Complex Networks as a tool to study human activity

INSTITUTE
FOR SCIENTIFIC INTERCHANGE
FOUNDATION

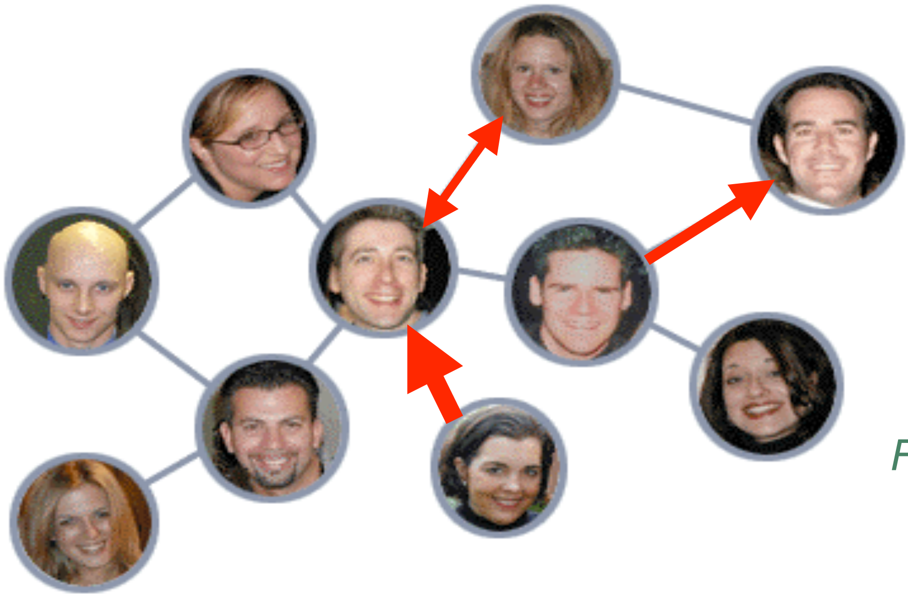
José Javier Ramasco

Bruno Gonçalves
Steven A. Morris
Romualdo Pastor-Satorras
Sergei Dorogovtsev

Outline

- Social Networks as weighted graphs
- Measures and meaning
- Social Inertia
- Results in Empirical Networks and models
- More information, Range
- Conclusions

Social Networks



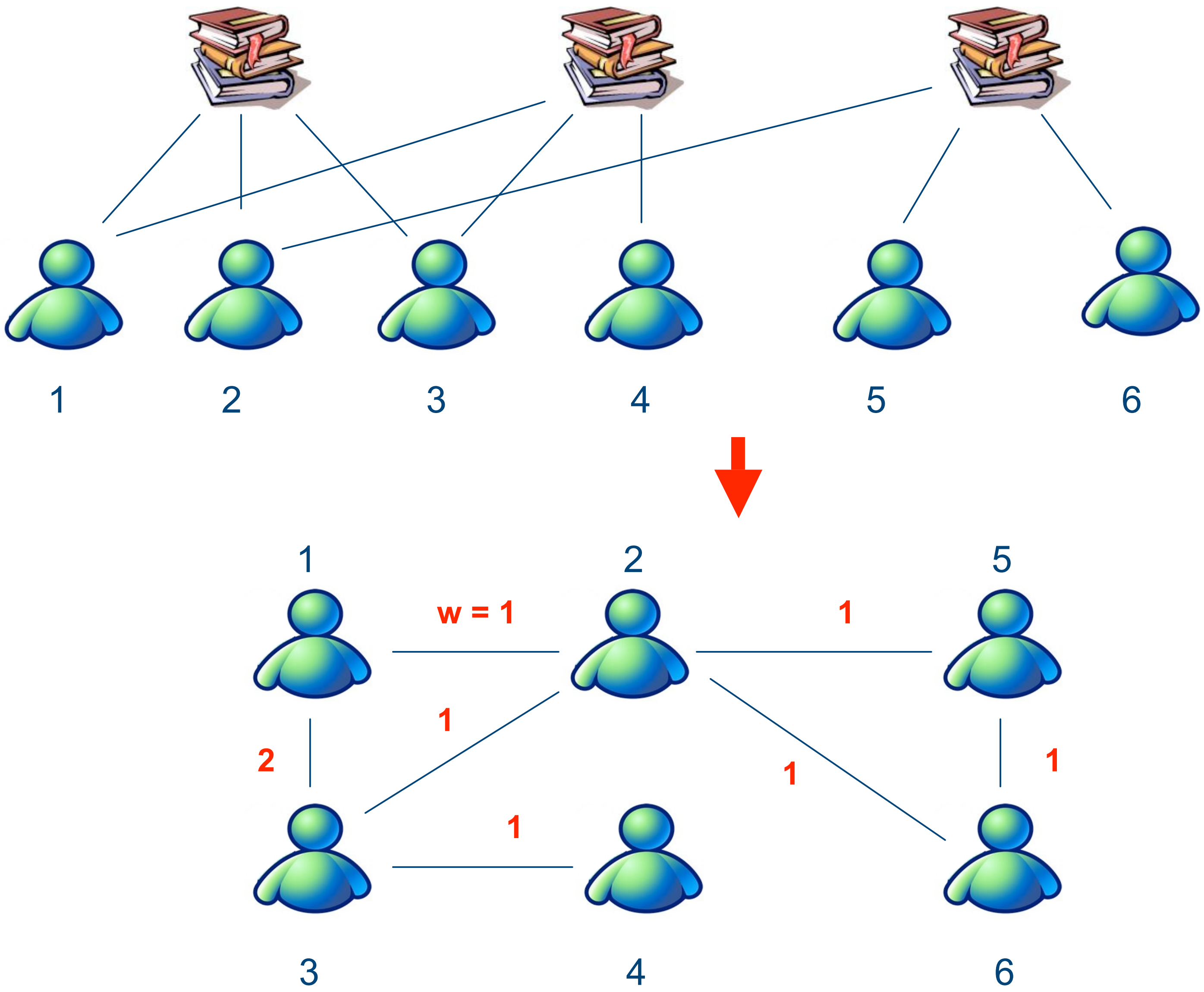
Friendster.com

Social Networks

Some characteristics that make social networks hard to study

- Arbitrariness of the definition
- Variability
- Size constrains on the social probes

Collaboration Networks



Collaboration Networks

The projected network contains less information than the original bipartite graph.

A way to avoid that in part is to use a weighted network in the projection.

The weight of a link may be defined as

$$w_{ij} = \sum_{k\ collab} \frac{\delta_i^k \delta_j^k}{n_k - 1}$$

$$w_{ij} = \sum_{k\ collab} \delta_i^k \delta_j^k = \text{num. common collab.}$$

Note however that there is still a certain loose of information.

M.E.J. Newman, PRE **64**, 016132 (2001).

Measures

We can define a number of distributions for the weighted network

- Degree distribution

$$P_k(k) \qquad \langle k \rangle \qquad C_k(k)$$

- Weight distribution of the edges

$$P_w(w) \qquad \langle w \rangle \qquad C_w(w)$$

- Strength distribution

$$s_i = \sum_{j \in \nu(i)} w_{ij}$$

$$P_s(s) \qquad \langle s \rangle \qquad C_s(s)$$

Measures

Other properties of the projected network that we are interested in

- Clustering

$$c_i = \frac{2t_i}{k_i(k_i - 1)} \qquad C = \langle c_i \rangle$$

$$c_i^w = \frac{1}{s_i(k_i - 1)} \sum_{j,m \in \nu(i)} \frac{w_{ij} + w_{im}}{2} a_{ij} a_{im} a_{jm} \qquad C^w = \langle c_i^w \rangle$$

- Correlations

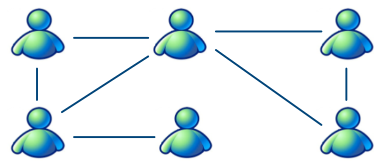
$$k_{nn,i} = \frac{1}{k_i} \sum_{j \in \nu(i)} k_j \qquad k_{nn}(k) = \langle k_{nn,i} \rangle_k$$

$$s_{nn}(s) = \langle s_{nn,i} \rangle_s$$

$$I_{nn}(I) = \langle I_{nn,i} \rangle_s$$

A. Barrat, *et al.*, PNAS **101**, 3747 (2004).

Social Inertia



Remember that

$$w_{ij} = \text{num. of works together}$$

that the strength means

$$s_i = \sum_{j \in \nu(i)} w_{ij} = \text{total num of partnerships}$$

And the degree of a node means

$$k_i = \text{num. of different partners}$$

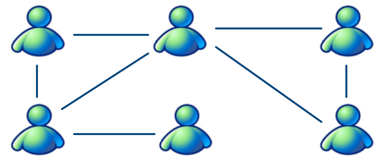
Let us define then the Social Inertia as

$$\boxed{\mathrm{I}_i = {}^{s_i}\!/_{k_i}}$$

J.J. Ramasco and S.A. Morris, PRE **73**, 016122 (2006).

Social Inertia

The extreme values of *I* are

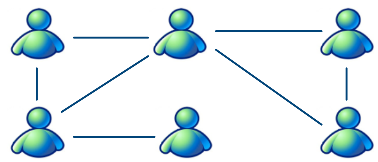


- $I_i \rightarrow 1$ if all collaborations happened with different partners.
- $I_i \rightarrow q_i$ if all works where done with the same team.

In general,   represents a measure of the eagerness of an author for collaborations with new people.

This concept could be generalized to other weighted graphs, its physical meaning (?).

Social Inertia

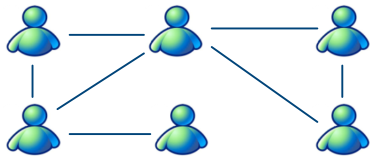


Warning: the social inertia is averaged over time. It is the same concept as an average velocity.

It should be possible to define an instant inertia but it requires a more detailed knowledge of the empirical databases.

$$\tilde{I}_i = \left\langle {}^{\Delta s_i}\!/_{\Delta k_i} \right\rangle_T \quad ?$$

Empirical results



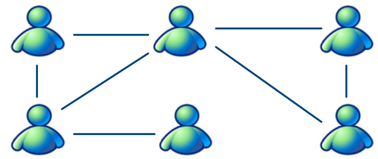
Field	N_c	N_a	$N_{ai}/N_a(\%)$
	IMDB movie database		
movies	127823	383640	0.37
	Scientific collaborations		
anthrax	2460	4320	8.9
atrial ablation	3091	6409	0.78
biosensors	5889	10993	1.1
botox	1560	3521	2.3
complex networks	900	1354	5.3
condmat	22002	16721	2.8
distance education	1389	2466	21.5
info science	14209	9399	40.4
info viz	2448	5520	12.4
scientometrics	3467	2926	21.04
self organized criticality	1631	2040	5.4
silicon on isolator	2381	4867	1.3
superconductors	1629	2981	6.5
superstrings	6643	3755	7.8

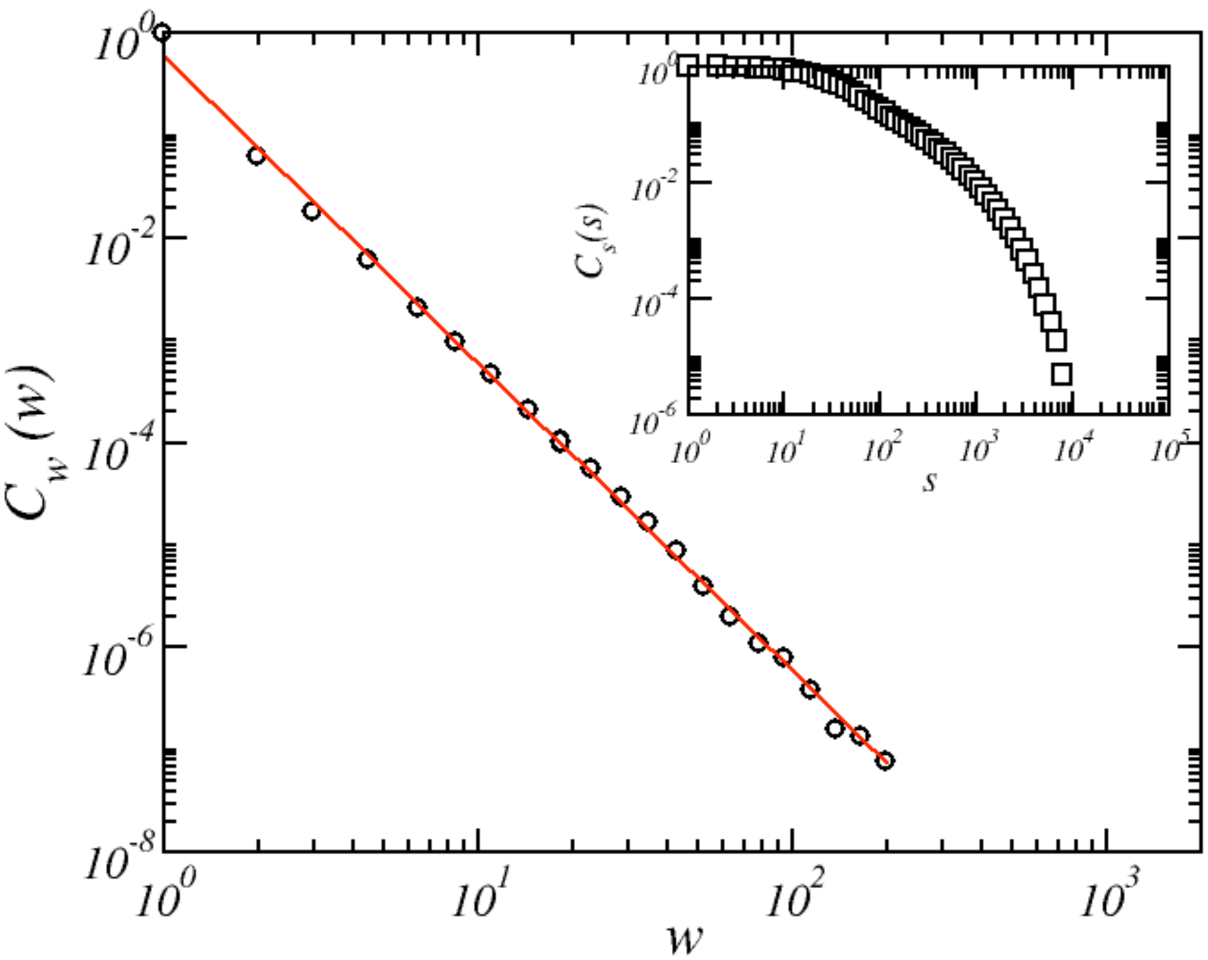
N_a = num. agents

N_c = num. collaborations

N_{ai} = num. agents that work alone

Empirical results





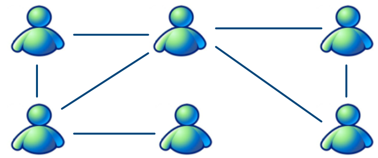
Main and Inset: movies

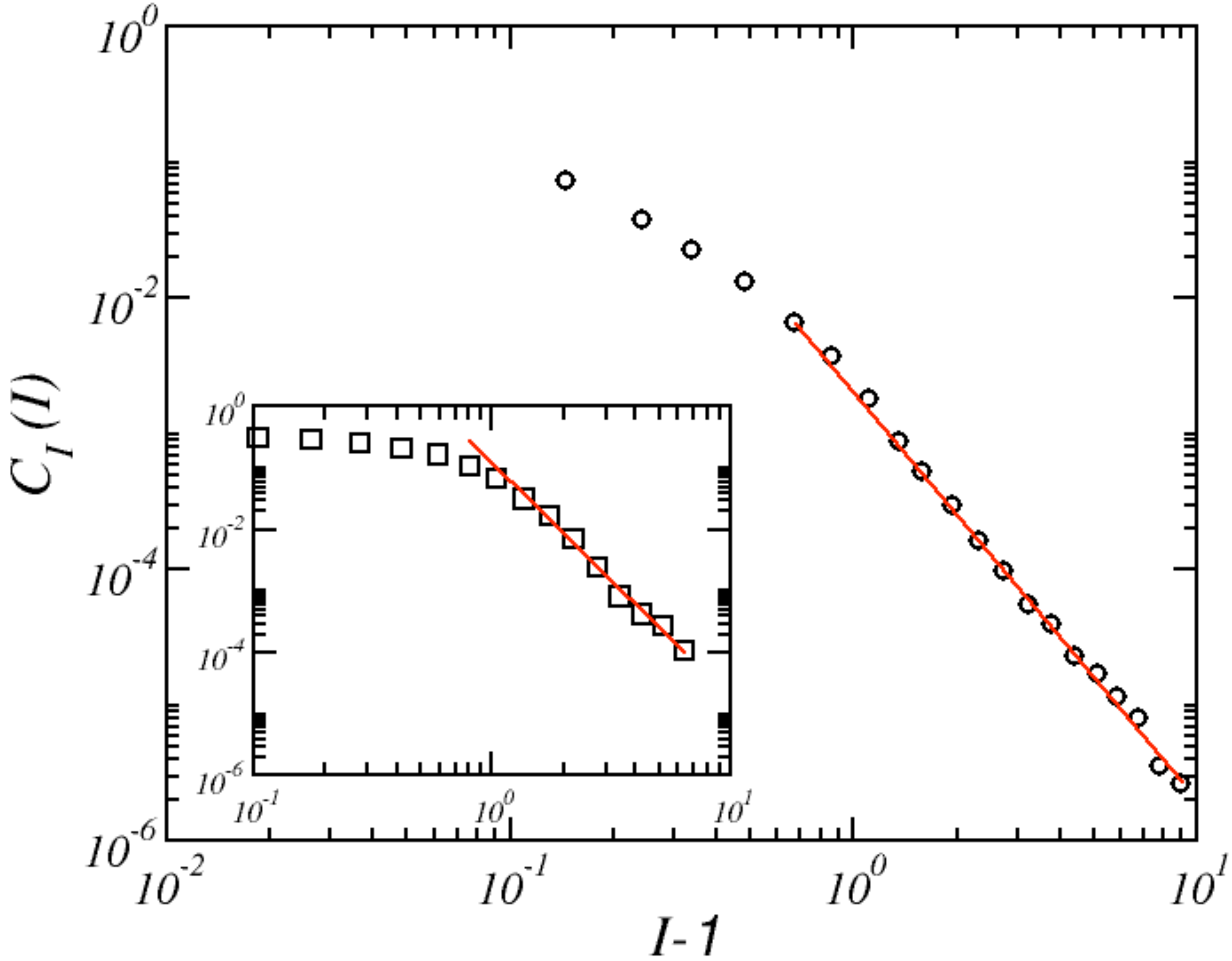
Slopes main
blue = -3 = 1-δ

$$P_w(w) \sim w^{-\delta}$$

$$C_w = \int_w^\infty P(w')dw' \sim w^{1-\delta}$$

Empirical results



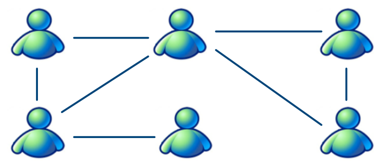


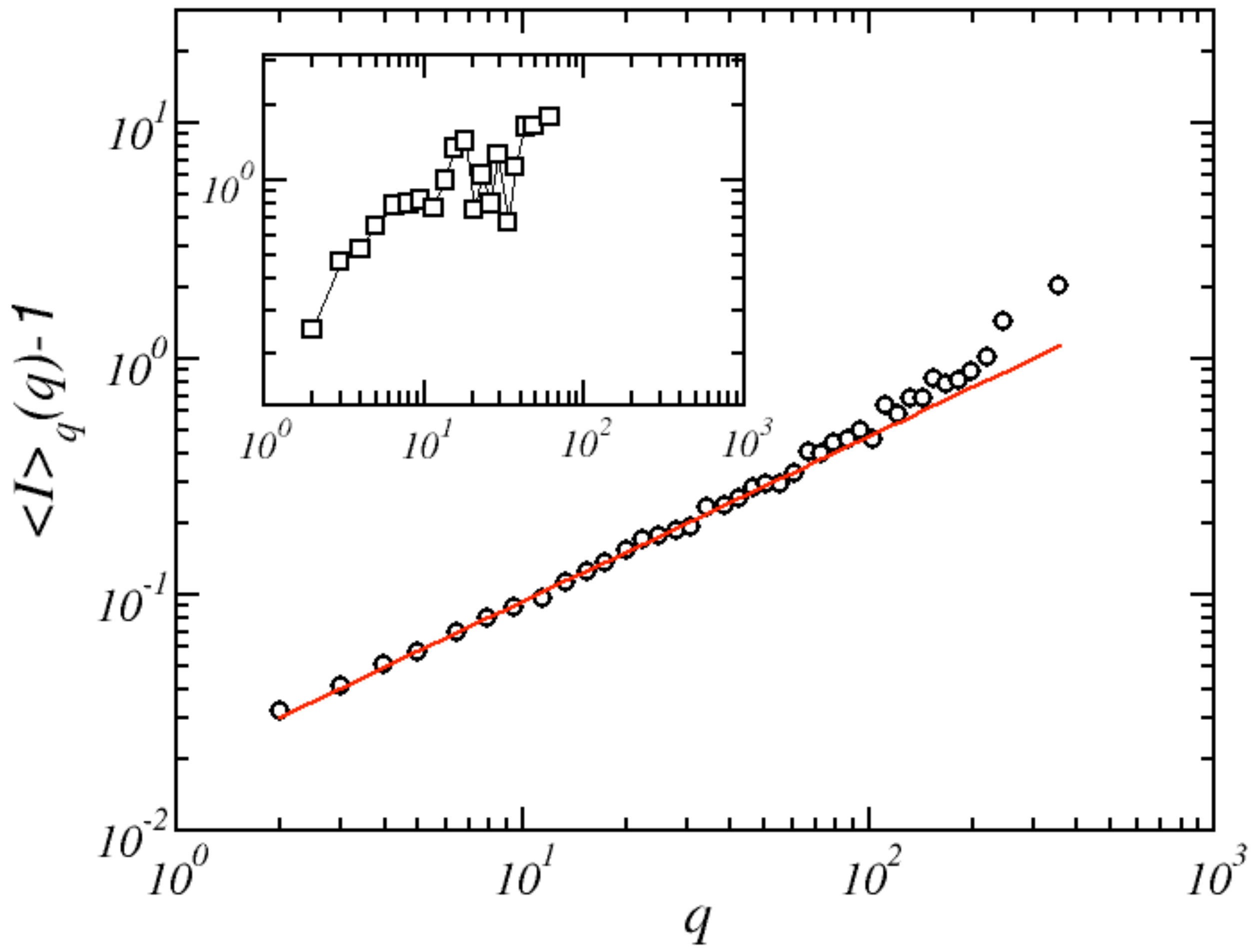
Main: movies
Inset: biosensors

Slopes main
red = -3

Slopes inset
red = -3.8

Empirical results

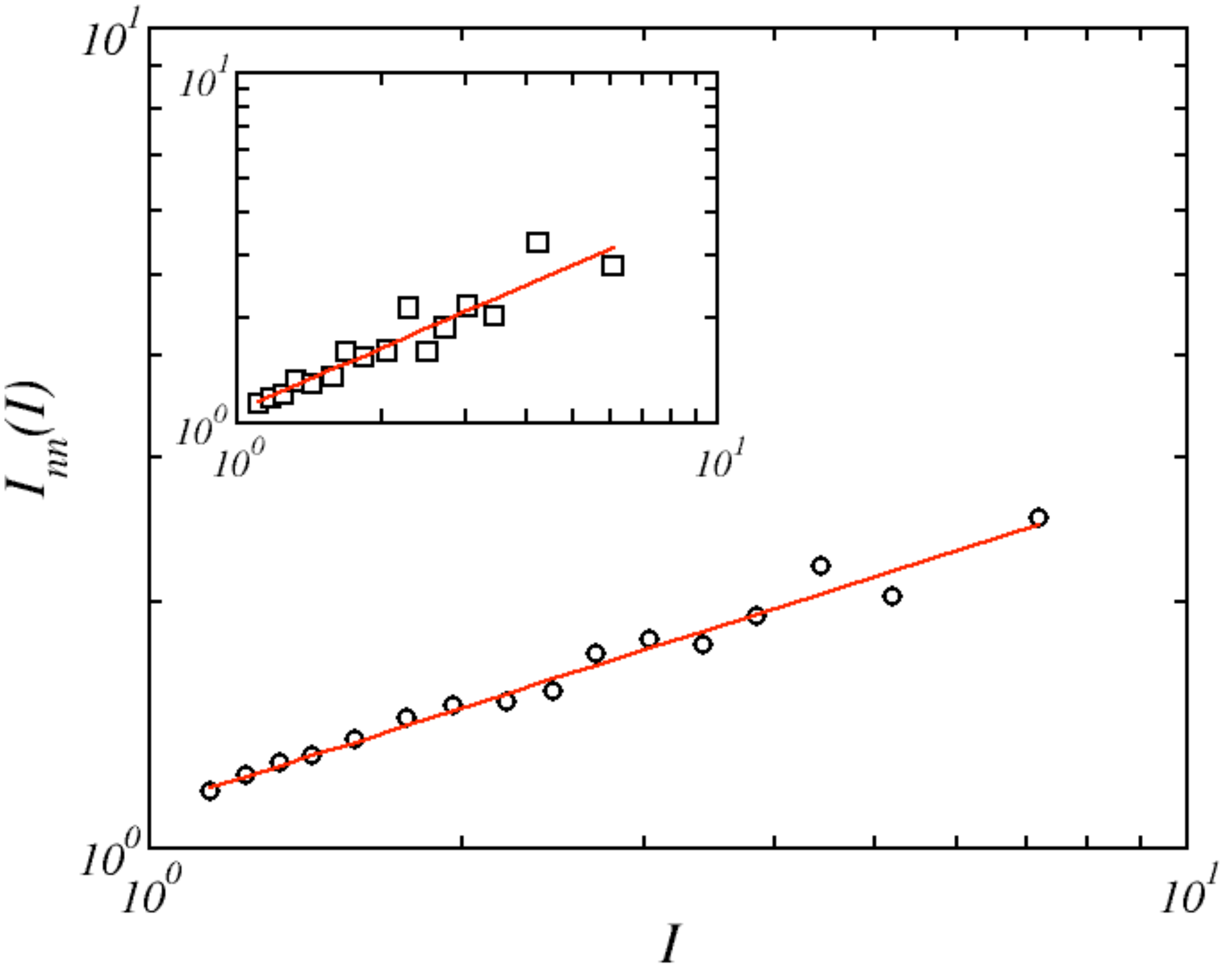




Main: movies
Inset: superstrings

Slopes main
red = 0.7

Empirical results

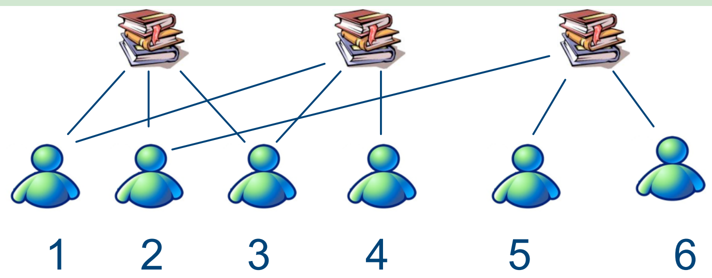


Main: movies
Inset: infoscience

Slopes main
red = 0.4

Slopes inset
red = 0.6

Models



Each step a new collaboration of size n is added. m of the agents are new, without experience.

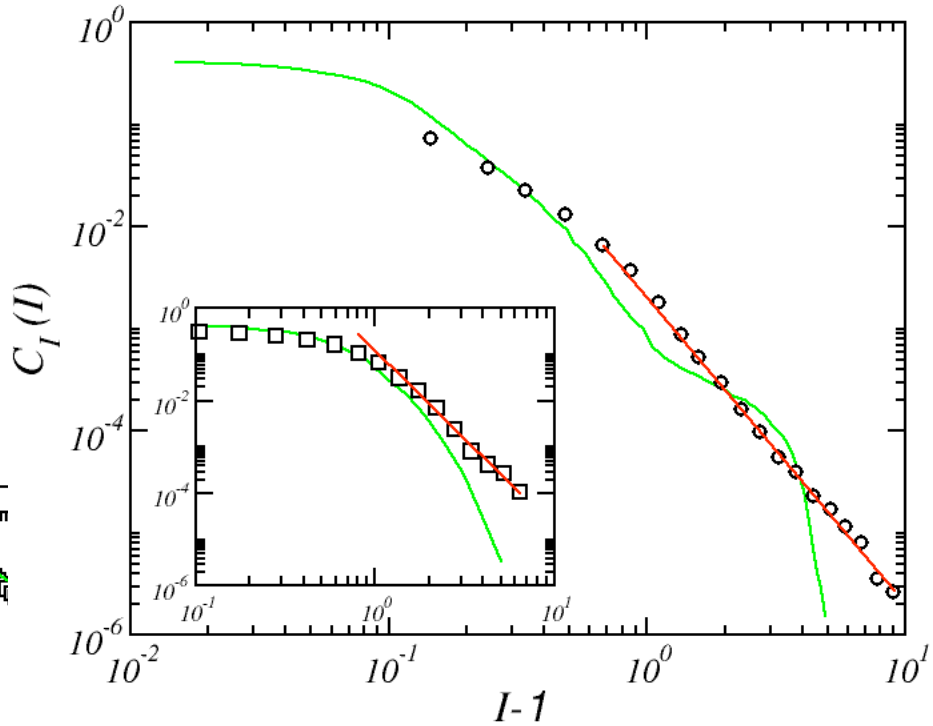
The rest $m-n$ are selected from the pool of old agents

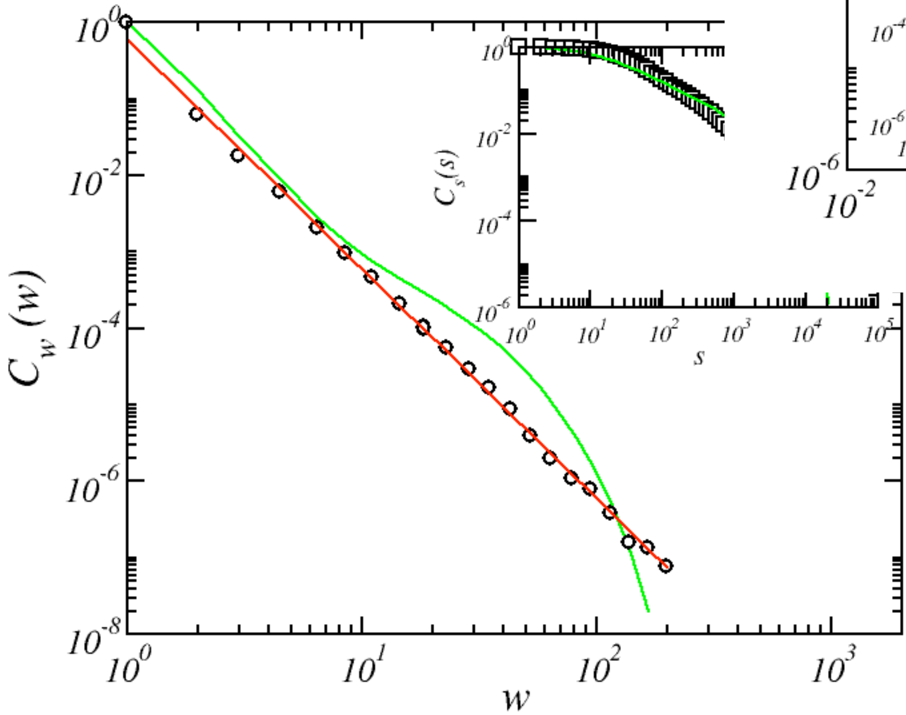
- Prob p → one of the previous partners of an "old" agent is chosen
- Prob 1-p → an "old" agent is chosen with prob proportional to q

After Q_c collaborations, the agents have a prob. $1/\tau$ of becoming inactive.

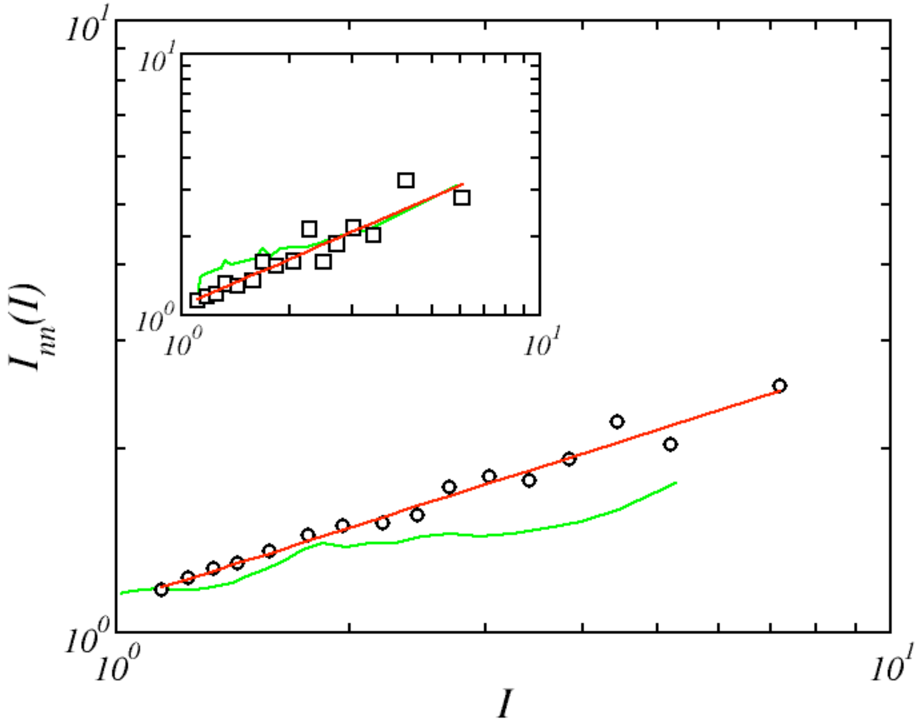
J.J. Ramasco *et al.*, PRE **70**, 036106 (2004).
R. Guimarà *et al.*, Science **308**, 697 (2005).

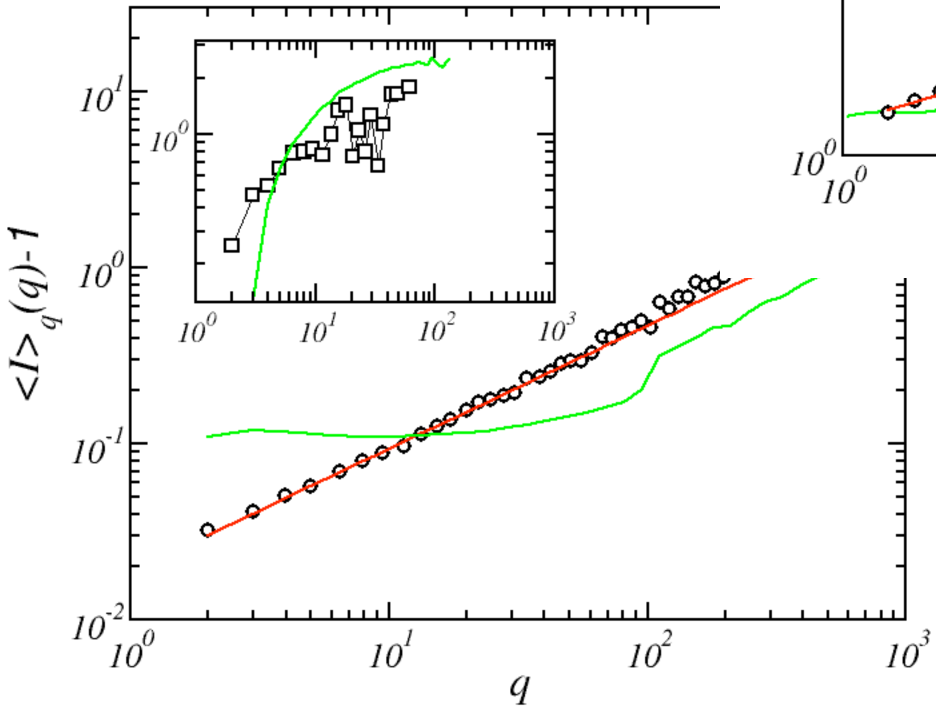
Simulation results



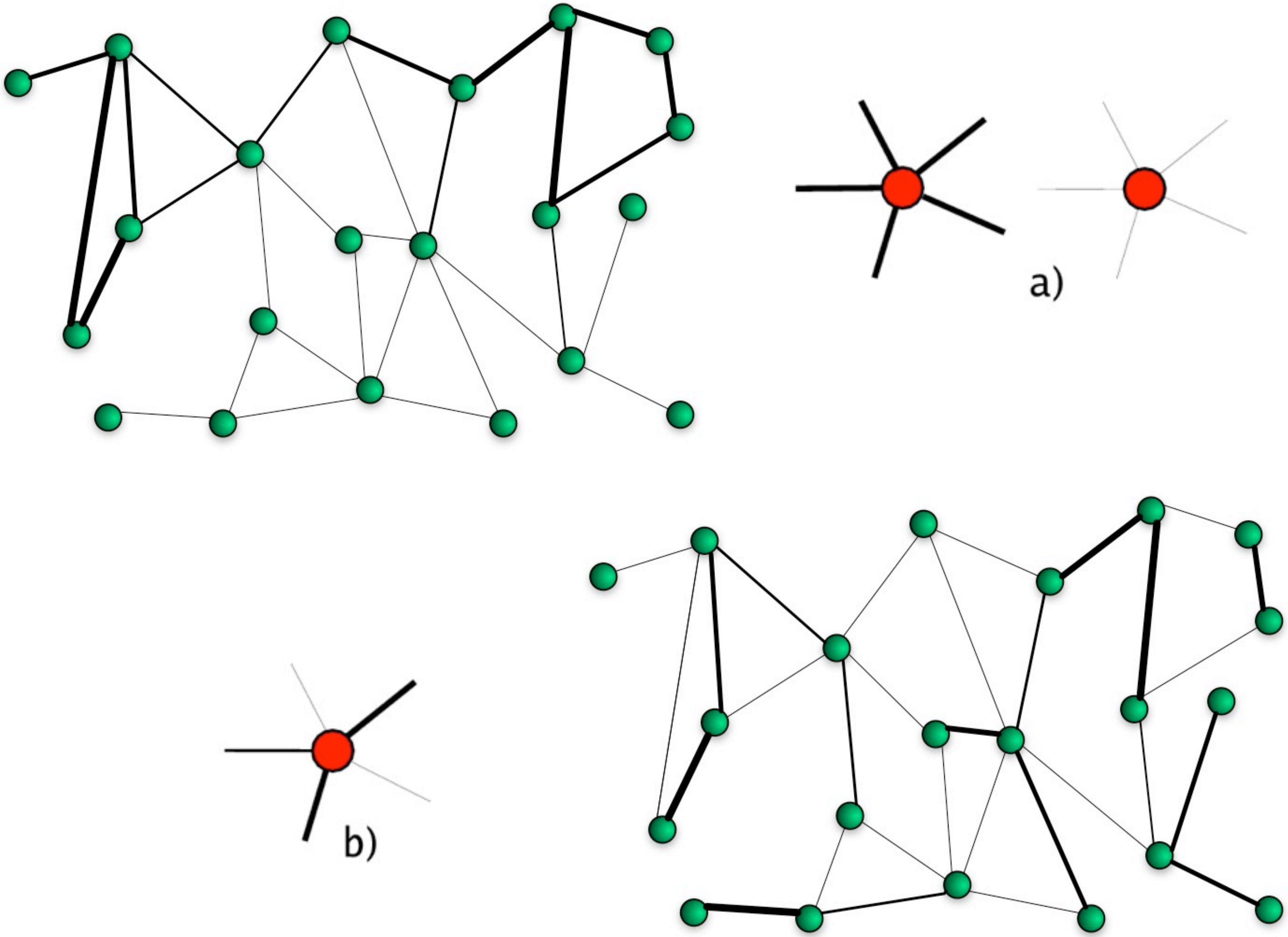


Simulation results



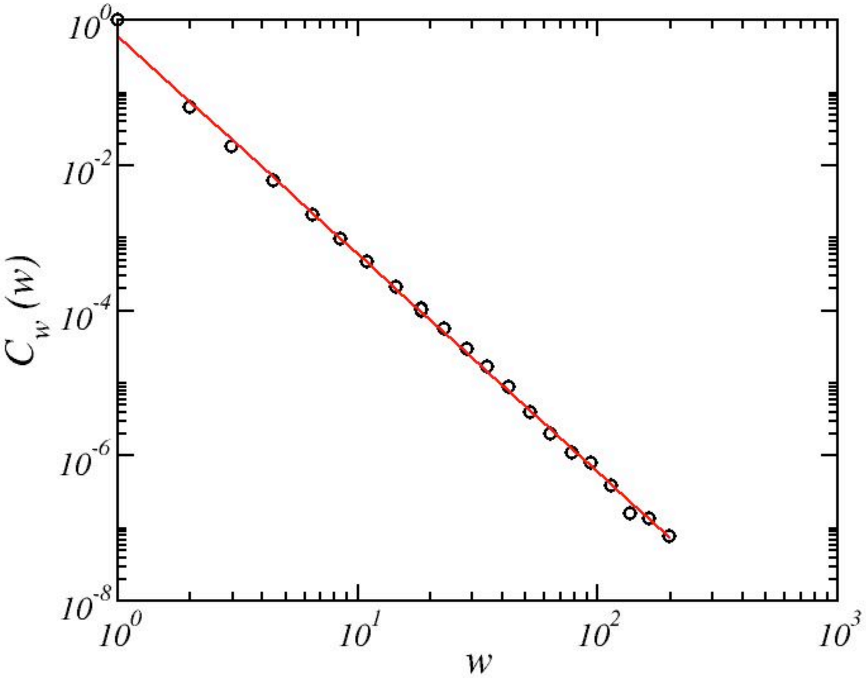


Range, different quality of the connections



Motivation

- IMDB Actor network, w_{im} = number of times *i* and *m* have worked together.



$$P_w(w) \sim w^{-\delta}$$

$$C_w = \int_w^\infty P(w')dw' \sim w^{1-\delta}$$

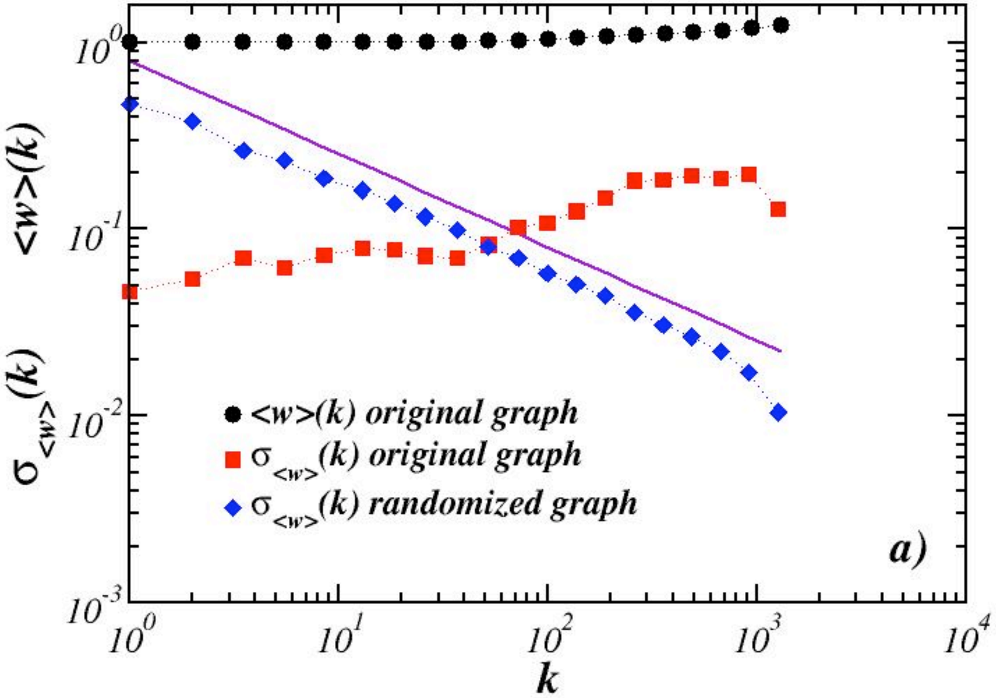
$$\text{fit } \delta = -4$$

Motivation

- P(w) has finite second moment.
- <w> does not depend on k
- Are the weight correlated?

$$s_i = \sum_{j \in \nu(i)} w_{ij} = w_{i\nu_1} + w_{i\nu_2} + \ldots + w_{i\nu_{k_i}}$$

$$\sigma_s(k) \sim k^{1/2} \Rightarrow \sigma_{<w>} \sim k^{-1/2}$$



(J.J. Ramasco, Eur. Phys. J. ST **143**, 47 (2007))

Models

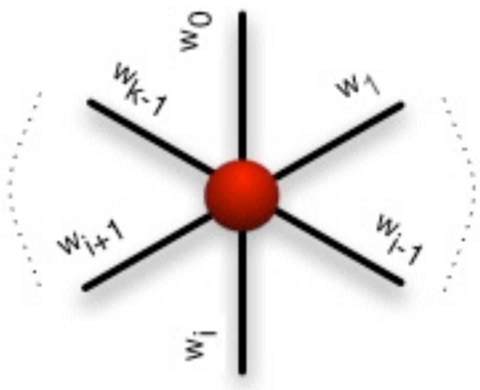
- We can try to replicate this situation in a toy model
- The simplest case requires P(w,w'), P(w) and P(w'|w)

and also a sets of rules for weight assignment.

- This is not a unique method.

$$P(w) = \int P(w,w')dw'$$

$$P(w'|w) = P(w,w')/P(w)$$



Models

- We chose three possible functional forms for P(w,w')

$$P_+(w,w') = \frac{X_+}{(w+w')^{2+\alpha}}$$

$$P_U(w,w') = \frac{X_U}{(ww')^{1+\alpha}}$$

$$P_-(w,w') = \frac{X_-}{(ww'+1)^{1+\alpha}}$$

- The reason was that

$$P(w) \sim w^{-1-\alpha}$$

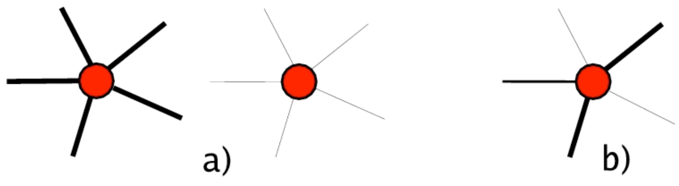
$$< w >_+ (w_0) = (1+\alpha+w_0)/\alpha$$

$$< w >_U = \alpha/(\alpha-1)$$

$$< w >_- (w_0) = (\alpha+1/w_0)/(\alpha-1)$$

Measures

- The goal is to find a measure to estimate the type/intensity of weight correlations and their intensity.
- P(w) is **fixed**.
- Compare actual pattern with a random configuration



$$\sigma_w^{\,2}(i) = \sum_j (w_{ij} - < w > (i))^2 \qquad \Sigma = \frac{< \sigma_w >_{original}}{< \sigma_w >_{random}}$$

$$Y_2(i) = \frac{\sum_j w_{ij}^{\,2}}{(\sum_j w_{ij})^2} \quad , \quad Y_2(k) \sim \begin{cases} cons. \\ 1/k \end{cases} \qquad \mathrm{Y} = \frac{< Y_2 >_{original}}{< Y_2 >_{random}}$$

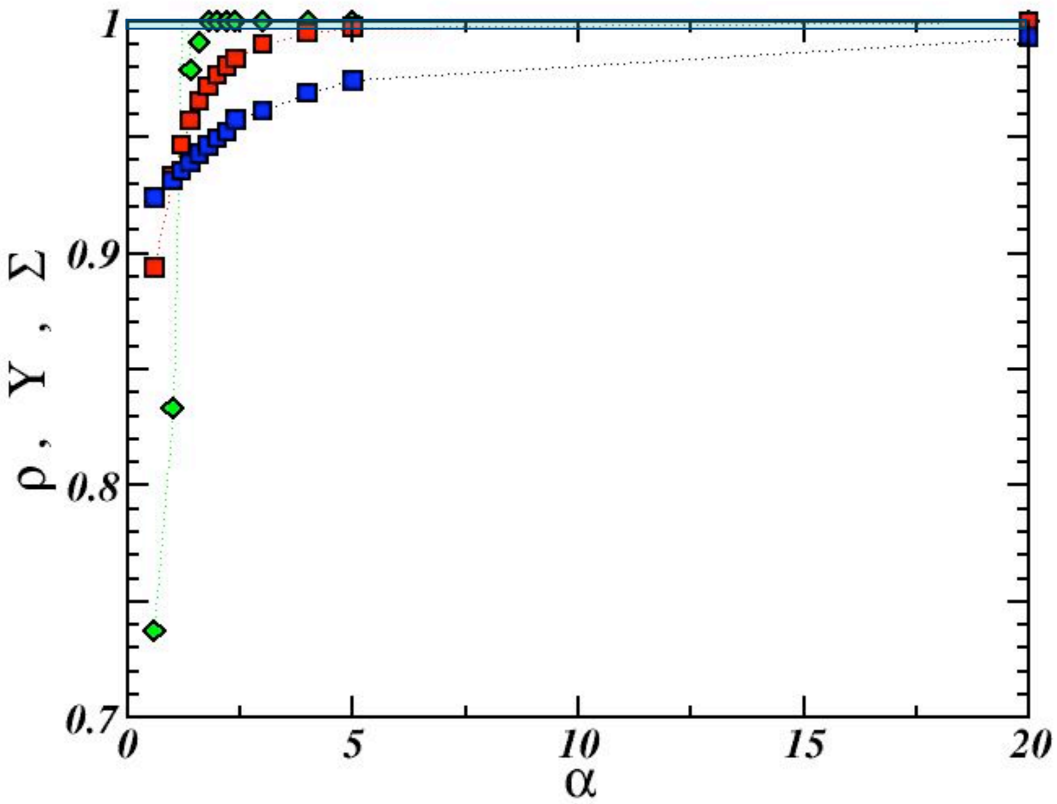
$$r(i) = \frac{w_{\max}(i) - w_{\min}(i)}{w_{\max}(i) + w_{\min}(i)} \qquad \rho = \frac{< r >_{original}}{< r >_{random}}$$

M. Barthelemy, Physica A **319**, 633 (2003).
J.J. Ramasco *et al.*, to appear PRE.

Measures

• The three magnitudes are able to detect up to some level weight correlations

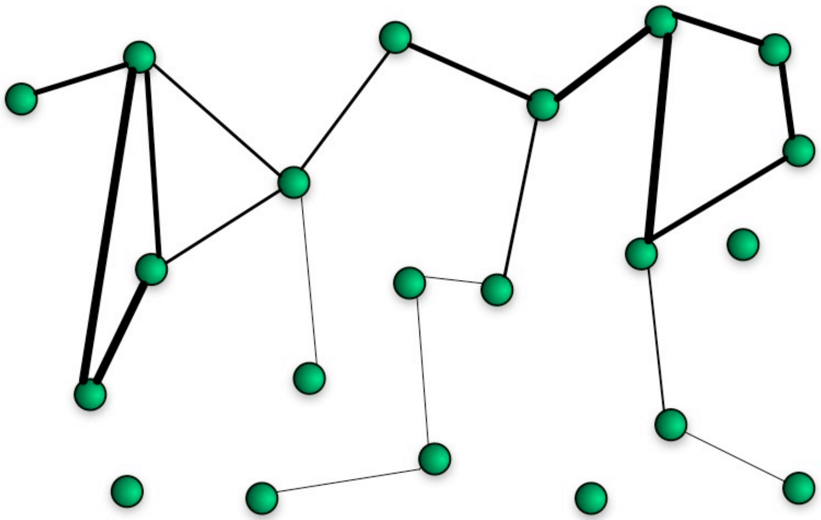
• But there is a resolution problem

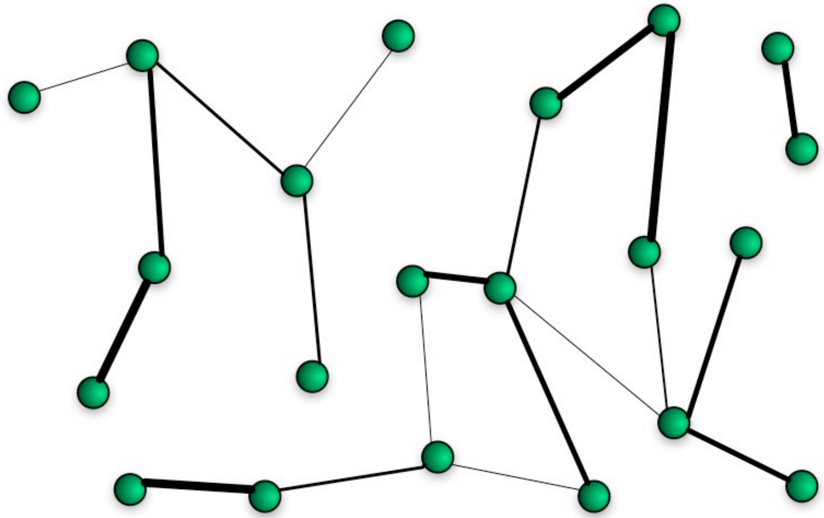


Optimal measure?

Transport

- There are many ways to describe network transport properties
- We focus on the so called Superhighways



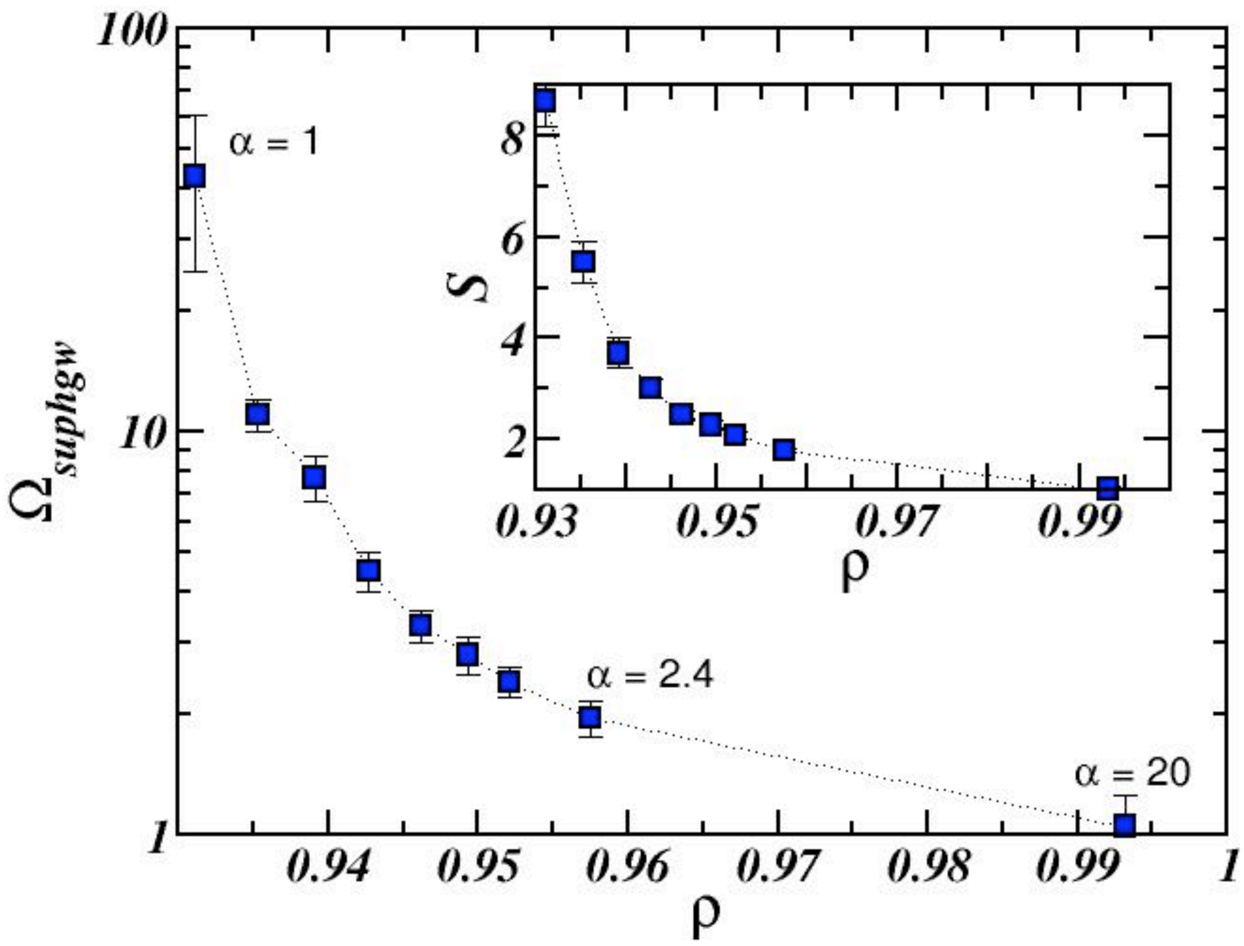


$$\Omega_{sphw} = \frac{W_{sph}(orig)}{W_{sph}(rand)}$$

$$S_{sphw} = \frac{\text{Num. nodes sphw (orig)}}{\text{Num. nodes sphw (rand)}}$$

Z. Wu et al., PRL **96**, 148702 (2006).

Transport



- Actor network

  ρ ≈ 0.268(1)

  Ω ≈ 20(7)

  S ≈ 3(1)

- US airport traffic

  ρ ≈ 0.983(1)

  Ω ≈ 2.6(1)

Conclusions

- We have studied collaborations networks using a weighted graph representation.
- This representation allows us to define the Social Inertia in a natural way.
- The Inertia (measured in a quantitative way) grows with the experience and the network is assortative for the inertia.
- The model is able to reproduce some of the quantitative features of the empirical networks but it is necessary more detail for a better quantitative result.

Conclusions

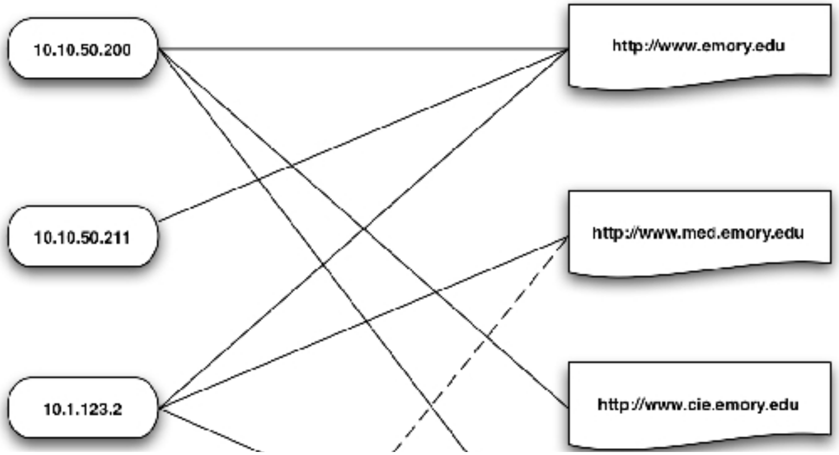
- Weight correlations appear in real-world networks
- We have propose a measures to quantify the level of correlations
- These correlations have a strong effect on the transport properties of the graph
- Open questions:
  - Origin of the phenomenon in real nets.
  - Effect on disease spreading.

J.J. Ramasco & B. Gonçalves, to appear in PRE.

Web Surfing

• The database is formed by the weblogs of Emory University from Apr. 1st 2005 to Jan. 17th 2006 (41 weeks).

• Each click in a web of the university is registered at the time resolution of 1 second.



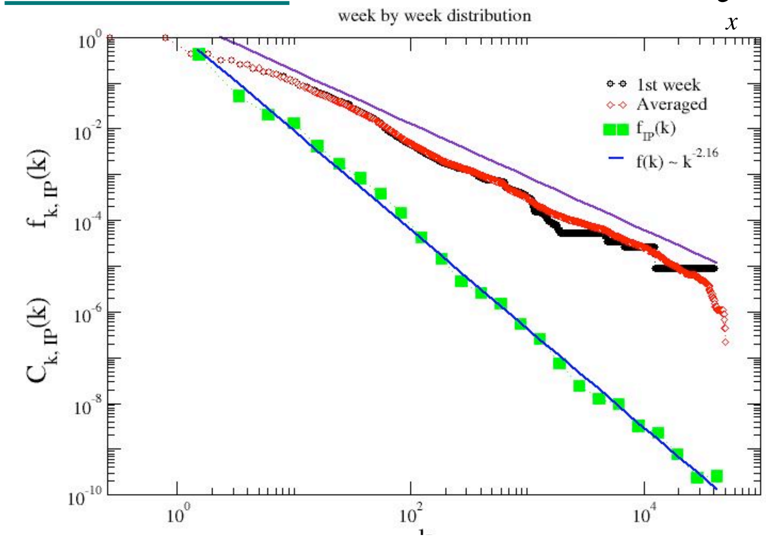
Number of IPs	N_{IP}	3,179,671
Number of URLs	N_{URL}	2,562,398
Total Number of page requests (weight)	Ω	53,582,121
Average number of IPs introduced per day	n_{IP}	10,742
Average number of URLs introduced per day	n_{URL}	8,396
Average number of edges introduced per day	e	77,569
Average wieght increment per day	$\Omega^{\dagger}$	186,350

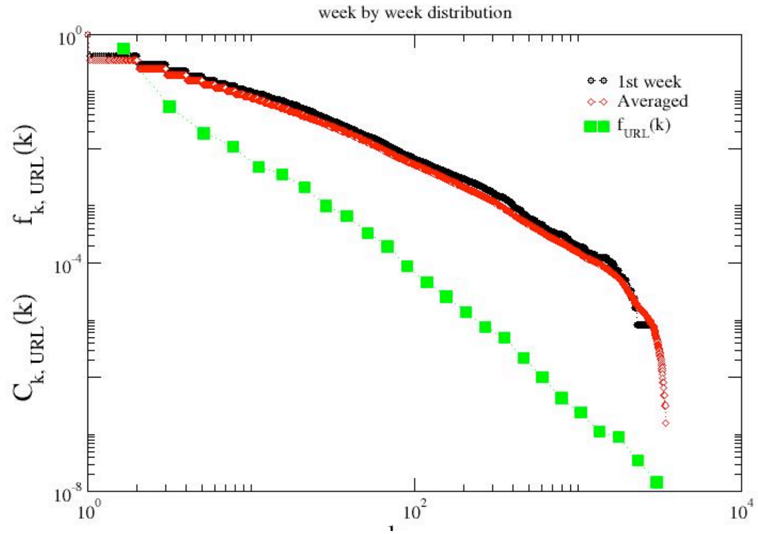
Statistics

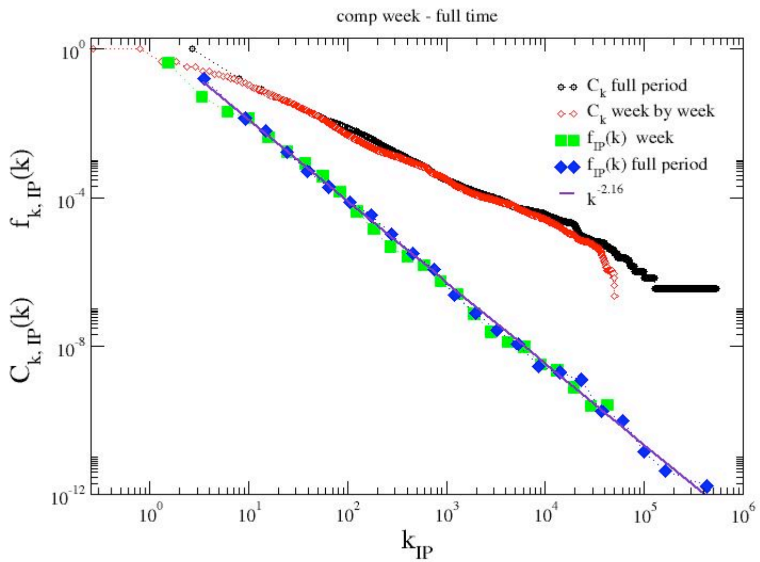
IPs

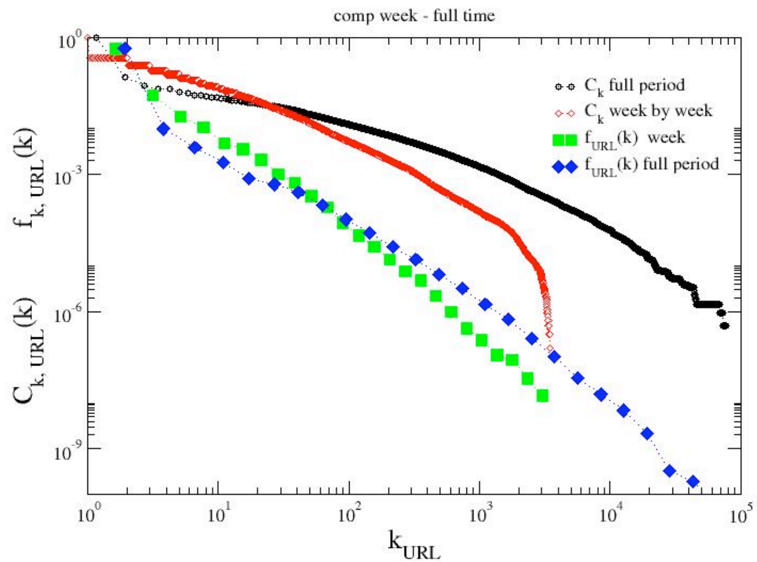
$$C_x = \int_x^{\infty} f_x(x')dx'$$

URLs

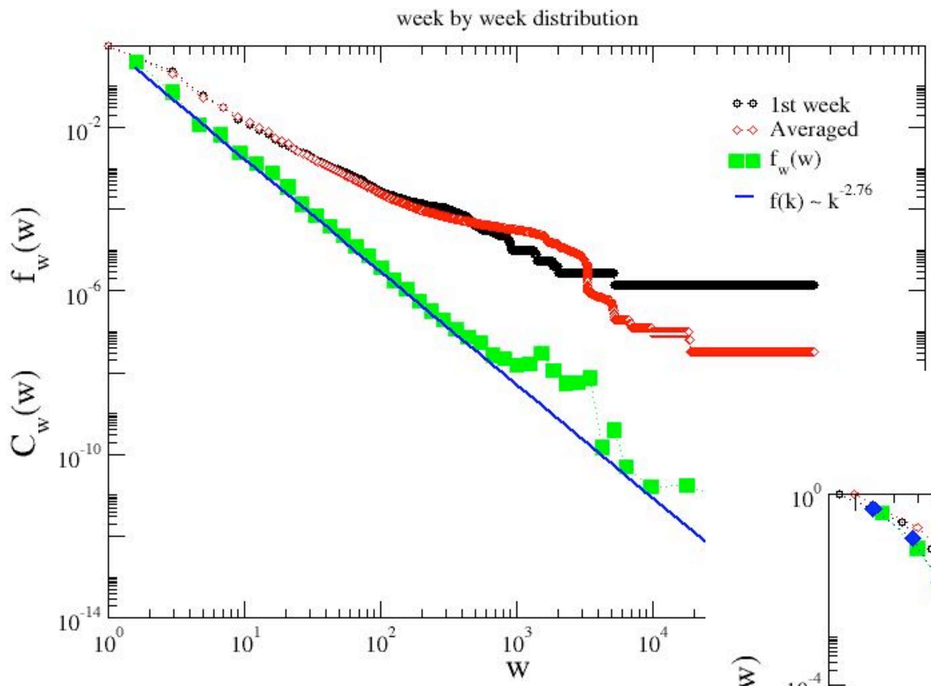




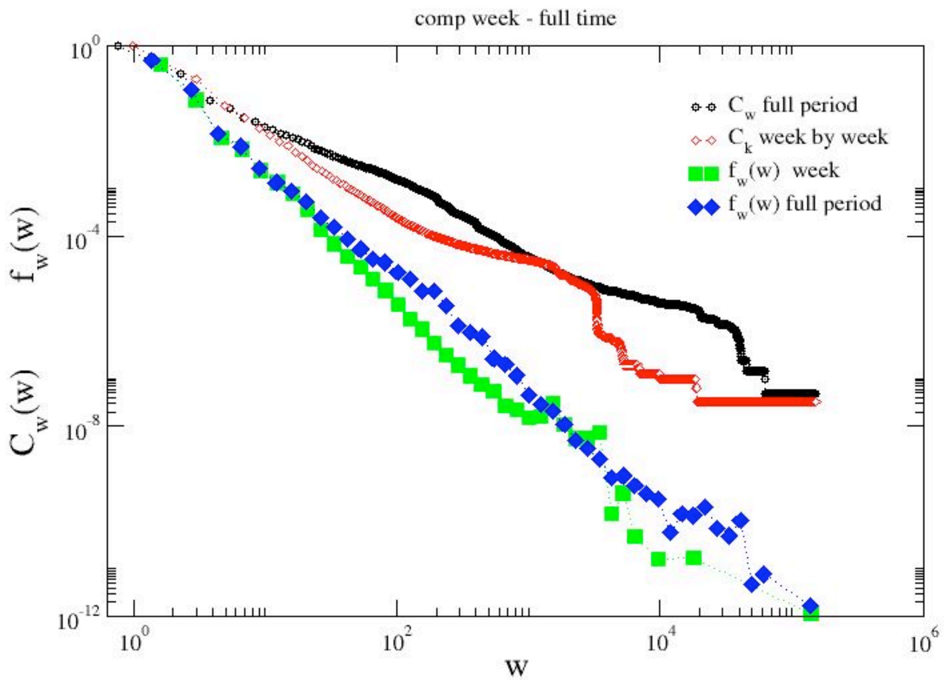




Statistics



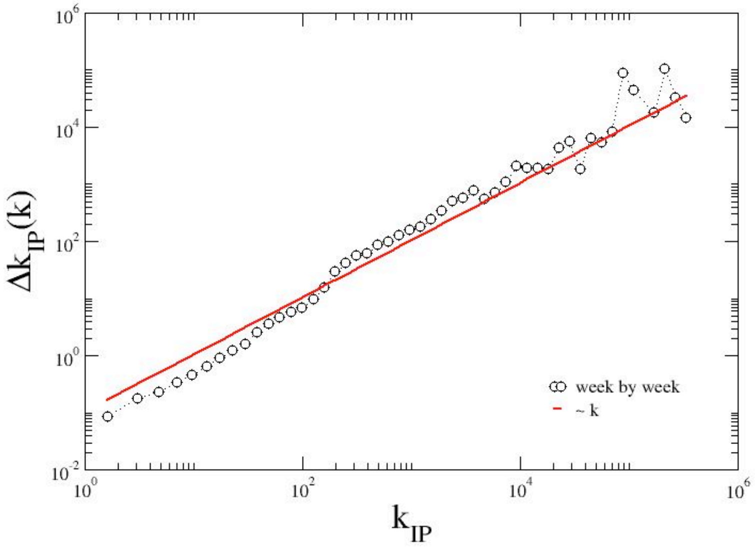
week by week distribution
1st week
Averaged
$f_w(w)$
$f(k) \sim k^{-2.76}$
$f_w(w)$
$C_w(w)$
w



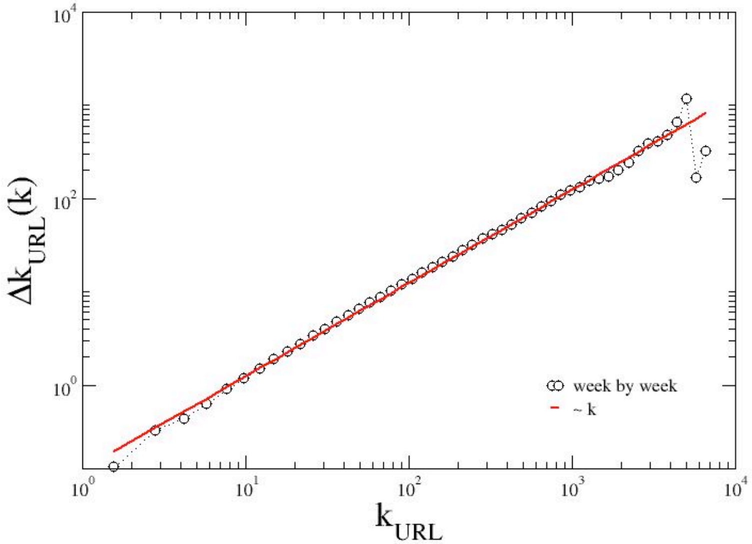
comp week - full time
C_w full period
C_k week by week
$f_w(w)$ week
$f_w(w)$ full period
$f_w(w)$
$C_w(w)$
w

Growth

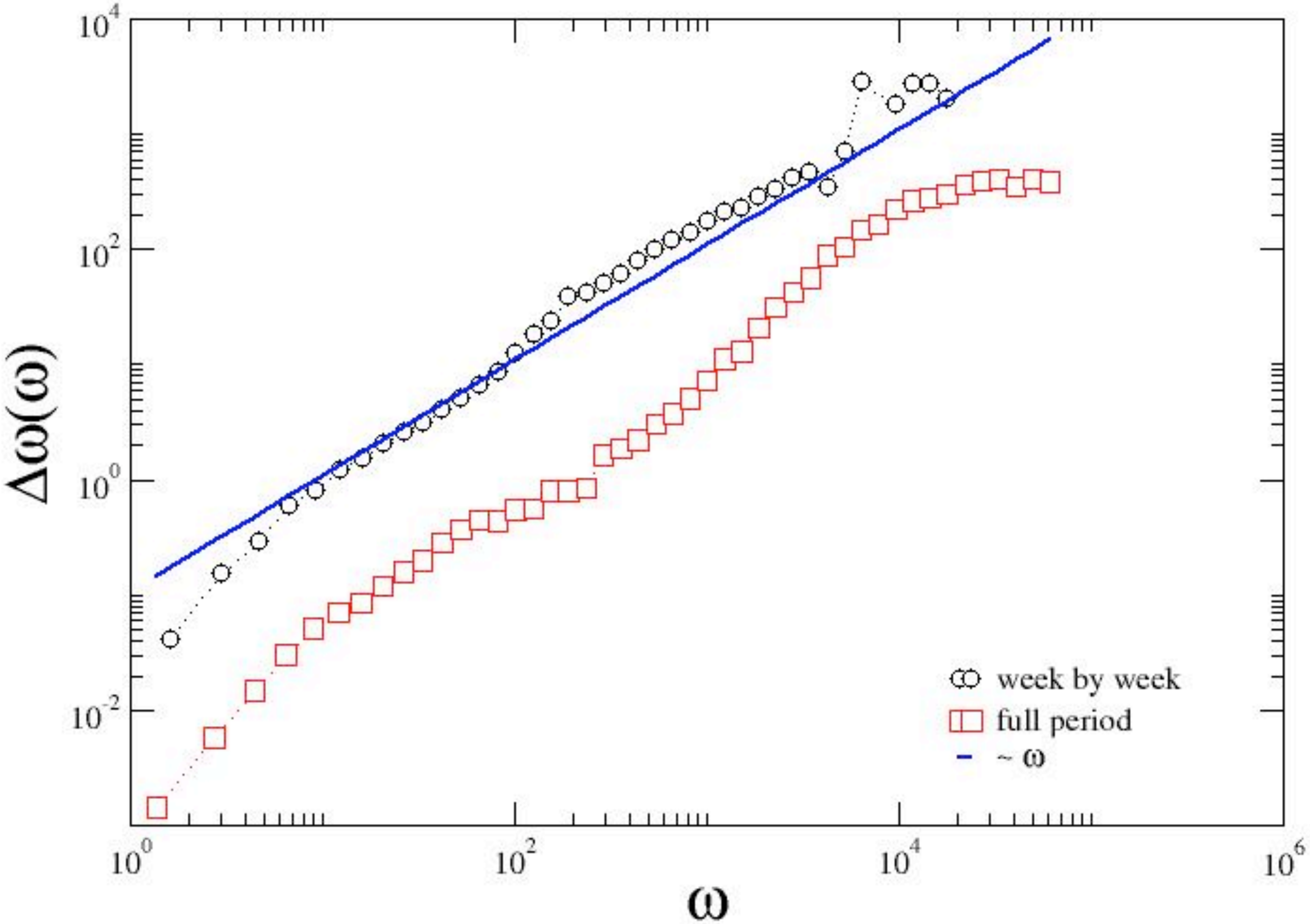
IPs



URLs

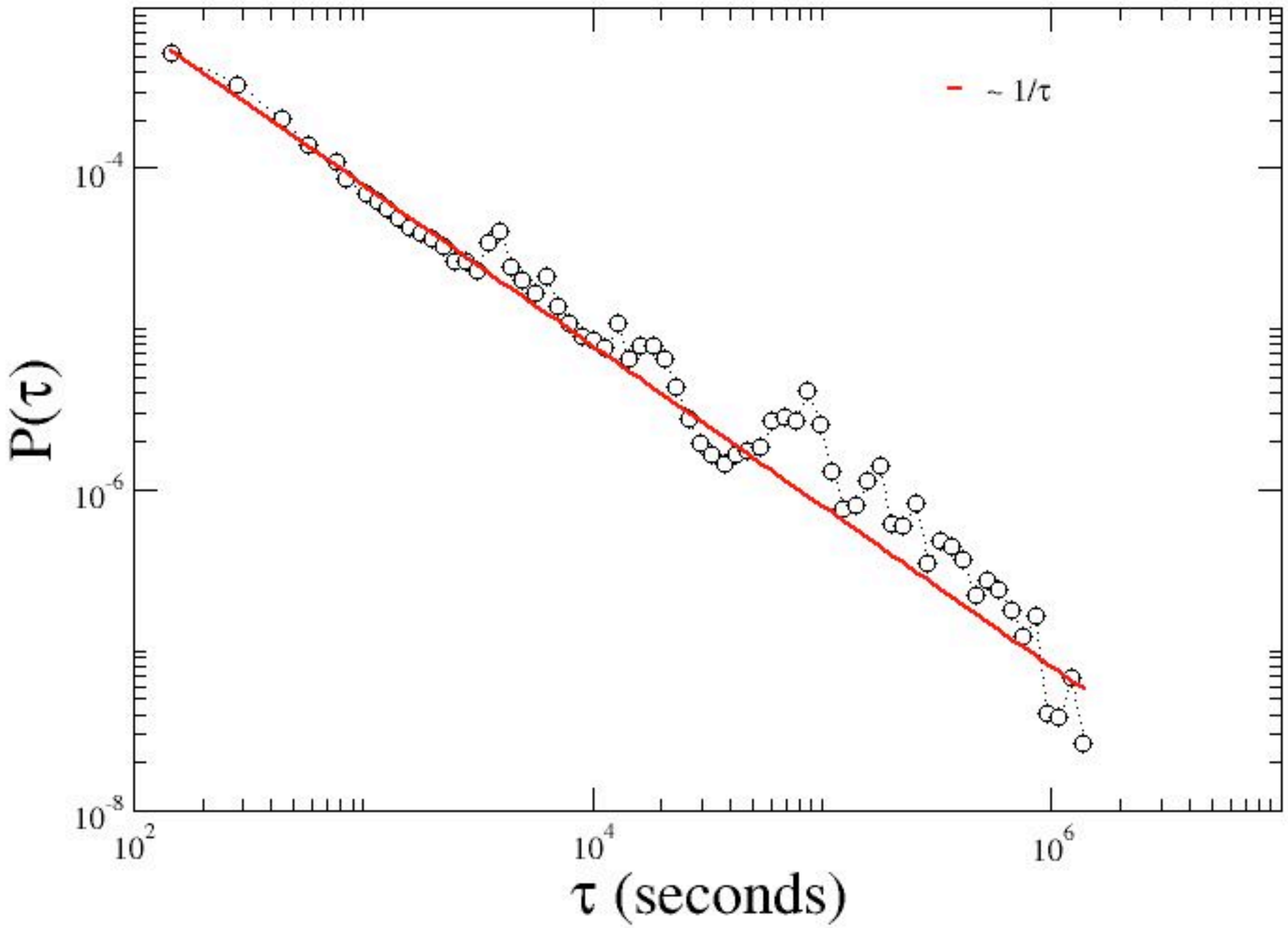


Growth

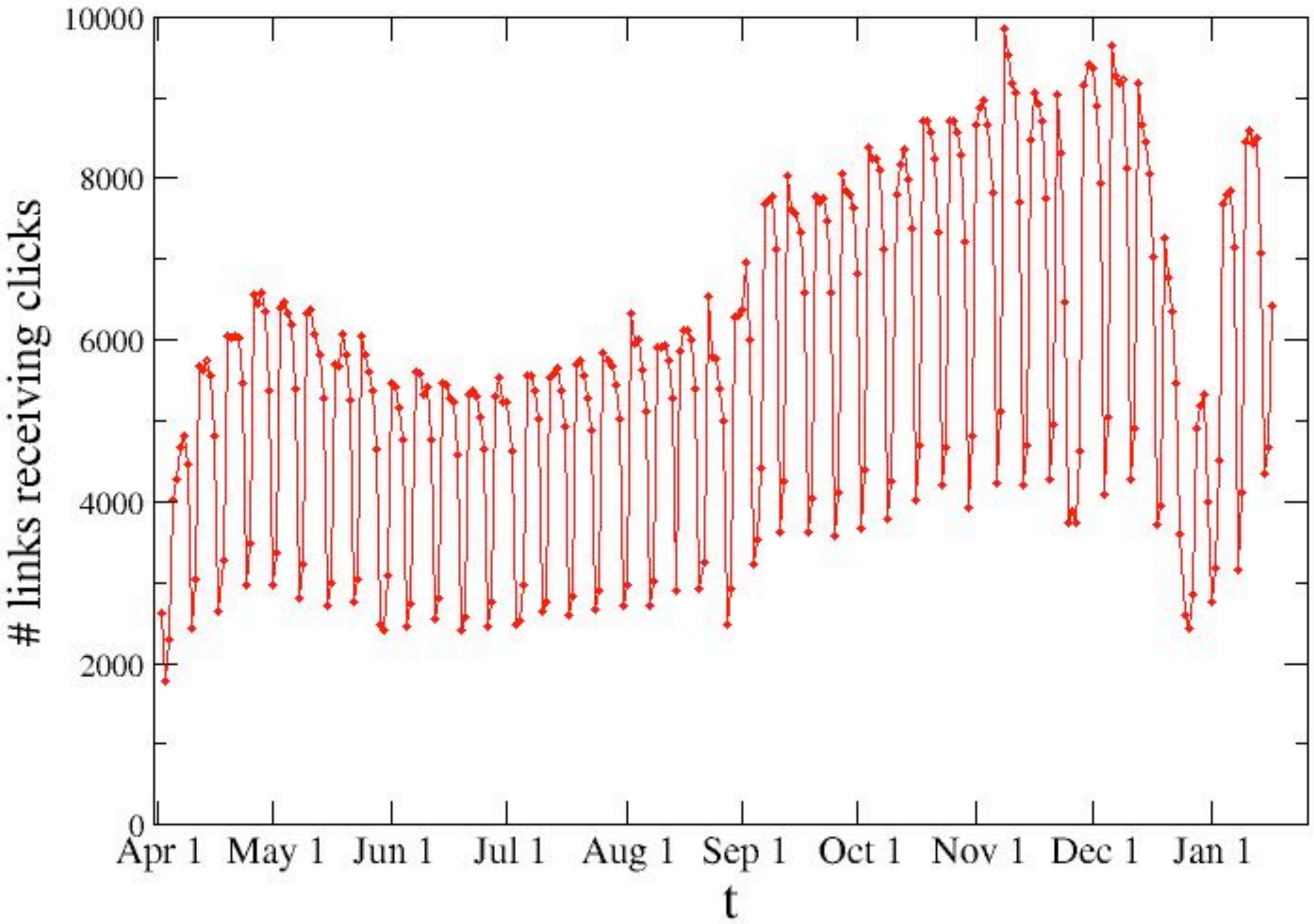


10^4
10^2
10^0
10^{-2}
$\Delta\omega(\omega)$
10^0
10^2
10^4
10^6
ω
week by week
full period
~ ω

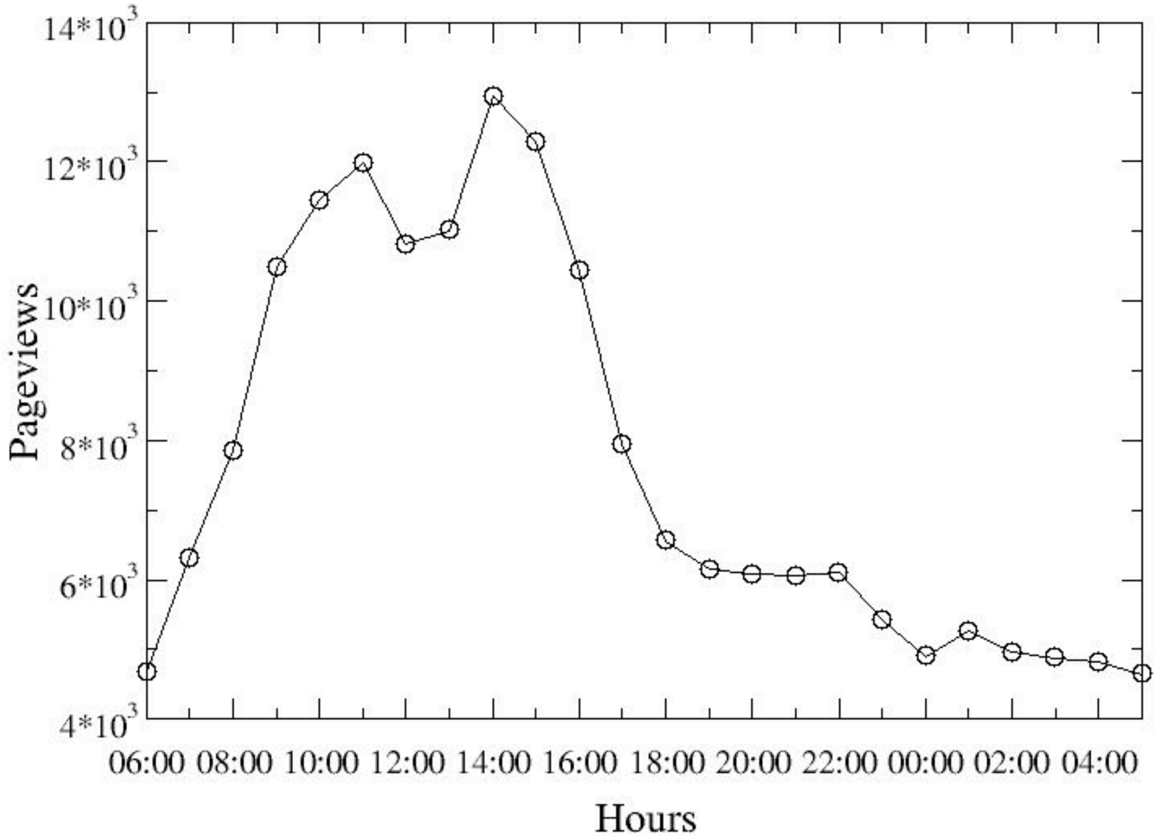
Other aspects



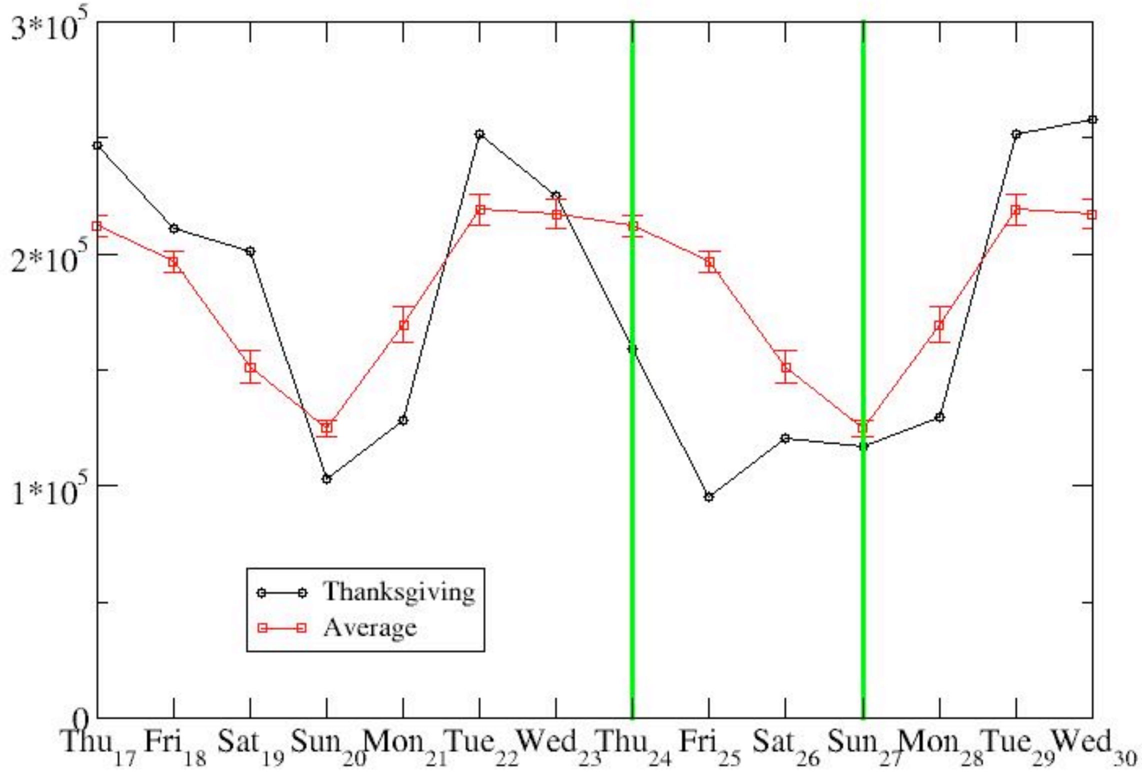
Other aspects



Other aspects



Other aspects



Conclusions

- We have studied an empirical database generating an evolving bipartite graph.
- Preferential attachment plays an important role in the evolution of the weights but so does aging of the connections.
- These data allow us to consider also the activity patterns of the community.
- Open questions …

B Gonçalves & JJ Ramasco, in preparation