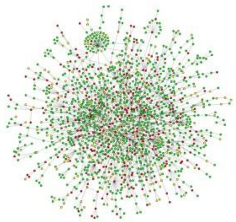
Applications of Complex Networks

Javier M. Buldú
http://www.complex.etsit.urjc.es/jmbuldu
Complex Systems Group
Universidad Rey Juan Carlos (Madrid, Spain)



CSG
Universidad Rey Juan Carlos

2.- Applications of Complex Networks

2.1.- Social Networks

Universidad Rey Juan Carlos

Social Networks:

General properties (if they exist!):

- Social networks show the small-world property.
- They are prone to exhibit power-law distribution and, as a consequence, the existence of highly connected nodes (hubs).
- The clustering coefficient is high (when compared with random networks).
- It is common to observe assortative mixing (i.e., most connected nodes like to be together)

CSG
Universidad Rey Juan Carlos

Vilfredo Pareto (1848 - 1923)

- Italian industrialist, sociologist, economist, and philosopher, he made several important contributions to economics, particularly in the study of income distribution.

- He analyzed the distribution of wealth (income, lands,…) in different countries, observing that all of them had the same distribution, a power law **$p(X>=x) \sim x^{-\beta}$**

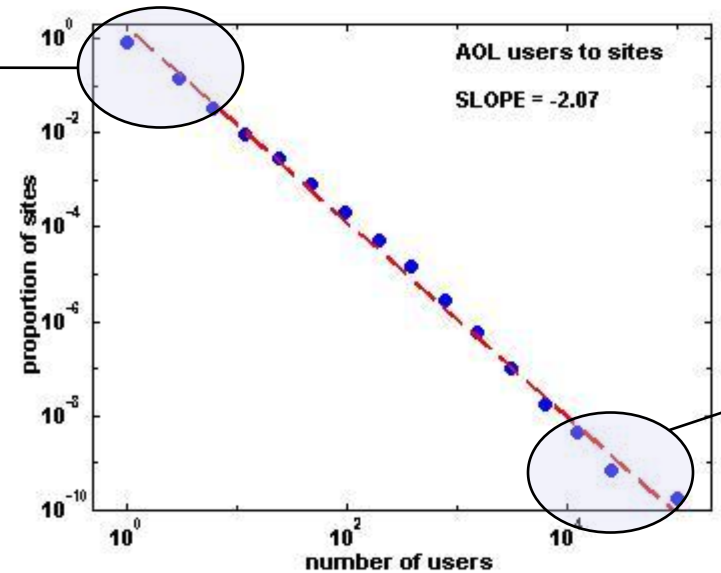
- The Pareto principle (also known as the 80-20 rule) states that, for many events, roughly 80% of the effects come from 20% of the causes.

- The Pareto principle has been applied in different fields, from economics to sociology.

❑ The 80-20 rule and the power law distribution

The 80-20 rule is a consequence of the power law (cumulative) distribution:

Most sites have a low number of visits



Few sites acumulate a huge number of visits

AOL users' visits to various sites on a December day in 1997. From L.A. Adamic

CSG Universidad Rey Juan Carlos

2.1.1.- Introduction

❑ George Kingsley Zipf, (1902-1950)

❑ An American linguist and philologist who studied statistical occurrences in different languages.

❑ He observed that the frequency of appearance of words in English language followed a ranking-distribution with a power-law shape:

$$N(r) \sim r^{-\gamma}$$

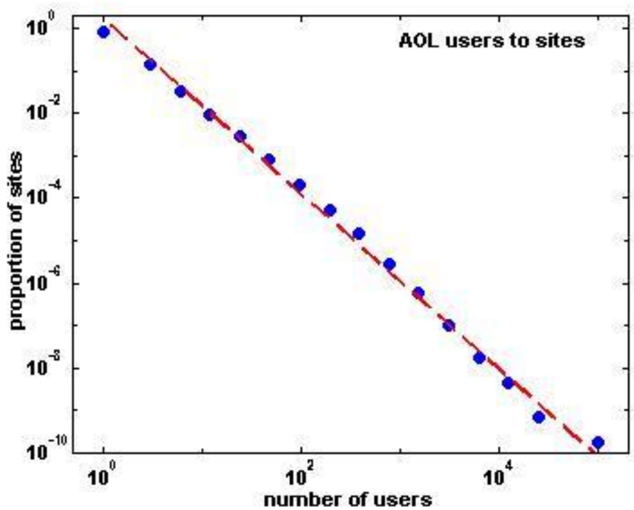
where *n* is the number of ocurrences and *r* is the ranking in the frequency distribution.

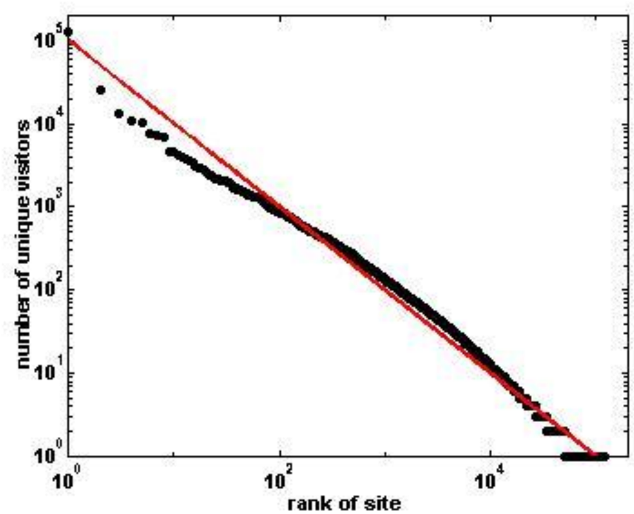
❑ The Zipfian distribution has been reported in many different fields such as linguistics, population distribution, income rankings, etc...

2.1.1.- Introduction

- Pareto and Zipf distribution are closely related

"The r^{th} largest city has n inhabitants" (Zipf) is equivalent to saying *"r cities have n or more inhabitants“ (Pareto)*





Pareto Distribution:

$r(X>=n) \sim n^{-\beta}$

Zipfian Distribution:

$n(r) \sim r^{-\gamma}$

Correspondence:

$\beta = 1/\gamma$

AOL users' visits to various sites on a December day in 1997.
From L.A. Adamic, *Zipf, Power-laws, and Pareto - a ranking tutorial* .

Universidad Rey Juan Carlos

❑ An explanation of the power-laws

- ❑ **Simon (1955)** explains the appearance of power law distributions as a consequence of the *"rich get richer"* phenomenon. [Simon, Biometrika, 42, 425 (1955)].

- ❑ **Price (1965),** gives the first theoretical model that explains the growth of a network with power-law distribution, in this case, the probability of citation in scientific papers. [Price, Science, 149, 510 (1965)]. Introduces the concept of ***cumulative advantage***.

- ❑ **Barabási-Albert (1999)** rediscovered the results of Price and introduced the concept of ***preferential attachment***. [Barabasi et al., Science, 286, 509 (1999)].

- ❑ Despite the BA model had been initially used to descibe the power-law distribution of scale-free networks, nowadays it is accepted that **more complete models** (or even different) may be behind the scale-free structure of several social networks.

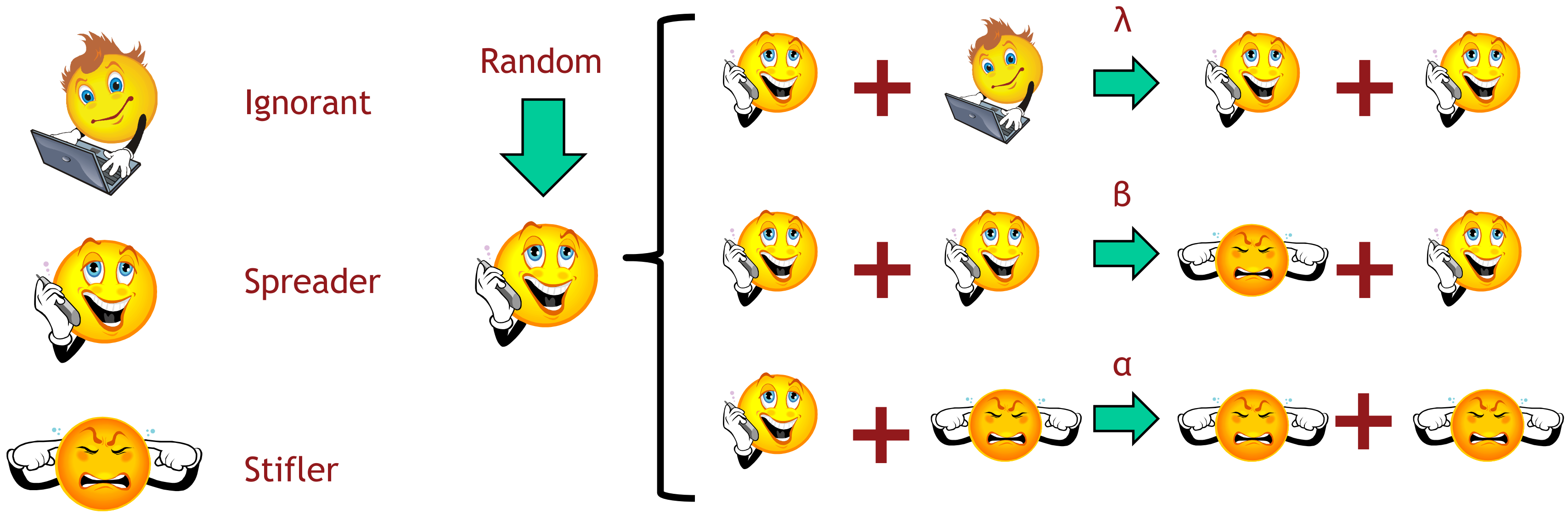
2.1.2.- Rumor/disease spreading

2.1.2.- Rumor/disease spreading

❑ How does a rumour spread (The Daley and Kendal model)?

- ❑ The Ignorant-Spreader-Stifler model (DK model):

At each time step a random Spreader *i* is selected and contacts one of its *j* neighbours. Then, (a) if *j* is Ignorant it turns *j* to Spreader and (b) if *j* is a Spreader/Stifler, then *i* turns to Stifler. We will introduce it in a small-world network.



Read more at:
Zanette, PRE, 64, 050901 (2001)

❑ How does a rumour spread in a small-world network?

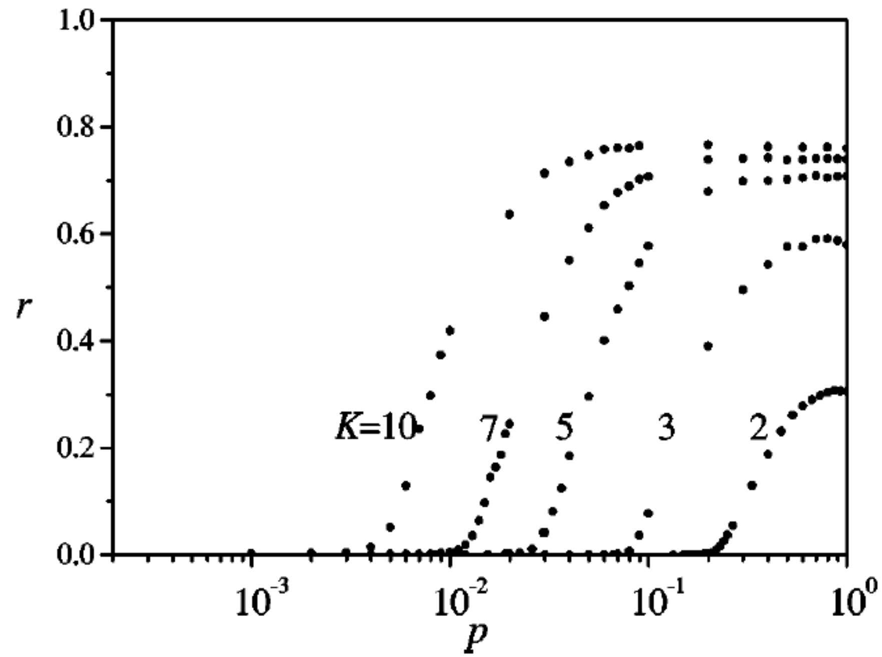
❑ The Ignorant-Spreader-Stifler model (DK model) in small-world networks:

After a certain transient, the rumor stops:

N= total number of nodes
r = $N_{stiflers}/N$
P= probability of rewiring
$\lambda = \beta = \alpha$ (for the case of the figure)

❑ Below a certain p_c the rumor does not reach a significant part of the network.

❑ Highest degree, highest probability for the onset of the rumor propagation.

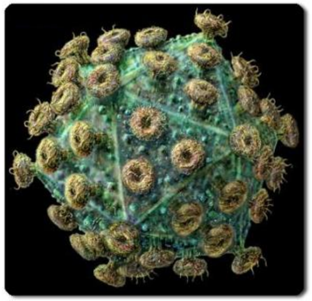


CSG
Universidad Rey Juan Carlos

Read more at:
Daley,
Epidemic Modelling (2000)

2.1.2.- Rumor/disease spreading

❑ How does a disease or a virus spread?



Among all epidemic models, two of them have been analyzed more deeply:

❑ The Susceptible-Infected-Susceptible (SIS) model (e.g., H1N1 flu)

$$S(i) + I(j) \xrightarrow{\lambda} I(i) + I(j)$$
$$I(i) \xrightarrow{\mu} S(i)$$

❑ The disease can persist indefinitely (endemic diseases).
❑ The epidemic transition, given by a critical value of the spread rate $\sigma=\lambda/\mu$, is the boundary between the persistance or exitintion.

❑ The Susceptible-Infected-Removed (SIR) model (e.g., measles)

$$S(i) + I(j) \xrightarrow{\lambda} I(i) + I(j)$$
$$I(i) \xrightarrow{\mu} R(i)$$

❑ The disease always disappears after a certain transient
❑ The epidemic transition, given by σ_c, is the boundary between affecting a significant amount of individuals or not.

❑ Homogeneous mixing in SIR

In the homogeneous mixing, individuals contact with other individuals at random (within the whole population).

$$\frac{\mathrm{d}s(t)}{\mathrm{d}t} = -\lambda\bar{k}\rho(t)s(t)$$

$$\frac{\mathrm{d}\rho(t)}{\mathrm{d}t} = -\mu\rho(t) + \lambda\bar{k}\rho(t)s(t)$$

$$\frac{\mathrm{d}r(t)}{\mathrm{d}t} = \mu\rho(t)$$

Where s(t), ρ(t) and r(t) are, respectively the fraction of susceptible, infected and removed. λ is the rate of infection, μ is the recovery rate, and κ is the number of contacts per time unit.

Initial condition

$$s(0) \simeq 1, \rho(0) \simeq 0, r(0) = 0$$

Number of infected individuals (epidemic incidence)

There exists a critical value of λ: $\lambda_c = \bar{k}^{-1}$

with $r_\infty = 0$ if $\lambda < \lambda_c$

and $r_\infty \sim (\lambda - \lambda_c)$ if $\lambda > \lambda_c$

Read more at:
Pastor-Satorras et al., PRE, 63, 066117 (2001)

❑ Homogeneous mixing in SIS

In the SIS model, we have one equation for the density of the infected individuals *ρ(t):*

$$\partial_t \rho(t) = -\rho(t) + \lambda \langle k \rangle \rho(t)[1-\rho(t)] + (\text{higher-order terms})$$

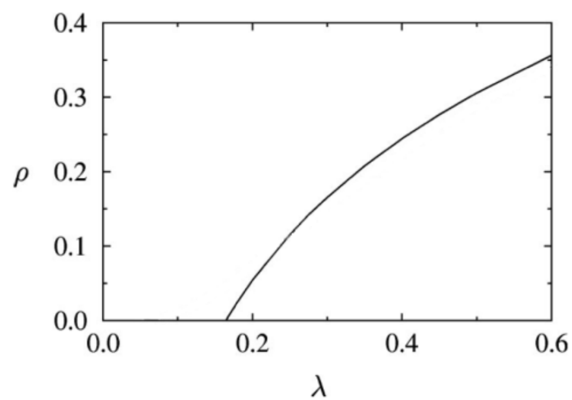
Where *ρ(t)* is the density of infected individuals, λ is the spreading rate of infection (λ=ν/δ, being ν the infecting rate and δ the recovery rate) and <κ> the number of contacts per time unit.

Number of infected individuals (epidemic incidence)

There exists a critical value of λ: $\lambda_c = \langle k \rangle^{-1}$

$$\rho = 0 \quad \text{if} \;\; \lambda < \lambda_c$$

$$\rho \sim \lambda - \lambda_c \quad \text{if} \;\; \lambda > \lambda_c$$



Density of infected nodes in a random network.
From Pastor-Satorras, PRE 63, 066117 (2001).

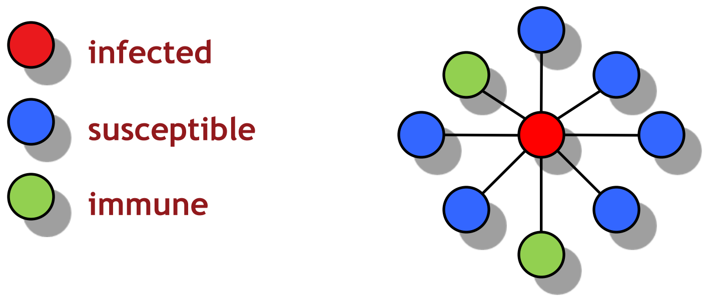
Read more at:
Pastor-Satorras, PRE, 63, 66117 (2001)

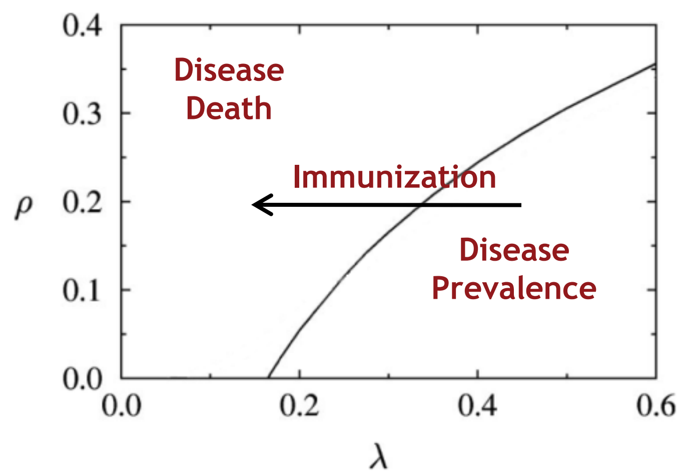
❑ How can I fight the disease?

Immunization of the population has been the traditional solution:

Random Inmunization:

λ= spreading rate
g= fraction of immune nodes

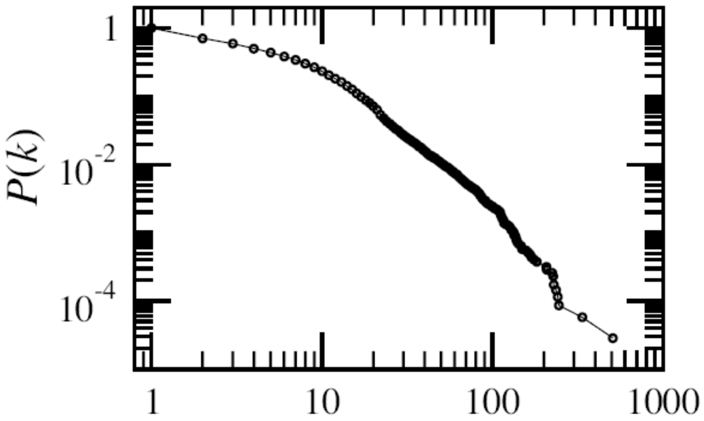




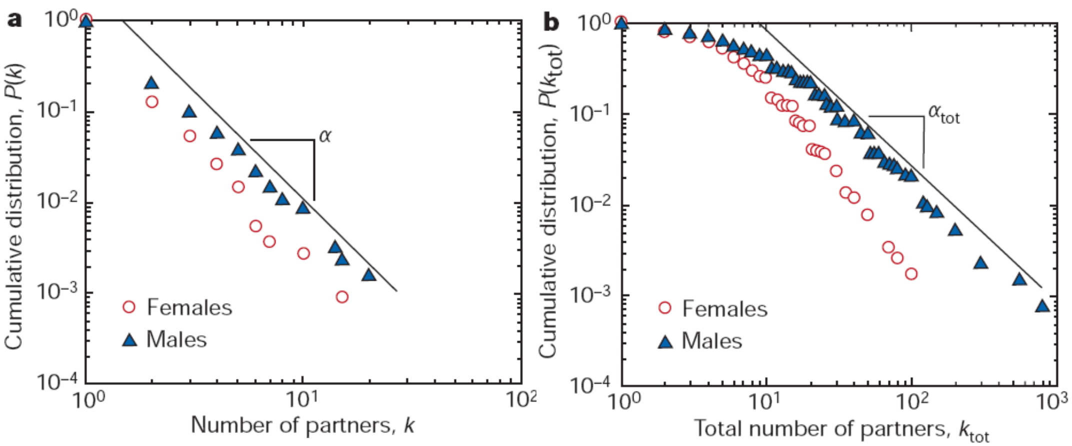
Effective Spreading Rate: $\lambda_g \rightarrow \lambda\,(1\text{-}g) \le \lambda_c$

2.1.2.- Rumor/disease spreading

❑ Social Networks are not homogeneous



Collaboration network between musical artists. From Park, IJBC, 17, 2281 (2007).



Network of sexual contacts. (a) Partners during the last year and (b) number of partners during the entire lifetime. Sample: 4781 Swedes. From Liljeros, Nature, 411, 907 (2001)

Wilt Chamberlain (NBA basketball player) wrote that he had had sex with approximately 20,000 women. (that means having sex with 1.2 women a day, every day since he was fifteen years old.) (no comments)

2.1.2.- Rumor/disease spreading

Read more at:
Pastor-Satorras et al., PRE, 65, 36104 (2002)

❑ SIS model in complex networks

Model definition:

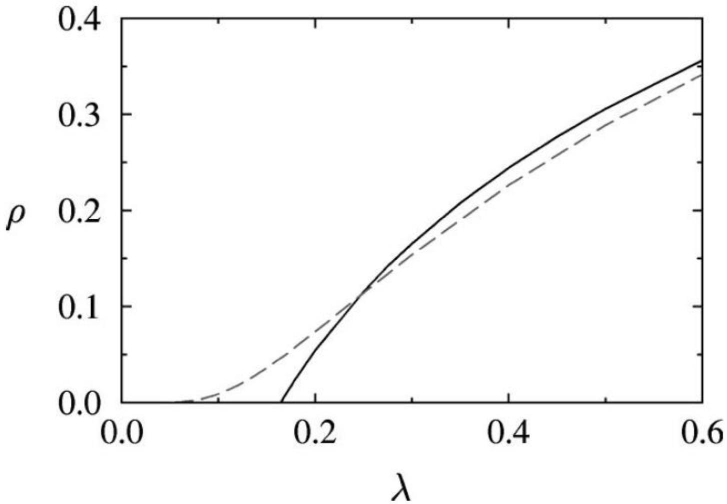
$$\partial_t \rho_k(t) = -\rho_k(t) + \lambda k[1-\rho_k(t)]\Theta(\rho(t))$$

Where $\Theta(\rho(t))$ is the probability that any given link points to an infected node, which **depends on the degree distribution**. In the steady state:

In a **BA network** we obtain:

$$\rho \sim e^{-1/m\lambda}$$

(the epidemic threshold disappears!)



Density of infected nodes in a random network (solid) and BA network (dashed). From Pastor-Satorras, PRE 63, 066117 (2001).

Universidad Rey Juan Carlos

Read more at:
Pastor-Satorras et al., PRE, 65, 36104 (2002)

❑ SIS model in complex networks

In general: $\lambda_c \sim \langle k\rangle / \langle k^2\rangle$

In a **network with equal connectivies:**

$\langle k^2\rangle = \langle k\rangle^2$

$\lambda_c = \langle k\rangle / \langle k\rangle^2 = \langle k\rangle^{-1}$

In a **BA networks:**

$P(k)=k^{-3}$

if $N \rightarrow \infty$ then $\langle k^2\rangle \rightarrow \infty$

and $\lambda_c \rightarrow 0$

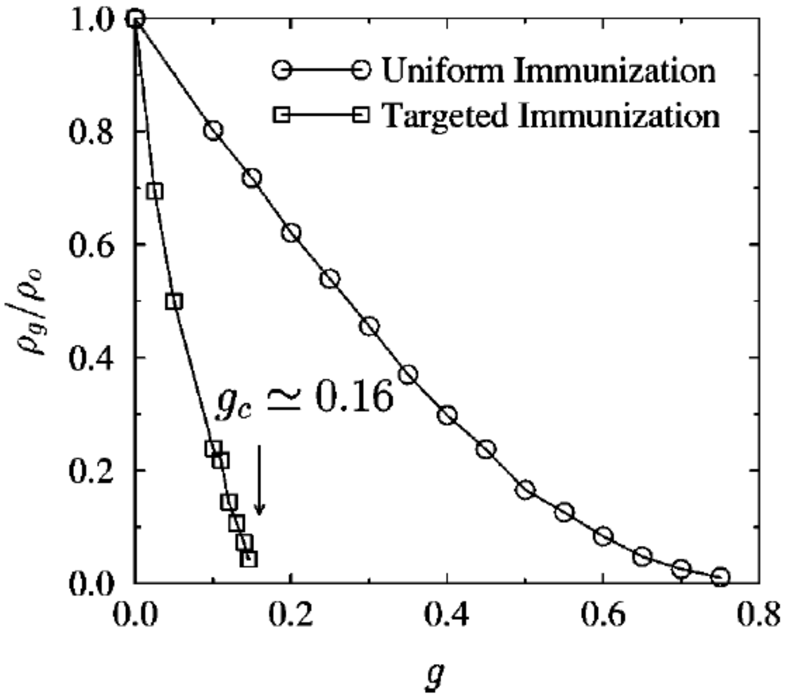
Read more at:
Pastor-Satorras et al., PRE, 65, 36104 (2002)

❑ How can I fight the disease in scale-free networks?

Targeted immunization: We select the most connected nodes as the candidates for immunization:

ρ_g/ρ_0 = Reduced prevalence

g= ratio of immune individuals



CSG

Universidad Rey Juan Carlos

2.1.3.- Communities and roles

2.1.3.- Communities and roles

❑ Communities and social networks: Zachary and the Karate club

Zachary (1977) studied the social relationships in a Karate club

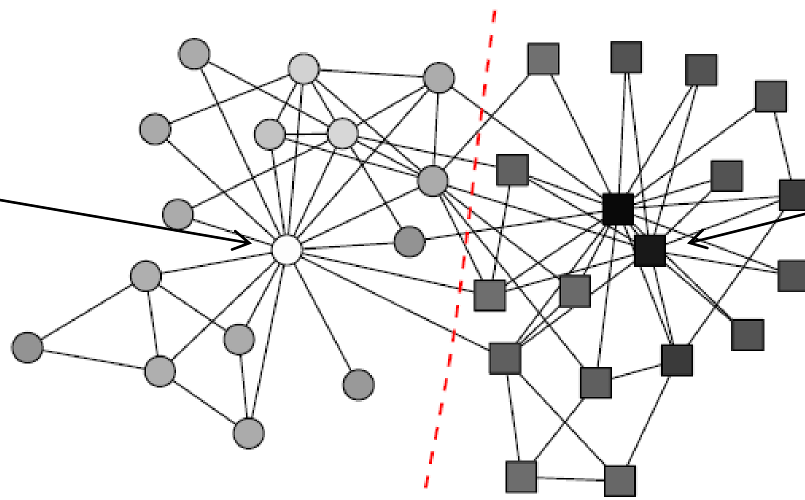
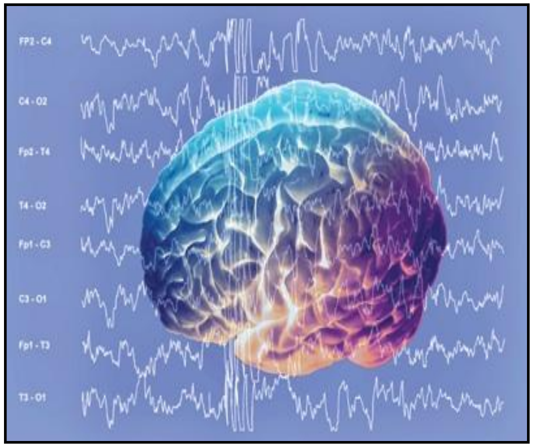


Figure from: M. E. J. Newman, Proc. Natl. Acad. Sci. USA 103, 8577 (2006)

❑ The modular structure of complex networks

- ❑ The existence of commnunities in the network is not reflected in the degree distributions, clustering, assortativity or other classical network parameters.
- ❑ Communities are related with the function of nodes in the network.
- ❑ The detection of commnunites is not a problem with a single solution.



Read more at:
Danon et al.,
World Scientific, 93 (2007)

❑ Algorithms for detecting community structures

There are several types of community detection algorithms:

A) **Link Removal methods**: Shortest path centrality, flow centrality, information centrality, random walk...

B) **Agglomerative methods**: Hierarchical clustering or L-shell methods.

C) **Modularity based algorithms**: Greedy algorithm, simulated annealing methods, extremal optimization, ...

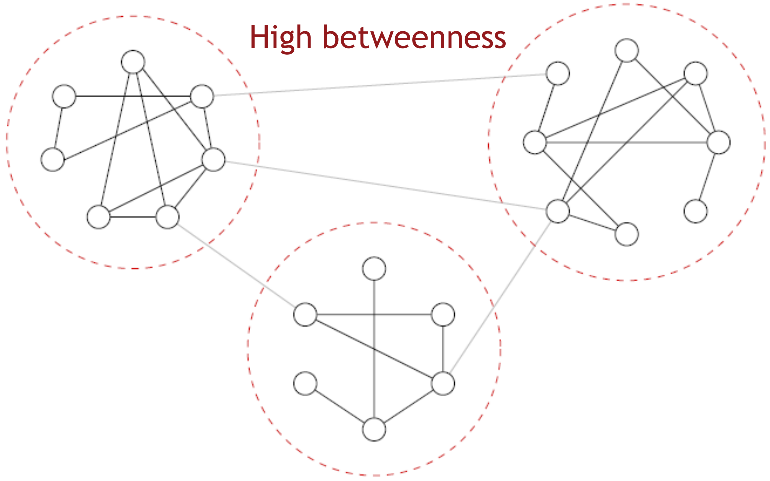
D) **Spectral based methods**: Spectral bisection, multi dimensional spectral analysis,...

2.1.3.- Communities and roles

❑ Link Removal methods

Most “weak” links of the network are eliminated until the network begins to break into separated components. The “weakness” of a link can be measured in different ways, depending on the type of algorithm (shortest path centrality, flow centrality, information centrality,...)

Example:
The weakest links are those with the highest betweenness.



CSG Universidad Rey Juan Carlos

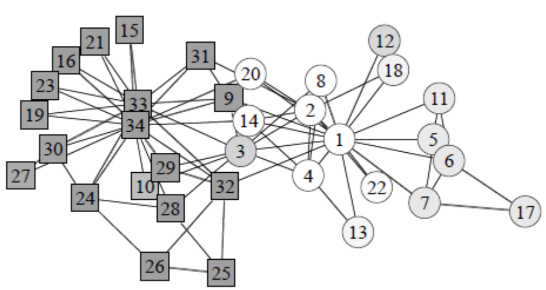
2.1.3.- Communities and roles

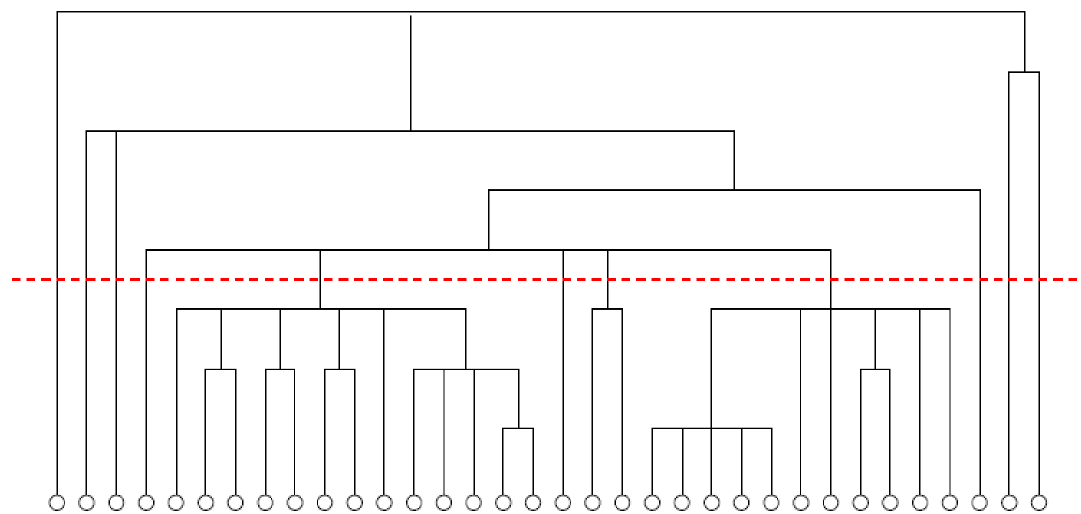
❑ Agglomerative methods

Community is construted from bottom to top: nodes are initially separated and those that are more “similar” are attached (sequentially).

Example:
The similarity is based on the *structural equivalence*

$$x_{ij} = \sqrt{\sum_{k\neq i,j} (A_{ik} - A_{ij})^2}$$





Hierachical clustering partition of the Zachary Karate Club.

2.1.3.- Communities and roles

❑ Modularity based algorithms

The modularity of a network *Q* is, up to a multiplicative constant, the number of edges falling within groups minus the expected number in an equivalent network with edges placed at random.

Example:
The simulated annealing algorithm is an stochastic algorihtm that searches for the maximum modularity by allowing negative (decreasing) variations of the modularity.

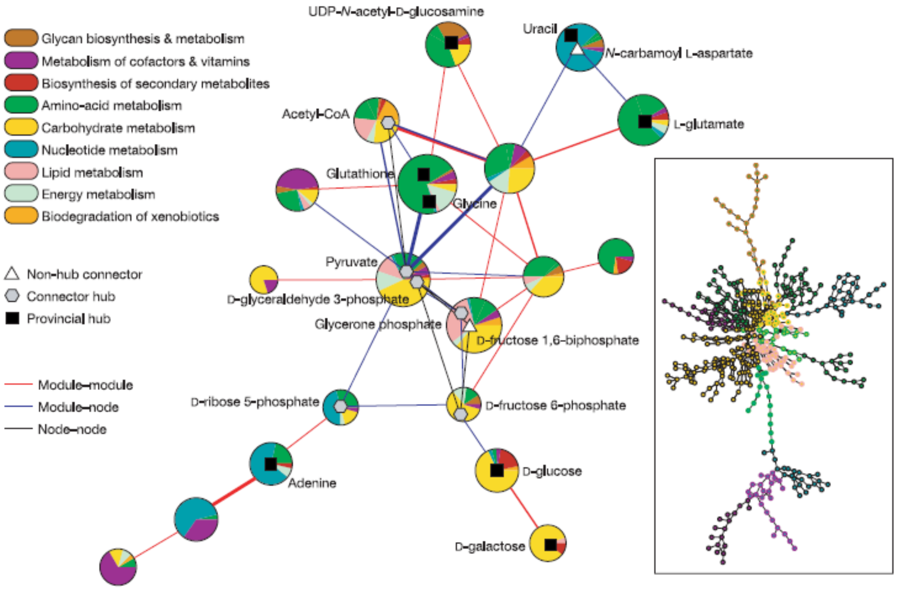


Figure: Partition of the metabollic network of E. Coli using AS algorithm. From Guimerà et al., Nature, 433,895 (2005).

Read more at:
Danon et al.,
World Scientific, 93 (2007)

❑ Spectral based methods

Splits the network into two (or more) components by analyzing the spectral properties of the Laplacian matrix (which contains all the topological information about the network).

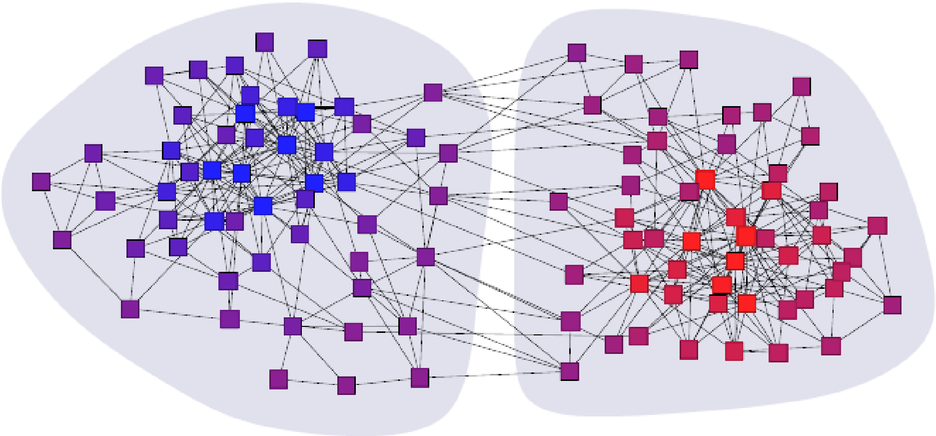
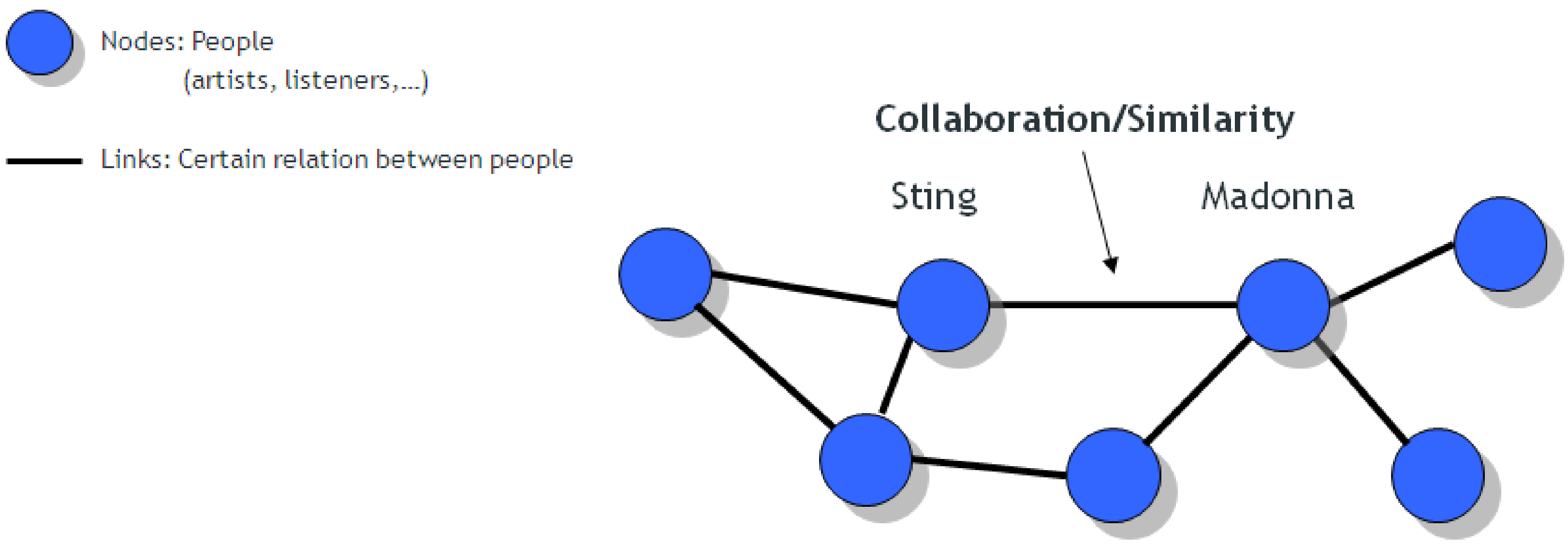


Figure: Division of a collection of US political books (blue, left wing) (red, right wing) made by spectral partitioning.
From Newman, PRE, 74, 36104 (2006).

❑ A practical application: community detection in music networks

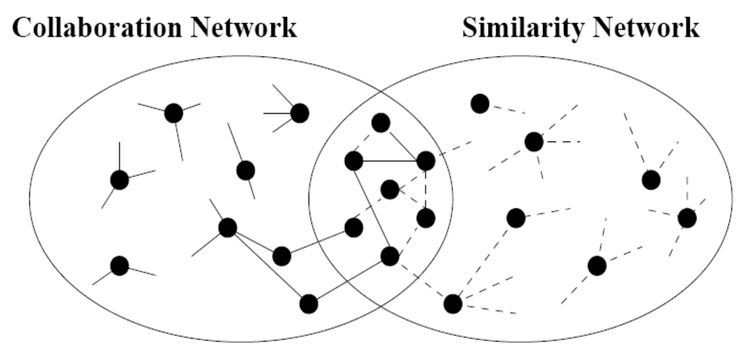
We obtain the collaboration and similarity network between musical artists



2.1.3.- Communities and roles

❑ A practical application: community detection in music networks

Both datasets are obtained from AllMusicGuide database (http://www.allmusicguide.com)



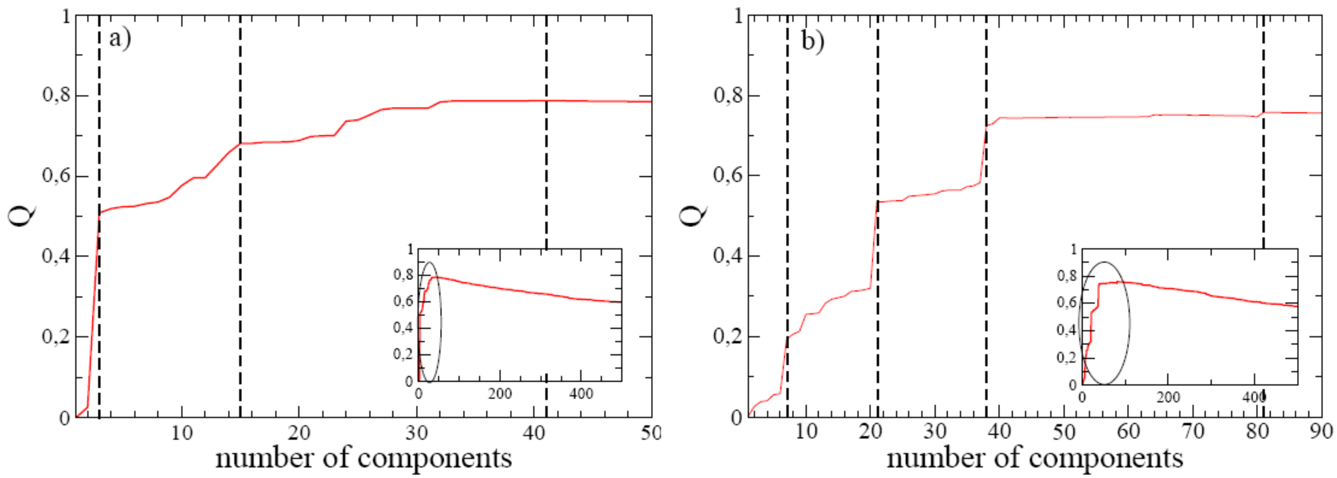
2.1.3.- Communities and roles

❑ A practical application: community detection in music networks

We combine the ***betweenness partition method*** with the modularity measure in order to obtain a good network division:

Similarity Network

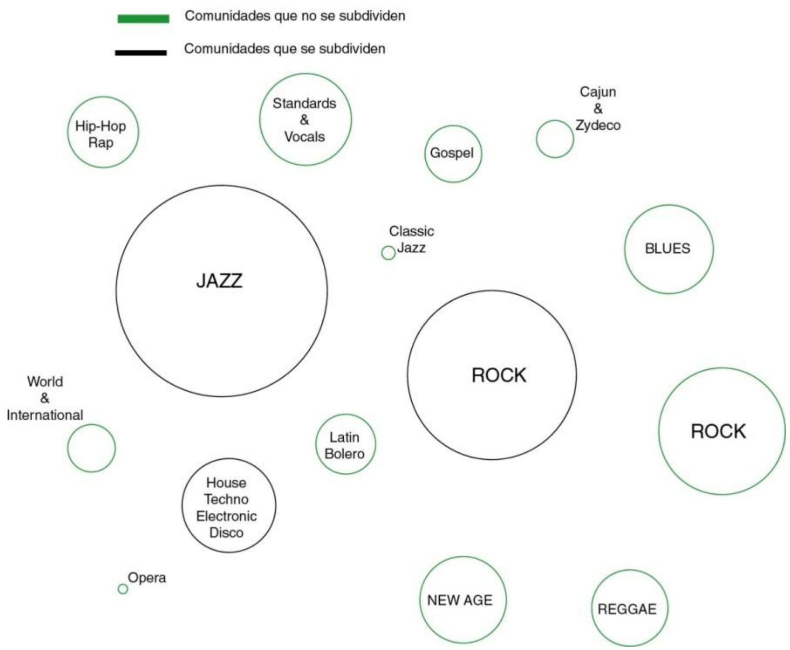
Collaboration Network



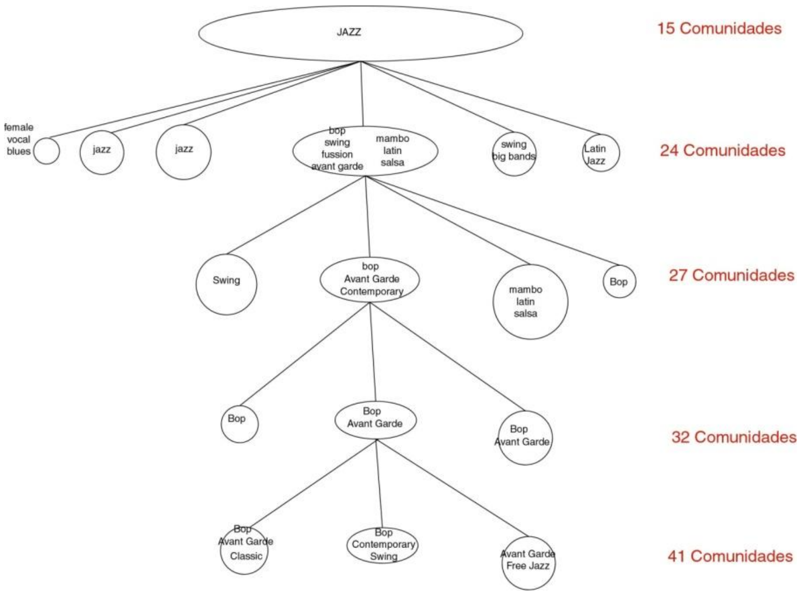
2.1.3.- Communities and roles

❑ A practical application: community detection in music networks

❑ The division of the similarity network is related with musical genres



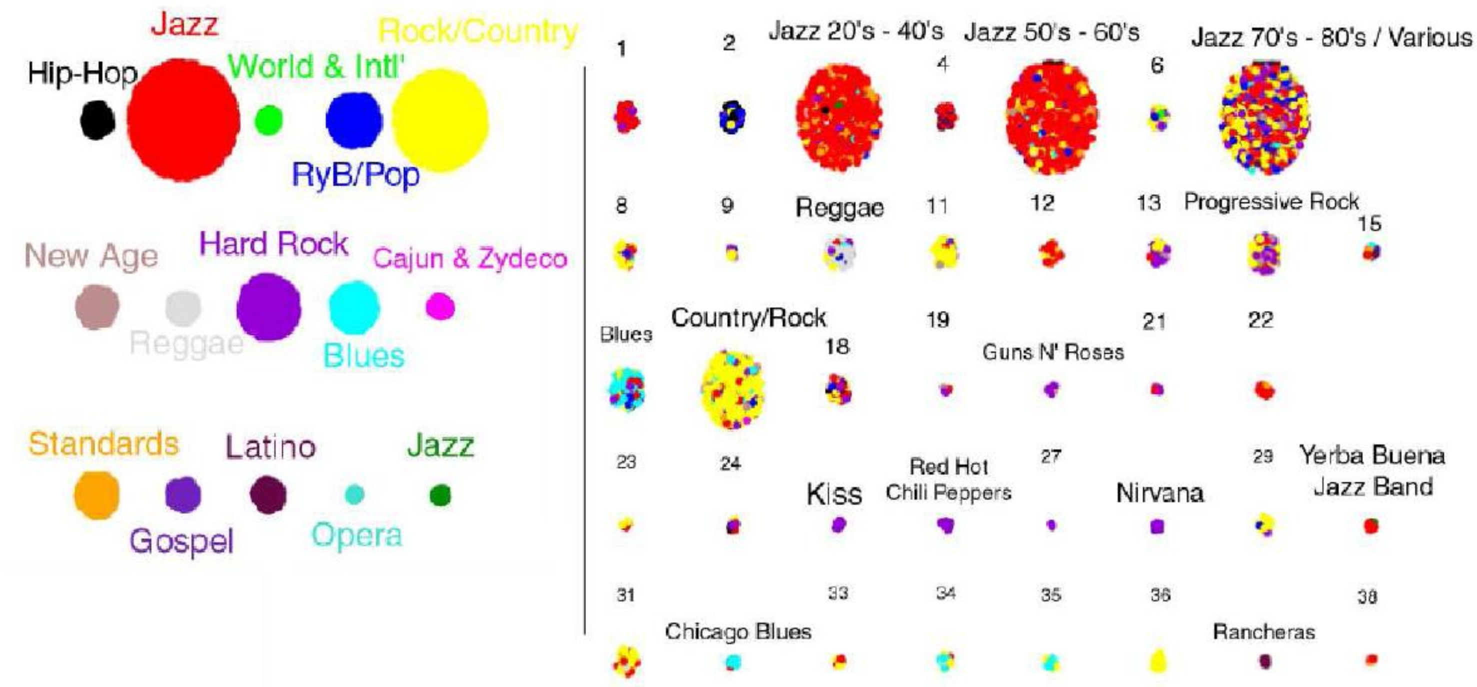
❑ The betwenneess method allows to split the network sequentially.



Read more at:
Teitelbaum et al.,
Chaos, 18, 043105 (2008).

❑ A practical application: community detection in music networks

This method allows to compare collaboration with similarity:



Universidad Rey Juan Carlos

Read more at:
Guimerà et al.,
Nature 433, 895 (2005)

❑ Defining the role of nodes with regard to the community structure

Guimerà et al., investigated the role of the nodes inside the community

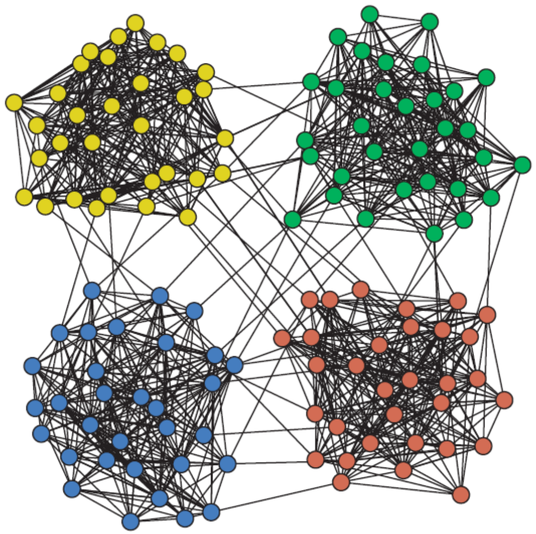
Functional cartography of complex metabolic networks

Roger Guimerà & Luís A. Nunes Amaral

NICO and Department of Chemical and Biological Engineering, Northwestern University, Evanston, Illinois 60208, USA

High-throughput techniques are leading to an explosive growth in the size of biological databases and creating the opportunity to revolutionize our understanding of life and disease. Interpretation of these data remains, however, a major scientific challenge. Here, we propose a methodology that enables us to extract

895



2.1.3.- Communities and roles

❑ Defining the role of nodes with regard to the community structure

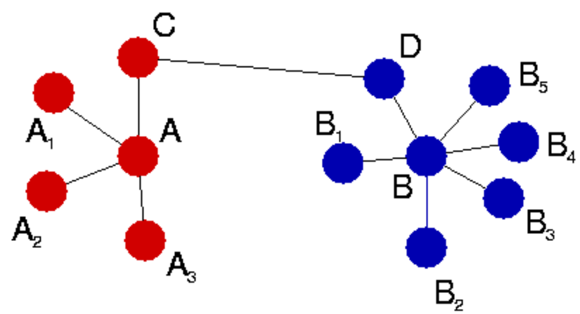
It is possible to evaluate the functionality of t he nodes from the topological properties:

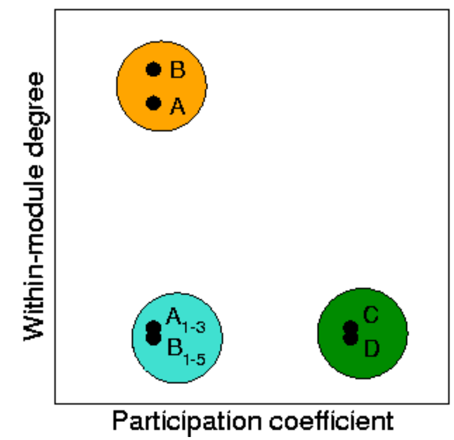
Within-module connectivity:

$$z_i = \frac{\kappa_i - \bar{\kappa}_{s_i}}{\sigma_{\kappa_{s_i}}}$$

Participation coefficient:

$$P_i = 1 - \sum_{s=1}^{N_M} \left(\frac{\kappa_{is}}{k_i} \right)^2$$

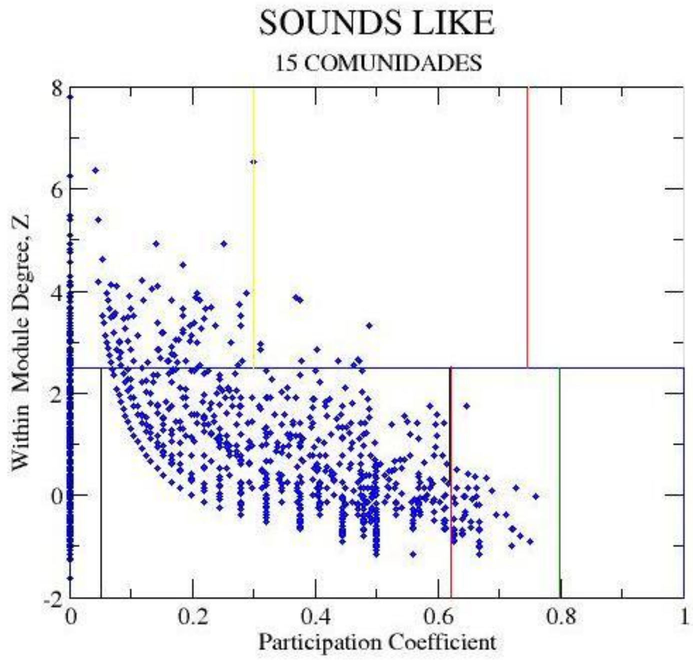


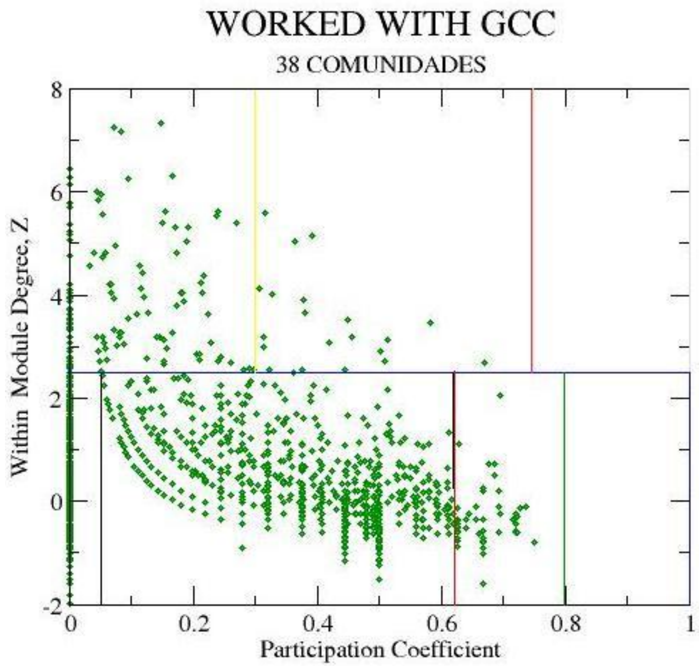


(Figures from R. Guimerà et al., Nature 433, 895 2005)

2.1.3.- Communities and roles

❑ A practical application: Music Networks



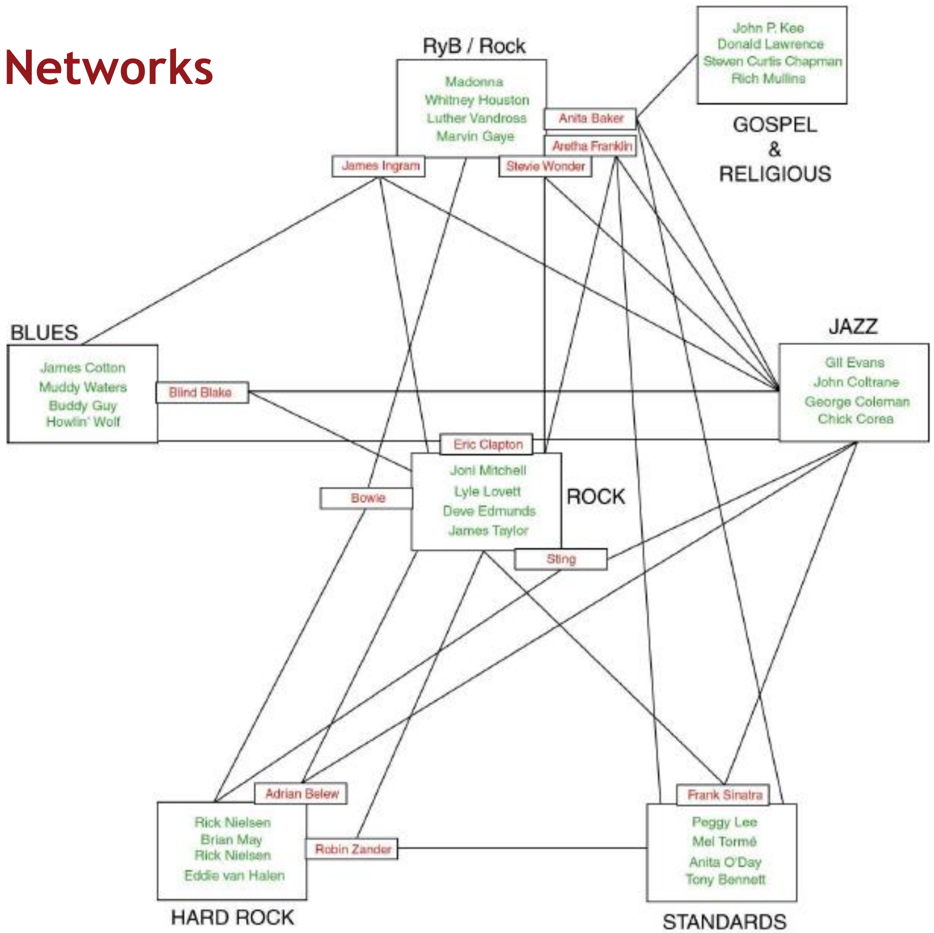


Universidad Rey Juan Carlos

2.1.3.- Communities and roles

- A practical application: Music Networks

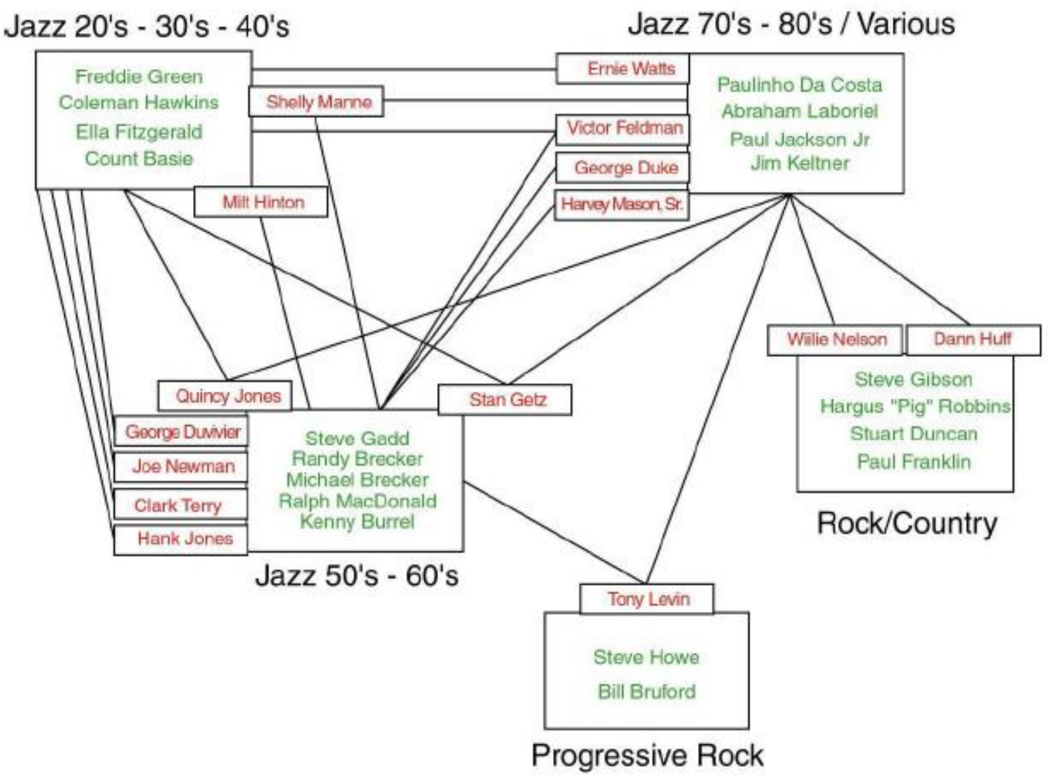
Music similarity cartography



Read more at:
Teitelbaum et al., Chaos, 18, 043105 (2008).

❑ A practical application: Music Networks

Collaboration cartography



2.1.4.- Recommendation networks

2.1.4.- Recommendation networks

❑ The importance of e-commerce

Video by Erik Qualman author of the book *Socialnomics*, about the importance of social networks on internet:

The importance of Social Networks (video)

La importancia de las Redes Sociales (video)

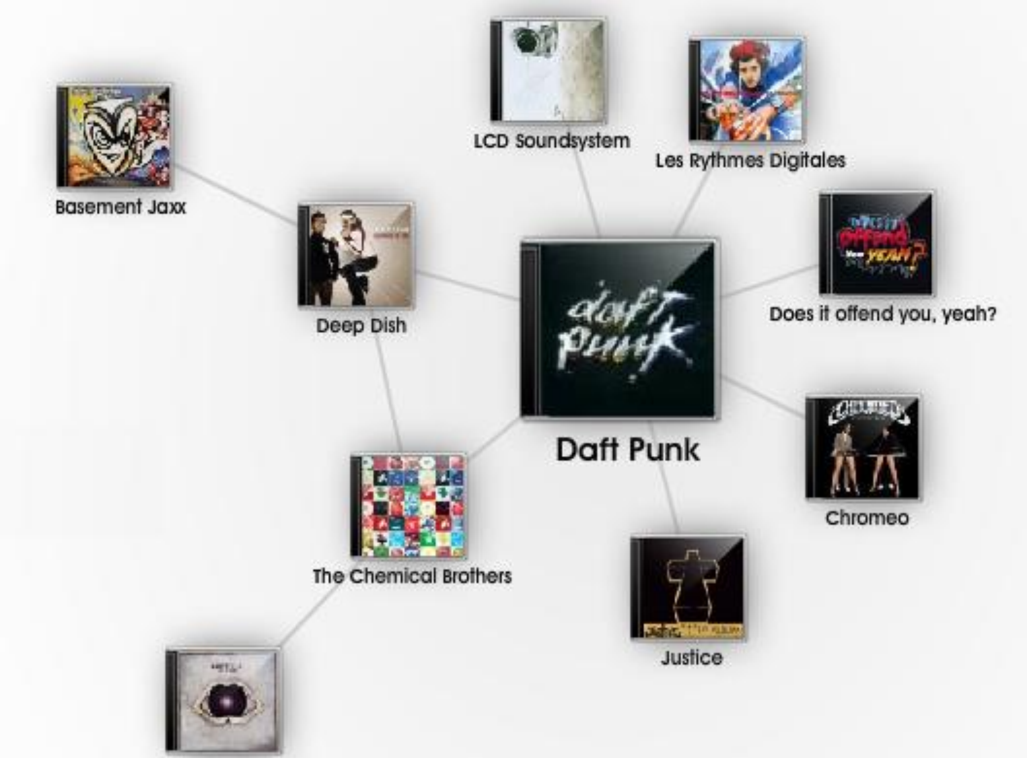
2.1.4.- Recommendation networks

- The long tail

With internet it is possible to do business in the tail:

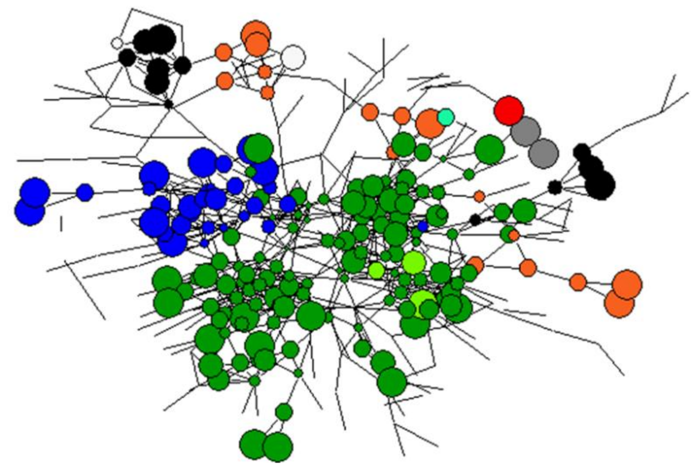
2.1.4.- Recommendation networks

- How is the structure of the underlying recommendation networks?



2.1.4.- Recommendation networks

- The artist similarity recommendation network:



Subgraph of 350 artists. Colors are musical genres and sizes are clustering coefficient. From Cano et al., Chaos, 16, 013107 (2006).

Universidad Rey Juan Carlos

2.1.4.- Recommendation networks

❑ The artist similarity recommendation network:

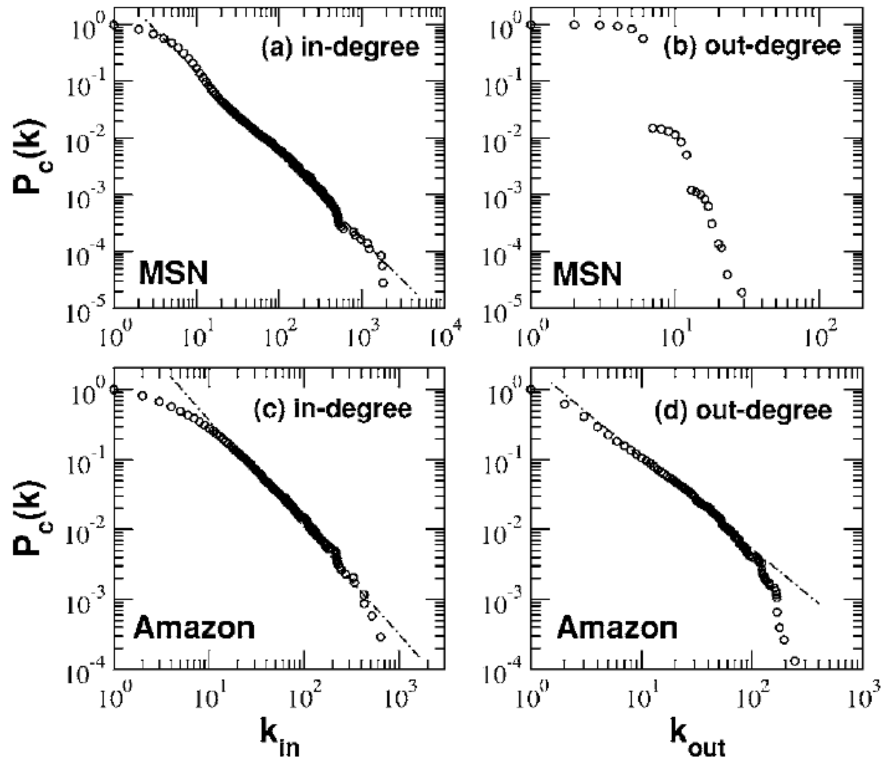
	n	<k>	C	C_r	d	d_r	r	γ_{in}	γ_{out}
MSN	51,616	5.5	0.54	$1.0\text{x}10^{-4}$	7.7	6.4	-0.07	2.4±.01	
Amazon	23,566	13.4	0.14	$5.7\text{x}10^{-4}$	4.2	3.9	-0.06	2.3±.02	2.4±.04
Yahoo!	16,302	62.8	0.38	$3.8\text{x}10^{-3}$	2.7	2.3	-0.21		
AMG	29,206	8.15	0.20	$2.8\text{x}10^{-4}$	6.2	4.9	0.18		

Network parameters for four different recommendation networks. From Cano et al., Chaos 2006

- ❑ All networks are small-world networks
- ❑ Two of them are scale-free, the other two are exponential

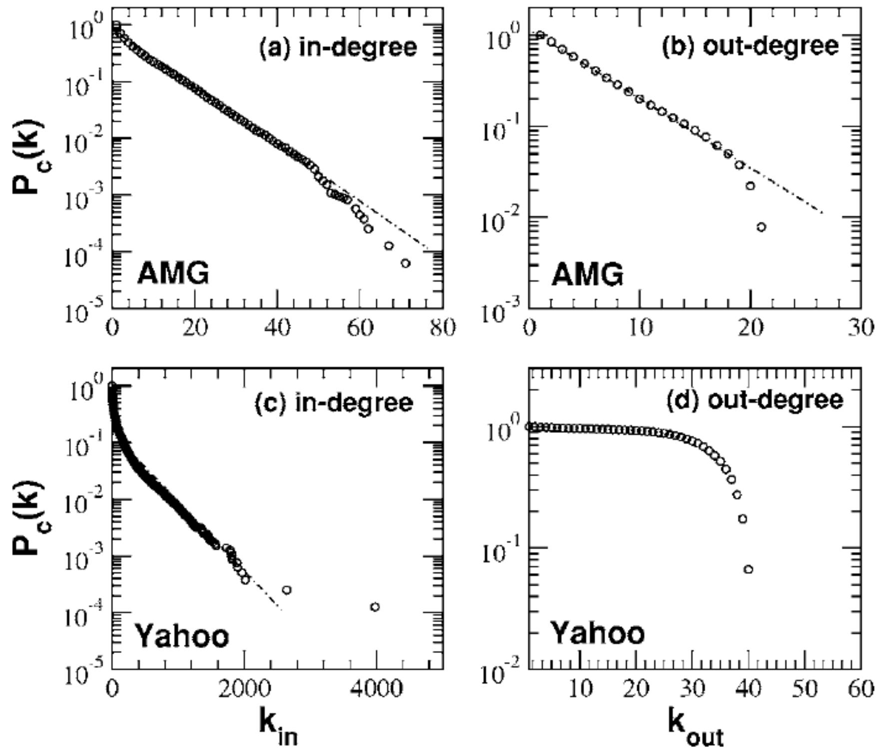
2.1.4.- Recommendation networks

- Two networks are power law



2.1.4.- Recommendation networks

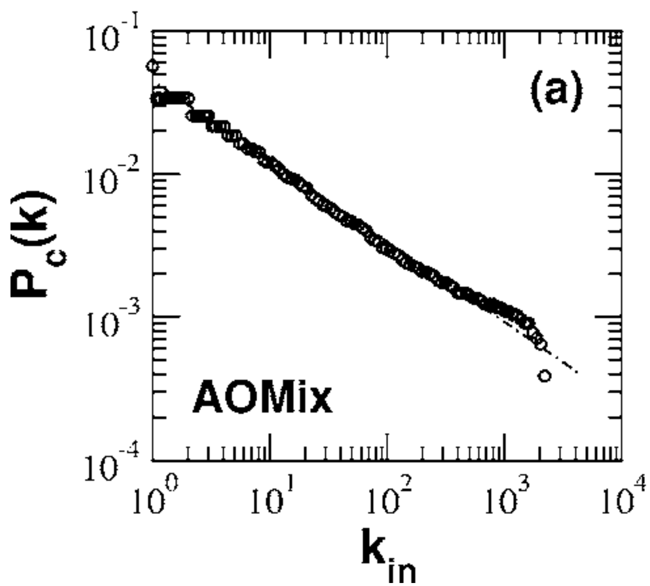
- Two networks are exponential



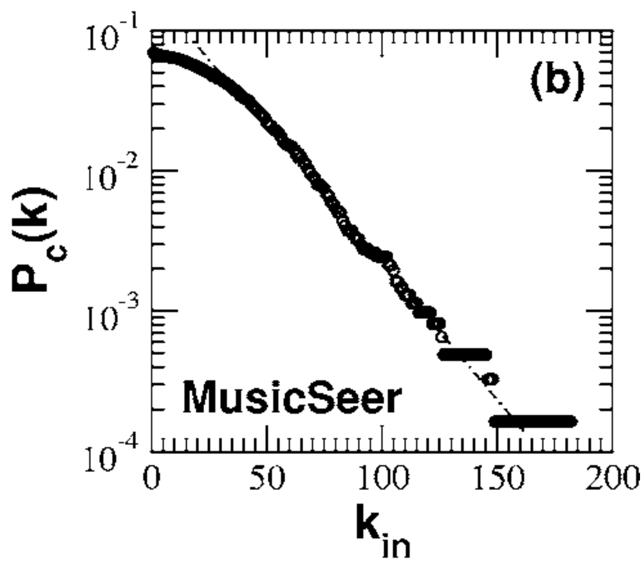
Read more at:
Cano et al.,
Chaos, 16, 13107 (2006).

2.1.4.- Recommendation networks

- ❑ **Two different experiments: AOMix and MusicSeer**



In AOM, users upload playlist with the favorite groups



In MusicSeer, users are asked to evaluate similiarity between artists in a guided list (made by musical experts).

Universidad Rey Juan Carlos

Read more at:
Cano et al.,
Chaos, 16, 13107 (2006).

❑ The network topology is a consequence of the network rules

Two different kind of structures are obtained:

- ❑ When the network is constructed by taking into account the ratings of other users, popularity controls the topology and the structure becomes scale-free. Hubs are the consequence of popularity, not music similarity (Amazon, MSN).

- ❑ When networks are created (strictly) by musical similarity, the network becomes exponential (AMG and Yahoo).

so...

... how do you prefer the recommendation?

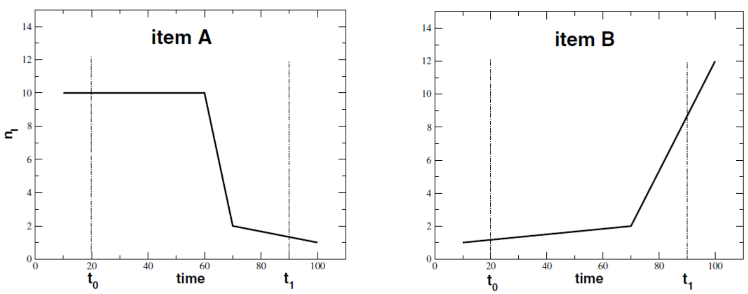
Read more at:
Zanin et al.,
IJBC, 2, 755 (2009).

2.1.4.- Recommendation networks

❑ Caution: information may be obsolete, scarce or abundant

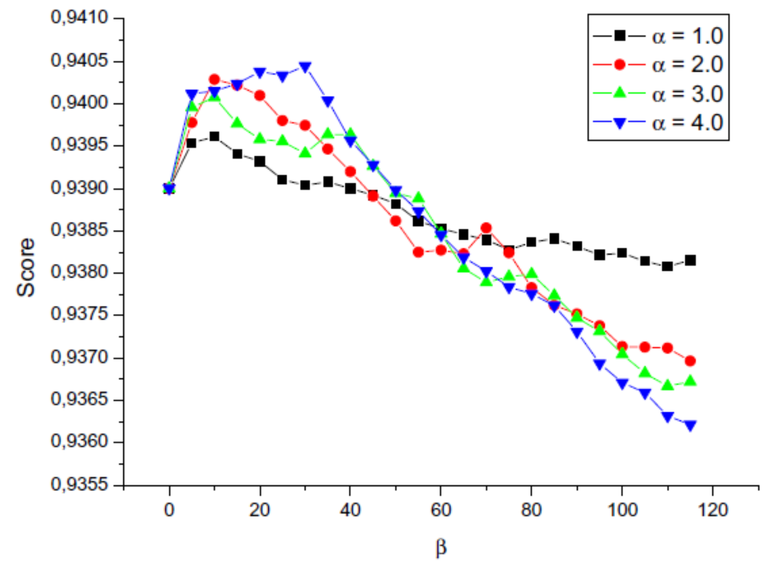
Zanin et al., analyze a movie recommendation network (Netflix)

❑ They report an influence of the age of the links in the recommendation.



When creating the recommendation ranking, they introduce an age-dependant weight:

$$W(i) = \begin{cases} 1, & a_i > \beta \\ 1 + \dfrac{\beta - a_i}{\beta}\alpha, & a_i \le \beta \end{cases}$$



Influence of β and α in the recommendation. Here *a=120* days. From Zanin et al., IJBC, 2, 755 (2009).

Universidad Rey Juan Carlos

2.1.- SOCIAL NETWORKS

Thanks for your attention

mañana más, pero no mejor!

CSG

Universidad Rey Juan Carlos