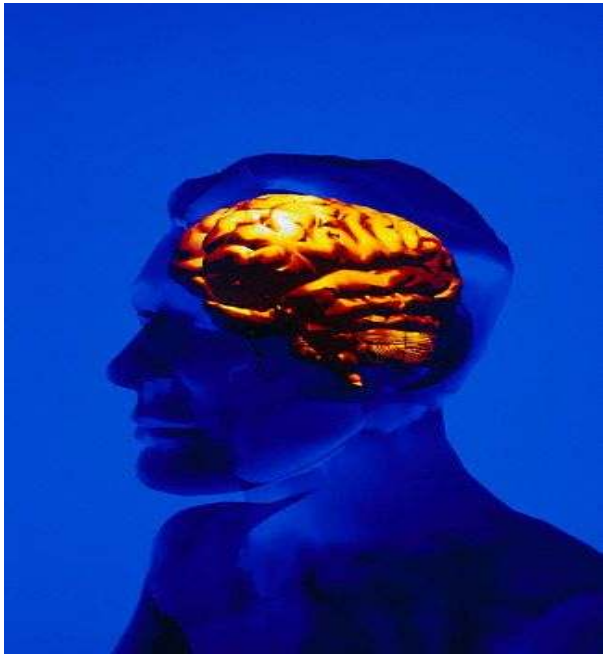
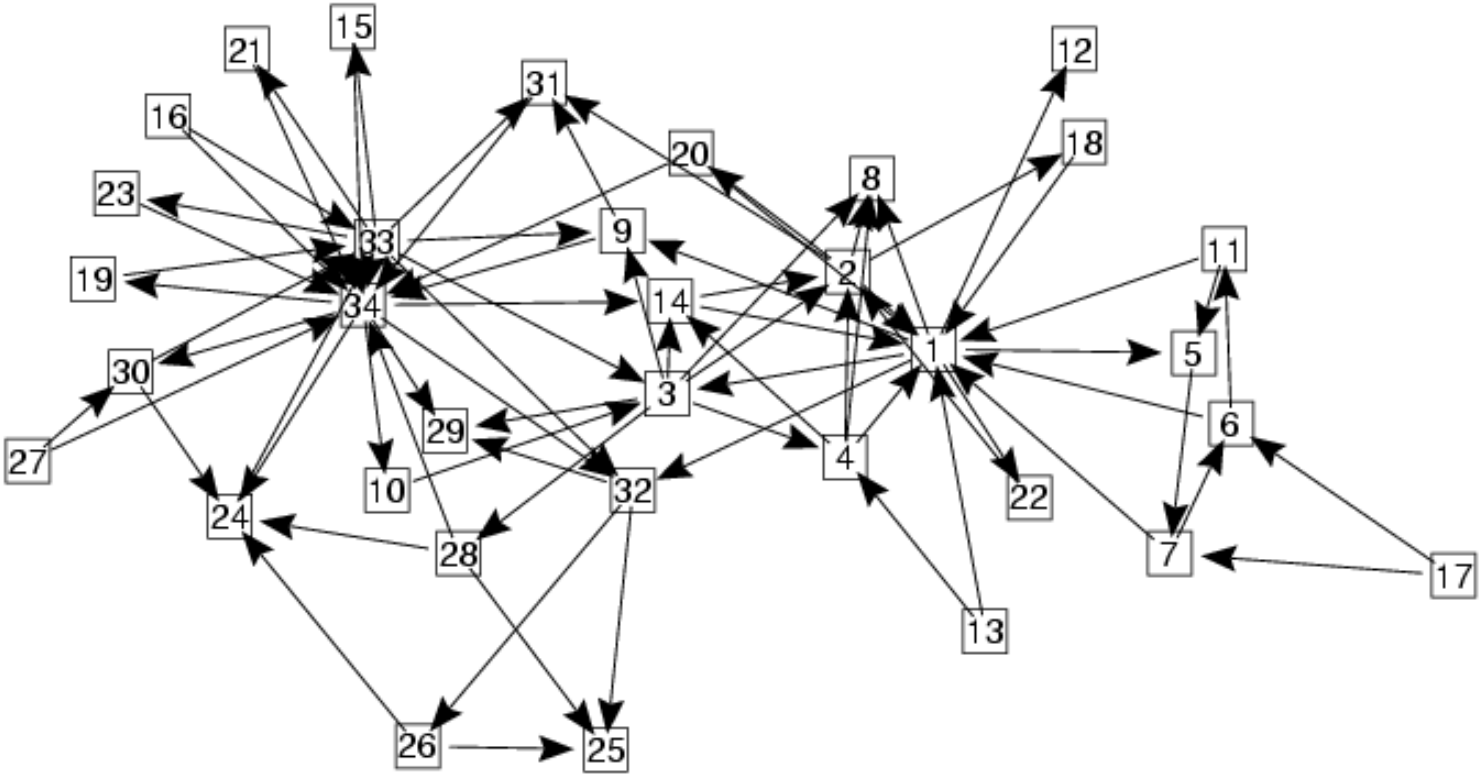
# Inferenza di networks

Sebino Stramaglia



# Se non c'è accesso alle connessioni fisiche tra i nodi?

Inferenza del network dai dati !



# Dati temporalmente correlati: Un sistema dinamico in ogni nodo

- Applicazioni più importanti
- Neuroscienze (fMRI, EEG, MEG, ECoG)
- System biology (gene regulatory networks)

# Paradigma generale (Kolaczyk 2009)

1) Scelta della misura dell'accoppiamento (bivariata o multivariata)
2) Test di significatività (p-value)
3) Controllo del false discovery rate

SPARSIFICAZIONE TRAMITE TEST STATISTICI

# Misure dell’accoppiamento:

Correlazioni

Mutua informazione

Sincronizzazione di fase

Sincronizzazione generalizzata

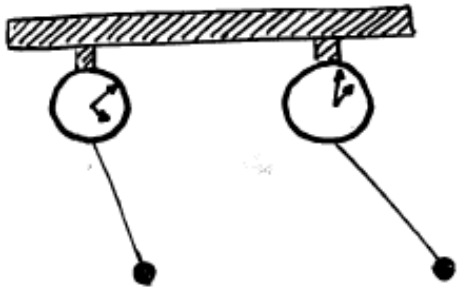
Coerenza spettrale

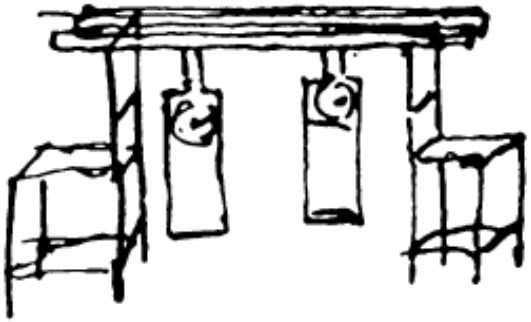
Entropia di trasferimento

Causalità alla Granger

La sincronizzazione è l’adattamento del ritmo di un sistema di oscillatori non-lineari dovuta alla loro debole interazione.

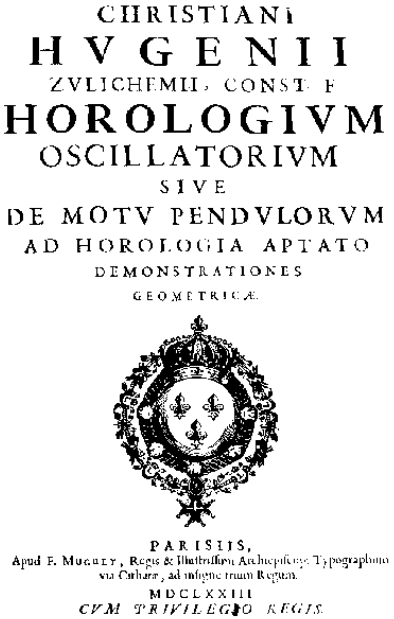
Esempio: gli orologi a pendolo (Huygens, 1665)





I due orologi (=oscillatori), separati, hanno velocità di avanzamento leggermente diverse.

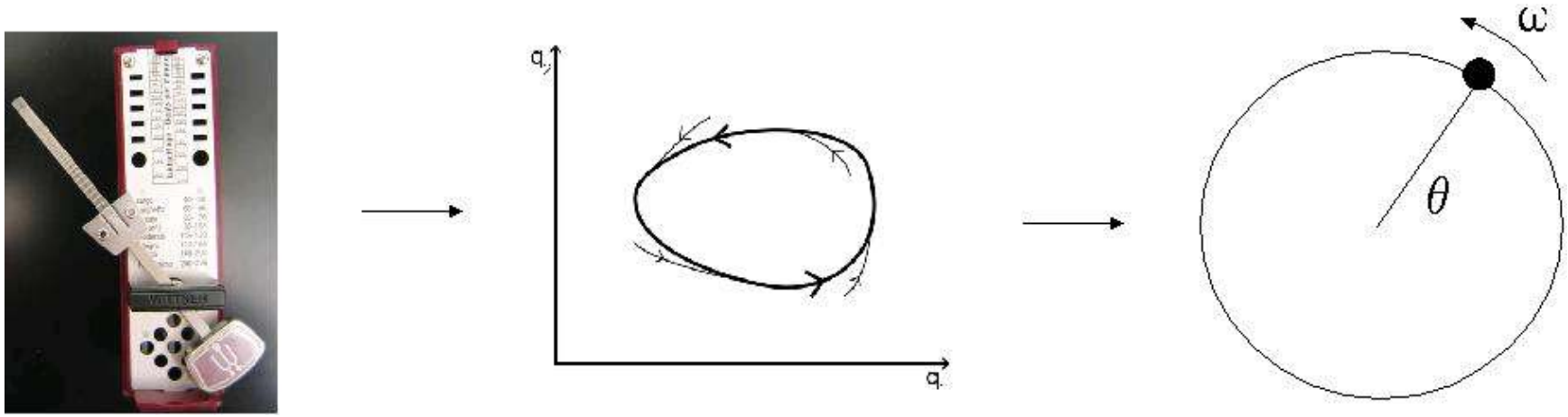
Connessi allo stesso supporto (=debole interazione), gli orologi avanzano a velocità perfettamente identica (=adattamento del ritmo).



CHRISTIANI
HVGENII
ZVLICHEMII, CONST. F
HOROLOGIVM
OSCILLATORIVM
SIVE
DE MOTV PENDVLORVM
AD HOROLOGIA APTATO
DEMONSTRATIONES
GEOMETRICÆ.

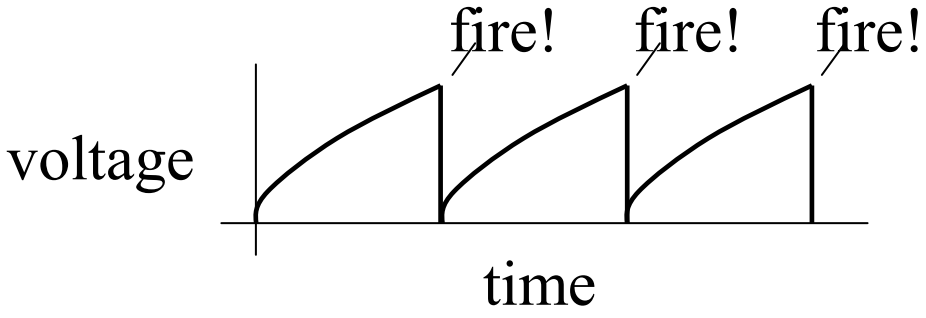
PARISIIS,
Apud F. Muguet, Regis & Illustrissimi Archiepiscopi Typographum
via Cithariæ, ad insigne trium Regum.
MDCLXXIII
CVM PRIVILEGIO REGIS.

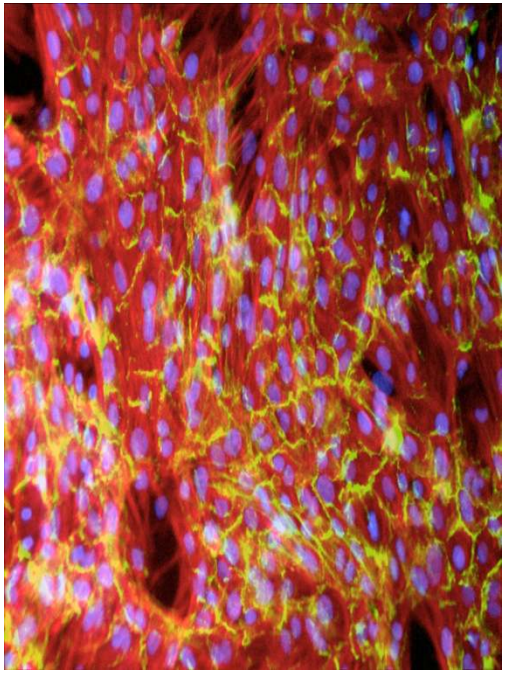
La dinamica di un sistema, anche non perfettamente periodico, può essere vista come la manifestazione di un punto che si muove su un ciclo limite.



# Battiti cardiaci

- Qualche migliaia di cellule sino-atriali
- Ognuna oscilla ad una sua frequenza caratteristica
- Peskin's model: quando una cellula emette un impulso, tutti le altre cellule hanno un piccolo salto nel voltaggio
- Questa debole interazione fa sincronizzare le cellule, che "sparano" sincrone dando origine al battito cardiaco



# Sincronizzazione nel cervello

- I neuroni attivi nella banda alfa presi individualmente lancerebbero segnali elettrici con frequenze tutte diverse tra loro.
- Presi assieme formano un vero e proprio orologio che batte con la frequenza di circa 10 pulsazioni al secondo

# Flash sincronizzati delle lucciole

- Scoperte in Malesia nell’800: colonie di migliaia di lucciole lampeggianti
- Flash usati nel corteggiamento rituale in gruppo
- Dopo un breve periodo il lampeggiamento diviene sincrono

# Spiegazioni

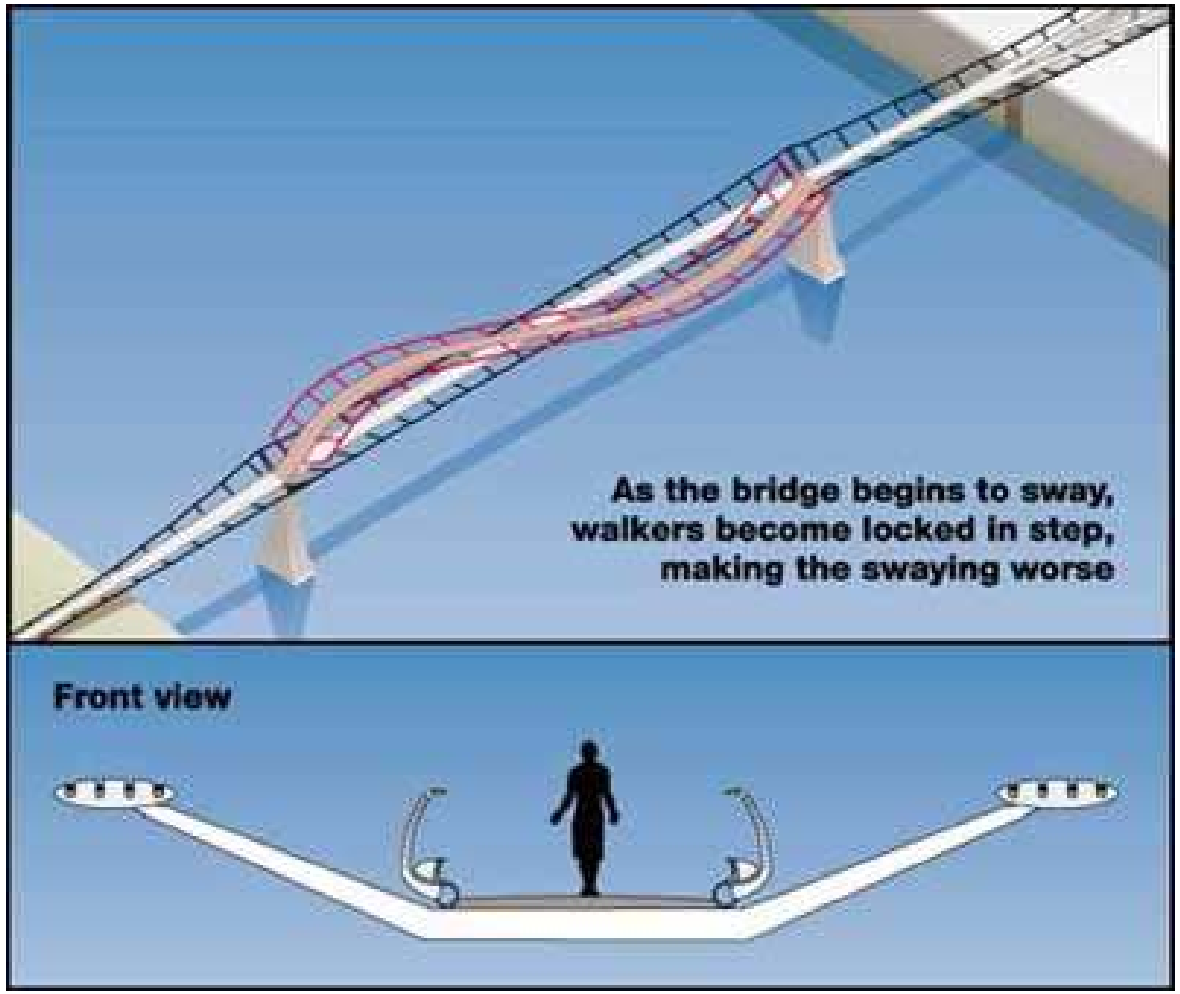
- Accidentale ?
- Effetto ottico (sbattere delle ciglia dell’osservatore) ?
- Colpi di vento?
- In realta’:
- Le lucciole hanno meccanismi neurali che vengono stimolati dai flash

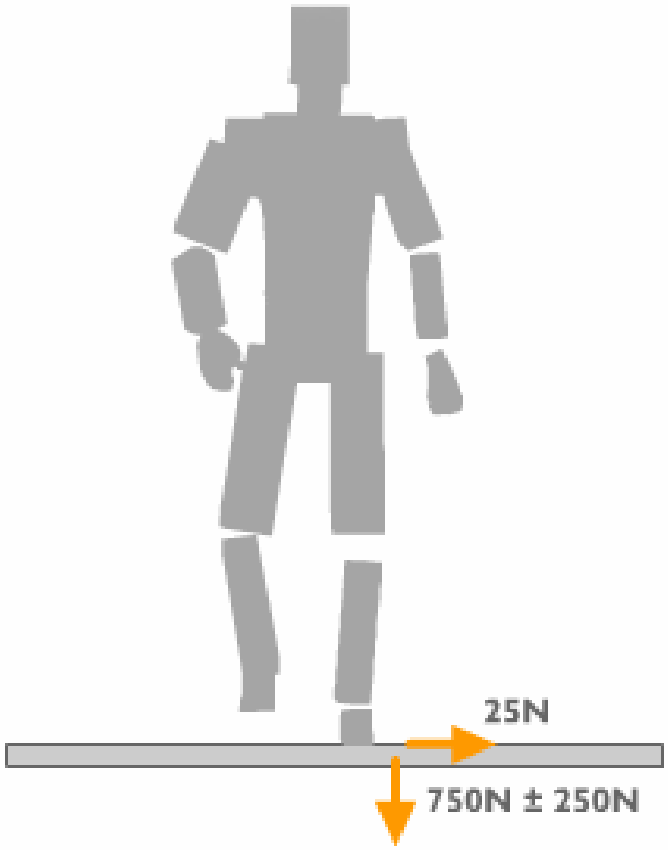
# Comportamenti simili negli esseri umani

- Chiudete e aprite gli occhi e battete un dito sul tavolo
- Quasi subito i due movimenti diventano sincroni
- Applausi
- Compagne di collegio

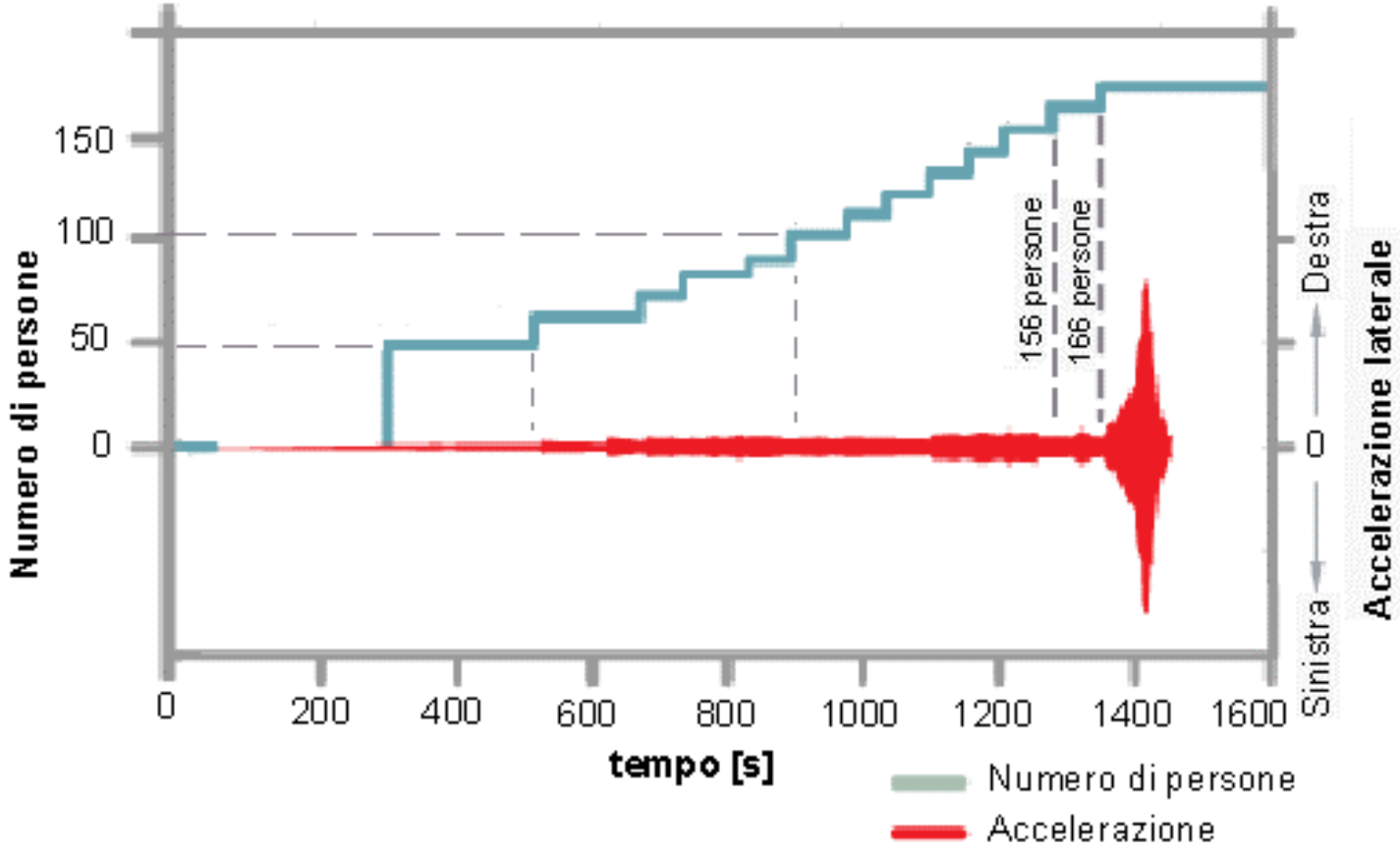
# Dark side of the moon

# Millennium bridge


As the bridge begins to sway,
walkers become locked in step,
making the swaying worse
Front view


25N
750N ± 250N

# Transizione di fase



# Onde α nel cervello

**EEG rithms:**

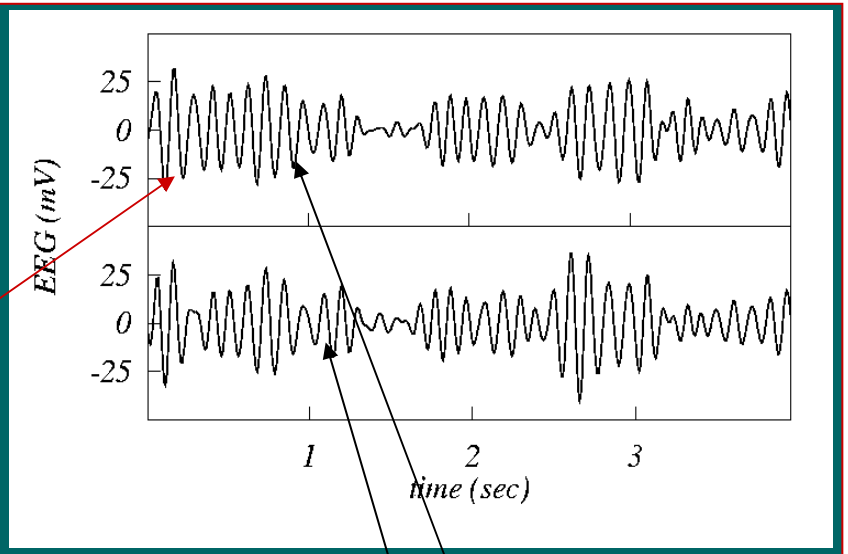
**δ(0.5-3.5 Hz):deep sleep**

**θ(3.5-7.5 Hz):sleep**

**α(7.5-12.5 Hz): wake,relax**

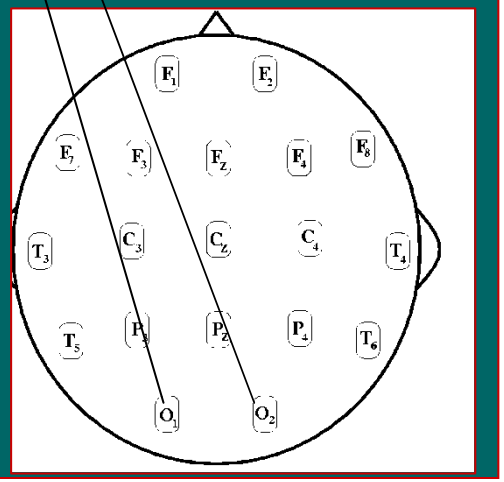
**β (12.5-30 Hz): tension**

**γ (30-60 Hz): Cellular level exp.**

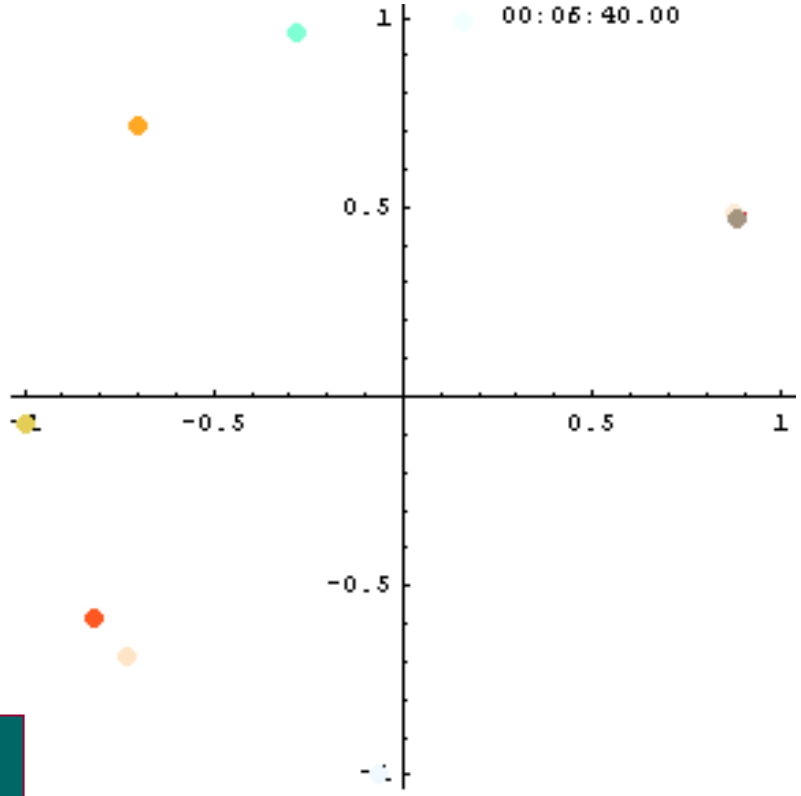



Periodicity is produced by spontaneous synchronization of microscopic signals

Synchronization between signals from different electrodes can be noted,

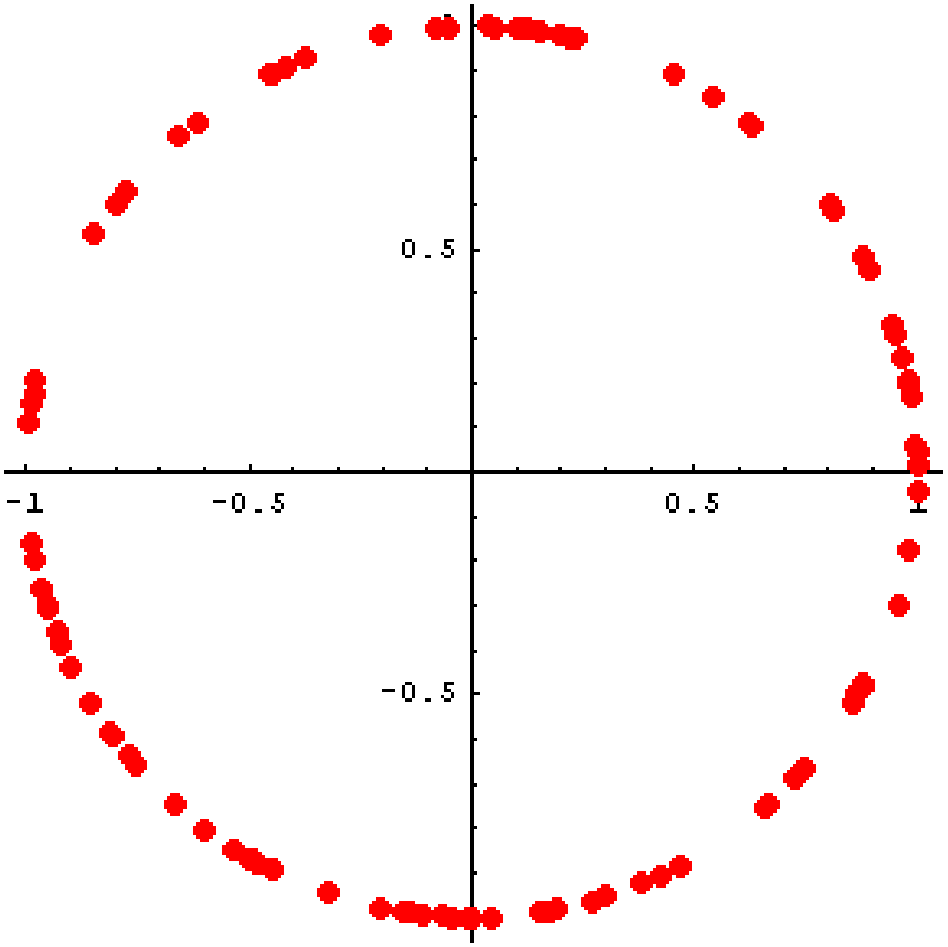


# Modello matematico: Eq. di oscillatori (simile a pendoli)

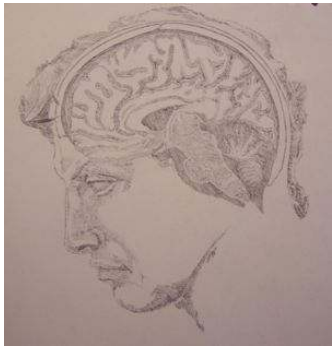



Oscillatori: punti che
si muovono su di una circonferenza
con velocita' diverse:

# (N grande, interazione non lineare)



# BINDING PROBLEM



- Il cervello mostra una specializzazione locale
- Compiti complessi richiedono cooperazione tra diverse aree del cervello
- La sincronizzazione è un meccanismo chiave per l'integrazione funzionale
- La sincronizzazione porta alla formazione di reti funzionali con struttura spaziale e temporale

# Binding problem

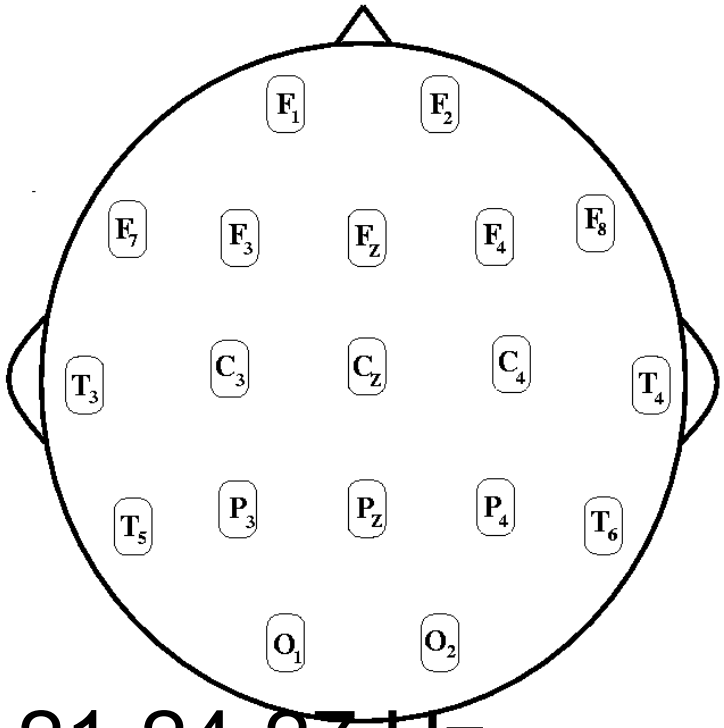
- Come comunicano tra loro diverse zone del cervello in presenza (o in assenza) di determinati stimoli?

# Emicrania

- Patologia molto complicata e diffusa
- Disordine del sistema nervoso centrale.
- Stress, sforzi fisici, cambiamenti di tempo, forti emozioni, luci intermittenti, …..
- Nessun modello valido sperimentale
- In cosa differisce la risposta a questi stimoli tra emicranici e persone sane?

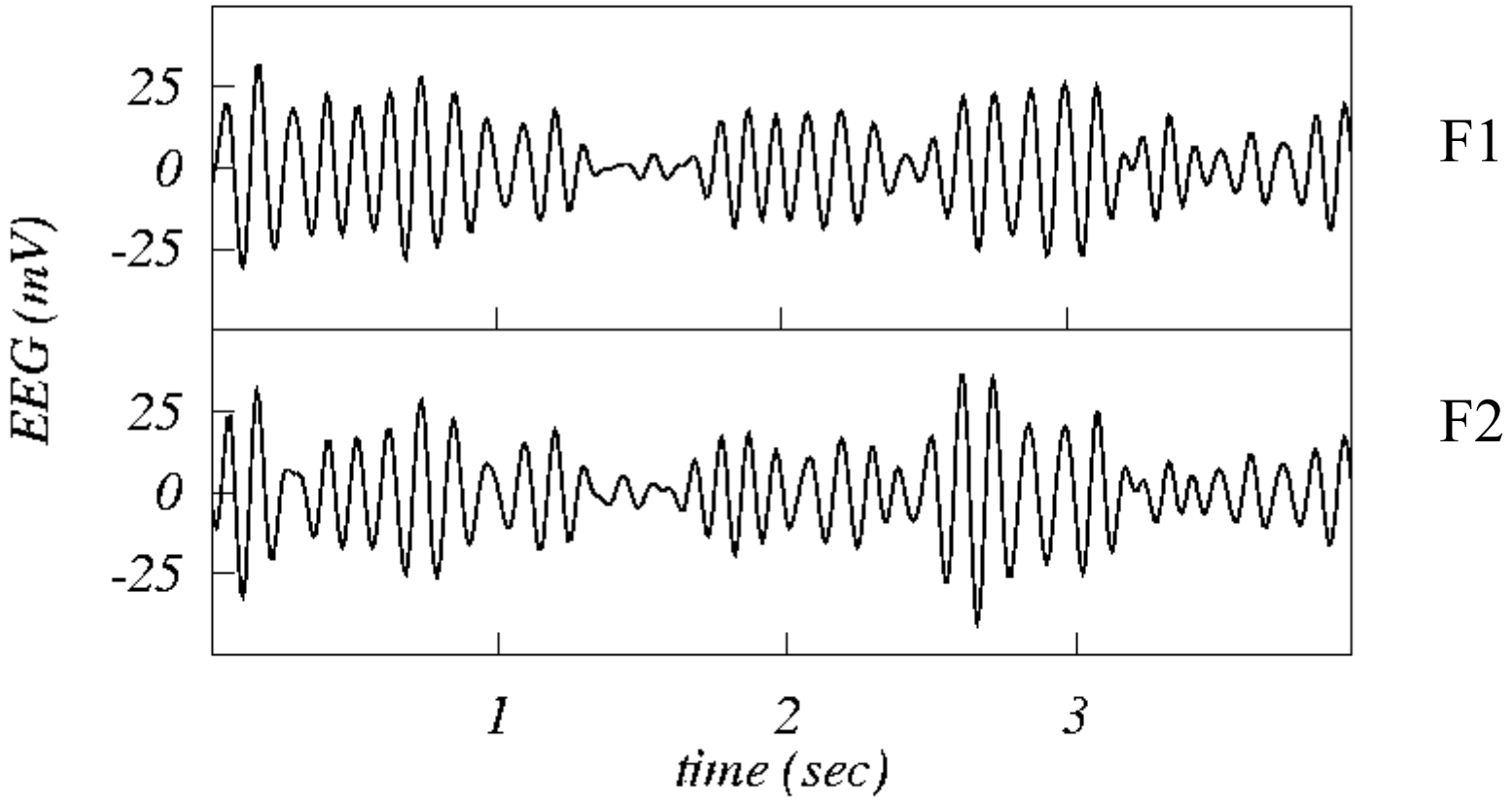
# Ritmi EEG

- Theta (4-7 Hz)
- Alpha(8-12.5 Hz)
- Beta (14-30 Hz)
- Gamma (>30 Hz)




- EEG spontaneo
- Flash: 3-6-9-12-15-18-21-24-27 Hz

•Sincronizzazione tra segnali provenienti da diverse regioni della corteccia
•Ruolo nella comunicazione tra le diverse regioni?



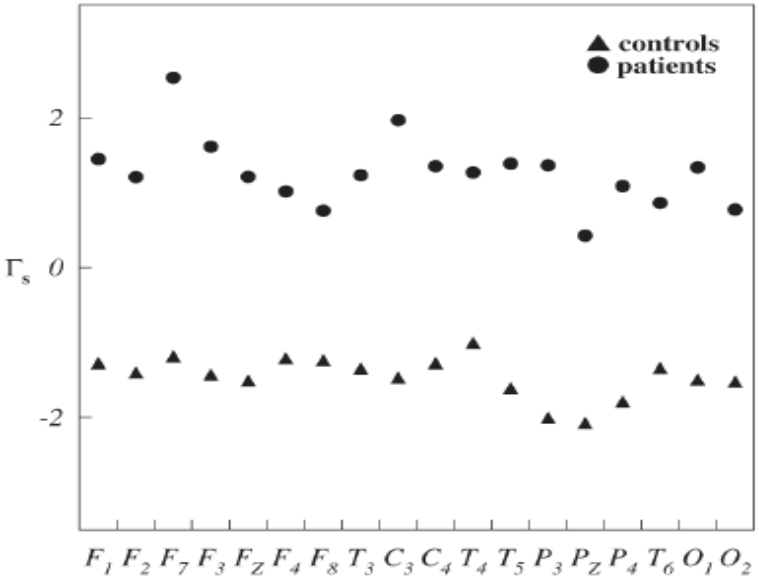
# Steady-State Visual Evoked Potentials and Phase Synchronization in Migraine Patients

L. Angelini,[1,2,3] M. De Tommaso,[1,4] M. Guido,[1] K. Hu,[5] P. Ch. Ivanov,[5,6] D. Marinazzo,[1] G. Nardulli,[1,2,3] L. Nitti,[1,7,3] M. Pellicoro,[1,2,3] C. Pierro,[1] and S. Stramaglia[1,2,3]

[1]*TIRES: Center of Innovative Technologies for Signal Detection and Processing, University of Bari, Italy*
[2]*Physics Department, University of Bari, Italy*
[3]*Istituto Nazionale di Fisica Nucleare, Sezione di Bari, Italy*
[4]*Department of Neurological and Psychiatric Sciences, University of Bari, Italy*
[5]*Center for Polymer Studies and Department of Physics, Boston University, Boston, Massachusetts, USA*
[6]*Institute of Solid State Physics, Bulgarian Academy of Sciences, 1784 Sofia, Bulgaria*
[7]*D.E.T.O., University of Bari, Italy*

(Received 8 October 2003; revised manuscript received 5 April 2004; published 15 July 2004)

We investigate phase synchronization in EEG recordings from migraine patients. We use the analytic signal technique, based on the Hilbert transform, and find that migraine brains are characterized by enhanced alpha band phase synchronization in the presence of visual stimuli. Our findings show that migraine patients have an overactive regulatory mechanism that renders them more sensitive to external stimuli.



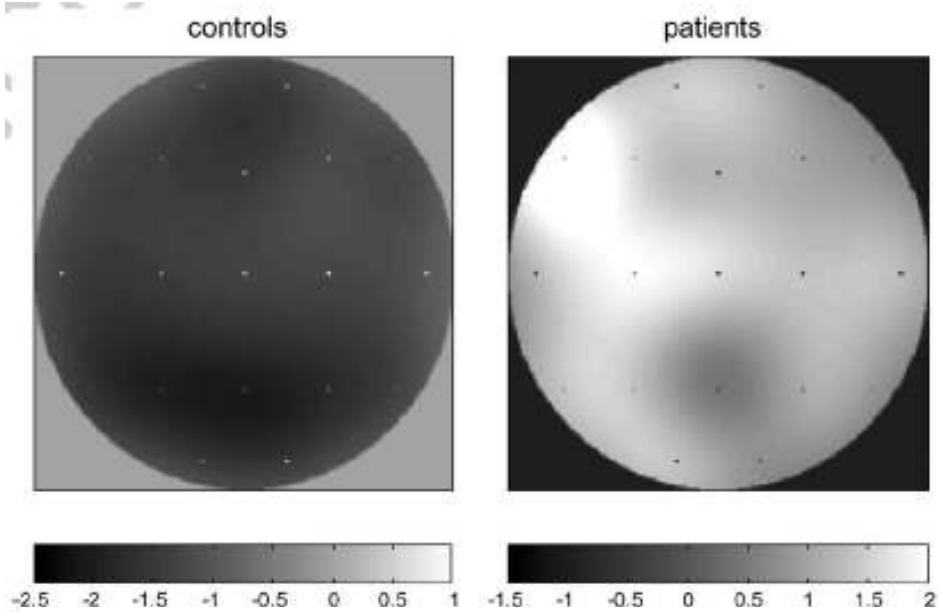



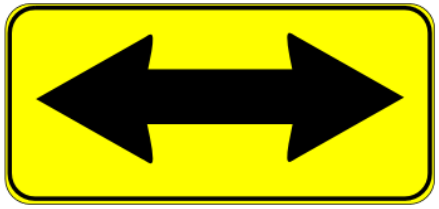

FIG. 3. In the case of 24 Hz stimuli, the mean of $\Gamma_s$ (over patients and over controls) is represented for all the 18 electrodes. On the average, phase synchronization increases for patients and decreases for controls.

# Relazioni di causa-effetto tra i sottosistemi

In molte situazioni siamo interessati non solo a rivelare stati sincronizzati, ma anche a mettere in evidenza relazioni di causa-effetto (drive-response) tra i sottosistemi di un sistema complesso.

# Esempio: sync vs causality

Sync:

RR  SAP

Causality:

RR SAP

# Esempi di applicazione

- Neuroscienze
- Epilessia
- Epidemiologia sperimentale
- Ecologia e climatologia
- Macroeconomia
- Sistemi caotici spazio-temporali
- Reo-caos
- Genomica
- ……..

Necessità di dare rigore statistico alla nozione di causa !

# Adamo ed Eva

## Esempio: Osserviamo una statua della regina Vittoria

- Perché la statua é qui?
- Causa materiale: la pietra con cui la statua é costruita
- Causa formale: la regina Vittoria
- Causa efficiente: lo scultore che ha fatto la statua
- Causa finale: volontá del popolo di esprimere gratitudine alla regina Vittoria

# Causalitá in fisica

In fisica l'applicazione della causalitá materiale e di quella formale non crea problemi di natura filosofica.

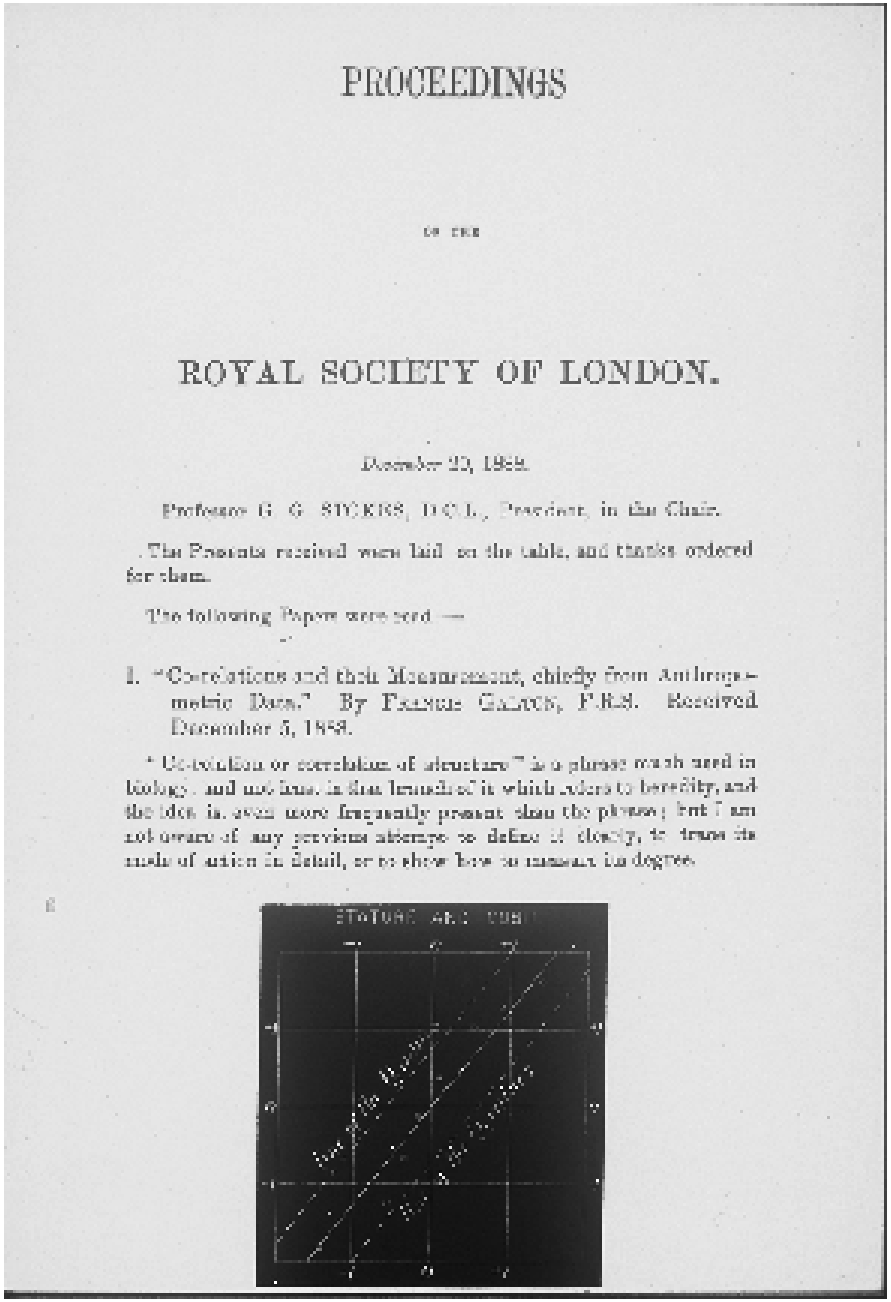
L'applicazione della causalitá efficiente, in fisica, ha un importante corollario: la causa deve precedere l'effetto (in modo invariante, cioé in tutti i sistemi di riferimento inerziali).

Esperimenti come EPR sulla non-localitá solo apparentemente mettono in crisi la causalitá efficiente,perché é stato dimostrato che effetti tipo EPR non possono essere utilizzati per mandare segnali con velocitá maggiore di c.

Causalitá finale in Fisica? Ha senso porsi questa domanda? Esempio: principio antropico.

# GALTON'S PLOT DI DATI CORRELATI

(1888)



PROCEEDINGS

OF THE

ROYAL SOCIETY OF LONDON.

December 20, 1888.

Professor G. G. STOKES, D.C.L., President, in the Chair.

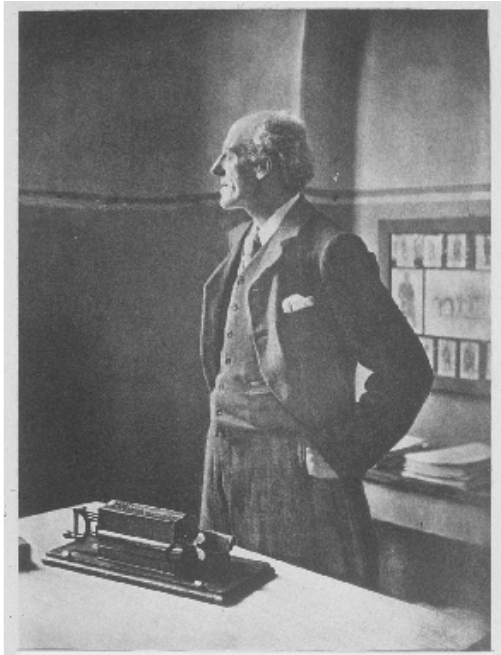
The Presents received were laid on the table, and thanks ordered for them.

The following Papers were read:—

I. "Correlations and their Measurement, chiefly from Anthropometric Data." By FRANCIS GALTON, F.R.S. Received December 5, 1888.

"Co-relation or correlation of structure" is a phrase much used in biology, and not less in that branch of it which refers to heredity, and the idea is even more frequently present than the phrase; but I am not aware of any previous attempt to define it closely, to trace its mode of action in detail, or to show how to measure its degree.

**Reazione di KARL PEARSON alla lettura del lavoro di Dalton(1934)**



KARL PEARSON, 1934

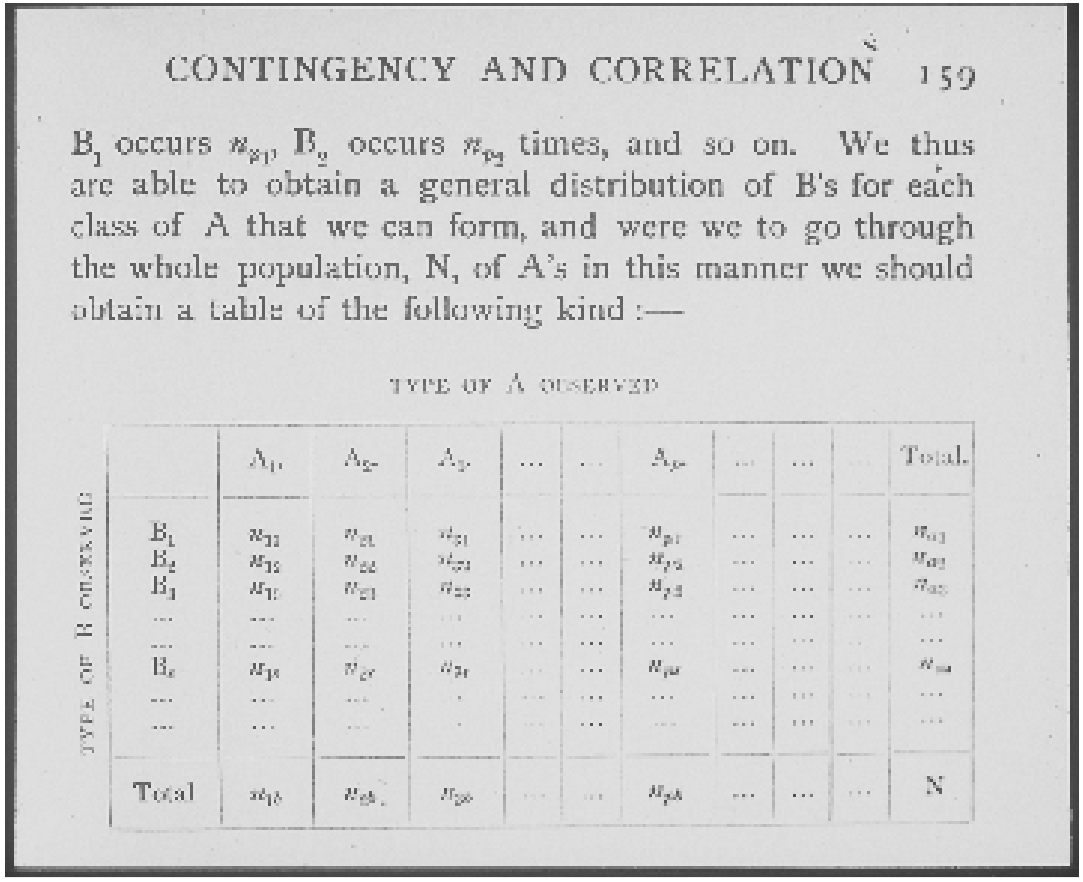
"I felt like a buccaneer of Drake's days -... I interpreted that sentence of Galton to mean that there was a category broader than causation, namely correlation, of which causation was only the limit, and that this new conception of correlation brought psychology, anthropology, medicine, and sociology in large parts into the field of mathematical treatment."

**TAVOLA DELLA CONTINGENZA (1911)**

Il 1911 vede la pubblicazione della terza edizione del libro di Pearson "The Grammar of Science". Contiene un nuovo capitolo "Contingency and correlation - the insufficiency of causation,“

Questo è quanto Pearson afferma:

"Beyond such discarded fundamentals as 'matter' and 'force' lies still another fetish amidst the inscrutable arcana of modern science, namely, the category of cause and effect."



CONTINGENCY AND CORRELATION 159

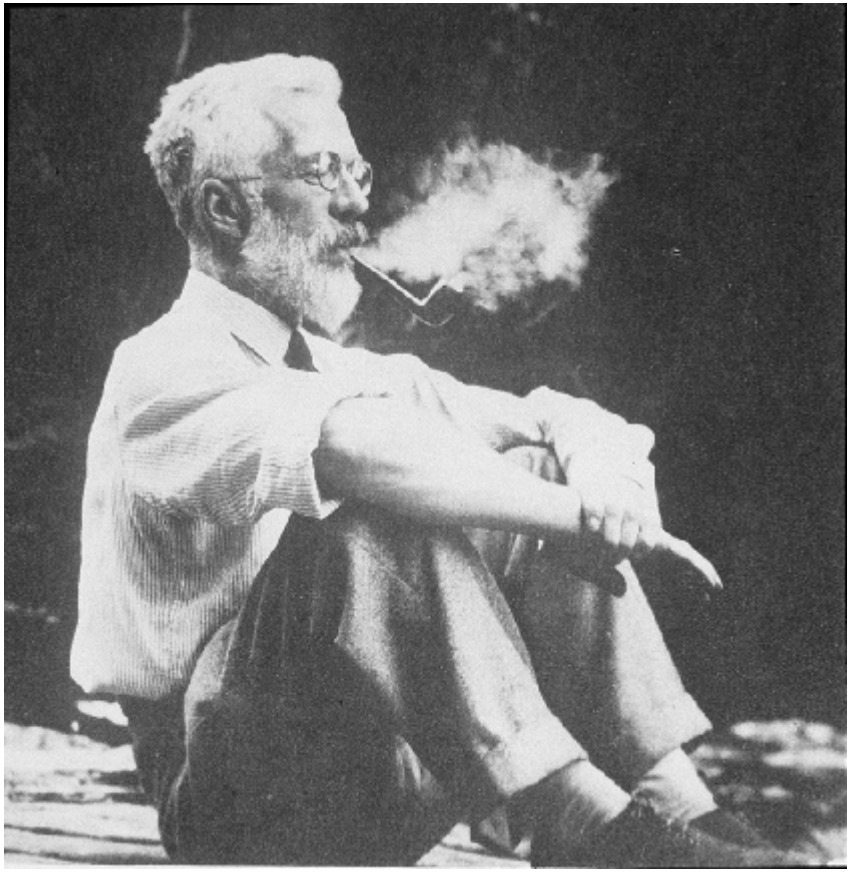
$B_1$ occurs $n_{p1}$, $B_2$ occurs $n_{p2}$ times, and so on. We thus are able to obtain a general distribution of B's for each class of A that we can form, and were we to go through the whole population, N, of A's in this manner we should obtain a table of the following kind :—

TYPE OF A OBSERVED

| TYPE OF B OBSERVED | $A_1$ | $A_2$ | $A_3$ | ... | ... | $A_p$ | ... | ... | ... | Total. |
|---|---|---|---|---|---|---|---|---|---|---|
| $B_1$ | $n_{11}$ | $n_{21}$ | $n_{31}$ | ... | ... | $n_{p1}$ | ... | ... | ... | $n_{01}$ |
| $B_2$ | $n_{12}$ | $n_{22}$ | $n_{32}$ | ... | ... | $n_{p2}$ | ... | ... | ... | $n_{02}$ |
| $B_3$ | $n_{13}$ | $n_{23}$ | $n_{33}$ | ... | ... | $n_{p3}$ | ... | ... | ... | $n_{03}$ |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| $B_s$ | $n_{1s}$ | $n_{2s}$ | $n_{3s}$ | ... | ... | $n_{ps}$ | ... | ... | ... | $n_{0s}$ |
| ... | ... | ... | . | ... | ... | ... | ... | ... | ... | ... |
| ... | ... | ... | . | ... | ... | ... | ... | ... | ... | ... |
| Total | $n_{10}$ | $n_{20}$ | $n_{30}$ | ... | ... | $n_{p0}$ | ... | ... | ... | N |

## Pearson rigetta categoricamente il bisogno del concetto indipendente di causa,oltre al concetto di correlazione.

Per i 25 anni successivi non si parlò più di relazioni causali in Statistica.
Fino a quando Sir Ronald Fisher formulò l'esperimento randomizzato, tuttora il solo metodo universalmente accettato per verificare relazioni causali dai dati.



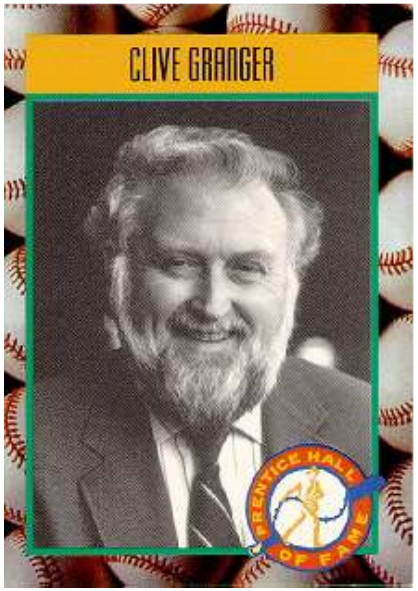
Fisher's classic example is that of the tea-drinking lady:

> "A lady declares that by tasting a cup of tea made with milk she can discriminate whether the milk or the tea infusion was first added to the cup. ... Our experiment consists in mixing eight cups of tea, four in one way and four in the other, and presenting them to the subject in random order. ... Her task is to divide the cups into two sets of 4, agreeing, if possible, with the treatments received. ... The element in the experimental procedure which contains the essential safeguard is that the two modifications of the test beverage are to be prepared "in random order." This is in fact the only point in the experimental procedure in which the laws of chance, which are to be in exclusive control of our frequency distribution, have been explicitly introduced. ... it may be said that the simple precaution of randomisation will suffice to guarantee the validity of the test of significance, by which the result of the experiment is to be judged."

L’esperimento randomizzato permetterebbe di stabilire che il fumo è causa del cancro ai polmoni e non viceversa.

- Quando non è possibile fare l’esperimento randomizzato (come nel caso del fumo) è ancora possibile inferire relazioni causali?

1969:




# Causalitá: definizione operativa di Granger

Clive Granger (premio Nobel 2003 per l'economia)

X é causa di Y se la conoscenza di X permette di fare previsioni piú accurate sui valori futuri di Y

Osserviamo che questa definizione ha senso solo per fenomeni irreversibili.

# Assenza di causalità: Generalized Markov property

$$P(x \mid X) = P(x \mid X, Y)$$

# Transfer entropy (Schreiber 2000)

Misura la violazione della generalized Markov property:

$$T(Y \rightarrow X) = \int P(x, X, Y) \log\left( \frac{P(x \mid X, Y)}{P(x \mid X)} \right) dx dX dY$$

T misura l' informazione che fluisce da una serie temporale all'altra

T è collegata ma non è equivalente all'accoppiamento

# Transfer entropy and regression

The minimizer of the risk functional

$$\mathsf{R}\left[f\right] = \int dX dx \left(x - f(X)\right)^2 p(X, x)$$

represents the best estimate of $x$, given X, and corresponds to the regression function

$$f^*(X) = \int dx p(x|X) x.$$

# Transfer entropy and regression

The best estimate of $x$, given $X$ and $Y$, is now:

$$g^*(X,Y) = \int dx p(x|X,Y)x.$$

If $p(x|X,Y) = p(x|X)$, then

$$f^*(X) = g^*(X,Y)$$

and the knowledge of $Y$ does not improve the prediction of $x$.

# Transfer entropy

$$S_X = -\int dxdXp(x,X)\log[p(x\,|\,X)]$$

$$S_{XY} = -\int dxdXdYp(x,X,Y)\log[p(x\,|\,X,Y)]$$

# Regression

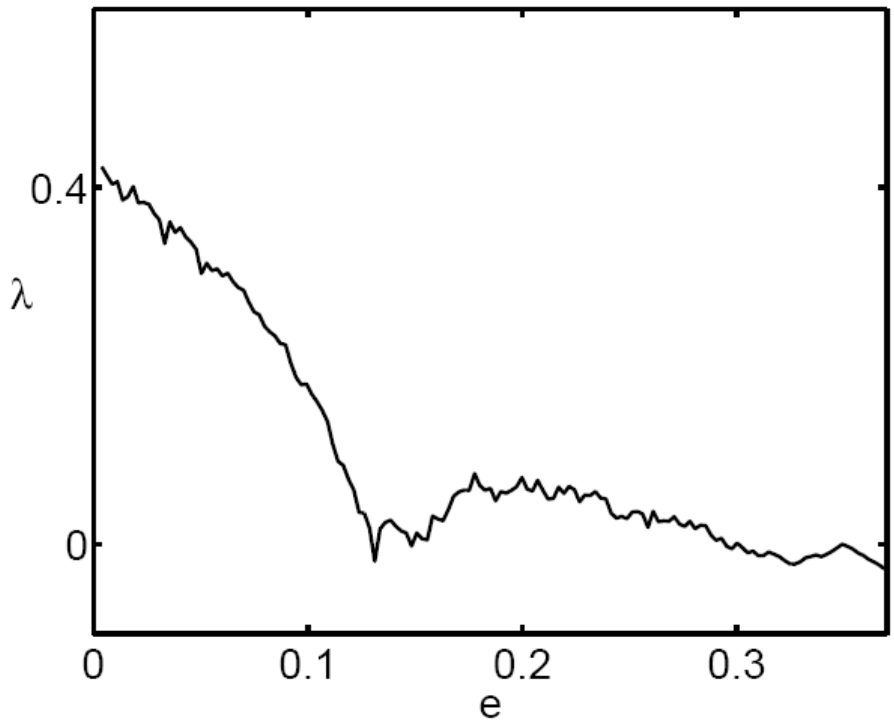
$$E_X = \int dxdX\,p(x,X)\left(x - \int dx'\,p(x'|\,X)x'\right)^2$$

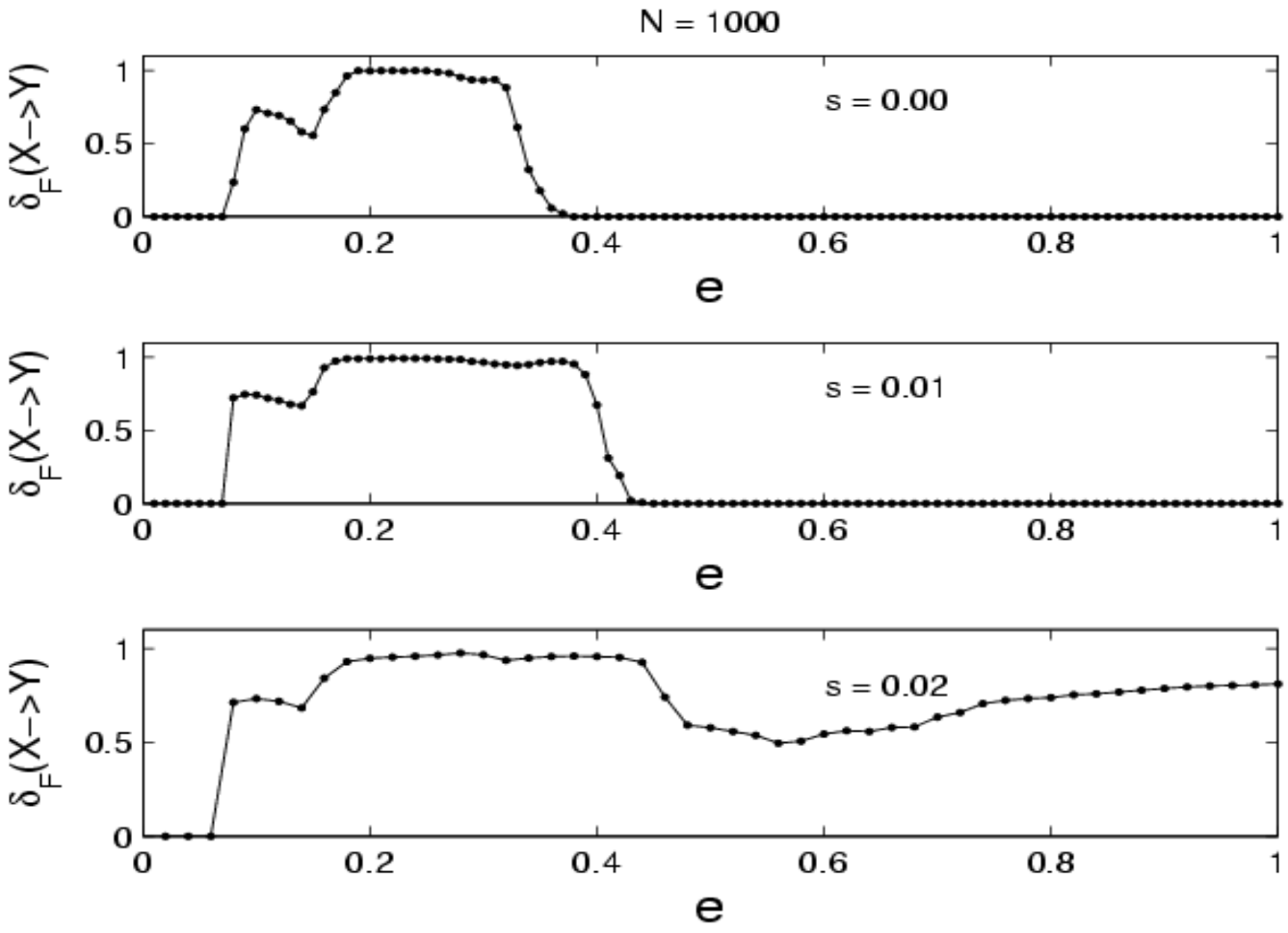
$$E_{XY} = \int dxdXdY\,p(x,X,Y)\left(x - \int dx'\,p(x'|\,X,Y)x'\right)^2$$

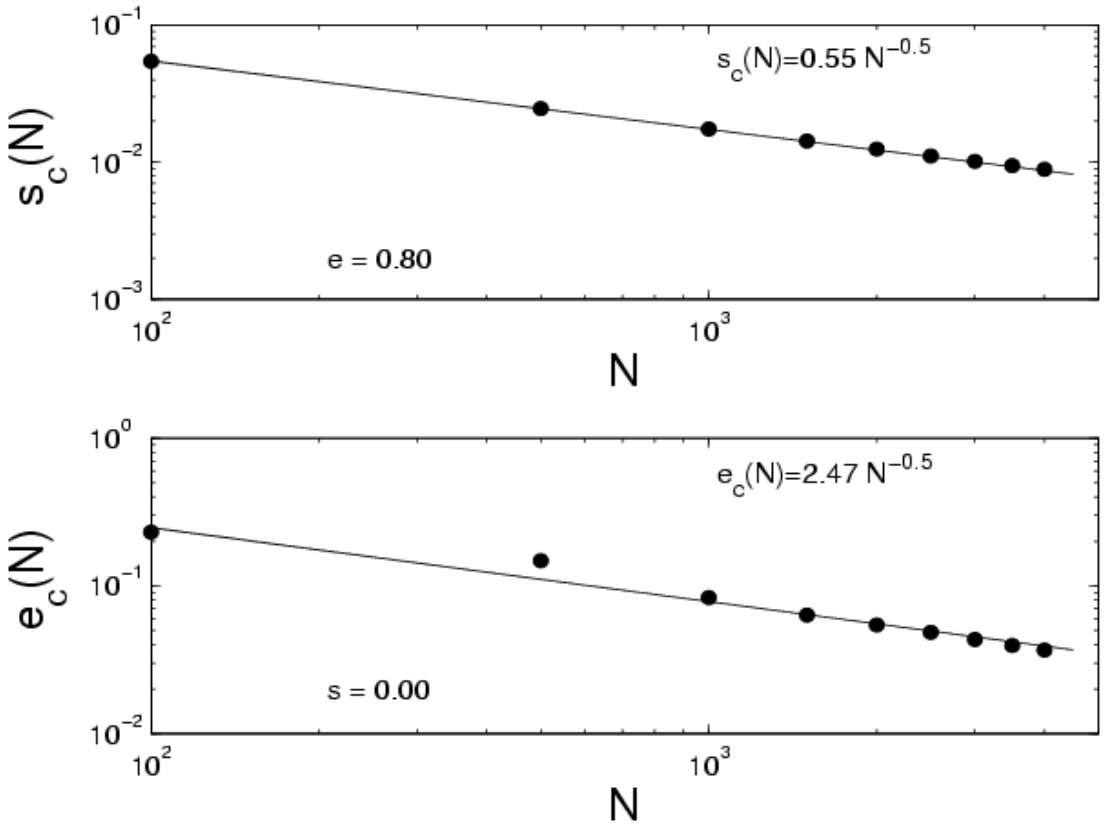
# Esempi di applicazione del nostro metodo

Pair of noisy logistic maps:

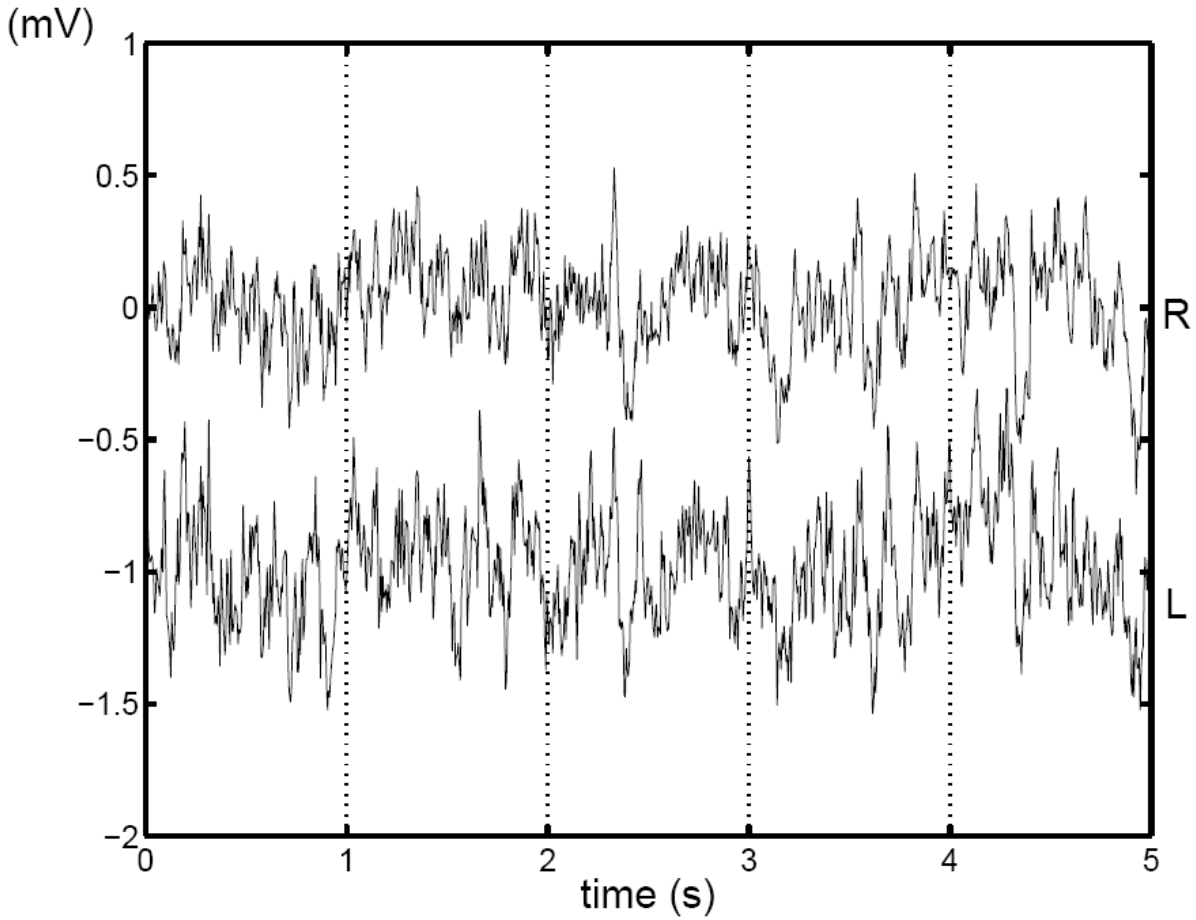
$$x_{n+1} = a\left(1 - x_n^2\right) + s\eta_{n+1},$$
$$y_{n+1} = (1-e)a\left(1 - y_n^2\right) + ea\left(1 - x_n^2\right) + s\xi_{n+1};$$



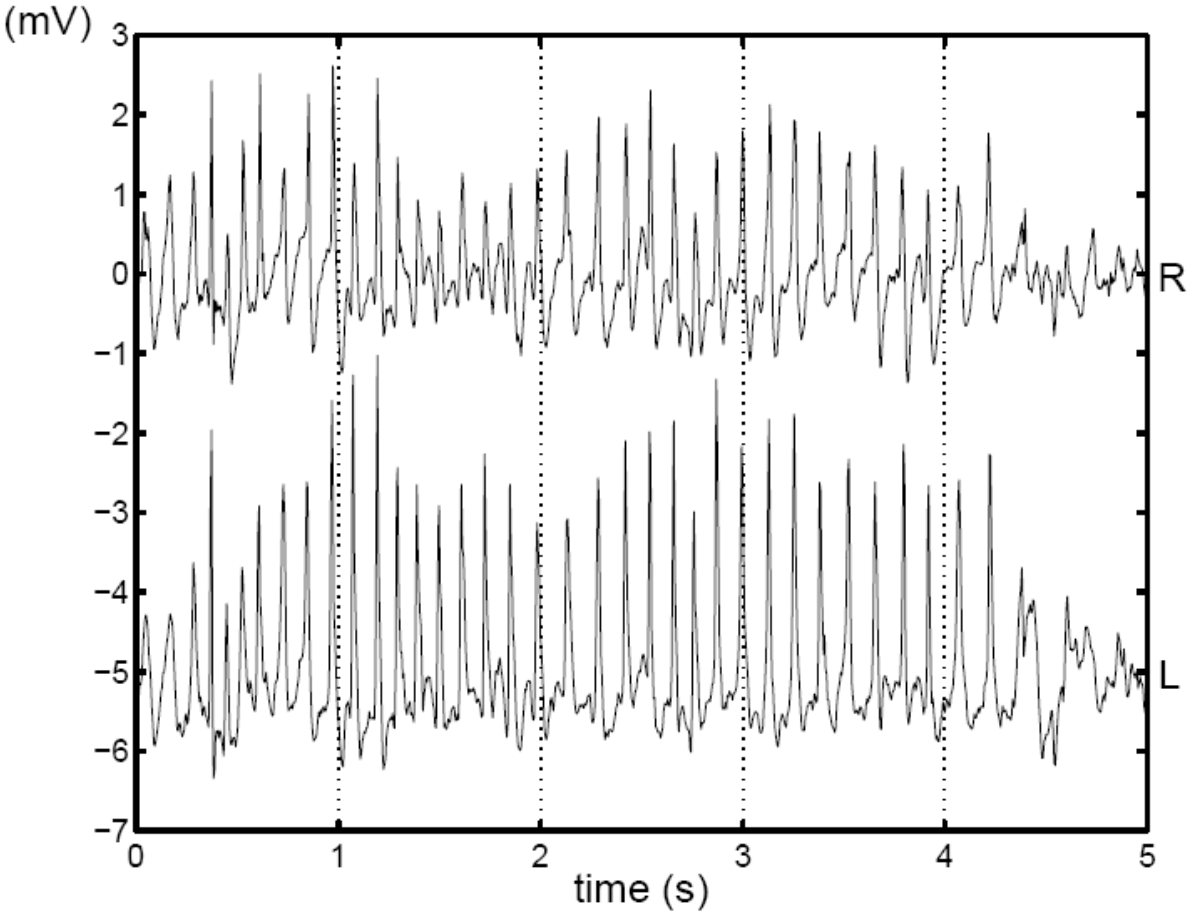

N = 1000
s = 0.00
s = 0.01
s = 0.02
δF(X->Y)
e
0
0.5
1
0.2
0.4
0.6
0.8

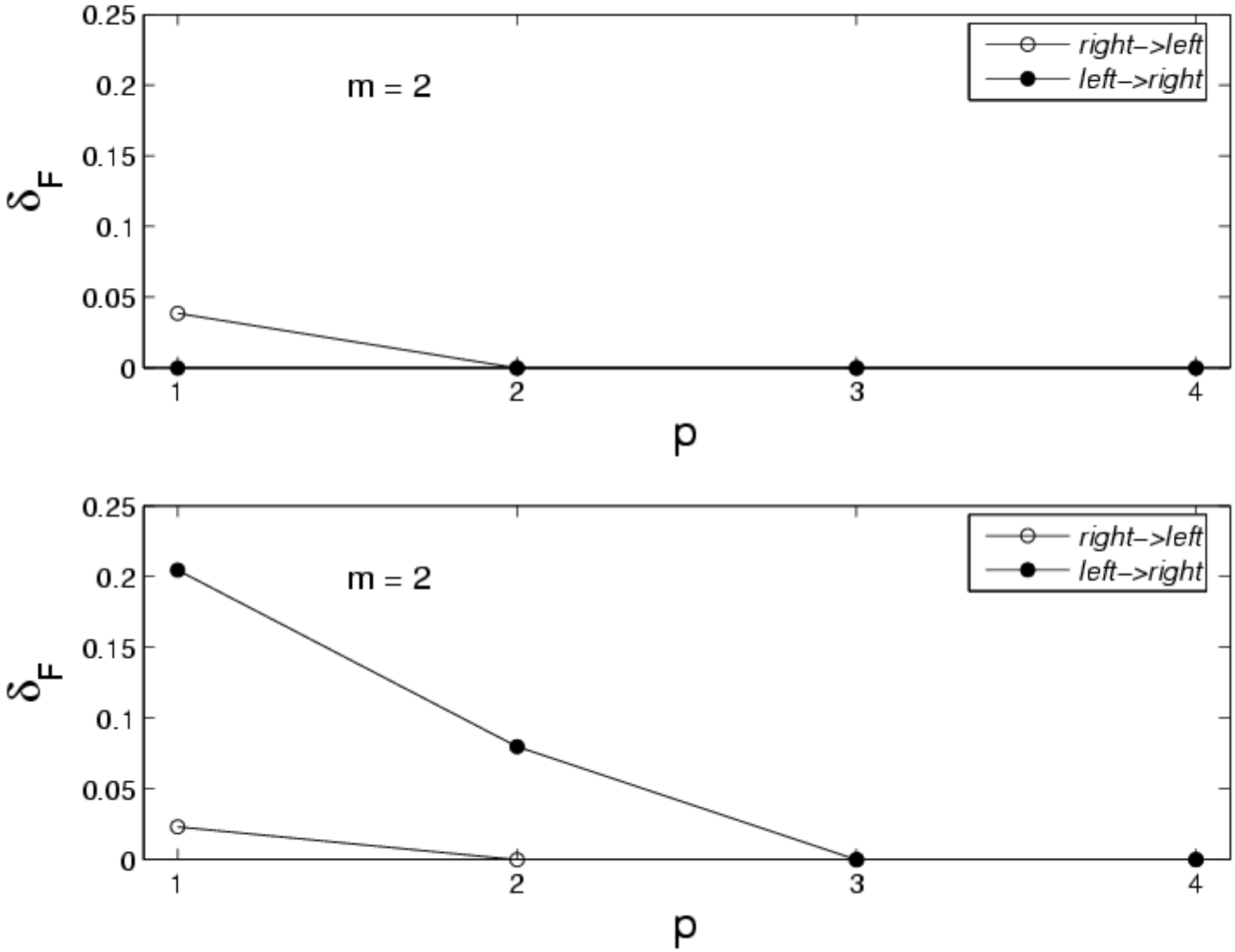

$s_c(N)=0.55\ N^{-0.5}$
e = 0.80
$s_c(N)$
N
$e_c(N)=2.47\ N^{-0.5}$
s = 0.00
$e_c(N)$
N

# Applicazione: EEG ratto

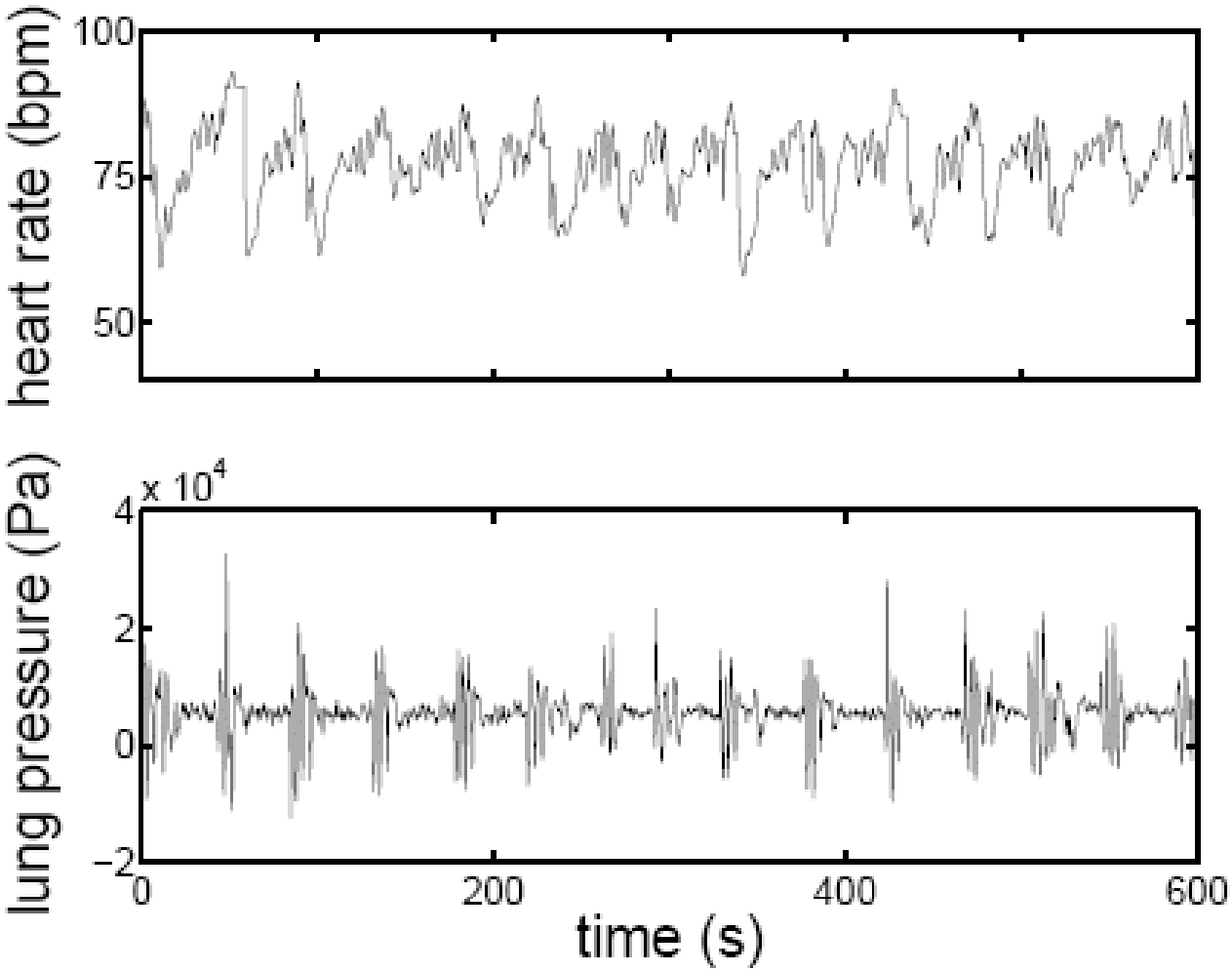


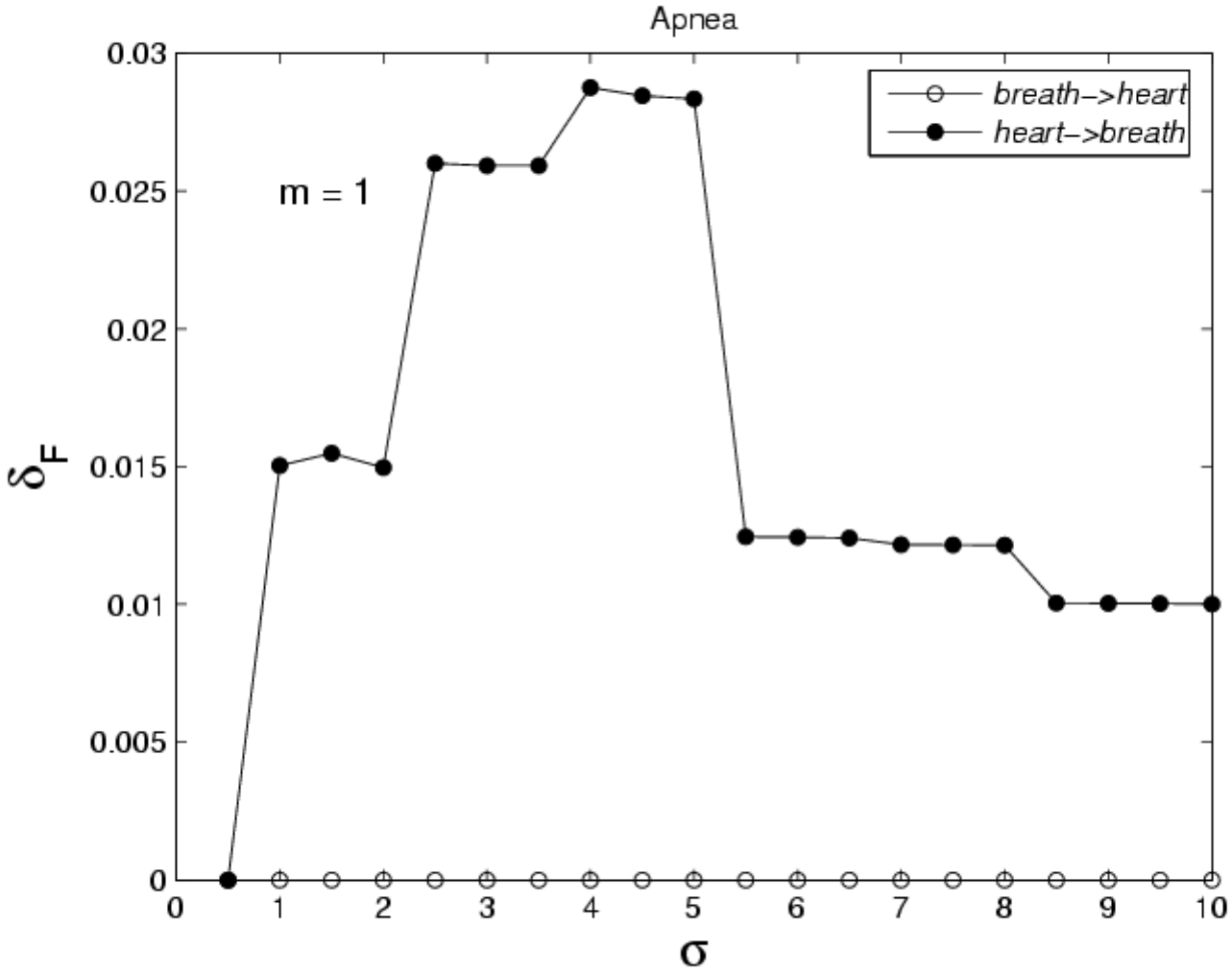
# EEG ratto: dopo la lesione




0.25
0.2
0.15
0.1
0.05
0
$\delta_F$
m = 2
right->left
left->right
1
2
3
4
p
0.25
0.2
0.15
0.1
0.05
0
$\delta_F$
m = 2
right->left
left->right
1
2
3
4
p

# Applicazione: apnea nel sonno



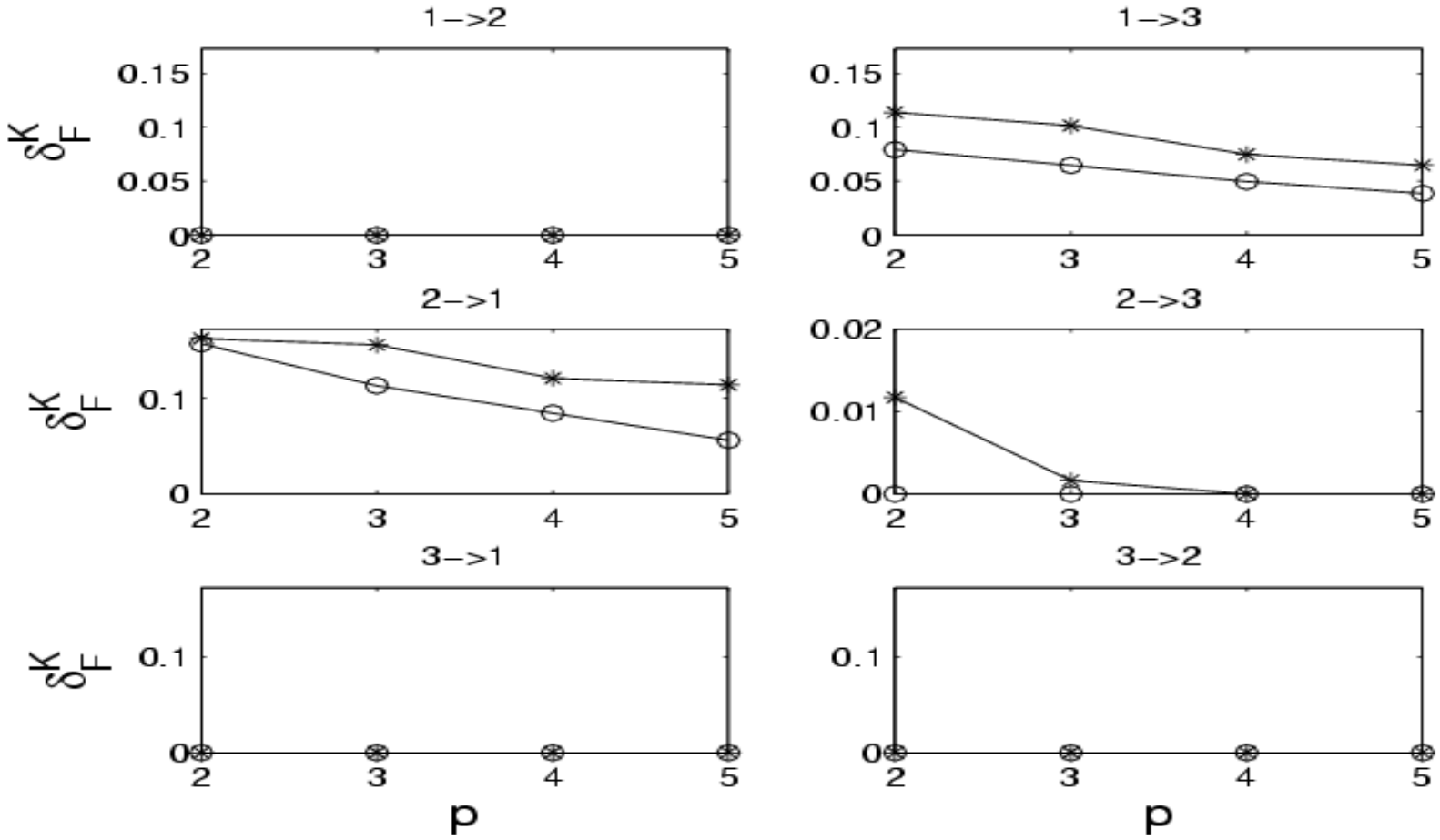

Apnea
0.03
0.025
0.02
0.015
0.01
0.005
0
0
1
2
3
4
5
6
7
8
9
10
σ
δF
m = 1
breath->heart
heart->breath

# Multivariate Granger causality and indirect relationships

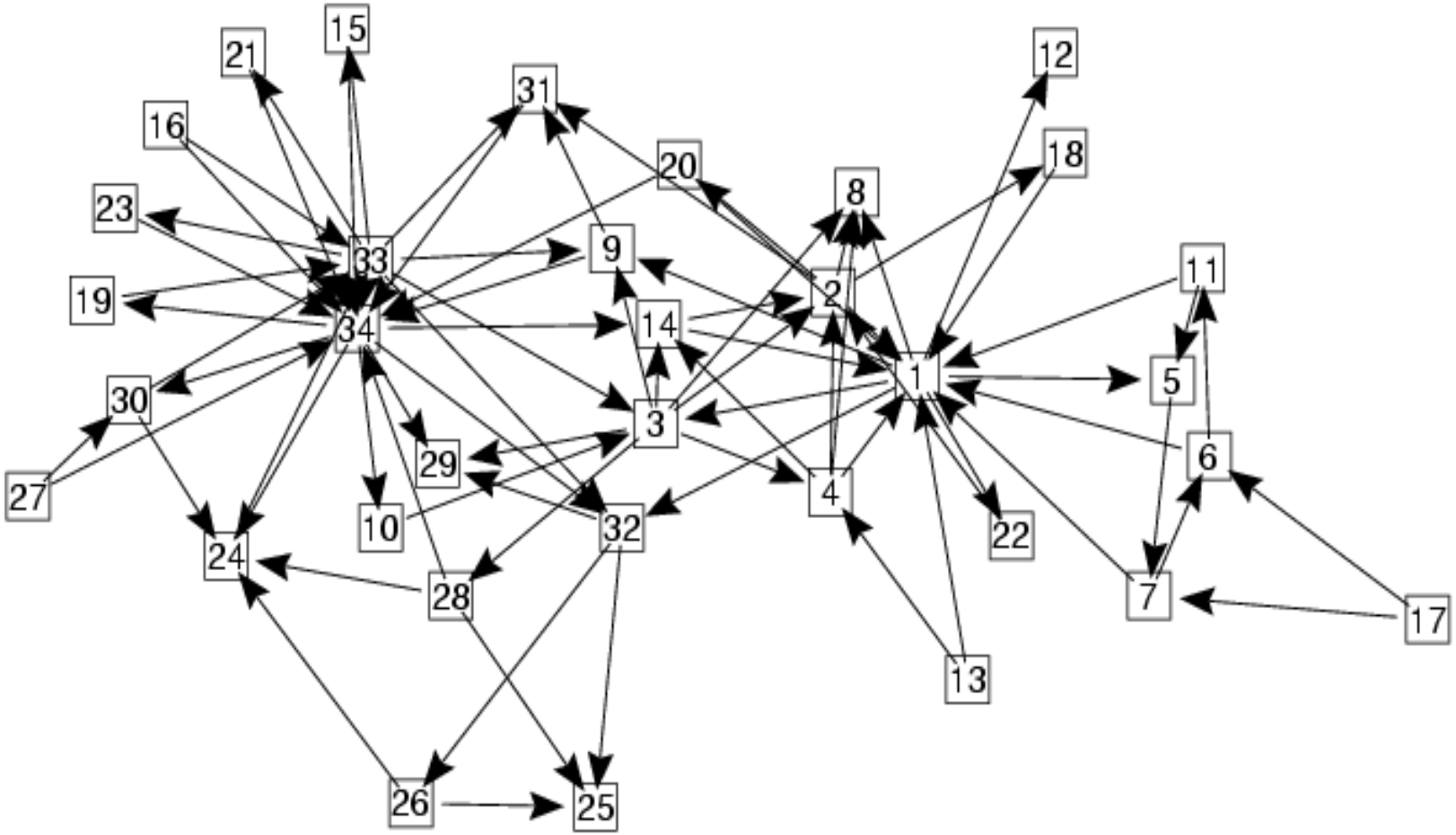
## Three logistic maps: 2→1→3

$$
\begin{aligned}
x_{1,t} &= 0.8\left(1 - ax_{1,t-1}^2\right) + 0.2\left(1 - ax_{2,t-1}^2\right) + s\tau_{1,t}, \\
x_{2,t} &= 1 - ax_{2,t-1}^2 + s\tau_{2,t}, \\
x_{3,t} &= 0.8\left(1 - ax_{3,t-1}^2\right) + 0.2\left(1 - ax_{1,t-1}^2\right) + s\tau_{3,t};
\end{aligned}
$$

# Three logistic maps: 2→1→3



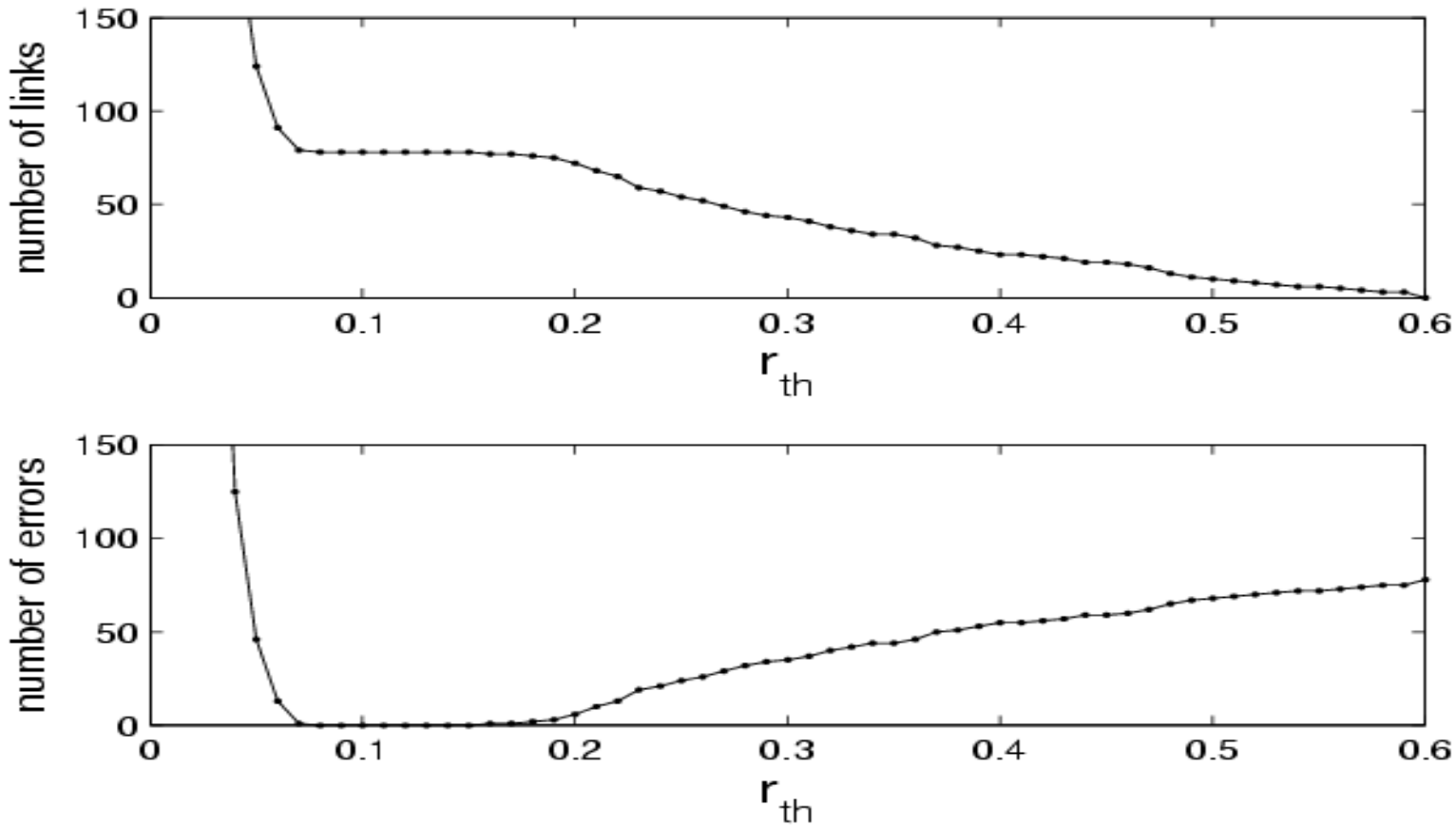
# Dynamical network



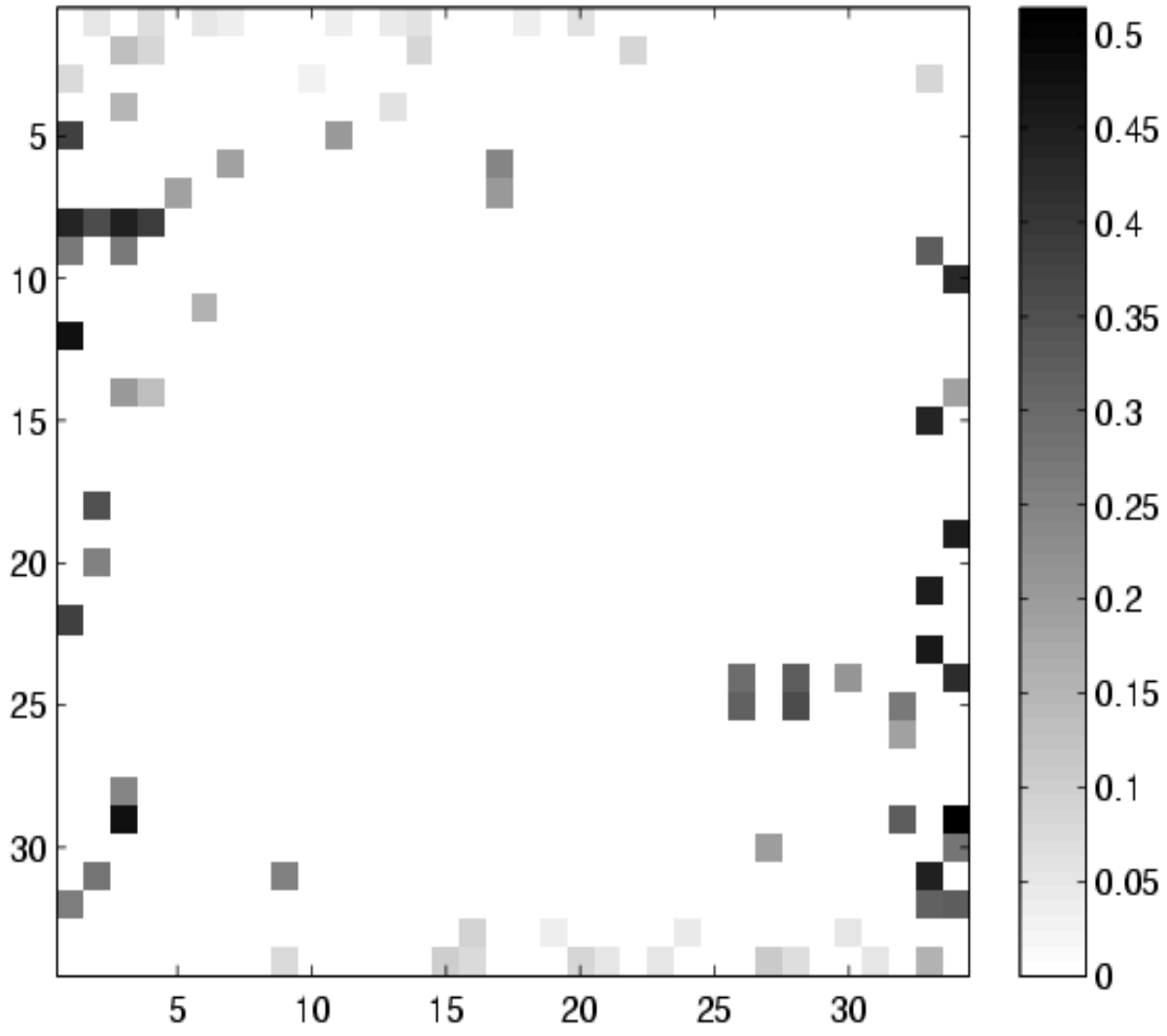
# Coupled Chaotic maps

$$x_{i,t} = \left(1 - \sum_{j=1}^{n} c_{ij}\right)\left(1 - a x_{i,t-1}^2\right) + \sum_{j=1}^{n} c_{ij}\left(1 - a x_{j,t-1}^2\right) + s\tau_{i,t},$$

# Results


number of links
150
100
50
0
0
0.1
0.2
0.3
0.4
0.5
0.6
r th
number of errors
150
100
50
0
0
0.1
0.2
0.3
0.4
0.5
0.6
r th

# Causalities vs coupling

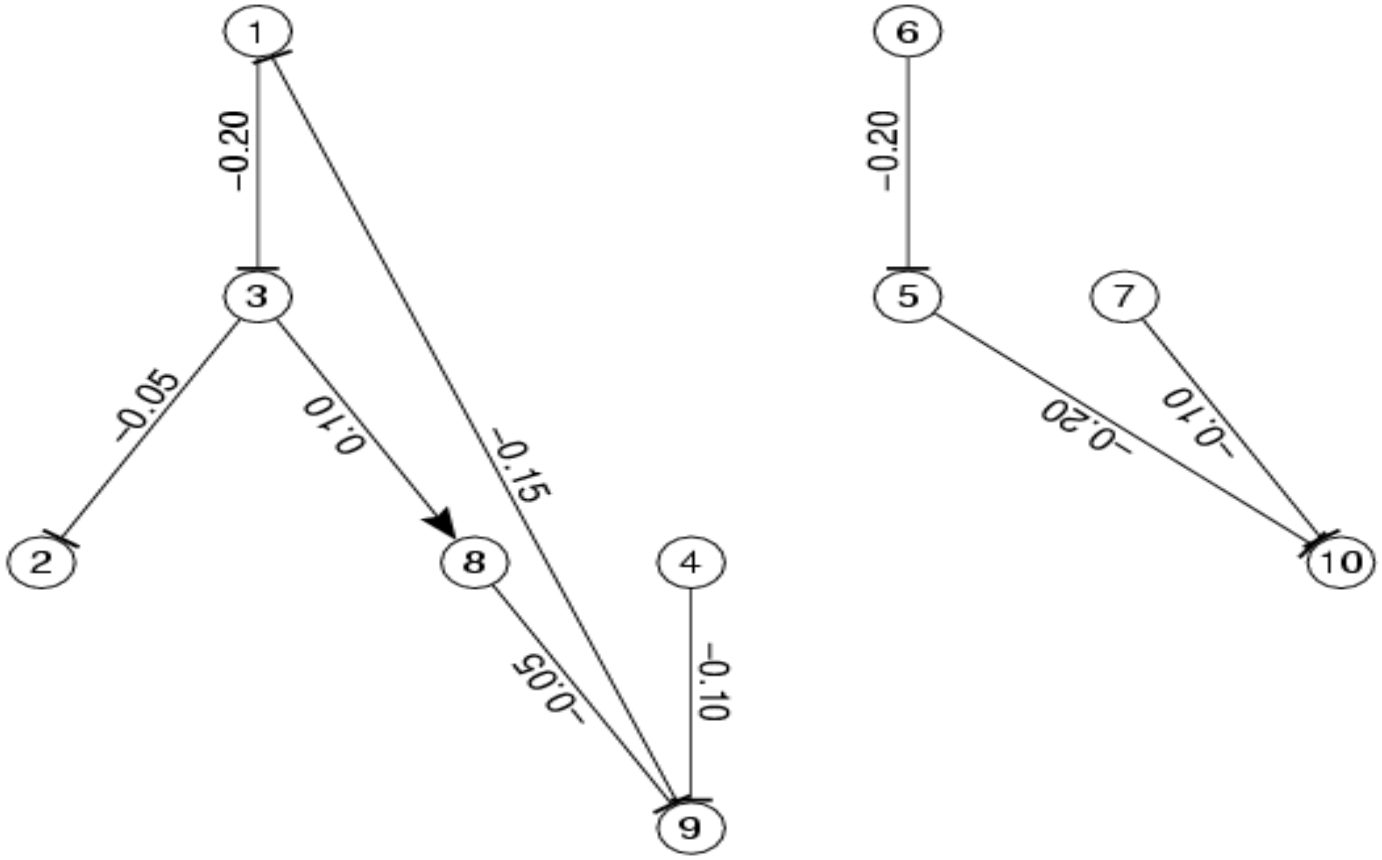


# Genetic regulatory network

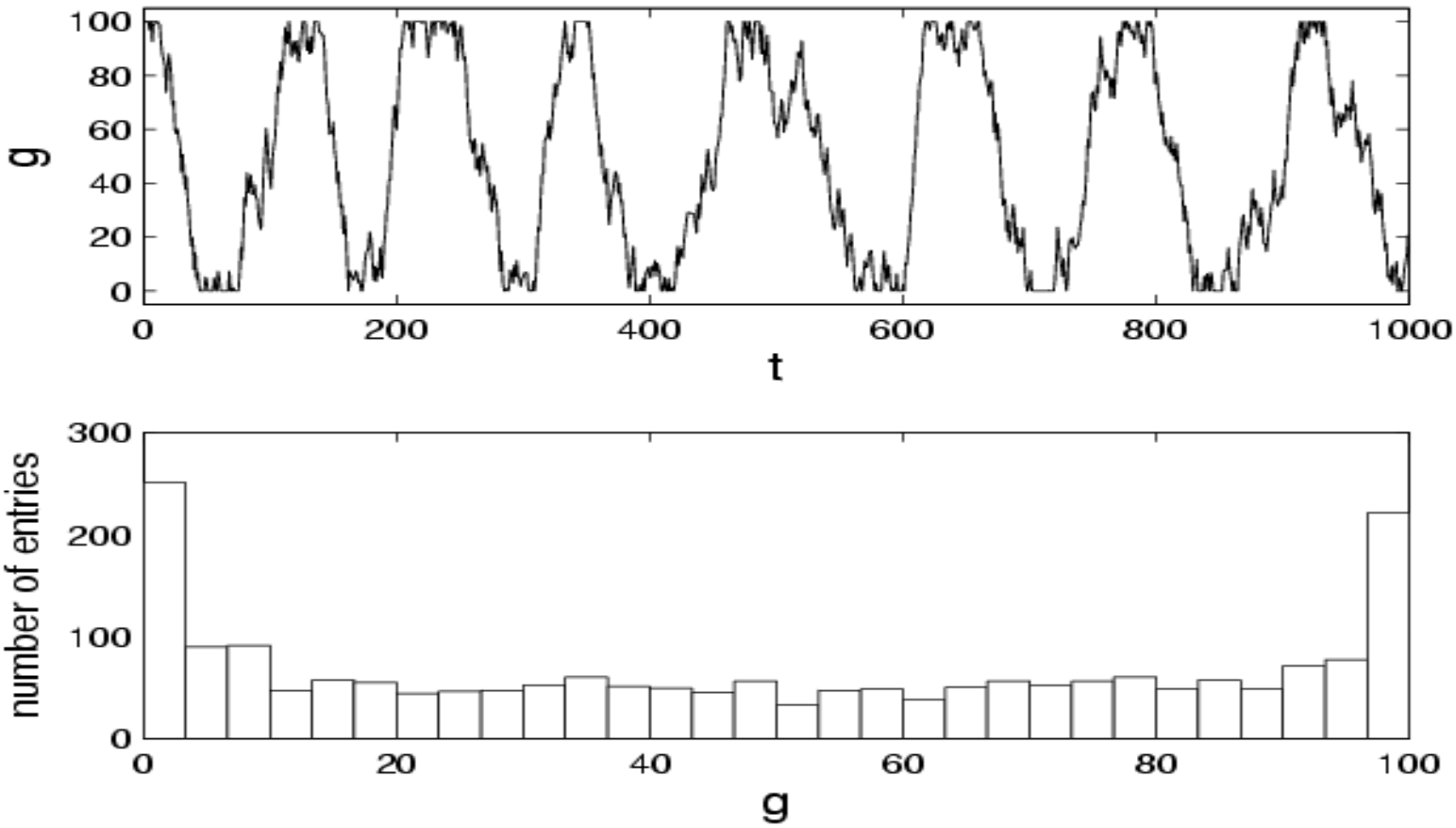
$$\mathbf{g}_{t+1} = \mathbf{g}_t + A\left(\mathbf{g}_t - T\mathbf{I}\right) + \mathbf{\Sigma}, \tag{1}$$

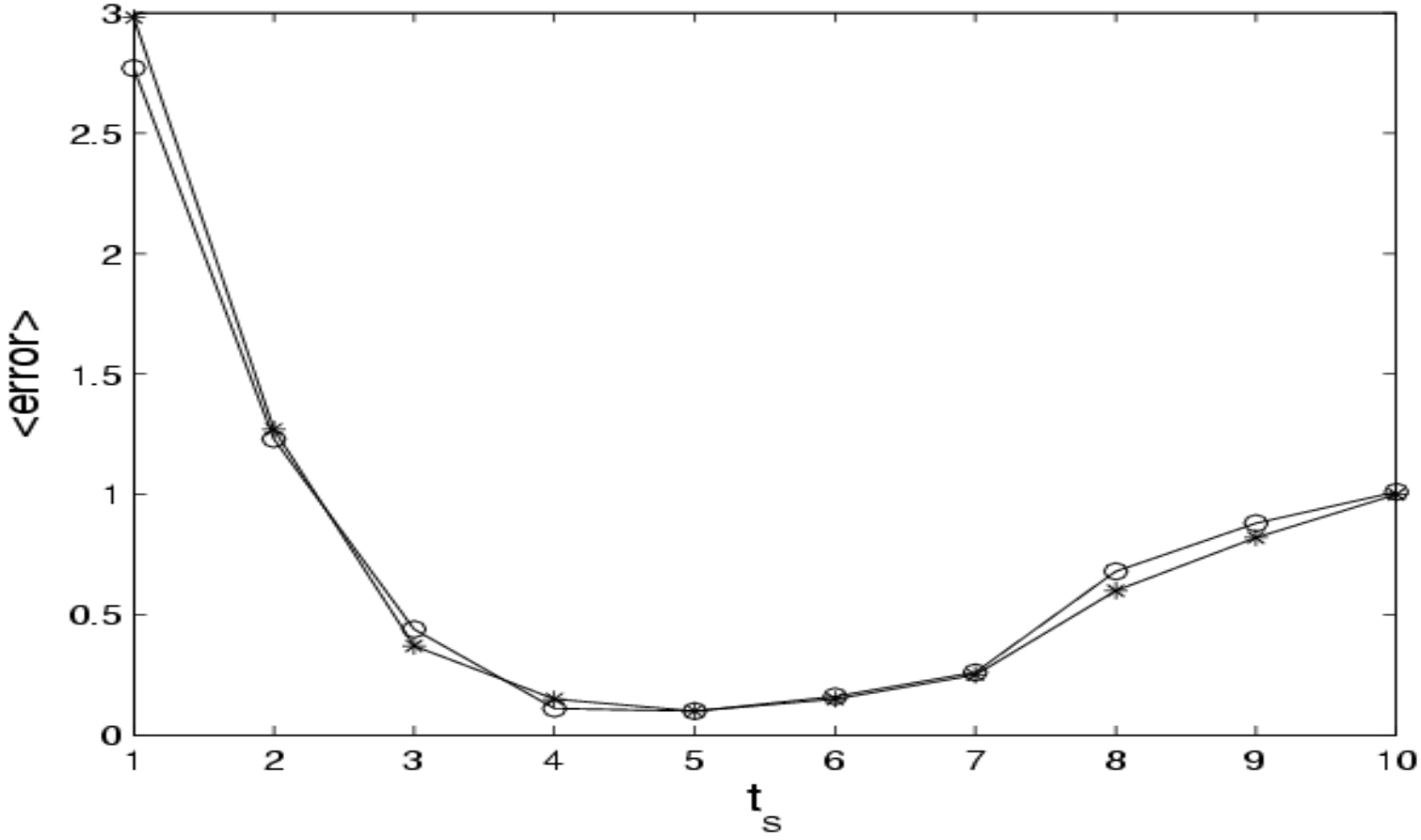
where $A$ is a connectivity strength matrix corresponding to the network, $T = 50$, $\mathbf{I}$ is the identity matrix and $\mathbf{\Sigma}$ is a vector of random variables uniform in $[-10, 10]$. The values of $\mathbf{g}$ are restricted by floor and ceiling function to range in $[0, 100]$: this constraint provides the nonlinear character of the model.

# Genetic regulatory network



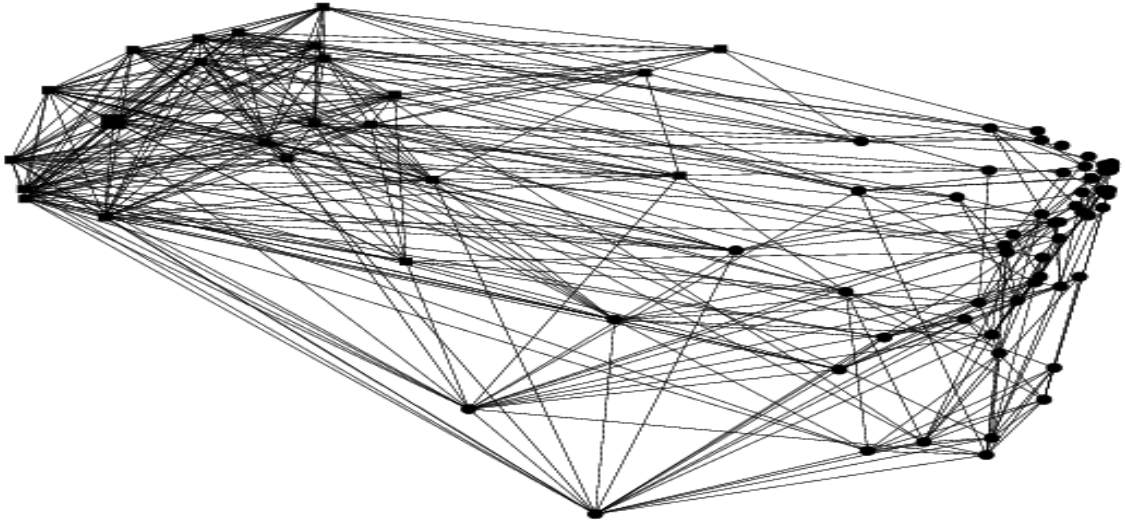
# Typical evolution of the expression of a gene




3
2.5
2
1.5
1
0.5
0
1
2
3
4
5
6
7
8
9
10
<error>
$t_s$

# HeLa cell cycle

- 94 genes and 48 time points (an hour interval separating two measures).
- Analysis of static correlations: 800 pairs of genes are significantly correlated
- The analysis of the undirected network shows two modules (23 and 62 genes resp.), corresponding to well known transcriptional factors (NFkB, p53-STAT3)



# Granger causality

- 19 causalities acting on the time scale of the sampling rate are significative

| | | | |
|---|---|---|---|
| TPD52L | → | TNF-a | (0.6166) |
| TSP1 | → | Bcl-XL | (0.5935) |
| TSP1 | → | c-myc | (0.5905) |
| IRF-2 | → | BRCA1 | (0.5669) |
| c-myc | → | FGFR4 | (0.5653) |
| R2 | → | c-myc | (0.5642) |
| PKIG | → | TSP1 | (0.5475) |
| IAP | → | NFkB | (0.5457) |
| NFkB | → | B99 | (0.5416) |
| Bcl-XL | → | OCT4 | (0.5354) |
| OCT4 | → | VCAM-1 | (0.5329) |
| VCAM-1 | → | TPD52L | (0.5315) |
| Killer/DR5 | → | c-myc | (0.5313) |
| A20 | → | Bcl-XL | (0.5284) |
| Cyclin E1 | → | E2F-1 | (0.5249) |
| PKIG | → | ICAM-1 | (0.5243) |
| IkappaBa | → | NFkB | (0.5156) |
| TPD52L | → | MASPIN | (0.5128) |
| FGF7 | → | MCP-1 | (0.5107) |