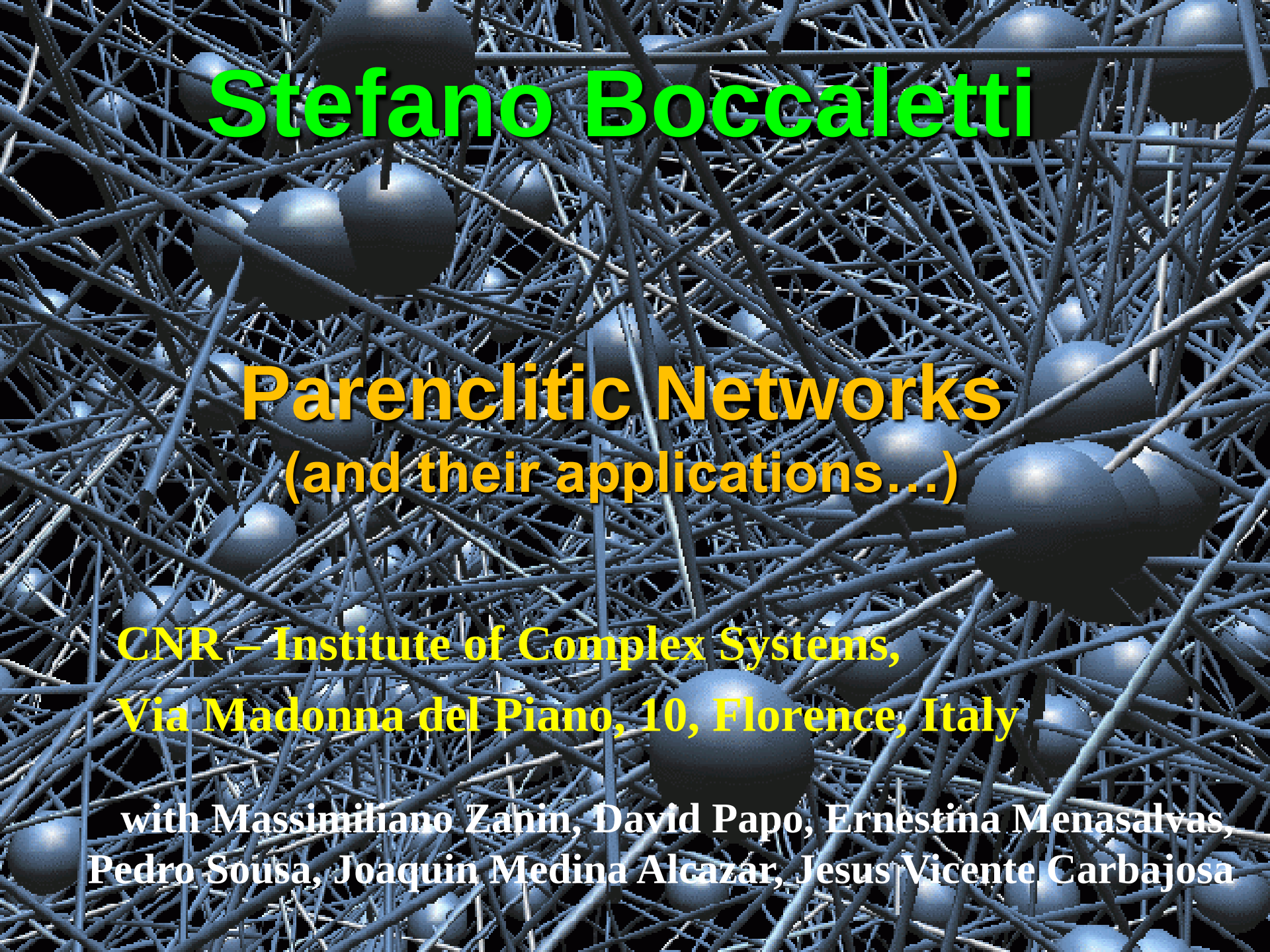


Stefano Boccaletti

Parenclitic Networks
(and their applications…)

CNR – Institute of Complex Systems,

Via Madonna del Piano, 10, Florence, Italy

with Massimiliano Zanin, David Papo, Ernestina Menasalvas, Pedro Sousa, Joaquin Medina Alcazar, Jesus Vicente Carbajosa

Network representation of complex systems

- 1) Physical networks
- 2) Functional networks
- 3) Parenclitic networks

Physical Networks are those where nodes and links can be directly extracted from the structure of interactions among the constituents of the system.

Society →

Nodes: individuals

Links: social relationship (family/work/friendship/etc.)

WWW →

Nodes: WWW documents

Links: URL links

Transportation →

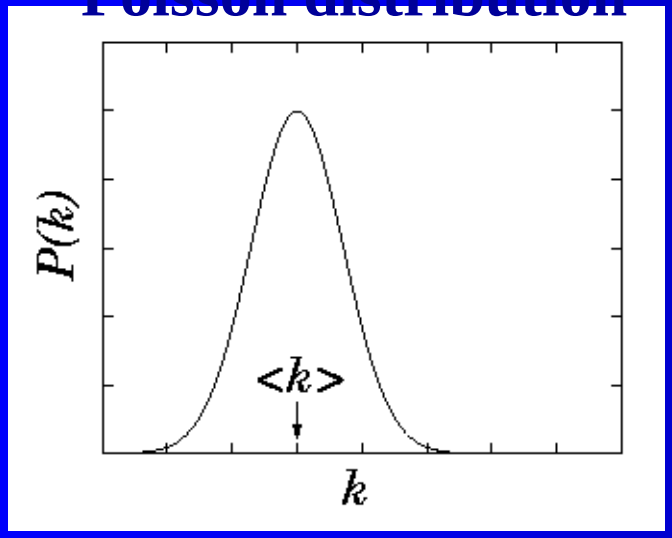
Nodes: airports

Links: flights

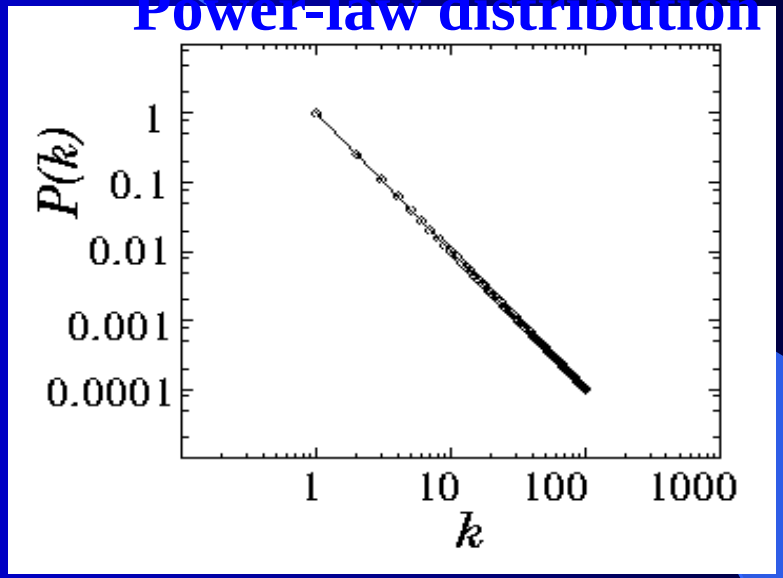
and thousands more…

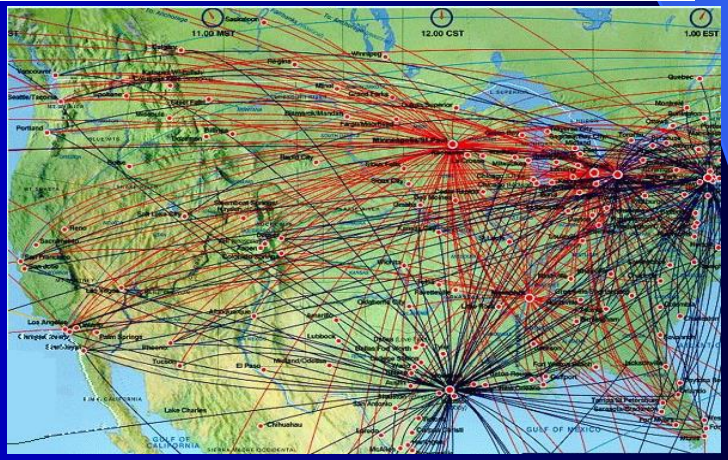
Road and Airline networks

Poisson distribution



Power-law distribution

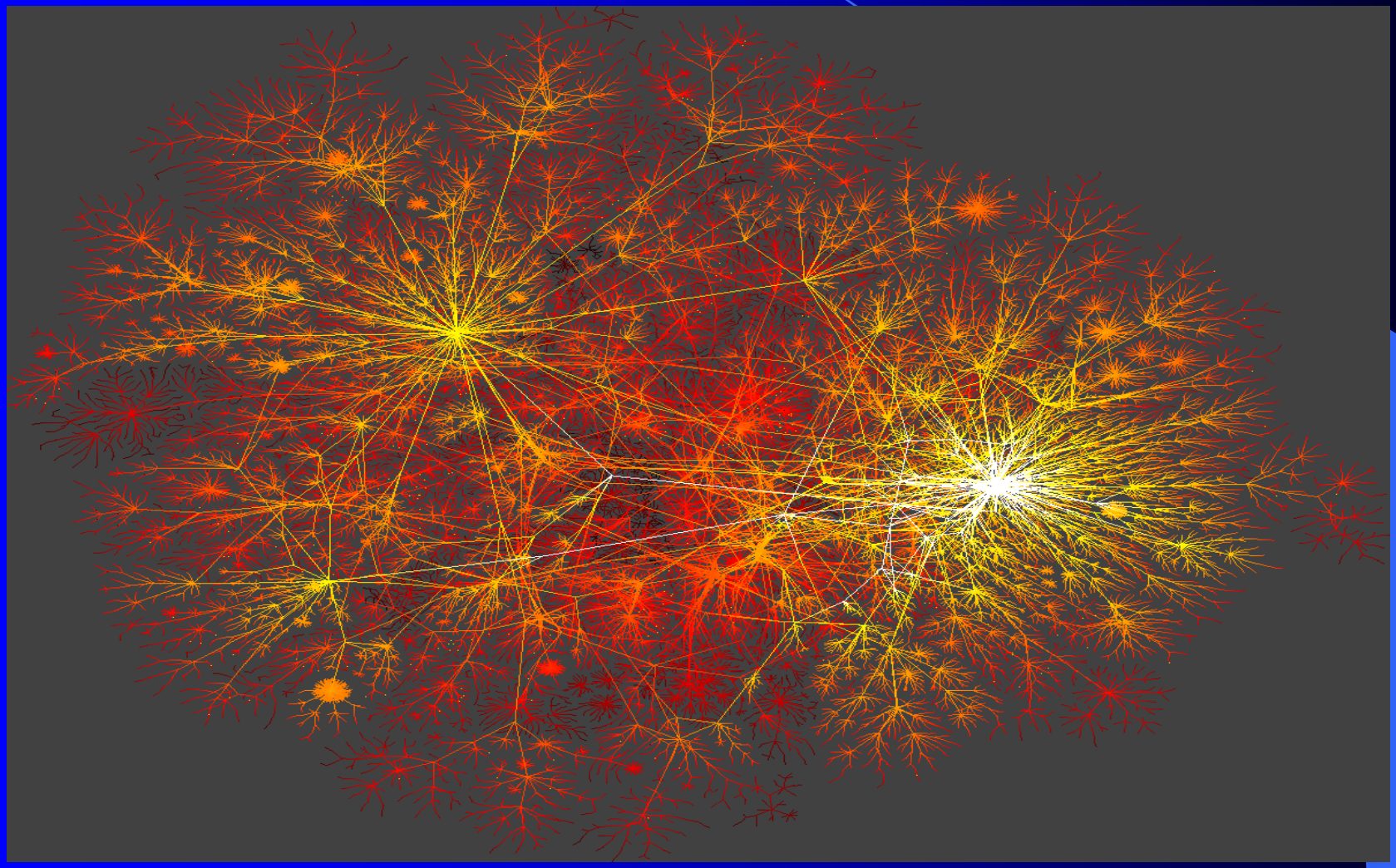




Exponential Network

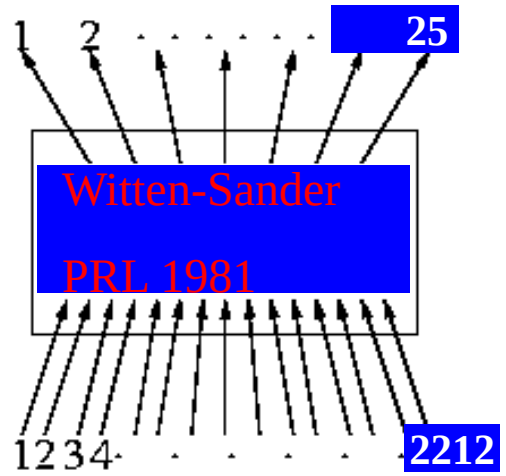
Scale-free Network

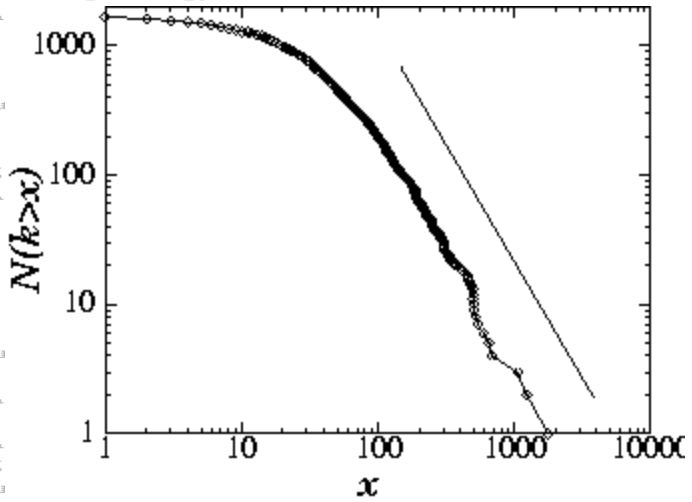
INTERNET BACKBONE



SCIENCE CITATION INDEX

Nodes papers
Links citations





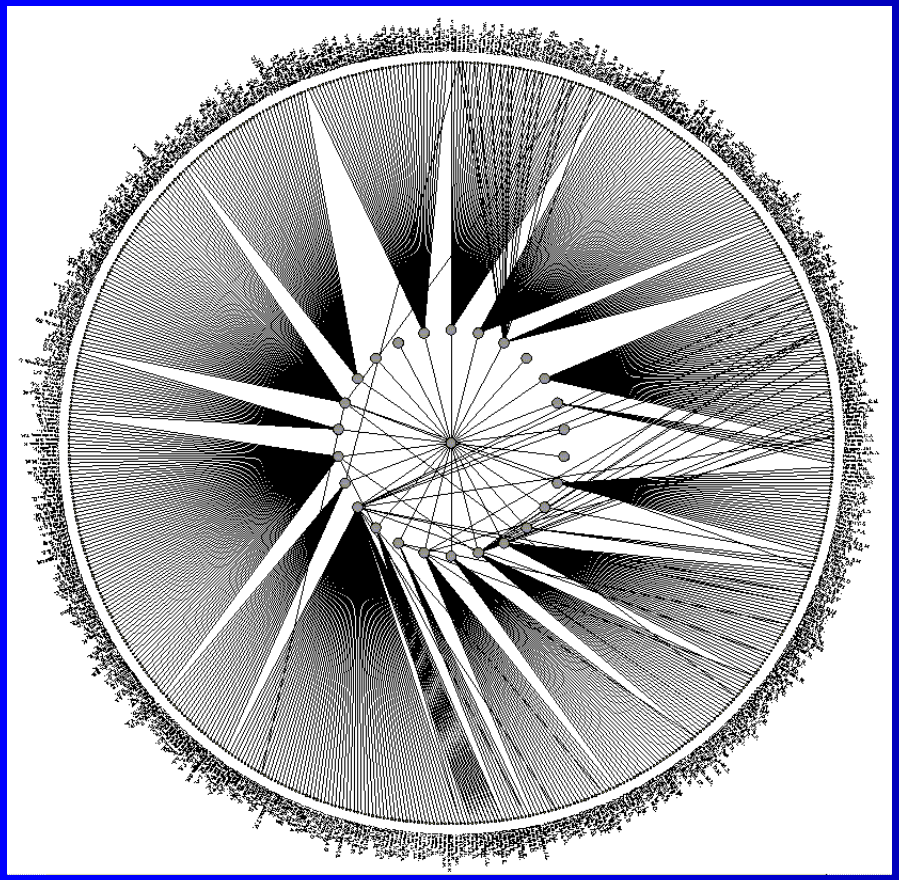
$P(k) \sim k^{-\gamma}$

1,000 Most Cited Physicists, 1981-June 1997
Out of over 500,000 Examined
(see http://www.sst.nrel.gov)

Author name		Institute	Country	Field	avg. cites	total art.	total cites	rank by total cit.
Wit		n (U)	USA, NJ	High-energy (T)	168	138	23235	1
Gos		CE	USA, CA					2
Cava	RJ	Bell Labs (I)	USA, NJ					3
Batl	B	Bell Labs (I)	USA, NJ					4
Ploo	K	ck (NL)	Germany					5
Blis	J	Euro Nuclear Cent.	Switzerla					6
Fisk	Z	Florida State (U)	USA, FL					7
Cardona	M	Max Planck (NL)	Germany					8
Na			USA, TX					9
He			USA, CA					10
Le								11
Su								12
An			USA, NJ					13
Su								14
Fre			USA, IL					15
Ta								16
Mi			Switzerla					17
Sel			USA, NJ					18
Ch			USA, CA					19
Mc			USA, IL					19
Mi			USA, CA					21
Ch		)	USA, TX					22
Be			Switzerla					23
Co		)	USA, CA	Solid State (T)	33	284	9311	23
Mc		)	USA, TX	Superconductivity (E)	86	108	9300	25
Wa			USA, NJ	Superconductivity (E)	57	162	9170	26
Sh			USA, NY	Superconductivity (E)	33	269	8841	27
Wi			USA, NJ	Semiconductors (E)	86	104	8822	28
Va			USA, NJ	Magnetism (E)	46	229	8686	29
Uc					25	301	8520	30
Ho		)	USA, TX	Superconductivity (E)	72	119	8512	31
Murphy	DW			Astronomy (E)	111	76	8439	32
Birgeneau	RJ	MIT (U)	USA, MA	Superconductivity (E)	41	286	8375	33
Jorgensen	JD	Argonne (NL)	USA, IL	Superconductivity (E)	50	167	8298	34
Hinks	DG	Argonne (NL)	USA, IL	Superconductivity (E)	37	223	8263	35

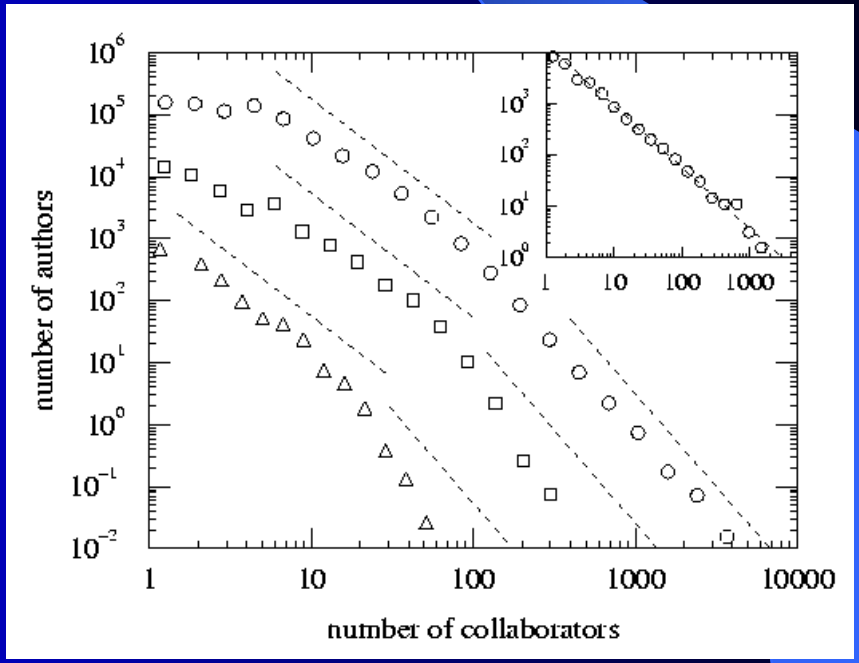
* citation total may be skewed because of multiple authors with the same name

SCIENCE COAUTHORSHIP

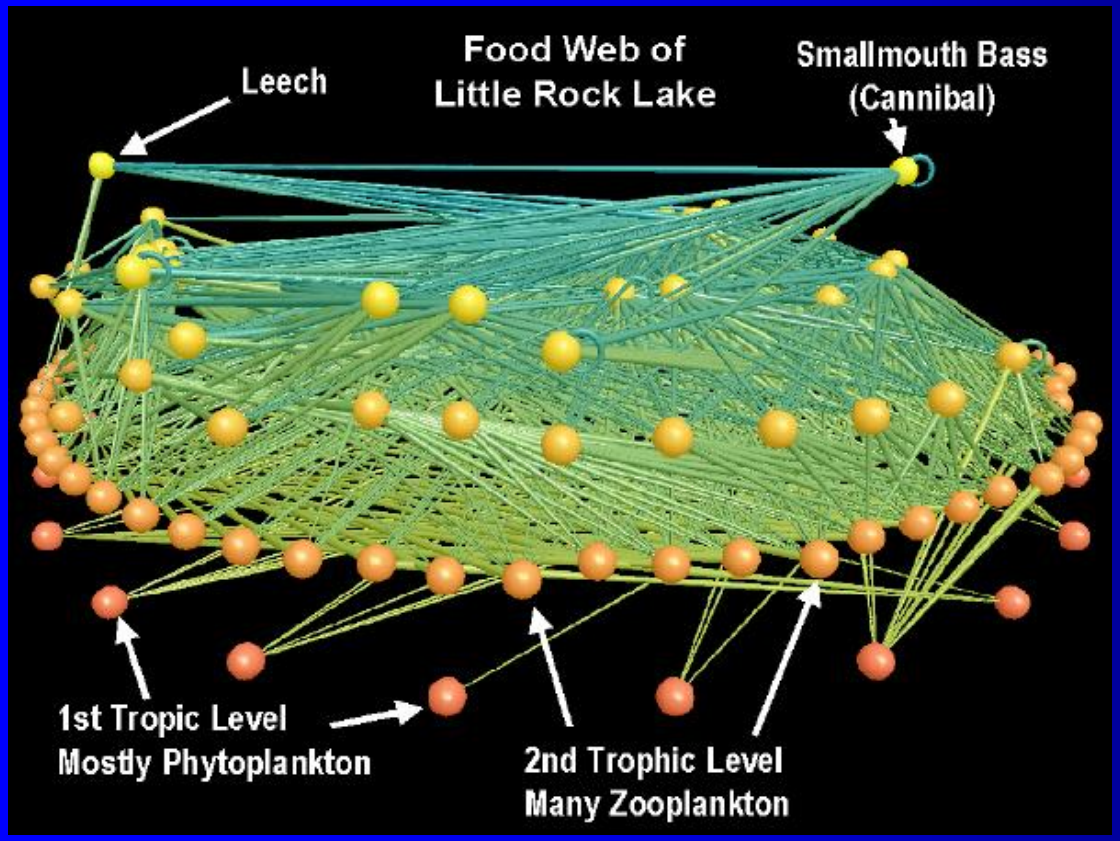


Nodes: scientist (authors)

Links: write paper together



FOOD WEBS



Nodes: trophic species

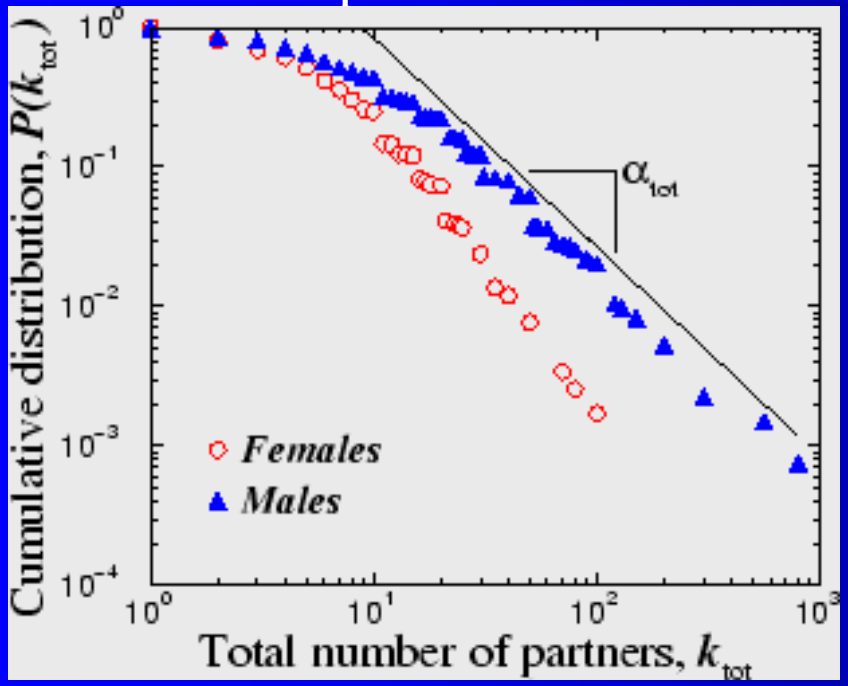
Links: trophic interactions

R.J. Williams, N.D. Martinez *Nature* (2000)

SEX WEBS

people (Females; Males)

Links: sexual relationships

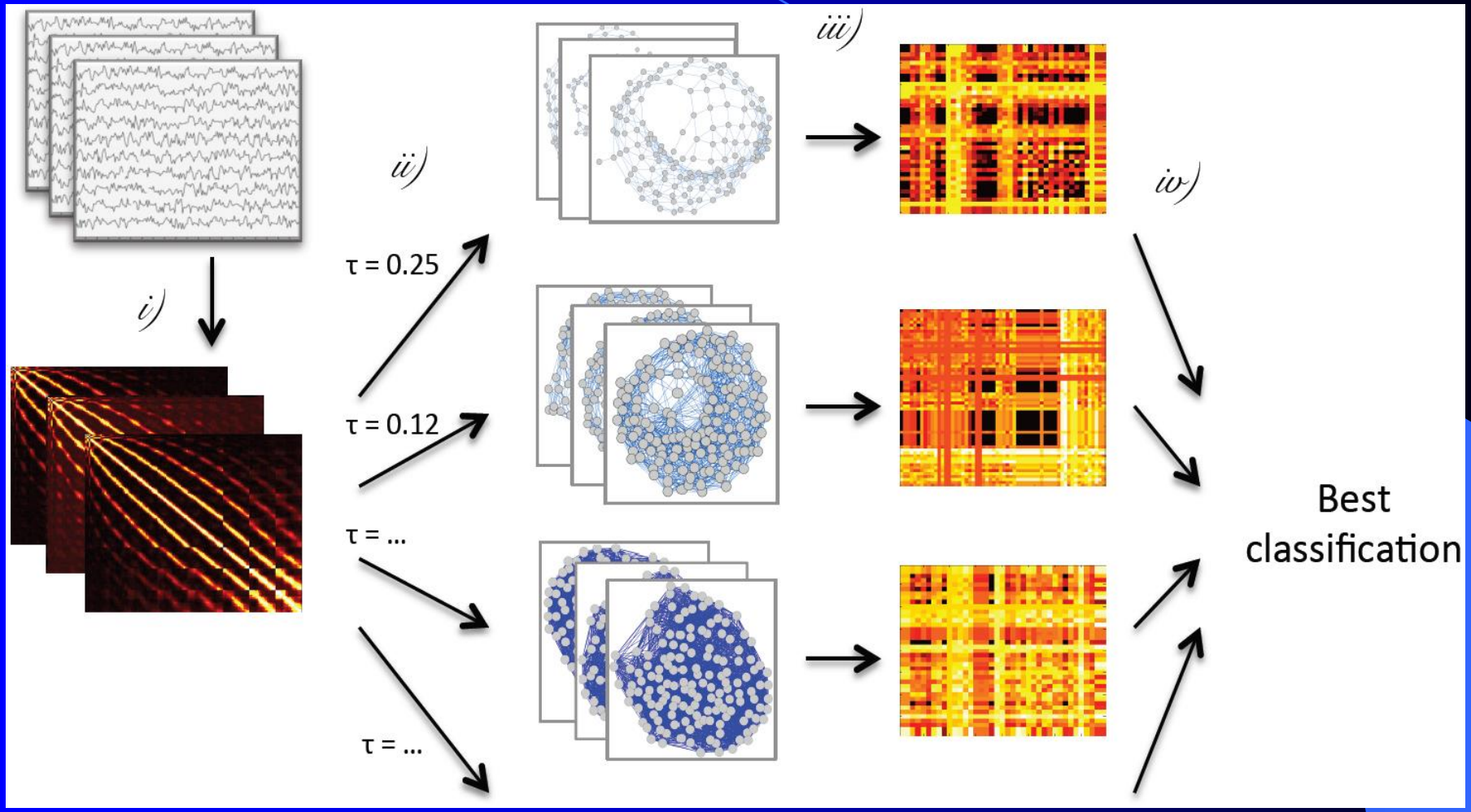


4781 Swedes; 18-74;

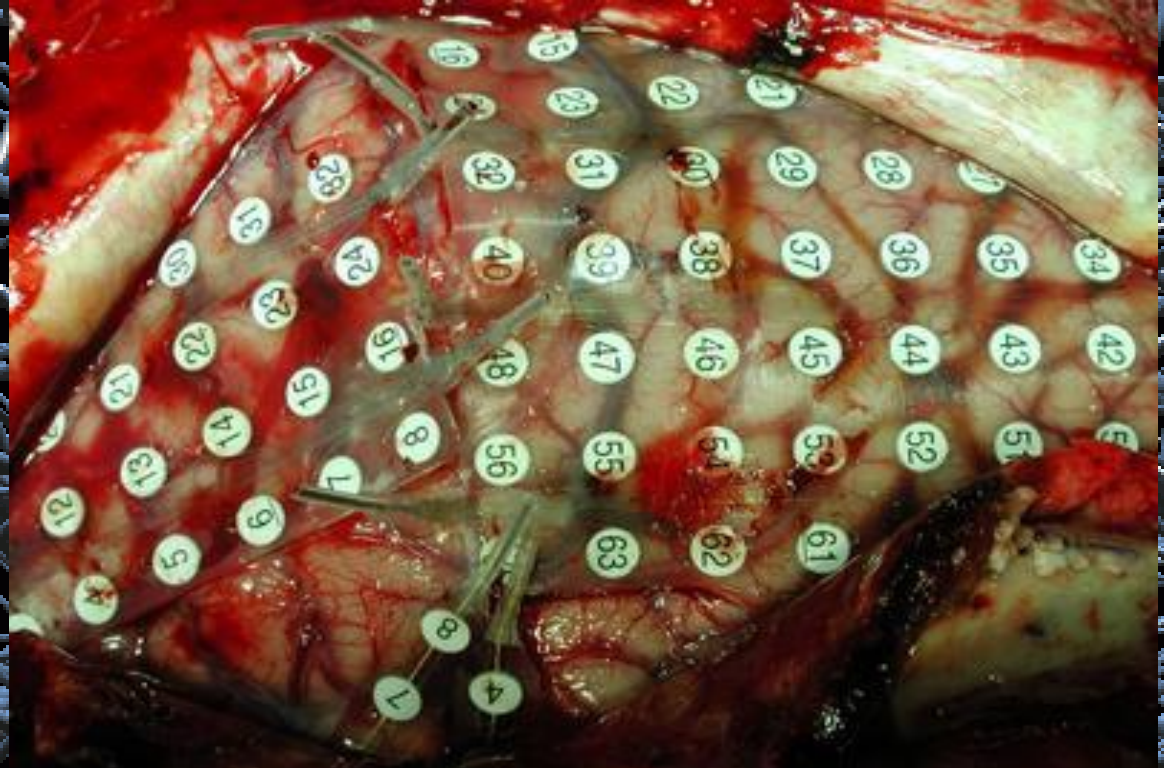
59% response rate.

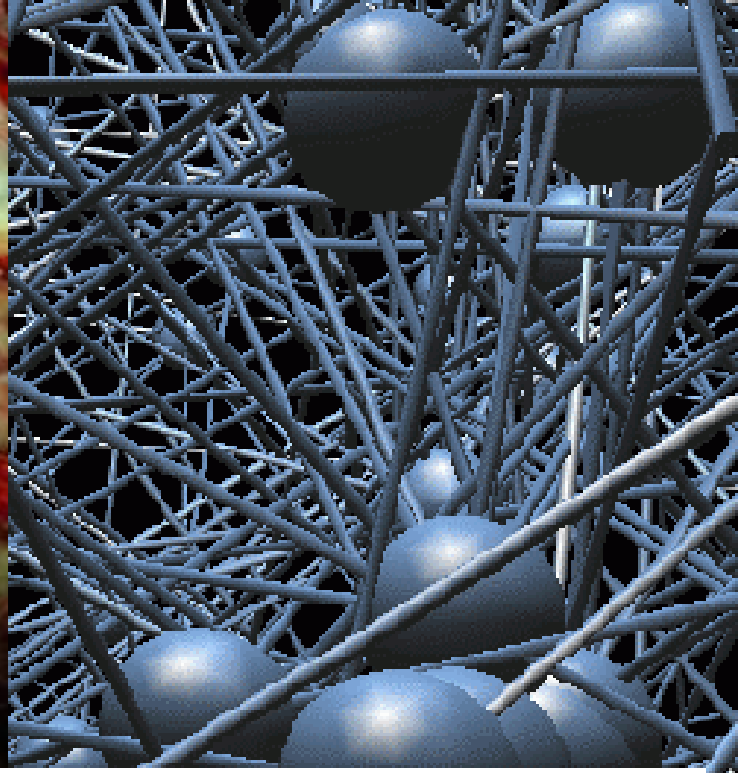
Liljeros et al. Nature 2001

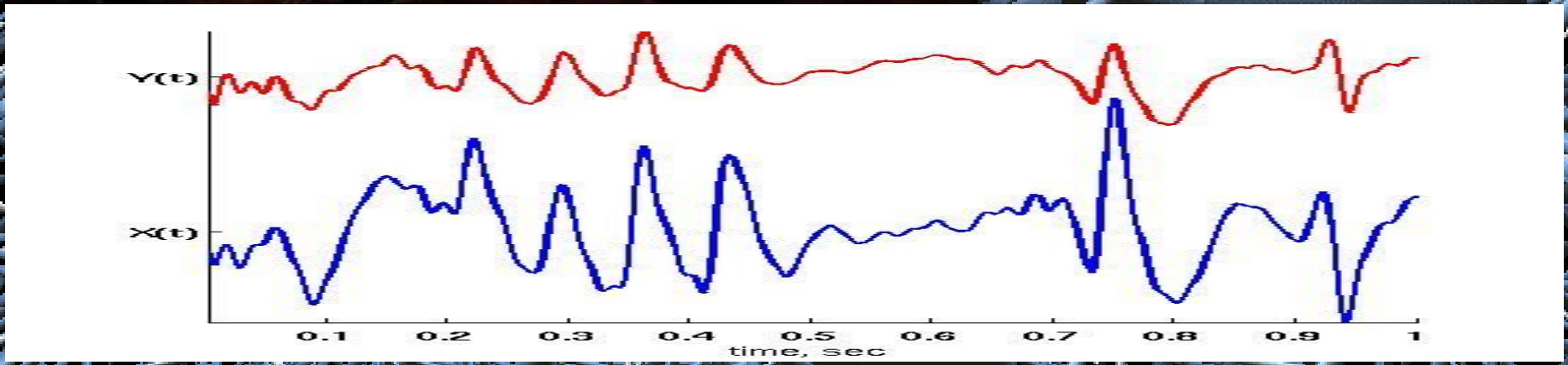
Functional Networks are those where links quantify correlation in multivariate time series analysis.



See: Nature Scientific Reports 2, 630 (2012)

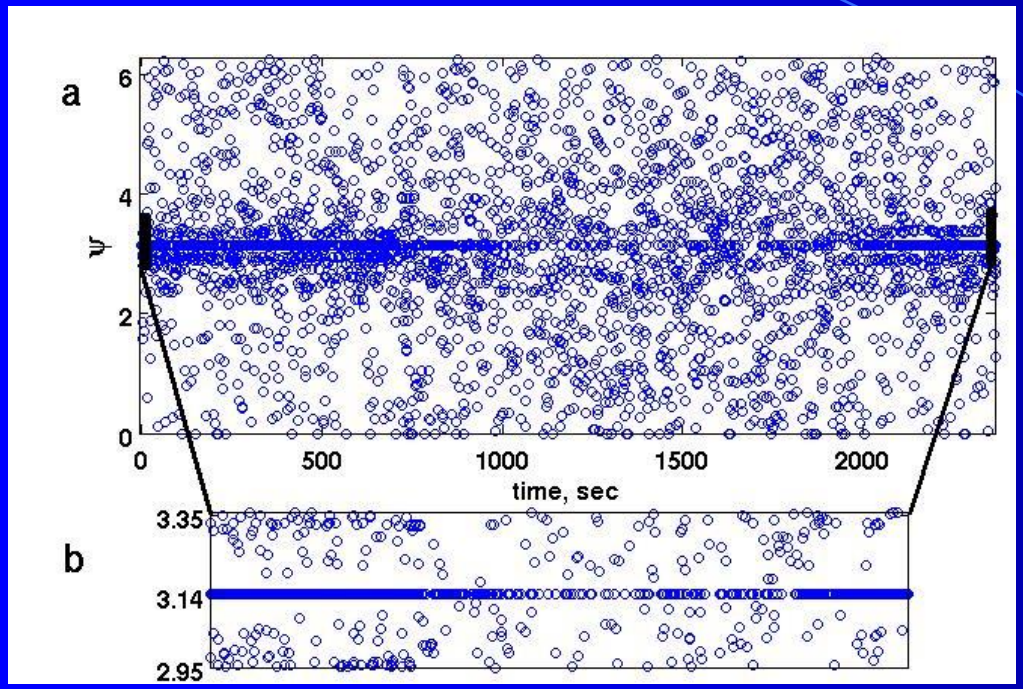


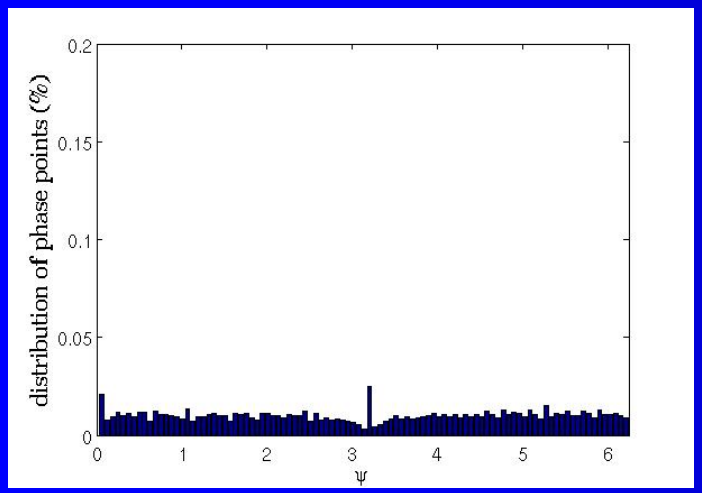


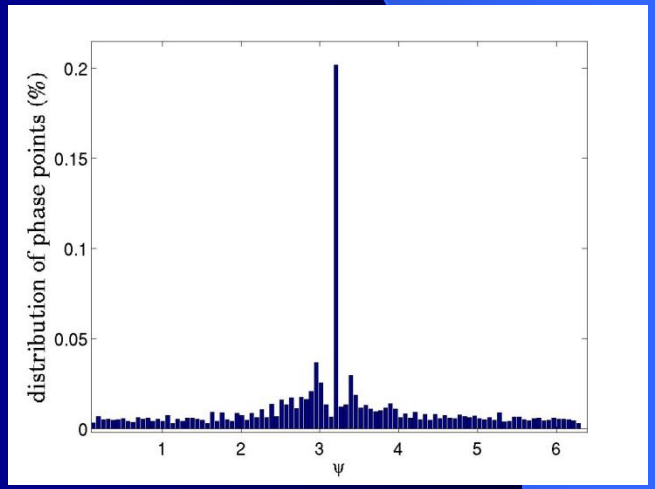


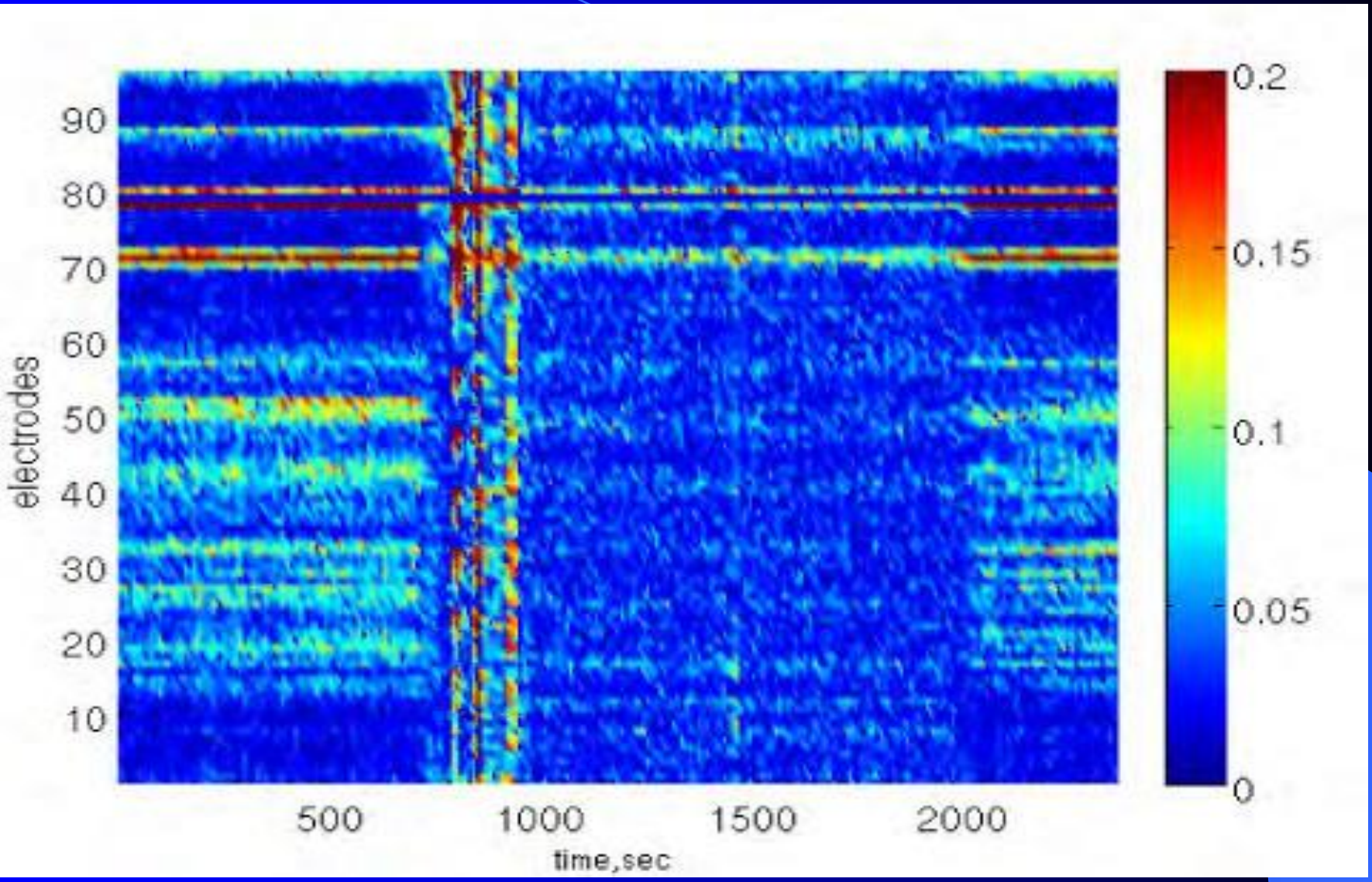
$\varphi(t) = 2\pi k + 2\pi\,(t-t_k)/(t_{k+1}-t_k)$ **Phase of the signal**

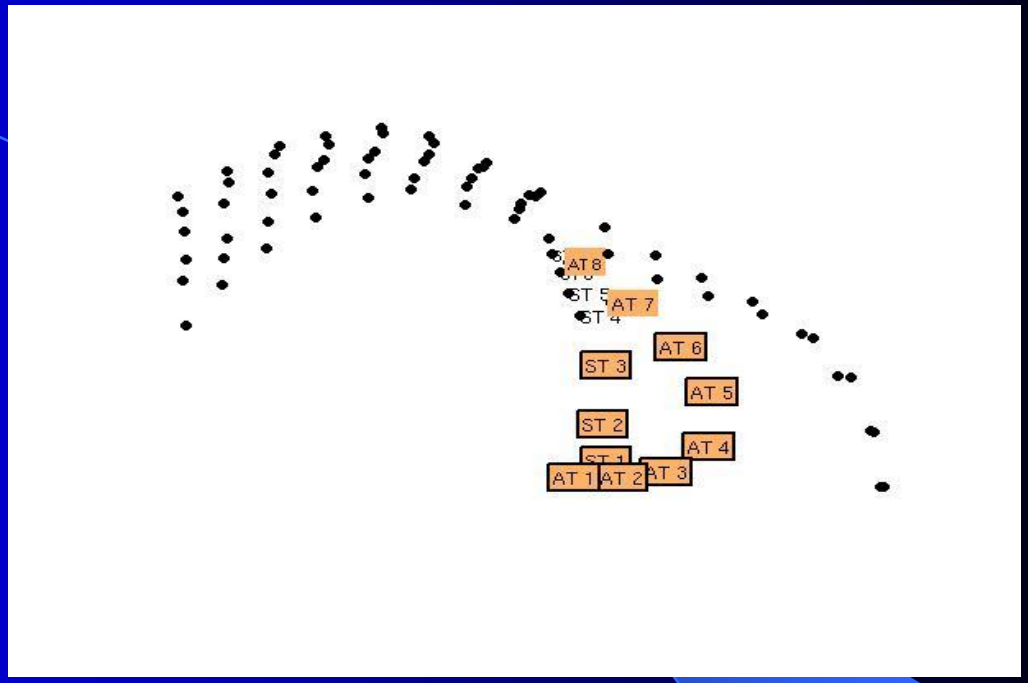
Synchrogram (stroboscopic view)

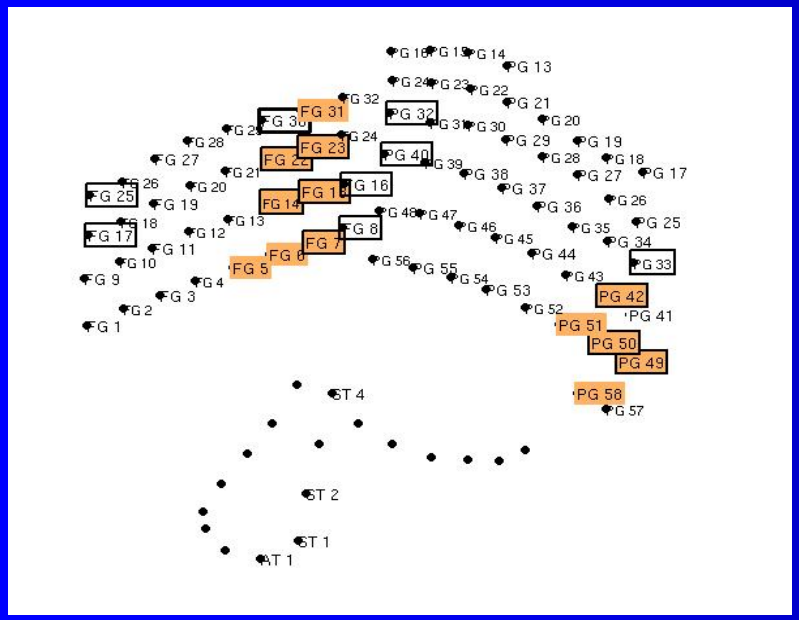


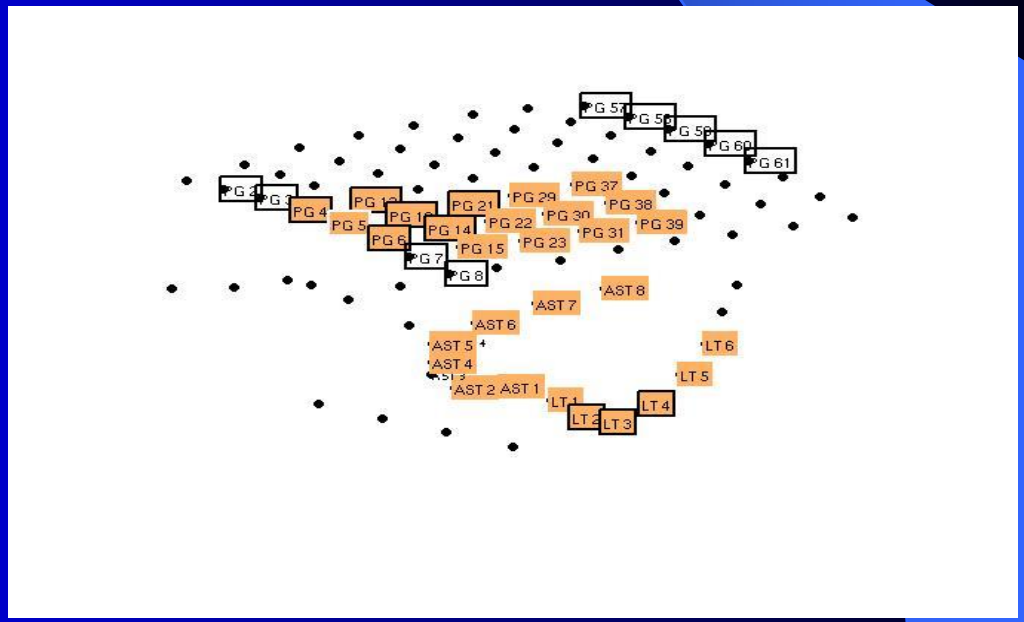





electrodes
90
80
70
60
50
40
30
20
10
500
1000
1500
2000
time,sec
0.2
0.15
0.1
0.05
0


AT 8
ST 5
AT 7
ST 4
AT 6
ST 3
AT 5
ST 2
AT 4
ST 1
AT 1
AT 2
AT 3


FG 31
FG 32
FG 23
FG 40
FG 16
FG 14
FG 7
FG 8
FG 5
PG 42
PG 51
PG 50
PG 49
PG 58
ST 4
ST 2
ST 1
AT 1


PG 37
PG 38
PG 39
PG 31
PG 23
PG 15
PG 8
AST 8
AST 7
AST 6
AST 5
AST 4
AST 2
AST 1
LT 6
LT 5
LT 4
LT 3

BUT……

There are thousands and thousands of relevant cases where either data are the static expressions of sets of variables/features…..

BLOOD TESTS

GENETIC EXPRESSIONS

METABOLIC DATA

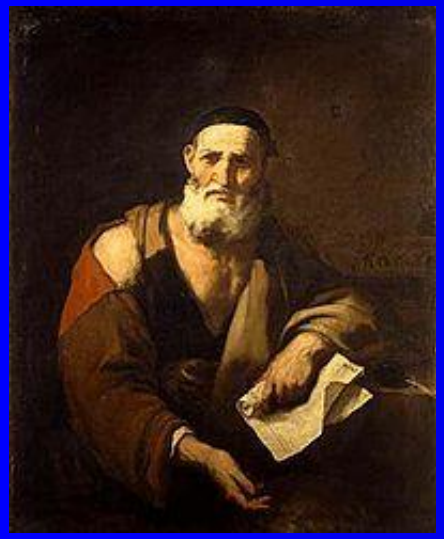
…..or the time resolution is not satisfactory

fMRI

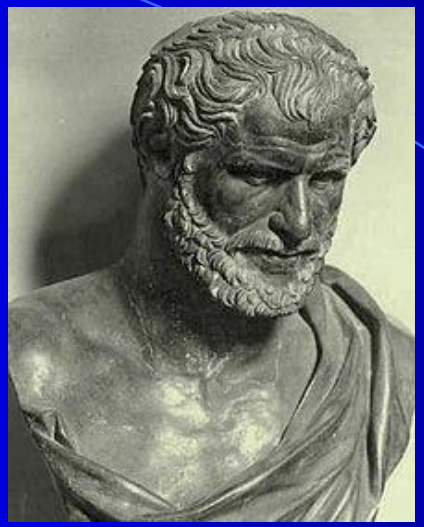
Evolution of National economies…

How can one obtain a proper network representation of such datasets? And…., what is good for?

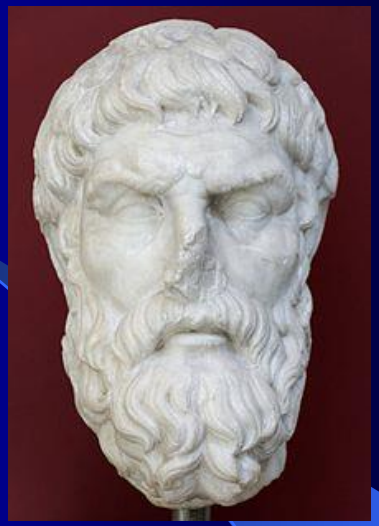
Parenclitic Networks….



Leukippos
(Mileto, V century a.C.)



Democritus
(Abdera, 460-360 a.C.)



Epicurus
(Samo 341-Athens 271 a.C)

Titus Lucretius Carus
(Pompei 94 –Roma 50 a.C)

Περὶ φύσεως

De rerum natura

Parenclisis (*παρέγκλισις*), Clinamen, and the ¨unpredictable swerve¨....

According to Lucretius:

¨When atoms move straight down through the void by their own weight, they deflect a bit in space at a quite uncertain time and in uncertain places, just enough that you could say that their motion has changed. But if they were not in the habit of swerving, they would all fall straight down through the depths of the void, like drops of rain, and no collision would occur, nor would any blow be produced among the atoms. In that case, nature would never have produced anything.¨

This indeterminacy provides the "free will which living things throughout the world have ¨

A set of *n* systems, or subjects $\{s_1, s_2,\ldots,s_n\}$, each one associated to one of *m* classes $\{c_1, c_2,\ldots,c_m\}$

Example: each system is a person, classified as *healthy* (or *control*)

In turn, each system *i* is identified by a vector of *p* features $f^i = (f_1^i, f_2^i,\ldots, f_p^i)$, i.e. a point in a *p*-dimensional space.

The **fundamental ansatz** is that each class corresponds to a ***set of constraints*** in the feature space. **In general, there will be *m* different set of constraints, one for each of the *m* classes, defining a MULTILAYER network.**

1) **Parenclitic network representation** is based on all possible bi-dimensional projections of the data.

2) For each pairs of features (*i*, *j=1,…,p)*, the values corresponding to subjects of a given class *C* are used to estimate the constraint

$$F_{ij}(f_i, f_j)=0.$$

All projected constraints (local models) can be obtained in several ways: **polynomial fit**, data mining methods as **Support Vector Machine** or **Artificial Neural Networks**, etc...

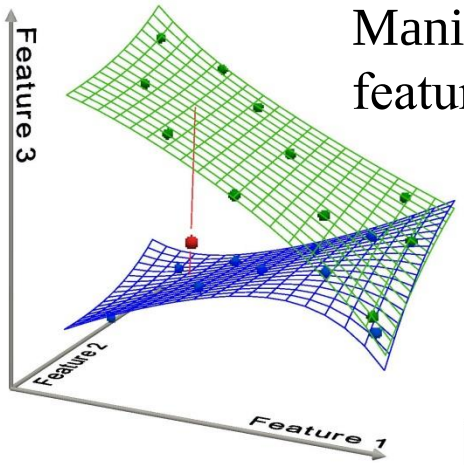
3) A unlabeled subject is then characterized by its position in such plane, and by its distance from the derived model. The link between nodes *i* and *j* *quantifies the deviation (parenclisis)* , in the *i-j* space, between the subject under analysis and the calculated projection function.

Parenclitic network reconstruction

Original data-set

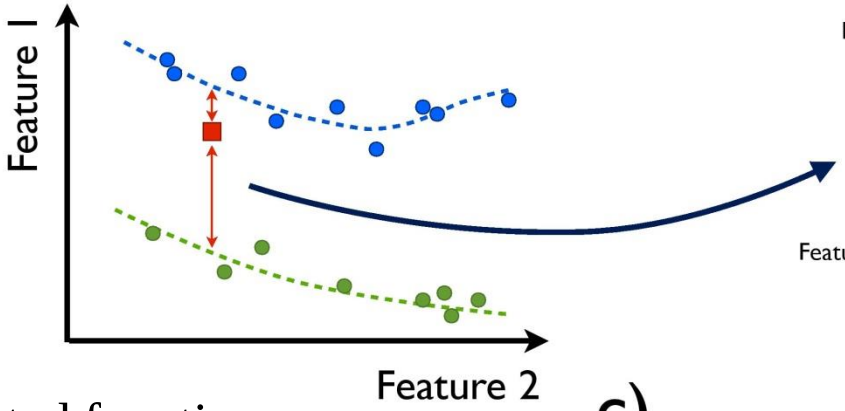
	Subject 1	Subject 2	Subject 3	...
Label	Healthy	Healthy	Healthy	...
Feature 1	0.435	0.532	0.635	...
Feature 2	0.332	0.827	0.518	...
Feature 3	0.342	0.927	0.381	...
...	...	...	...	...

a)



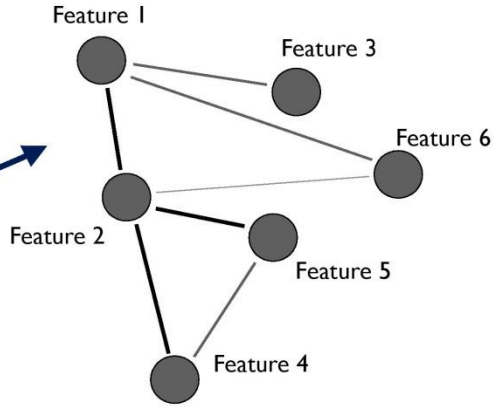
Manifolds in the full features' space

b)



Projected functions

c)

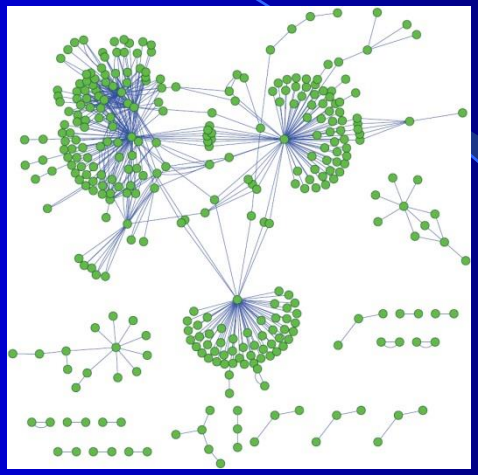


Parenclitic network

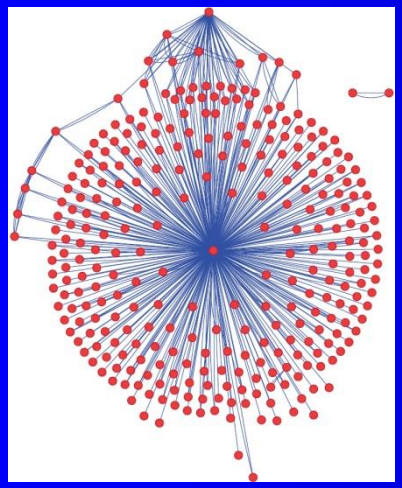
d)

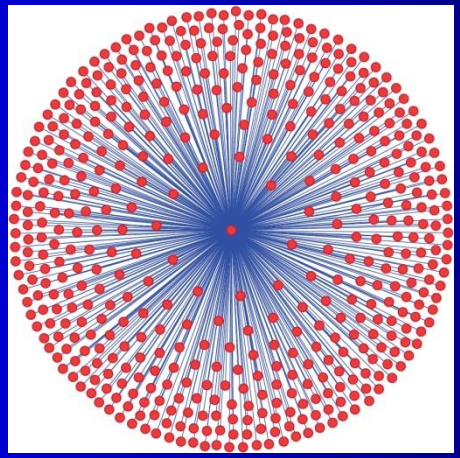
Early diagnosis of obstructive nephropathy (ON)

Control



ON



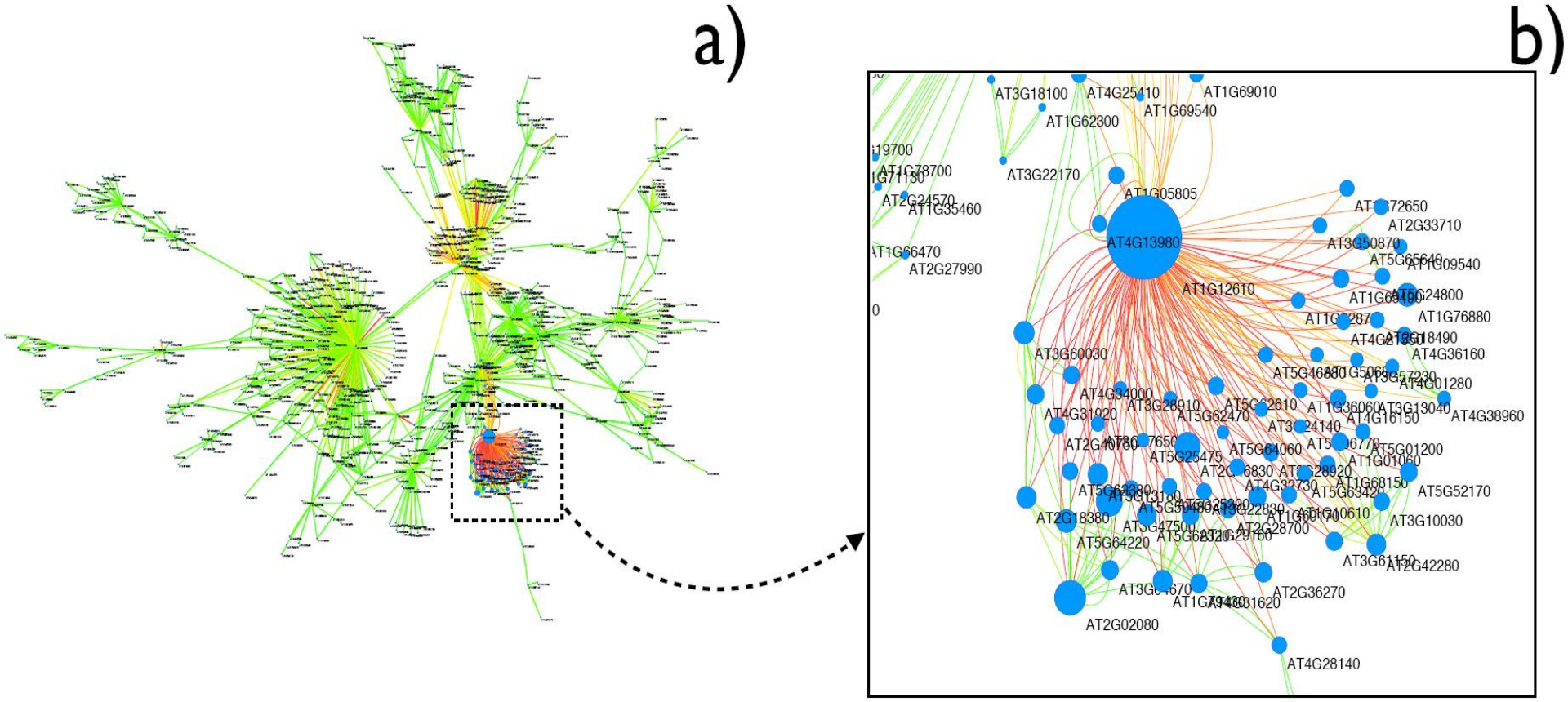


852 metabotites and 834 regulators (microRNAs) from 10 ON patients and a 10-member control group. For each metabolite (or regulator) a parenclitic network is constructed embodying how far every other metabolite (or regulator) deviated from its statistically expected correlation. Control group networks had amorphous topologies, whereas ON networks had star-like topologies centered on the most abnormal metabolite or regulator.

See: Chaos 21, 033103 (2011), and Physics Today 64, 20 (2011)

Expression levels of genes of the *Arabidopsis thaliana* under osmotic stress

Data consist of expression levels of 22,591 gene loci of the plant *Arabidopsis thaliana*, measured at different times (5, 10, 60, 120, 240 minutes) after the exposure of the plant to a osmotic stress.



Ranking	Gene	Centrality	Ranking	Gene	Centrality
5 minutes			1 hour		
1	AT5G04340	1	1	AT5G18090	1
2	AT2G17040	0.892558	2	AT5G16560	0.956474
3	AT1G13300	0.88111	3	AT4G14410	0.896794
4	AT5G51910	0.729679	4	AT4G36930	0.857808
5	AT1G07520	0.7176	5	AT2G23290	0.785677
10 minutes			2 hour		
1	AT1G44830	1	1	AT3G49760	1
2	AT1G43950	0.556858	2	AT5G07210	0.302456
3	AT5G60470	0.443326	3	AT4G38340	0.300498
4	AT2G22300	0.356115	4	AT4G09820	0.274155
5	AT3G59470	0.304019	5	AT3G61910	0.264721
30 minutes			6 hour		
1	AT1G12610	1	1	AT2G28450	1
2	AT2G02080	0.35273	2	AT2G35430	0.903528
3	AT3G47500	0.281975	3	AT1G09540	0.709785
4	AT2G46830	0.271497	4	AT1G10170	0.665937
5	AT1G76880	0.201452	5	AT5G35550	0.648662

Table 1: **Most central genes at different time steps.**

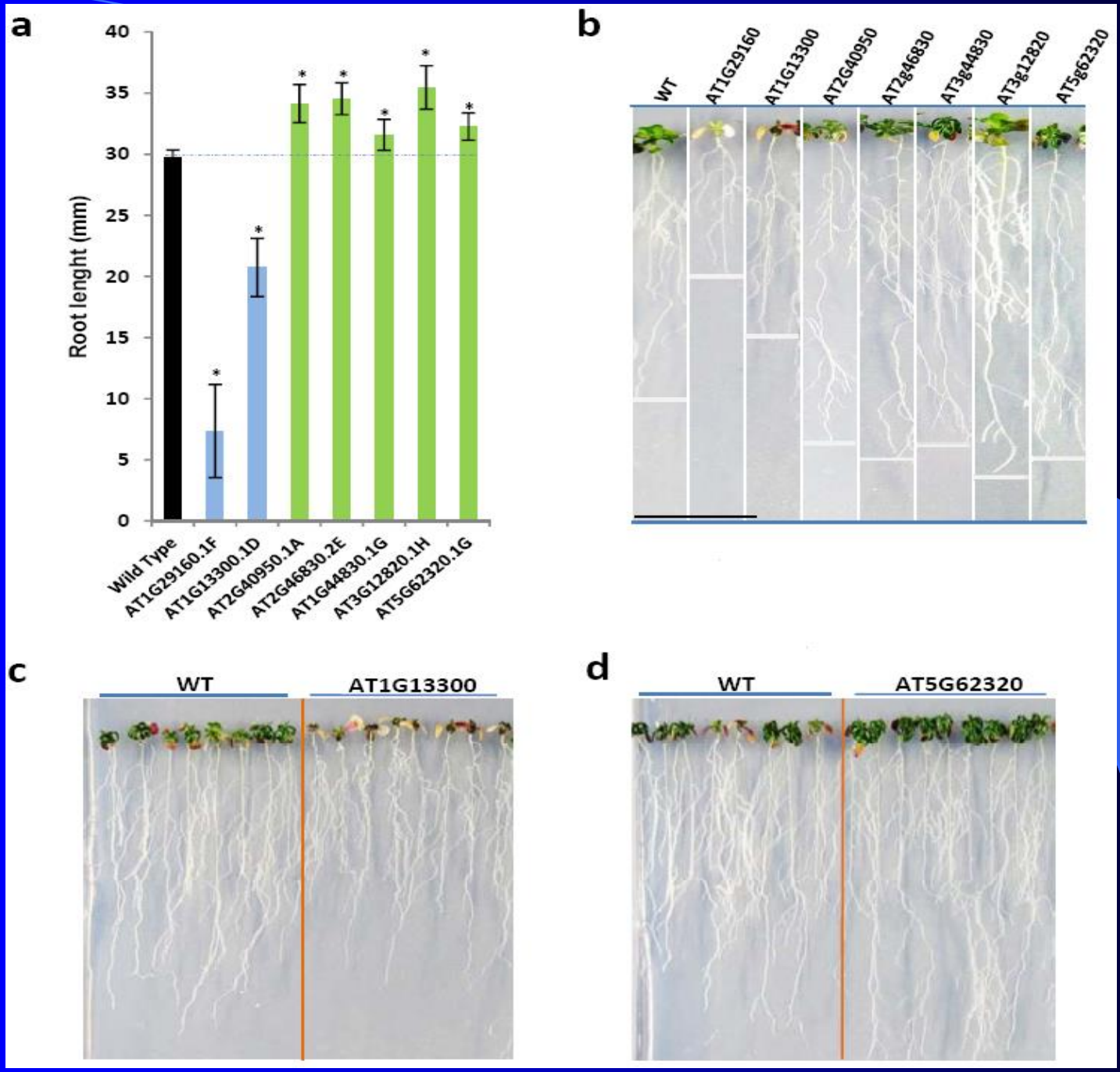
Aim: identification of relevant nodes (i.e., genes)

The more central is a node, the more connections with high weight it has, and therefore it frequently shows a behavior different from the expected one.

We used *α-centrality*

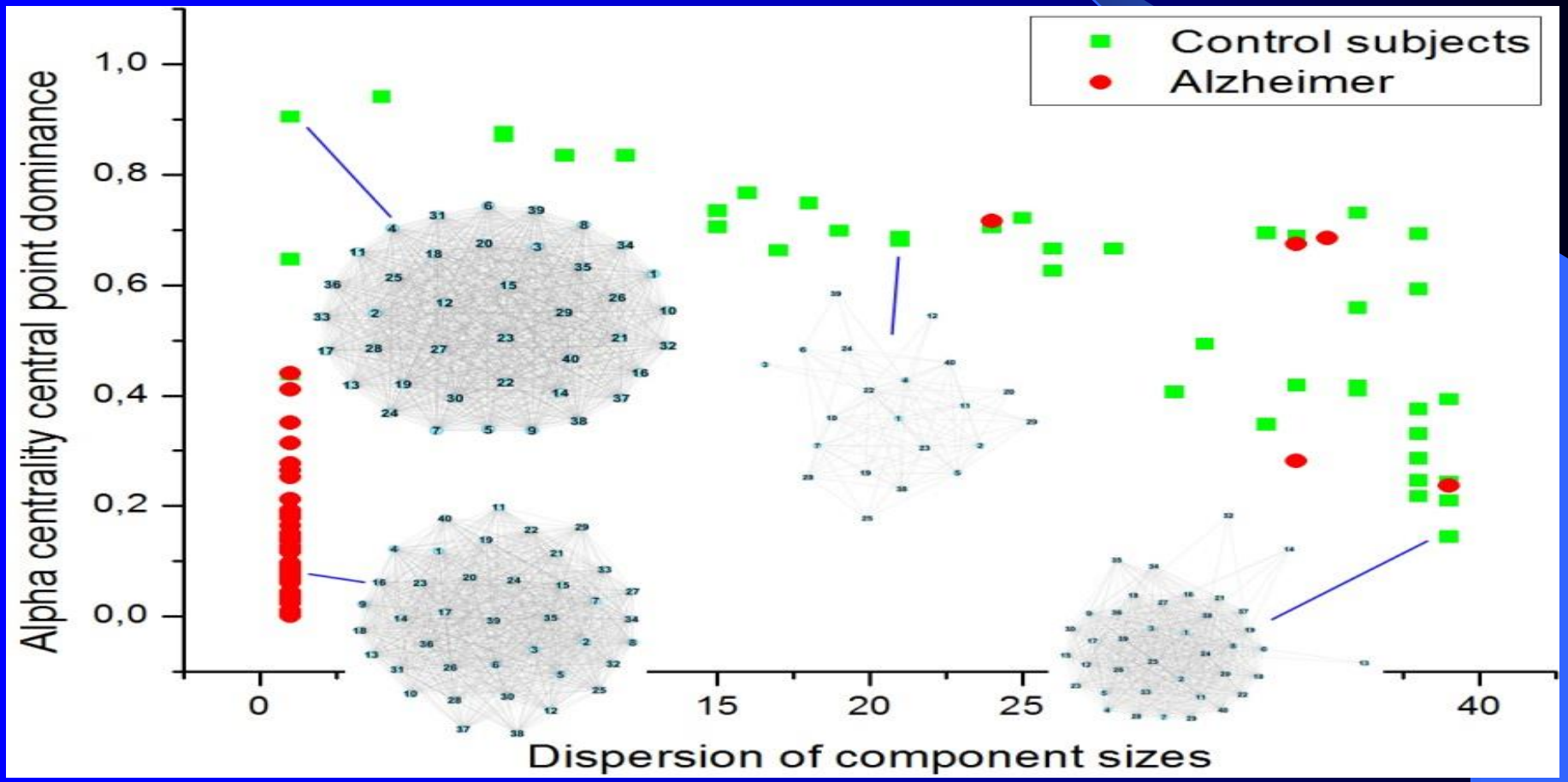
If *xi* is the centrality of the *i*-th node, then

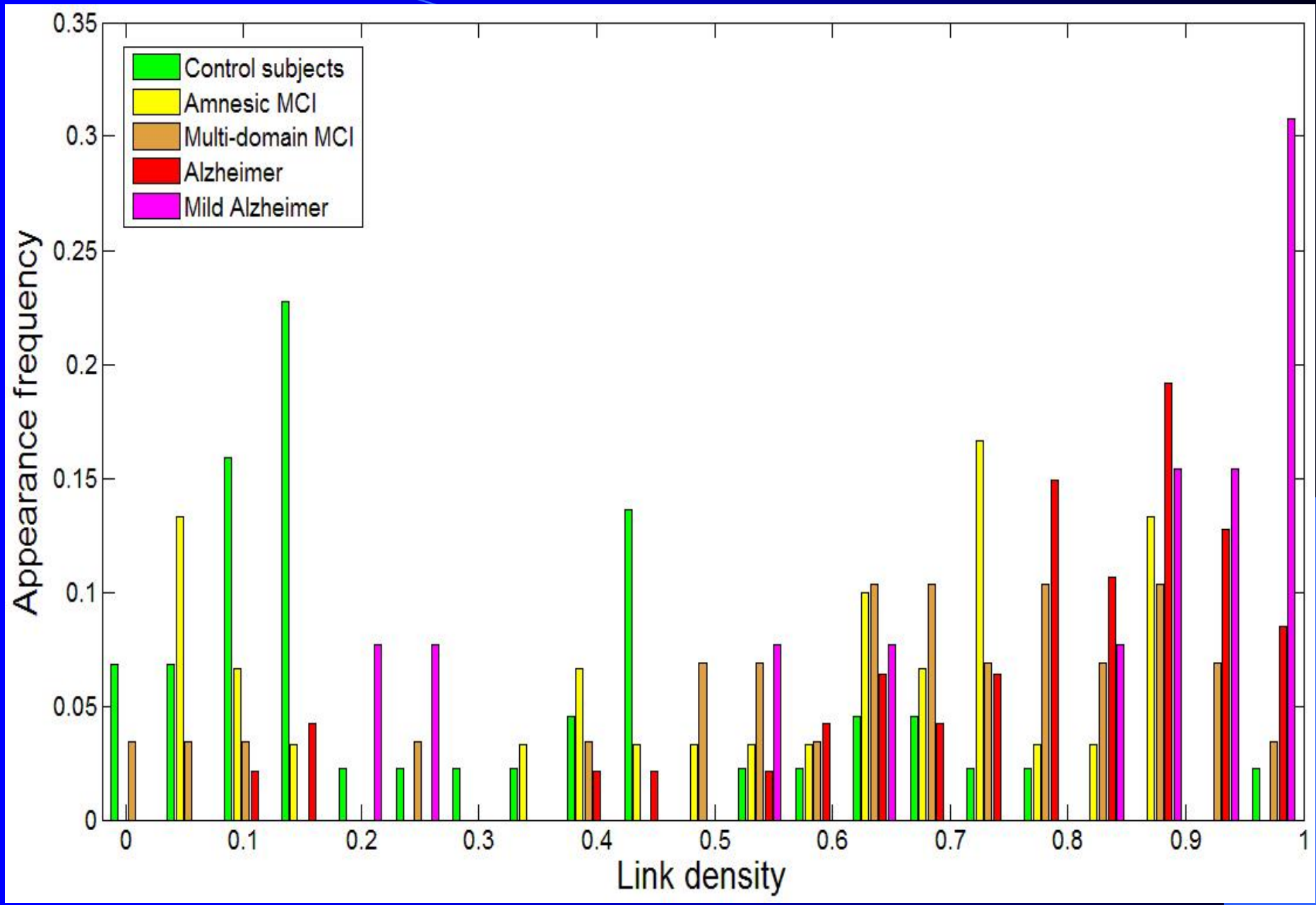
$$x_i = \sum_j x_j (W_{ij} + \alpha)$$


a
Root lenght (mm)
0
5
10
15
20
25
30
35
40
Wild Type
AT1G29160.1F
AT1G13300.1D
AT2G40950.1A
AT2G46830.2E
AT1G44830.1G
AT3G12820.1H
AT5G62320.1G
b
WT
AT1G29160
AT1G13300
AT2G40950
AT2g46830
AT3g44830
AT3g12820
AT5g62320
c
WT
AT1G13300
d
WT
AT5G62320

Application to fMRI data

Data consist of fMRI recordings of the temporal average of the oxygen consumption (3 variables) of 12 subcortical regions. **The total number of nodes is 36**. Five types of subjects are considered (representing different stages of the Alzheimer's disease progression): **control subjects** (44), **Amnesic MCI** (30), **Multidomain** MCI (29), **Mild Alzheimer's** (13) and **normal Alzheimer's** (48) patients.




Control subjects
Amnesic MCI
Multi-domain MCI
Alzheimer
Mild Alzheimer
Appearance frequency
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Link density

Conclusions

1) We introduced a novel network representation of a set of pre-labeled subjects that unveils the presence of reference relationships between nodes. When the data corresponding to a subject deviates from this reference situation, links are created, in a way that the resulting topology quantifies the distance of that subject from the initially defined class(es).

In analogy to the spirit of the doctrine from Democritus and Epicurus, we call it the *parenclitic network representation* of a system.

2) Potential applications include tissues and organic sample analysis, like blood analysis or spectrography; genetic expression levels of individuals, without evolution through time; biomedical analysis like neuro-imaging techniques; or social network analysis, when just a snapshot of the system evolution is available.