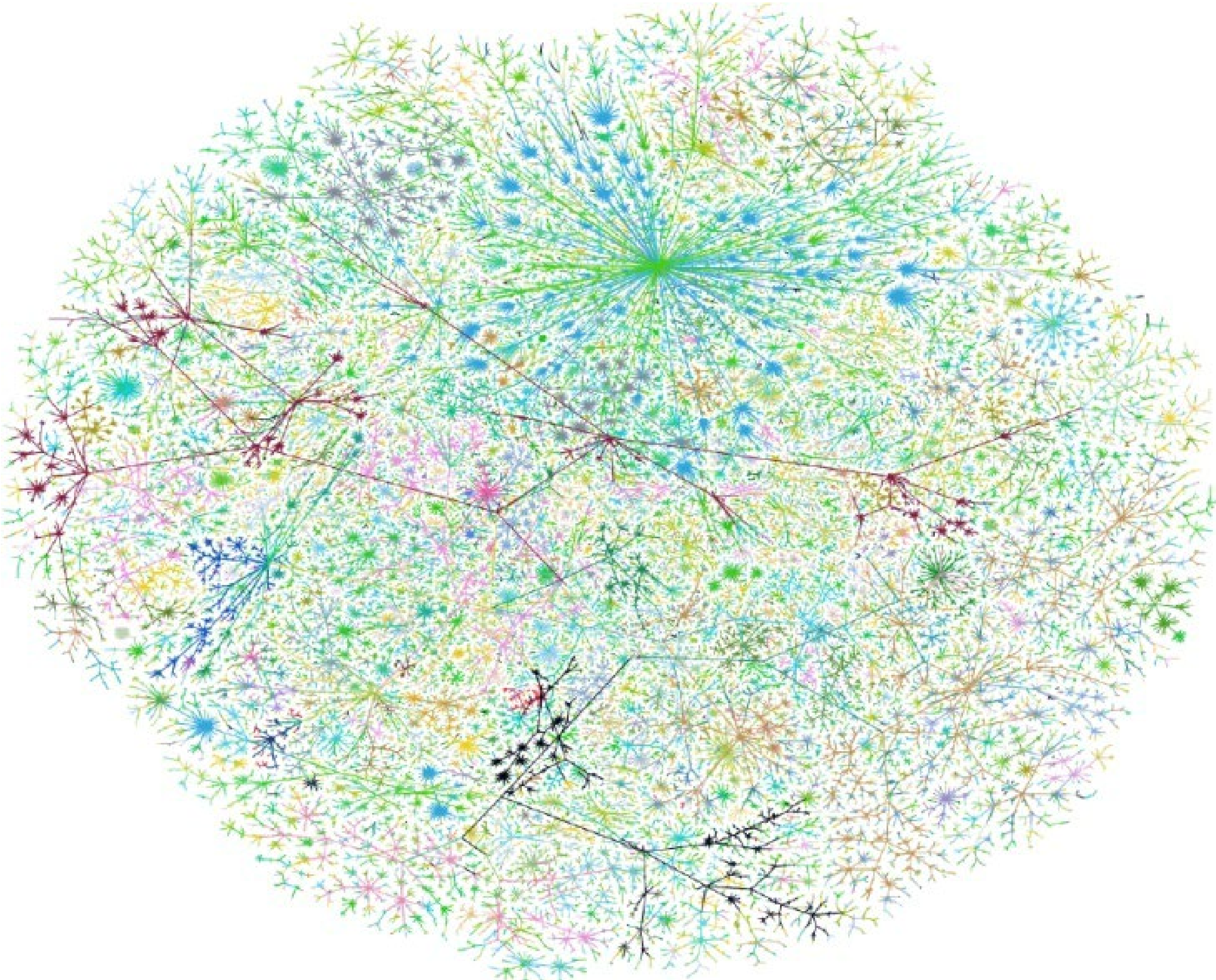
# Analisi di Network

# Network e sistemi complessi

- I sistemi complessi sono attorno a noi:
  - La Società è una collezione di 6 miliardi di individui;
  - I sistemi di comunicazione  collegano device elettronici;
  - L'informazione e la  conoscenza sono legati
  - L'interazione tra migliaia di geni che regolano la vita
  - I nostri pensieri sono legati da connessioni nascoste tra miliardi di neuroni nel nostro cervello
- Cosa hanno in comune?
- Come possiamo rappresentarli?

# La network

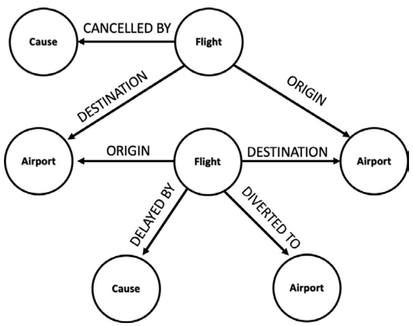


- Dietro diversi sistemi c’è un intricato diagramma di connessioni, **una network**, che definisce le interazioni tra le componenti.

- **Non capiremo mai questi sistemi fino a quando non capiremo le reti che ci stanno dietro.**

- Una network è una collezione di oggetti dove alcune coppie di oggetti sono connesse da link.

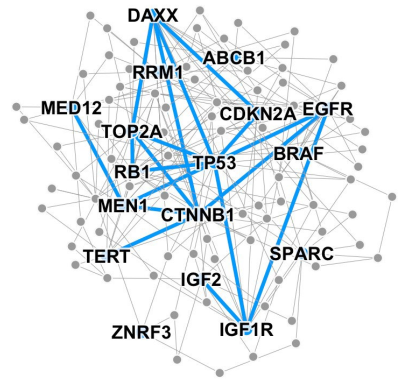
# Tipi di reti




**Grafi di Eventi**

Immagine da: SalientNetworks

**Computer Network**




**Pathway**

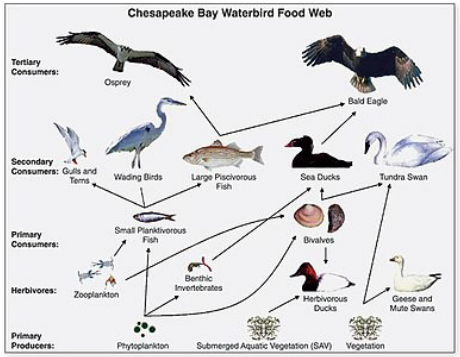



Immagine da: Wikipedia

**Food Web**

Immagine da: Pinterest

**Particle Networks**

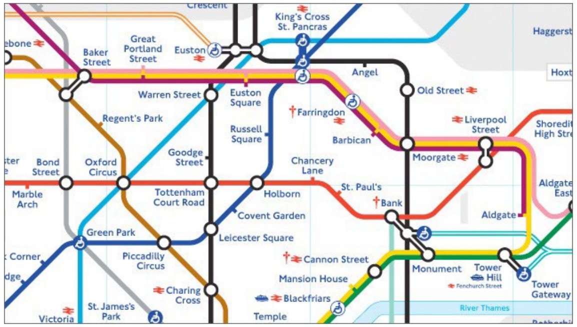



Immagine daa: visitlondon.com

**Metro netwrok**

# Tipi di reti

Immagine da: Medium

**Social Network**

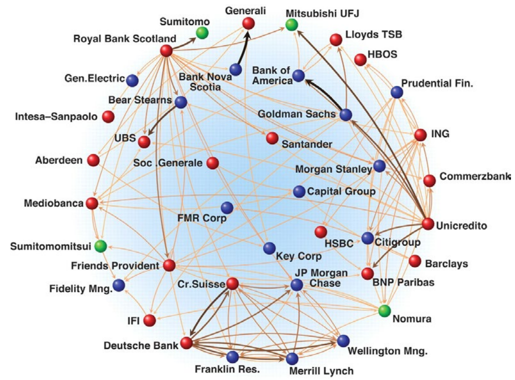
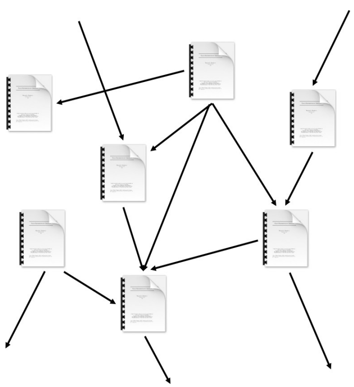



Immagine da: Science

**Economic Network**

Immagine da : Lumen Learning

**Communication Network**



**Citation Network**


Immagine da : La7

**Internet**

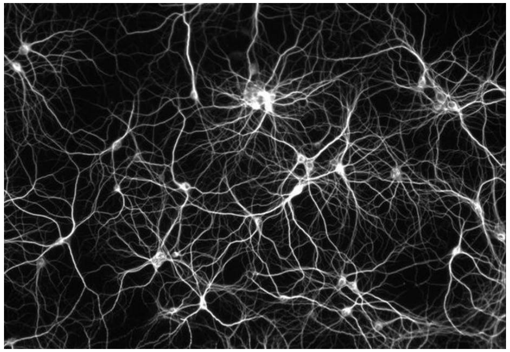


Immagine da: The Conversation

**Networks of Neuron**

# Tipi di reti

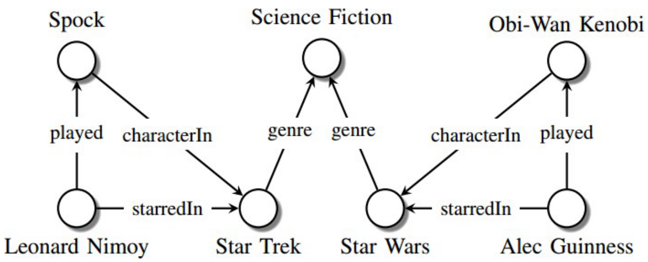



Immagine da: Maximilian Nickel et al

**Knowledge Graph**

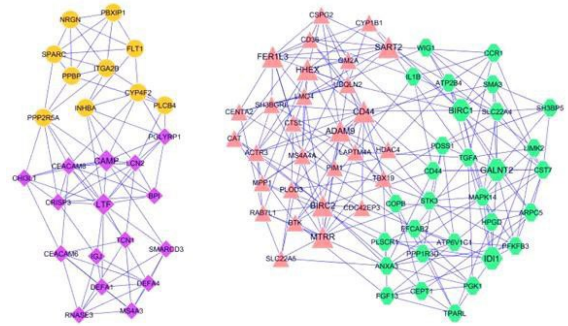


Immagine da: ese.wustl.edu

**Regulatory Network**

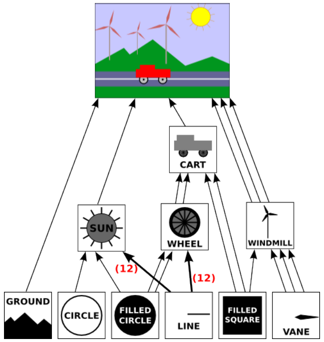



Immagine da: math.hws.edu

**Scene Graph**

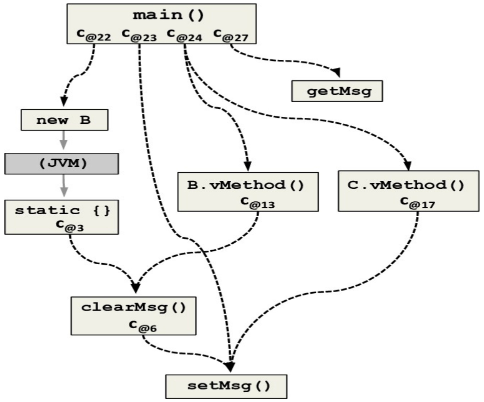



Immagine da: ResearchGate

**Codice**

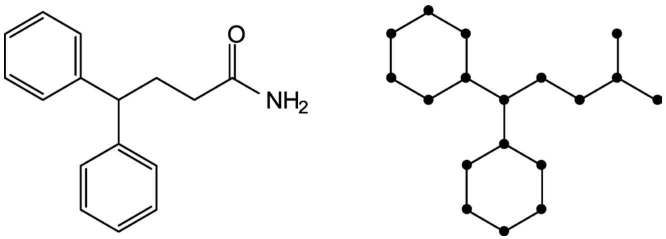



Immagine da: MDPI

**Molecole**

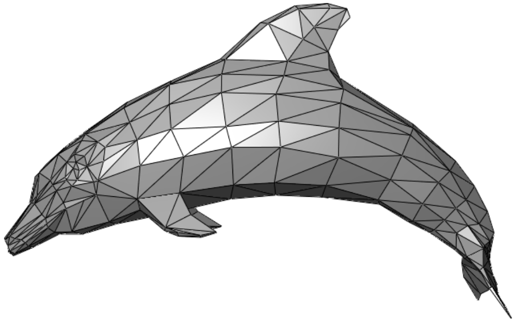
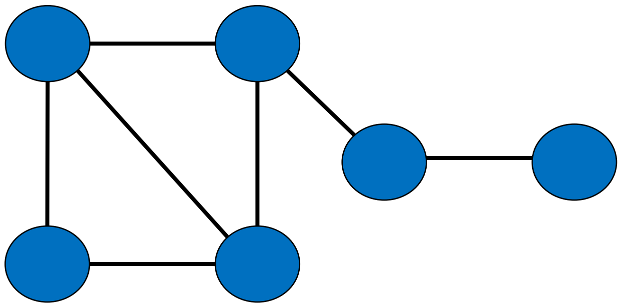


Immagine da: Wikipedia

**3D Shape**

# Componenti di una network

- Oggetti: nodi, vertici ***N***
- Interazioni: archi, link ***E***
- Sistema: Grafo network ***G(N,E)***



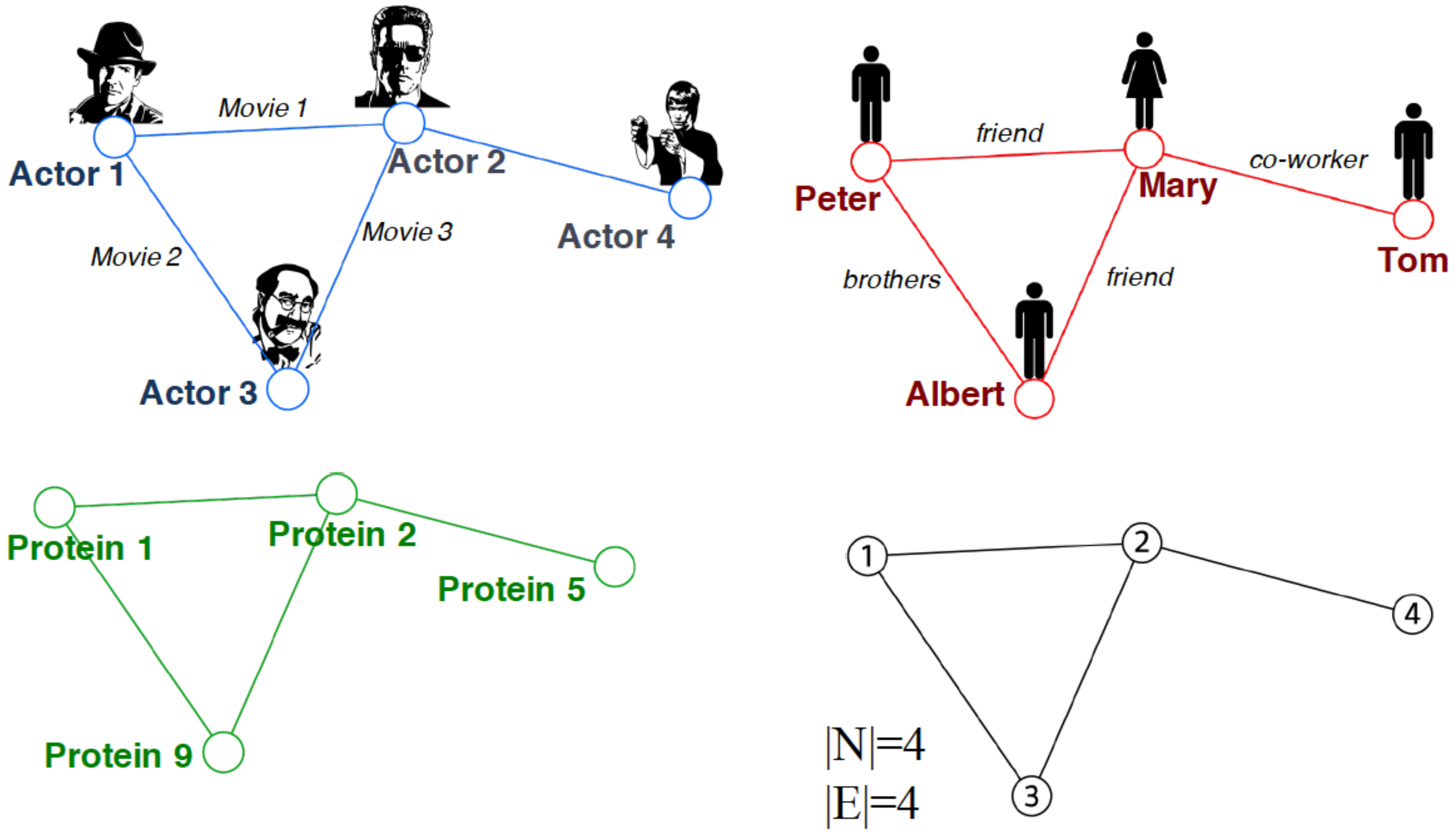
# Network o Grafo?

- **Network** fa spesso riferimento a sistemi reali
  - Web, social network, Network Metabolico
  - Linguaggio: Network, nodo, link

- **Grafo** è la rappresentazione matematica di una network
  - Grafo del web, social graph (termine di Facebook)
  - Linguaggio: Grafo, vertice, arco

- Solitamente li useremo come termini intercambiabili, in ogni caso faremo questa distinzione se appropriato.

# Network: linguaggio comune

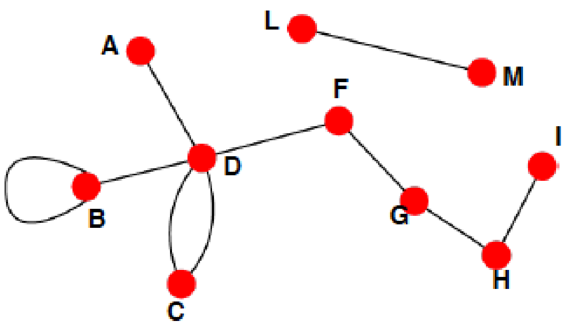


# Come definire una rete?

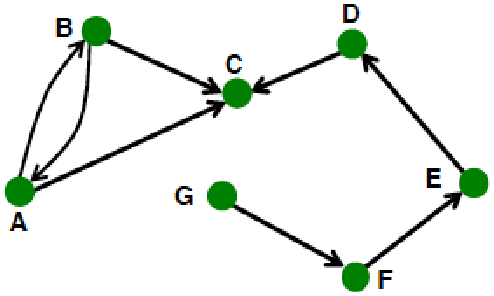
- Come costruire un grafo?
  - Quali sono i nodi?
  - Quali sono gli archi?
  - Scegliere la rappresentazione appropriata di un domino/problema determina la nostra abilità nell'usare con successo le network.
    - In alcuni casi c'è una unica non ambigua rappresentazione
    - In altri casi la rappresentazione non è unica
    - Il modo in cui si assegnano i link determina la questione che si desidera studiare

# Grafi diretti vs non direzionati

**Non direzionati**




**Diretti**




**Esempi:**

Collaborazioni

Amicizia

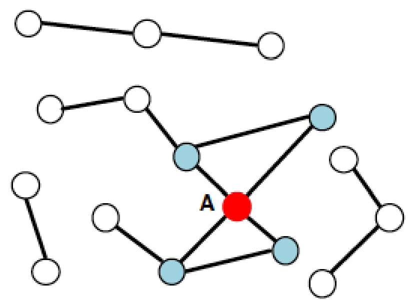
**Esempi**

Chiamate telefoniche

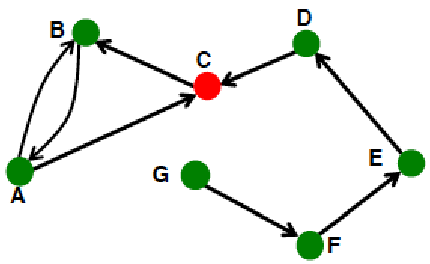
Following su Twitter

# Grado dei nodi

**Non direzionati**




**Diretti**




Sorgente: $k^{in}$
Destinazione $k^{out}$

- **Grado (degree)** di un nodo $k_i$: numero di archi adiacenti al nodo *i*
- $k_A$=4
- ***Degree Medio***:

$$\bar{k} = \langle k \rangle = \frac{1}{N}\sum_{i=1}^{N} k_i = \frac{2E}{N}$$

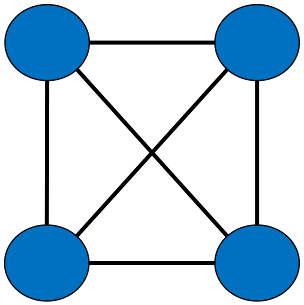
- *Nelle reti dirette definiamo in-degree e out-degree. Il grado del nodo è la somma di in-degree e out-degree.*
- $k_c^{in} = 2 \;\; k_c^{out} = 1 \;\; k_c = 3$
- :$\bar{k} = \frac{E}{N} \qquad \bar{k}^{in} = \bar{k}^{out}$

# Grafo completo

- Il numero massimo di archi in un grafo non direzionato con N nodi è

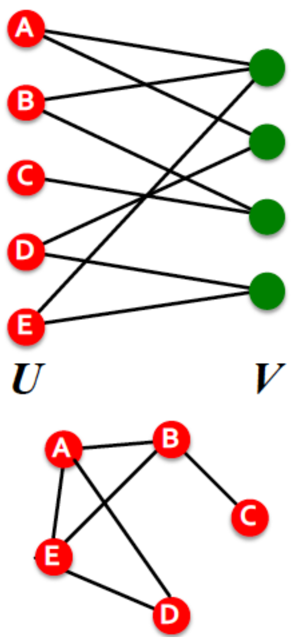
$$E_{max} = \binom{N}{2} = \frac{N(N-1)}{2}$$

Un grafo non diretto con $\boldsymbol{E = E_{max}}$ è chiamato grafo completo ed il suo degree medio è ***N-1***



# Grafo bipartito

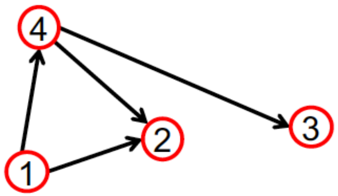
- In un **Grafo Bipartito** i nodi sono divisi in due insiemi disgiunti: U e V tale che ogni link connette un nodo di U ad un nodo di V




- Esempi:
  - Utenti-Film (pesata)
  - Articoli-Autori
  - Attori-Film
- Proiezione della rete su U (o su V)
  - Rete di collaborazione
  - Co-rating dei film

# Rappresentazione di un grafo

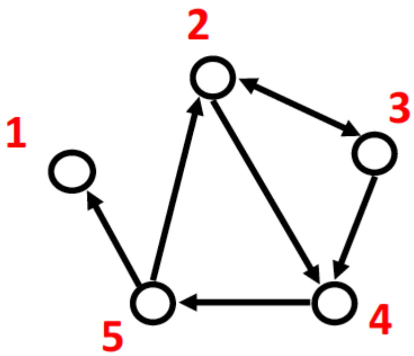
**Matrice di Adiacenza** **Lista di archi** **Lista di adiacenza**




$$A = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

- (2, 3)
- (2, 4)
- (3, 2)
- (3, 4)
- (4, 5)
- (5, 2)
- (5, 1)

- 1:
- 2: 3, 4
- 3: 2, 4
- 4: 5
- 5: 1, 2



- La maggior parte delle reti reali sono sparse

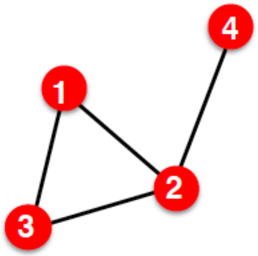
$$E << E_{max} \quad (\bar{k} \ll N - 1)$$

# Attributi sugli archi

- Possibili opzioni
  - Peso
  - Rank (migliore amico, secondo migliore amico, ecc)
  - Tipo (Amico, parente, collega)
  - Segno (Vero vs Falso)

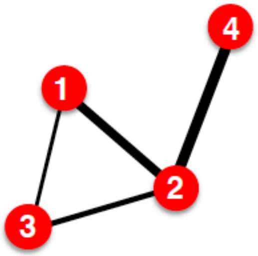
# Tipi di grafi

**Non pesata**



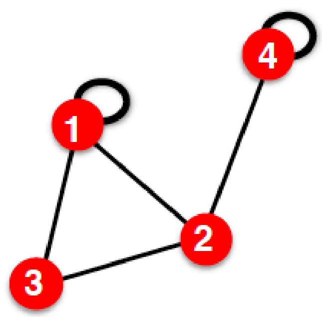

$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

**Pesata**



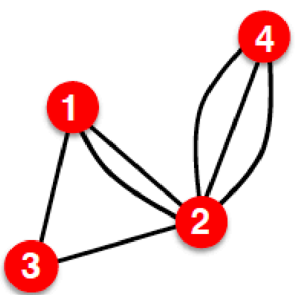

$$A_{ij} = \begin{pmatrix} 0 & 2 & 0.5 & 0 \\ 2 & 0 & 1 & 4 \\ 0.5 & 1 & 0 & 0 \\ 0 & 4 & 0 & 0 \end{pmatrix}$$

## Self-edge (self loop)




$$A_{ij} = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

## Multigrafo




$$A_{ij} = \begin{pmatrix} 0 & 2 & 1 & 0 \\ 2 & 0 & 1 & 3 \\ 1 & 1 & 0 & 0 \\ 0 & 3 & 0 & 0 \end{pmatrix}$$

# Grafi multilayer e multiplex

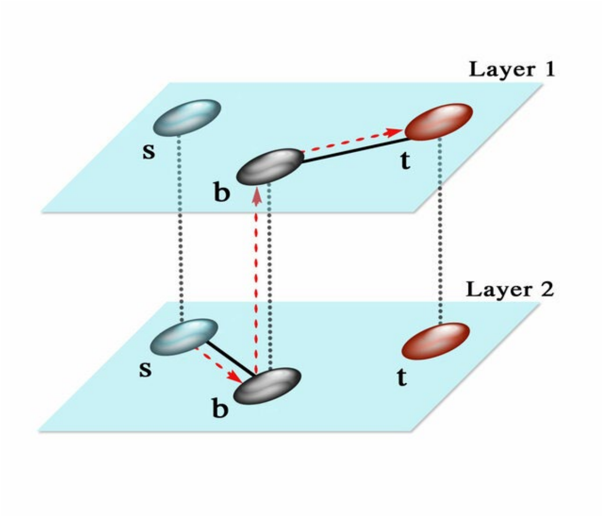



Immagine da: [ResearchGate]

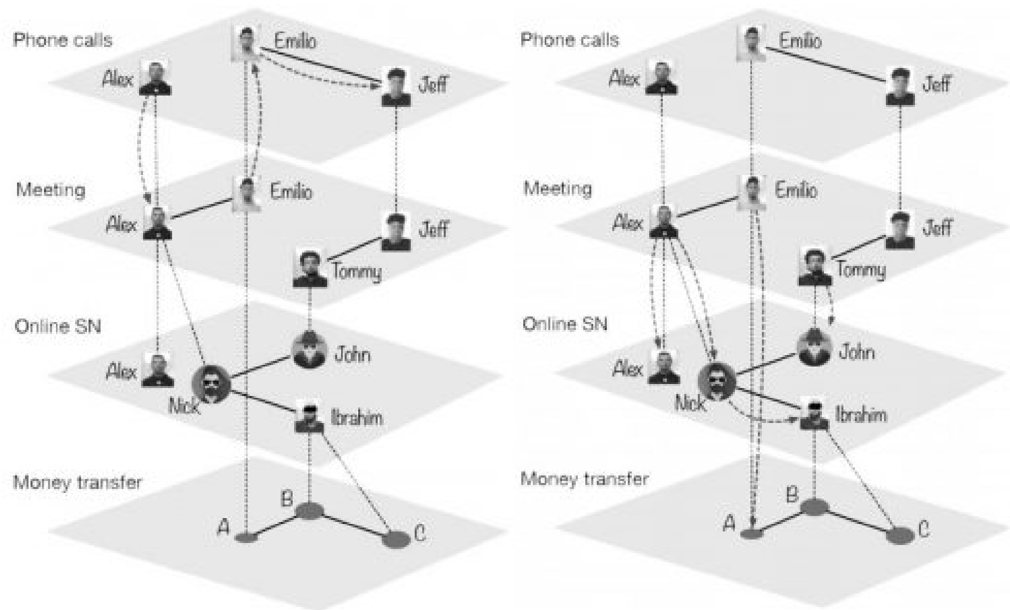
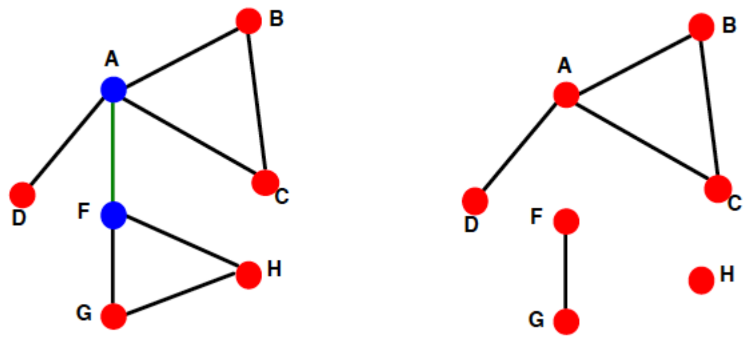



Immagine da:
https://www.knowledge-share.eu/

# Connettività dei grafi non direzionati

- Grafo connesso:
  - Due vertici qualsiasi possono essere connessi da un cammino
- Un grafo disconnesso è fatto da più componenti connesse



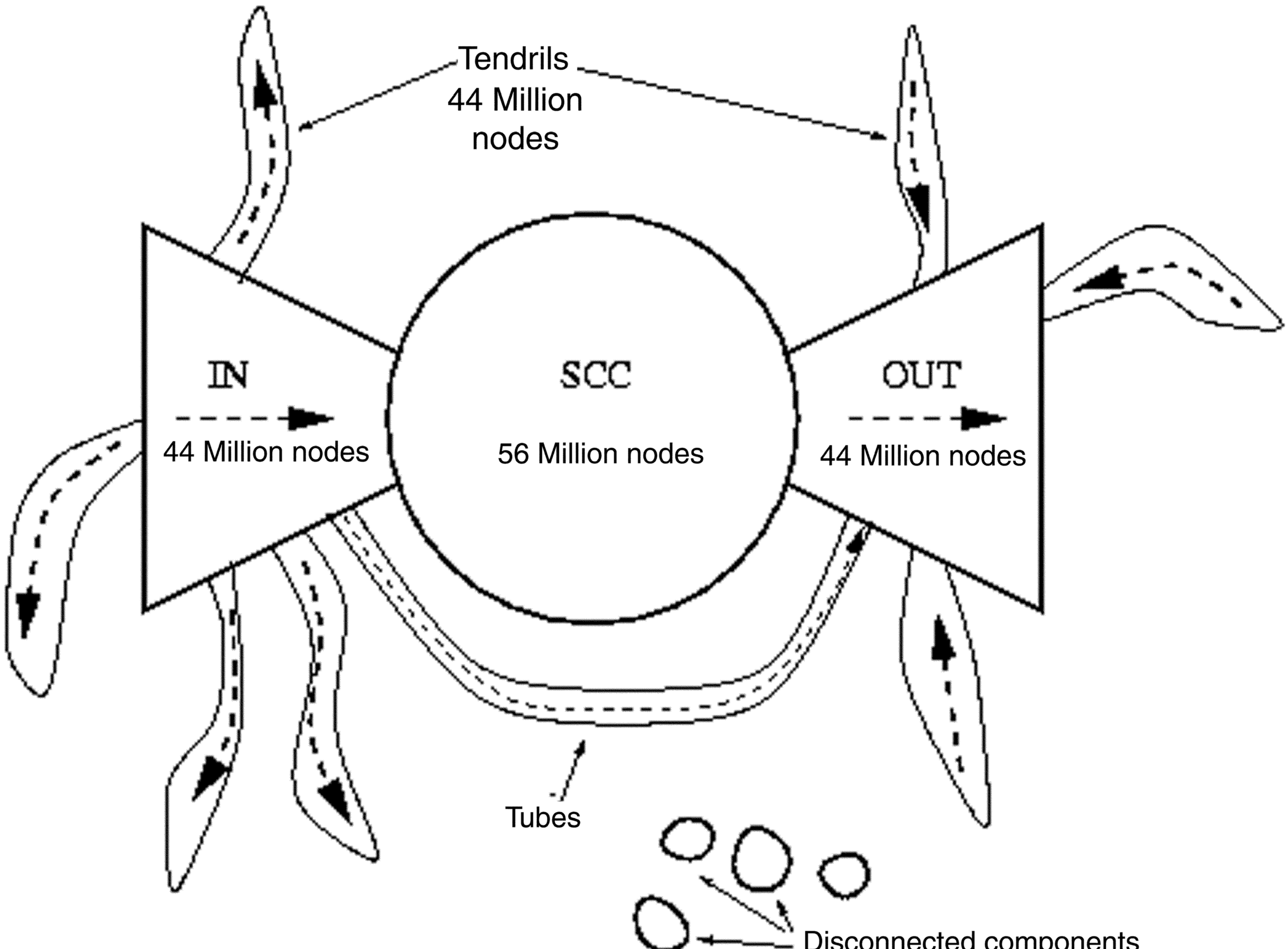

- **Arco ponte**: Se cancelliamo l'arco verde il grafo diventa disconnesso
- **Punto di articolazione**: se lo cancelliamo il grafo diventa disconnesso
- Componente connessa più grande: **Componente connessa gigante**
- Nodo isolato (H)

# Connettività dei grafi diretti

- Componente fortemente connessa
  - Per ogni coppia di nodi esiste un path che li connette (A → B e B→A)
- Grafi diretti con connessioni deboli
  - Connesso se togliamo la direzione agli archi

- WWW → Multigrafo diretto con self-edge
- Facebook amicizia → non direzionato non pesato
- Rete citazione → non pesato, diretto, aciclico
- Rete collaborazioni → non direzionato, multigrafo o pesato
- Rete chiamate cellulare → Diretto, pesato, multigrafo
- Protein Protein Interaction Network → non direzionata, non pesata, self-edge

# Struttura del Web (203M pagine 1.5B archi) [Broder et al. 2000]



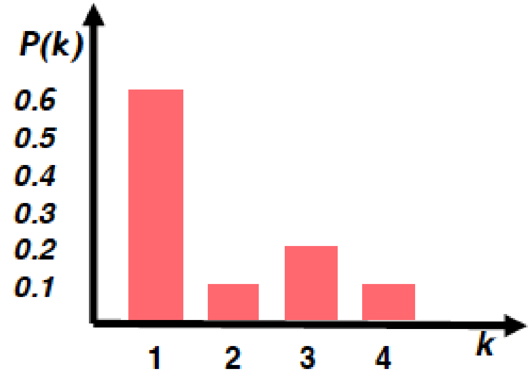
# Proprietà di una rete

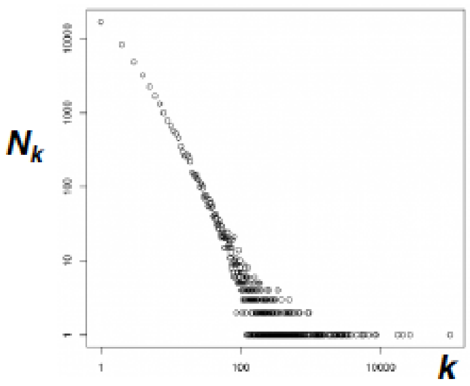
- Come possiamo descrivere una rete?
- Cosa possiamo misurare in una rete?
- Proprietà chiave di una rete:
  - Distribuzione dei degree: $P(k)$
  - Lunghezza dei path: $h$
  - Coefficiente di Clustering: $C$

# Degree distribution

- ***Degree distribution P(k):*** Probabilità che preso a caso un nodo questo abbia degree *k*
  - $N_k$ = *numero di nodi con degree*
- Istogramma normalizzato:

$$P(k) = \frac{N_k}{N}$$



# Cammini in un grafo

- Un cammino (path) è una sequenza di nodi dove ogni nodo è collegato la successivo
- $P_n = \{i_0, i_1, i_2, \dots, i_n\}$ $\quad P_n = \{(i_0, i_1), (i_1, i_2), \dots, (i_{n-1}, i_n)\}$
- Un path può intersecare se stesso e passare più volte dallo stesso arco.
- Nei grafi diretti ha direzione degli archi

# Numero di cammini

- Come calcolare il numero di cammini tra u e v?
- Lunghezza $h = 1$: se esiste l'arco tra $u$ e $v$ $A_{uv} = 1$ altrimenti $A_{uv} = 0$
- Lunghezza h=2: Se esiste un path di lunghezza 2 tra u e v allora $A_{uk} A_{kv}=1$ altrimenti $A_{uk} A_{kv}=0$

$$H_{uv}^{(2)} = \sum_{k=1}^{N} A_{uk} A_{kv} = [A^2]_{uv}$$

- Lunghezza h, il numero di path tra u e v di lunghezza h è:

$$H_{uv}^{(h)} = [A^h]_{uv}$$

# Distanza in un grafo

- Cammino minimo
- Come lo calcoliamo?

# Diametro di una rete

- **Diametro**: massima distanza (shortest path) tra una coppia qualsiasi di nodi del grafo
- Lunghezza media di un path per un grafo connesso (o componente) o una componente fortemente connessa di un grafo diretto

$$\bar{h} = \frac{1}{2E_{max}} \sum_{i \neq j} h_{ij}$$

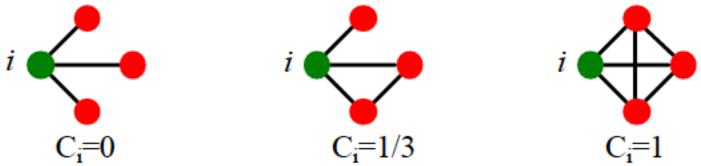
- Dove $h_{ij}$ è la distanza tra i nodi i e j

# Coefficienet di clustering

- Coefficiente di clustering:
  - Quale porzione dei vicini del nodo *i* sono connessi?
  - Il nodo i ha degree $k_i$
  - $C_i \in [0,1]$

$$C_i = \frac{2e_i}{k_i(k_i-1)}$$

  - dove $e_i$ *è il numero di archi tra i vicini del nodo* ***i***




  - **Coefficiente di clustering medio**

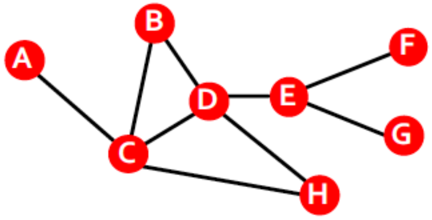
$$C = \frac{1}{N}\sum_{i}^{N} C_i$$

# Coefficiente di clustering

- Coefficiente di clustering:
  - Quale porzione dei vicini del nodo *i* sono connessi?
  - Il nodo i ha degree $k_i$

$$C_i = \frac{2e_i}{k_i(k_i-1)}$$

  - dove $e_i$ è *il numero di archi tra i vicini del nodo* ***i***



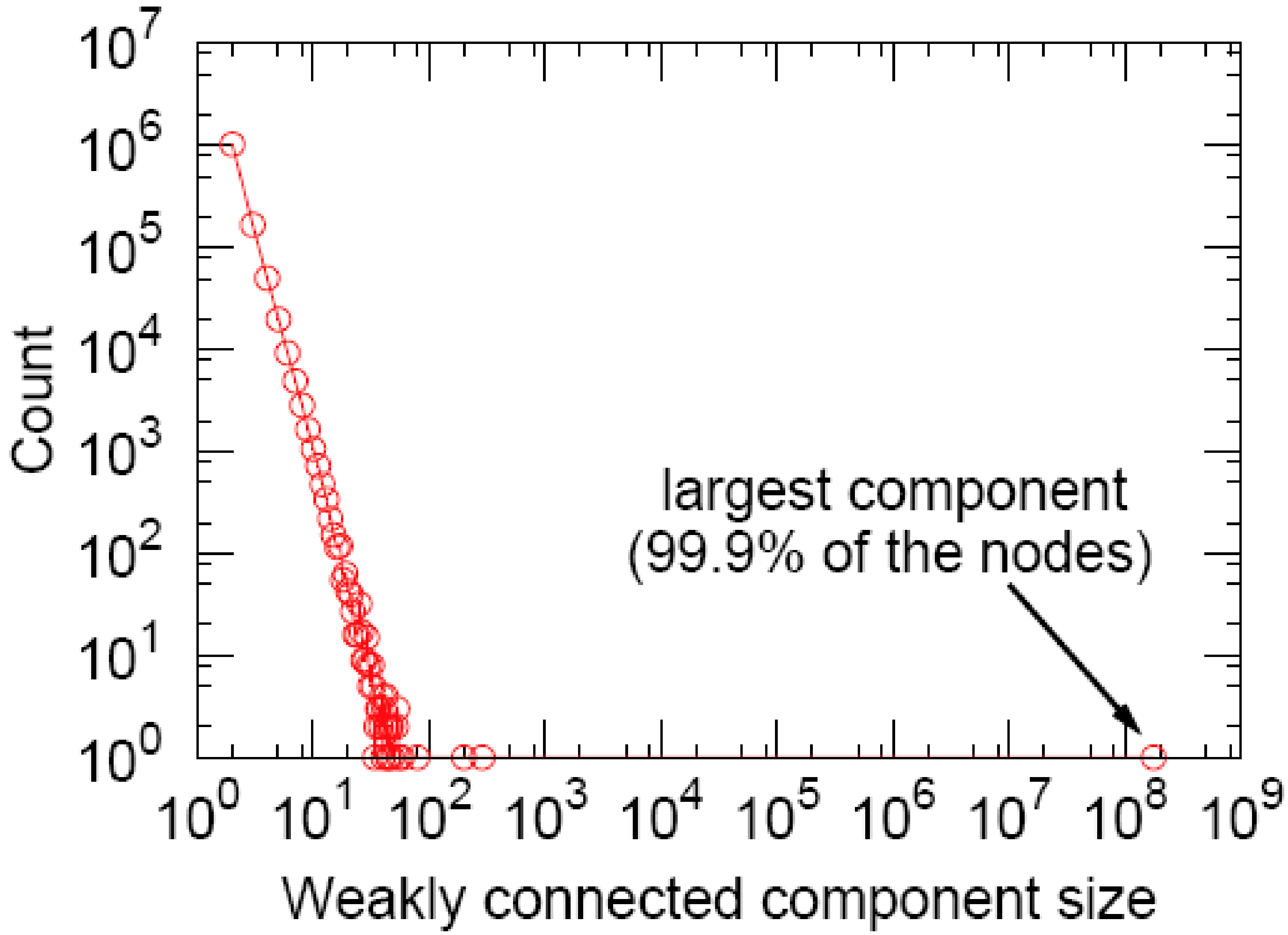

$k_B=2,\ e_B=1,\ C_B=2/2 = 1$

$k_D=4,\ e_D=2,\ C_D=4/12 = 1/3$

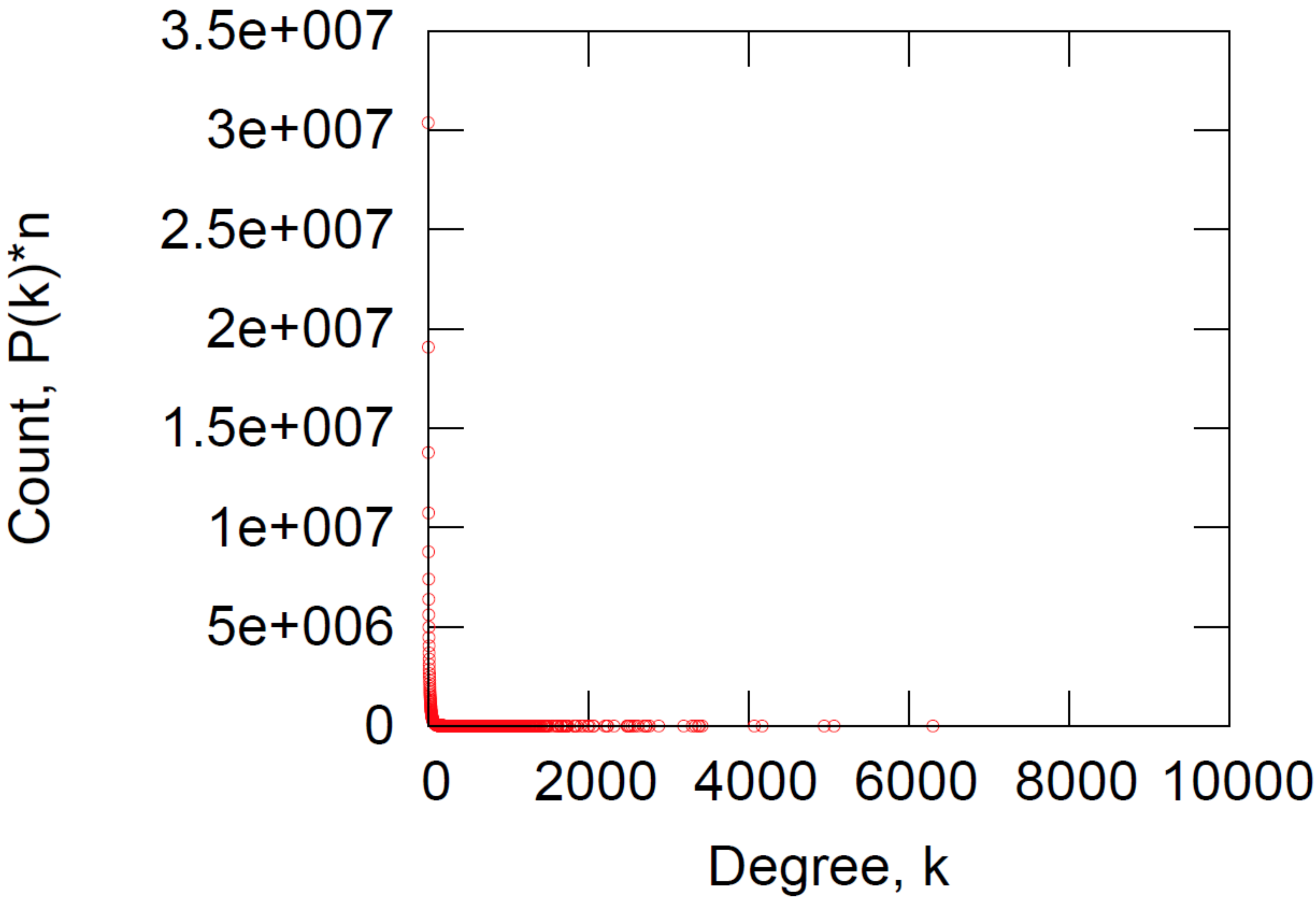
# Un esempio reale: MSN Messenger

- Attività MSN Messenger giugno 2006
  - 245M utenti
  - 180M coinvolti in conversazioni
  - Più di 30 miliardi di conversazioni
  - Più di 255 miliardi di messaggi scambiati

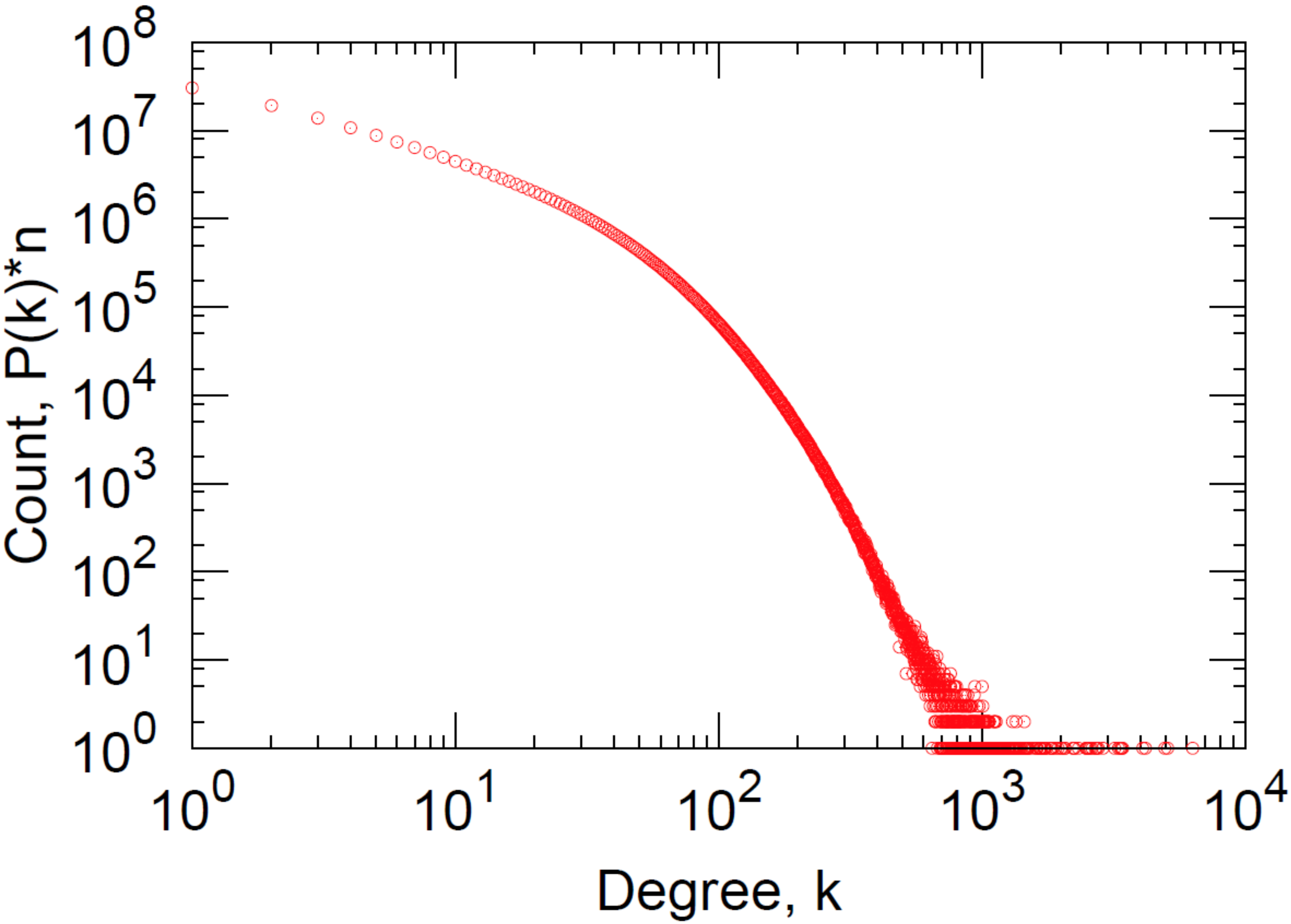
# Connettività



# Degree distribution

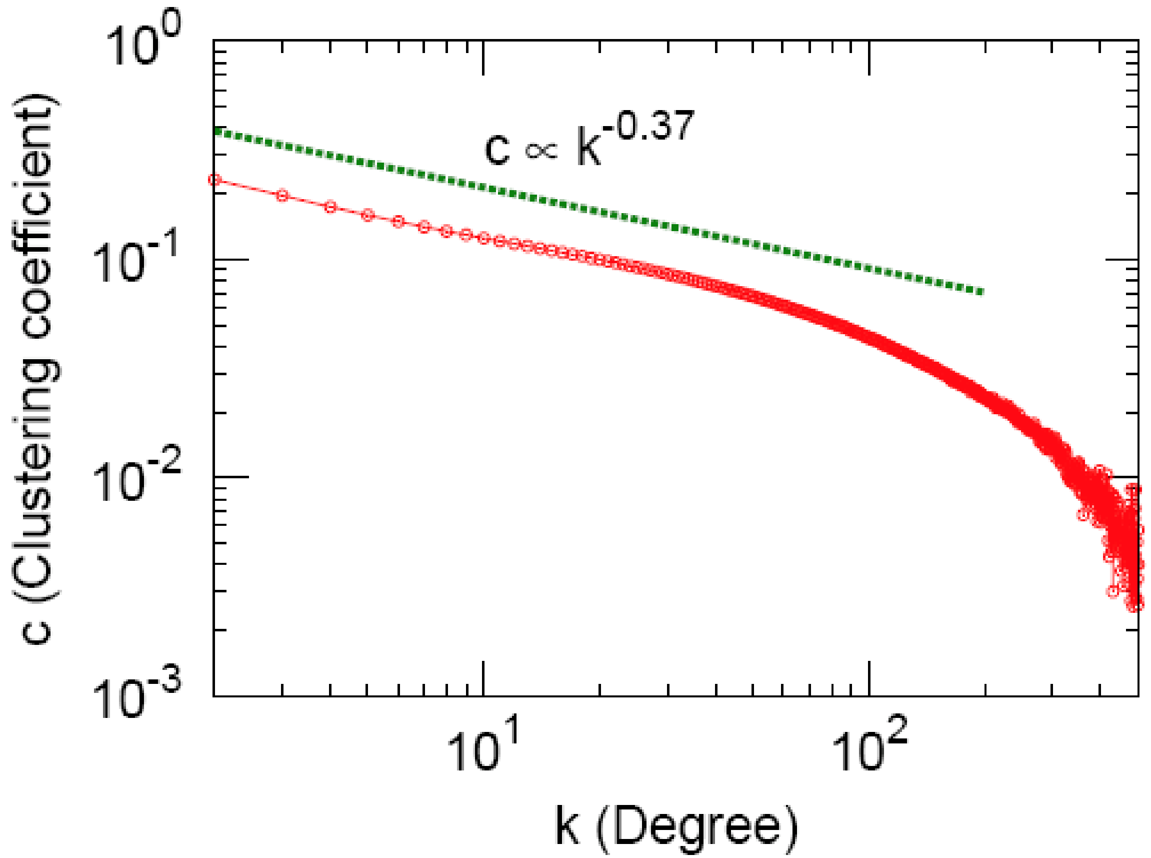


# Log-Log Degree distribution

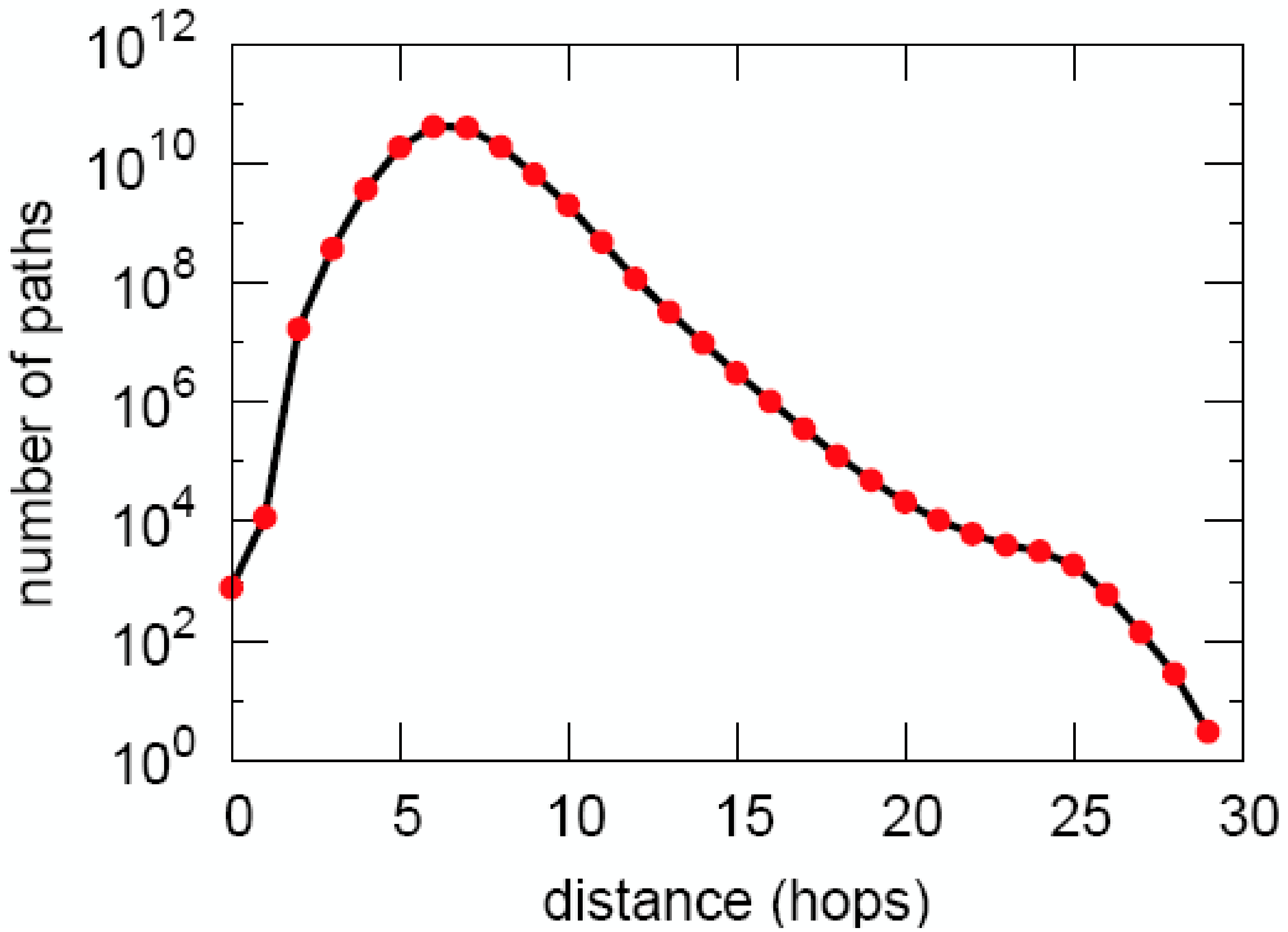


# Coefficiente Clustering

- $C_k$: $C_i$ *medio dei nodi i con degree* $k$, $C_k = \frac{1}{N_k}\sum_{i:k_i=k} C_i$



# Diametro



$10^{12}$
$10^{10}$
$10^{8}$
$10^{6}$
$10^{4}$
$10^{2}$
$10^{0}$
number of paths
0
5
10
15
20
25
30
distance (hops)

# Proprietà chiave MSN

- Distribuzione dei degree: ***Degree medio 14.4***
- Lunghezza dei path: ***6.6***
- Coefficiente di Clustering: ***0.11***

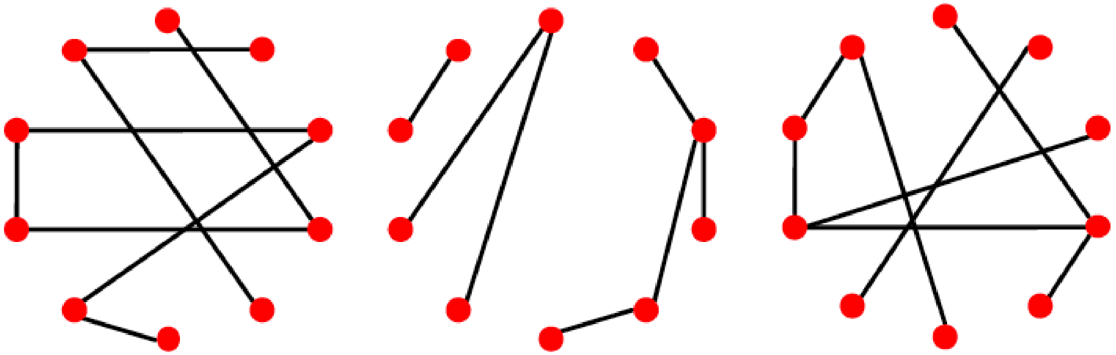
- Come possiamo valutare queste proprietà?
  - Attese?
  - Sorprendenti?
- Per rispondere a tale domanda abbiamo bisogno di un **modello nullo!**

# Erdös-Renyi Random Graph Model

- Due Varianti
  - $G_{n,p}$: grafo non direzionato con *n* nodi e ogni arco *(u,v)* appare con una probabilità *p*
  - $G_{n,m}$: grafo non direzionato con *n* nodi e *m* archi selezionati in modo random

# Random Graph model

- **n e p non determinano in modo univoco il grafo**
  - Il grafo è il risultato di un processo random
  - Possiamo ottenere diversi grafi con gli stessi *n* e *p*



n = 10
p= 1/6

# Quanti archi?

- Con quale probabilità il grafo avrà E archi?
- P(E) probabilità che $G_{n,p}$ generi un grafo con esattamente E archi:

$$P(E) = \binom{E_{max}}{E} p^E (1-p)^{E_{max}-E}$$

- Dove $E_{max} = n(n-1)/2$ è il numero massimo di archi in un grafo indiretto con n nodi.
- P(E) segue una distribuzione Binomiale

# Degree dei nodi nel grafo Random

- Qual è il grado atteso di un nodo?
- Sia $X_v$ una variabile random che misuri il grado del nodo $v$
- Vogliamo calcolare $E[X_v] = \sum_{j=0}^{N-1} jP(X_v = j)$
- Decomponiamo $X_v$ come $X_v = X_{v,1} + X_{v,2} + \cdots + X_{v,n-1}$
- Dove $X_{v,u}$ è una variabile random che assume i valori {0,1} e ci dice se l'arco esiste o meno.
- $E[X_v] = \sum_{u=1}^{n-1} E[X_{vu}] = (n-1)p$

# Analizziamo le proprietà chiave di $G_{n,p}$

- La distribuzione di $G_{n,p}$ è Binomiale
- P(k) frazione di nodi con degree k

$$P(k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}$$

- $\bar{k} = p(n-1)$
- $\sigma^2 = p(1-p)(n-1)$

# Coefficiente di clustering di $G_{n,p}$

- Ricordiamo che: $C_i = \frac{2e_i}{k_i(k_i-1)}$ con $e_i$ numero di archi tra i vicini di i
- Gli archi in $G_{n,p}$ appaiono con probabilità p
- Deduciamo quindi che: $e_i = p\frac{k_i(k_i-1)}{2}$
- Ne segue che:
- $C_i = \frac{pk_i(k_i-1)}{k_i(k_i-1)} = p = \frac{\bar{k}}{n-1} \approx \frac{\bar{k}}{n}$
- Il coefficiente di clustering di un grafo random è piccolo
- Per un degree medio fissato, *C* decresce con la dimensione *n* del grafo

# Proprietà di $G_{n,p}$

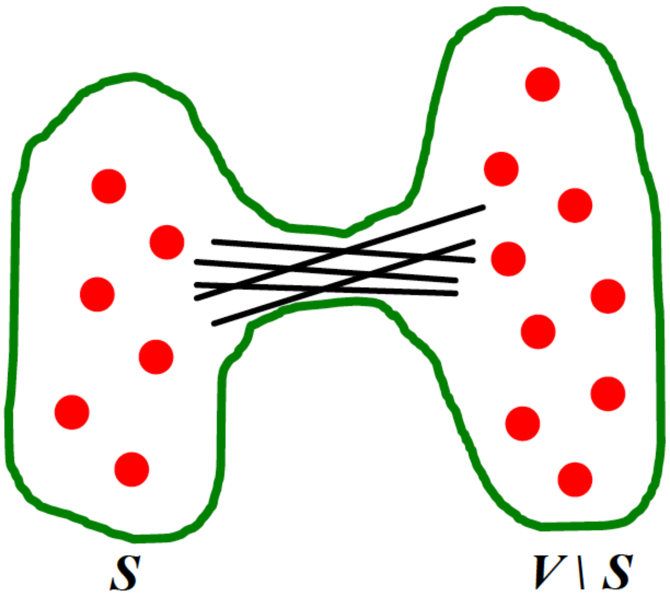
- Distribuzione dei degree: $P(k) = \binom{n-1}{k} p^k (1-p)^{n-1-j}$
- Lunghezza dei path: ?
- Coefficiente di Clustering: $C = \frac{\bar{k}}{n}$

# Definizione di espansione

- Un grafo G(V,E) ha una espansione $\alpha$ se $\forall\, S \subseteq V$ si ha:

  #archi uscenti da S $\geq \alpha \min(|S|, |V \setminus S|)$

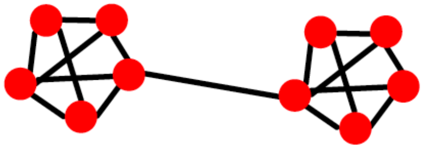
- Alternativamente:
  - $\alpha = \min_{S \subseteq V} \frac{\#\ edges\ leaving\ S}{\min(|S|,|V \setminus S|)}$



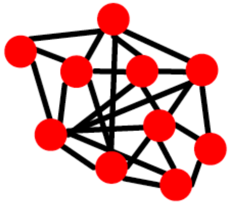
# Random k-regular graph

- Un grafo k-regolare ha tutti i vertici con grado k.

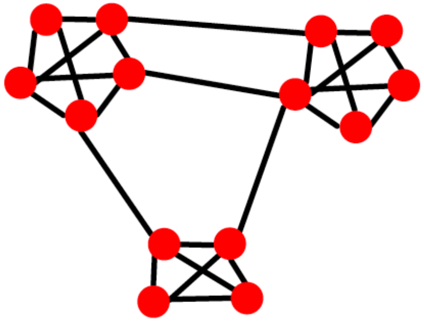
- L'espansione misura la robustezza della rete:
  - Per disconnettere $l$ nodi dobbiamo rimuovere $\geq \alpha * l$ archi
  - Bassa Espansione



  - Espansione elevata



  - Reti Sociali:
    - "comunità"



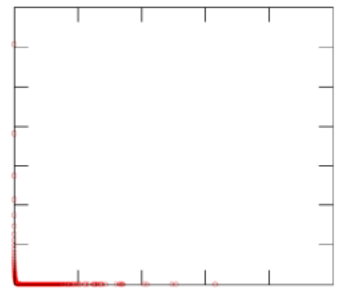
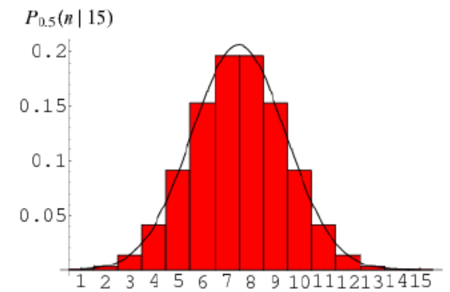
- Fatto:
  - Per un random **3-regular graph** con *n* nodi esiste una costante α (α > 0 indipendente da n) tale che l'espansione del grafo è ≥ α

- Fatto
  - In un grafo con *n* nodi con espansione α per tutte le coppie di nodi s e t esiste un path con O((log n) / α) archi che li connette.

# Proprietà di $G_{n,p}$

- Distribuzione dei degree: $P(k) = \binom{n-1}{k} p^k (1-p)^{n-1-j}$
- Lunghezza dei path: *O(log n)*
- Coefficiente di Clustering: $\mathrm{C} = \frac{\bar{k}}{n}$

# MSN vs $G_{n,p}$

| | **MSN** | $G_{np}$ |
|---|---|---|
| Distribuzione dei degree: |  |  |
| Lunghezza dei path: | **6.6** | $O(\log n)$ $\approx 8.2$ |
| Coefficiente di Clustering: | $0.11$ | $\bar{k} / n$ $\approx 8 \cdot 10^{-8}$ |

# Reti reali vs $G_{n,p}$

- Le reti reali sono simili a quelle random?
  - Lunghezza media di un cammino (si)
  - Componente connessa gigante (si)
  - Coefficiente di clustering (no)
  - Degree distribution (no)

  - La risposta è:
    - le reti reali sono diverse da $G_{n,p}$

# Configuration Model

- Obiettivo:
  - Generare una rete con una sequenza di degree fissata k1, k2,…,kn
- Configuration model
  - Generiamo per ogni nodo i rispettivi degree
  - Accoppiamo i nodi in modo random
  - Otteniamo un grafo random
- Utile come modello nullo, per confrontare una rete reale G con una rete G’ che ha la stessa sequenza di gradi.

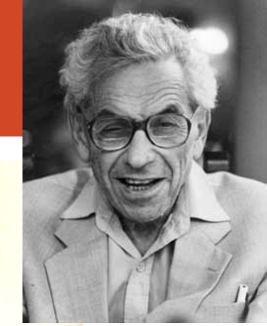
## Il modello Small-World

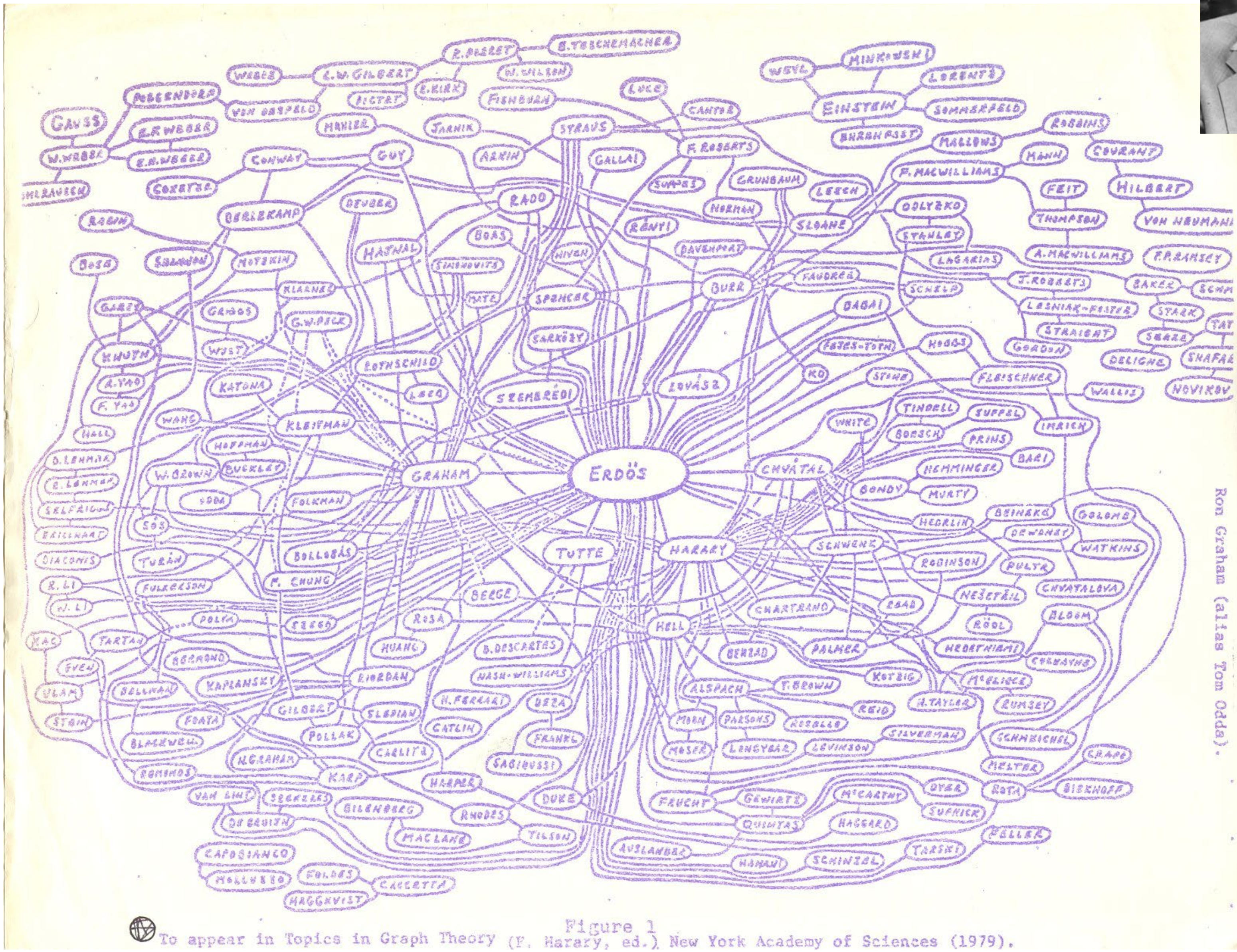
- Possiamo avere alto coefficiente di Clustering e allo stesso tempo cammini con lunghezza bassa?

# Idea dello Small-World

- Bacon Number
  - Creare la rete degli attori di Hollywood
  - Connettere due attori se hanno recitato nello stesso film
  - Bacon Number:
    - Numero di step per arrivare a Kevin Bacon
    - Dic. 2007: il più alto Bacon number calcolato è 8
    - Solo il 12% degli attori non può essere collegato a Bacon


# Erdos number

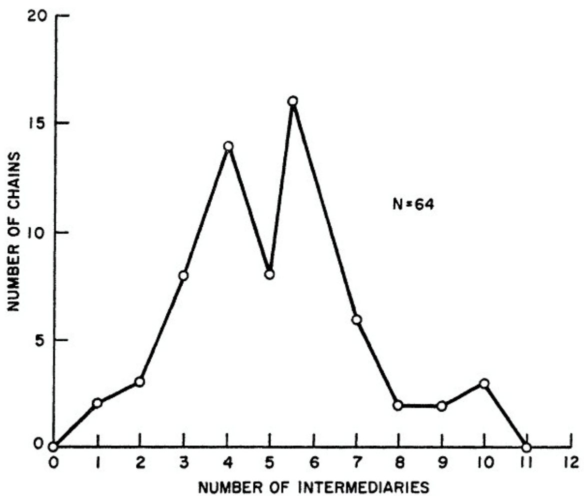





http://www.ams.org/mathscinet/collaborationDistance.html

# Esperimento Small-World [Milgram '67]

- Qual è la lunghezza tipica dello shortest path tra due persone?
  - Scelte 300 persone in alcuni stati americani.
  - Chiesto  di inviare una lettera ad un broker di Boston solo passandola agli amici
  - Quanti step sono stati effettuati?
  - In media 6.2 "6 degrees of separation"

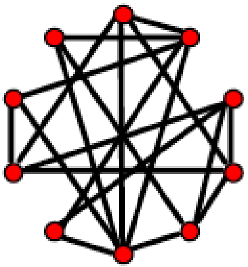



# E' sorprendente?

- Assumiamo che ogni persona conosce 100 altre persone
  - Step1: raggiugiamo 100 persone
  - Step2: raggiungiamo 100*100: 10000 persone
  - Step3: raggiungiamo 100*100*100: 1000000 persone
  - Step4: raggiungiamo 100*100*100*100: 100M persone
  - Step5: Raggiungiamo 10 Miliardi di persone

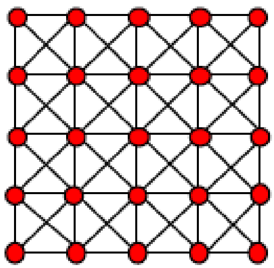
# Una controversia:

- Conseguenza dell’espansione:
  - Short path O(log n)
    - Se abbiamo un degree costante
    - Clustering coefficient basso



Low diameter
Low clustering coefficient

- Ma le reti hanno una struttura locale:
  - Chiusura Triadica
    - Amico di un mio amico è mio amico
    - Coefficiente di clustering alto ma diametro alto.



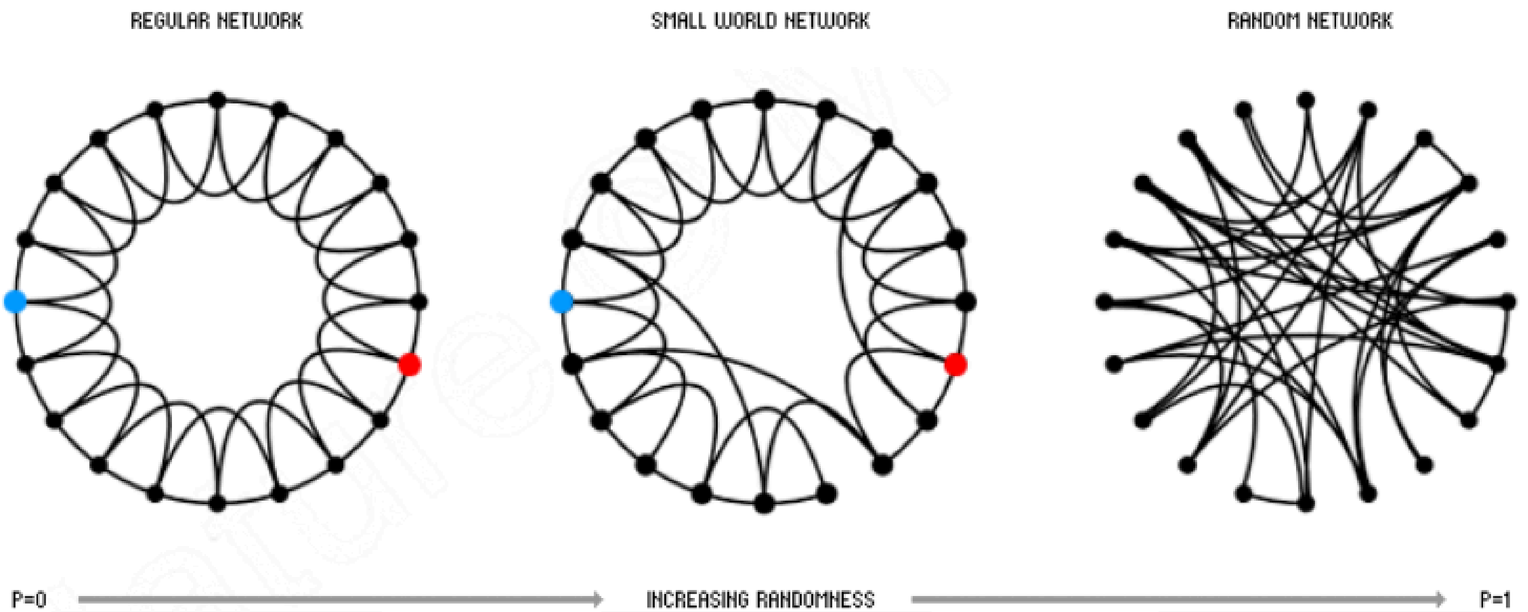
High clustering coefficient
High diameter

- Possiamo avere entrambe le cose?

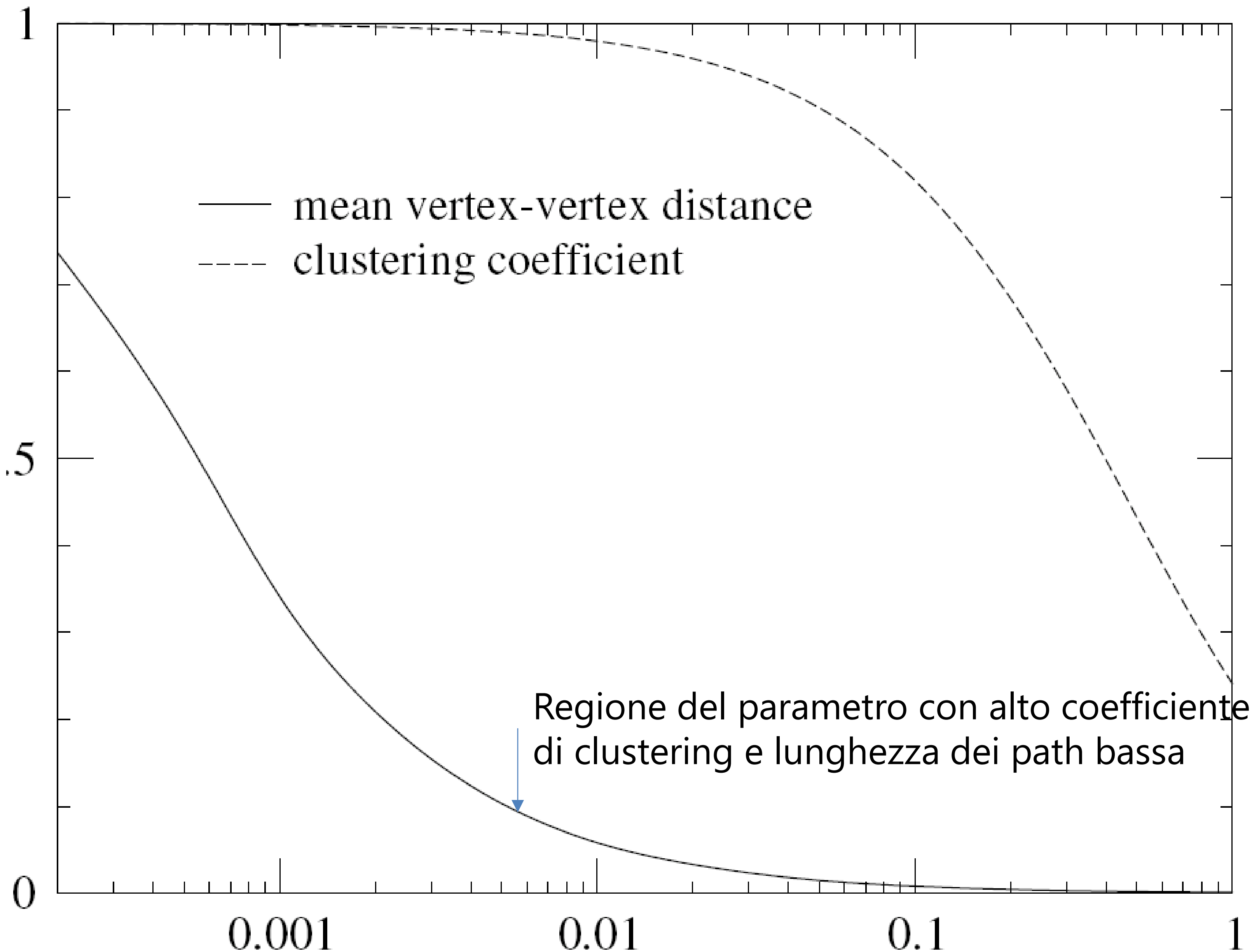
# Small-world come? [Watts-Strogatz '98]

- Due componenti del modello:
  - (1) Iniziamo con un **lattice regolare a bassa dimensione**
    - Alto coefficiente di Clustering
  - Introduciamo la **randomizzazione** (shortcuts)
  - (2) Facciamo un **rewiring** della rete
    - Aggiungiamo e rimuoviamo degli archi per creare collegamenti brevi tra parti remote del lattice
    - Per ogni arco con probabilità ***p*** muoviamo la sua destinazione ad un altro nodo random

# Il modello Small-World



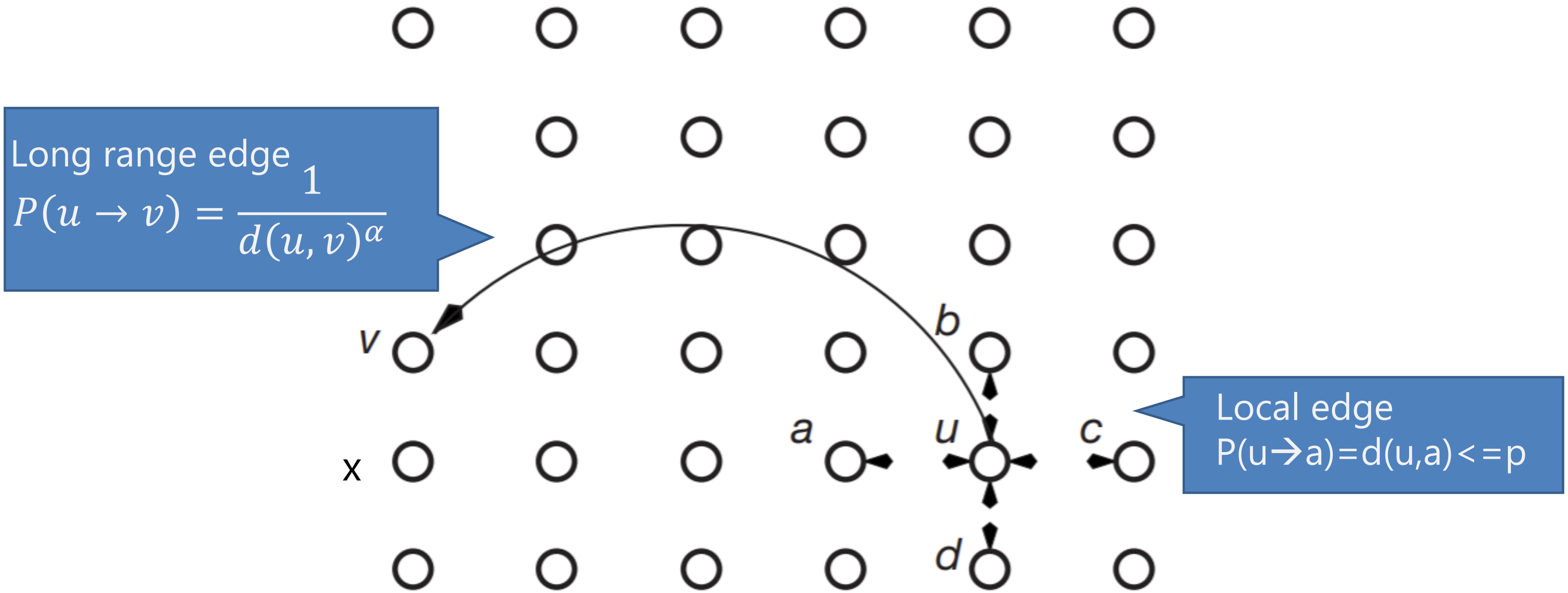

| High clustering<br>High diameter | High clustering<br>Low diameter | Low clustering<br>Low diameter |
|---|---|---|
| $h = \frac{N}{2\bar{k}}$ $\quad C = \frac{3}{4}$ | | $h = \frac{\log N}{\log \alpha}$ $\quad C = \frac{\bar{k}}{N}$ |


1
.5
0
0.001
0.01
0.1
1
mean vertex-vertex distance
clustering coefficient
Regione del parametro con alto coefficiente
di clustering e lunghezza dei path bassa

- Il modello di Watts e Strogatz
  - Coefficiente di clustering alto e small world.
  - Cattura la struttura di diverse reti reali
  - Non produce la distribuzione di degree corretta
  - Non ammette la **navigazione:** Identificare cammini brevi in una rete sociale
    - Modello di Kleinberg risolve il problema
      - https://www.nature.com/articles/35022643.pdf

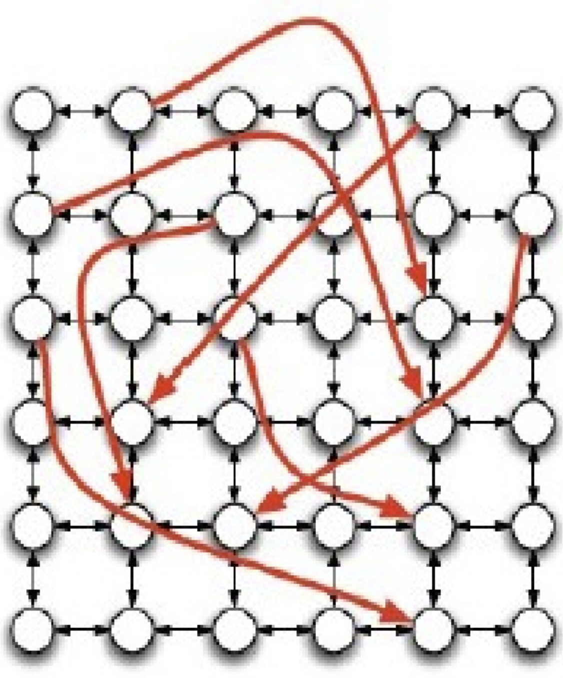
# Ricerca decentralizzata

- Il contesto
  - S conosce solo le posizioni dei suoi amici e quella del target.
  - S non conosce i link di tutti ma solo quelli suoi
  - Navigazione geografica:
    - S naviga ad un nodo geograficamente vicino a t
    - Tempo di ricerca T: numero di step per raggiunger t

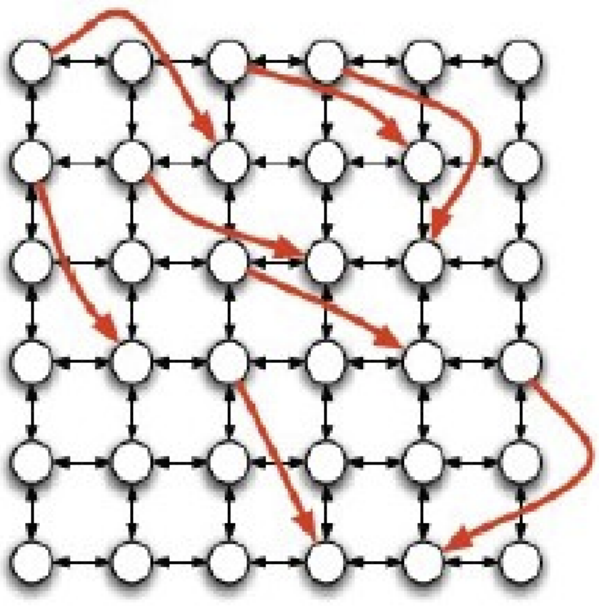



*Distanza nel Lattice*

$$\mathrm{d(u, v) = |d - x| + |x - v|}$$



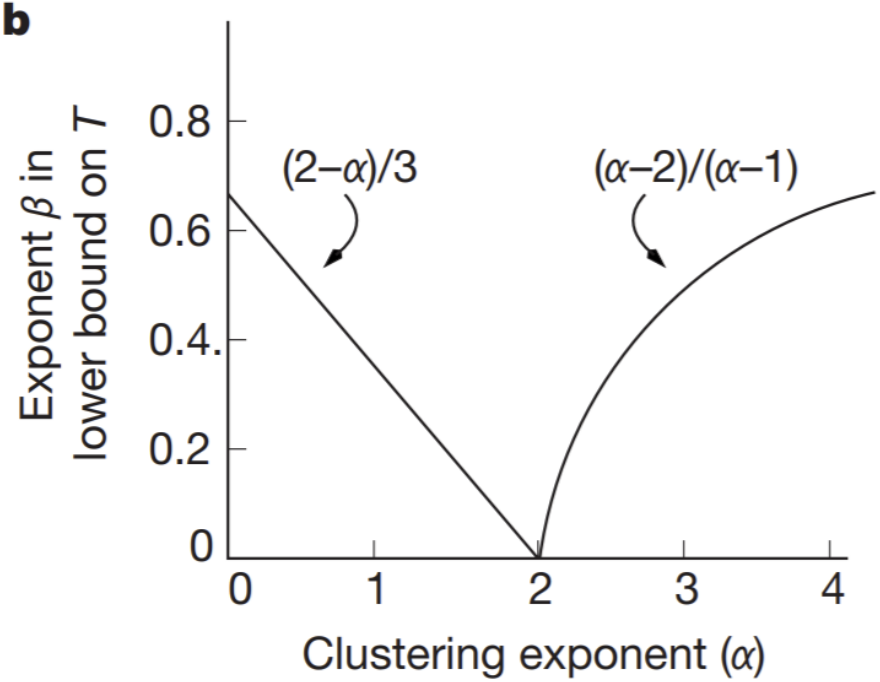
Esponente di clustering basso



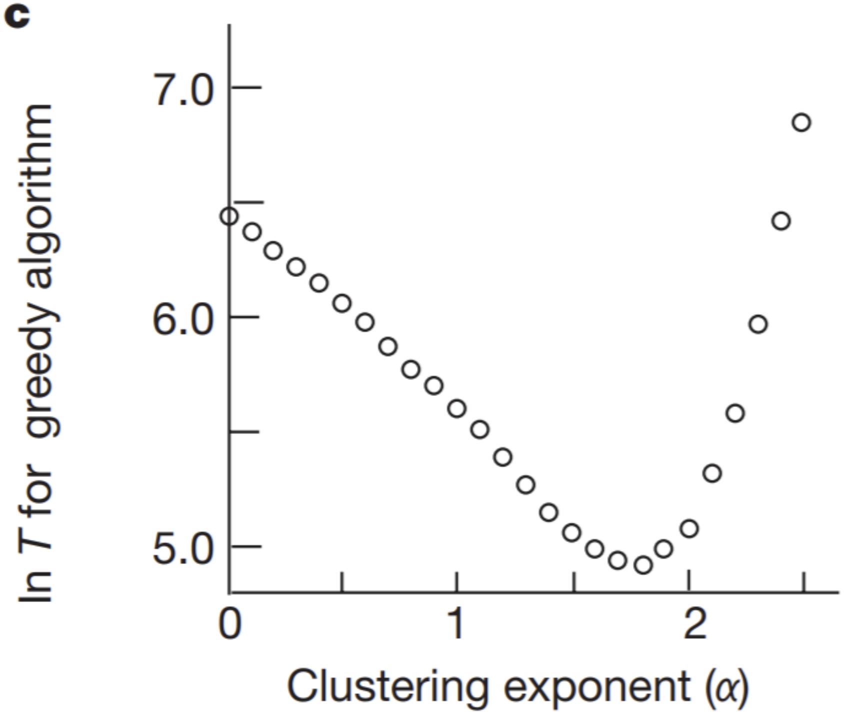
Esponente di clustering alto

- i collegamenti random devono essere generati in modo da avere lo stesso numero di collegamenti per ogni fascia di distanze
- Sia d il numero di dimensioni della griglia. Modificando il valore di α si ottengono diversi gradi di navigabilità
  - α=0: i link sono scelti in modo uniforme su tutta le rete. Modello di Watts Strogatz
  - α>d: si tende a scegliere links con nodi vicini, i long range links non sono sufficienti.
    - la ricerca decentralizzata trova rapidamente un target nelle vicinanze, ma lo raggiunge 'lentamente' se è lontano
  - α <d si tende ad avere molti link con nodi lontani
    - La ricerca decentralizzata si avvicina rapidamente al target ma poi rallenta fino a raggiungere il target
  - α=d navigabilità ottima della rete

**b**


Exponent $\beta$ in lower bound on $T$
0.8
0.6
0.4
0.2
0
(2–α)/3
(α–2)/(α–1)
0
1
2
3
4
Clustering exponent (α)

**c**


ln $T$ for greedy algorithm
7.0
6.0
5.0
0
1
2
Clustering exponent (α)

# Teorema

- Sia d(u,v) la distanza tra due nodi di una griglia bidimensionale di n nodi. Se tra due vertici u e v viene aggiunto un arco remoto con probabilità $P(u \to v) = \frac{1}{d(u,v)^{\alpha}}$ , con $\alpha = 2$ allora esiste un algoritmo di routing decentralizzato A ed una costante $\alpha_2$ indipendente da n tali che quando p =1 e q=1 (p è la distanza dei punti vicini al nodo u, q è il numero di connesisoni a lunga distanza) il numero di passi impiegati da A per trasmettere un messaggio tra una coppia qualsiasi di nodi è al più $\alpha_2(\log n)^2$

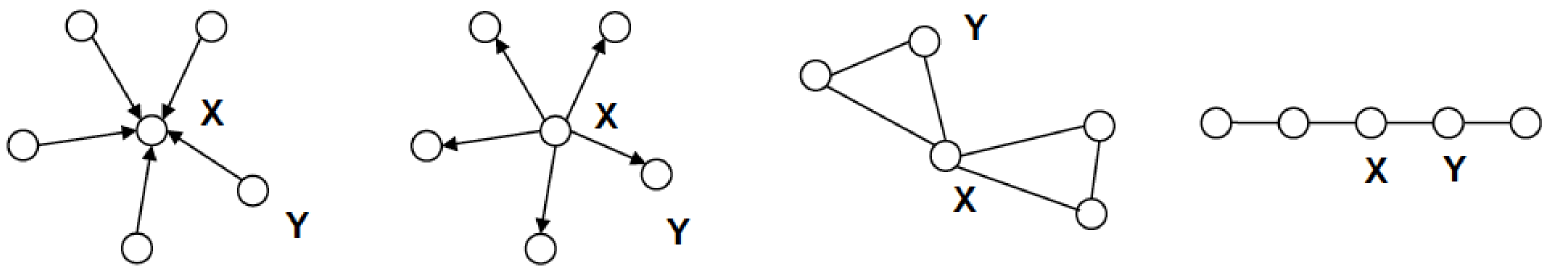
[Kleinberg's Model of Small-Worlds | An Explorer of Things (chih-ling-hsu.github.io)](chih-ling-hsu.github.io)

| | Kleinberg's Model | Watts-Strogatz Model | Erdos-Renyi Model |
|---|---|---|---|
| Navigable? | Yes<br>$T = O((\log n)^\beta)$ | No<br>$T = O(n^\alpha)$ | No<br>$T = O(n^\alpha)$ |
| Search Time | $O((\log n)^2)$ | $O(n^{\frac{2}{3}})$ | $O(n)$ |

[SmallWorldSearch (ladamic.com)]

# Misure per la centralità

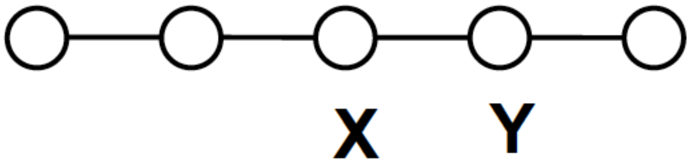
- Nelle seguenti reti X ha una centralità più elevata di Y in base ad una particolare misura:




indegree outdegree betweenness closeness

# Catturiamo la mediazione

- Intuizione:
  - Quante coppie di individui devono passare attraverso te per raggiungersi nel minimo numero di passi?



# Definizione di Betweenness

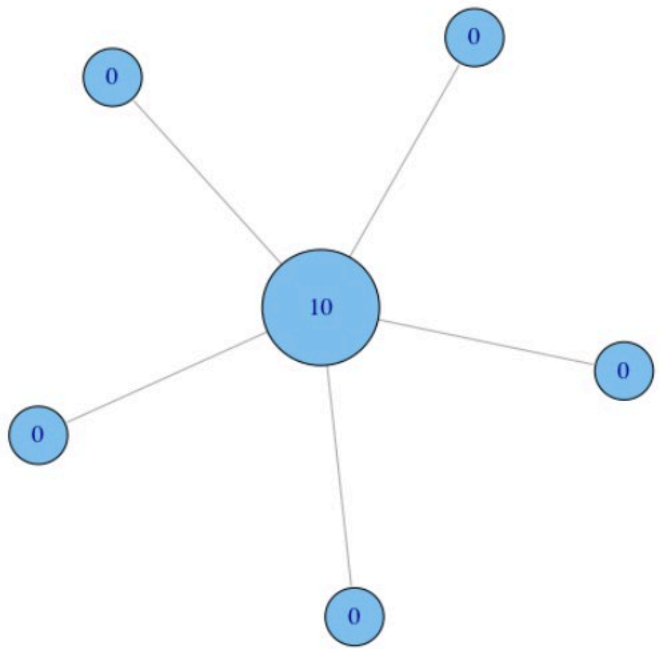
$$C_B(i) = \sum_{j<k} g_{jk}(i)/g_{jk}$$

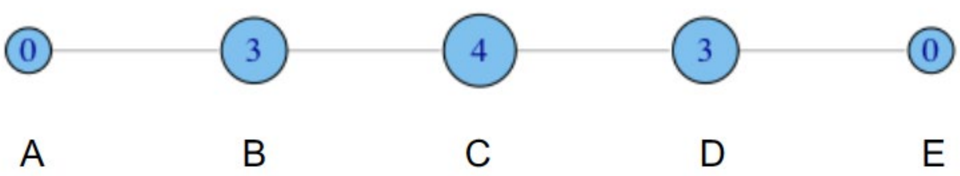
- Dove:
  - $g_{jk}$ = il numero di shortest path che connettono j e k
  - $g_{jk}(i)$ = quanti di questi contengono il nodo i

- Usualmente normalizzata come

$$C'_B(i) = C_B(i)/[\frac{(n-1)(n-2)}{2}]$$

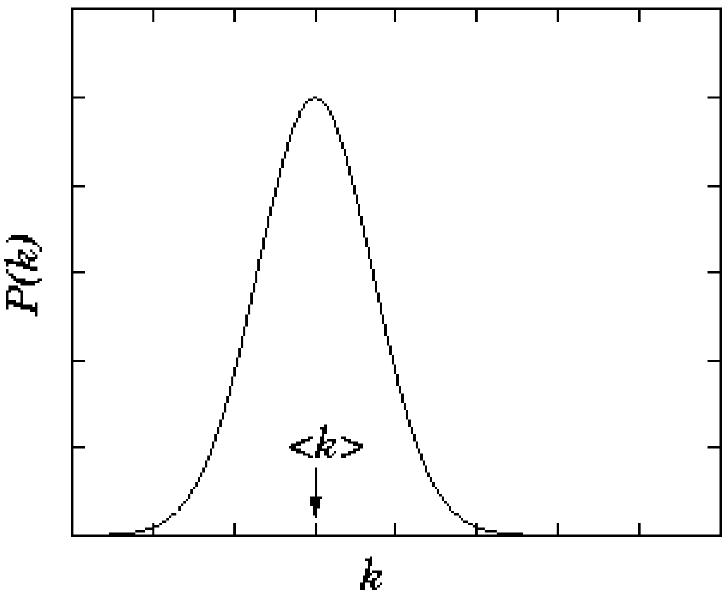
# Esempi



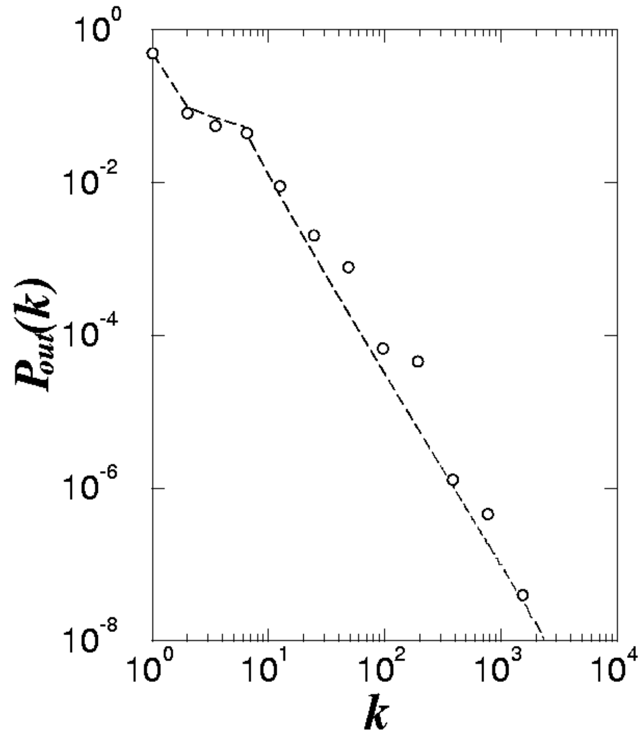



# Processo di formazione delle reti

- Modello Small-Word
  - Diametro
  - Coefficiente di Clustering
- Preferential Attachment
  - Distribuzione del degree dei nodi
    - Quale frazione di nodi ha degree k (come funzione di k)?
    - Predizione attraverso un modello semplice:
      - $p(k) = funzione\ esponenziale\ di\ k$
      - Osservazione: spesso una power-law: $p(k) \propto k^{-\alpha}$



P(k)
<k>
k

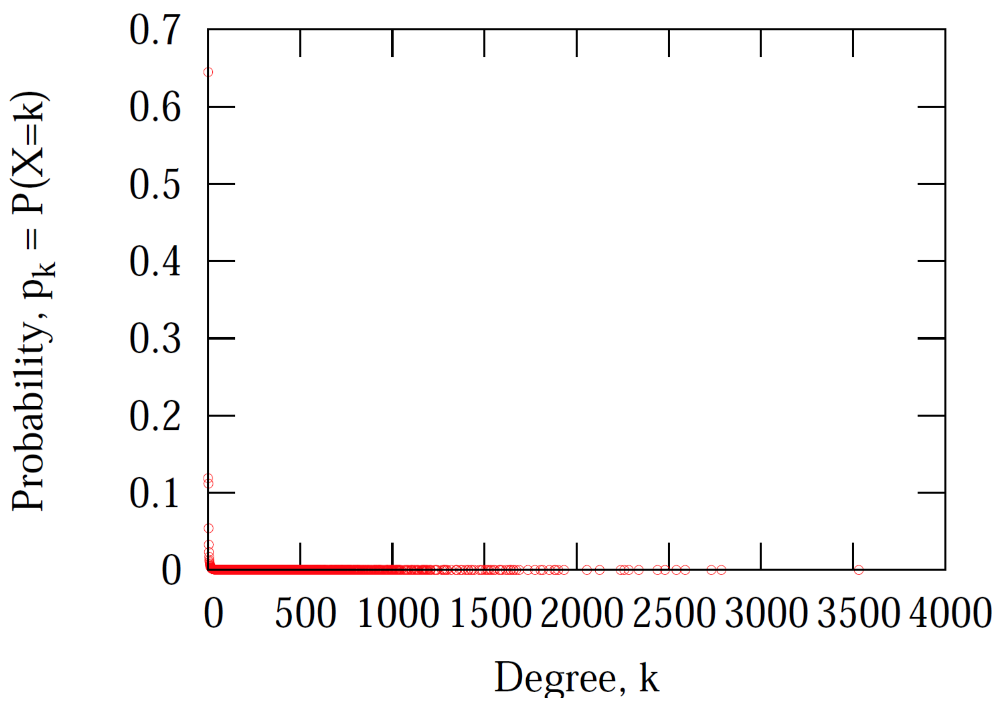
$P_{out}(k)$
$10^0$
$10^{-2}$
$10^{-4}$
$10^{-6}$
$10^{-8}$
$10^0$
$10^1$
$10^2$
$10^3$
$10^4$
k

# Esercizio

[Leskovec et al. KDD '08]

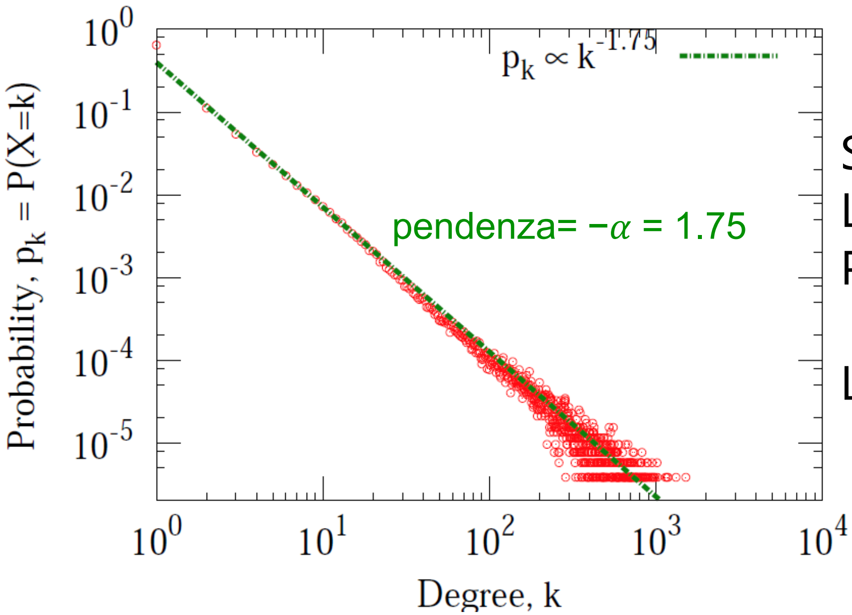
- Prendere una rete e plottare un istogramma p(k) vs k

$$p(k) = \frac{|\{u | d_u = k\}|}{N}$$




Flickr social network
n= 584,207,
m=3,555,115

- Plottare il tutto nello spazio log-log

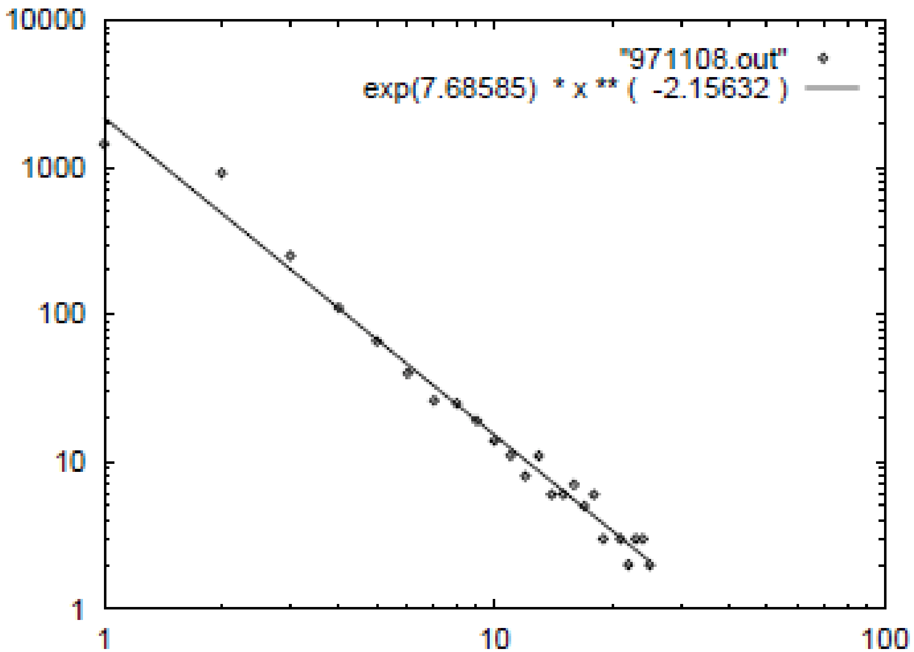



Sul piano log-log
La power-low sembra una retta con
Pendenza - $\alpha$

Log(y) = - $\alpha$ log(x)

http://konect.cc/

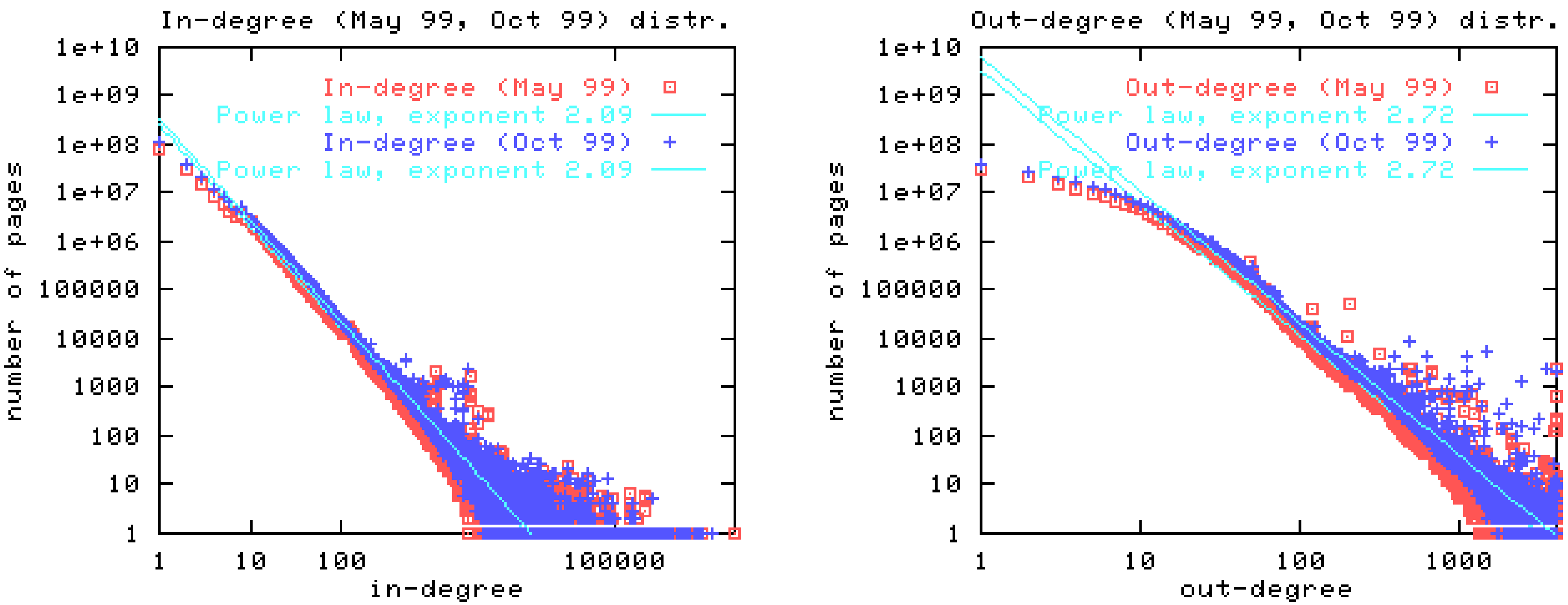
- Internet Autonomous Systems [Faloutsos, Faloutsos and Faloutsos, 1999]



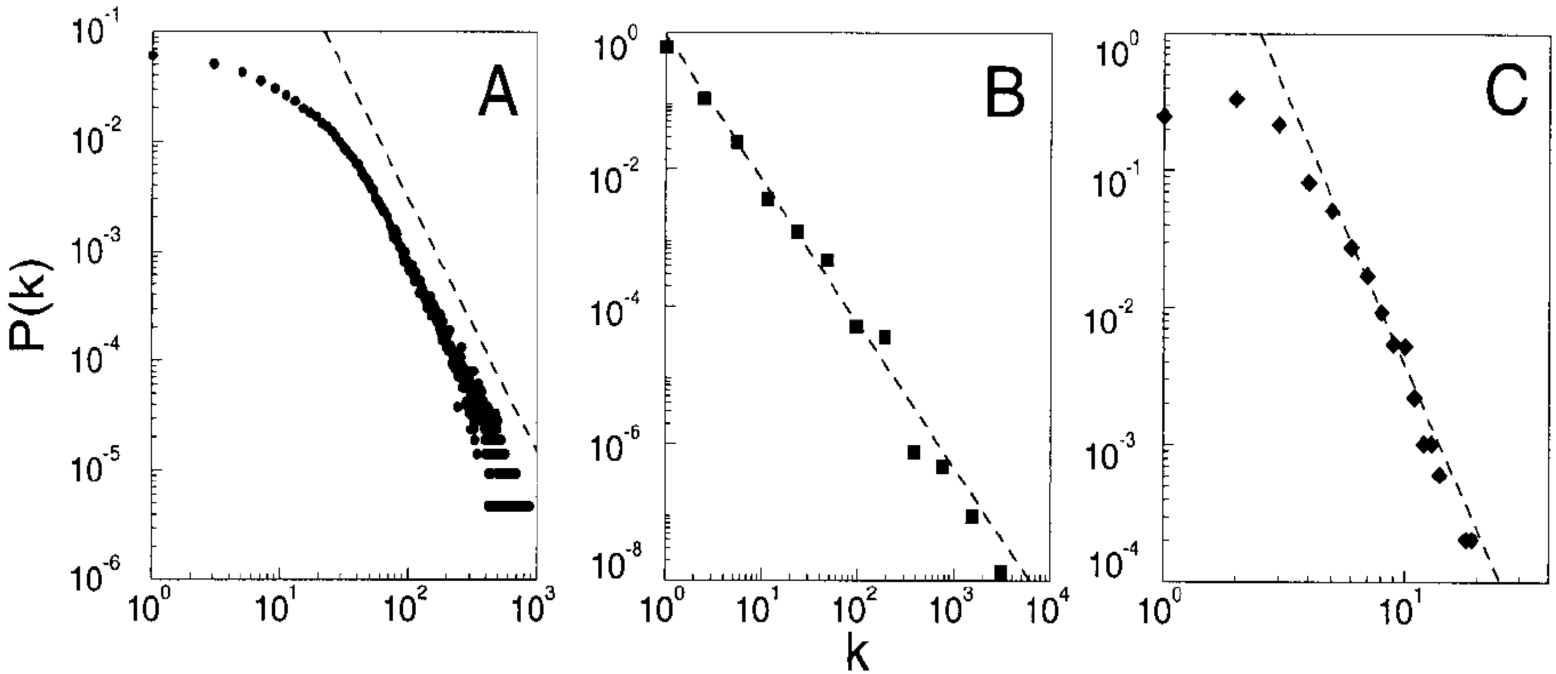

(a) Int-11-97

Internet domain topology

# Il World Wide Web [Broder et al., 2000]



# [Barabasi-Albert, 1999]

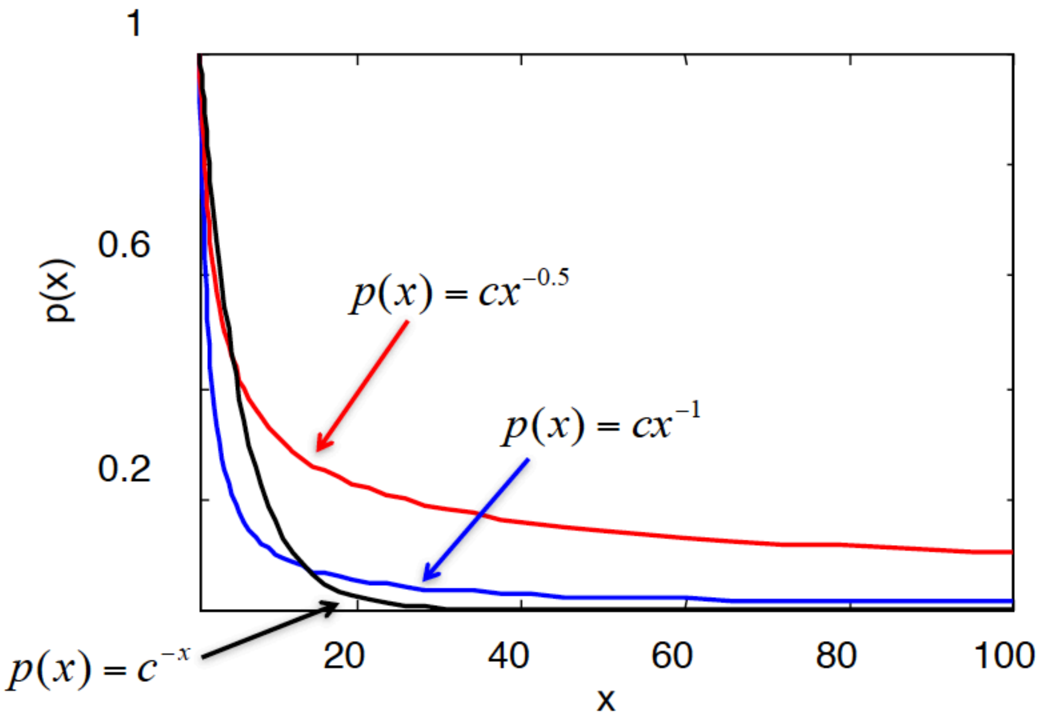



Acctor collaborations

Web graph

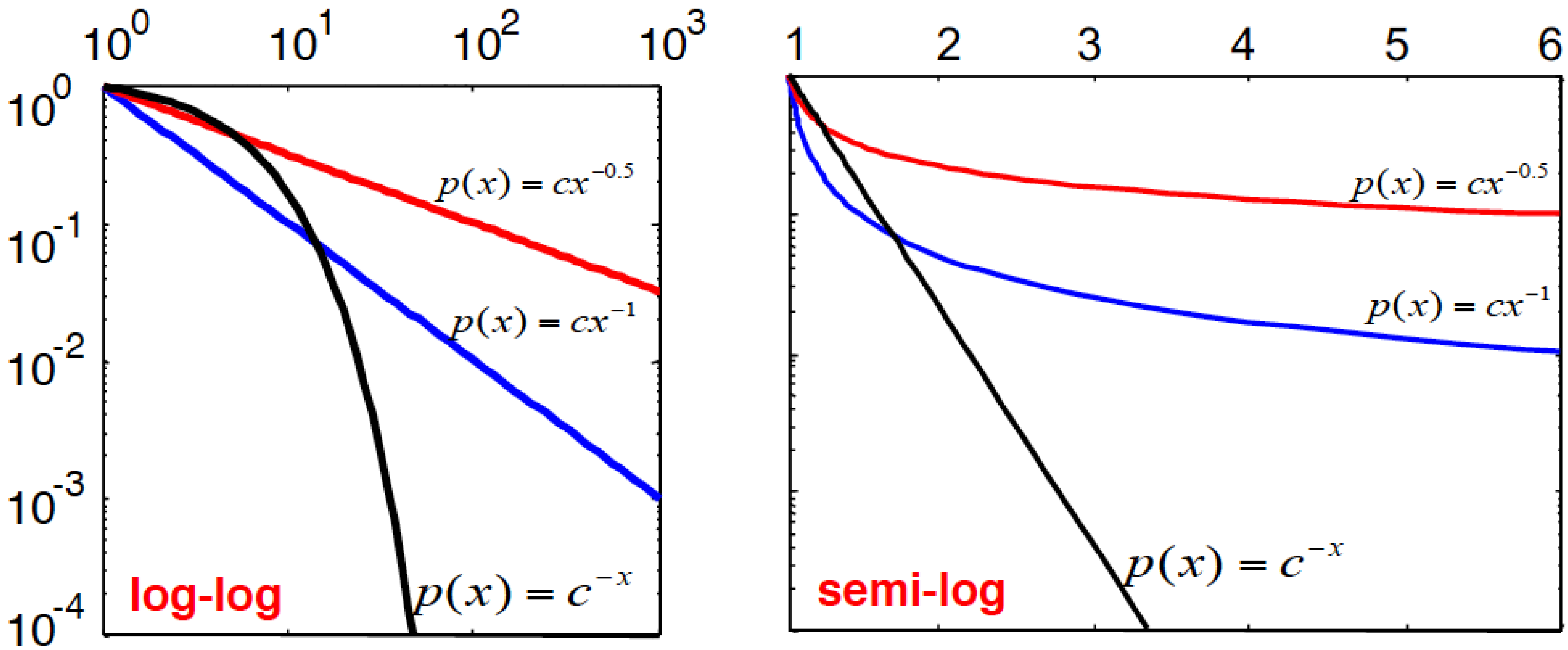
Power-grid

# Curva Esponenziale vs Power-Law




Sopra un certo valore x la legge di potenza è sempre più alta dell'esponenziale

# [Clauset-Shalizi-Newman 2007]



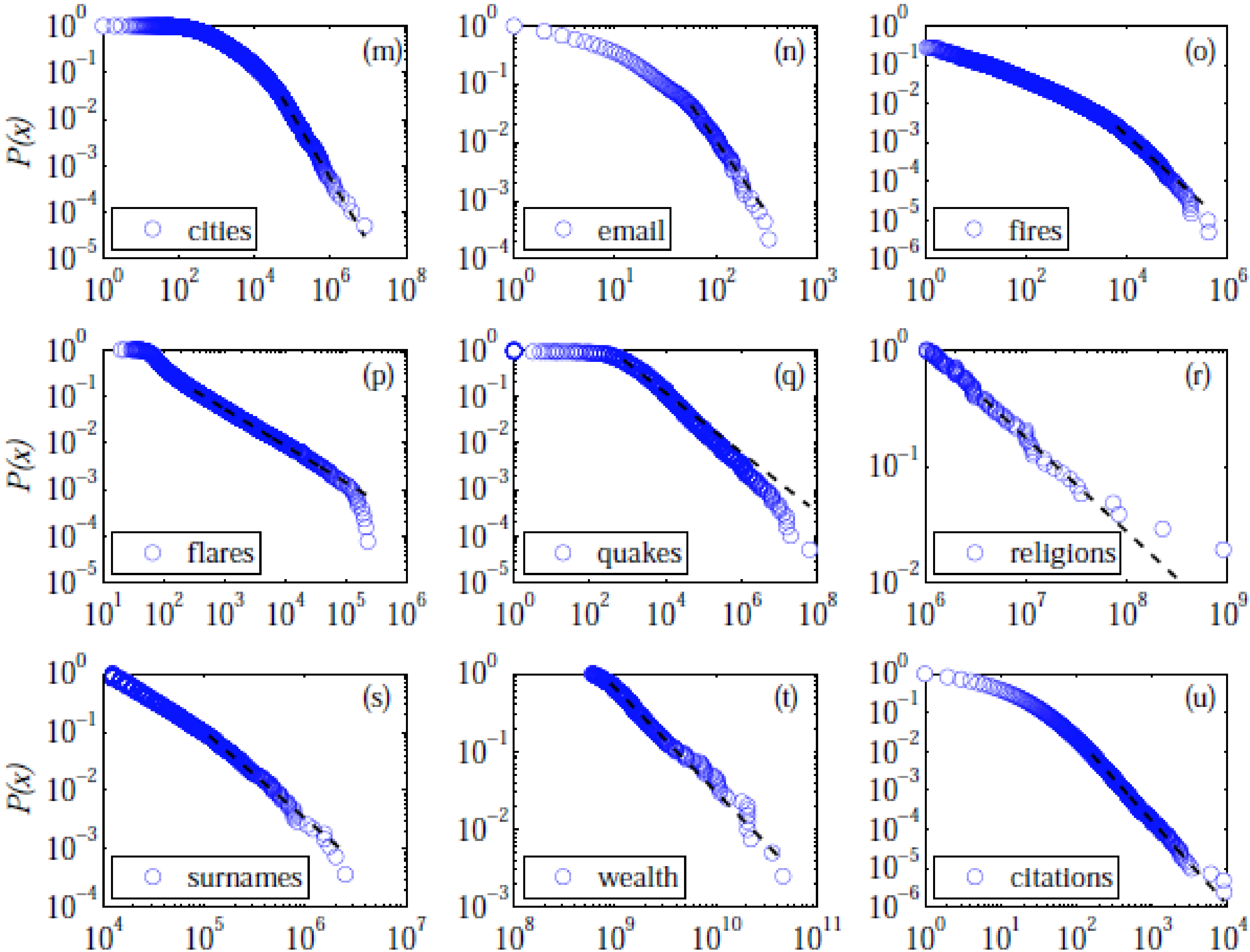

log-log vs semi-log (log-lin)

# Ricapitolando

- Esponente Power-law tipicamente $2 < \alpha < 3$
- Web graph:
  - $\alpha_{in} = 2.1$, $\alpha_{out} = 2.4$ [Broder et al. 00]
- Autonomous systems:
  - $\alpha = 2.4$ [Faloutsos3, 99]
- Actor-collaborations:
  - $\alpha = 2.3$ [Barabasi-Albert 00]
- Citations to papers:
  - $\alpha \approx 3$ [Redner 98]
- Online social networks:
  - $\alpha \approx 2$ [Leskovec et al. 07]

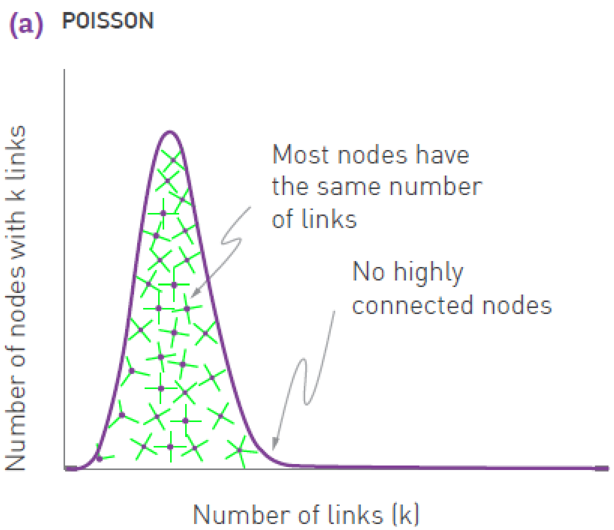
# Network scale-free

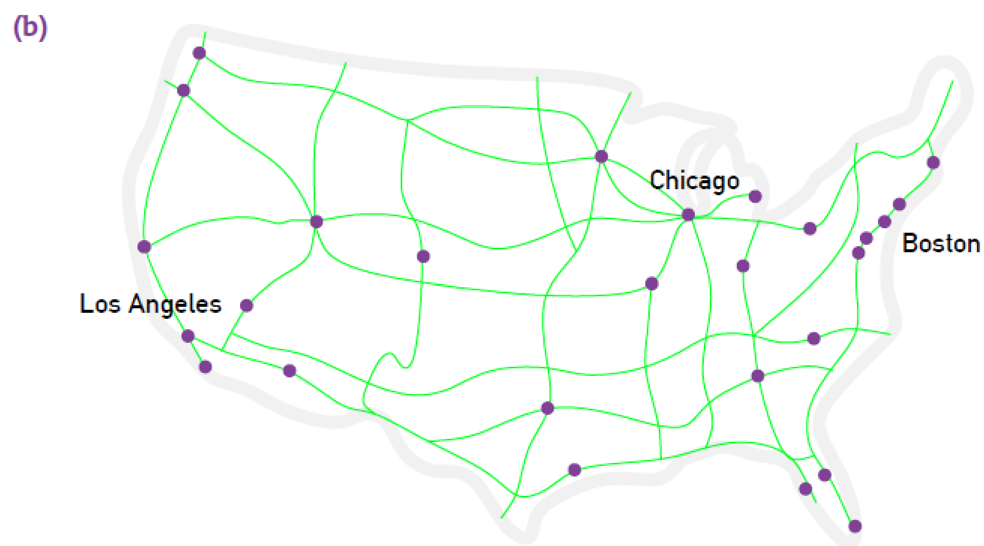
- Definizione:
  - Le network con una coda power-law nella distribuzione dei degree sono chiamate “Network Scale-Free”
- Da dove deriva il nome:
  - Invarianza di scala: non c'è una scala caratteristica
  - Funzione Scale free $f(\alpha x) = \alpha^{\lambda} f(x)$
  - Funzione Power-law: $f(\alpha x) = \alpha^{\lambda} x^{\lambda} = \alpha^{\lambda} f(x)$


(m)
cities
(n)
email
(o)
fires
(p)
flares
(q)
quakes
(r)
religions
(s)
surnames
(t)
wealth
(u)
citations
P(x)

# Robustezza delle reti scale-free - esempio

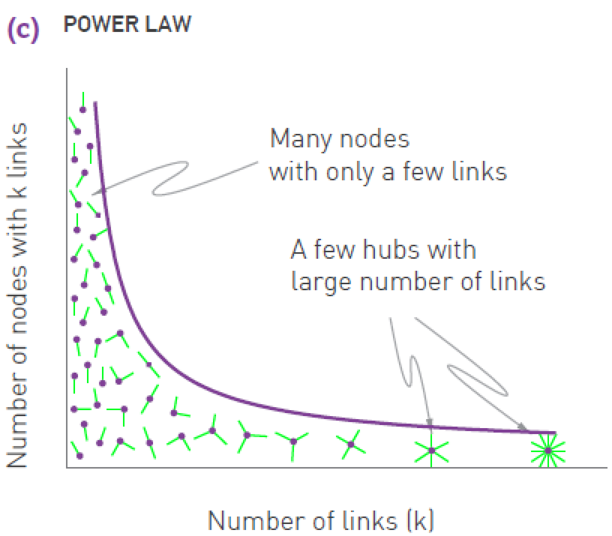
- Rete autostradale americana: simile ad una rete random

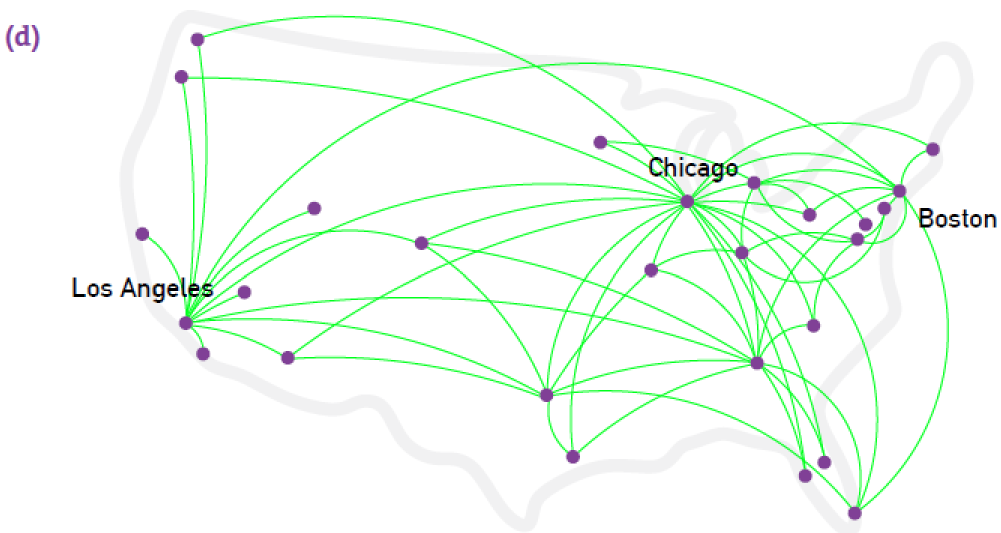


# Robustezza delle reti scale-free - esempio

- Rete di traffico aereo americano: simile ad una scale-free



# Matematica delle Power-Low

- I degree hanno una coda lunga :
  - La distribuzione P(X>x) ha una coda lunga se:

$$\lim_{x\to\infty}\frac{P(X>x)}{e^{-\lambda x}}=\infty$$

Notiamo che:

La distribuzione Normale $p(x)=\frac{1}{\sqrt{2\pi\sigma}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

La distribuzione Esponenziale $p(x)=\lambda e^{-\lambda x}$

Non sono a coda lunga

# Distribuzione a coda lunga (heavy tailed)

- Vari nomi, tipi e forme
  - Long tail, heavy tail, Zipf's law, Pareto's law
  - Nelle distribuzioni Heavy tailed
    - P(x) è proporzionale a
    - power law $P(x) \propto x^{-\alpha}$
    - power law con cutoff $x^{-\alpha}e^{-\lambda x}$
    - Stretched exponential $x^{\beta-1}e^{-\lambda x^{\beta}}$
    - Log normal $\frac{1}{x}\exp[-\frac{(lnx-\mu)^2}{2\sigma^2}]$

- Come si ottiene la costante di normalizzazione?

$$\mathrm{p}(x) = Z\,x^{-\alpha} \quad Z = ?$$

p(x) è una distribuzione, quindi ne segue che l'area sotto la curva è 1

$$1 = \int_{x_m}^{\infty} p(x)dx = Z\int_{x_m}^{\infty} x^{-\alpha}dx$$

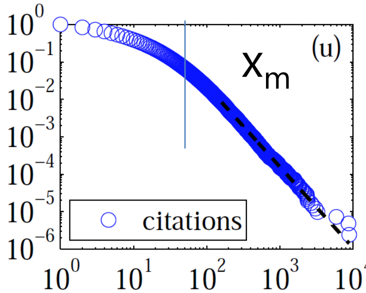
Risolvendo rispetto a Z

$$Z = (\alpha - 1)x_m^{\alpha-1}$$

Quindi

$$p(x) = \frac{\alpha - 1}{x_m^{1-\alpha}}x^{-\alpha}$$

P(x) diverge al tendere di x a 0. $x_m$ è il minimo valore della power-law $x_m$



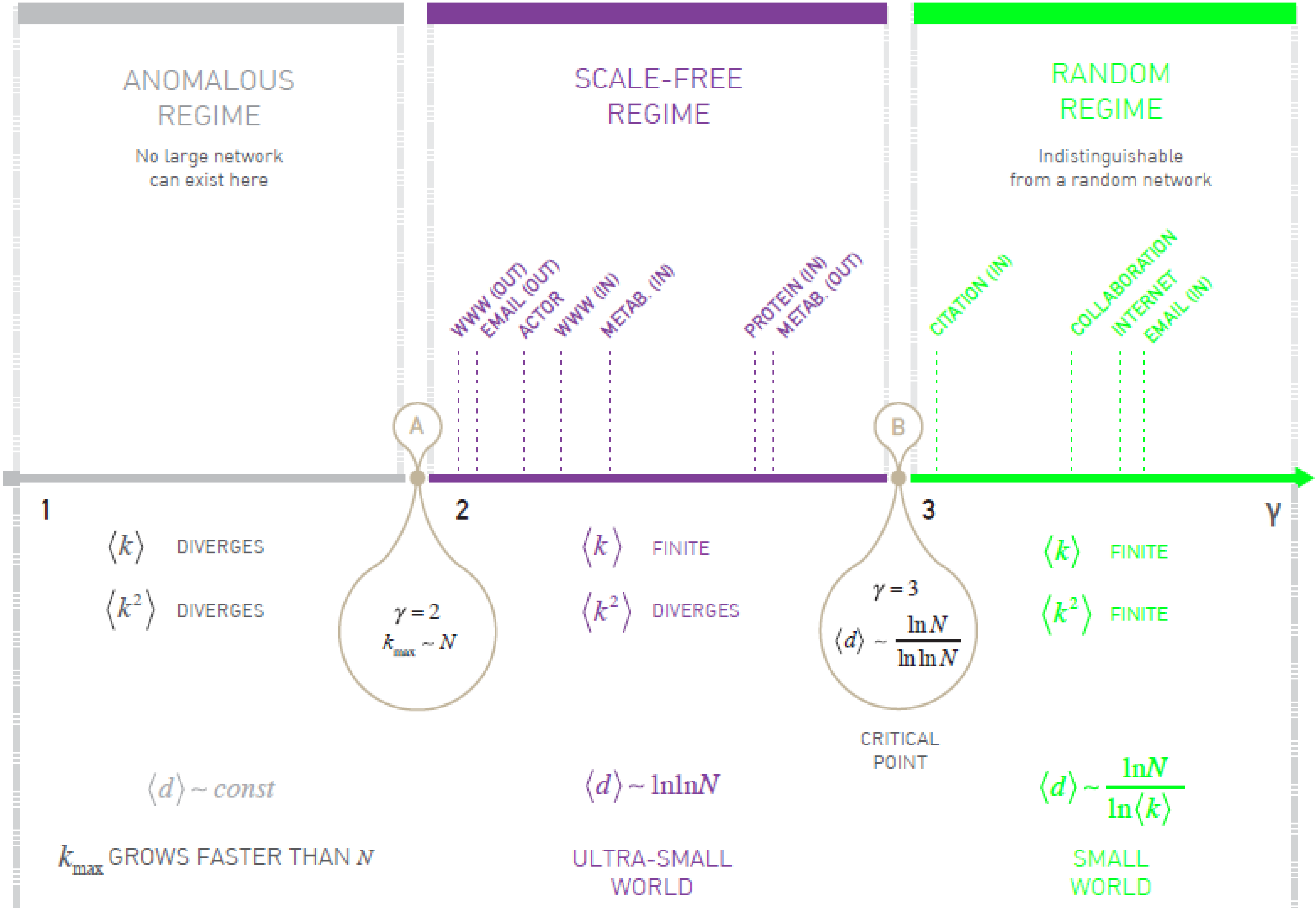
# Momenti

- $E[X] = \frac{\alpha-1}{\alpha-2} x_m$
- Se $\alpha \leq 2$: $E[X] = \infty$
- Se $\alpha \leq 3$: $Var[X] = \infty$

# Proprietà ultra small-world

- La presenza degli hub in una rete scale-free ha un'altra importante conseguenza: riduce la distanza media tra i nodi.
- Per valori dell'esponente di grado $2 < \gamma < 3$ (valori tipici di molte reti reali) si parla di ultra small-world.
- In questo scenario la distanza media tra i nodi è più piccola della distanza media osservata nella rete random (che è di per sé già una rete small-world).
- Per $\gamma > 3$, la rete è small-world ed ha proprietà che la rendono molto simile ad una rete random.
- Il caso $\gamma < 2$ è anomalo, in quanto per N che tende a infinito, sia la media che la varianza divergono. Ciò implica che non possono esistere reti grandi con $\gamma < 2$.

# Riepilogo



ANOMALOUS REGIME
No large network can exist here
SCALE-FREE REGIME
RANDOM REGIME
Indistinguishable from a random network
WWW (OUT)
EMAIL (OUT)
ACTOR
WWW (IN)
METAB. (IN)
PROTEIN (IN)
METAB. (OUT)
CITATION (IN)
COLLABORATION
INTERNET
EMAIL (IN)
A
B
1
2
3
γ
$\langle k \rangle$ DIVERGES
$\langle k^2 \rangle$ DIVERGES
$\gamma = 2$
$k_{max} \sim N$
$\langle k \rangle$ FINITE
$\langle k^2 \rangle$ DIVERGES
$\gamma = 3$
$\langle d \rangle \sim \frac{\ln N}{\ln \ln N}$
$\langle k \rangle$ FINITE
$\langle k^2 \rangle$ FINITE
CRITICAL POINT
$\langle d \rangle \sim const$
$\langle d \rangle \sim \ln \ln N$
$\langle d \rangle \sim \frac{\ln N}{\ln \langle k \rangle}$
$k_{max}$ GROWS FASTER THAN $N$
ULTRA-SMALL WORLD
SMALL WORLD

- Stimare $\alpha$ dai dati
- Usare un maximum likelihood sui dati osservati

$$L(\alpha) = \ln \left(\prod_{i}^{n} p(d_i)\right) = \sum_{i}^{n} \ln p(d_i) = \sum_{i}^{n} \left( \ln(\alpha - 1) - \ln(x_m) - \alpha \ln\left(\frac{d_i}{x_m}\right) \right)$$

Vogliamo trovare $\alpha$ che massimizza $L(\alpha)$ settiamo $\frac{dL(\alpha)}{d\alpha} = 0$

Ne segue che:

$$\frac{n}{\alpha - 1} - \sum_{i}^{n} \ln\left(\frac{d_i}{x_m}\right) = 0$$

$$\hat{\alpha} = 1 + n \left[ \sum_{i}^{n} \ln\left(\frac{d_i}{x_m}\right) \right]^{-1}$$

# Preferential attachment

- Nodi arrivano in ordine 1,2,…,n
- Allo step *j*, sia $d_i$ il degree del nodo $i<j$
- Un nuovo nodo j arriva e crea m archi uscenti
- Probabilità che j si leghi a uno dei precedenti nodi i è proporzionale al degree $d_i$ del nodo i

$$P(j \to i) = \frac{d_i}{\sum_k d_k}$$

# Rich get richer

- I nodi nuovi verosimilmente si collegano a nodi con degree alto

- Il nodo j  crea un arco in uscita ad un nodo i scelto:
  - Con probabilità p, j si collega a i scelto uniformemente random
  - Con probabilità 1-p il nodo j si collega ad un nodo i con probabilità proporzionale a $d_i$

- Il modello genera reti dove la frazione di nodi con in-degree k scala come:

$$p(d_i = k) \propto k^{-\left(1+\frac{1}{1-p}\right)}$$