$\Longrightarrow$ systems X and Y are correlated, $X \sim Y$

$r_{XY}(\tau) \neq 0$:

$\Longrightarrow$ (linear) correlation of x_t and $y_{t+\tau}$

$\Longrightarrow$ X effects the future of Y

$\Longrightarrow$ $X \rightarrow Y$

$r_{XY}(-\tau) \neq 0 \quad \Longrightarrow \quad Y \rightarrow X$

$r_{XY}(0) \neq 0$:

$\Longrightarrow$ (linear) correlation of x_t and y_t

$\Longrightarrow$ systems X and Y are correlated, $X \sim Y$

$r_{XY}(\tau) \neq 0$:

$\Longrightarrow$ (linear) correlation of x_t and $y_{t+\tau}$

$\Longrightarrow$ X effects the future of Y

$\Longrightarrow$ $X \rightarrow Y$

$r_{XY}(-\tau) \neq 0 \quad \Longrightarrow \quad Y \rightarrow X$

Thus $r_{XY}(\tau) = r(X_t, Y_{t+\tau})$ and $I_{XY}(\tau) = I(X_t, Y_{t+\tau})$ indicate the direction of interaction.

$r_{XY}(0) \neq 0$:

$\Longrightarrow$ (linear) correlation of x_t and y_t

$\Longrightarrow$ systems X and Y are correlated, $X \sim Y$

$r_{XY}(\tau) \neq 0$:

$\Longrightarrow$ (linear) correlation of x_t and $y_{t+\tau}$

$\Longrightarrow$ X effects the future of Y

$\Longrightarrow$ $X \rightarrow Y$

$r_{XY}(-\tau) \neq 0 \quad \Longrightarrow \quad Y \rightarrow X$

Thus $r_{XY}(\tau) = r(X_t, Y_{t+\tau})$ and $I_{XY}(\tau) = I(X_t, Y_{t+\tau})$ indicate the direction of interaction.

Can they also be used as causality measures?

$r_{XY}(0) \neq 0$:

$\Longrightarrow$ (linear) correlation of x_t and y_t

$\Longrightarrow$ systems X and Y are correlated, $X \sim Y$

$r_{XY}(\tau) \neq 0$:

$\Longrightarrow$ (linear) correlation of x_t and $y_{t+\tau}$

$\Longrightarrow$ X effects the future of Y

$\Longrightarrow$ $X \rightarrow Y$

$r_{XY}(-\tau) \neq 0 \quad \Longrightarrow \quad Y \rightarrow X$

Thus $r_{XY}(\tau) = r(X_t, Y_{t+\tau})$ and $I_{XY}(\tau) = I(X_t, Y_{t+\tau})$ indicate the direction of interaction.

Can they also be used as causality measures?

How can we approximate: $r(X_t; Y_{t+1}|Y_t)$, $r(X_t; Y_{t+1}|Y_t)$?

$r_{XY}(0) \neq 0$:

$\Longrightarrow$ (linear) correlation of x_t and y_t

$\Longrightarrow$ systems X and Y are correlated, $X \sim Y$

$r_{XY}(\tau) \neq 0$:

$\Longrightarrow$ (linear) correlation of x_t and $y_{t+\tau}$

$\Longrightarrow$ X effects the future of Y

$\Longrightarrow$ $X \to Y$

$r_{XY}(-\tau) \neq 0 \quad \Longrightarrow \quad Y \to X$

Thus $r_{XY}(\tau) = r(X_t, Y_{t+\tau})$ and $I_{XY}(\tau) = I(X_t, Y_{t+\tau})$ indicate the direction of interaction.

Can they also be used as causality measures?

How can we approximate: $r(X_t; Y_{t+1}|Y_t)$, $r(X_t; Y_{t+1}|Y_t)$?

...or even $r(X_t; Y_{t+1}|Y_t, Z_t)$, $r(X_t; Y_{t+1}|Y_t, Z_t)$?

Linear causality measures (direct and indirect)

Idea of Granger causality $X \to Y$: [Brandt & Williams, 2007, Chp 2]
predict Y better when including X in the regression model.

Measure 1a: Granger Causality Index (GCI)

Bivariate time series $\{x_t, y_t\}_{t=1}^{n}$
driving system: X, response system: Y

Model 1 (restricted, R, X absent in the model):

$$y_t = \sum_{i=1}^{p} a_i y_{t-i} + e_{R,t}$$

Model 2 (unrestricted, U, X present in the model):

$$y_t = \sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=1}^{p} b_i x_{t-i} + e_{U,t}$$

$$\text{GCI}_{X \to Y} = \ln \frac{\text{Var}(\hat{e}_{R,t})}{\text{Var}(\hat{e}_{U,t})} \qquad \text{GCI}_{X \to Y} > 0 \Rightarrow X \to Y \text{ holds}$$

Parametric significance test for GCI

$\text{GCI}_{X\to Y} > 0$? $\Rightarrow$ Significance test

If X does not Granger causes Y then the contribution of X-lags in the unrestricted model should be insignificant $\Rightarrow$
the terms of X should be insignificant

H_0: $b_i = 0$, for all $i = 1, \ldots, p$
H_1: $b_i \neq 0$, for any of $i = 1, \ldots, p$

Snedecor-Fisher test (F-test):

$$F = \frac{(\text{SSE}^R - \text{SSE}^U)/p}{\text{SSE}^U/\text{ndf}}$$

SSE: sum of squared errors
ndf: number of degrees of freedoms, $\text{ndf} = (n-p) - 2p$,
$n-p$: number of equations,
$2p$: number of coefficients in the U-model.

Linear causality measures (direct and indirect)

Measure 1b: Conditional Granger Causality Index (CGCI)

K time series $\{x_t, y_t\}_{t=1}^n$ and $\{\mathbf{z}_t\}_{t=1}^n = \{z_{1,t}, z_{2,t}, \ldots, z_{K-2,t}\}_{t=1}^n$
driving system: X, response system: Y,
conditioning on system Z, $Z = \{Z_1, Z_2, \ldots, Z_{K-2}\}$

Model 1 (restricted, R, X absent in the model):

$$y_t = \sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=1}^{p} A_i \mathbf{z}_{t-i} + e_{R,t}$$

Model 2 (unrestricted, U, X present in the model):

$$y_t = \sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=1}^{p} b_i x_{t-i} + \sum_{i=1}^{p} A_i \mathbf{z}_{t-i} + e_{U,t}$$

$$\text{CGCI}_{X\to Y|Z} = \ln \frac{\text{Var}(\hat{e}_{R,t})}{\text{Var}(\hat{e}_{U,t})}$$

Parametric significance test for CGCI

$\text{CGCI}_{X\to Y|Z} > 0$? $\Rightarrow$ Significance test as for GCI

H_0: $b_i = 0$, for all $i = 1, \dots, p$
H_1: $b_i \neq 0$, for any of $i = 1, \dots, p$

$$F = \frac{(\text{SSE}^R - \text{SSE}^U)/p}{\text{SSE}^U/\text{ndf}}$$

$\text{ndf} = (n-p) - Kp$,
$n-p$: number of equations,
Kp: number of coefficients in the U-model.

Model order and embedding parameters

VAR model for Y

$$y_t = \sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=1}^{p} b_i x_{t-i} + e_{U,t}$$

Model order and embedding parameters

VAR model for Y

$$y_t = \sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=1}^{p} b_i x_{t-i} + e_{U,t}$$

$$y_{t+1} = \sum_{i=1}^{p} a_i y_{t-i+1} + \sum_{i=1}^{p} b_i x_{t-i+1} + e_{U,t+1}$$

y_{t+1} is given in terms of $\mathbf{y}_t = [y_t, y_{t-1}, \ldots, y_{t-p+1}]$ and $\mathbf{x}_t = [x_t, x_{t-1}, \ldots, x_{t-p+1}]$, $y_{t+1} = \mathbf{F}(\mathbf{y}_t, \mathbf{x}_t) + e_{t+1}$

Model order and embedding parameters

VAR model for Y

$$y_t = \sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=1}^{p} b_i x_{t-i} + e_{U,t}$$

$$y_{t+1} = \sum_{i=1}^{p} a_i y_{t-i+1} + \sum_{i=1}^{p} b_i x_{t-i+1} + e_{U,t+1}$$

y_{t+1} is given in terms of $\mathbf{y}_t = [y_t, y_{t-1}, \ldots, y_{t-p+1}]$ and $\mathbf{x}_t = [x_t, x_{t-1}, \ldots, x_{t-p+1}]$, $\quad y_{t+1} = \mathbf{F}(\mathbf{y}_t, \mathbf{x}_t) + e_{t+1}$

Let the lag step be $\tau \geq 1 \quad \Rightarrow \quad \mathbf{y}_t = [y_t, y_{t-\tau}, \ldots, y_{t-(p-1)\tau}]$:

τ, p: embedding parameters (generally different for X and Y)

Model order and embedding parameters

VAR model for Y

$$y_t = \sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=1}^{p} b_i x_{t-i} + e_{U,t}$$

$$y_{t+1} = \sum_{i=1}^{p} a_i y_{t-i+1} + \sum_{i=1}^{p} b_i x_{t-i+1} + e_{U,t+1}$$

y_{t+1} is given in terms of $\mathbf{y}_t = [y_t, y_{t-1}, \ldots, y_{t-p+1}]$ and $\mathbf{x}_t = [x_t, x_{t-1}, \ldots, x_{t-p+1}]$, $\quad y_{t+1} = \mathbf{F}(\mathbf{y}_t, \mathbf{x}_t) + e_{t+1}$

Let the lag step be $\tau \geq 1 \quad \Rightarrow \quad \mathbf{y}_t = [y_t, y_{t-\tau}, \ldots, y_{t-(p-1)\tau}]$:

τ, p: embedding parameters (generally different for X and Y)

State space reconstruction:

$\mathbf{x}_t = [x_t, x_{t-\tau_x}, \ldots, x_{t-(m_x-1)\tau_x}]'$, embedding parameters: m_x,τ_x

$\mathbf{y}_t = [y_t, y_{t-\tau_y}, \ldots, y_{t-(m_y-1)\tau_y}]'$, embedding parameters: m_y,τ_y

y_{t+1}: future state of Y

Nonlinear causality measures (direct and indirect)

(Shannon) Entropy: $H(X) = -\sum_x p(x) \log p(x)$
Mutual Information of X and Y:
$I(X;Y) = H(X) + H(Y) - H(X,Y)$

Nonlinear causality measures (direct and indirect)

(Shannon) Entropy: $H(X) = -\sum_x p(x) \log p(x)$
Mutual Information of X and Y:
$I(X;Y) = H(X) + H(Y) - H(X,Y)$

Transfer Entropy (TE) [Schreiber, 2000]

Measure the effect of X on Y at one time step ahead, accounting (conditioning) for the effect from its own current state

$$\begin{aligned}\mathrm{TE}_{X\to Y} &= I(y_{t+1}; \mathbf{x}_t | \mathbf{y}_t) \\ &= H(\mathbf{x}_t, \mathbf{y}_t) - H(y_{t+1}, \mathbf{x}_t, \mathbf{y}_t) + H(y_{t+1}, \mathbf{y}_t) - H(\mathbf{y}_t) \\ &= \sum p(y_{t+1}, \mathbf{x}_t, \mathbf{y}_t) \log \frac{p(y_{t+1}|\mathbf{x}_t, \mathbf{y}_t)}{p(y_{t+1}|\mathbf{y}_t)}\end{aligned}$$

Nonlinear causality measures (direct and indirect)

(Shannon) Entropy: $H(X) = -\sum_x p(x) \log p(x)$
Mutual Information of X and Y:
$I(X;Y) = H(X) + H(Y) - H(X,Y)$

Transfer Entropy (TE) [Schreiber, 2000]

Measure the effect of X on Y at one time step ahead, accounting (conditioning) for the effect from its own current state

$$\begin{aligned}
\mathrm{TE}_{X \to Y} &= I(y_{t+1}; \mathbf{x}_t | \mathbf{y}_t) \\
&= H(\mathbf{x}_t, \mathbf{y}_t) - H(y_{t+1}, \mathbf{x}_t, \mathbf{y}_t) + H(y_{t+1}, \mathbf{y}_t) - H(\mathbf{y}_t) \\
&= \sum p(y_{t+1}, \mathbf{x}_t, \mathbf{y}_t) \log \frac{p(y_{t+1}|\mathbf{x}_t, \mathbf{y}_t)}{p(y_{t+1}|\mathbf{y}_t)}
\end{aligned}$$

Joint entropies (and distributions) can have high dimension!

Entropy estimates from nearest neighbors [Kraskov et al, 2004]

Nonlinear causality measures (direct and indirect)

(Shannon) Entropy: $H(X) = -\sum_x p(x) \log p(x)$
Mutual Information of X and Y:
$I(X;Y) = H(X) + H(Y) - H(X,Y)$

Transfer Entropy (TE) [Schreiber, 2000]

Measure the effect of X on Y at one time step ahead, accounting (conditioning) for the effect from its own current state

$$\begin{aligned}\mathrm{TE}_{X\rightarrow Y} &= I(y_{t+1};\mathbf{x}_t|\mathbf{y}_t) \\ &= H(\mathbf{x}_t,\mathbf{y}_t) - H(y_{t+1},\mathbf{x}_t,\mathbf{y}_t) + H(y_{t+1},\mathbf{y}_t) - H(\mathbf{y}_t) \\ &= \sum p(y_{t+1},\mathbf{x}_t,\mathbf{y}_t)\log\frac{p(y_{t+1}|\mathbf{x}_t,\mathbf{y}_t)}{p(y_{t+1}|\mathbf{y}_t)}\end{aligned}$$

Joint entropies (and distributions) can have high dimension!

Entropy estimates from nearest neighbors [Kraskov et al, 2004]

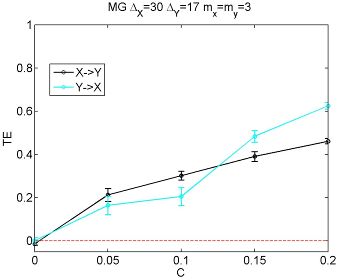
TE is equivalent to GCI when the stochastic process of (X, Y) is Gaussian [Barnett et al, PRE 2009]

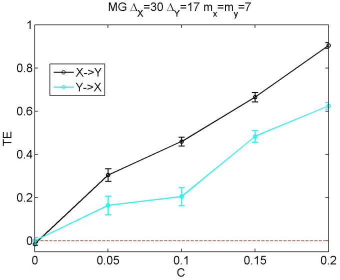
Entropy estimates from nearest neighbors [Kraskov et al, 2004]

What are the appropriate embedding parameters?

Example: Unidirectionally coupled Mackey-Glass system

$$
\begin{aligned}
\dot{x}(t) &= \frac{0.2x(t-\Delta_x)}{1+x(t-\Delta_x)^{10}} - 0.1x(t) \\
\dot{y}(t) &= \frac{0.2y(t-\Delta_y)}{1+y(t-\Delta_y)^{10}} - 0.1y(t) + C\frac{x(t-\Delta_x)}{1+x(t-\Delta_x)^{10}}.
\end{aligned}
$$





Nonlinear causality measures (direct)

driving system: X, response system: Y,
conditioning on system Z, $Z = \{Z_1, Z_2, \ldots, Z_{K-2}\}$
join all $K-2$ z-reconstructed vectors: $\mathbf{Z}_t = [\mathbf{z}_{1,t}, \ldots, \mathbf{z}_{K-2,t}]$

Partial Transfer Entropy (PTE) [Vakorin et al, 2009; Papana et al, 2012]

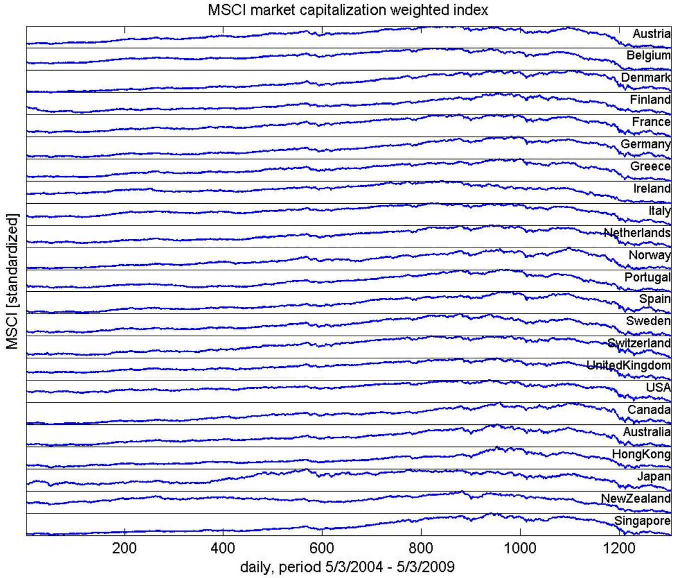
Measure the effect of X on Y at T times ahead, accounting (conditioning) for the effect from its own current state and the current state of the other variables except X.

$$\mathrm{PTE}_{X \to Y|Z} = I(y_{t+1}; \mathbf{x}_t | \mathbf{y}_t, \mathbf{Z}_t)$$

$$= H(\mathbf{x}_t, \mathbf{y}_t | \mathbf{Z}_t) - H(y_{t+1}, \mathbf{x}_t, \mathbf{y}_t | \mathbf{Z}_t) + H(y_{t+1}, \mathbf{y}_t | \mathbf{Z}_t) - H(\mathbf{y}_t | \mathbf{Z}_t)$$

Joint entropies (and distributions) can have very high dimension!

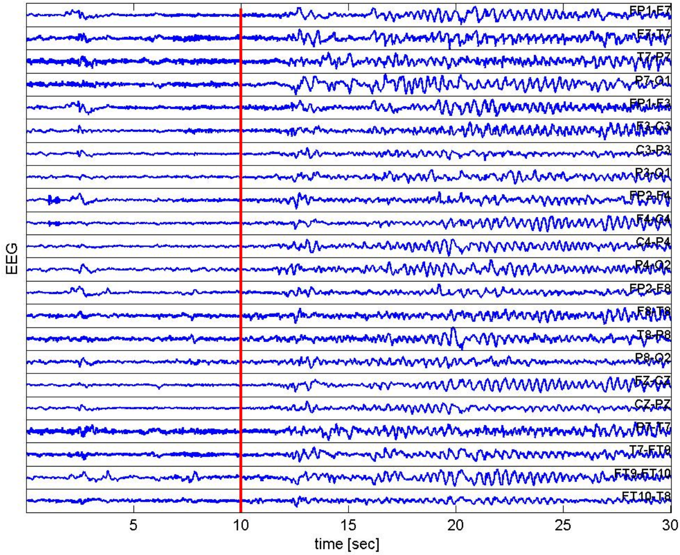
Example: Global financial market



Data source: `https://www.msci.com/market-cap-weighted-indexes`

Network ?

Example: Brain dynamical system

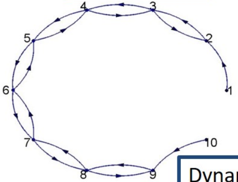


Data source: `https://physionet.org/pn6/chbmit/chb08/`

Network ?

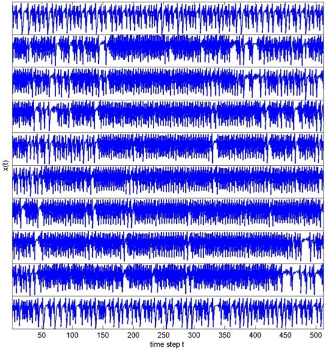
Complex networks from multivariate time series

True network



Dynamics
$\boldsymbol{x}_{t+1} = \boldsymbol{f}(\boldsymbol{x}_t)$

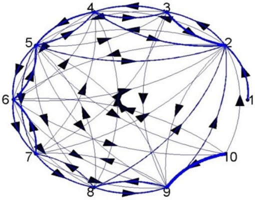
Multivariate time series



Causality measures
$X \rightarrow Y$, $X \rightarrow Y|Z$

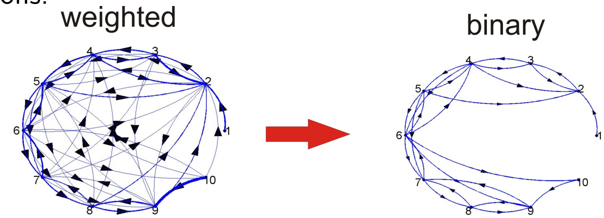
GCI, CGCI
TE, PTE

Estimated network (weighted connections)



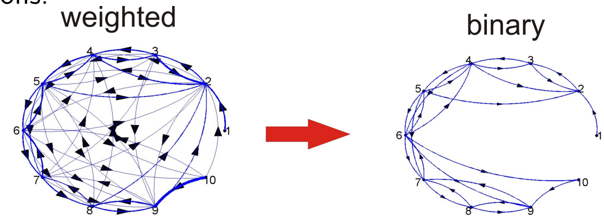
How to assess the presence of a connection?

Three possible ways to convert a network of weighted connections (the Granger causality measure) to a network of binary connections:



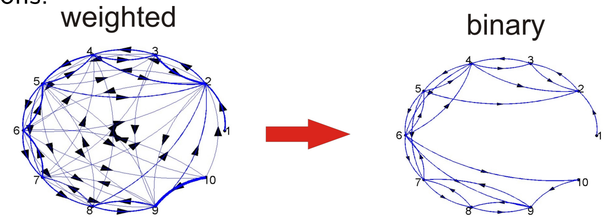
How to assess the presence of a connection?

Three possible ways to convert a network of weighted connections (the Granger causality measure) to a network of binary connections:



1. Threshold on the measure magnitude, $q(i \to j) > \text{thr}$.

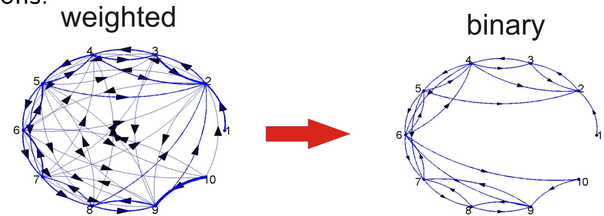
Three possible ways to convert a network of weighted connections (the Granger causality measure) to a network of binary connections:



1. Threshold on the measure magnitude, $q(i \rightarrow j) > \text{thr}$.
2. Threshold on the network density, only the $d\%$ largest $q(i \rightarrow j)$.

How to assess the presence of a connection?

Three possible ways to convert a network of weighted connections (the Granger causality measure) to a network of binary connections:

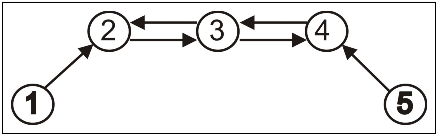


1. Threshold on the measure magnitude, $q(i \to j) > \text{thr}$.
2. Threshold on the network density, only the $d\%$ largest $q(i \to j)$.
3. Significance test on each $q(i \to j)$. Threshold, e.g. $\alpha = 0.05$ on the p-value of the test.
   Parametric or resampling test (resampling test for a nonlinear causality measure).

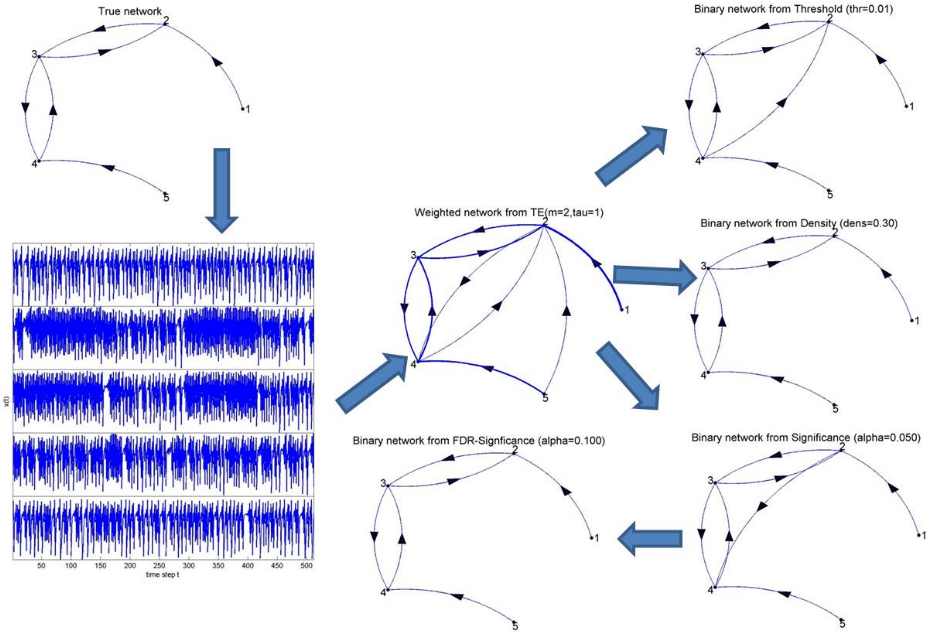
$$
\begin{aligned}
x_{1,t+1} &= 1.4 - x_{1,t}^2 + 0.3x_{1,t-1} \\
x_{i,t+1} &= 1.4 - (0.5C(x_{i-1,t} + x_{i+1,t}) + (1-C)x_{i,t})^2 + 0.3x_{i,t-1} \\
x_{K,t+1} &= 1.4 - x_{K,t}^2 + 0.3x_{K,t-1}
\end{aligned}
$$

C: coupling strength [Politi & Torcini, 1992]

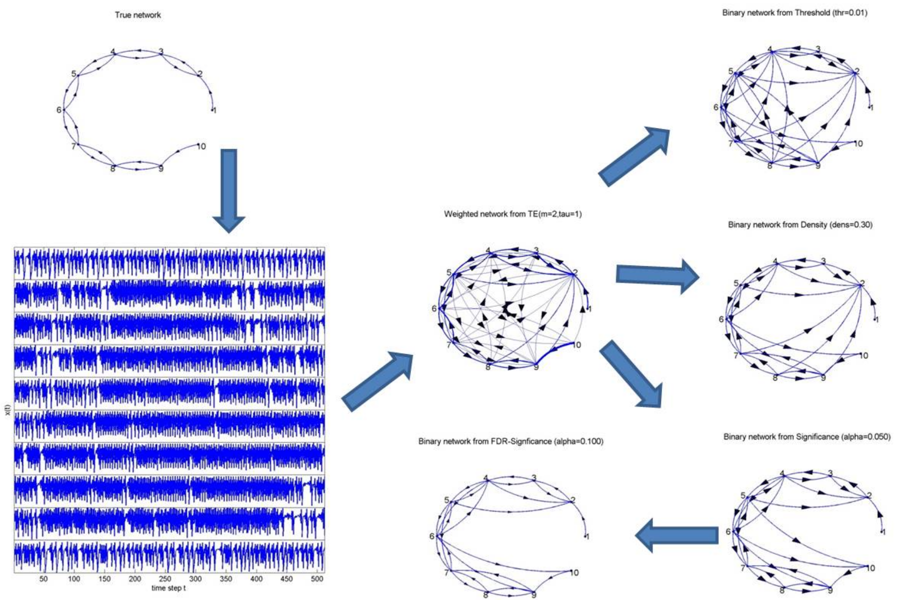
Network structure for $K = 5$



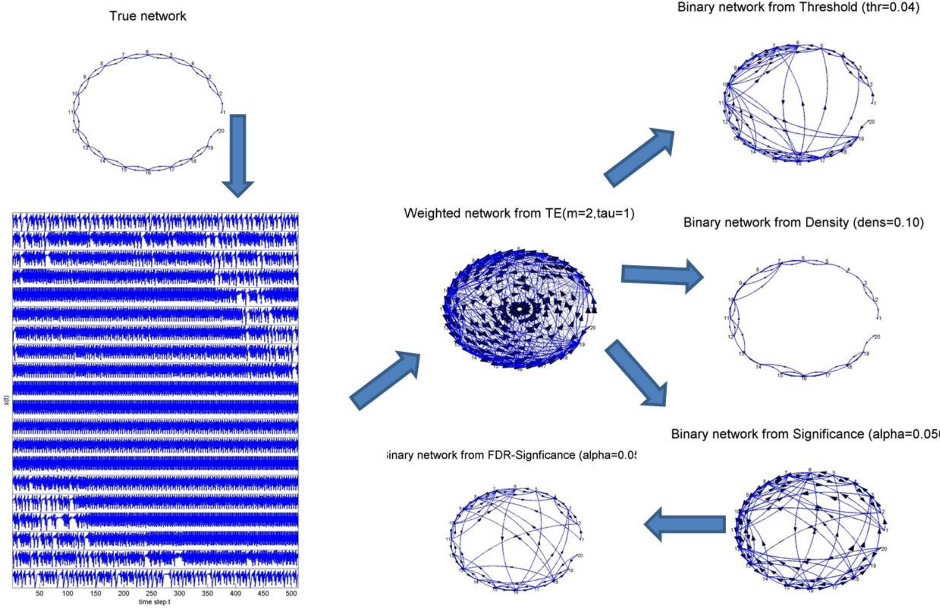
Example, TE, $K = 5$



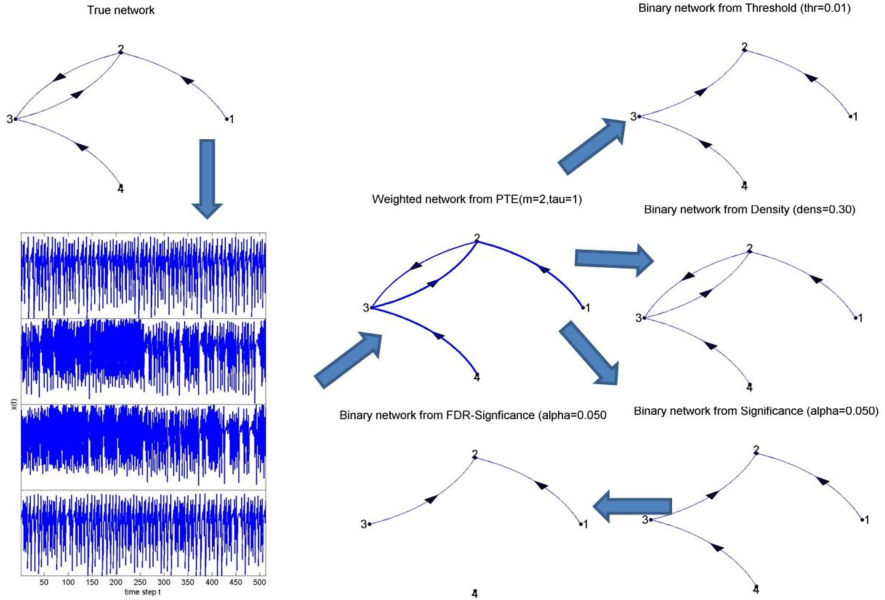
Example, TE, $K = 10$



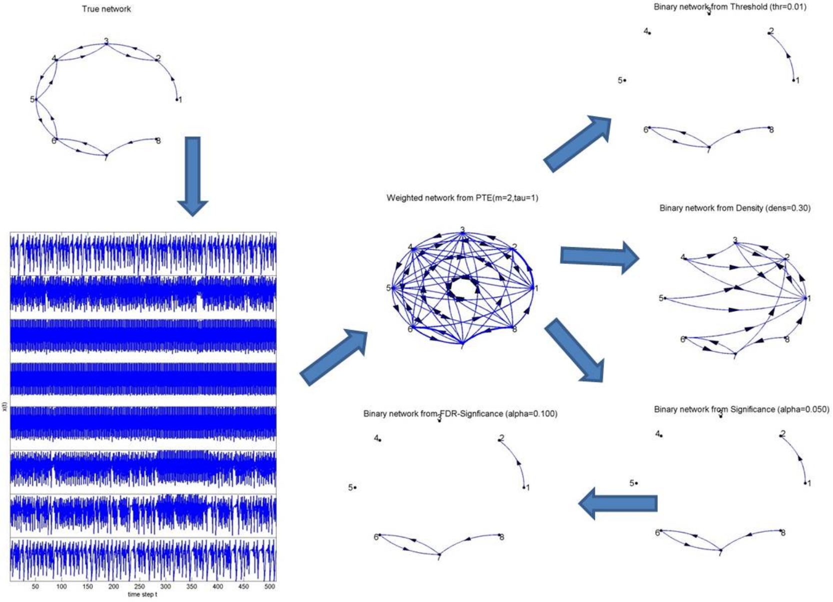
Example, TE, $K = 20$



Example, PTE, $K = 4$



Example, PTE, $K = 8$



MIME applies dimension reduction and then uses conditional mutual information. The idea: [Vlachos & Kugiumtzis, PRE, 2010]

MIME applies dimension reduction and then uses conditional mutual information. The idea: [Vlachos & Kugiumtzis, PRE, 2010]

1. Find a subset $\mathbf{w}_t$ of lagged variables from X and Y that explains best the future of Y, y_{t+1}.

MIME applies dimension reduction and then uses conditional mutual information. The idea: [Vlachos & Kugiumtzis, PRE, 2010]

1. Find a subset $\mathbf{w}_t$ of lagged variables from X and Y that explains best the future of Y, y_{t+1}.
2. Quantify the information on Y ahead that is explained by the X-components in this subset.

MIME applies dimension reduction and then uses conditional mutual information. The idea: [Vlachos & Kugiumtzis, PRE, 2010]

1. Find a subset $\mathbf{w}_t$ of lagged variables from X and Y that explains best the future of Y, y_{t+1}.
2. Quantify the information on Y ahead that is explained by the X-components in this subset.

If there are no components of X in $\mathbf{w}_t$, then MIME $= 0$.

Mutual Information from Mixed Embedding - 2

$$\mathbf{w}_t = (\underbrace{x_{t-\tau_{x1}}, x_{t-\tau_{x2}}, \ldots, x_{t-\tau_{xm_x}}}_{\mathbf{w}_t^x}, \underbrace{y_{t-\tau_{y1}}, y_{t-\tau_{y2}}, \ldots, y_{t-\tau_{ym_y}}}_{\mathbf{w}_t^y})$$

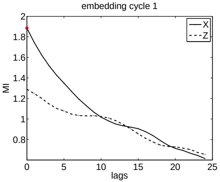
Measure 3a: The causality measure MIME

$$R_{X\to Y} = \frac{I(\mathbf{y}_t^T; \mathbf{w}_t^x \mid \mathbf{w}_t^y)}{I(\mathbf{y}_t^T; \mathbf{w}_t)}$$

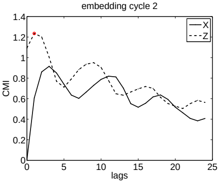
- $R_{X\to Y}$: information of Y explained only by X-components of the embedding vectors, normalized against the total mutual information (in order to give a value between 0 and 1).
- If $\mathbf{w}_t$ contains no components from X, then $R_{X\to Y} = 0$ and X has no effect on the future of Y.

Example: Embedding from X and Z variables of the chaotic Lorenz system to explain X, $W_t = \{x_t, \ldots, x_{t-24}, z_t, \ldots, z_{t-24}\}$ $\mathbf{y}_t^T = (x_{t+1}, \ldots, x_{t+5})$, $N = 10000$, sampling time $\tau_s = 0.05$

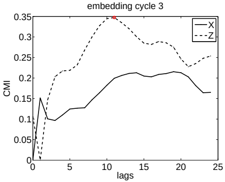
$x_t = \arg\max \left\{ I\left(\mathbf{y}_t^T; w_t\right) \right\}$,
$w_t \in W_t$



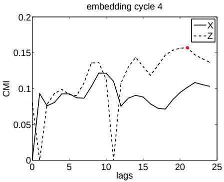
$z_{t-1} =$
$\arg\max \left\{ I\left(\mathbf{y}_t^T; w_t \mid x_t\right) \right\}$

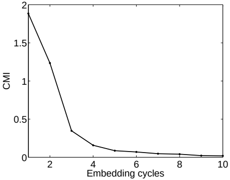


$z_{t-11} = \arg\max \left\{ I\left(\mathbf{y}_t^T; w_t \mid x_t, z_{t-1}\right) \right\}$



Too small increase in CMI



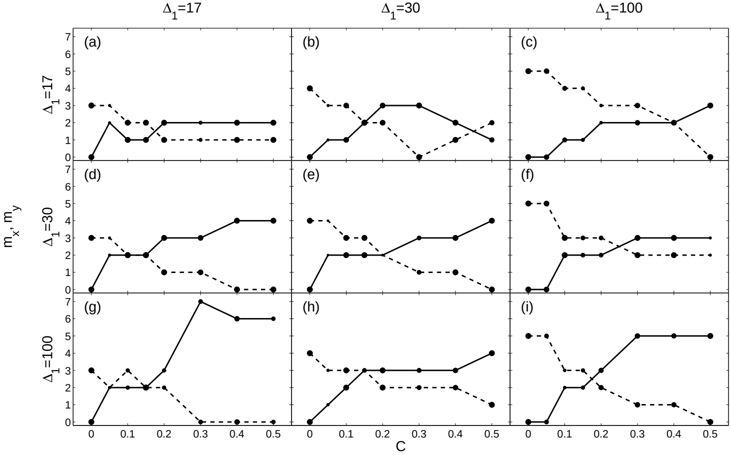


Embedding vector:
$\mathbf{w}_t = (x_t, z_{t-1}, z_{t-11})$

Example: Coupled Mackey-Glass system

$\Delta = 17, 30, 100, \qquad N = 4096$

$\mathbf{y}_t^T = \{y_{t+1}, y_{t+\tau_1}, y_{t+\tau_2}\}, \qquad L_x = L_y = 50$



solid line: driving system dashed line: response system

Partial Mutual Information from Mixed Embedding - 1

driving system: X, response system: Y,
conditioning on system Z, $Z = \{Z_1, Z_2, \ldots, Z_{K-2}\}$

The same non-uniform embedding scheme for explaining $\mathbf{y}_t^T$ from vector of lags of all $X, Y, Z_1, Z_2, \ldots, Z_{K-2}$,

$W_t =$
$\{x_t, \ldots, x_{t-L_x-1}, y_t, \ldots, y_{t-L_y-1}, z_{1,t}, \ldots, z_{1,t-L_z-1}, \ldots, z_{K-2,t-L_z-1}\}$
e.g., for $K = 3$, X, Y, Z:

$$\mathbf{w}_t = (\underbrace{x_{t-\tau_{x1}}, \ldots, x_{t-\tau_{xm_x}}}_{\mathbf{w}_t^x}, \underbrace{y_{t-\tau_{y1}}, \ldots, y_{t-\tau_{ym_y}}}_{\mathbf{w}_t^y}, \underbrace{z_{t-\tau_{z1}}, \ldots, z_{t-\tau_{zm_z}}}_{\mathbf{w}_t^z})$$

Partial Mutual Information from Mixed Embedding - 2

The non-uniform embedding vector of lags of all X, Y, Z for explaining y_{t+1}:

$$\mathbf{w}_t = (\underbrace{x_{t-\tau_{x1}}, \ldots, x_{t-\tau_{xm_x}}}_{\mathbf{w}_t^x}, \underbrace{y_{t-\tau_{y1}}, \ldots, y_{t-\tau_{ym_y}}}_{\mathbf{w}_t^y}, \underbrace{z_{t-\tau_{z1}}, \ldots, z_{t-\tau_{zm_z}}}_{\mathbf{w}_t^z})$$

Partial Mutual Information from Mixed Embedding - 2

The non-uniform embedding vector of lags of all X, Y, Z for explaining y_{t+1}:

$$\mathbf{w}_t = (\underbrace{x_{t-\tau_{x1}}, \ldots, x_{t-\tau_{xm_x}}}_{\mathbf{w}_t^x}, \underbrace{y_{t-\tau_{y1}}, \ldots, y_{t-\tau_{ym_y}}}_{\mathbf{w}_t^y}, \underbrace{z_{t-\tau_{z1}}, \ldots, z_{t-\tau_{zm_z}}}_{\mathbf{w}_t^z})$$

The causality measure PMIME

$$R_{X \to Y|Z} = \frac{I(y_{t+1}; \mathbf{w}_t^x \mid \mathbf{w}_t^y, \mathbf{w}_t^Z)}{I(y_{t+1}; \mathbf{w}_t)}$$

- $R_{X \to Y|Z}$: information on the future of Y explained only by X-components of the embedding vector (given the components of Y and Z), normalized with the mutual information of the future of Y and the embedding vector.

Partial Mutual Information from Mixed Embedding - 2

The non-uniform embedding vector of lags of all X, Y, Z for explaining y_{t+1}:

$$\mathbf{w}_t = (\underbrace{x_{t-\tau_{x1}}, \ldots, x_{t-\tau_{xm_x}}}_{\mathbf{w}_t^x}, \underbrace{y_{t-\tau_{y1}}, \ldots, y_{t-\tau_{ym_y}}}_{\mathbf{w}_t^y}, \underbrace{z_{t-\tau_{z1}}, \ldots, z_{t-\tau_{zm_z}}}_{\mathbf{w}_t^z})$$

The causality measure PMIME

$$R_{X\rightarrow Y|Z} = \frac{I(y_{t+1}; \mathbf{w}_t^x \mid \mathbf{w}_t^y, \mathbf{w}_t^Z)}{I(y_{t+1}; \mathbf{w}_t)}$$

- $R_{X\rightarrow Y|Z}$: information on the future of Y explained only by X-components of the embedding vector (given the components of Y and Z), normalized with the mutual information of the future of Y and the embedding vector.
- If $\mathbf{w}_t^Z = \emptyset$, then $R_{X\rightarrow Y|Z} = R_{X\rightarrow Y}$.

Partial Mutual Information from Mixed Embedding - 2

The non-uniform embedding vector of lags of all X, Y, Z for explaining y_{t+1}:

$$\mathbf{w}_t = (\underbrace{x_{t-\tau_{x1}}, \ldots, x_{t-\tau_{xm_x}}}_{\mathbf{w}_t^x}, \underbrace{y_{t-\tau_{y1}}, \ldots, y_{t-\tau_{ym_y}}}_{\mathbf{w}_t^y}, \underbrace{z_{t-\tau_{z1}}, \ldots, z_{t-\tau_{zm_z}}}_{\mathbf{w}_t^z})$$

The causality measure PMIME

$$R_{X \to Y|Z} = \frac{I(y_{t+1}; \mathbf{w}_t^x \mid \mathbf{w}_t^y, \mathbf{w}_t^Z)}{I(y_{t+1}; \mathbf{w}_t)}$$

- $R_{X \to Y|Z}$: information on the future of Y explained only by X-components of the embedding vector (given the components of Y and Z), normalized with the mutual information of the future of Y and the embedding vector.
- If $\mathbf{w}_t^Z = \emptyset$, then $R_{X \to Y|Z} = R_{X \to Y}$.
- If $\mathbf{w}_t$ contains no components from X, then $R_{X \to Y|Z} = 0$ and X has no direct effect on the future of Y.

Three main advantages of PMIME

Partial Mutual Information from Mixed Embedding - 3

Three main advantages of PMIME

- $R_{X \to Y|Z} = 0$ when no significant causality is present, and $R_{X \to Y|Z} > 0$ when it is present
  [no significance test, no issues with multiple testing!]

Partial Mutual Information from Mixed Embedding - 3

Three main advantages of PMIME

- $R_{X \to Y|Z} = 0$ when no significant causality is present, and $R_{X \to Y|Z} > 0$ when it is present
  [no significance test, no issues with multiple testing!]
- mixed embedding for all variables is formed as part of the measure [it does not require the determination of embedding parameters for each variable]

Partial Mutual Information from Mixed Embedding - 3

Three main advantages of PMIME

- $R_{X \to Y|Z} = 0$ when no significant causality is present, and $R_{X \to Y|Z} > 0$ when it is present
  [no significance test, no issues with multiple testing!]
- mixed embedding for all variables is formed as part of the measure [it does not require the determination of embedding parameters for each variable]
- inclusion of more confounding variables only slows the computation and has no effect on statistical accuracy
  [**no "curse of dimensionality"** for any dimension of Z, only slow computation time]

Partial Mutual Information from Mixed Embedding - 3

Three main advantages of PMIME

- $R_{X \to Y|Z} = 0$ when no significant causality is present, and $R_{X \to Y|Z} > 0$ when it is present
  [no significance test, no issues with multiple testing!]
- mixed embedding for all variables is formed as part of the measure [it does not require the determination of embedding parameters for each variable]
- inclusion of more confounding variables only slows the computation and has no effect on statistical accuracy
  [**no "curse of dimensionality"** for any dimension of Z, only slow computation time]

$\Rightarrow$ good candidate for causality analysis with many variables

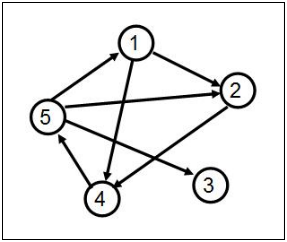
Example: linear coupled system

$K = 5$ linear Vector Autoregressive process, VAR(4) in 5 variables

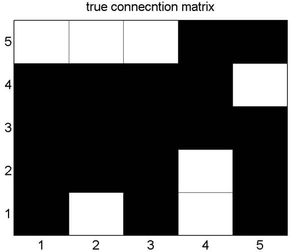
$$
\begin{aligned}
x_{1,t} &= 0.4x_{1,t-1} - 0.5x_{1,t-2} + 0.4x_{5,t-1} + e_{1,t} \\
x_{2,t} &= 0.4x_{2,t-1} - 0.3x_{1,t-4} + 0.4x_{5,t-2} + e_{2,t} \\
x_{3,t} &= 0.5x_{3,t-1} - 0.7x_{3,t-2} - 0.3x_{5,t-3} + e_{3,t} \\
x_{4,t} &= 0.8x_{4,t-3} + 0.4x_{1,t-2} + 0.3x_{2,t-2} + e_{4,t} \\
x_{5,t} &= 0.7x_{5,t-1} - 0.5x_{5,t-2} - 0.4x_{4,t-1} + e_{5,t}
\end{aligned}
$$

[Schelter et al, 2006]

Network with directed links

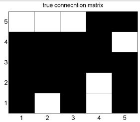


Causality matrix

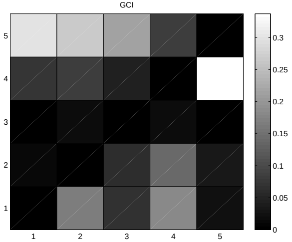


Linear VAR(4) in 5 variables

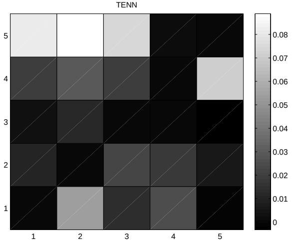
$n = 1000,\ p = m = L_x = L_y = 5,\ T = 1$



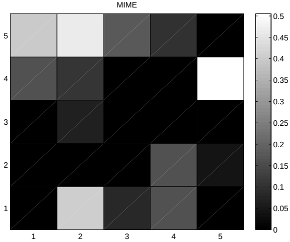
GCI



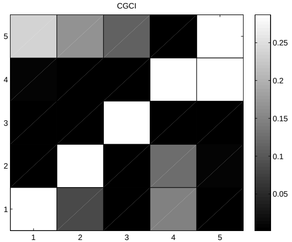
TE



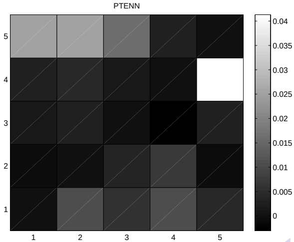
MIME



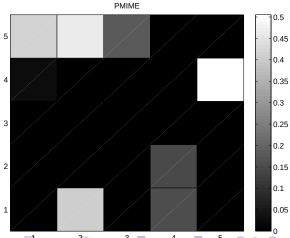
CGCI



PTE



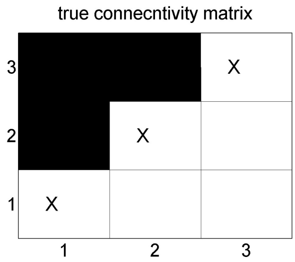
PMIME

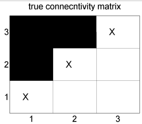


Nonlinear stochastic map:

$$
\begin{aligned}
x_{1,t} &= 3.4x_{1,t-1}(1 - x_{1,t-1}^2)e^{-x_{1,t-1}^2} + 0.4e_{1,t} \\
x_{2,t} &= 3.4x_{2,t-1}(1 - x_{2,t-1}^2)e^{-x_{2,t-1}^2} + 0.5x_{1,t-1}x_{2,t-1} + 0.4e_{2,t} \\
x_{3,t} &= 3.4x_{3,t-1}(1 - x_{3,t-1}^2)e^{-x_{3,t-1}^2} + 0.3x_{2,t-1} + 0.5x_{1,t-1}^2 + 0.4e_{3,t}
\end{aligned}
$$

$n = 512$ [Model 7, Gourevich et al, 2006]

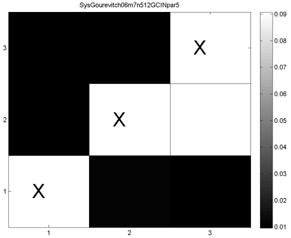




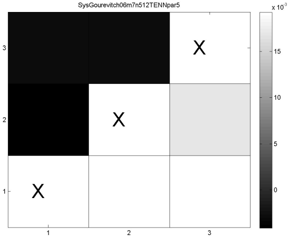
Nonlinear stochastic map

$n = 512,\ p = m = 5,\ L_x = L_y = 5,\ T = 1$

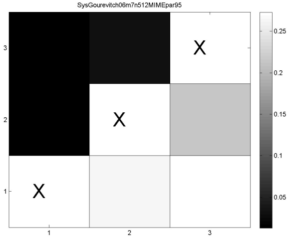
GCI



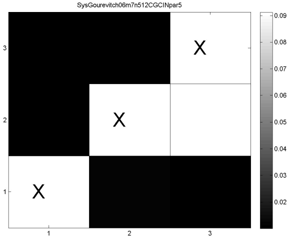
TE



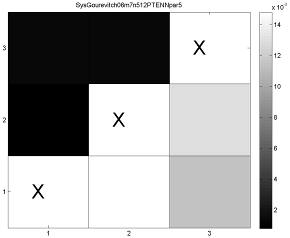
MIME



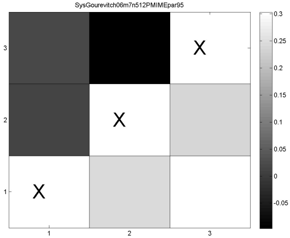
CGCI



PTE



PMIME

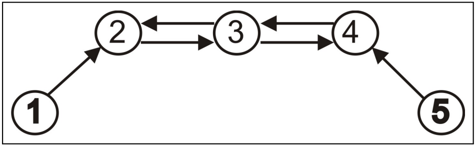


$K = 5$ Henon coupled maps

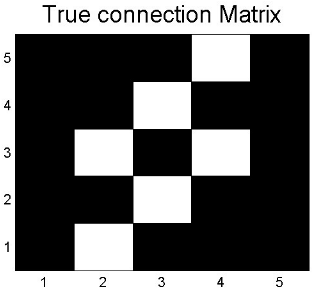
$$\begin{aligned} x_{1,t+1} &= 1.4 - x_{1,t}^2 + 0.3x_{1,t-1} \\ x_{i,t+1} &= 1.4 - (0.5C(x_{i-1,t} + x_{i+1,t}) + (1-C)x_{i,t})^2 + 0.3x_{i,t-1} \\ x_{K,t+1} &= 1.4 - x_{K,t}^2 + 0.3x_{K,t-1} \end{aligned}$$

coupling strength: $C = 0, \ldots, 0.9, \quad n = 4096$

Network with directed links

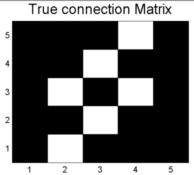


Causality matrix (not symmetric)



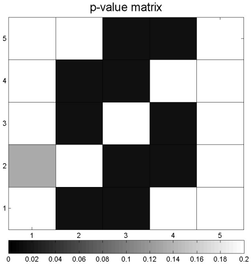
$N = 4096$, $M = 100$ weak coupling $C = 0.2$

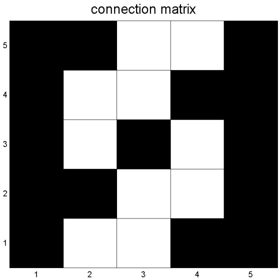
$\text{TE}_{X \to Y}$, $m = 2$



matrix of p-values from the randomization test

$$p(\text{TE}_{X \to Y}) = \begin{bmatrix} & 0.0133 & 0.0133 & 0.2106 & 0.3093 \\ 0.1317 & & 0.0133 & 0.0133 & 0.3685 \\ 0.7237 & 0.0133 & & 0.0133 & 0.7040 \\ 0.7632 & 0.0133 & 0.0133 & & 0.5264 \\ 0.3685 & 0.4080 & 0.0133 & 0.0133 & \end{bmatrix}$$



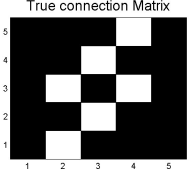


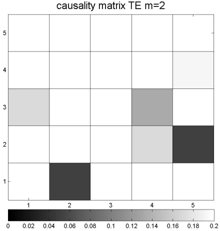
$N = 4096,\ M = 100$

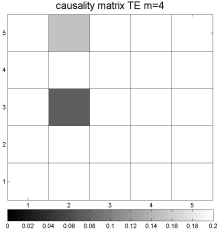
TE for $m_x = m_y = 2$ and $m_x = m_y = 4$

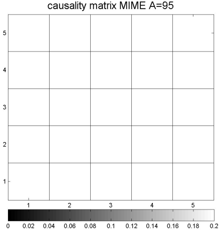
MIME for $L_x = L_y = 5$

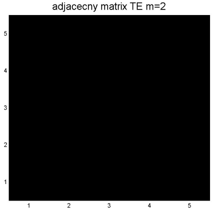
$C = 0.0$

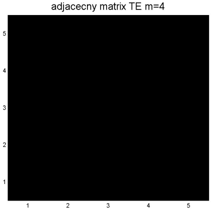


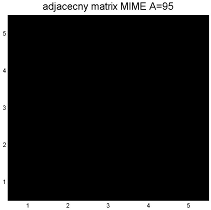










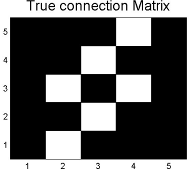


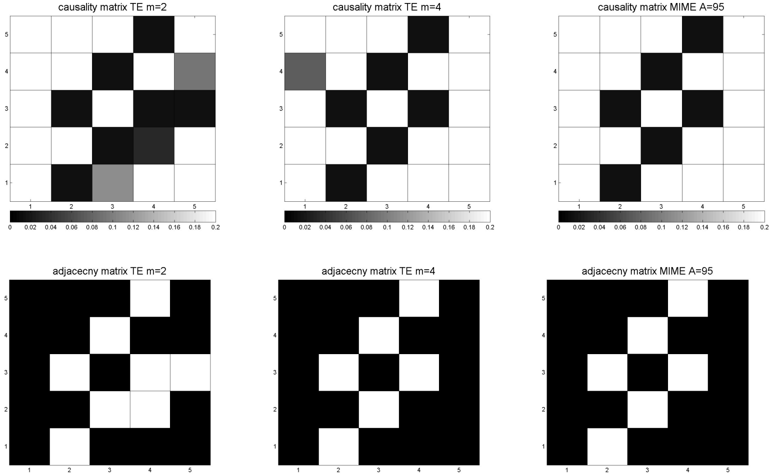
$N = 4096, \; M = 100$

TE for $m_x = m_y = 2$ and $m_x = m_y = 4$

MIME for $L_x = L_y = 5$

$C = 0.1$



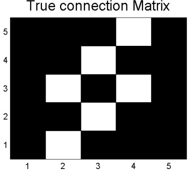


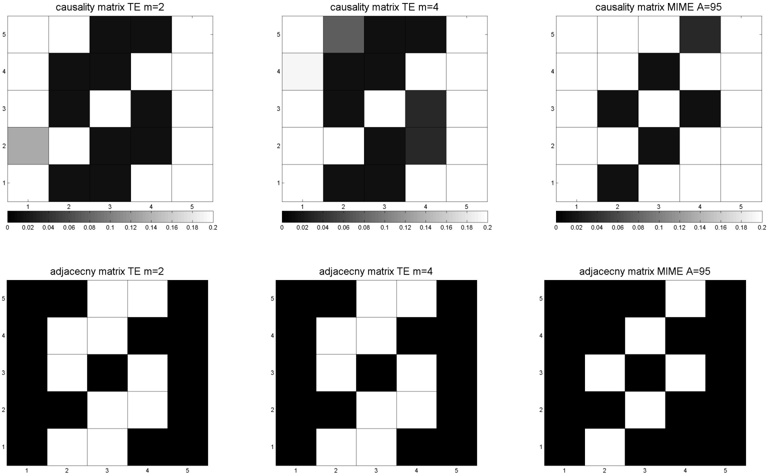
$N = 4096,\ M = 100$

TE for $m_x = m_y = 2$ and $m_x = m_y = 4$

MIME for $L_x = L_y = 5$

$C = 0.2$



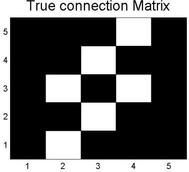


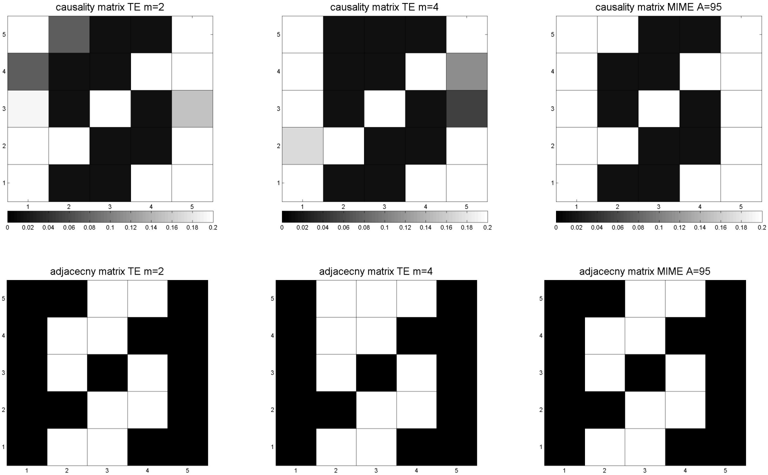
$N = 4096$, $M = 100$

TE for $m_x = m_y = 2$ and $m_x = m_y = 4$

MIME for $L_x = L_y = 5$

$C = 0.4$



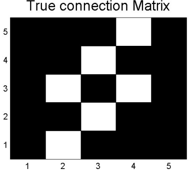


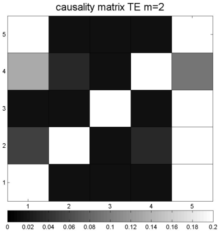
$N = 4096,\ M = 100$

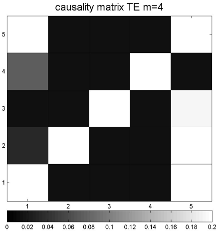
TE for $m_x = m_y = 2$ and $m_x = m_y = 4$

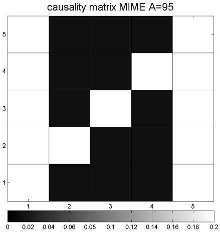
MIME for $L_x = L_y = 5$

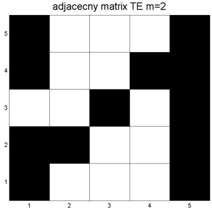
$C = 0.6$

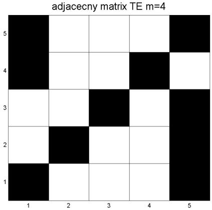


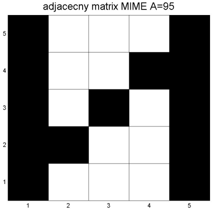










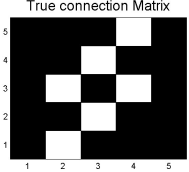


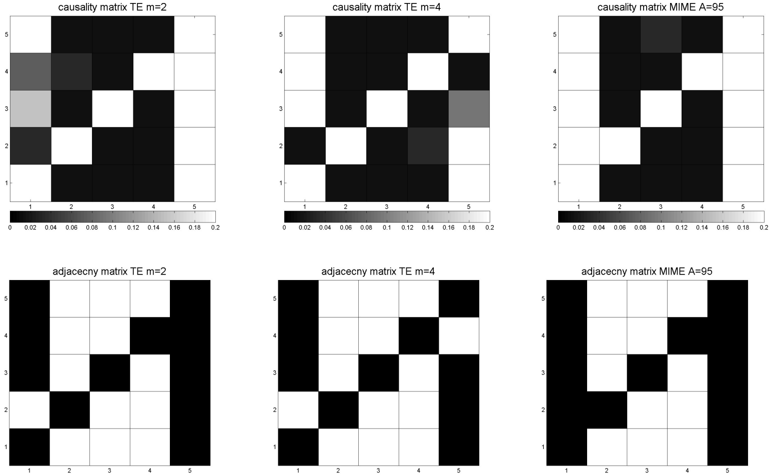
$N = 4096, \; M = 100$

TE for $m_x = m_y = 2$ and $m_x = m_y = 4$

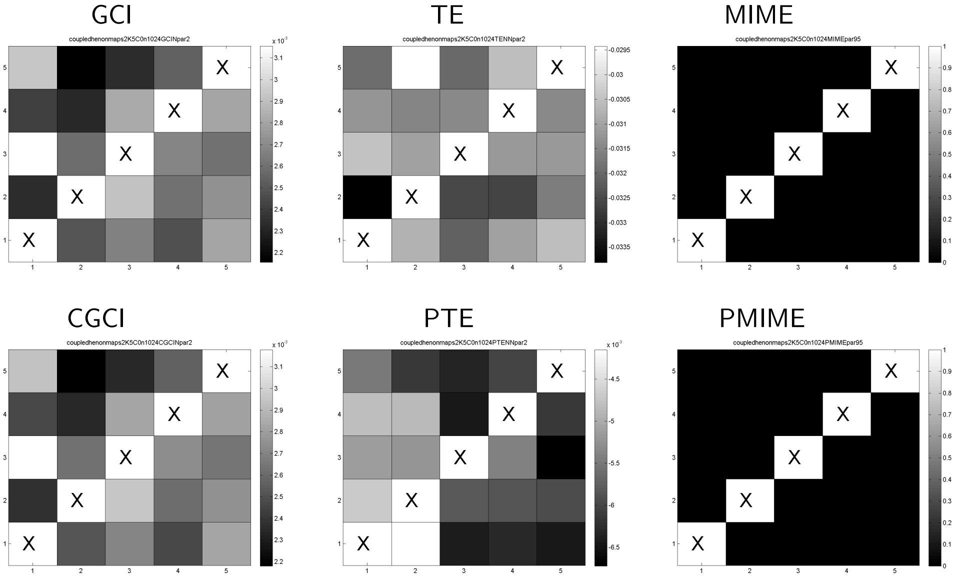
MIME for $L_x = L_y = 5$

$C = 0.9$

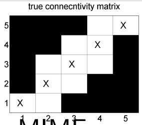




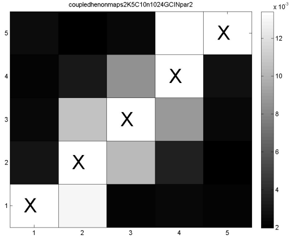
$C = 0.0$ $n = 1024,\ p = 5,\ m = 2,\ L = 5,\ T = 1$



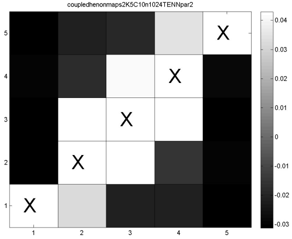
$C = 0.1 \; n = 1024, \; p = 5, \; m = 2, \; L = 5, \; T = 1$



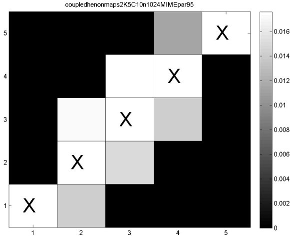
GCI



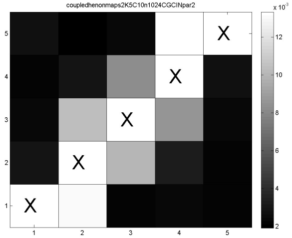
TE



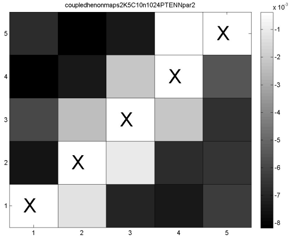
MIME



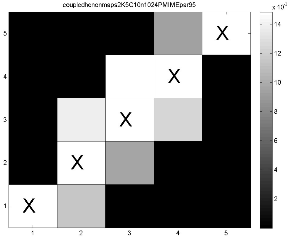
CGCI



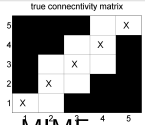
PTE



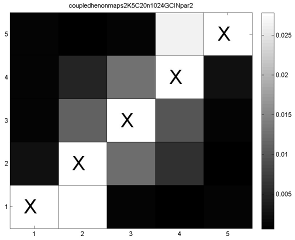
PMIME



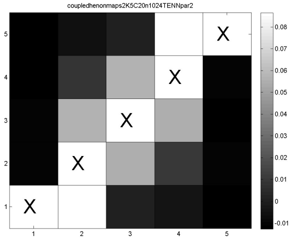
$C = 0.2$ $n = 1024,\ p = 5,\ m = 2,\ L = 5,\ T = 1$



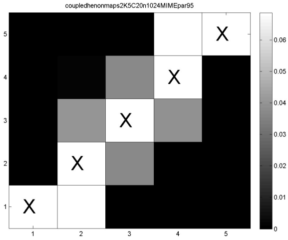
GCI



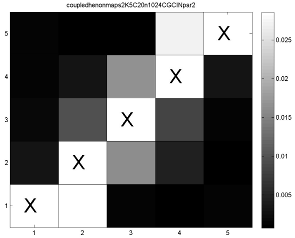
TE



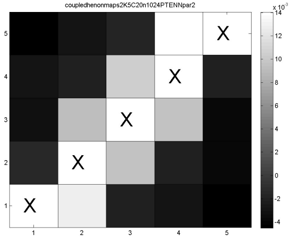
MIME



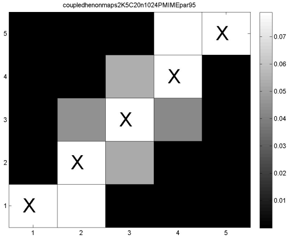
CGCI

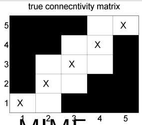


PTE



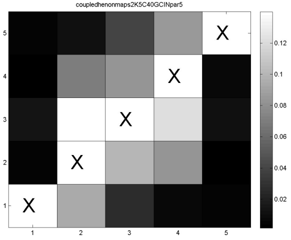
PMIME



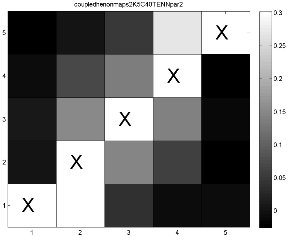


$C = 0.4$ $n = 1024,\ p = 5,\ m = 2,\ L = 5,\ T = 1$

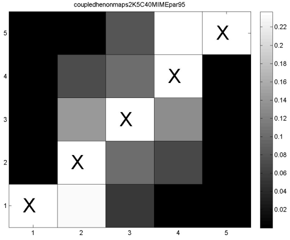
GCI



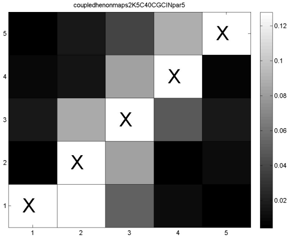
TE



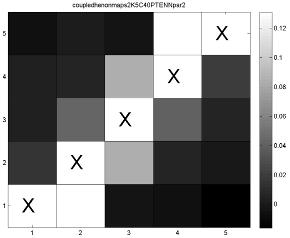
MIME



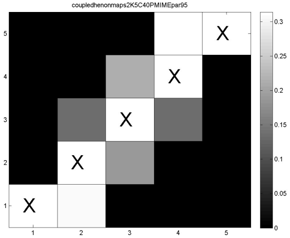
CGCI



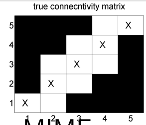
PTE



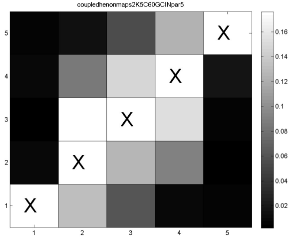
PMIME



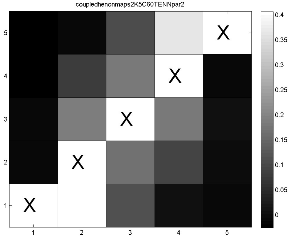
$C = 0.6$ $n = 1024,\ p = 5,\ m = 2,\ L = 5,\ T = 1$



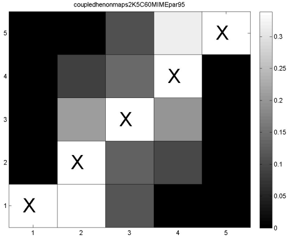
GCI



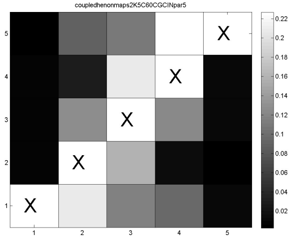
TE



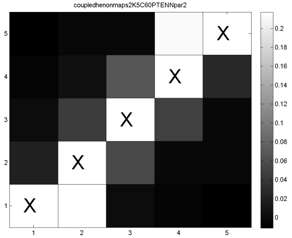
MIME



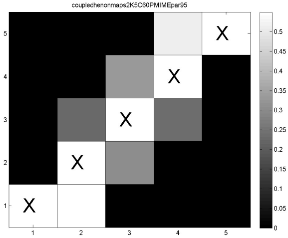
CGCI



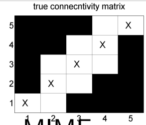
PTE



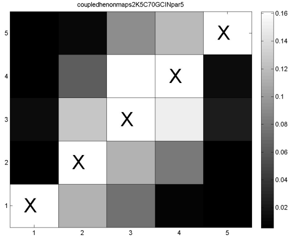
PMIME



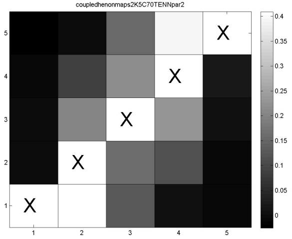
$C = 0.7$ $n = 1024,\ p = 5,\ m = 2,\ L = 5,\ T = 1$



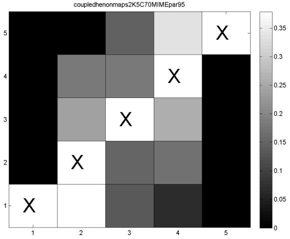
GCI



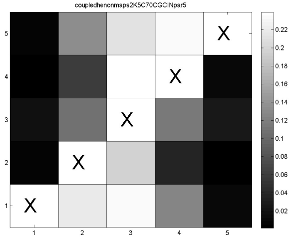
TE



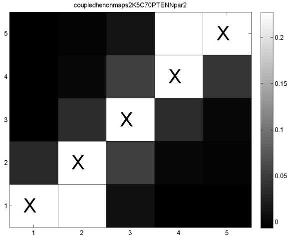
MIME



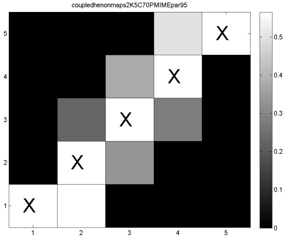
CGCI

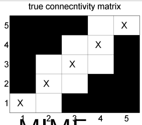


PTE



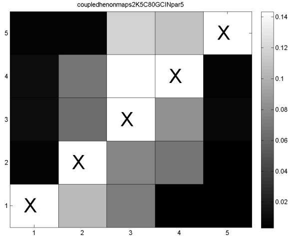
PMIME



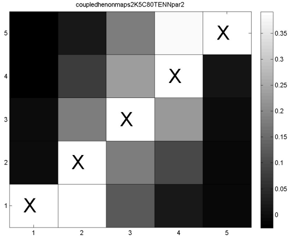


$C = 0.8$ $n = 1024, \; p = 5, \; m = 2, \; L = 5, \; T = 1$

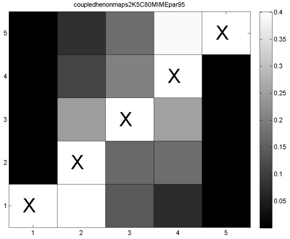
GCI



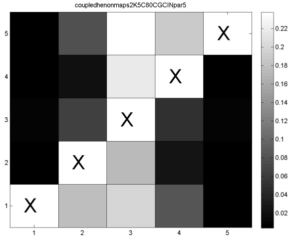
TE



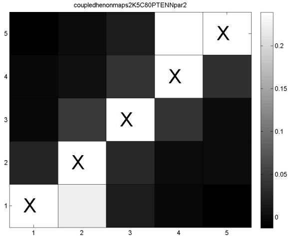
MIME



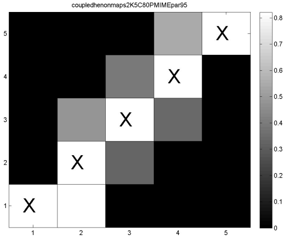
CGCI



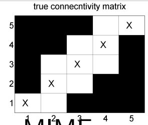
PTE



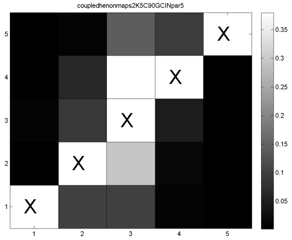
PMIME



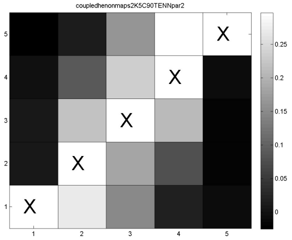
$C = 0.9$ $n = 1024,\ p = 5,\ m = 2,\ L = 5,\ T = 1$



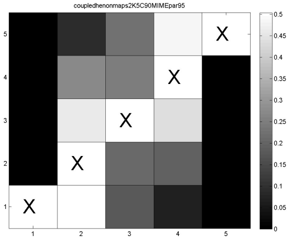
GCI



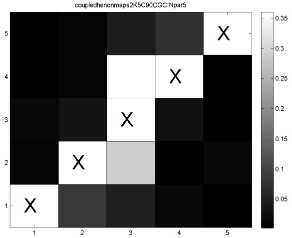
TE



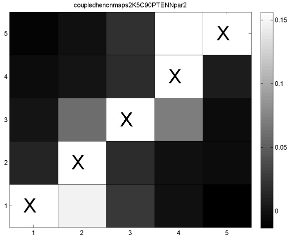
MIME



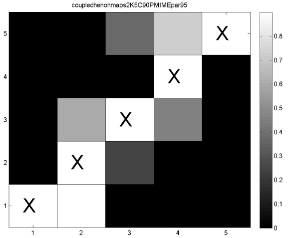
CGCI



PTE



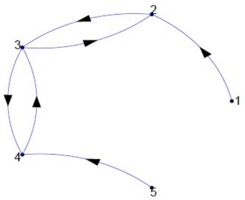
PMIME



Coupled identical Mackey-Glass delayed differential equations

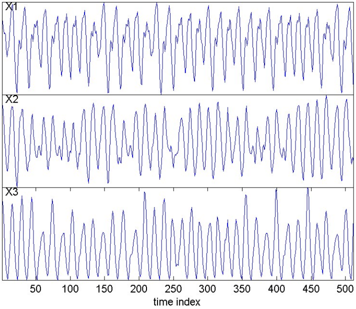
$$\dot{x}_i(t) = -0.1 x_i(t) + \sum_{j=1}^{K} \frac{C_{ij} x_j(t-\Delta)}{1 + x_j(t-\Delta)^{10}} \quad \text{for} \quad i = 1, \ldots, K$$

$K = 5$



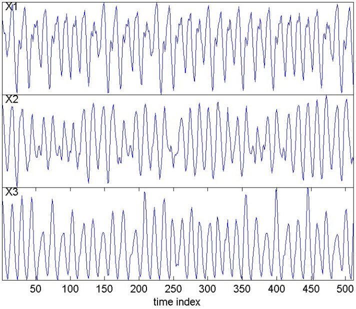
Mackey-Glass, $C = 0.2$

$\Delta = 20$

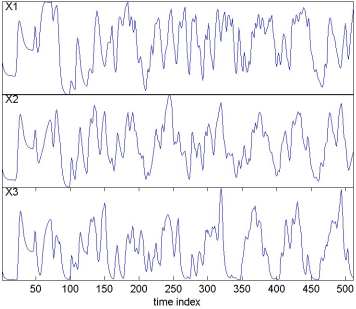


Mackey-Glass, $C = 0.2$

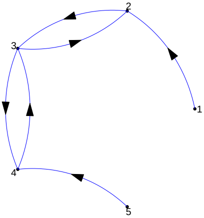
$\Delta = 20$

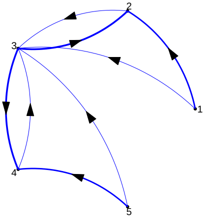


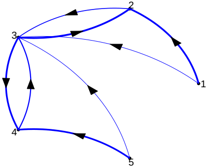
$\Delta = 100$



$K = 5$ True from PMIME ($\Delta = 20$) from PMIME ($\Delta = 100$)

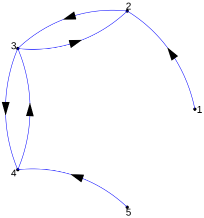


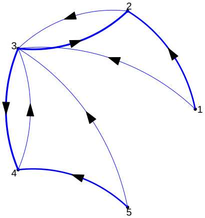


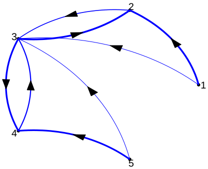


Mackey-Glass: true/estimated network [Kugiumtzis and Kimiskidis, IJNS 2015]

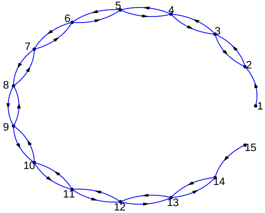
$K = 5$ True | from PMIME ($\Delta = 20$) | from PMIME ($\Delta = 100$)

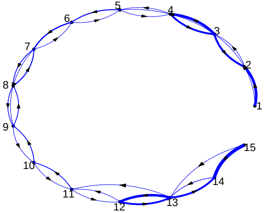


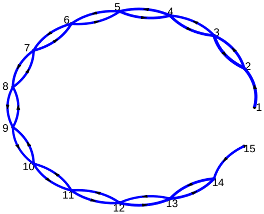




$K = 15$ True | from PMIME ($\Delta = 20$) | from PMIME ($\Delta = 100$)







Network indices

Symbol	Description
deg^m	degree distribution, m=mean,std,skewness,kurtosis
str^m	strength distribution, m=mean,std,skewness,kurtosis
TrR_k	transitivity ratio, k=binary undirected (bu),binary directed (bd) weighted directed (wd)
$EigC^m$	eigenvector centrality distribution, m=mean,std
λ_k	characteristic path length, k=bd,wd
GE_k	global efficiency, k=bd,wd
ϵ_k^m	eccentricity distribution, m=mean,std and k=bd,wd
rad_k	radius, k=bd,wd
d_k	diameter, k=bd,wd
C_k^m	clustering coefficient distribution,m=mean,std and k=bd,wd
g_k^m	betweenness centrality distribution,m=mean,std and k=bd,wd
$e-g_k^m$	edge betweenness centrality distribution,m=mean,std and k=bd,wd
LE_k^m	local efficiency distribution,m=mean,std and k=bd,wd
$3motif(i)$	i^{th} motif of 3 nodes, i=1,2,...13
$modul(i)$	modularity for i modules, i=2,3,5
$r_{deg}(i,j)$	assortativity coefficient in terms of the degree, i=in,out and j=in,out or i,j=und
$r_{str}(i,j)$	assortativity coefficient in terms of the strength, i=in,out and j=in,out or i,j=und
p_{top}	Rent exponent:topological
p_{ph}	Rent exponent:physical
p_{ee}	Rent exponent:efficient embedding
SW_k	small-worldness, k=bd,wd
kcs	k-core size, k=90-percentile of degree distribution
scs	s-core size, k=90-percentile of strength distribution
ϕ_k	Rich club coefficient, k=bd,wd
$cycprob_1$	fraction of 3-cycles out of 3-paths
$cycprob_2$	probability: non-cyclic 2-path extend to 3-cycle

Simulation: Random, Small-World, Scale-Free networks

Simulation example:

- coupled Mackey-Glass system, $K = 25$, $\Delta = 100$, $C = 0.2$
- Three network types: Random (RAND), Small-World (SW), Scale-Free(SCF)
- Different realizations of the same network type

Mutivariate time series record with structural changes

Simulation: Random, Small-World, Scale-Free networks

Simulation example:

- coupled Mackey-Glass system, $K = 25$, $\Delta = 100$, $C = 0.2$
- Three network types: Random (RAND), Small-World (SW), Scale-Free(SCF)
- Different realizations of the same network type

Mutivariate time series record with structural changes

Estimation of networks with PMIME at sliding windows

Simulation: Random, Small-World, Scale-Free networks

Simulation example:

- coupled Mackey-Glass system, $K = 25$, $\Delta = 100$, $C = 0.2$
- Three network types: Random (RAND), Small-World (SW), Scale-Free(SCF)
- Different realizations of the same network type

Mutivariate time series record with structural changes

Estimation of networks with PMIME at sliding windows

Estimation of network characteristics on the PMIME networks

Simulation example:

- coupled Mackey-Glass system, $K = 25$, $\Delta = 100$, $C = 0.2$
- Three network types: Random (RAND), Small-World (SW), Scale-Free(SCF)
- Different realizations of the same network type

Mutivariate time series record with structural changes

Estimation of networks with PMIME at sliding windows

Estimation of network characteristics on the PMIME networks

Structural change detection [Slow]
[Middle]
[Fast]
[Very fast]

Analysis of EEG

Practical problems to overcome:

- Application on small time windows ⇒ limited data size
- scalp EEG ⇒ many channels ⇒ many variables in Z to account for
- Brain system is complex: the connectivity measure has to deal

  with { high dimensionality
  nonlinearity?
  sensitivity on free parameters?

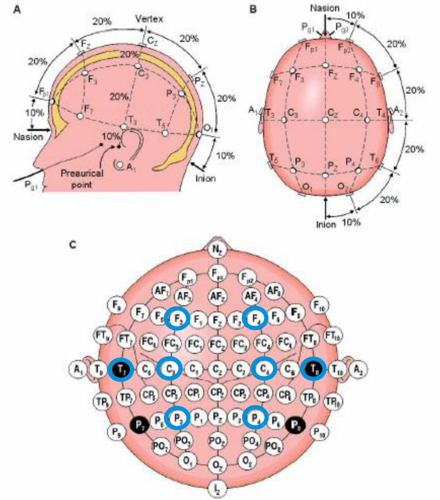
Analysis of epileptic EEG

Scalp EEG from Rikshospital, Norway.
Use 8 channels:
C3, C4, T7, T8, F3, F4, P3, P4

Subtract the average value of the four neighboring channels.

Non-overlapping segments of 20 sec.

TE for $m_x = m_y = 5$
MIME for $L_x = L_y = 15$

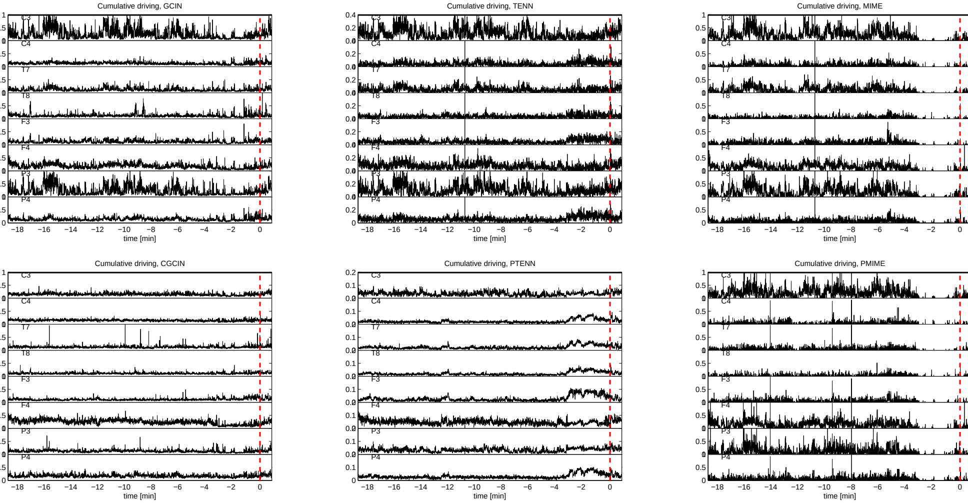


Recording: 19h 45min of scalp multi-channel EEG.

No 1 **No 2** **No 3** **No 4** **No 5** **No 6** **No 7**

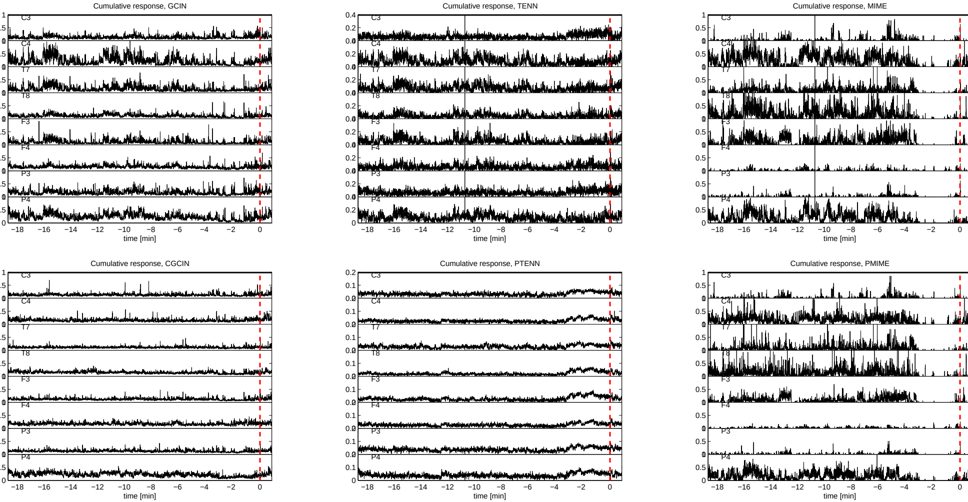
EEG, cumulative driving

Cumulative driving for channel i (e.g., $\sum_{j \neq i} \mathrm{TE}_{i \to j}$)



EEG, cumulative response

Cumulative response for channel i: (e.g., $\sum_{j \neq i} \mathrm{TE}_{j \to i}$)

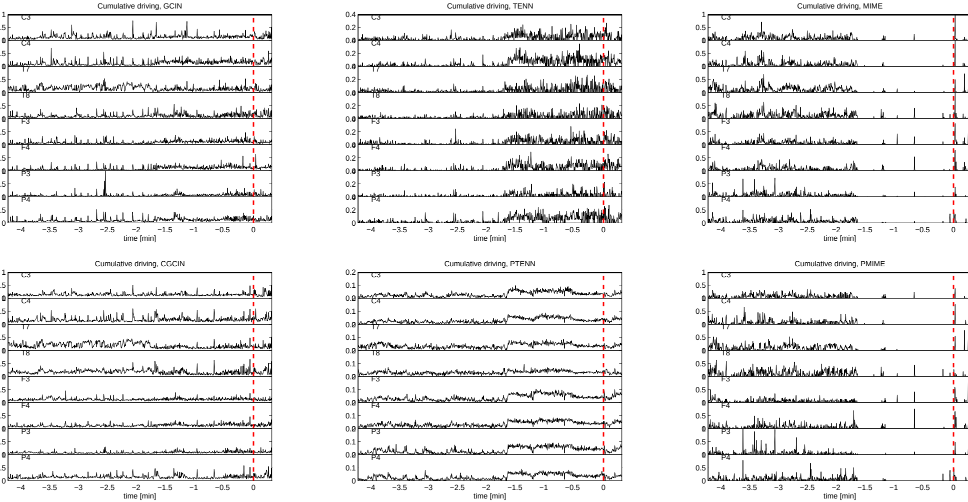


EEG, driving and response

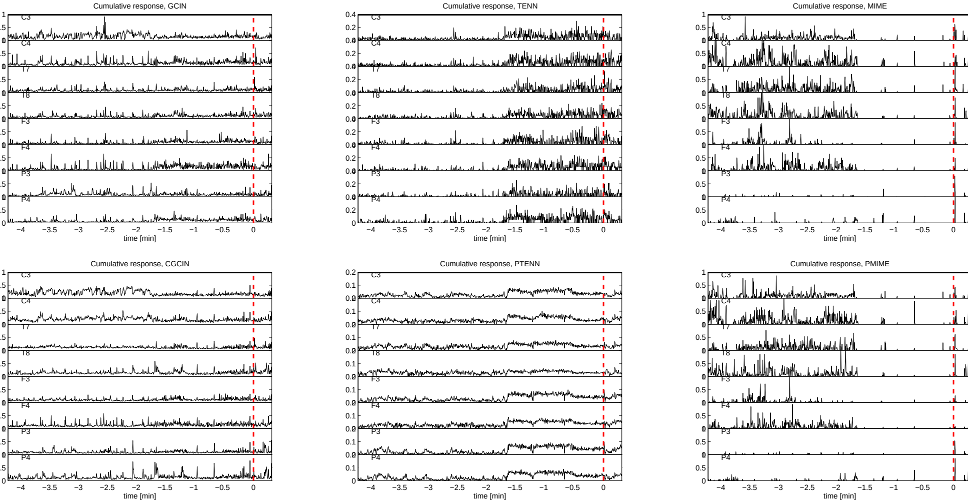
Recording: 4h 35min of scalp multi-channel EEG.

No 1 **No 2**

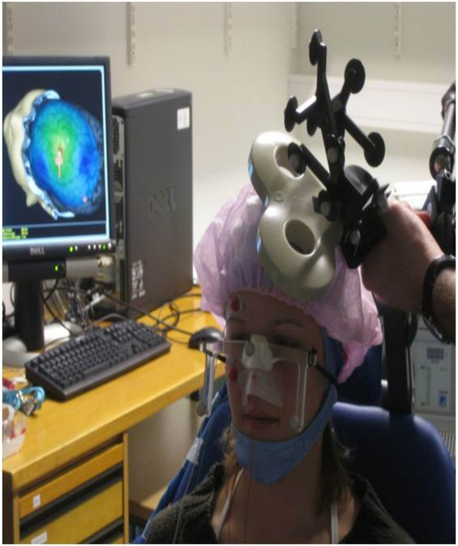
EEG, cumulative driving

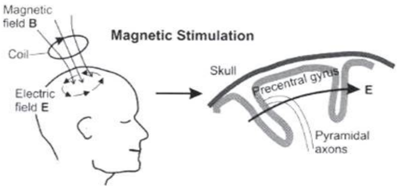


EEG, cumulative response



Transcranial Magnetic Stimulation (TMS)





EEG - TMS: brain connectivity analysis

Many issues related to processing of EEG-TMS data:

1. Rejection of corrupted EEG channels [by visual inspection, initially 60 channels]
2. Elimination of TMS artifact [*forward-backward nearest neighbor smoothing*]
3. Removal of artifacts [ICA]
4. Filtering [FIR, lowpass 0.3Hz, highpass 40Hz, order 60]
5. Re-referencing [from mastoid to infinite reference, REST]
6. Sampling frequency [downsampling from 1450 Hz to 200 Hz]

TMS was administered in blocks of 5 at frequency 3Hz after epileptic discharges were visually detected.

Computation of PMIME was done on sliding windows (length: 2 sec, sliding step: 1 sec).

Connectivity is reduced during ED and is regained by the end of ED

ED no TMS 1

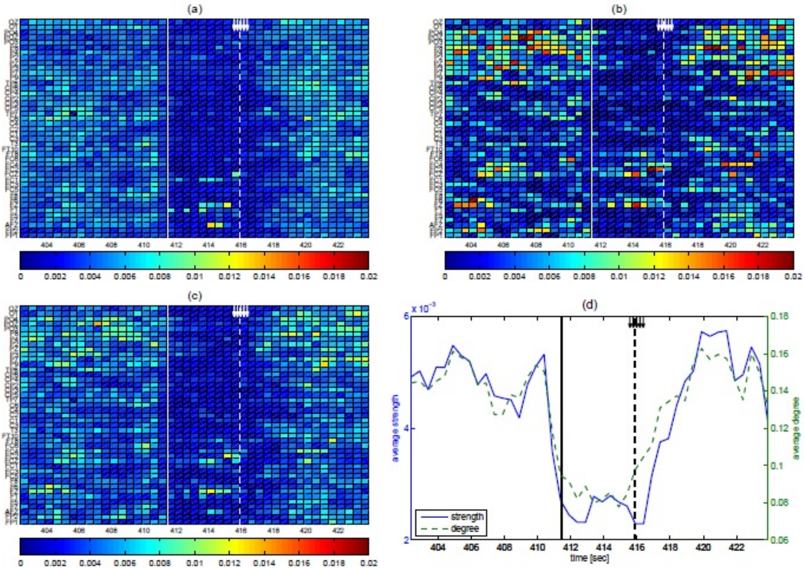
ED no TMS 2

TMS terminates ED and regains connectivity

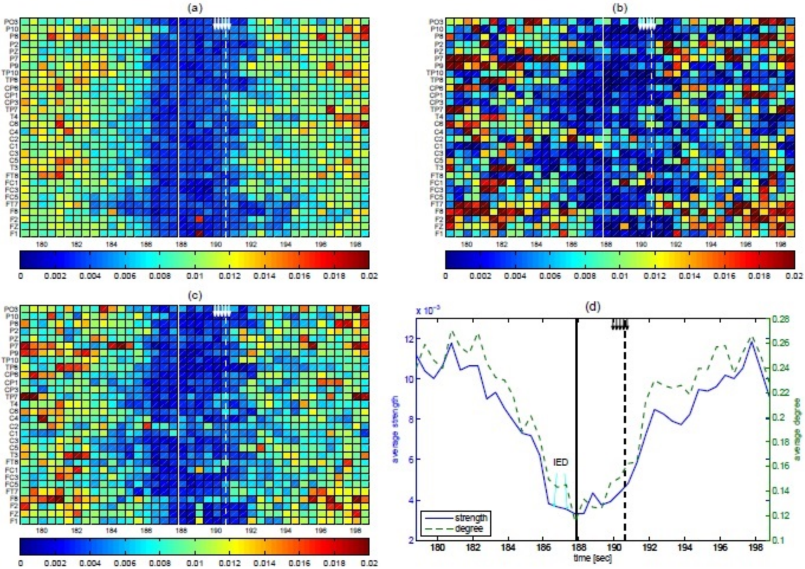
ED with TMS 1

ED with TMS 2

Network measure: Average Strength



Network measure: Average Strength



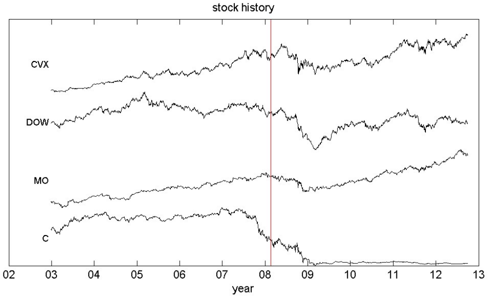
Strength of connection is reduced at ED and regained by TMS

Financial markets

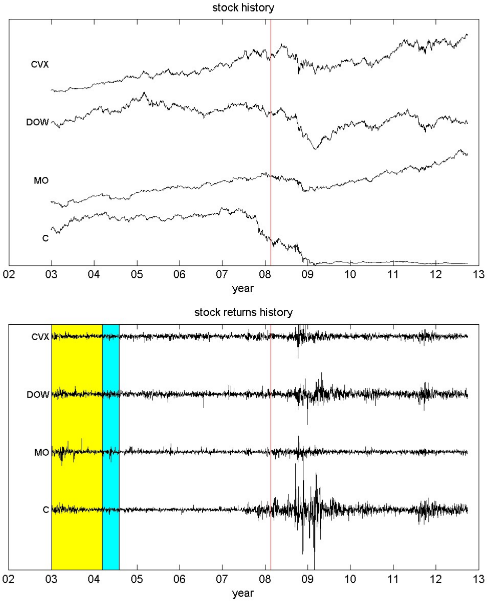
5 top stocks for each of the 8 sectors of the US economy
Daily closing index in the period 30/12/2002 - 28/9/2012

Sector	Symbol	Name	Sector	Symbol	Name
basic materials	XOM	Exxon Mobil Corp	Healthcare	JNJ	Johnson & Johnson
	CVX	Chevron corporation		PFE	Pfizer Inc.
	SLB	Schlumberger N.V.		MRK	Merck & Company
	COP	ConocoPhillips Co.		BMY	Bristol-Myers
	OXY	Occidental Petrol		AMGN	Amgen Inc.
Conglomerates	UTX	United Technological	Industrials	GE	General Electric
	MMM	3M		HON	Honeywell
	CAT	Caterpillar Inc.		CAT	Caterpillar Inc.
	DOW	Dow Chemical Company		EMR	Emerson
	MGT	MGT Capital Investments		LMT	Lockheed Martin
Consumer	PG	Proctor & Gamble	Services	WMT	Wal-Mart Stores
	KO	Coca Cola		AMZN	Amazon.com
	PM	Phillip Morris Int		DIS	Walt Disney Company
	PEP	Pepsico Inc.		HD	Home Depot
	MO	Altria Group Inc.		CMCSA	Comcast Corporation
Financials	BRK-B	Berkshire Hathaway	Technology	AAPL	Apple Inc.
	WFC	Wells Fargo		GOOG	Google
	JPM	JP Morgan Chase		MSFT	Microsoft
	C	Citigroup		IBM	IBM
	BAC	Bank of America		T	AT&T

Excluded:
CAT (Caterpillar Inc.): doubled (Conglomerates and Industrials)
PM (Phillip Morris Int): index starts 31/3/2008
GOOG (Google): index starts 19/8/2004



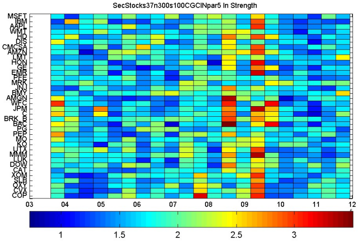
Structural change in 2008,
e.g. change point on 22/2/2008
[Dehling et al, 2013]



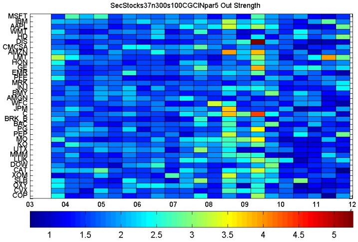
Structural change in 2008,
e.g. change point on 22/2/2008
[Dehling et al, 2013]

causality measure computed on log-returns at
windows of 300 days,
sliding step 100 days

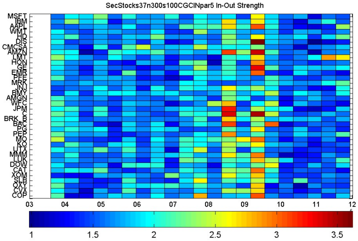
In-Strength



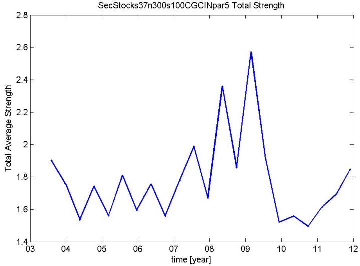
Out-Strength



In-Out-Strength

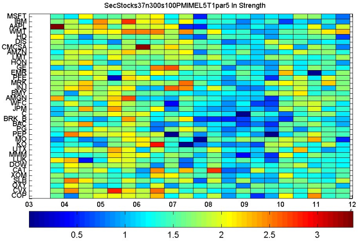


Average strength

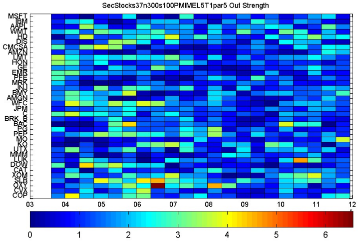


Network from PMIME

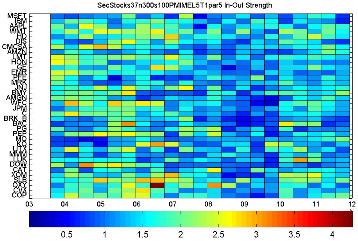
In-Strength



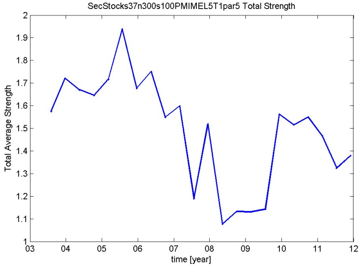
Out-Strength



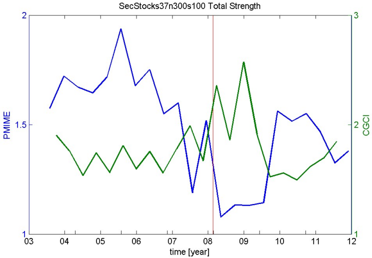
In-Out-Strength



Average strength

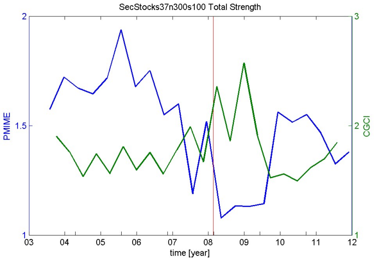


Average strength of PMIME / CGCI

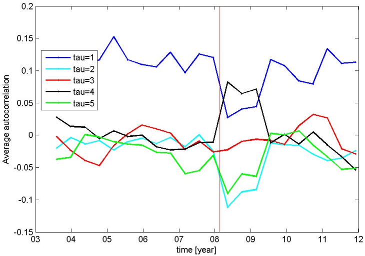


Network from PMIME

Average strength of PMIME / CGCI

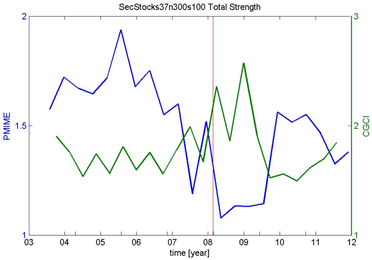


Average strength of autocorrelation

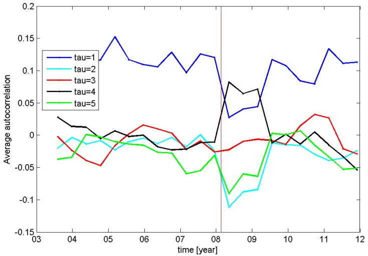


Network from PMIME

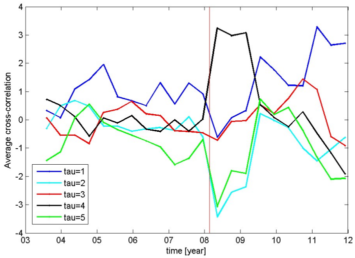
Average strength of PMIME / CGCI



Average strength of autocorrelation

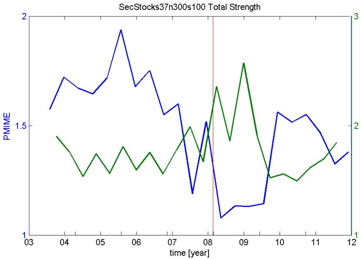


Average strength of cross-correlation

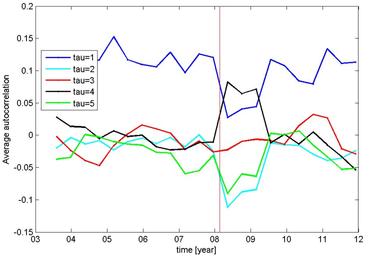


Network from PMIME

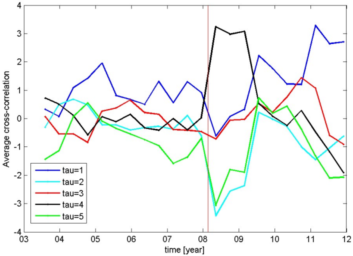
Average strength of PMIME / CGCI



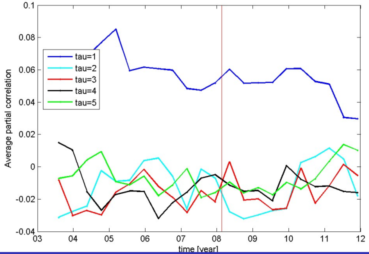
Average strength of autocorrelation



Average strength of cross-correlation



Average strength of partial correlation



Summary

Summary

- Granger causality measures can capture the inter-dependence structure of a multivariate complex system / stochastic process.

Summary

• Granger causality measures can capture the inter-dependence structure of a multivariate complex system / stochastic process.

• Granger causality measures are good candidates to detect structural changes.

Summary

• Granger causality measures can capture the inter-dependence structure of a multivariate complex system / stochastic process.

• Granger causality measures are good candidates to detect structural changes.

• Many measures of causality:

Summary

• Granger causality measures can capture the inter-dependence structure of a multivariate complex system / stochastic process.

• Granger causality measures are good candidates to detect structural changes.

• Many measures of causality:

best: these that can capture also nonlinear and direct causal effects at the presence of many variables...

Summary

• Granger causality measures can capture the inter-dependence structure of a multivariate complex system / stochastic process.

• Granger causality measures are good candidates to detect structural changes.

• Many measures of causality:

best: these that can capture also nonlinear and direct causal effects at the presence of many variables... but practically hard to estimate reliably.

Summary

• Granger causality measures can capture the inter-dependence structure of a multivariate complex system / stochastic process.

• Granger causality measures are good candidates to detect structural changes.

• Many measures of causality:

best: these that can capture also nonlinear and direct causal effects at the presence of many variables... but practically hard to estimate reliably.

1. More advanced measures (nonlinear, direct effects) involve more (and depend more on) **free parameters**.
2. Harder to establish **statistical significance** of the measures when many variables are present (many nodes in the network). Correction for multiple testing requires many many surrogates.
3. Statistical accuracy of the direct causality measures decreases with the **number of confounding variables**.

References

Brandt PT & Williams JT (2007) *Multiple Time Series Models*, Sage Publications

Dehling H, Vogel D, Wendler M, Wied D (2013) "An efficient and robust test for a change-point in correlation, *arXiv:1203.4871*

Kimiskidis V K, Kugiumtzis D, Papagiannopoulos S, Vlaikidis N (2013) "Transcranial Magnetic Stimulation (TMS) Modulates Epileptiform Discharges in Patients with Frontal Lobe Epilepsy: a Preliminary EEGTMS Study", *International Journal of Neural Systems*, 23, 1250035

Kraskov A, Stögbauer H & Grassberger P (2004) "Estimating Mutual Information", *Physical Review E*, 69(6): 066138

Kugiumtzis D (2013) "Direct Coupling Information Measure from Non-uniform Embedding", *Physical Review E*, 87: 062918

Papana A, Kugiumtzis D & Larsson PG (2011) "Reducing the bias of causality measures", *Physical Review E*, 83: 036207

Papana A, Kugiumtzis D & Larsson PG (2012) "Detection of direct causal effects and application in the analysis of electroencephalograms from patients with epilepsy", *International Journal of Bifurcation and Chaos*, to be published

Schreiber T (2000) "Measuring Information Transfer", *Physical Review Letters*, 85(2): 461–464

Vlachos I & Kugiumtzis D (2010) "Non-uniform state space reconstruction and coupling detection", *Physical Review E*, 82: 016207