Complex Networks

Floriana Gargiulo

6/03/2024

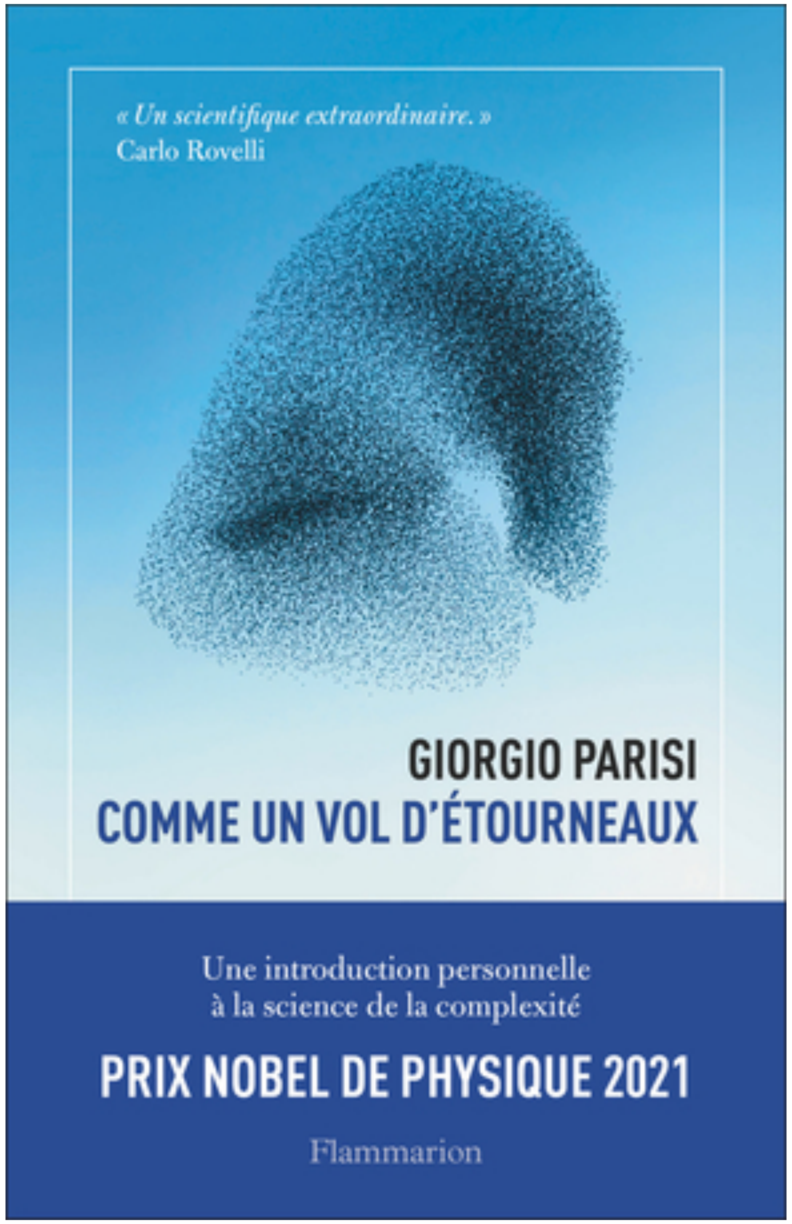
APPROACHES TO STUDY NETWORKS

Approaches	Original discipline	Focus	Objects
Graph theory	Mathematics ~1750 Euler	Algorithms for graph topology	Usually small regular graphs without attributes
Social Network Analysis (SNA)	Sociology and Economics ~1900 Durkheim ~1930 Moreno	• Global indicators for network structure (i.e. connectivity), centrality measures • correlation between the attributes and the structure	Small real graphs with several attributes (rich)
Complex Networks	Physics and Computer Science (Complex Systems) ~1998 Watts - Strogats ~1999 Barabasi - Albert	• statistical properties of networks • role of topology on dynamical processes • Morphogenesis	Large real graphs with no (or few) attributes (poor)

COMPLEX SYSTEMS

COMPLEX SYSTEMS

EMERGENT BEHAVIOR:
Macroscopic phenomenon appearing from the collective interaction of the components of the system, that is not present when components are isolated and that does not depend on the "choices" of the components.



COMPLEX SYSTEMS

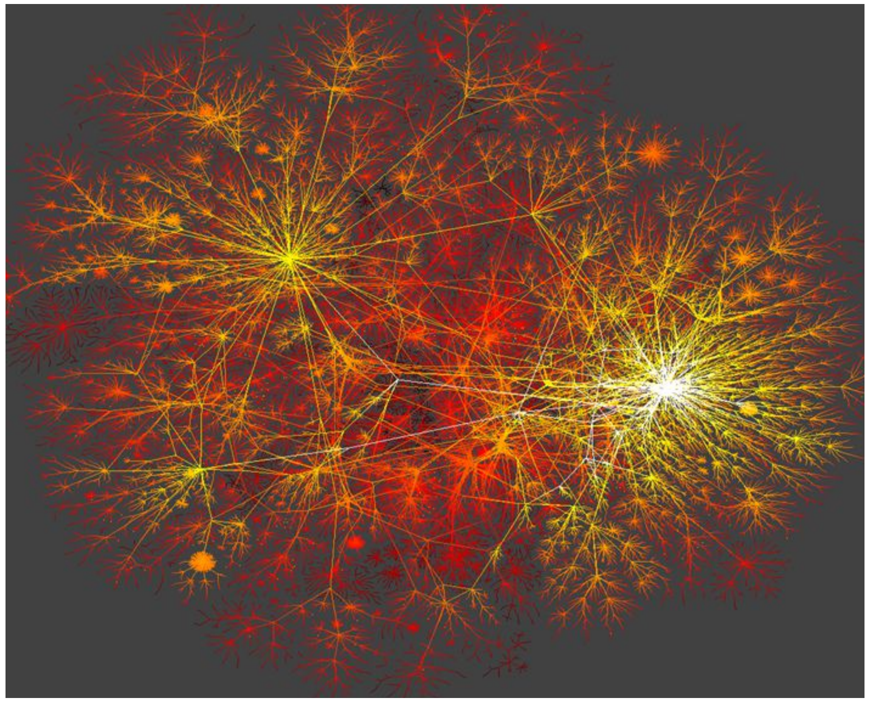
EMERGENT BEHAVIOR:

Macroscopic phenomenon appearing from the collective interaction of the components of the system, that is not present when components are isolated and that does not depend on the "choices" of the components.

Identification of emergent patterns through the **statistical description** of the system

Modeling:
From micro-interactions to macro-behaviors

COMPLEX NETWORKS



Identification of emergent patterns through the **statistical description** of the system

From 1990 we see the growth of **internet** and of the **WWW**. The web grows extremely fast and **unregulated**.

The **experiments** aimed at measuring the Web's graph structure are based on Web crawler, programs that starting from a source page, detect and store all the links they encounter, then follow them iteratively to build up a set of pages reachable from the starting one

At the end of the 90s several data collections from the web crawlers become available. These are the first LARGE NETWORKS datasets.

Albert, Réka, Hawoong Jeong, and Albert-László Barabási. "Diameter of the world-wide web." *nature* 401.6749 (1999): 130-131.

Adamic, Lada A., and Bernardo A. Huberman. "The Web's hidden order." *Communications of the ACM* 44.9 (2001): 55-60.

COMPLEX NETWORKS

At the end of the 90s two milestones in complex network literature, from the point of view of network morphogenesis processes: how microscopic peer interactions can reproduce observed patterns in networks.

Modeling:
From micro-interactions to macro-behaviors

- In 1998 **Duncan Watts and Jacob Strogatz** developed a model for small-world networks

Watts, Duncan J., and Steven H. Strogatz. "Collective dynamics of ‘small-world’networks." *nature* 393.6684 (1998): 440.

- In 1999 **Albert Lazlo Barabasi and Reka Albert** implement the a model with cumulative advantage

Barabási, Albert-László, and Réka Albert. "Emergence of scaling in random networks." *science* 286.5439 (1999): 509-512.

Part 1:

Identification of emergent patterns through the **statistical description** of the system

Statistical characterization of networks

In large systems we cannot look at the individual local properties to understand the network regularities. One has to shift the attention to **statistical measures that take into account the global behavior of the local indicators.**

We look for **classes of nodes and links** sharing similar features: the basic assumption is that nodes in the same class (for example, with similar degree) have the same behavior

Statistical characterization of networks

Social Network Analysis (SNA)

Few dozens of nodes

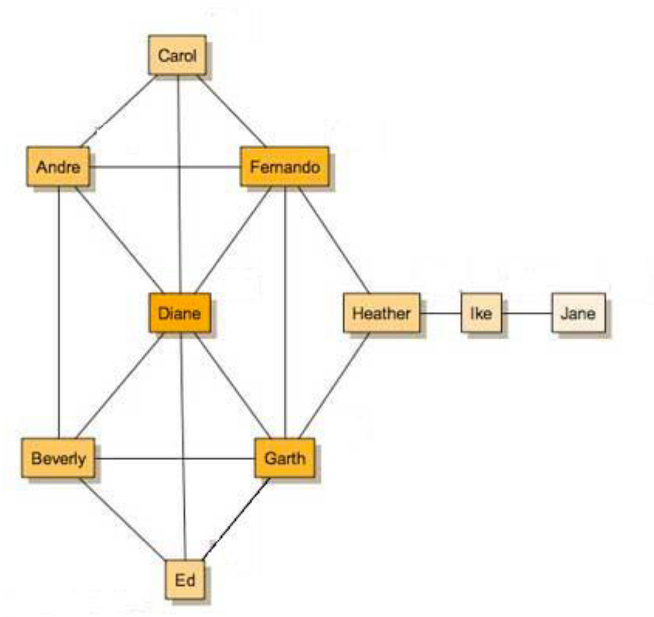
Complex Networks

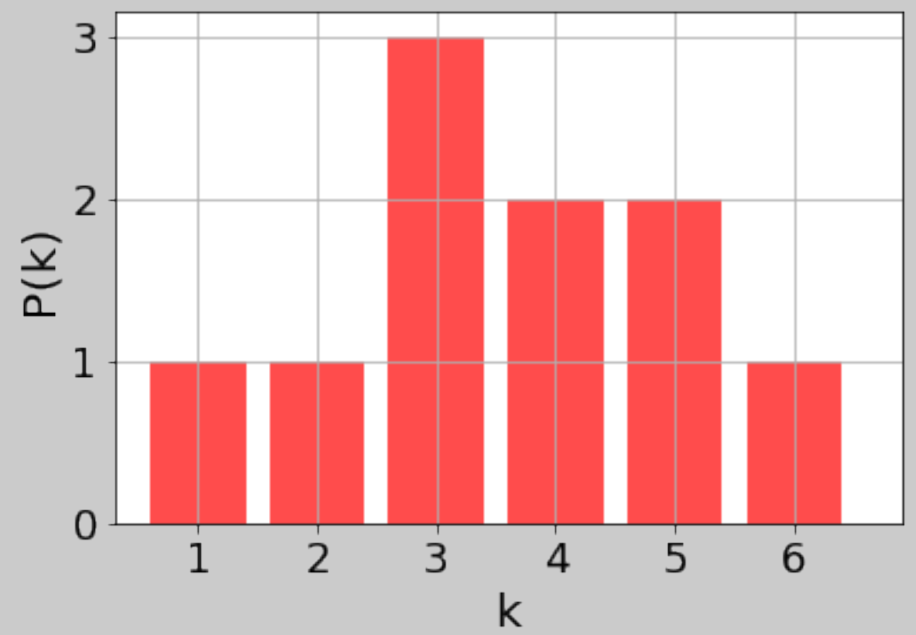
Billions of nodes

Centrality measures → Distributions of centrality measures

Degree centrality ranking:

Diane	6
Garth	5
Fernando	5
Beverly	4
Andre	4
Ed	3
....	





Statistical characterization of networks

1 Statistical properties of the degree

- Degree distribution/heterogeneity
- Degree assortativity
- Weighted networks
- Rich club

1 Statistical properties of the degree

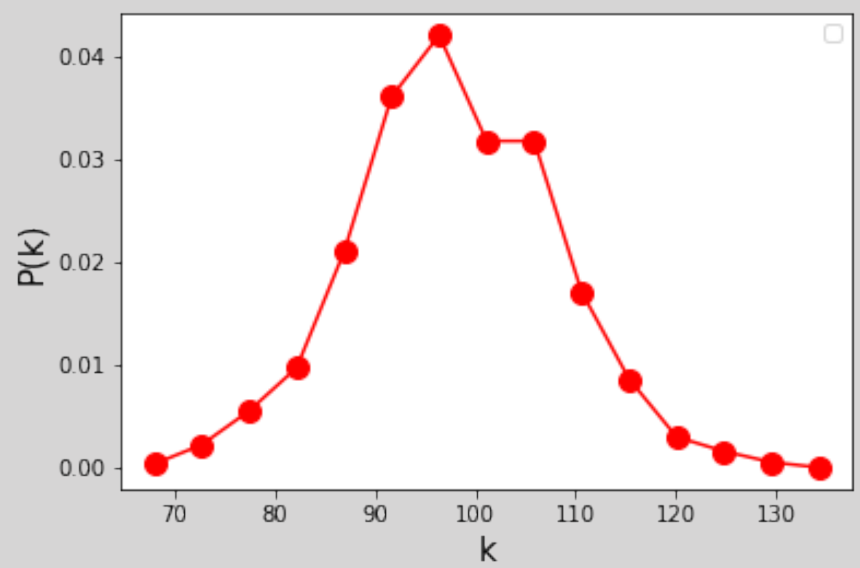
- **Degree distribution/heterogeneity**
- Degree assortativity
- Weighted networks
- Rich Club

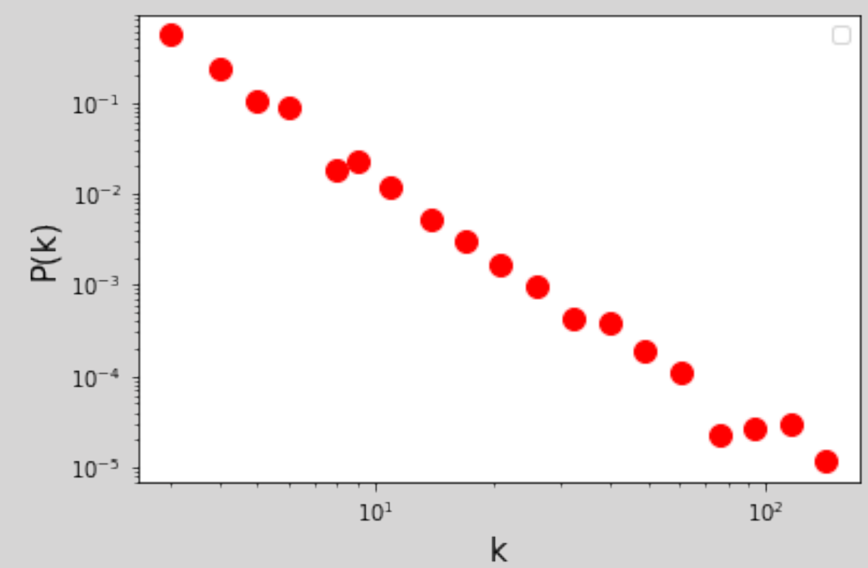
Degree centrality → **Degree distribution**

number of links connected to node i

$$k_i = \sum_j x_{ij}$$

Probability that a randomly chosen vertex has degree k.
It is constructed by the normalized histogram of the list of nodes degrees.



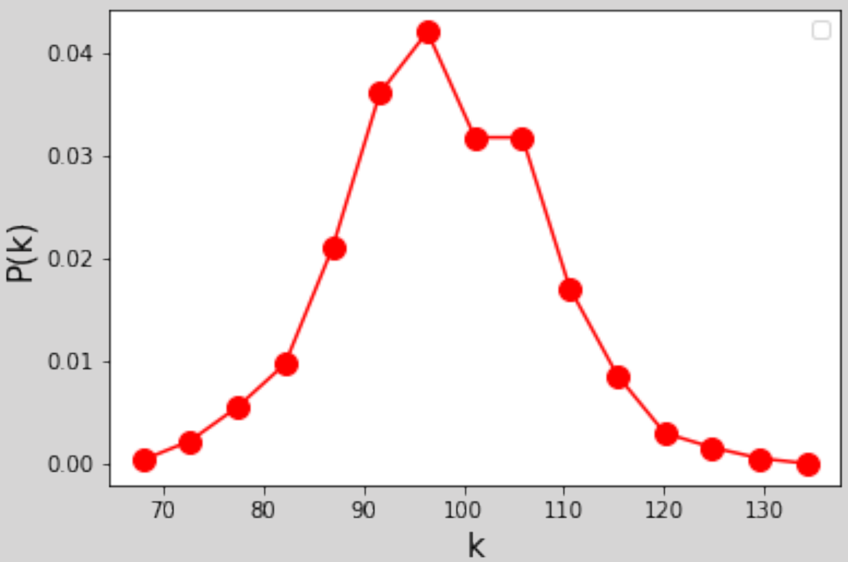


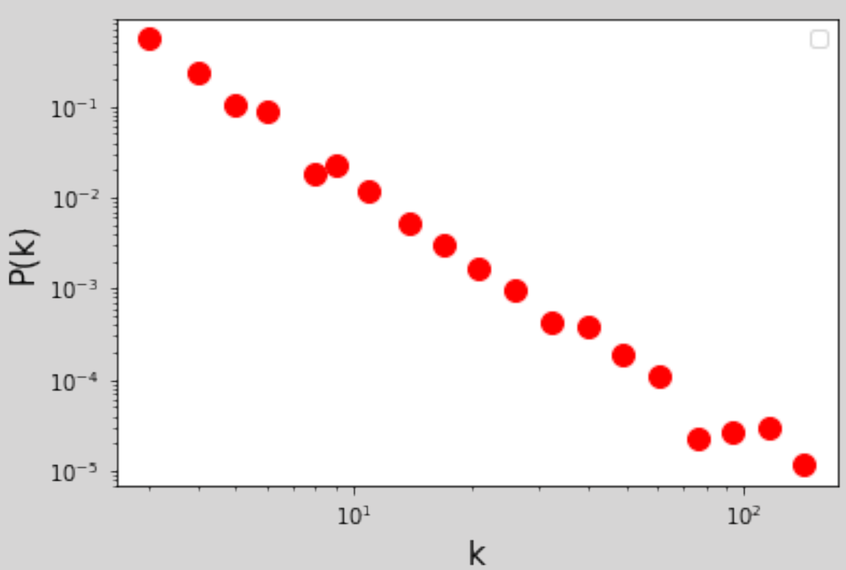
The shape of the degree distribution is a key fingerprint of the network structure and gives important hints on the dynamical processes on the network.

1 Statistical properties of the degree

- **Degree distribution/heterogeneity**
- Degree assortativity
- Weighted networks
- Rich Club

Degree distribution





Average degree of a graph

$$\langle k \rangle = \frac{1}{N} \sum_i k_i = \sum_k k P(k)$$
$$\rightarrow \int k P(k) dk$$

Sometimes the average is not sufficient to characterize a distribution

n-th moment of the distribution $$\langle k^n \rangle = \sum_k k^n P(k)$$

variance $$var(k) = \sigma_k^2 = \langle k^2 \rangle - \langle k \rangle^2$$

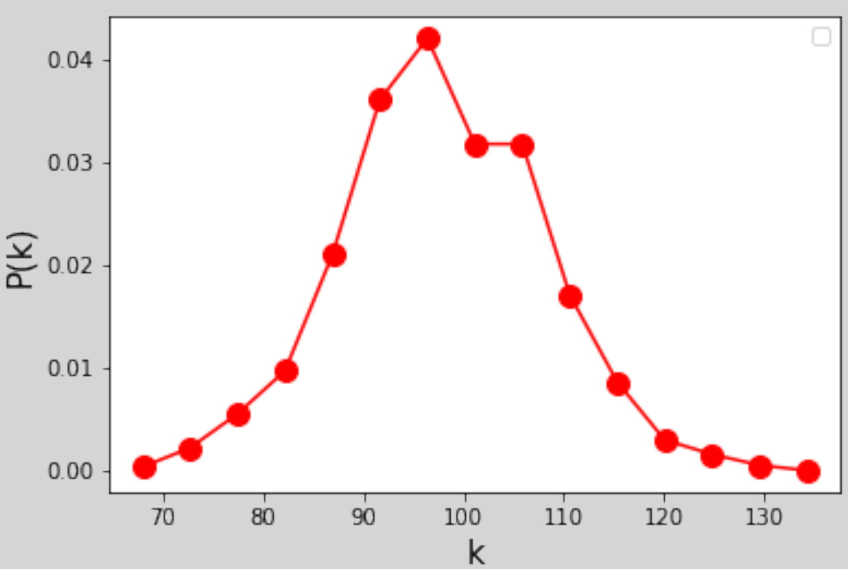
Standard deviation

1 Statistical properties of the degree

- **Degree distribution/heterogeneity**
- Degree assortativity
- Weighted networks
- Rich Club

Degree distribution

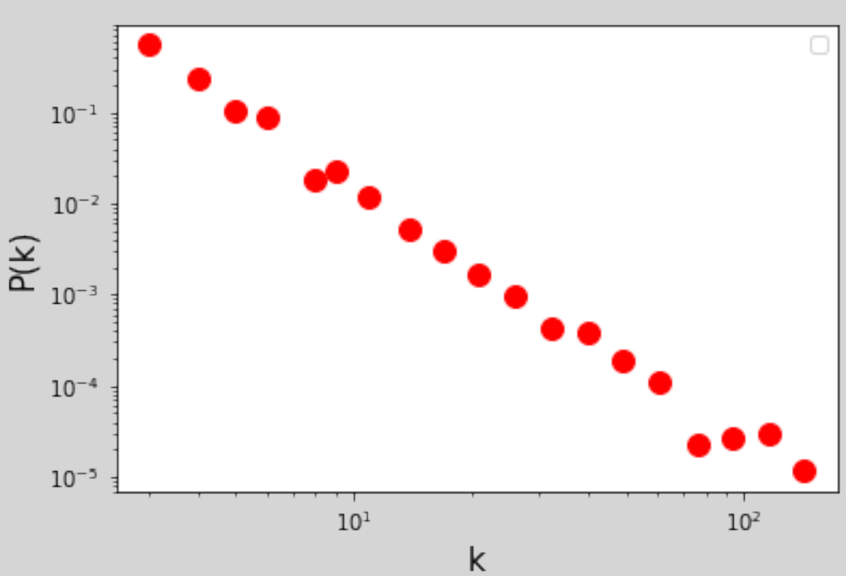
Two classes of networks



Homogeneous networks

Small standard deviation. Fluctuations around the average are small

$$\langle k^2 \rangle \sim \langle k \rangle^2$$



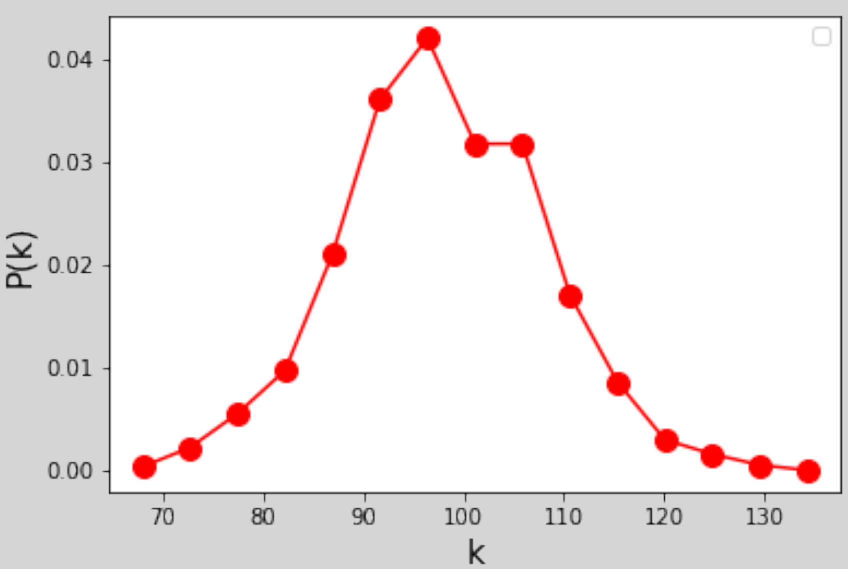
Heterogeneous (scale free) networks

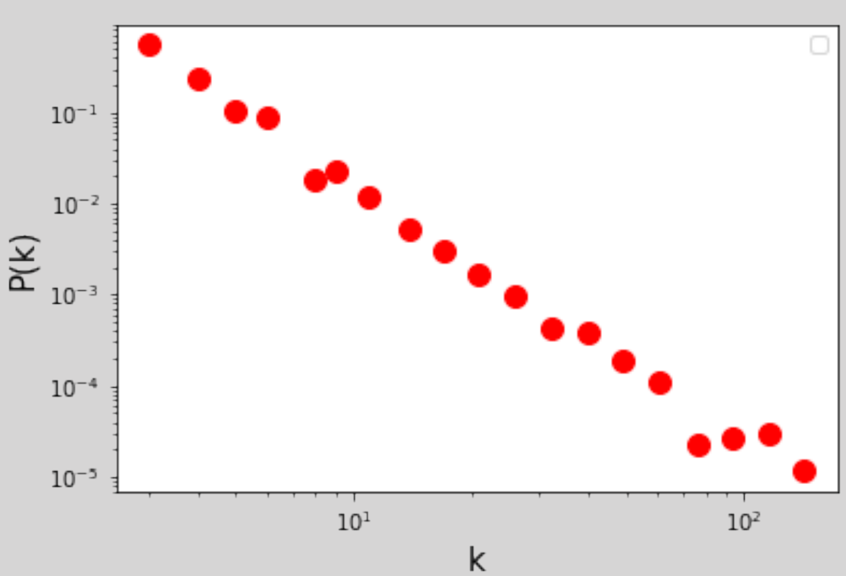
Large standard deviation. There is not a limit to the amplitude of fluctuations

1 Statistical properties of the degree

- **Degree distribution/heterogeneity**
- Degree assortativity
- Weighted networks
- Rich Club

Degree distribution





Two classes of networks

Homogeneous networks

Small standard deviation. Fluctuations around the average are small

$$\langle k^2 \rangle \sim \langle k \rangle^2$$

$$K \sim 1$$

Heterogeneous (scale free) networks

Large standard deviation. There is not a limit to the amplitude of fluctuations

$$K >> 1$$

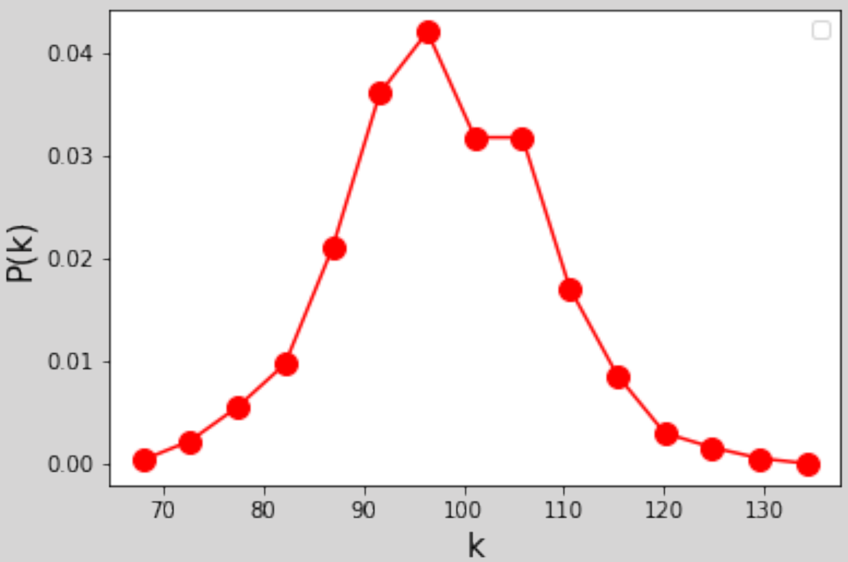
Heterogeneity indicator

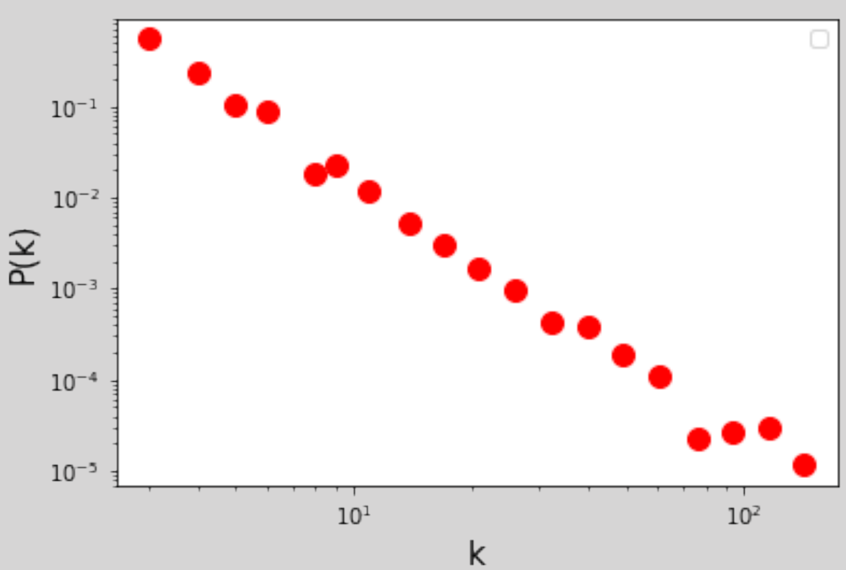
$$K = \frac{\langle k^2 \rangle}{\langle k \rangle^2}$$

1 Statistical properties of the degree

- **Degree distribution/heterogeneity**
- Degree assortativity
- Weighted networks
- Rich Club

Degree distribution





Two classes of networks

Homogeneous networks

Small standard deviation. Fluctuations around the average are small

$$\langle k^2 \rangle \sim \langle k \rangle^2$$

$$K \sim 1$$

Ex. Poisson distribution, Gaussian distribution

Heterogeneous (scale free) networks

Large standard deviation. There is not a limit to the amplitude of fluctuations

$$K >> 1$$

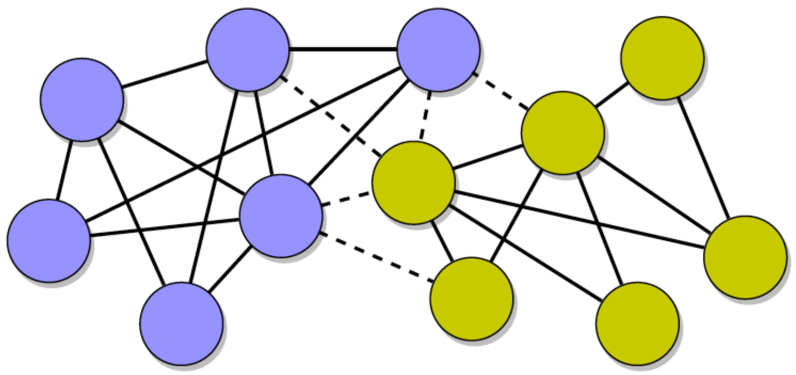
Ex. Power law (Pareto) distribution

How the properties of a node are connected to the properties of their neighbors.

In social networks, for example relatives may live near each other, and friends may have similar interests.

This is called *HOMOPHILY: "birds of a feather flock together"*

Because of assortativity, we are able to make predictions about a person's qualities by inspecting their neighbors.



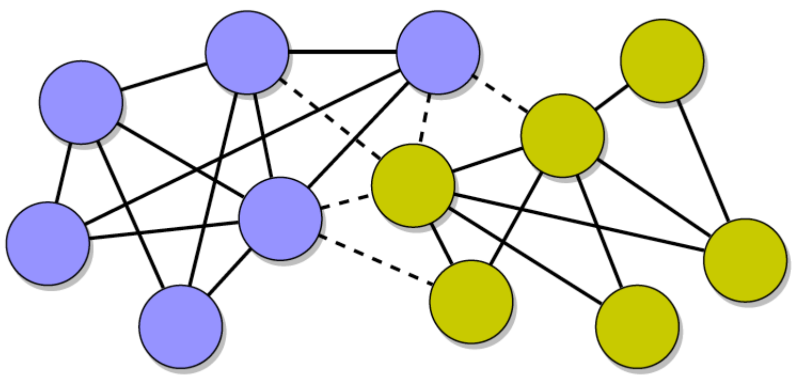
Two possible processes are responsible for homophily:

1)people select similars for connecting
2)people become similar for social influence

1 Statistical properties of the degree

- Degree distribution/heterogeneity
- **Degree assortativity**
- Weighted networks
- Rich Club

How the properties of a node are connected to the properties of their neighbors.



We can calculate the assortativity of a network based on a given node attribute, using the Newman's correlation coefficient. In Networkx:

For categorical attributes:

```
assort_a = nx. attribute_assortativity_coefficient (G, category )
```

For numerical attributes:

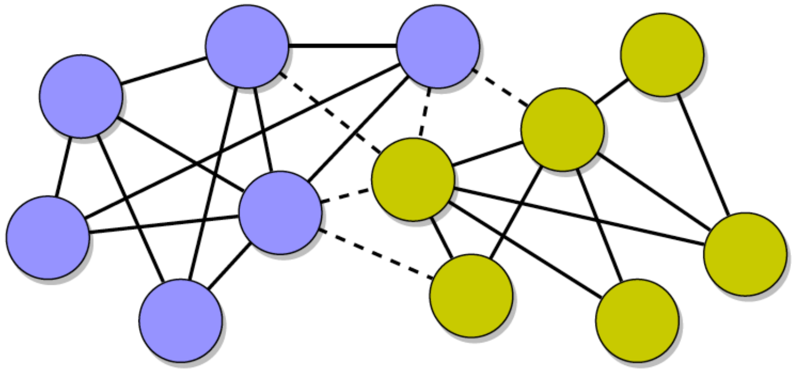
```
assort_n = nx. numeric_assortativity_coefficient (G, quantity )
```

M. E. J. Newman, Mixing patterns in networks, Physical Review E, 67 026126, 2003

1 Statistical properties of the degree

- Degree distribution/heterogeneity
- **Degree assortativity**
- Weighted networks
- Rich Club

How the properties of a node are connected to the properties of their neighbors.



Nodes in any network have the fundamental property of degree.
Assortativity based on degree is called **degree assortativity** or **degree correlation**

An intuitive measure: the average nearest neighbors degree

$$k_{nn,i} = \frac{1}{k_i} \sum_{j \in \mathcal{V}_i} k_j$$

1 Statistical properties of the degree

- Degree distribution/heterogeneity
- **Degree assortativity**
- Weighted networks
- Rich Club

Degree correlation → **Degree Mixing**

How the properties of a node are connected to the properties of their neighbors.

An intuitive measure: the average nearest neighbors degree

$$k_{nn,i} = \frac{1}{k_i} \sum_{j \in \mathcal{V}_i} k_j$$

To analyze this measure from a statistical point of view we calculate the average value of this measure for all the nodes of degree k

$$k_{nn}(k) = \frac{1}{N_k} \sum_{i | k_i = k} k_{nn,i}$$

Number of nodes with degree k

1

Statistical properties of the degree

- Degree distribution/heterogeneity
- **Degree assortativity**
- Weighted networks
- Rich Club

Degree correlation → **Degree Mixing**

How the properties of a node are connected to the properties of their neighbors.

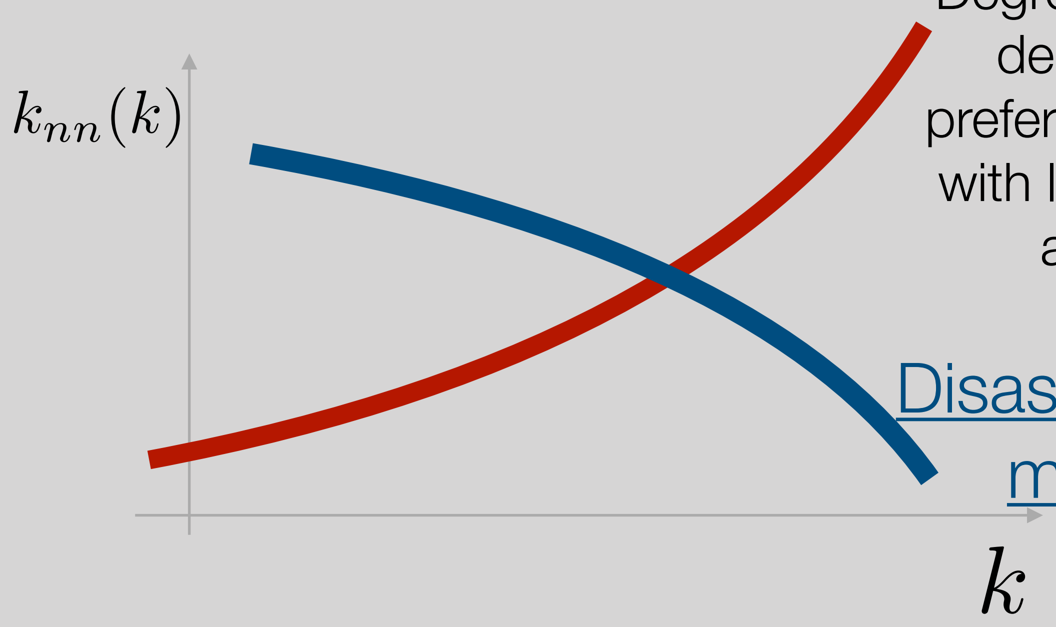
An intuitive measure: the average nearest neighbors degree

$$k_{nn,i} = \frac{1}{k_i} \sum_{j \in \mathcal{V}_i} k_j$$

To analyze this measure from a statistical point of view we calculate the average value of this measure for all the nodes of degree k

$$k_{nn}(k) = \frac{1}{N_k} \sum_{i | k_i = k} k_{nn,i}$$

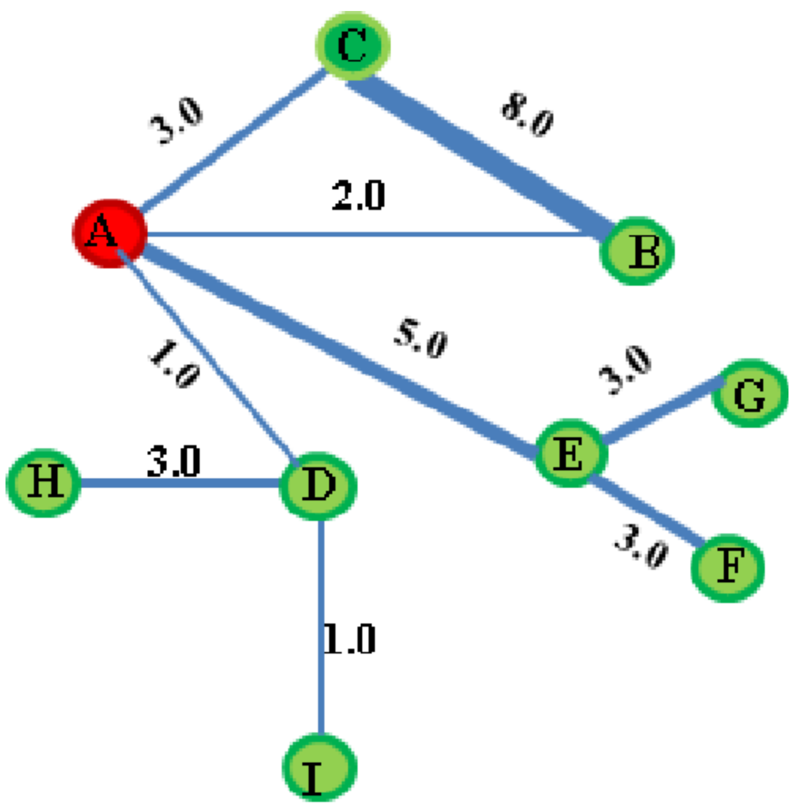
Two possible behaviors



Assortative mixing

Degree homophily: low degree nodes are preferentially connected with low degree nodes and viceversa

Disassortative mixing



Weighted networks have a number (weight, w_{ij}) associated to each link.
The entries of the adjacency matrix are therefore the values associated to the weights.

For weighted networks it is important to characterize the relationship between the topology (the link structure) and the weights.

We define "strength" of a node as the sum of the weights on the connected links:

$$s_i = \sum_j w_{ij}$$

For directed weighted networks, in analogy with the degree, we can define in-strength and out-strength

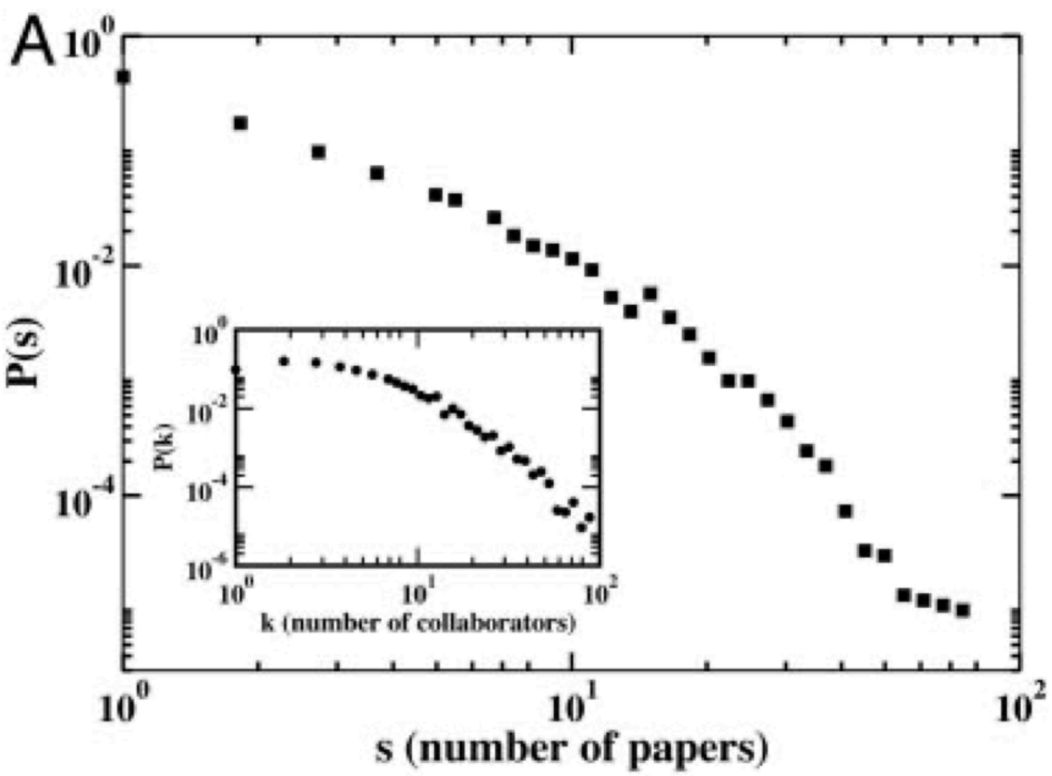
Barrat, A., Barthelemy, M., Pastor-Satorras, R., & Vespignani, A. (2004). The architecture of complex weighted networks. *Proceedings of the national academy of sciences*, *101*(11), 3747-3752.

1 Statistical properties of the degree

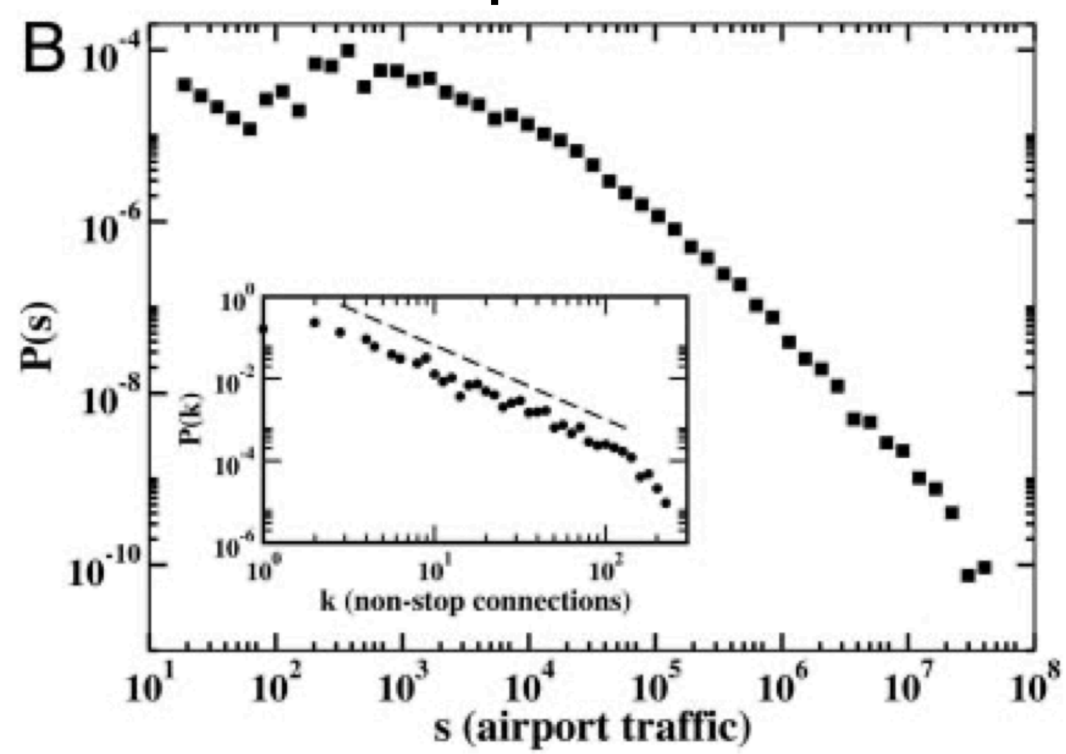
- Degree distribution/heterogeneity
- Degree assortativity
- **Weighted networks**
- Rich Club

For weighted networks we can calculate
- the probability distribution of the weights (P(w) -vs - w)
-the probability distribution of the node strengths

Scientific collaboration network



Airport traffic



The degree distribution and the strength distribution seem to be quite similar among them.

1 Statistical properties of the degree

- Degree distribution/heterogeneity
- Degree assortativity
- **Weighted networks**
- Rich Club

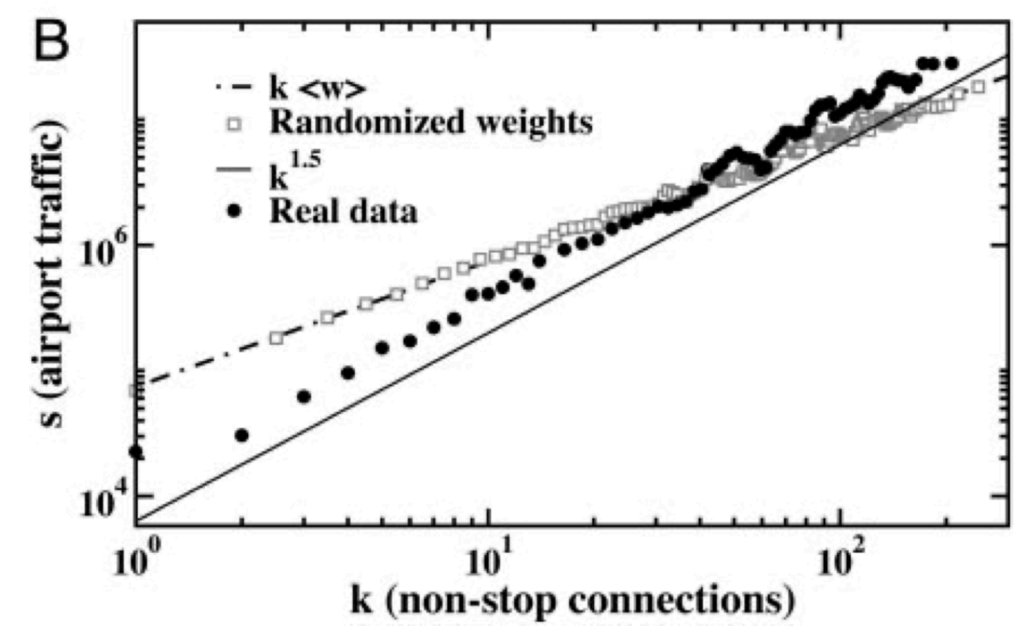
For weighted networks we can calculate
- the probability distribution of the weights
-the probability distribution of the node strengths
-the weight-structure correlation

For both the case: $s \sim k^{\beta}$.

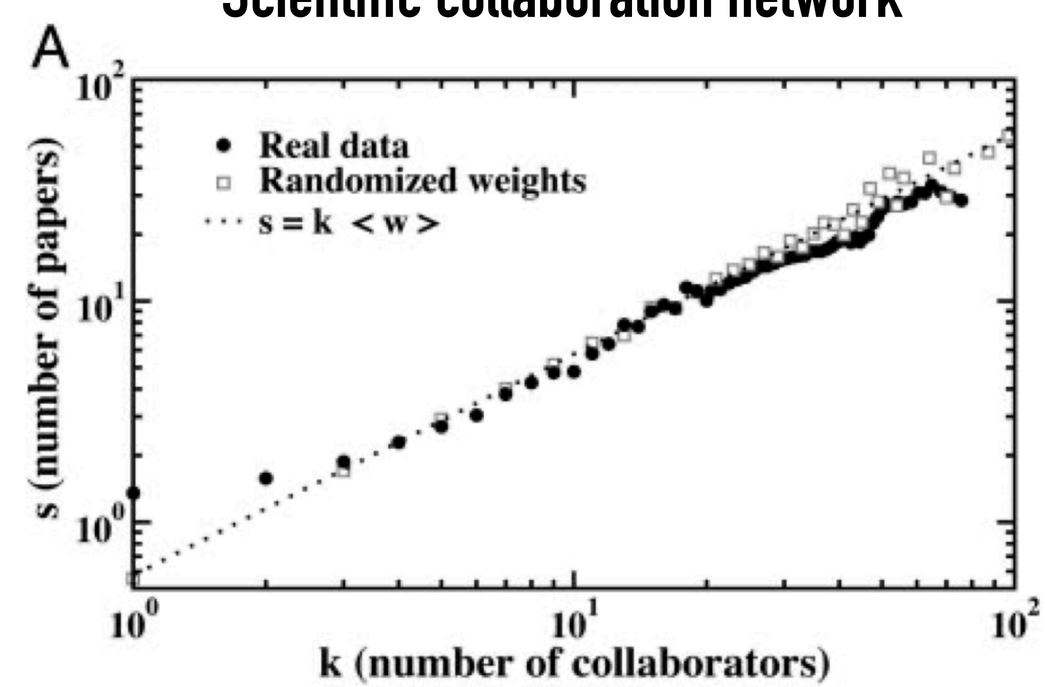
For scientific collaboration $\beta \sim 1$: there is correlation between the weights and the node degree. The strength and the degree give the same information.

For airports $\beta \sim 1.5$: Stronger weights are associated to higher degree nodes. The strength and the degree don't give the same information.

Airport traffic



Scientific collaboration network



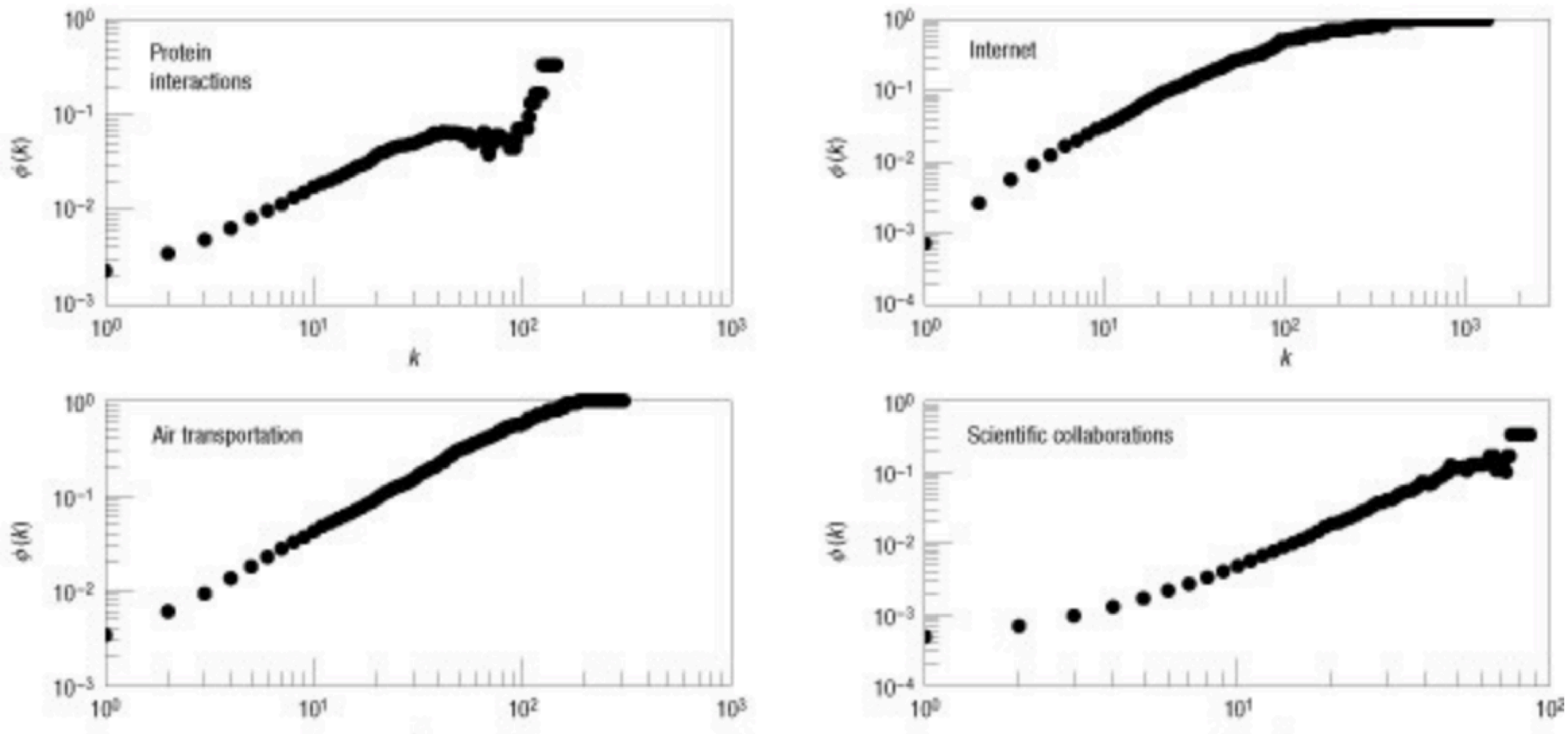
The **rich club** property measures the extent to which well-connected nodes also connect to each other.

$$\phi(k) = \frac{2E_{>k}}{N_{>k}(N_{>k} - 1)}$$

Colizza, V. and Flammini, A. and Serrano, M. A. and Vespignani, A. (2006). "Detecting rich-club ordering in complex networks". *Nature Physics*. 2. **2** (2): 110–115

1 Statistical properties of the degree

- Degree distribution/heterogeneity
- Degree assortativity
- Weighted networks
- **Rich Club**



Colizza, V. and Flammini, A. and Serrano, M. A. and Vespignani, A. (2006). "Detecting rich-club ordering in complex networks". *Nature Physics*. 2. **2** (2): 110–115

1 Statistical properties of the degree

- Degree distribution/heterogeneity
- Degree assortativity
- Weighted networks
- **Rich Club**

The **rich club** index exists in a normalized form: comparing the actual value with the value observed in a randomized version of the network.

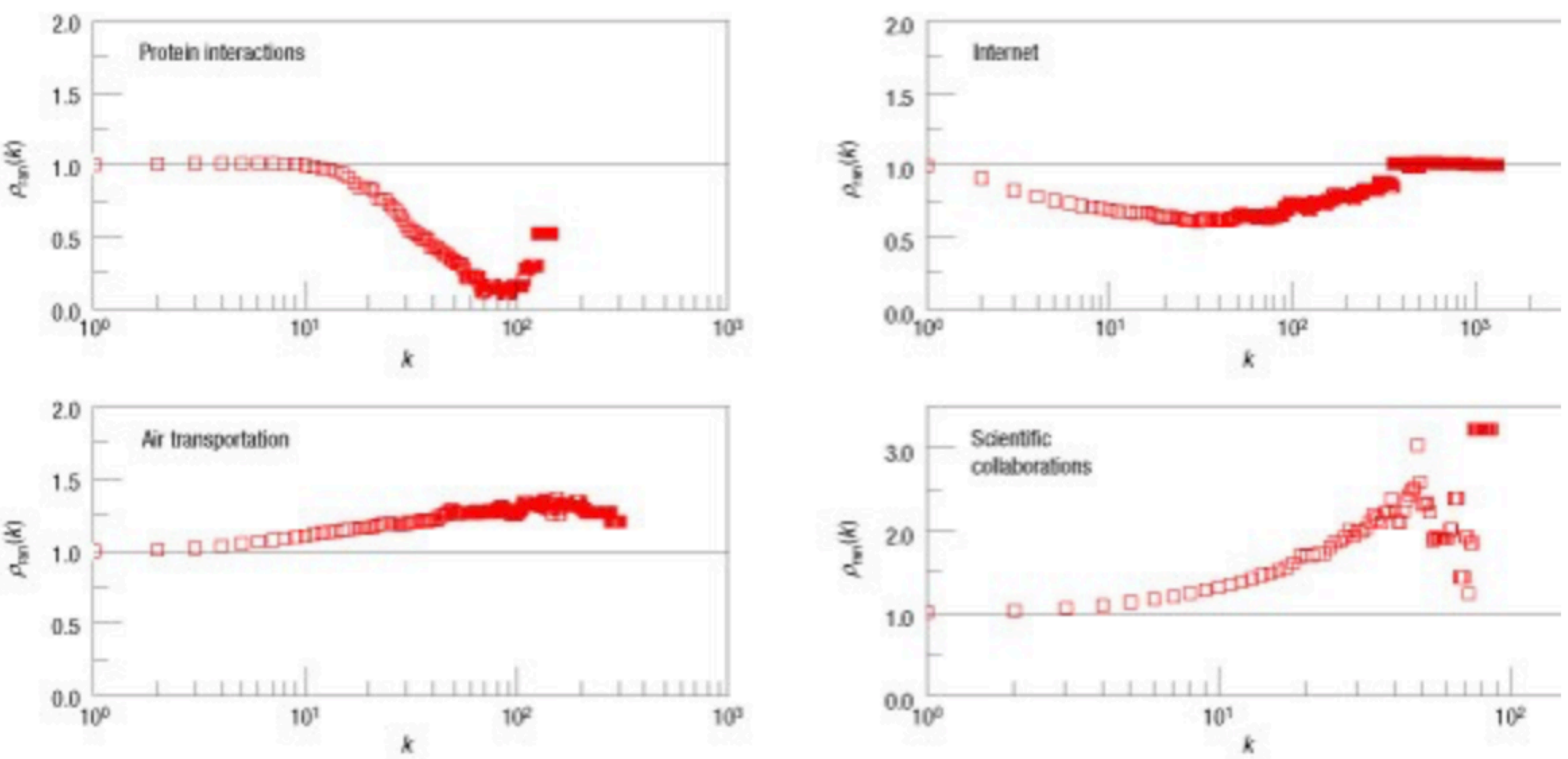
$$\rho_{rand}(k) = \frac{\phi(k)}{\phi_{rand}(k)}$$

$$\phi_{rand}(k) \sim \frac{k^2}{\langle k \rangle N}$$

Colizza, V. and Flammini, A. and Serrano, M. A. and Vespignani, A. (2006). "Detecting rich-club ordering in complex networks". *Nature Physics*. 2. **2** (2): 110–115

1 Statistical properties of the degree

- Degree distribution/heterogeneity
- Degree assortativity
- Weighted networks
- **Rich Club**



Colizza, V. and Flammini, A. and Serrano, M. A. and Vespignani, A. (2006). "Detecting rich-club ordering in complex networks". *Nature Physics*. 2. **2** (2): 110–115

Statistical characterization of networks

1. **Statistical properties of the degree**
    - Degree distribution/heterogeneity
    - Degree assortativity
    - Weighted networks
    - Rich club
2. **The clustering spectrum**

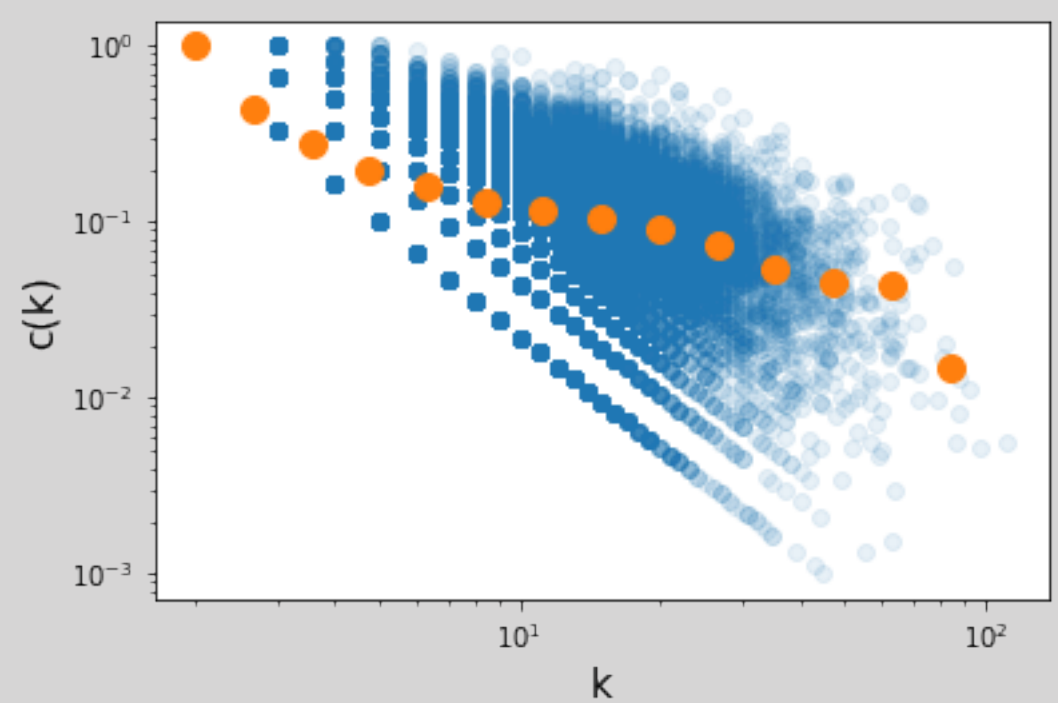
2 The clustering spectrum

Clustering coefficient

Clustering spectrum

To analyze the clustering properties from a statistical point of view we calculate the average value of this measure for all the nodes of degree k

$$c(k) = \frac{1}{N_k} \sum_{i|k_i=k} C(i)$$



Statistical characterization of networks

1. **Statistical properties of the degree**
   - Degree distribution/heterogeneity
   - Degree assortativity
   - Weighted networks
   - Rich club
2. **The clustering spectrum**
3. **What "small world" means in a large network?**

3 What "small world" means in a large network?

Average shortest path length

Small world Property

In real-world networks, any two members of the network are usually connected via short paths. In other words, the average path length is small

Six degrees of separation:

Stanley Milgram In the well-known small-world experiment conducted in the 1960's conjectured that people around the world are connected to one another via a path of at most 6 individuals

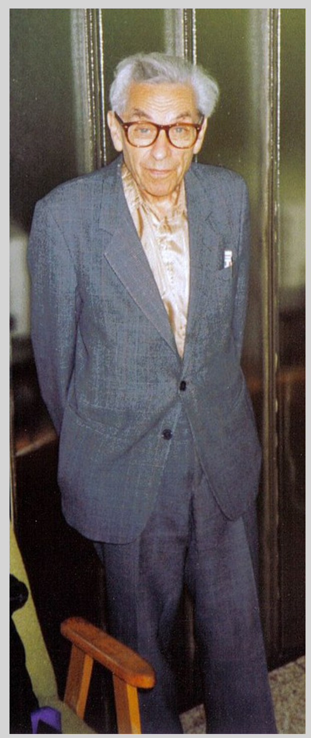
Four degrees of separation:

Lars Backstrom et al. in May 2011, the average path length between individuals in the Facebook graph was 4.7. (4.3 for individuals in the US)

Web	Facebook	Flickr	LiveJournal	Orkut	YouTube
16.12	4.7	5.67	5.88	4.25	5.10

3 What "small world" means in a large network?

Average shortest path length → Small world Property



Erdos Number:
Paul Erdos was a famous mathematician who made a critical contribution to the science of networks.
Mathematicians like to study their distance, in the network of coauthors, with the particular node corresponding to Erdos.
This distance is called "Erdos Number".
Many mathematicians have a very small Erdos number: the average for Fields medalist is 3.1.

MR Erdos Number = 4

Floriana Gargiulo	coauthored with	Sergio Ferrara	MR1997477
Sergio Ferrara	coauthored with	John Scherk	MR0516919
John Scherk	coauthored with	Vijaya Kumar Murty	MR1298275
Vijaya Kumar Murty	coauthored with	Paul Erdős[1]	MR0880469

3 What “small world” means in a large network?

Six degrees of Kevin Bacon

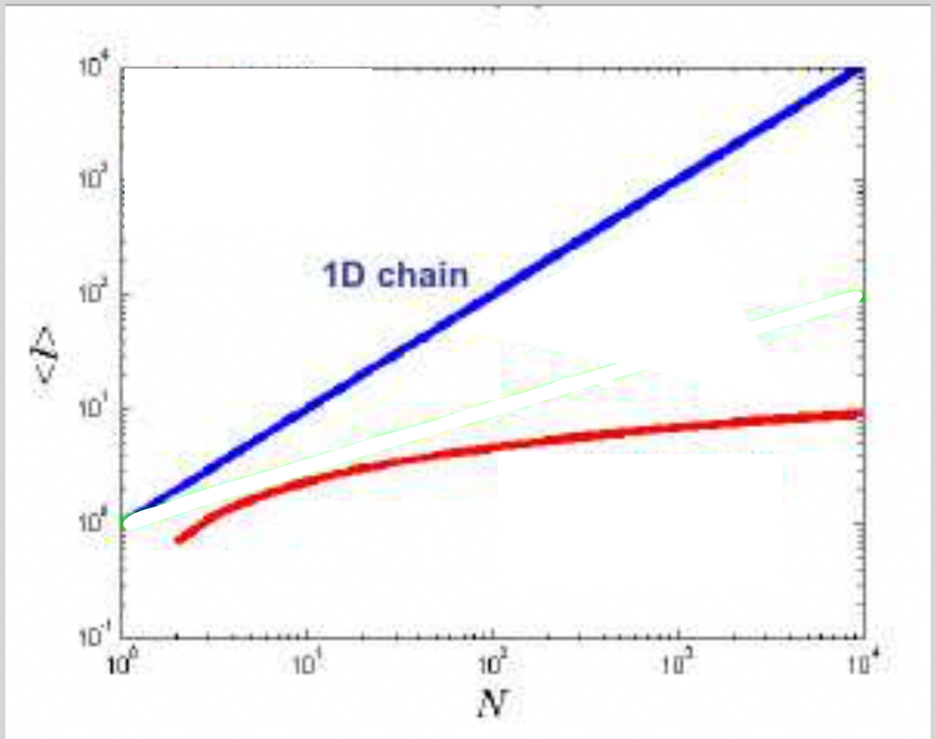
On the site 'oracleofbacon.org' you can calculate the distance in the co-starring network between Kevin Bacon and all other actors…

3 What "small world" means in a large network?

Average shortest path length

Small world Property

How we can "quantify" small?



In a linear chain:

$$\langle l \rangle \sim N$$

We define '**small world**' a network the networks for which:

$$\langle l \rangle \sim log(N)$$

Statistical characterization of networks

1. **Statistical properties of the degree**
   - Degree distribution/heterogeneity
   - Degree assortativity
   - Weighted networks
   - Rich club
2. **The clustering spectrum**
3. **What “small world” means in a large network?**
4. **A dynamical network property: the robustness**

Dynamical measures on network: robustness

Network robustness

Electric Grids

Internet

Banking systems

The stability of complex networks is a problem of crucial importance

How networks remain functional under attacks or failure?

We study the effect of progressive node (or edge) removal on networks:
we remove a fraction *f* of nodes and we observe which part of the network remains connected.

The simplest quantitative measure of damage is given by the relative size of the largest connected component of the remaining network

$$S_f/N$$

Dynamical measures on network: robustness

Network robustness

The simplest quantitative measure of damage is given by the relative size of the largest connected component of the remaining network

$$S_f/N << 1$$

The network is broken in many microscopic part and it is no more functional

Two types of node removal:

<u>Targeted attacks:</u> nodes are removed according to their centrality

<u>Random failures:</u> nodes are removed randomly

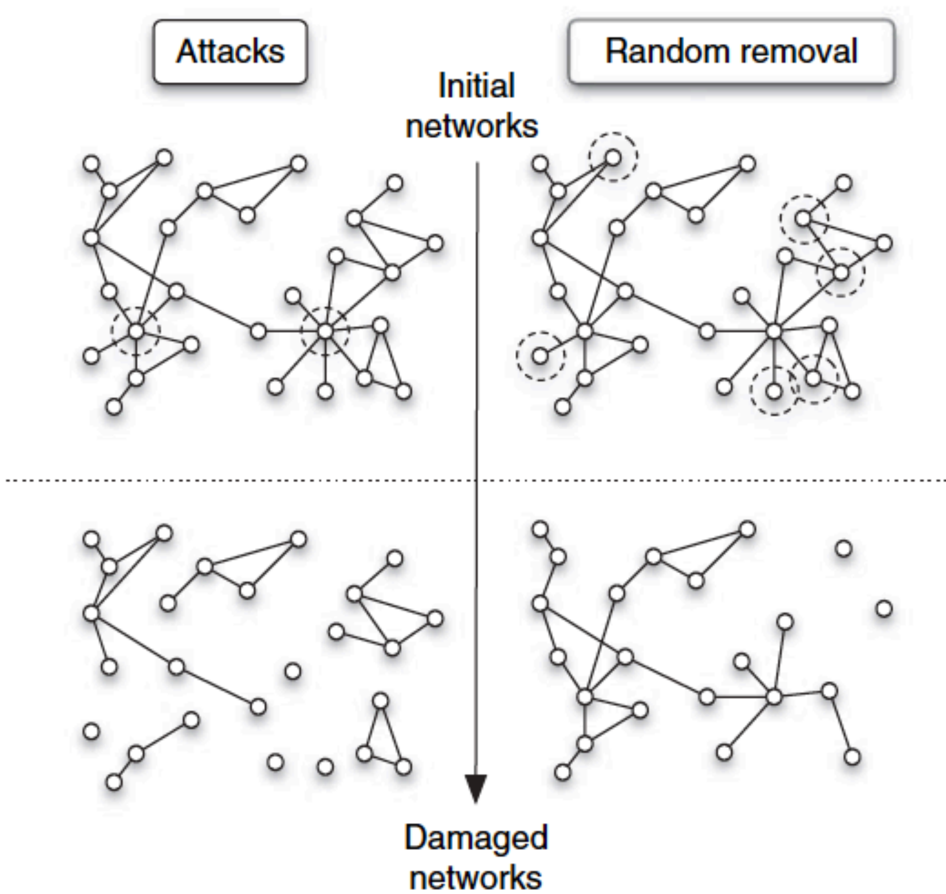
Dynamical measures on network: robustness

Two models for node removal:

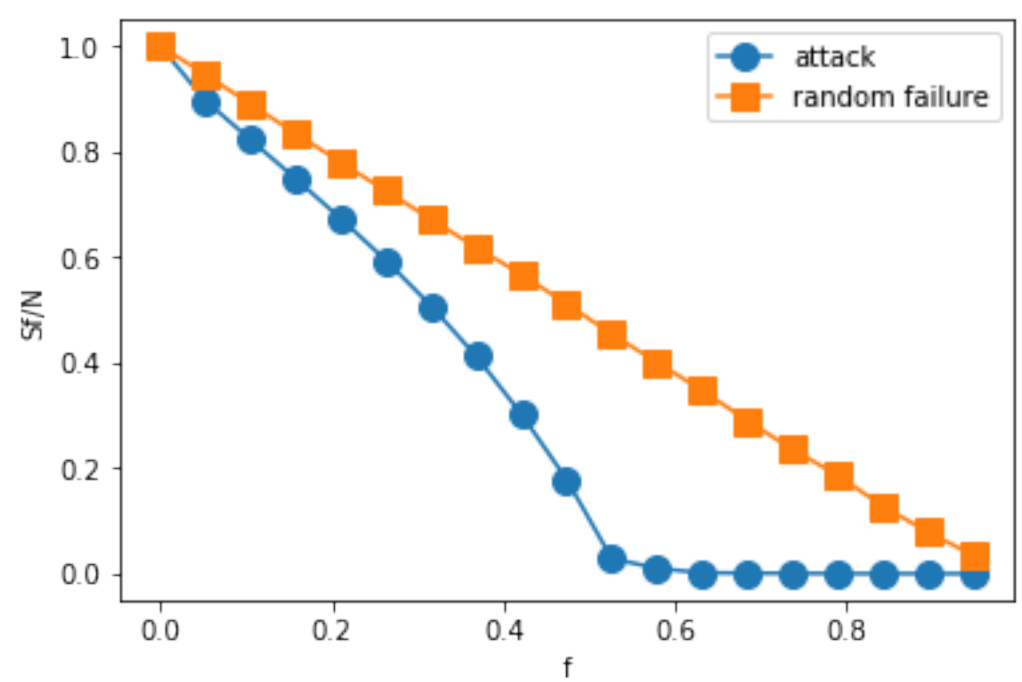
Network robustness

Targeted attacks: nodes are removed according to their centrality

Random failures: nodes are removed randomly



At each iteration we remove a fraction of nodes (according to the selected model) and we calculate the fraction of nodes that remain connected.



Pareto distributions

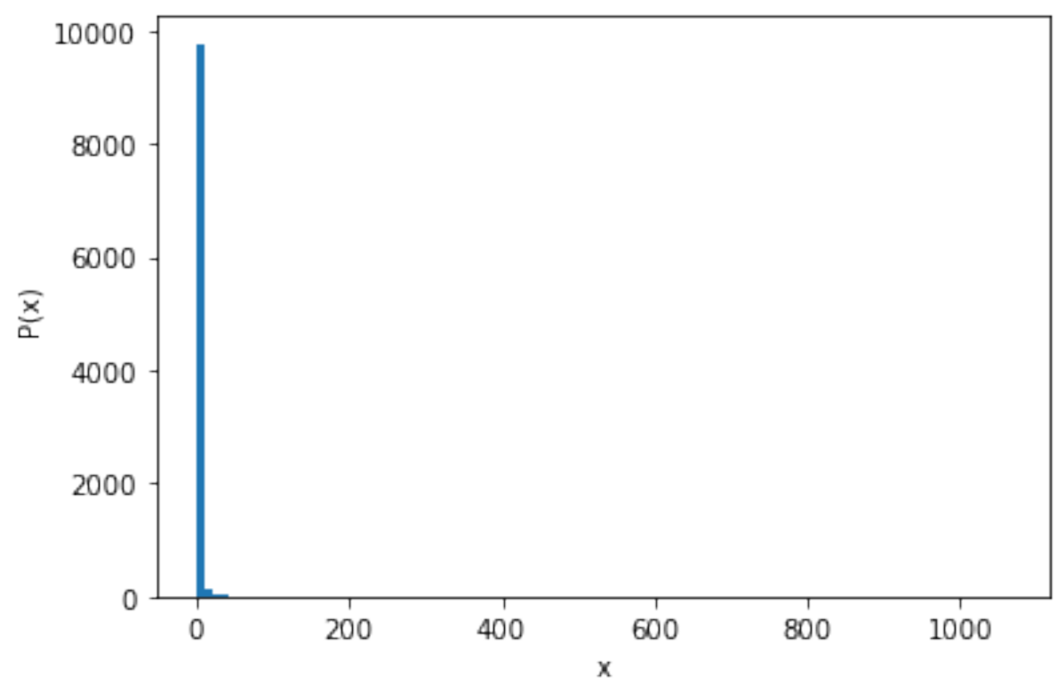
In the following lessons we will see that often measures from real network datasets follow Pareto distributions:

$$P(x) = \alpha x^{-\gamma}$$

A first way to identify Pareto distribution is trough a visual inspection. But is is not a trivial task…

Pareto distributions

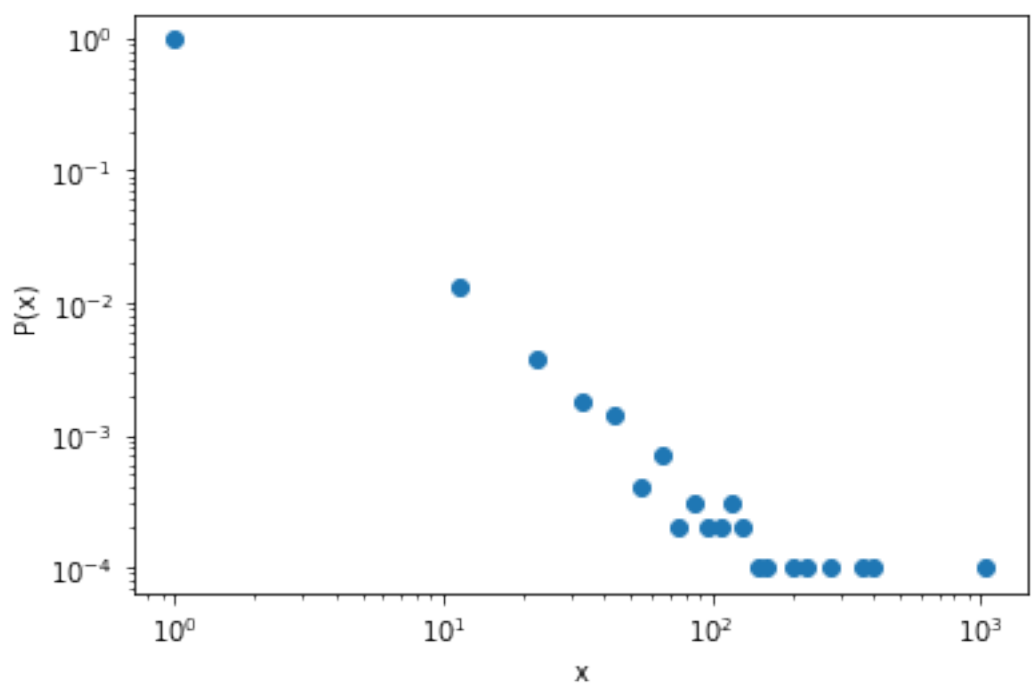
1- If we plot a simple an histogram of the distribution we observe a result of this type:



A binned degree histogram will give poor statistics at the tail of the distribution: for large values, in every bin there will be only a few samples. Impossible to observe something visually.

Pareto distributions

2- To better visualize the distribution we can therefore use a logarithmic plot log(P(x))-vs- log(x)



$$P(x) = \alpha x^{-\gamma}$$

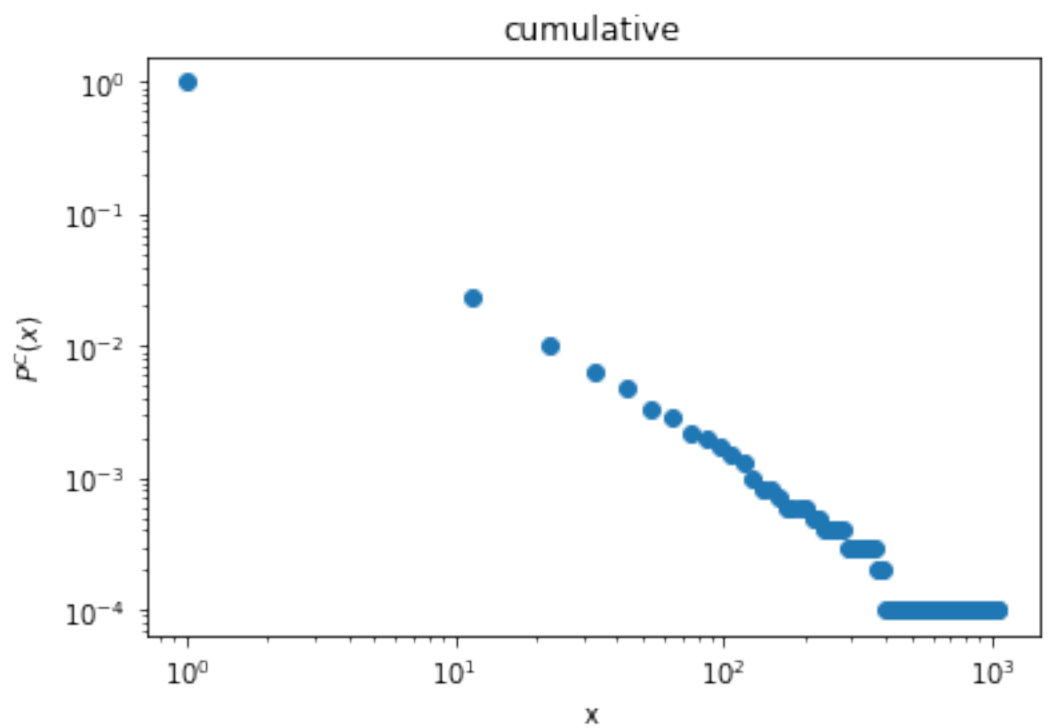
$$log(P(x)) = log(\alpha x^{-\gamma}) =$$
$$= log(a) - \gamma log(x)$$

In the log-log space we expect to see a linear behavior.

However the poor statistics for large values still makes difficult to identify the linear behavior in the tail.

Pareto distributions

3- A first trick to improve visualization is to go to the cumulative distribution



$$P(x) = \alpha x^{-\gamma}$$

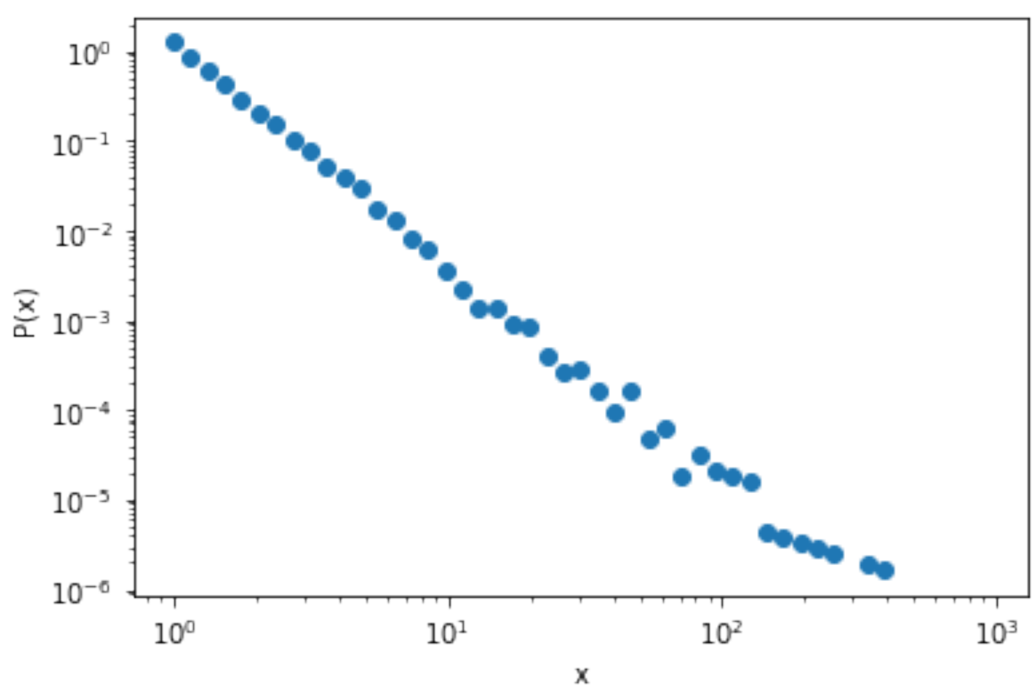
$$P^C(x) = \int \alpha x^{-\gamma} dx =$$

$$= \frac{\alpha}{1-\gamma} x^{1-\gamma}$$

In the log-log space we expect again to see a linear behavior.

Pareto distributions

4- The best solution is to use logarithmic binning: we divide the x range in bins that are larger in the tail, in order to have a better statistics for large values



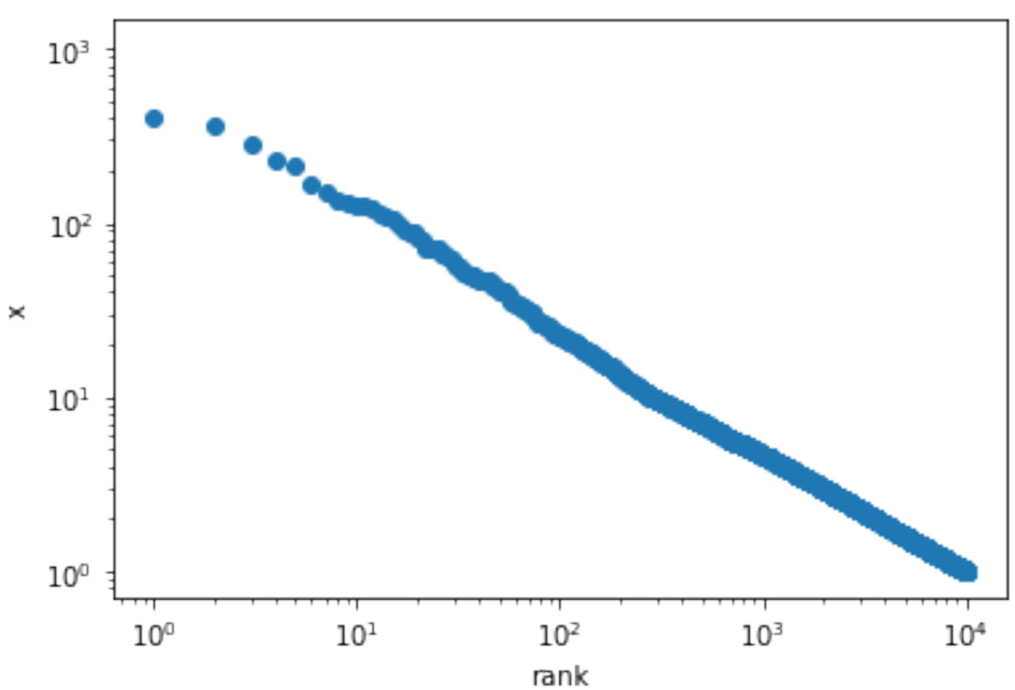
In this scheme each bin is made wider than its predecessor by a constant factor:

- First bin $b^0 \leq x < b^1$
- Second bin $b^1 \leq x < b^2$
- n-th bin $b^{n-1} \leq x < b^n$

Careful at normalizing the bins correctly: a bin of width b^2 will get b times more samples compared to the previous bin of width b! Divide the count number by the bin width!

Pareto distributions

5- Another possibility, if the sample is not too big, is to plot the values according to their ranking, from the larger to the smaller.



This is the representation in form of the Zipf's law.