Dottorato Internazionale in Fisica Applicata
Nonlinear Stochastic Dynamics in Complex Systems

Dipartimento di Fisica e Tecnologie Relative - Università degli Studi di Palermo
Mercoledì 14 marzo 2007

DYNAMICAL CLUSTERING:
The role of Synchronization for detecting Community Structures in Complex Networks

Alessandro Pluchino*

in collaboration with A. Rapisarda, V. Latora*, M. Ivanchenko** and S. Boccaletti****

CACTUS Chaos And Complexity Theoretical University Study Group Catania

* Dipartimento di Fisica e Astronomia and INFN sezione di Catania University of Catania, Italy - Group web page: www.ct.infn.it/~cactus

** Moscow University

*** Nat.Ist. Applied Optycs - Florence

Outline

The Problem:
Finding Community Structures in Complex networks

The Approach:
Synchronization of Dynamical Oscillators in Weighted Networks

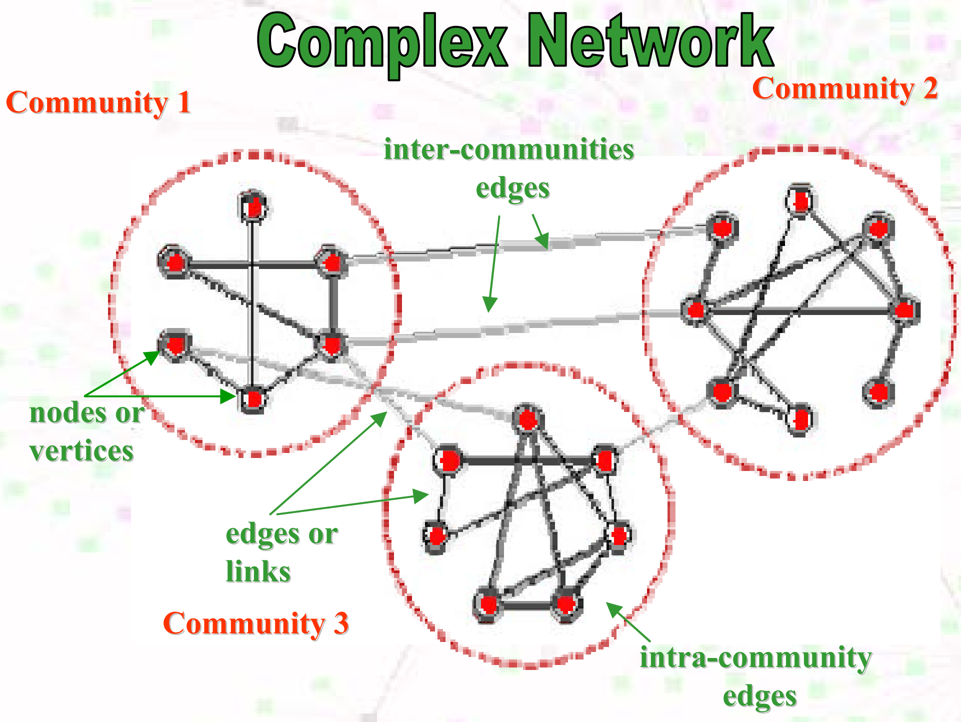
The Solution:

Dynamical Clustering algorithm for the identification of Community Structures in Trial and Real Networks

Discussion and Numerical Results

The problem: Finding Community Structures in Complex Networks

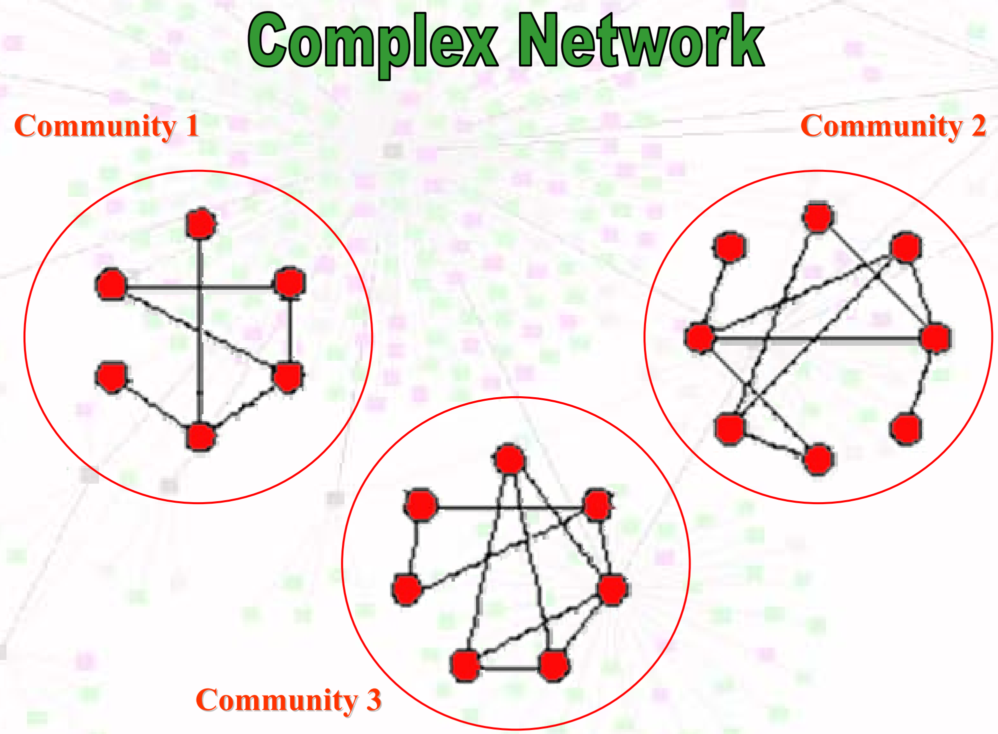
An important open problem in complex networks analysis is the identification of modular structures.



Distinct modular structures, usually called Communities, can loosely be defined as subset of nodes (vertices) which are more densely linked, when compared to the rest of the network.

The problem: Finding Community Structures in Complex Networks

In a limiting case, communities can be also defined as non-connected clusters of interconnected nodes:



Communities, of course, are fundamental in social networks (parties, cultures, elites), but are also very important in biochemical, metabolic or neural networks (functional groups), in the world wide web (thematic pages), in economic networks, food webs, computer clusters and so on…

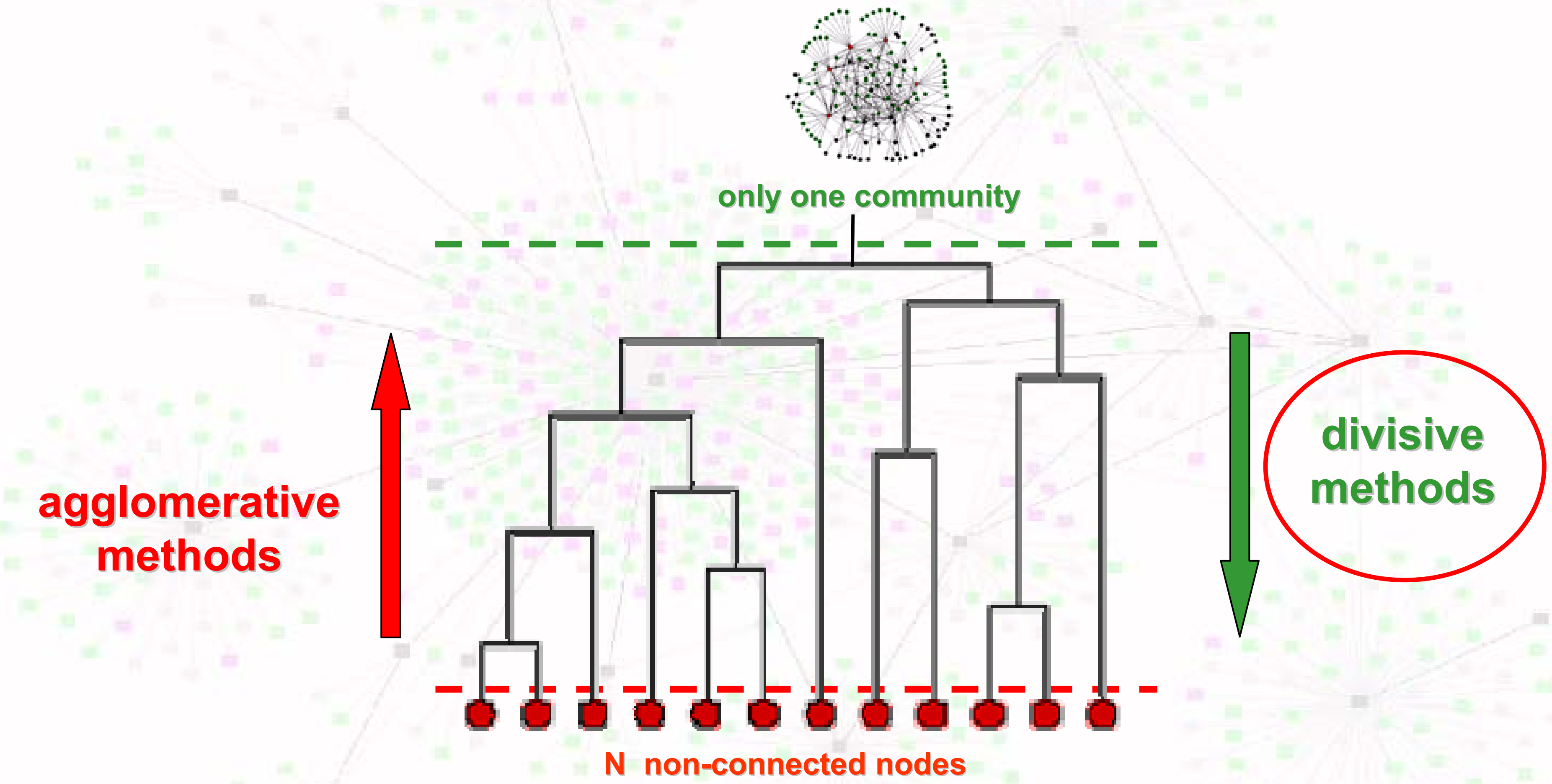
A useful set of techniques for the detection of community structures was firstly developed in social network analysis and is known as

HIERARCHICAL CLUSTERING METHODS.

These techniques are aimed at discovering natural divisions of networks into groups, based on various metric of similarity or strength of connection between vertices.

They fall into two broad classes: **agglomerative** and **divisive** methods, depending on whether they focus on the addition or the removal of edges to or from the network…

HIERARCHICAL TREE (DENDROGRAM)



Divisive methods progressively remove the edges of the networks in terms of their importance:

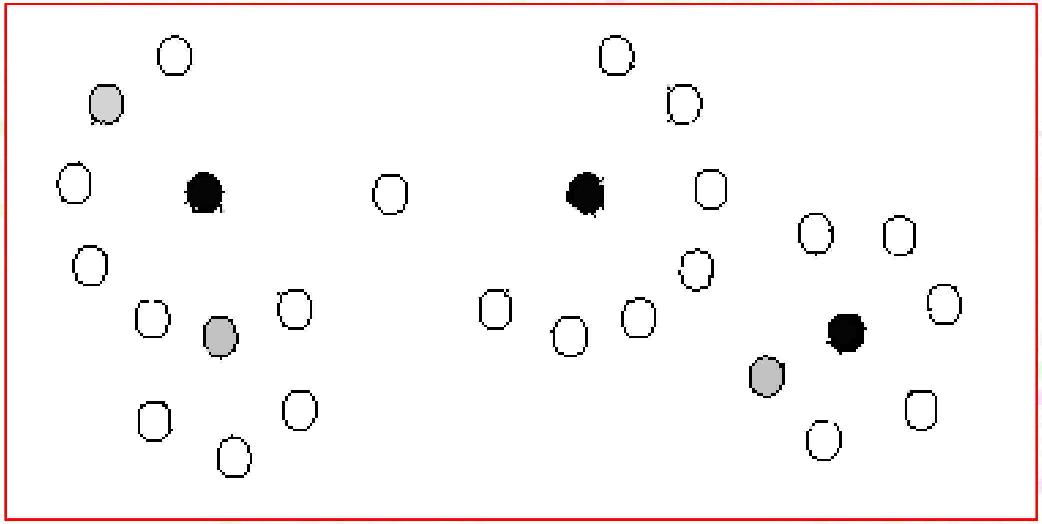
- in propagating some information over the network (*information centrality*)

S.Fortunato, V.Latora, M.Marchiori, 2004 *Phys. Rev. E* **70** 056104

- in connecting many pairs of nodes (*shortest-path edge betweenness*, i.e. the number of shortest paths which are making use of a given edge)

M.E.J.Newman and M.Girvan, 2004 *Phys. Rev. E* **69** 026113

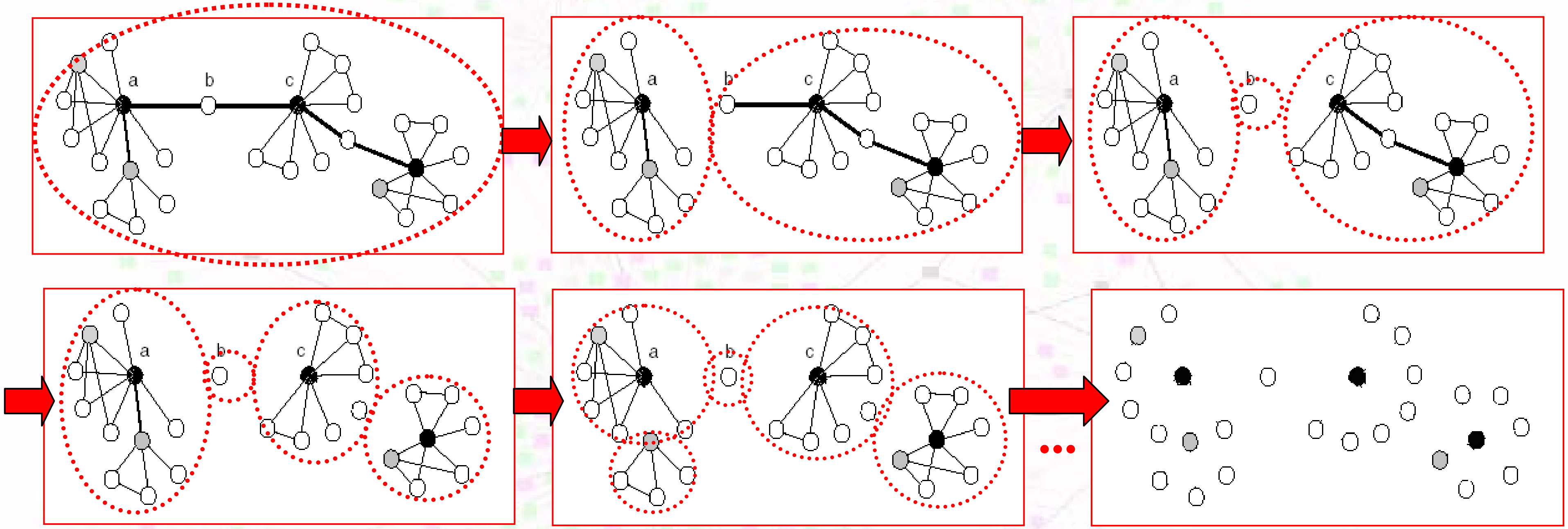
- other quantities...



By doing this repeatedly, and recalculating the betweenness at each step, the network breaks iteratively into smaller and smaller groups of nodes…

…until it breaks into a collection of single non-connected nodes...

The divisive algorithm produces a hierarchy of subdivisions of the network in isolated groups of interconnected nodes (communities)...



But which subdivision level does give the best communities configuration for a given network?

Clearly we need some parameter to quantify the reliability of each communities configuration…

This parameter is the "modularity" **Q** *, a quantity that, at each step, compares the fraction of intra-community edges with the expected value of the same quantity in an equivalent network with random connections, and allows us to test which communities configuration found by the divisive algorithm is the best one:

modularity

$$\mathrm{Q} = \sum_{i=1}^{n_c} (\mathrm{e}_{\mathrm{ii}} - b_i^2)$$

fraction of edges that connect vertices in community i

fraction of edges that connect vertices in community i for a random network

n_c is the number of communities

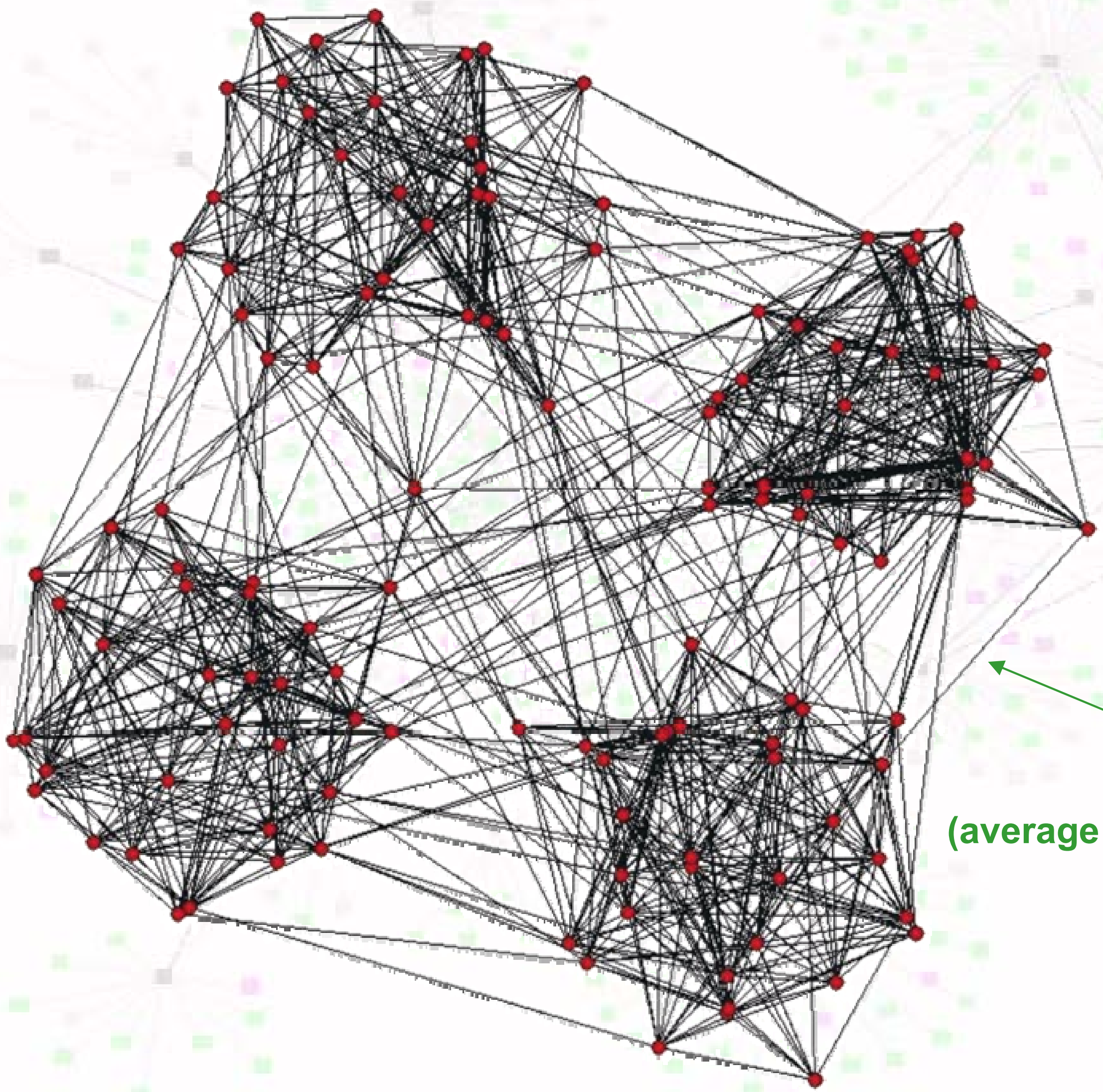
||e|| is a n_c x n_c matrix whose elements e_{ij} represent the fraction of total edges connecting a node in community-i with a node in community-j

$b_i = \sum_j e_{ij}$ represents the fraction of total edges connected to a node in community-i

Usually 0.3 < Q < 0.7

*M.E.J.Newman and M.Girvan, 2004 *Phys. Rev. E* **69** 026113

Modularity in “ad hoc” random trial networks (N=128, <k>=16, 4 communities)

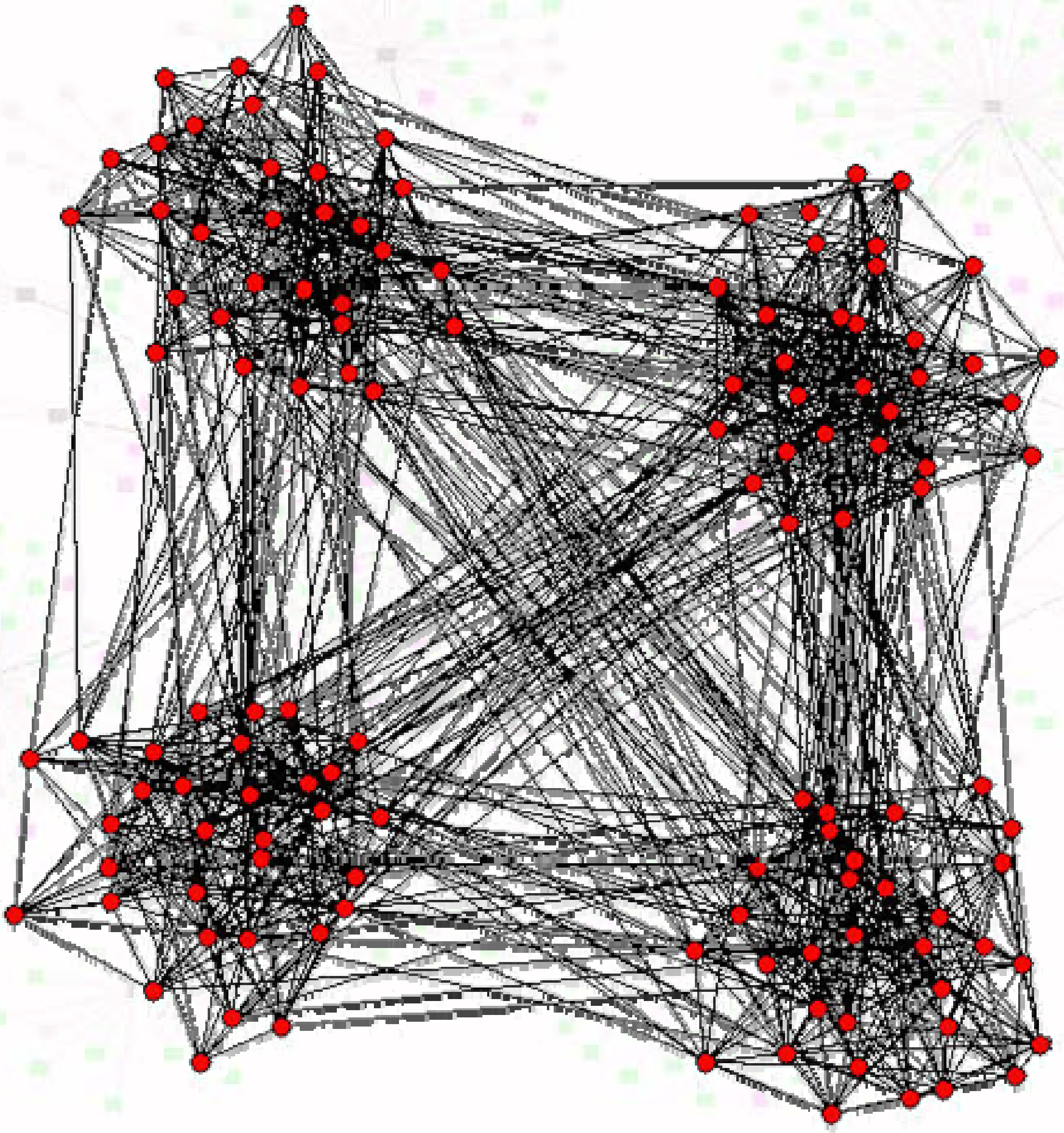


zout = 2
Q ~ 0.7

zout

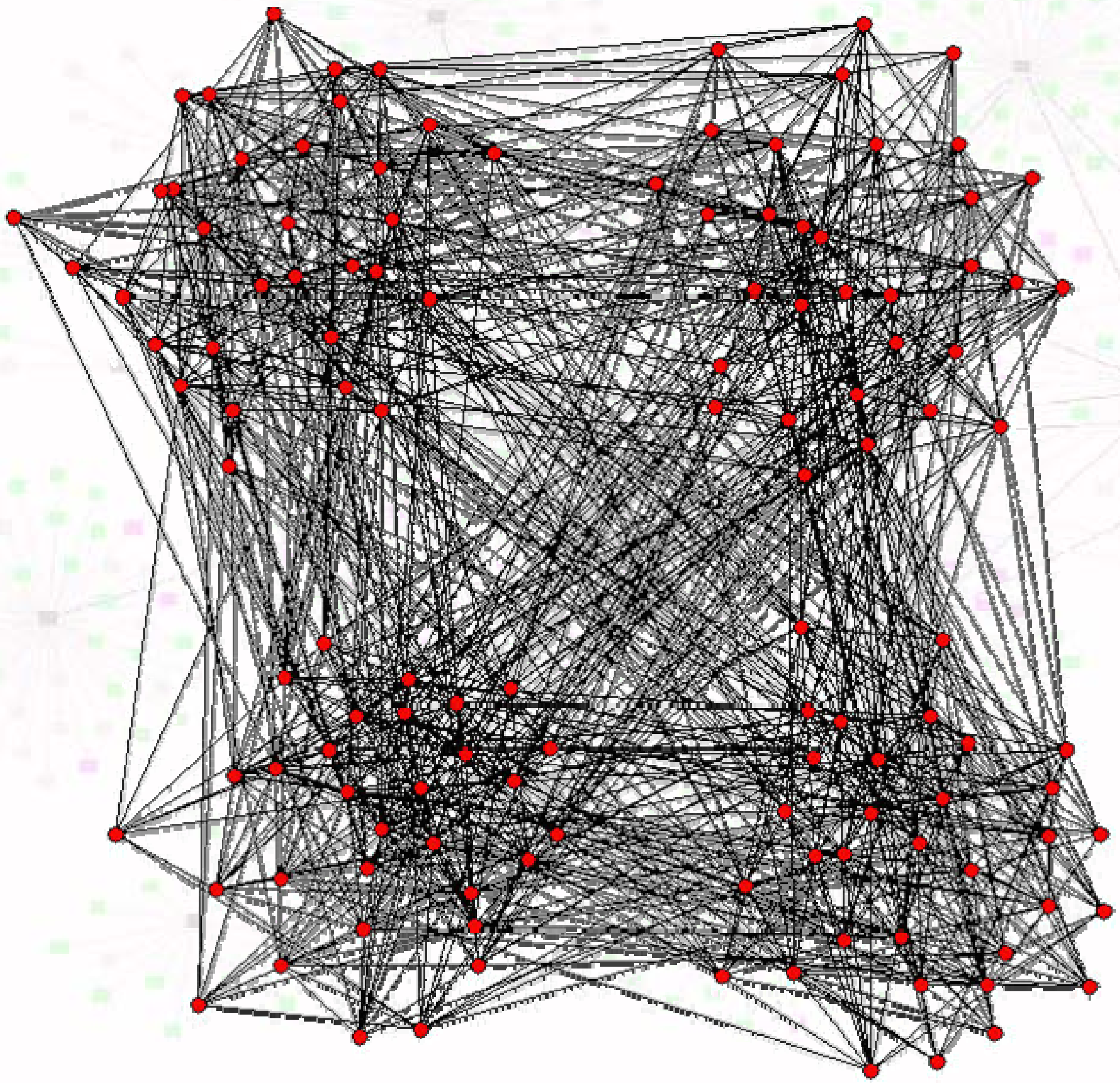
(average number of inter-community edges per node)

Modularity in "ad hoc" random trial networks
(N=128, <k>=16, 4 communities)



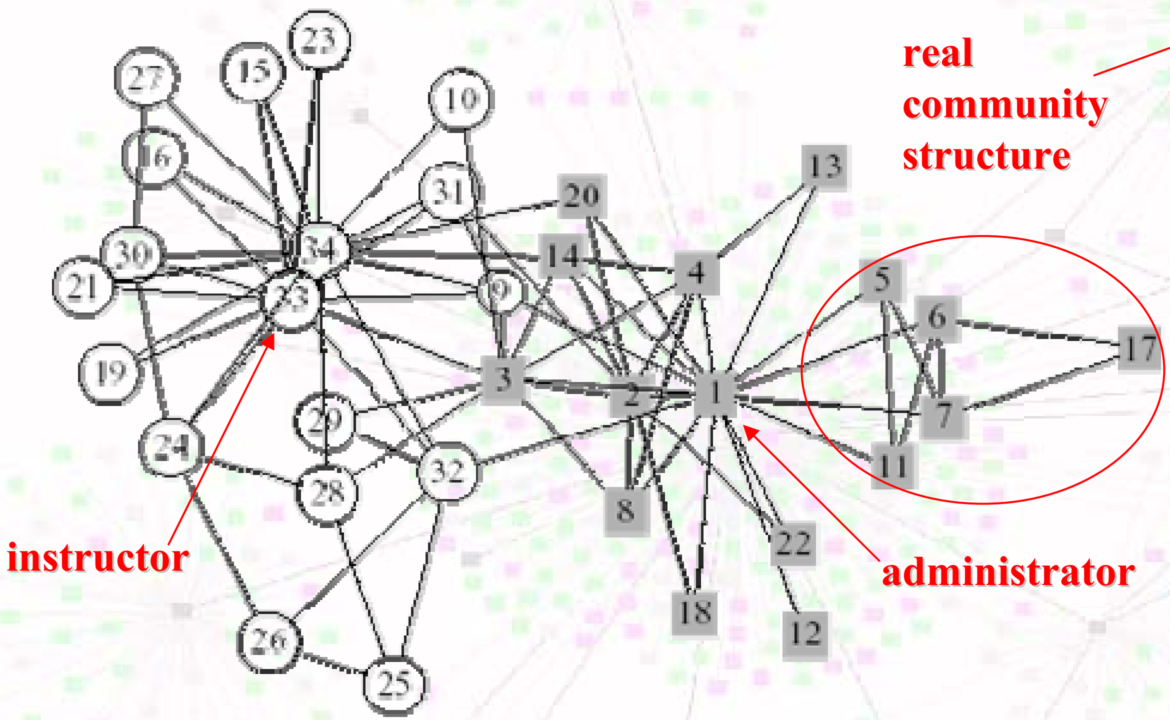
zout = 4
Q ~ 0.6

Modularity in “ad hoc” random trial networks
(N=128, <k>=16, 4 communities)

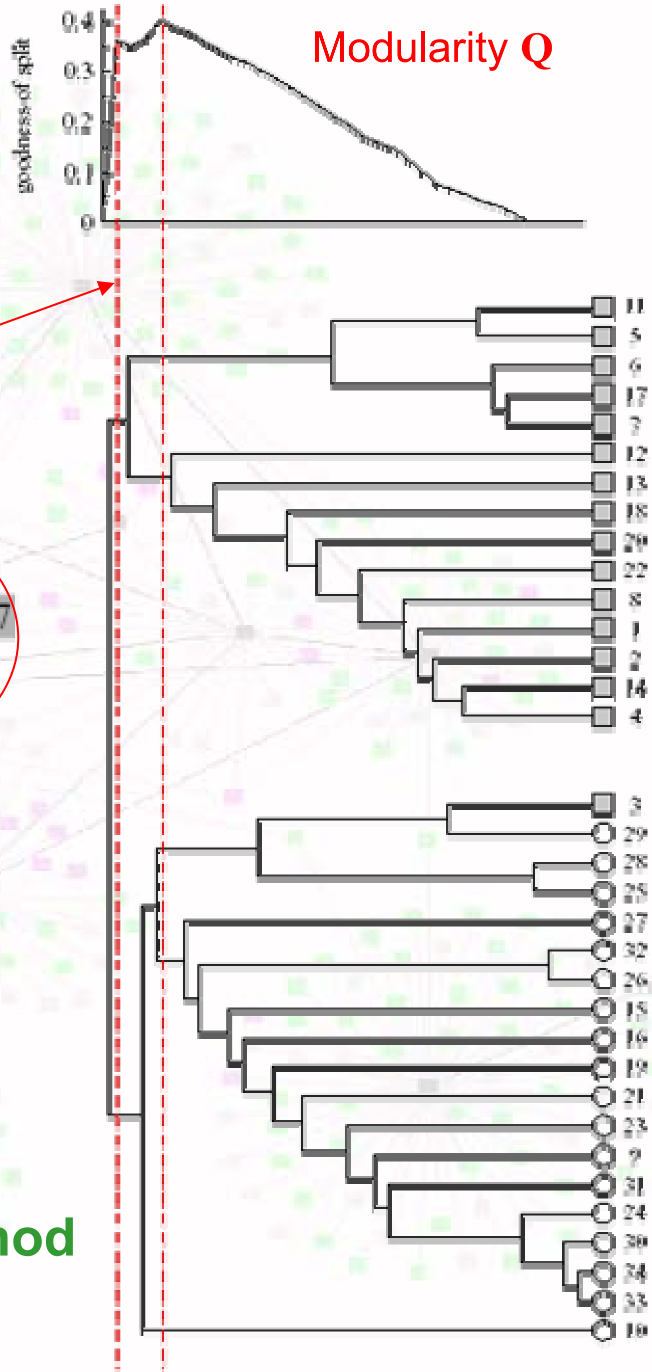


zout = 6
Q ~ 0.5

Zachary's Karate Club friendships network



Community 2 (18 nodes) **Community 1 (16 nodes)**



Girvan Newman
Shortest-path edge-betweenness divisive method

M.E.J.Newman and M.Girvan, 2004 *Phys. Rev. E* **69** 026113

W.Zachary (1977) *J.Anthropol.Res.* **33** 452-473

Topological Divisive Algorithms like GN have the problem of recalculating betweennesses at each step.

Since a single-step calculus of all the edge-betweennesses takes $O(N^2)$ operations, and the whole process takes N steps, these algorithms are quite slow – $O(N^3)$ –

DIFFERENT APPROACH:
Synchronization of Dynamical Oscillators in Weighted Networks

The Kuramoto model*

The simplest models for synchronization available on the market is the celebrate Kuramoto model and consists of N fully connected phase oscillators with natural frequencies ω_i and coupling parameter K:

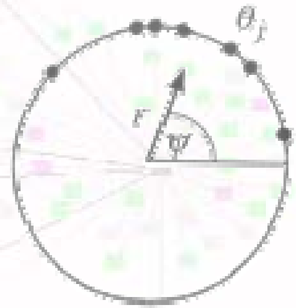
coupling strenght

$$\frac{d\vartheta_i(t)}{dt} = \omega_i + \frac{K}{N}\sum_{j=1}^{N}\sin(\vartheta_j - \vartheta_i) \;, \quad i = 1,\ldots.N$$

$\vartheta_i(t) \in [0, 2\pi)$

natural frequencies

phases of oscillators



The coherence of the system is measured by the mean field order parameter r $(0 \leq r(t) \leq 1)$:

$$re^{i\psi} = \frac{1}{N}\sum_{j=1}^{N} e^{i\vartheta_j}$$

Through r the rate equations can be rewritten as:

$$\frac{d\vartheta_i(t)}{dt} = \omega_i + Kr\sin(\psi - \vartheta_i), \quad i = 1,\ldots.N$$

*proposed by Y.Kuramoto in 1975

The Kuramoto model (2)

As Kuramoto showed analitically in a beautiful analysis, one observes synchronization above a certain *critical value* of the control parameter K_c …

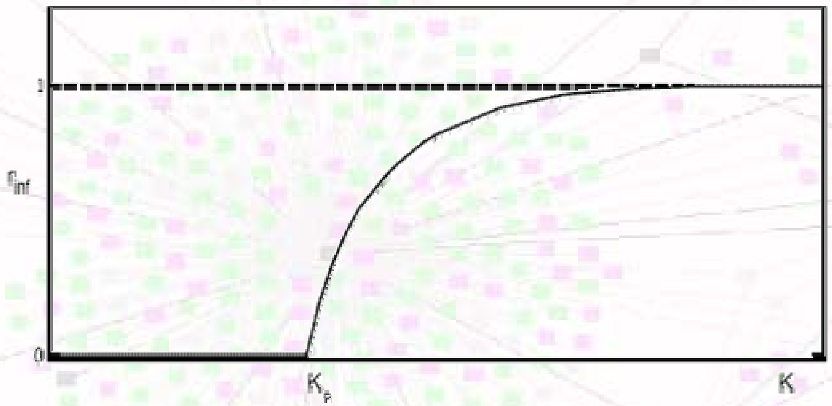
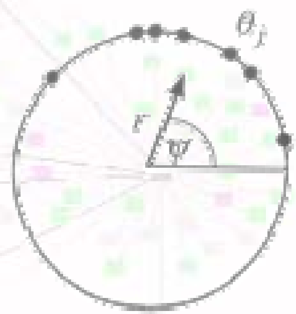


Fig. 1. Asymptotic order parameter r_∞ as a function of the coupling in the Kuramoto model

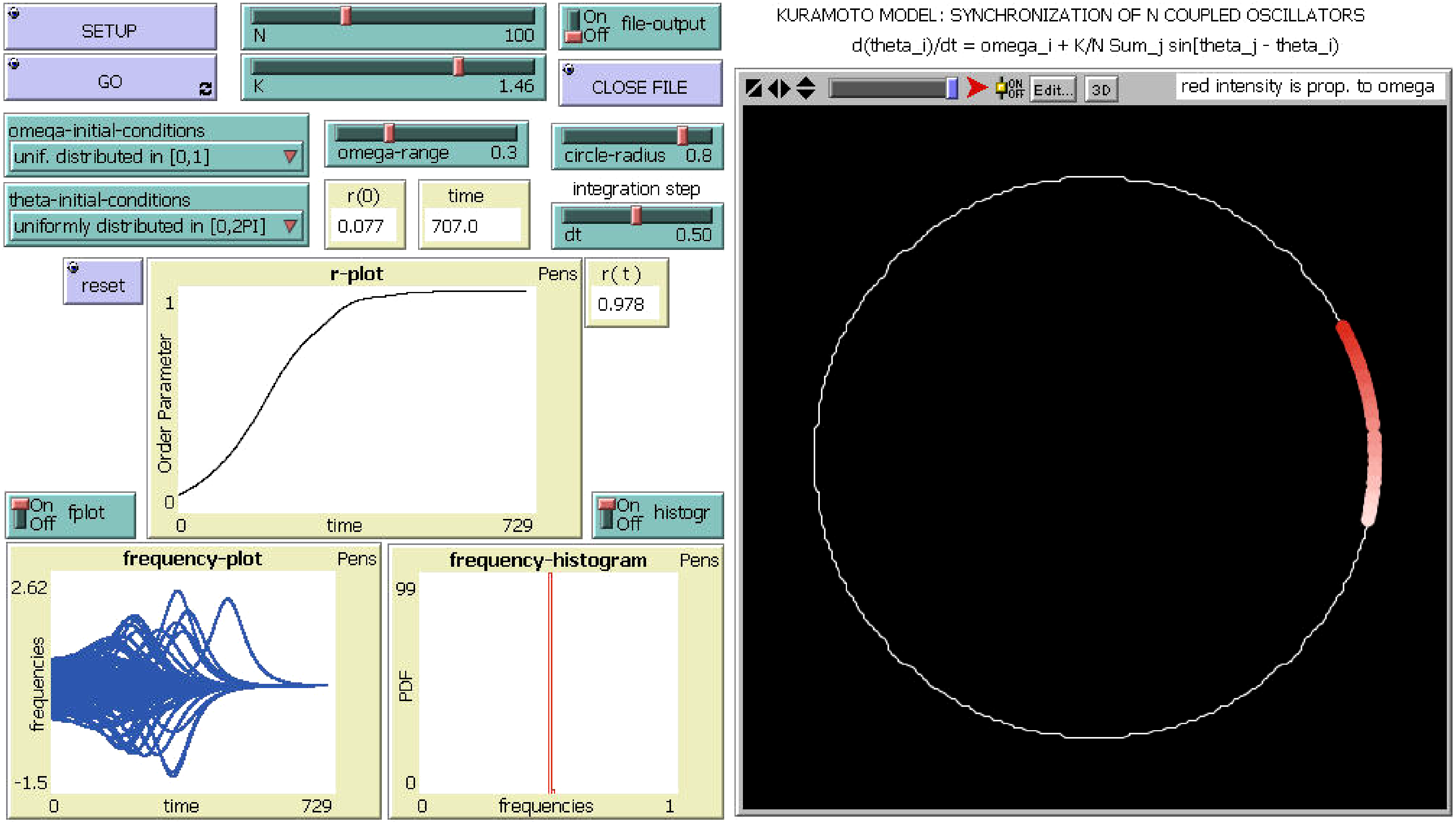


$K \to 0 \qquad \vartheta_i(t) \approx \omega_i t + \vartheta_i(0) \qquad r \to 0$ Incoherent phase

$K \to \infty \qquad \vartheta_i(t) \approx \psi(t) \qquad r \to 1$ Global synchronization

Global syncronization
- phase syncronization (phase locking)
- frequency syncronization (frequency locking)

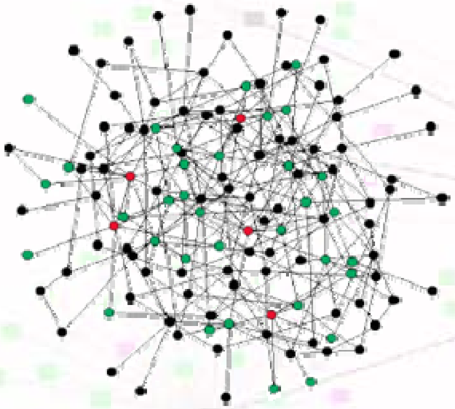
The Kuramoto model (3)



THE MASTER STABILITY FUNCTION APPROACH TO ENHANCE SYNCHRONIZATION IN COMPLEX NETWORKS

Suppose to have a (unweighted, undirected) network of N linearly coupled identical oscillators*. The equation of motion reads:

Network with N nodes



coupling strenght

$$\underbrace{\dot{\vec{x}}_i = \vec{F}(\dot{\vec{x}}_i)}_{\text{dynamical system defined over each node of the network}} - \sigma \sum_{j=1}^{N} G_{ij}\, \vec{H}[\vec{x}_i - \vec{x}_j] \qquad i = 1,\ldots,N$$

coupling matrix

coupling function

If G has a real spectrum of eigenvalues λ_i (for symmetric coupling) and if we associate λ_1 to the state $\mathbf{x}_s(t)$, the stability of the synchronous manifold ($\mathbf{x}_i(t)=\mathbf{x}_s(t)$, $\forall$ i) requires that all the conditional Lyapunov exponents Λ associated with $\lambda_2 \leq \ldots \leq \lambda_i \leq \ldots \leq \lambda_N$ would be negative...

*M.Chavez, D.U.Hwang, A.Amann, H.G.E.Hentschel and S.Boccaletti, *Phys. Rev. Lett.* **94** 218701 (2005)

THE MASTER STABILITY FUNCTION APPROACH TO ENHANCE SYNCHRONIZATION IN COMPLEX NETWORKS (2)

Network with N nodes

coupling strenght

$$\underbrace{\dot{\vec{x}}_i = \vec{F}(\dot{\vec{x}}_i)}_{\text{dynamical system defined over each node of the network}} - \sigma \sum_{j=1}^{N} G_{ij} \vec{H}[\vec{x}_i - \vec{x}_j] \qquad i = 1,\ldots,N$$

coupling matrix **coupling function**

Defining the Master Stability Function (MSF) as the largest Lyapunov exponent Λ_{max} versus a parameter $\nu=\sigma\lambda$, it can be shown* that, for a large class of dynamical systems, the MSF is negative in a finite parameter interval (ν_1, ν_2).

Thus the condition for synchronization stability is governed by the ratio λ_N/λ_2: the more packed the eigenvalues of G are, the higher is the chance of having all Lyapunov exponents into the stability range for some σ.

*M.Chavez, D.U.Hwang, A.Amann, H.G.E.Hentschel and S.Boccaletti, *Phys. Rev. Lett.* **94** 218701 (2005)

One can use the master stability function approach:

1) to find the best synchronization condition of a given network*
2) to tune the synchronization of a network in order to identify community structures** (DYNAMICAL CLUSTERING)

Both the results can be realized with an opportune choice of the coupling matrix G_{ij} in the network equation, by means of a weighting procedure that assignes to each edge a 'load' l_{ij} equal to its betweenness (i.e. the number of shortest paths that are making use of that edge):

coupling matrix G=G(α)

$$\dot{\vec{x}}_i = \vec{F}(\dot{\vec{x}}_i) - \sigma \sum_{j \in K_i} \frac{l_{ij}^{\ \alpha}}{\sum_{j \in K_i} l_{ij}^{\ \alpha}} \vec{H}[\vec{x}_i - \vec{x}_j] \qquad i = 1, \ldots, N$$

where α is a real tunable parameter and K_i is the set of neighbors of the i^{th} node.

master stability function arguments apply

*M.Chavez, D.U.Hwang, A.Amann, H.G.E.Hentschel and S.Boccaletti, *Phys. Rev. Lett.* **94** 218701 (2005)

** A.Pluchino., M.Ivanchenko, V.Latora, A.Rapisarda and S.Boccaletti, *in preparation*

Most common dynamical systems defined over the network

$$\dot{\vec{x}}_i = \vec{F}(\vec{x}_i) - \frac{\sigma}{\sum_{j\in K_i} l_{ij}^{\alpha}} \sum_{j\in K_i} l_{ij}^{\alpha} \vec{H}[\vec{x}_i - \vec{x}_j] \qquad i = 1,\dots,N$$

Chaotic Rössler identical 3D oscillators

$$\begin{cases} \dot{x}_i = -\omega y_i - z_i - \dfrac{\sigma}{\sum_{j\in K_i} l_{ij}^{\alpha(t)}} \sum_{j\in K_i} l_{ij}^{\alpha(t)} (x_i - x_j) \\ \dot{y}_i = \omega x_i + 0.165 y_i \\ \dot{z}_i = 0.2 + z_i (x_i - 10) \end{cases} \qquad i = 1,\dots,N$$

Kuramoto's non identical 1D oscillators

$$\dot{\vartheta}_i = \omega_i + \frac{\sigma}{\sum_{j\in K_i} l_{ij}^{\alpha(t)}} \sum_{j\in K_i} l_{ij}^{\alpha(t)} \sin(\theta_j - \theta_i) \qquad i = 1,\dots,N$$

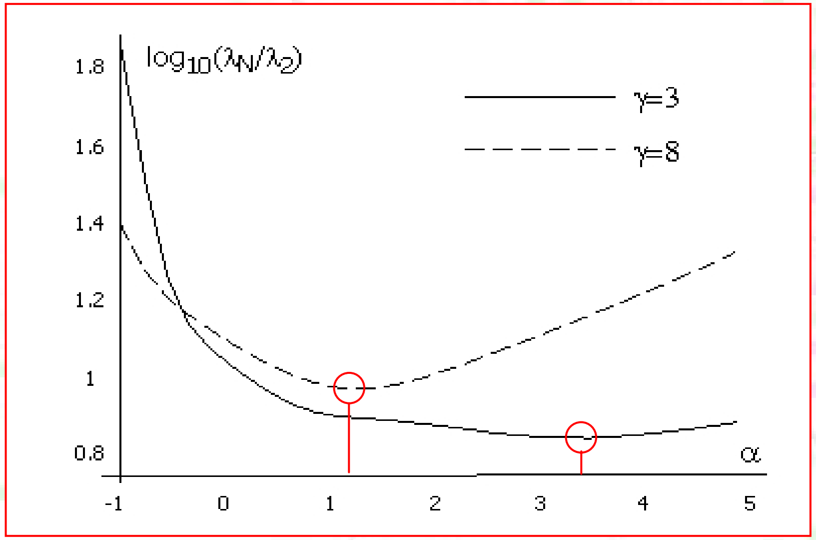
Sine-Circle Map: non identical 1D oscillators

$$x_i(n+1) = x_i(n) + \omega_i + \frac{\sigma}{\sum_{j\in K_i} l_{ij}^{\alpha(t)}} \sum_{j\in K_i} l_{ij}^{\alpha(t)} \sin(x_j - x_i) \qquad i = 1,\dots,N$$

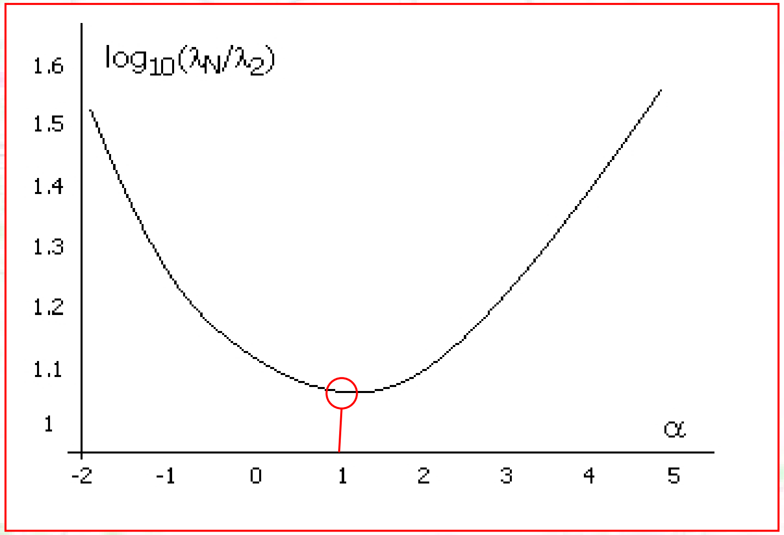
1. Finding the best synchronization condition for a network of Rössler oscillators

$$\dot{\vec{x}}_i = \vec{F}(\dot{\vec{x}}_i) - \frac{\sigma}{\sum_{j\in K_i} l_{ij}{}^{\alpha}} \sum_{j\in K_i} l_{ij}{}^{\alpha}\, \vec{H}[\vec{x}_i - \vec{x}_j] \qquad i = 1,\ldots,N$$

Scale free networks



Random networks



M.Chavez, D.U.Hwang, A.Amann, H.G.E.Hentschel and S.Boccaletti, *Phys. Rev. Lett.* **94** 218701 (2005)

2. Tuning the synchronization of a network of oscillators for finding community structures

DYNAMICAL CLUSTERING ALGORITHMS

$$\dot{\vec{x}}_i = \vec{F}(\dot{\vec{x}}_i) - \frac{\sigma}{\sum_{j \in K_i} l_{ij}^{\alpha(t)}} \sum_{j \in K_i} l_{ij}^{\alpha(t)} \vec{H}[\vec{x}_i - \vec{x}_j] \qquad i = 1, \dots, N$$

1. At variance with the topological methods we calculate the edge betweennesses (i.e. the edge's loads l_{ij}) of the network only one time before starting the simulation;

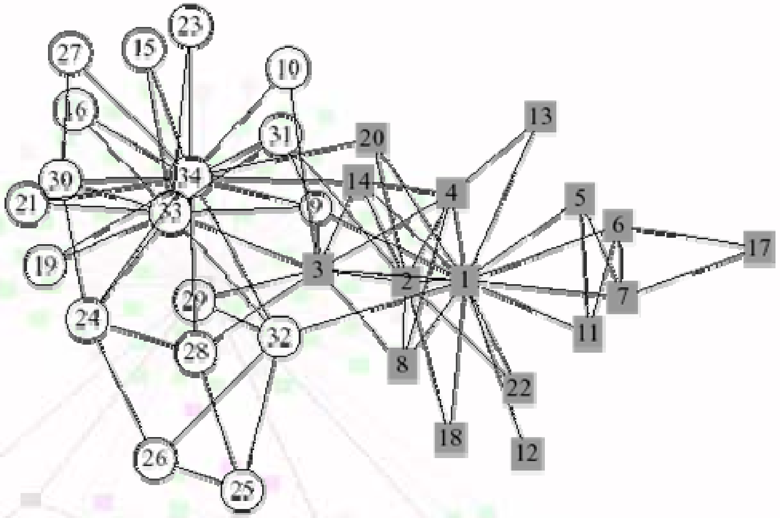
2. $t = 0$: $\alpha(0) \sim 0$ The system starts from a state with perfectly synchronized frequencies for a given value of the coupling strenght.;

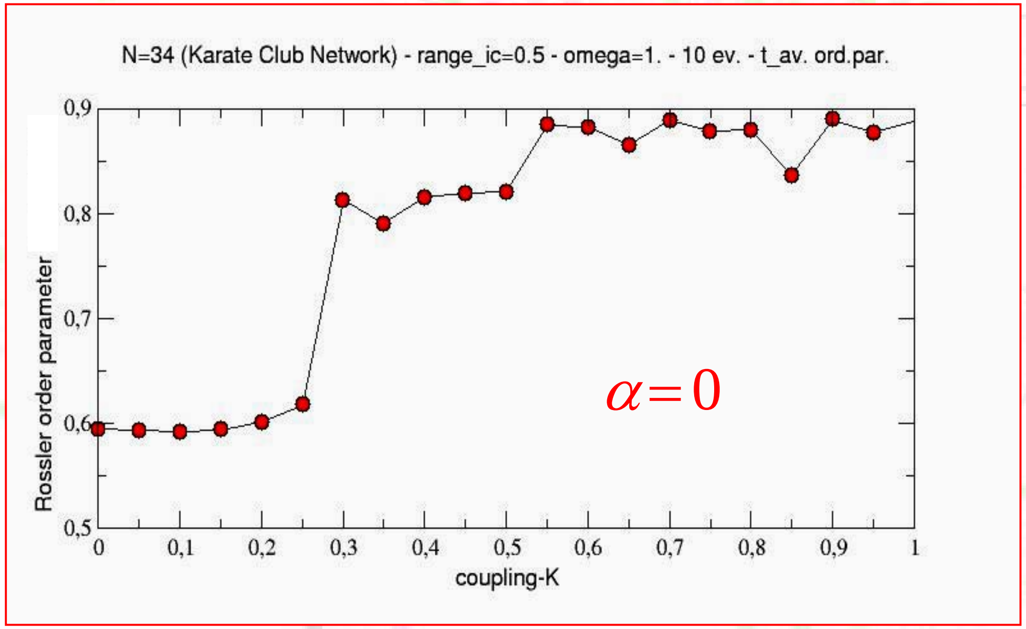
3. $t > 0$: $\alpha(t) \to -\infty$ Decreasing α, the edges with the greatest betweenness are weighted less and less and the oscillators progressively desynchronize;

4. We look to clusters of nodes oscillating with a common frequency (communities) and we select the clusters configuration with the highest modularity Q.

Chaotic Rössler identical 3D oscillators

$$
\begin{cases}
\dot{x}_i = -\omega y_i - z_i - \dfrac{K}{\sum_{j\in K_i} l_{ij}^{\alpha}} \sum_{j\in K_i} l_{ij}^{\alpha}(x_i - x_j) \\
\dot{y}_i = \omega x_i + 0.165 y_i \\
\dot{z}_i = 0.2 + z_i(x_i - 10)
\end{cases}
\qquad i = 1,\ldots,N
$$





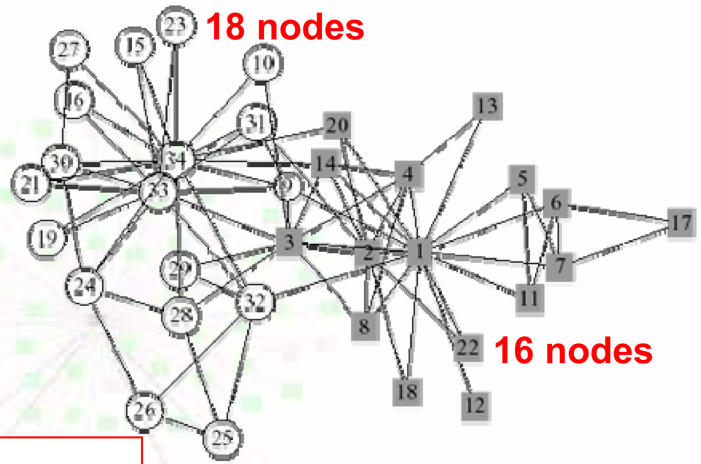
$$
\psi_{\inf} = \left\langle \frac{1}{N}\left|\sum_{i=1}^{N} e^{j\Phi i(t)}\right|\right\rangle_{t\to\infty}
$$

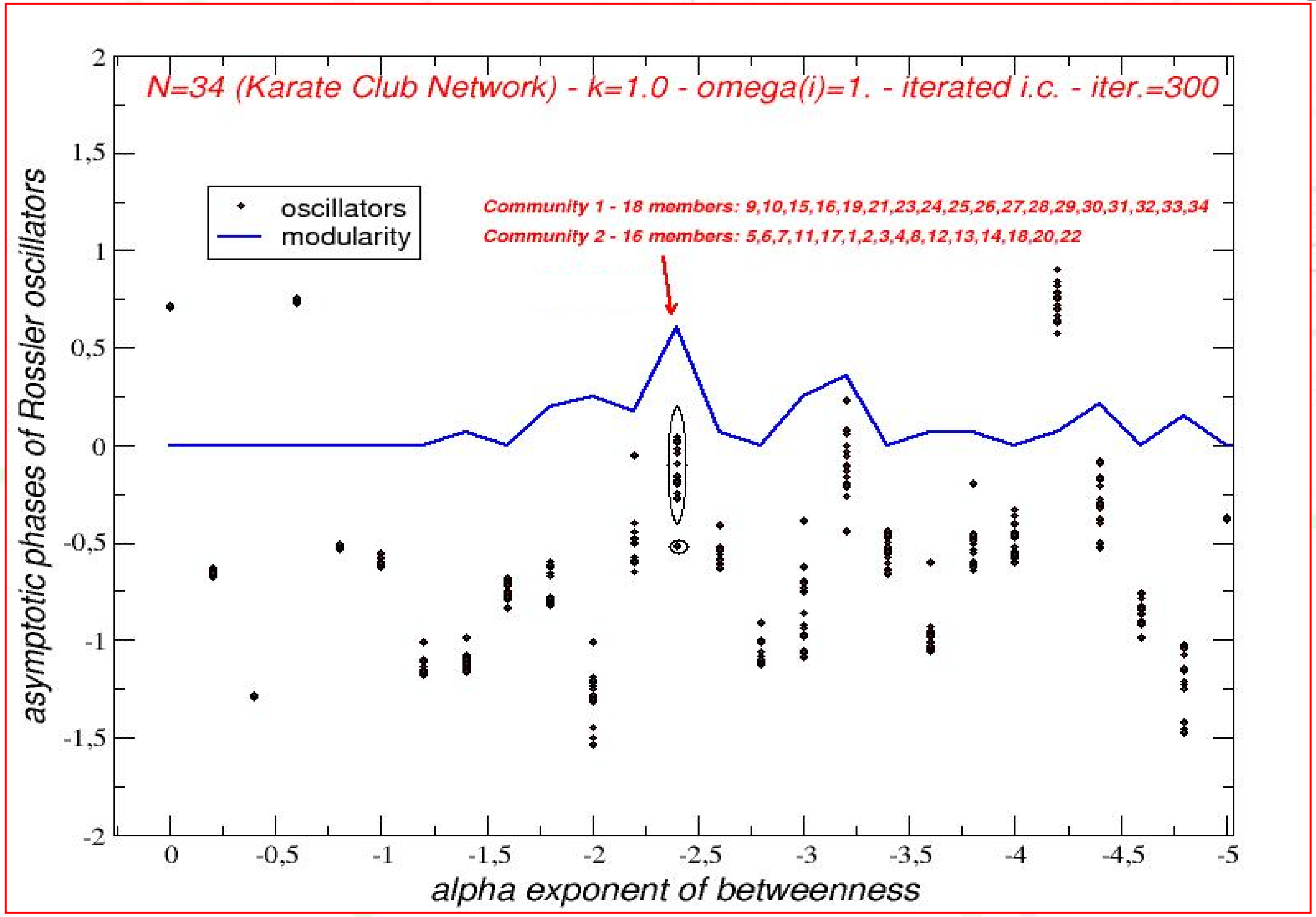
$$
\phi_i(t) = \arctan\left[\frac{y_i(t)}{x_i(t)}\right]
$$

A.Pluchino, V.Latora, A.Rapisarda and S.Boccaletti, *in preparation*

Chaotic Rössler identical 3D oscillators

$$
\begin{cases}
\dot{x}_i = -\omega y_i - z_i - \dfrac{K}{\sum_{j\in K_i} l_{ij}^{\alpha}} \sum_{j\in K_i} l_{ij}^{\alpha}(x_i - x_j) \\
\dot{y}_i = \omega x_i + 0.165 y_i \qquad\qquad i = 1,\dots,N \\
\dot{z}_i = 0.2 + z_i(x_i - 10)
\end{cases}
$$



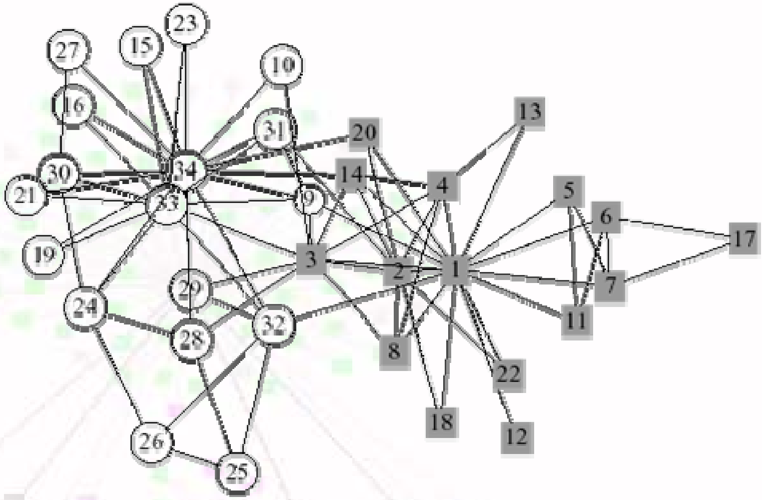


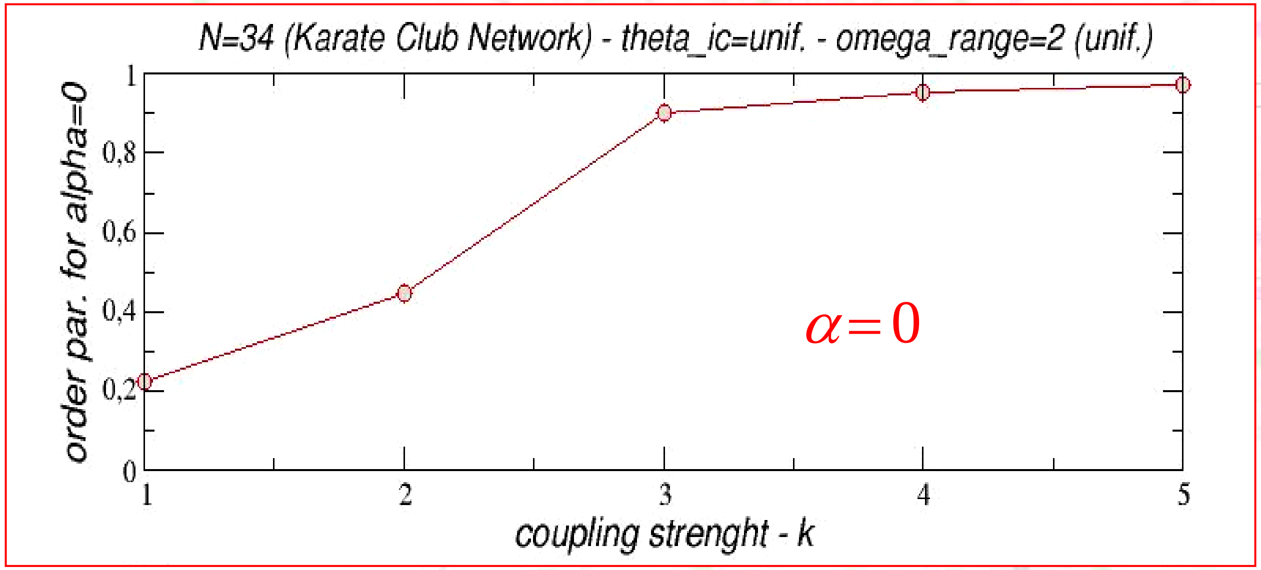
In this case the two main communities of Zachary's network have been correctly recognized

A.Pluchino, V.Latora, A.Rapisarda and S.Boccaletti, *in preparation*

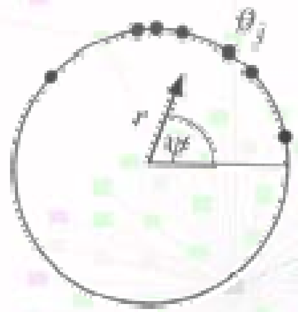
Kuramoto's non identical 1D oscillators

$$\dot{\vartheta}_i = \omega_i + \frac{K}{\sum_{j \in K_i} l_{ij}^{\alpha}} \sum_{j \in K_i} l_{ij}^{\alpha} \sin(\theta_j - \theta_i) \qquad i = 1, \ldots, N$$





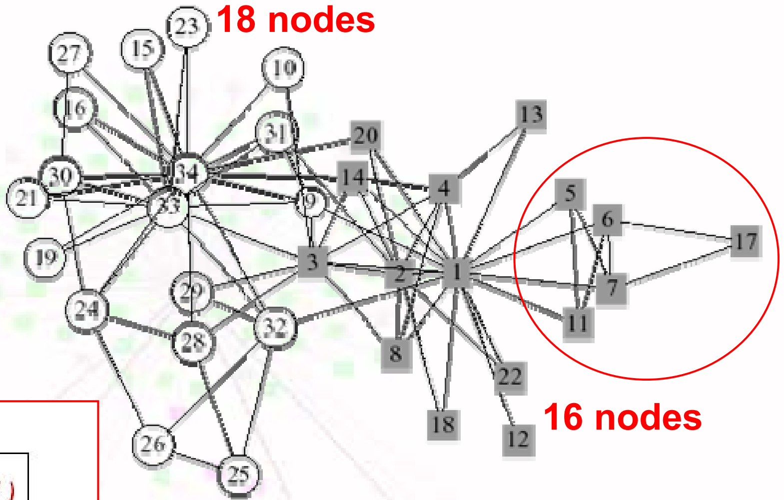
$$r_{\text{inf}} = \left\langle \frac{1}{N} \left| \sum_{i=1}^{N} e^{j\vartheta_i(t)} \right| \right\rangle_{t \to \infty}$$

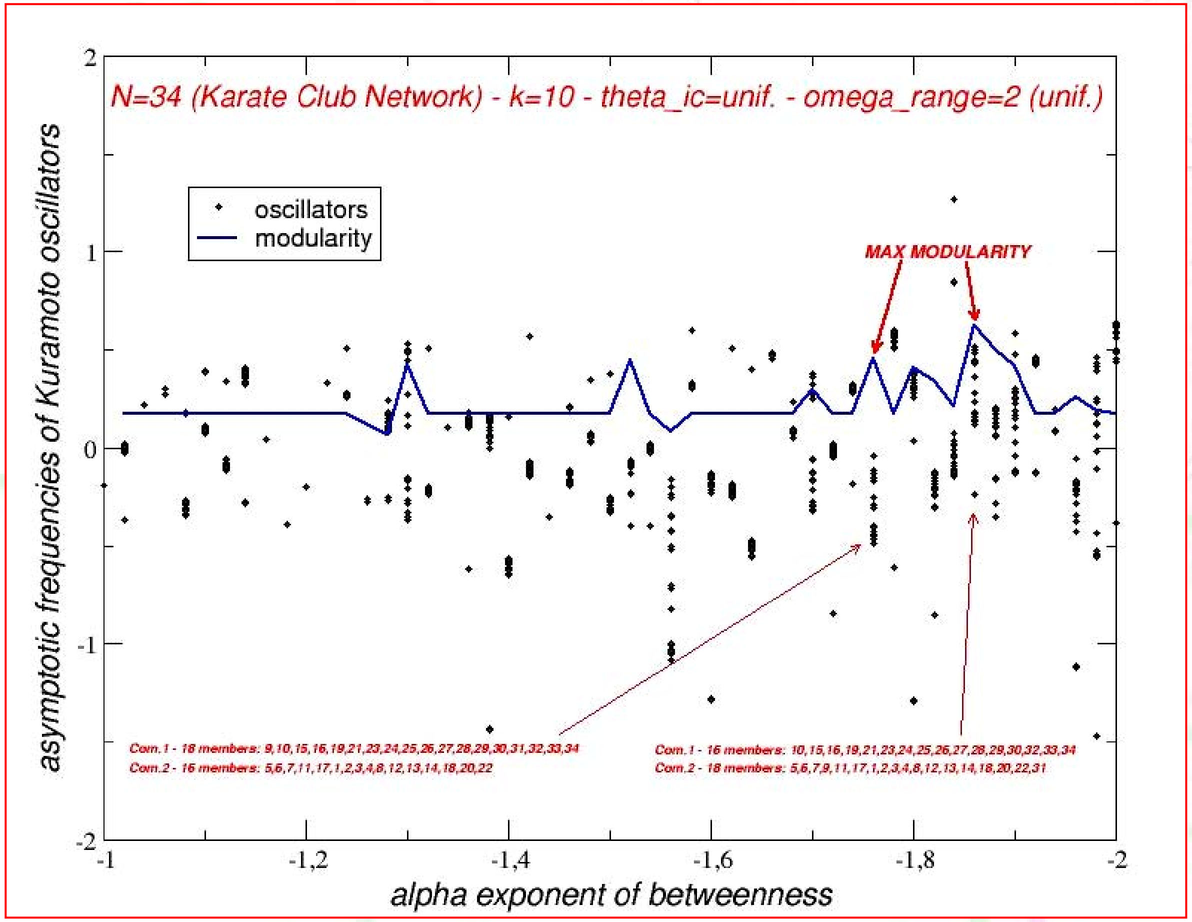


A.Pluchino, V.Latora, A.Rapisarda and S.Boccaletti, *in preparation*

Kuramoto's non identical 1D oscillators

$$\dot{\vartheta}_i = \omega_i + \frac{K}{\sum_{j\in K_i} l_{ij}^{\alpha}} \sum_{j\in K_i} l_{ij}^{\alpha} \sin(\theta_j - \theta_i) \qquad i = 1,\ldots,N$$

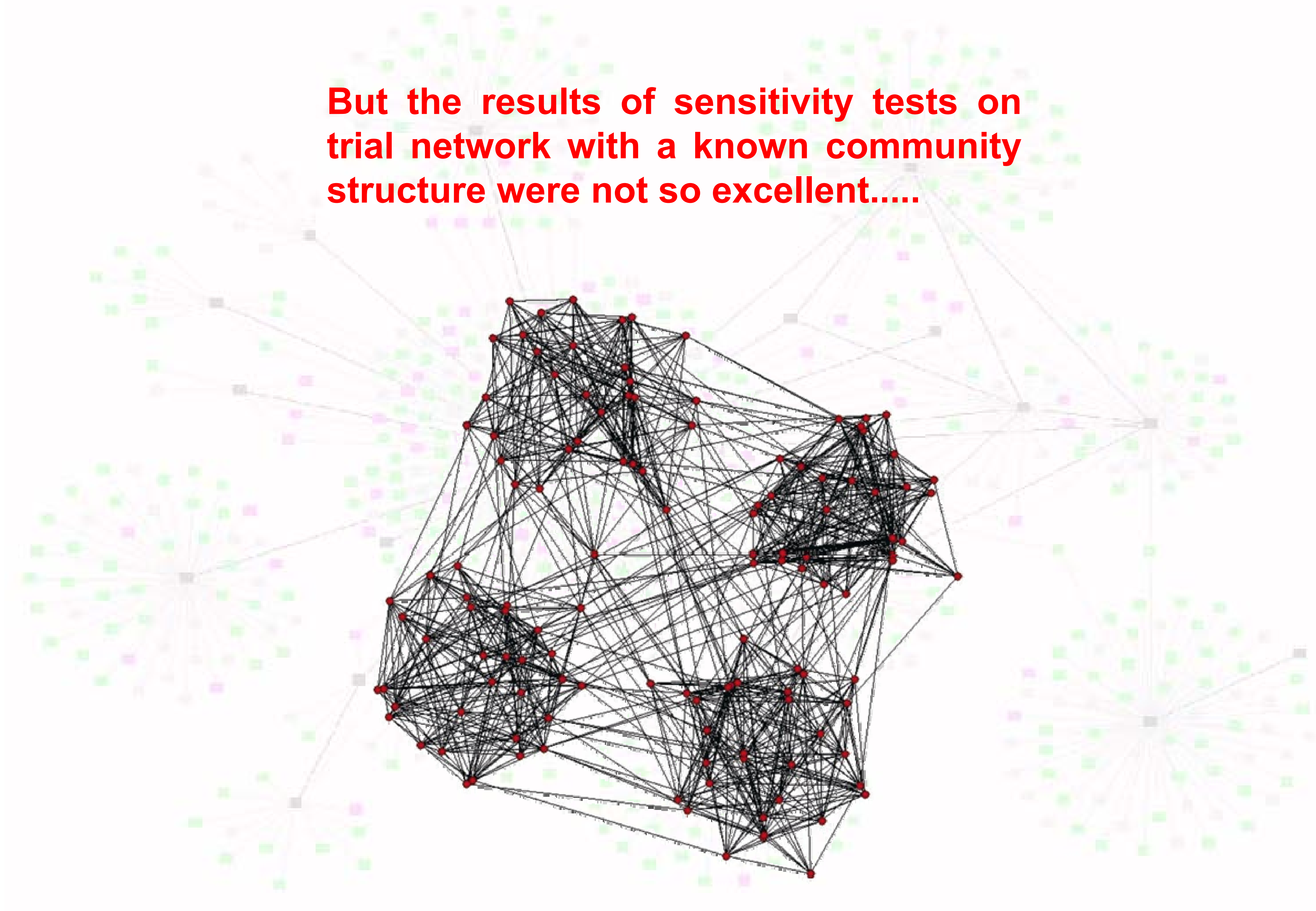




The two main communities of Zachary's network have been correctly recognized

A.Pluchino, V.Latora, A.Rapisarda and S.Boccaletti, *in preparation*

But the results of sensitivity tests on trial network with a known community structure were not so excellent.....



The Opinion Changing Rate (OCR) model*

It is a modification of the Kuramoto model and consists of the following rate equations describing the opinions evolution of N fully interacting agents:

coupling strenght

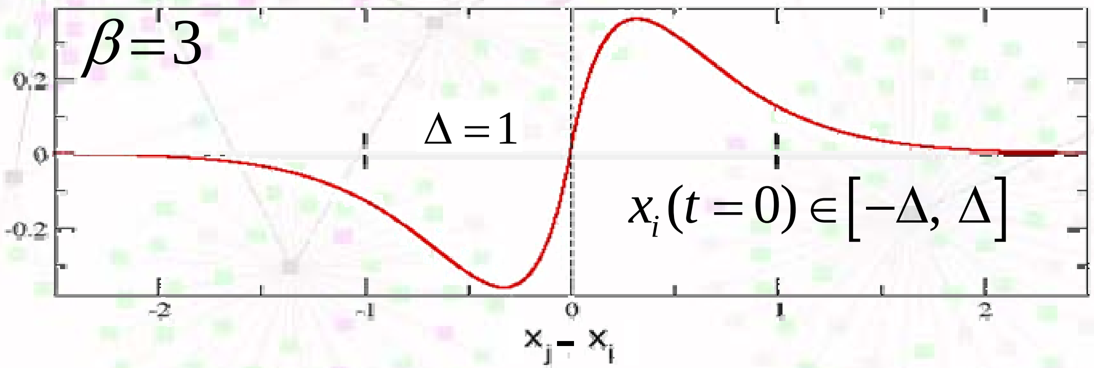
$$\dot{x}_i = \omega_i + \frac{\sigma}{N}\sum_{j=1}^{N}\beta \sin(x_j - x_i)\, e^{-\beta|x_j - x_i|}, \quad i = 1,....N$$

intrinsic frequencies **opinions**

$$x_i(t) \in \left]-\infty, +\infty\right[$$

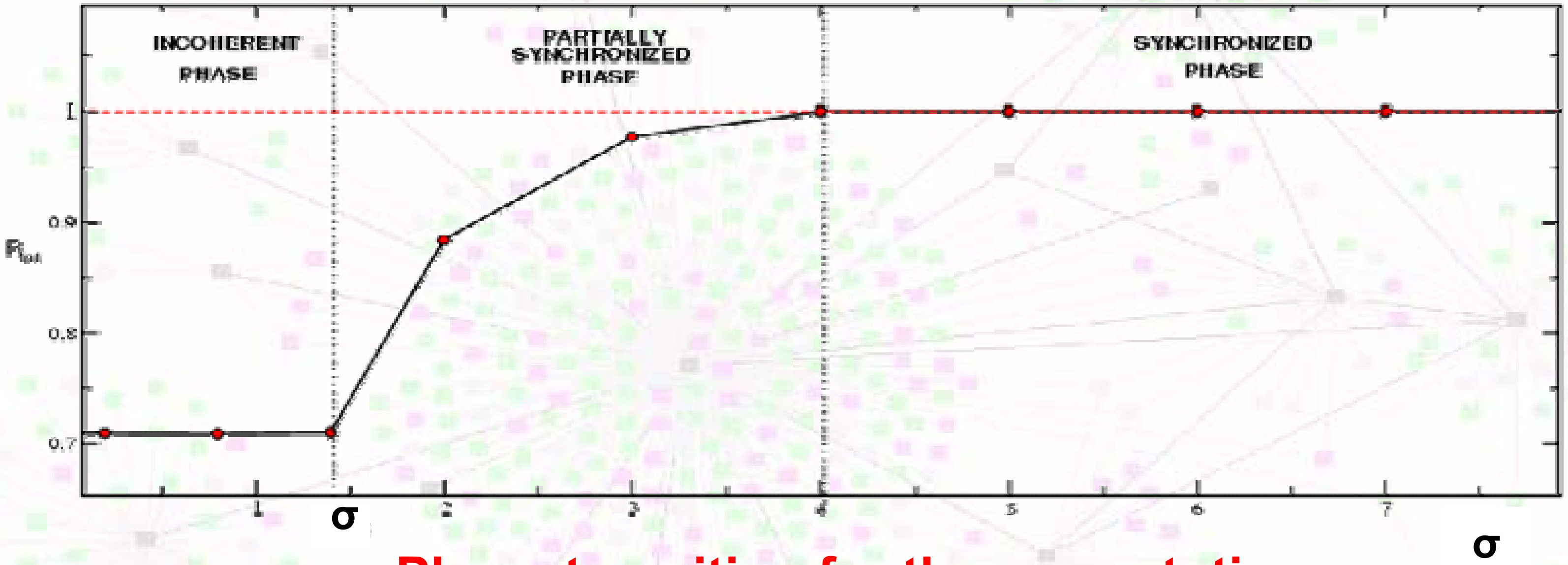
$$\omega_i \in [0, 1]$$

The interaction potential decreases for distant opinions:



*A.Pluchino, V.Latora, A.Rapisarda, *Int.Journ.of Mod.Phys. C* **16** 515 (2005)

The Opinion Changing Rate (OCR) model*



Phase transition for the asymptotic order parameter R_{inf} at σ_C~1.4

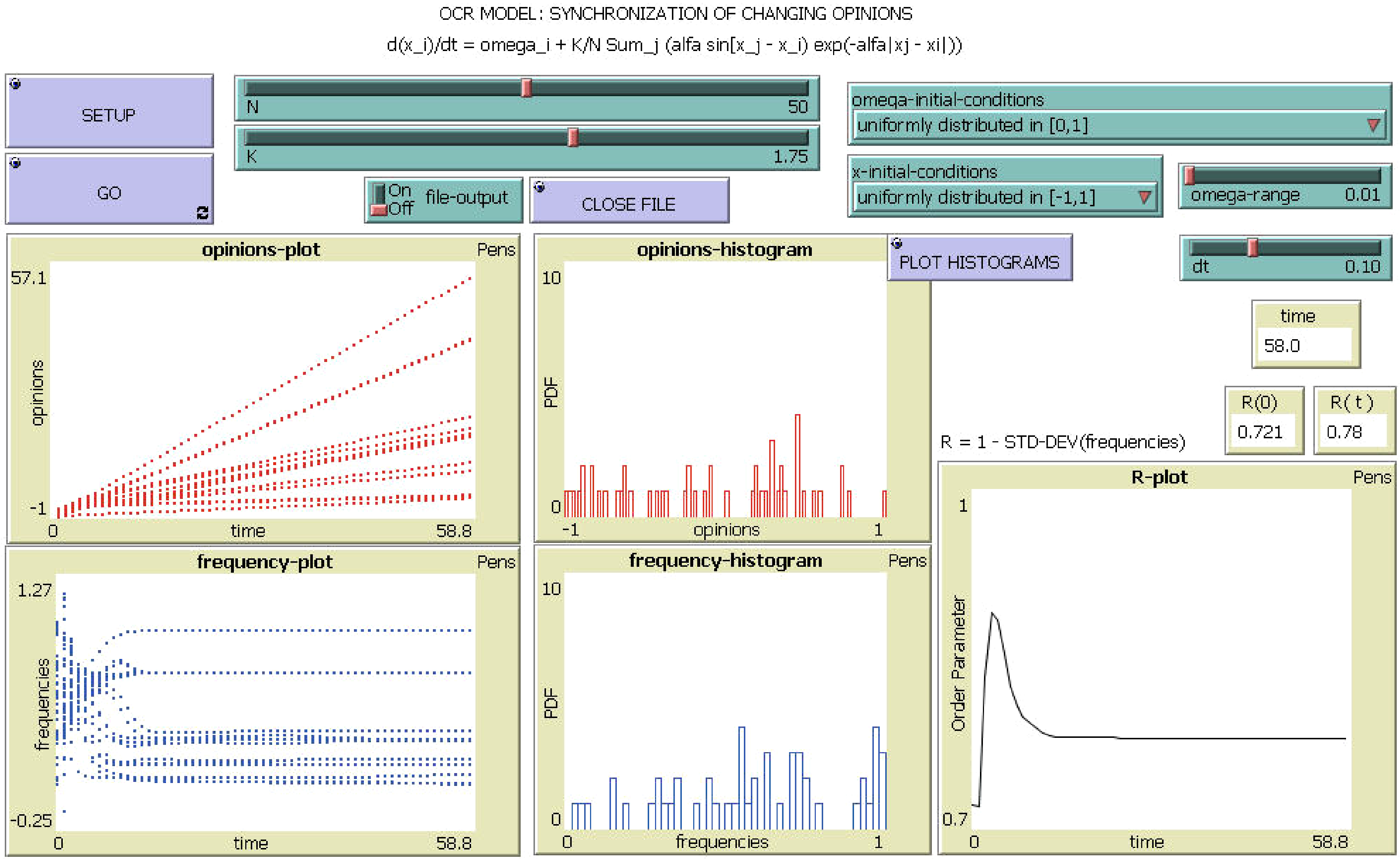
$$\boxed{R(t) = 1 - \sqrt{\frac{1}{N}\sum_{j=1}^{N}(\dot{x}_i - \dot{X})^2}}$$

with

$$\boxed{\dot{X} = \frac{1}{N}\sum_{i=1}^{N}(\dot{x}_i)^2}$$

*A.Pluchino, V.Latora, A.Rapisarda, *Int.Journ.of Mod.Phys. C* **16** 515 (2005)

The Opinion Changing Rate (OCR) model*



Here we further modified the standard OCR model forcing the oscillators natural frequencies to follow the so called Heigselmann-Krause dynamics...

$$\dot{x}_i = \omega_i + \frac{\sigma}{N}\sum_{j=1}^{N} \beta \sin(x_j - x_i)\; e^{-\beta |x_j - x_i|}, \quad i = 1, \ldots. N$$

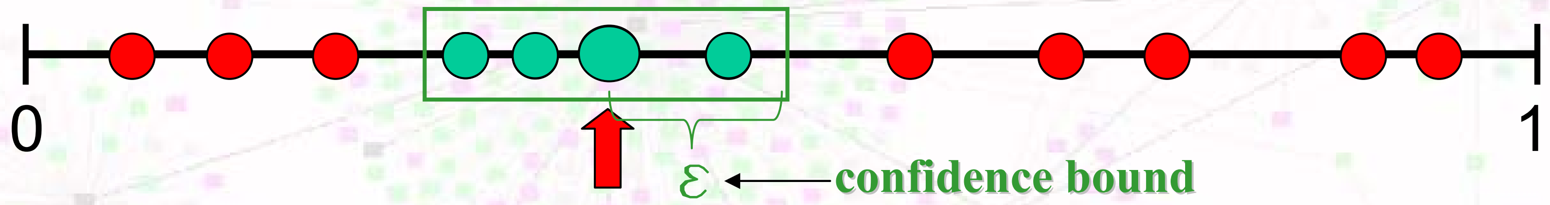
intrinsic frequencies

$\omega_i = \omega_i(t)$ time dependent

Heigselmann-Krause Dynamics

The Hegselmann-Krause (HK) opinion dynamics* is based on the presence of a parameter ε, called "confidence bound", which expresses the range of compatibility of opinions of agents put on a network (real space).

The 1-D opinion space is represented by the points of a [0,1] line, where the opinions are randomly distributed:



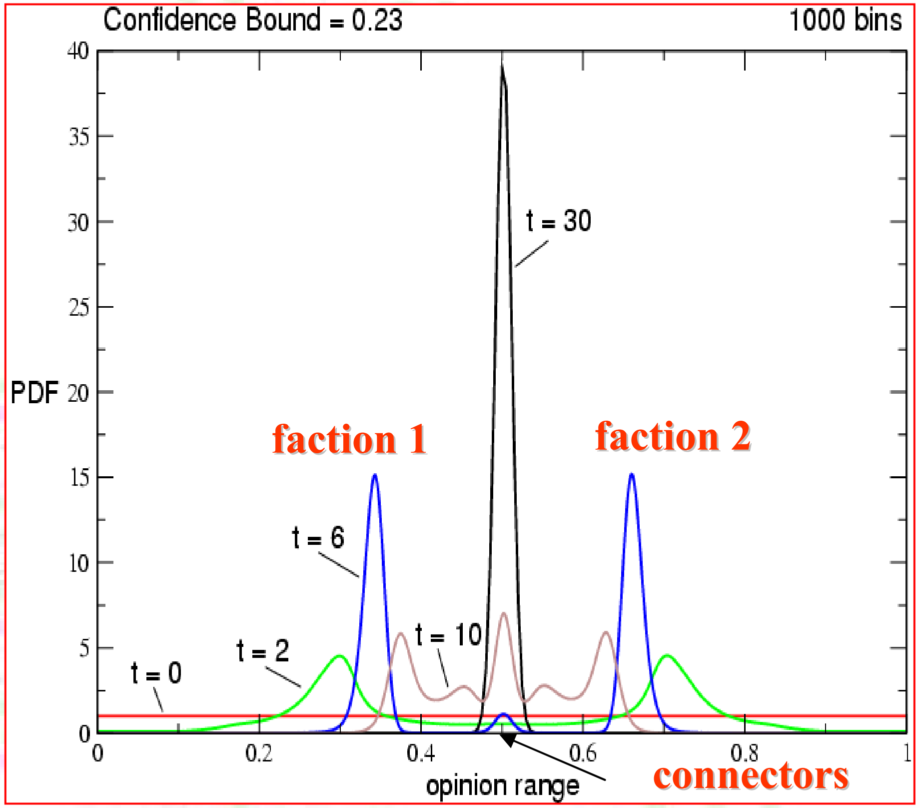
At each step, one chooses at random one opinion and checks how many opinions, belonging to first neighbours agents on the network, are compatible with him, i.e. are inside the confidence bound…

…at the next step, the agent takes the **average opinion** of its compatible neighbours...

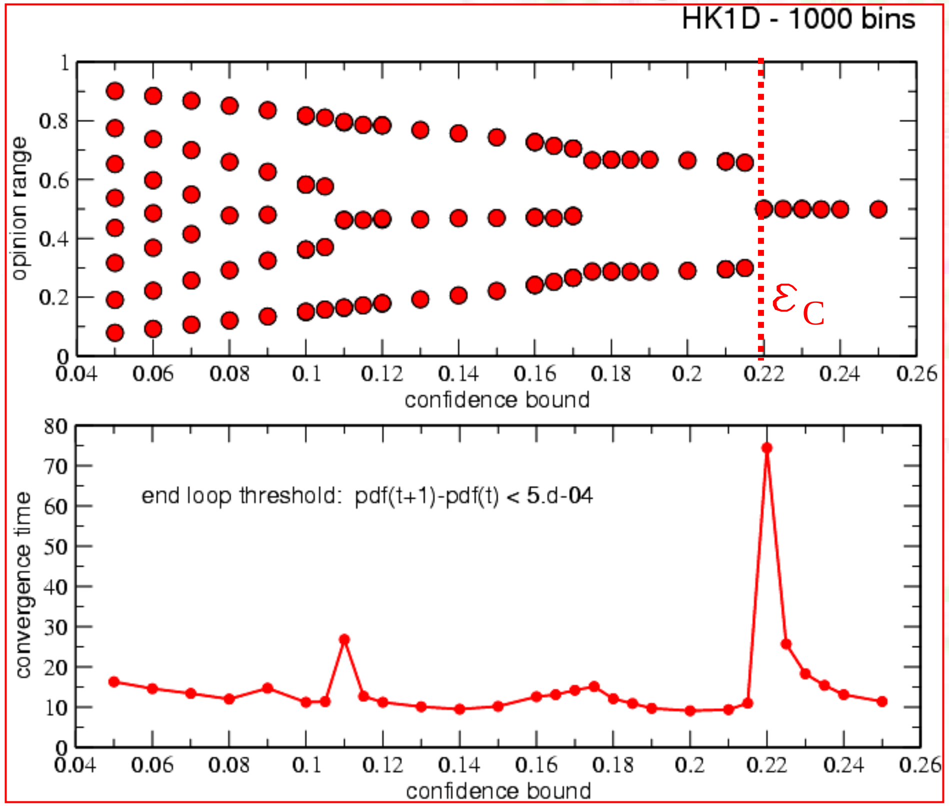
*R. Hegselmann and U. Krause, *Journal of Articial Societies and Social Simulation 5*, issue 3, paper 2 (jasss.soc.surrey.ac.uk) (2002);

Continuous version of Heigselmann-Krause model

Time evolution above the consensus threshold

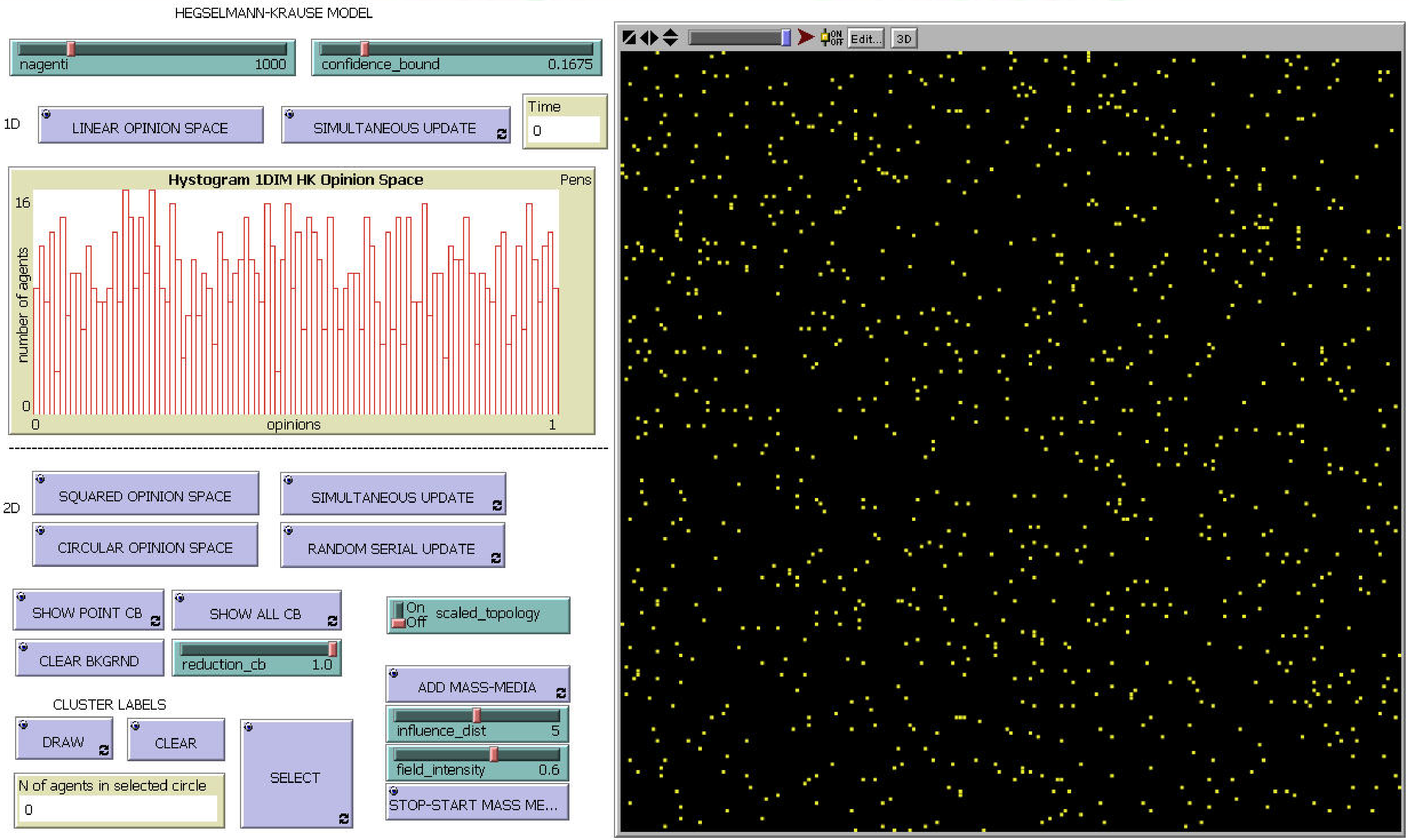


Clusters fusion and Convergence time



S.Fortunato, V.Latora, A.Pluchino, A.Rapisarda, *"Vector Opinion Dynamics in a bounded confidence consensus model"* Int.Journ.of Mod.Phys. C 16 (2005) 1535

Heigselmann-Krause Dynamics



OCR-HK
Dynamical Clustering Algorithm

istantaneous frequencies
(opinion changing rates)

loads (betweennesses)

tuning parameter

$$\dot{x}_i(t) = \omega_i(t) + \frac{\sigma}{\sum_{j \in K_i} l_{ij}^{\alpha(t)}} \sum_{j=1}^{N} \beta\, l_{ij}^{\alpha(t)} \sin(x_j - x_i)\, e^{-\beta |x_j - x_i|}, \quad i = 1, \ldots N$$

neighbours of node -i in the selected netwrok

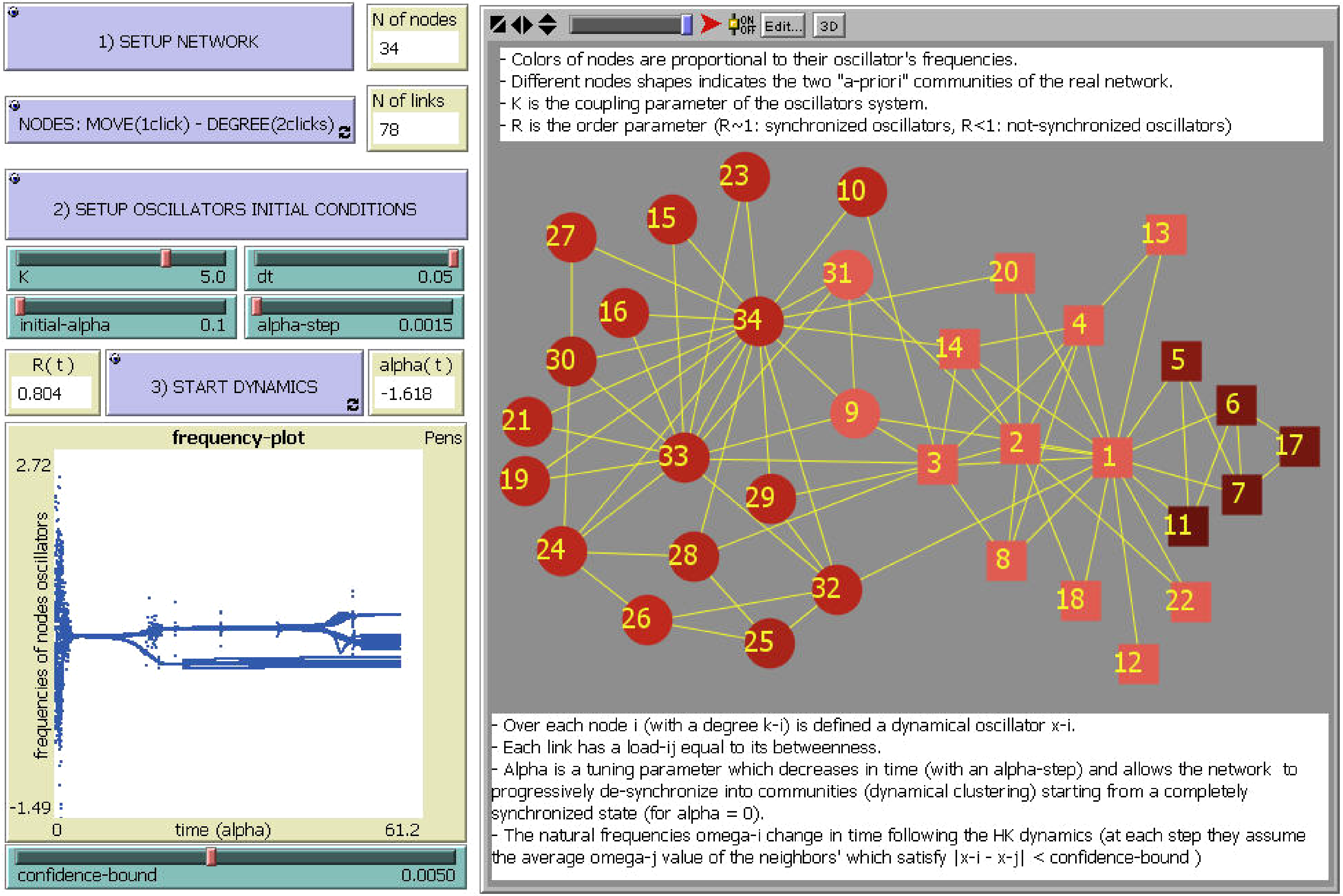
intrinsic frequencies, no more constant
but updated with HK dynamics

Starting from α=0 (synchronized state) we let α to decrease in time and we look at the evolution of clusters in frequency during a single run (frequency desynchronization).

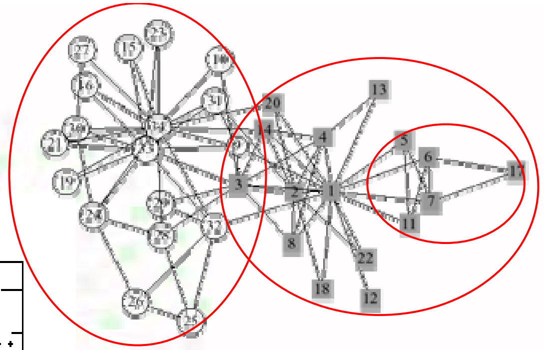
We repeat the procedure for several runs, with different initial conditions, then we select the configuration with the highest modularity Q

A.Pluchino, M.Ivanchenko, V.Latora, A.Rapisarda and S.Boccaletti - Submitted to PRE (2007) - *physics/0607179*

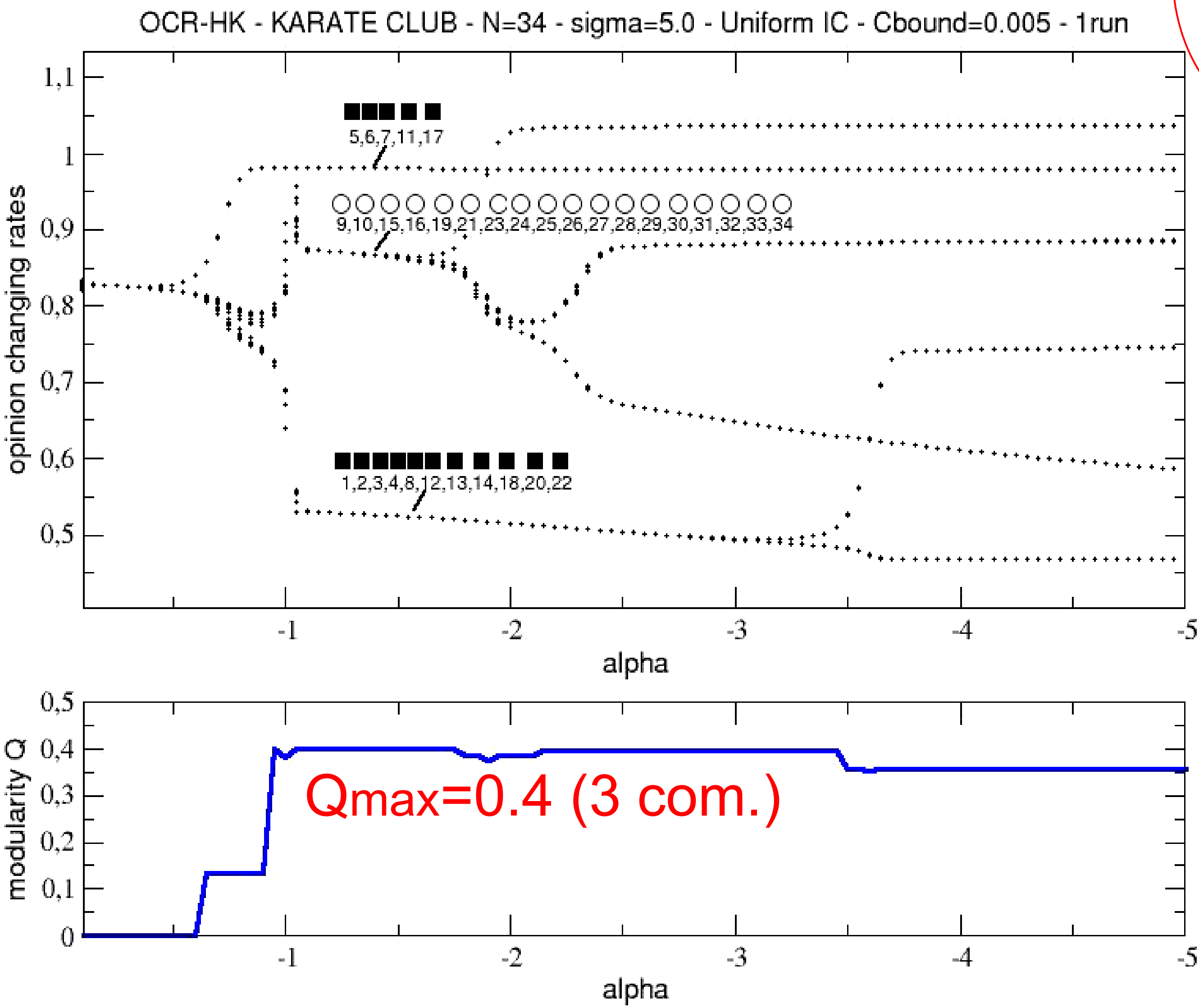
OCR-HK model: Karate Club



OCR-HK model: Karate Club



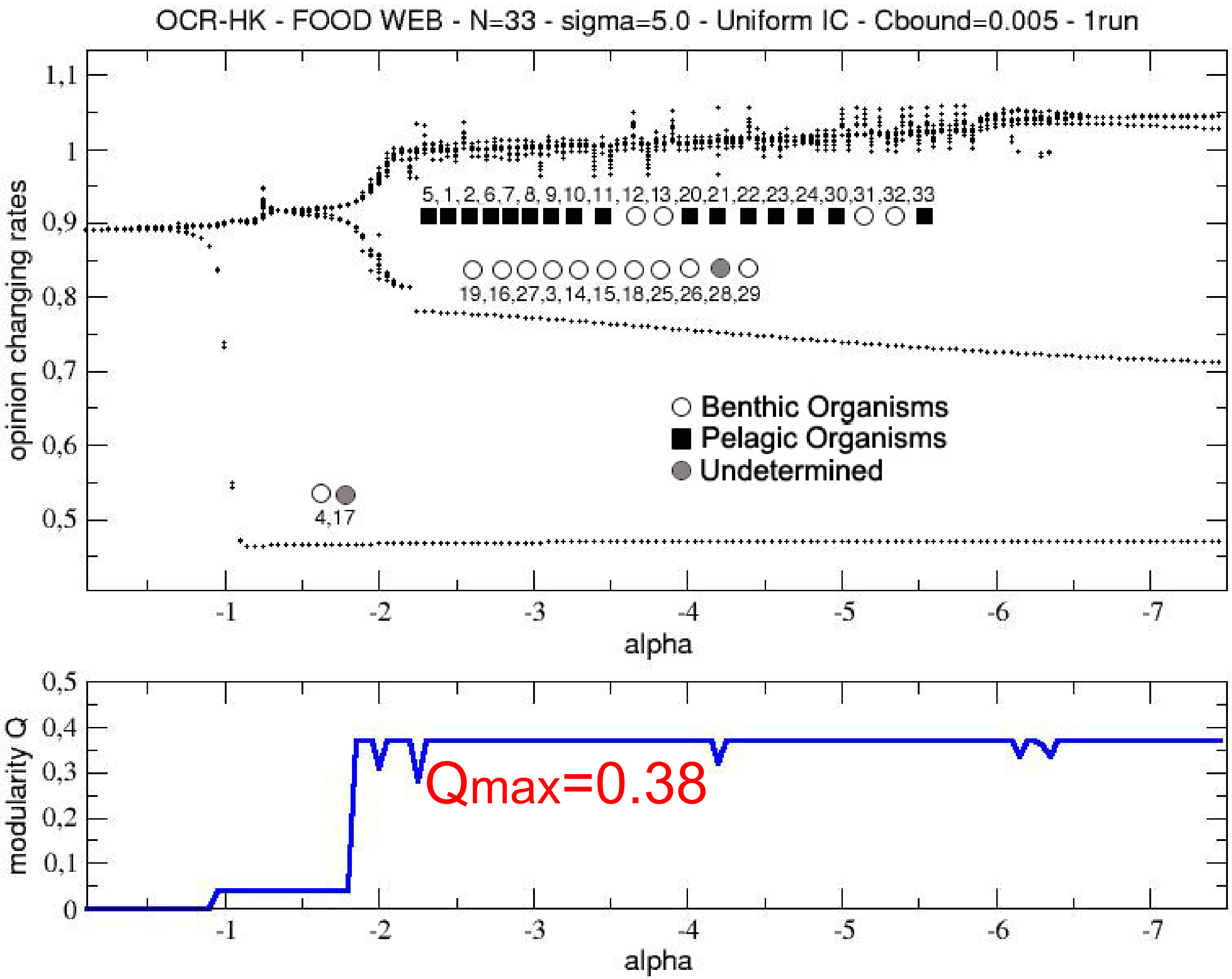
Q(2 com.)=0.37



W.Zachary (1977) *J. Anthropol.Res.* **33** 452-473

A.Pluchino, M.Ivanchenko, V.Latora, A.Rapisarda and S.Boccaletti - Submitted to PRE (2007) - *physics/0607179*

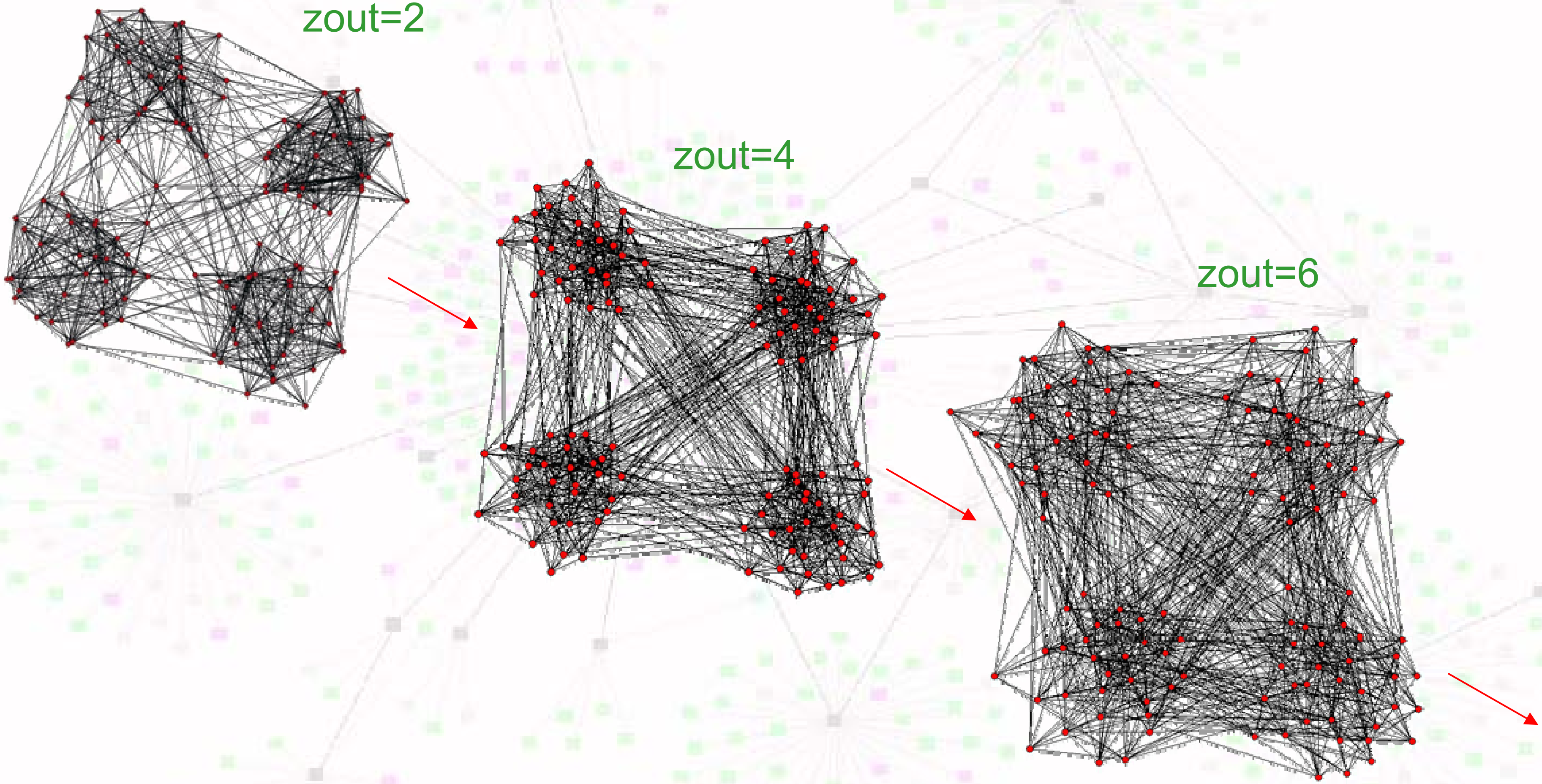
OCR-HK model: Chesapeake Bay food web (USA)

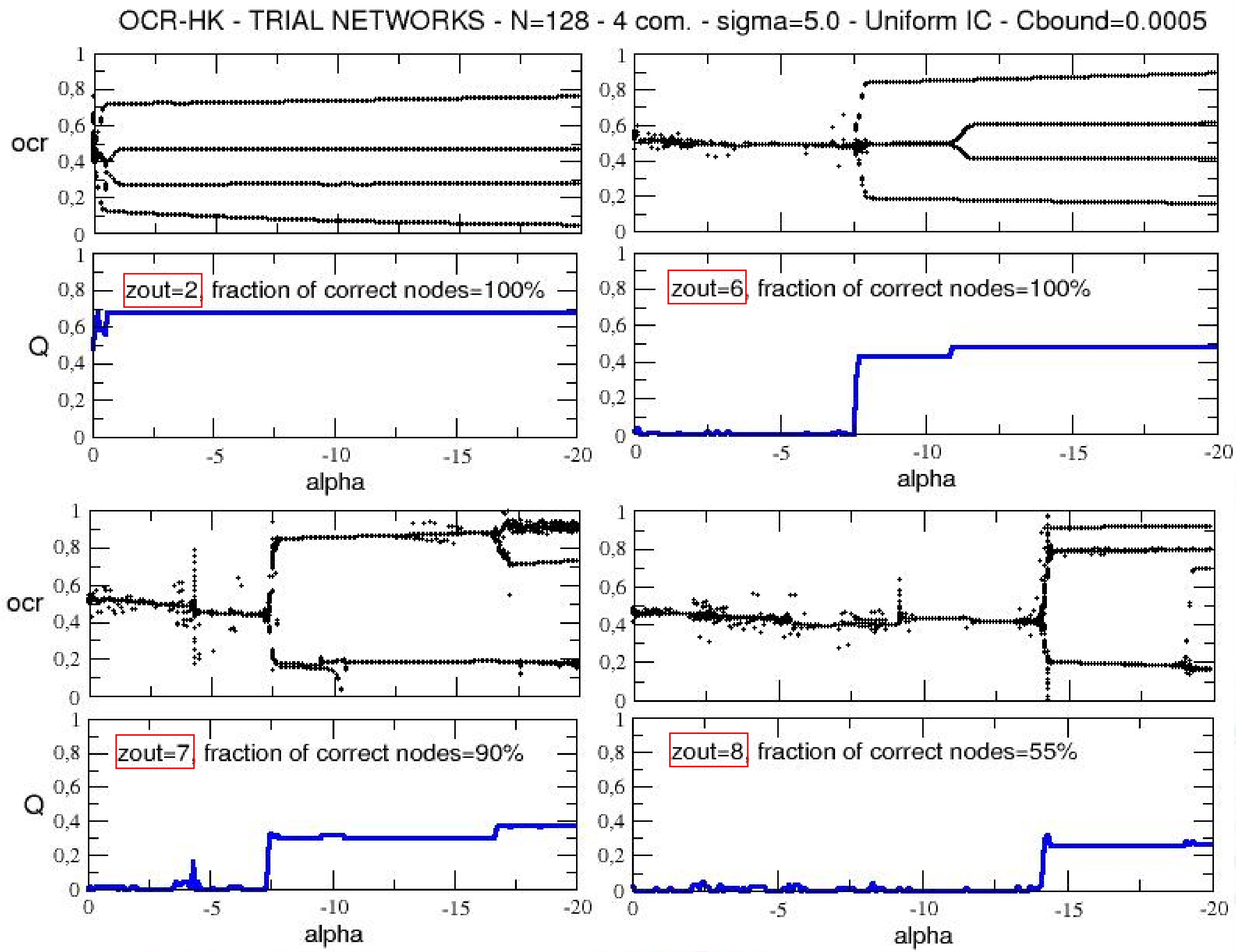


D.Baird & R.Ulanowicz (1989) *Ecol.Monogr*. **59** 329-364

A.Pluchino, M.Ivanchenko, V.Latora, A.Rapisarda and S.Boccaletti - Submitted to PRE (2007) - *physics/0607179*

OCR-HK model: “ad hoc” random trial networks (N=128, <k>=16, 4 communities) with increasing zout

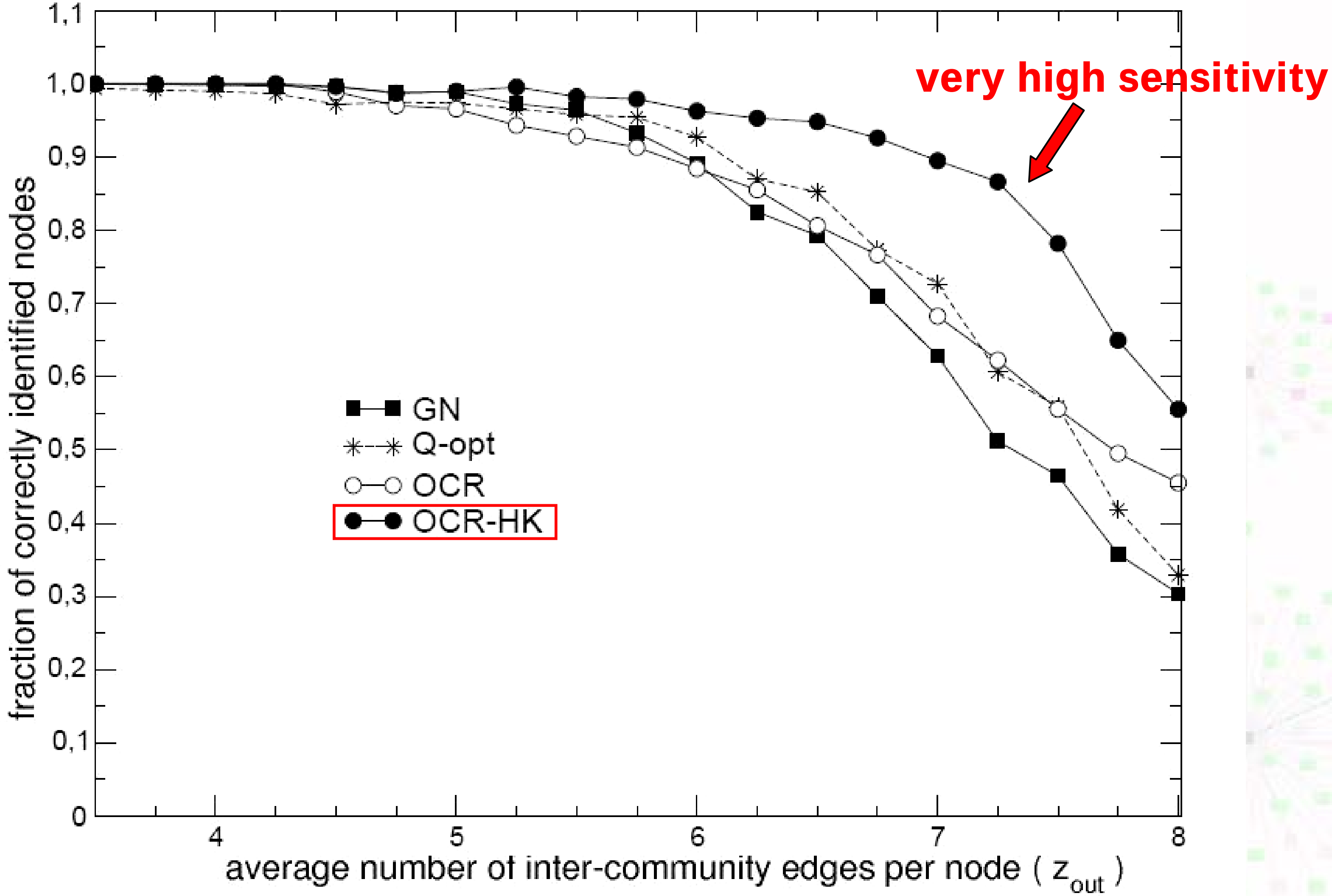




A.Pluchino, M.Ivanchenko, V.Latora, A.Rapisarda and S.Boccaletti - Submitted to PRE (2007) - *physics/0607179*

OCR-HK model on "ad hoc" random trial networks

Sensitivity test



L.Danon, A.Diaz-Guilera, J.Duch and A.Arenas *J.of Stat.Mech.: Theory and Exp.* (2005)

A.Pluchino, M.Ivanchenko, V.Latora, A.Rapisarda and S.Boccaletti - Submitted to PRE (2007) - *physics/0607179*

OCR-HK model on “ad hoc” random trial networks

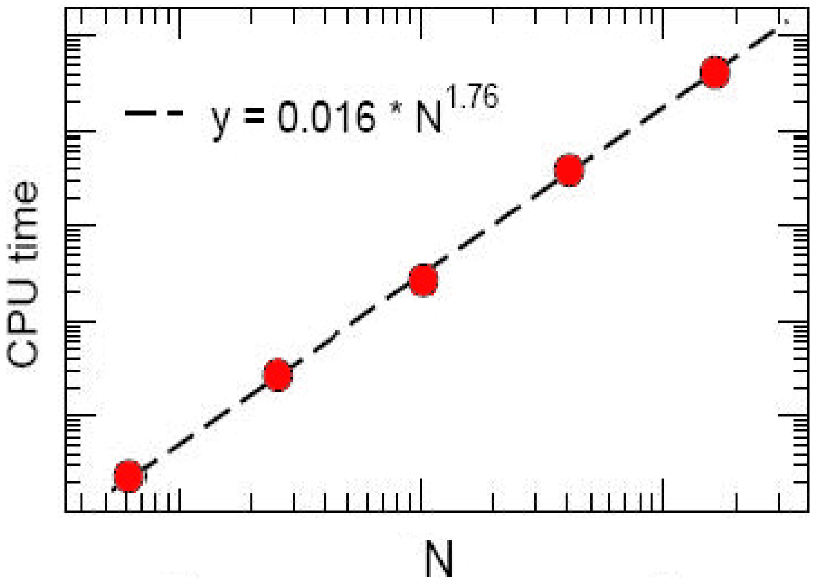
Computational cost

1.initial betwenness calculation: $O(N^2)$

+

2.dynamical clustering evolution time: $O(N^{1.76})$

very low global computational cost $O(\sim N^2)$



The best hierarchical methods scales with network size as $O(N\log^2 N)$

A.Pluchino, M.Ivanchenko, V.Latora, A.Rapisarda and S.Boccaletti - Submitted to PRE (2007) - *physics/0607179*

Conclusions

- The problem of **finding the best community structure** subdivision of a network is very important

- Divisive **topological methods** have a good sensitivity but have also an high computational cost

- We developed a new algorithm based on a **dynamical clustering** tecnique that shows a very high sensitivity and at the same time is very fast

- It makes also an interesting **bridge** between researches in **complex network** and those in **synchronization** of dynamical systems

- Future investigations regard the application of our algorithm to large real networks (genetic networks, social networks, etc...) – **work in preparation...**

Thanks for the attention!

CACTUS

Chaos And Complexity Theoretical University Study

Group

Catania

http://www.ct.infn.it/~cactus