

**Lezione n.18**  
**SCALE FREE NETWORKS**  
**19/5/2009**

**Materiale didattico:**  
**Materiale distribuito a Lezione**  
**Peer-to-Peer Systems**  
**and Applications**  
**Capitolo 6**

# DISTRIBUZIONE POISSONIANA

## Processo di Poisson

- **eventi indipendenti** che accadono in un intervallo di tempo fissato
- processo che avviene con probabilità costante al variare del tempo
- l'occorrenza di un evento non aumenta ne diminuisce la probabilità di occorrenza di un altro evento
- si conosce la **frequenza media  $\lambda$**  dell'evento nell' intervallo di tempo considerato
- conoscendo il valore della media, il processo di Poisson calcola la probabilità di scostarsi dal valor medio.

$$X \sim \text{Poisson}(\lambda)$$

$$P_X(k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad k = 0, 1, 2, \dots$$

$$\mathbb{E}(X) = \lambda \quad \text{Var}(X) = \lambda$$

# DISTRIBUZIONE POISSONIANA

Processi poissoniani:

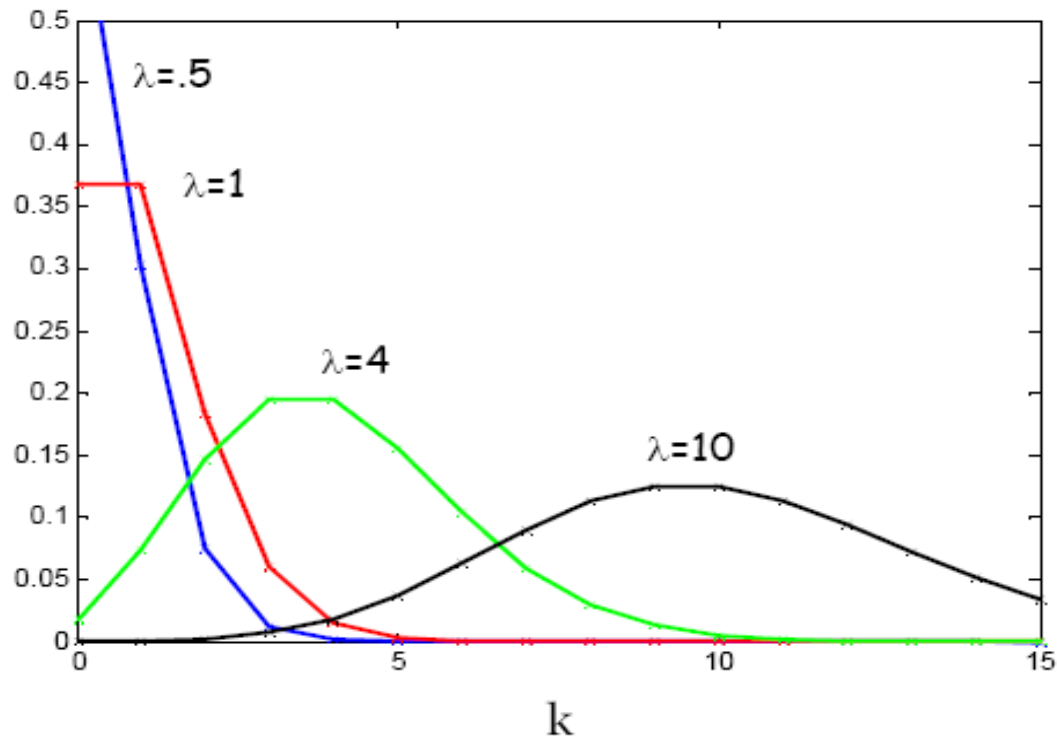
- mail ricevute in un giorno: se ricevo in media 4 mail al giorno, quale è la probabilità che ne riceva 4, 5,6,7... oppure 3,2...
- numero di auto che passano in un certo punto di una strada durante un certo intervallo di tempo
- numero di accessi ad un web server durante un certo intervallo di tempo
- **esempio storico:** numero di soldati uccisi da un calcio di cavallo nei reggimenti di cavalleria prussiani

# DISTRIBUZIONE POISSONIANA

$$X \sim \text{Poisson}(\lambda)$$

$$P_X(k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad k = 0, 1, 2, \dots$$

$$\mathbb{E}(X) = \lambda \quad \text{Var}(X) = \lambda$$



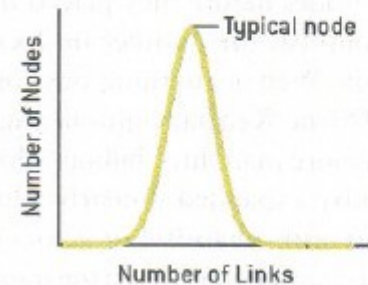
# DISTRIBUZIONE POISSONIANA

- **Random Graphs** (Erdos-Renyi): i link tra i nodi vengono stabiliti in modo casuale, utilizzando una distribuzione binomiale
- La maggior parte dei nodi **possiedono lo stesso numero di links**
- La distribuzione dei links dei nodi può essere descritta mediante una **poissoniana**
- La probabilità che un nodo sia connesso ad altri  $k$  **decresce esponenzialmente al crescere di  $k$**
- Esempio di grafo random: sistema autostradale americano

Random Network



Bell Curve Distribution of Node Linkages



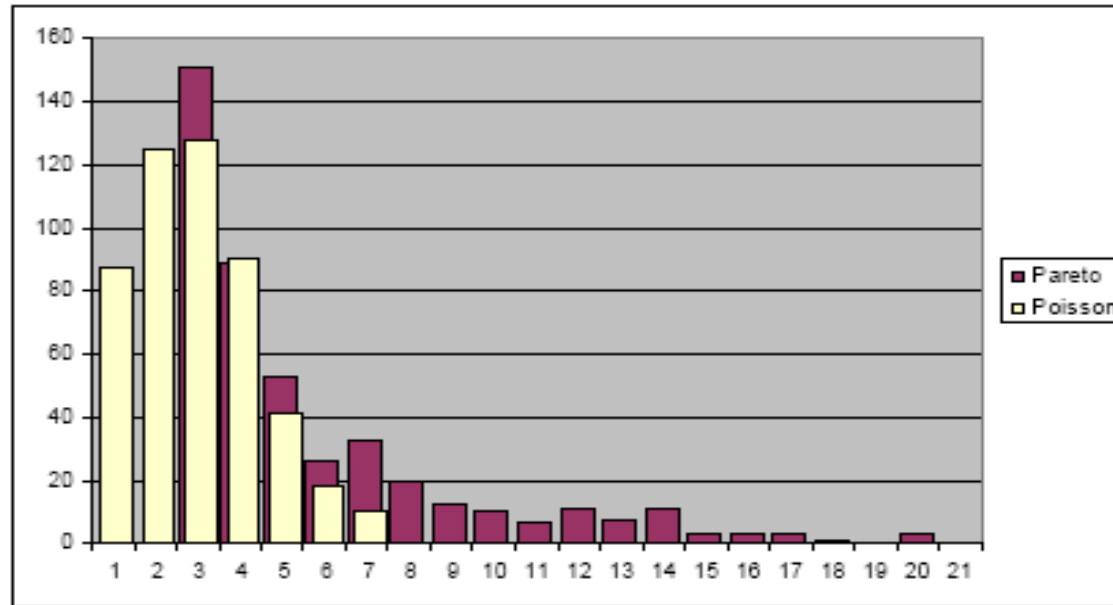
# POWER LAWS

- **Power Law** : è una funzione  $y=f(x)$  in cui il valore  $y$  della funzione è proporzionale ad una **potenza** del **valore in ingresso  $x$**

$$f(x) \sim x^{-\alpha}$$

- **Power Law Distribution**: una distribuzione di probabilità che può essere descritta con una power law
- La probabilità **diminuisce** al crescere del valore di  $x$ , **ma rimane diversa da 0** per valori molto grandi di  $x$ .
- Proprietà di una **distribuzione power law**
  - possiede una '**lunga coda**': right skew, heavy tail
  - asimmetria
  - alto rapporto tra valore **massimo** e valore **minimo**

# POWER LAW VS. POISSON DISTRIBUTION



## Poisson Distribution:

- la probabilità dei valori presenti nella coda si avvicina a 0 in maniera esponenziale la probabilità di trovare un valore molto grande diventa rapidamente 0

## Power Law Distribution: esempio Distribuzione di Pareto

- presenta una lunga coda (heavy tail, right skew)
- la probabilità di trovare un valore molto alto è bassa, ma non nulla

# POWER LAWS: HEAVY TAILED DISTRIBUTION

- **Esempio:** distribuzione delle dimensioni delle città degli Stati Uniti
  - New York, 8 milioni di persone
  - Duffield, 52 abitanti
- **Right Skew**
  - la maggior parte delle città ha un basso numero di abitanti. La distribuzione si concentra su valori bassi di  $x$ .
  - ma...esiste un **basso numero** di città con una popolazione molto alta
  - la distribuzione mostra una **'lunga coda' a destra**
- **Alto rapporto tra valore massimo e valore minimo**
  - NYC: 8 milioni, Duffield, Virginia 52, rapporto: 150000



# POWER LAW SU SCALA LOGARITMICA

- Una **power law** viene rappresentata **come una retta** se si utilizza un grafico con scala logaritmica per entrambi gli assi (**log-log graph**)
- Consideriamo la seguente power law

$$y = k * x^\alpha$$

se si passa ai logaritmi si ottiene

$$\log(y) = \alpha \log(x) + \log(k)$$

- per  $m = \alpha$  e  $q = \log(k)$  possiamo ricondurci alla retta

$$y = mx + q$$

- $m$ , il coefficiente angolare della retta, dipende da  $\alpha$

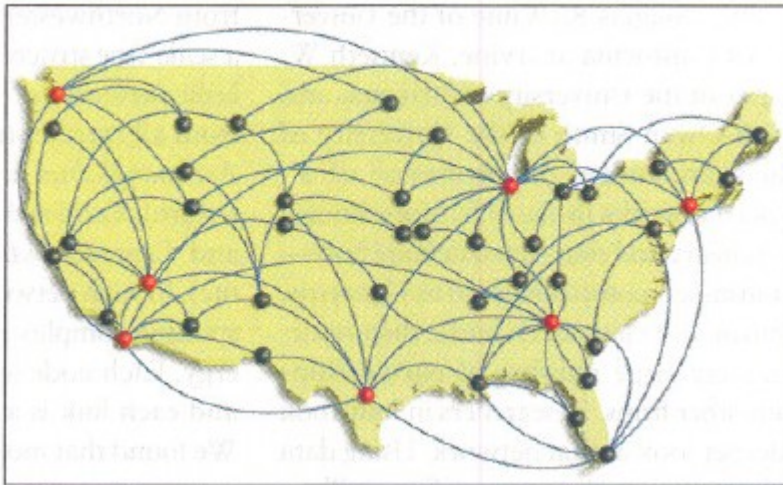
# SCALE FREE NETWORKS

- **Scale free Network:** Una rete viene indicata come Scale Free Network se la **distribuzione dei gradi dei nodi** (probabilità  $P(K_i)$  che un nodo scelto in modo random abbia grado  $K_i$ ) segue una **power law**.
- **Scale Free:** la 'forma' della curva di distribuzione di probabilità non dipende dalla 'scala' con cui vengono misurati i dati
- In queste reti:
  - la **maggior parte** dei nodi hanno un **grado basso**
  - un **numero molto limitato di nodi** della rete sono caratterizzati da un **alto grado**

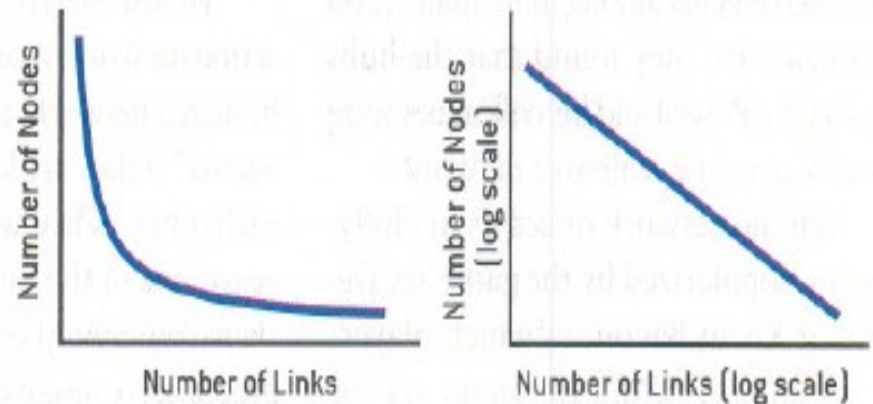
# SCALE FREE NETWORKS

- esempio di rete scale free: sistema aereo statunitense
- contiene alcuni nodi (hubs, rappresentati in rosso) e caratterizzati da un alto numero di links
- la distribuzione dei nodi, rappresentata in scala log-log è rappresentata da una retta

Scale-Free Network



Power Law Distribution of Node Linkages



# SCALE FREE NETWORKS: IL WEB

## Esperimento di Barabasi:

- obiettivo: esaminare una parte del web
- utilizza un crawler che memorizza in un database, tutte le URL individuate in un documento e poi, ricorsivamente, segue queste URL per reperire i documenti puntati
- costruisce la rete corrispondente, in cui i nodi corrispondono alle **pagine**, gli archi **ai links (URLs)** tra pagine.
- ipotesi iniziale di Barabasi: il web può essere descritto mediante **un grafo random**, perchè ogni persona:
  - sceglie in modo indipendente, secondo i propri interessi, i links da inserire nella propria pagina
  - ha, in genere, interessi diversi
  - può scegliere tra un altissimo numero di pagine da linkare

# SCALE FREE NETWORKS: IL WEB

- l'esperimento contraddice l'ipotesi iniziale di Barabasi: più dell'80% delle pagine ha meno di 4 links, ma lo 0.01 per cento delle pagine include più di 1000 links
- esistono alcuni hubs (es: Yahoo o Google), che 'dominano' la rete
- $P_{out}(k)$  e  $P_{in}(k)$ , probabilità che un documento abbia,  $k$  archi rispettivamente in uscita ed in ingresso possono essere descritte come delle power law

$$P(k) \cong k^{-\gamma}$$

con valori tipici di  $\gamma$  compresi tra 2 e 3.

Esperimento dei fratelli [Faloutsos](#):

- Esamina la rete definita dai routers e dalle connessioni esistenti tra di loro
- Anche in questo caso si rileva una distribuzione di tipo power law

# SCALE FREE NETWORKS: COSTRUZIONE

Caratteristiche delle scale free networks che favoriscono il formarsi di nodi caratterizzati da alto grado

- **Crescita dinamica della rete**

- Il numero di **nodi cresce** con il tempo. Ad esempio il web cresce continuamente
- I nodi 'più vecchi' hanno maggior opportunità di acquisire nuovi links
- Al contrario, nella costruzione di un Grafo Random si suppone che tutti i nodi siano disponibili quando comincia la generazione casuale degli archi tra i nodi della rete

- **Preferential Attachment**

- I links non vengono generati in modo random, alcuni nodi vengono scelti più di frequente
- Esempio: in internet nuovi hosts tendono a connettersi a routers che hanno già molte connessioni, perchè dispongono di una maggior banda



# SCALE FREE NETWORKS

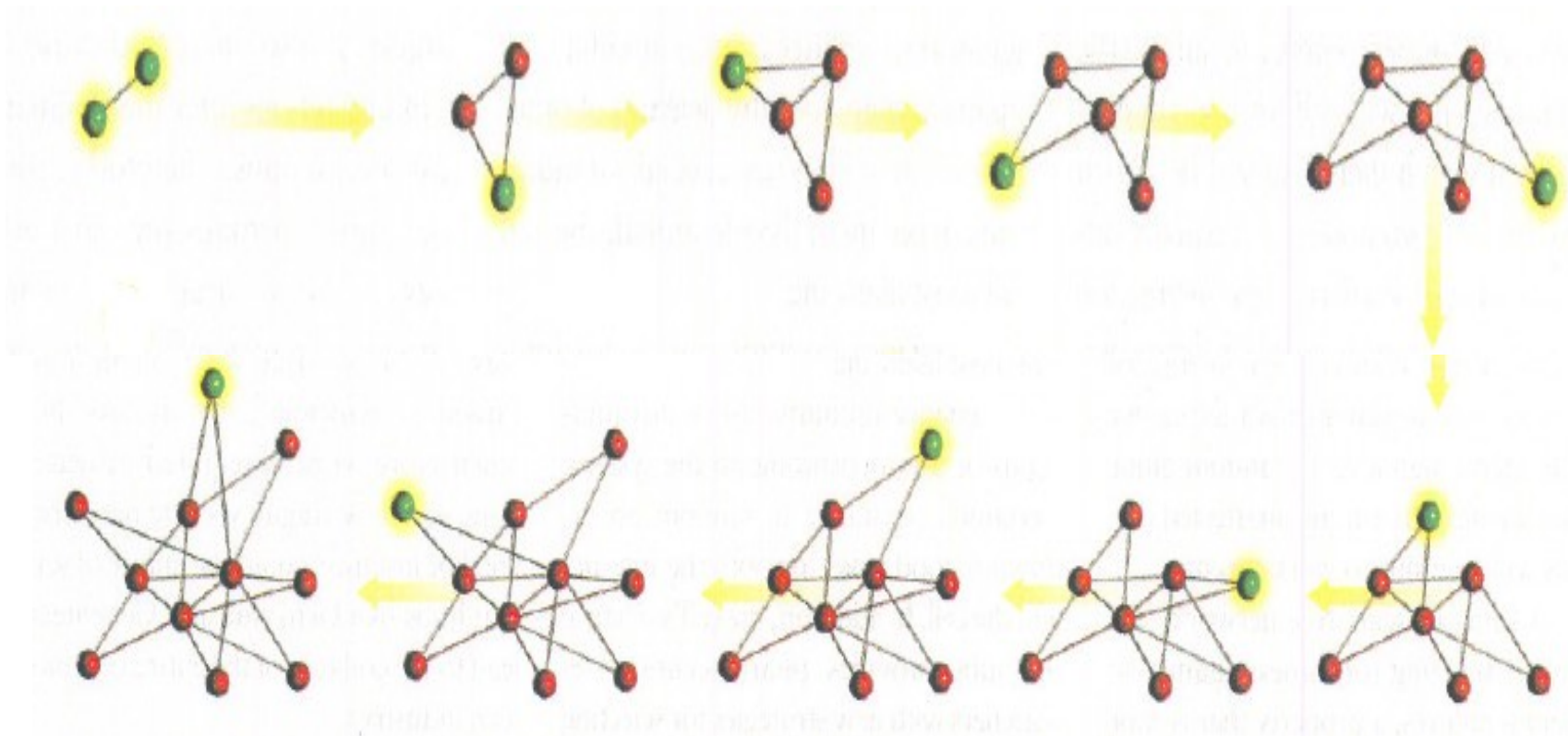
## Growth + Preferential Attachment

- Il tempo è discretizzato: la rete cresce nel tempo, ad ogni passo un nuovo nodo entra nella rete
- la generazione di un arco tra un nuovo nodo  $w$  ed uno vecchio  $v$  non avviene in modo casuale, ma segue il seguente principio
- più alto è il grado di  $v$  , più alta è la probabilità che  $w$  stabilisca un collegamento con  $v$
- Con una metafora.....**the rich get richer**

**Preferential Attachment:** La probabilità  $\Pi(v)$  che si stabilisca un collegamento tra il vertice  $v$  ed il nuovo vertice  $w$  è definita come

$$\Pi(v) = \frac{\deg(v)}{\sum_{w \in V} \deg(w)}$$

# GENERAZIONE DI UNA SCALE FREE NETWORK





# SCALE FREE NETWORKS: VANTAGGI

Vantaggio : robustezza rispetto a guasti.

- se si rimuove in **modo casuale** (distribuzione uniforme) un nodo  $v$  da una rete scale-free, con alta probabilità, tale vertice è un vertice di grado basso
- la rimozione di un vertice di grado basso non costituisce un evento "disastroso"
- le scale-free networks presentano buone caratteristiche di **tolleranza ai guasti**
- fino all'80% dei routers in Internet può essere soggetta a crash senza pregiudicare la connettività della rete
  - esiste un cammino tra due host qualsiasi anche se l'80% dei routers si rende indisponibile
- Scale free networks sono caratterizzate da una notevole robustezza che deriva dalla loro disomogeneità.

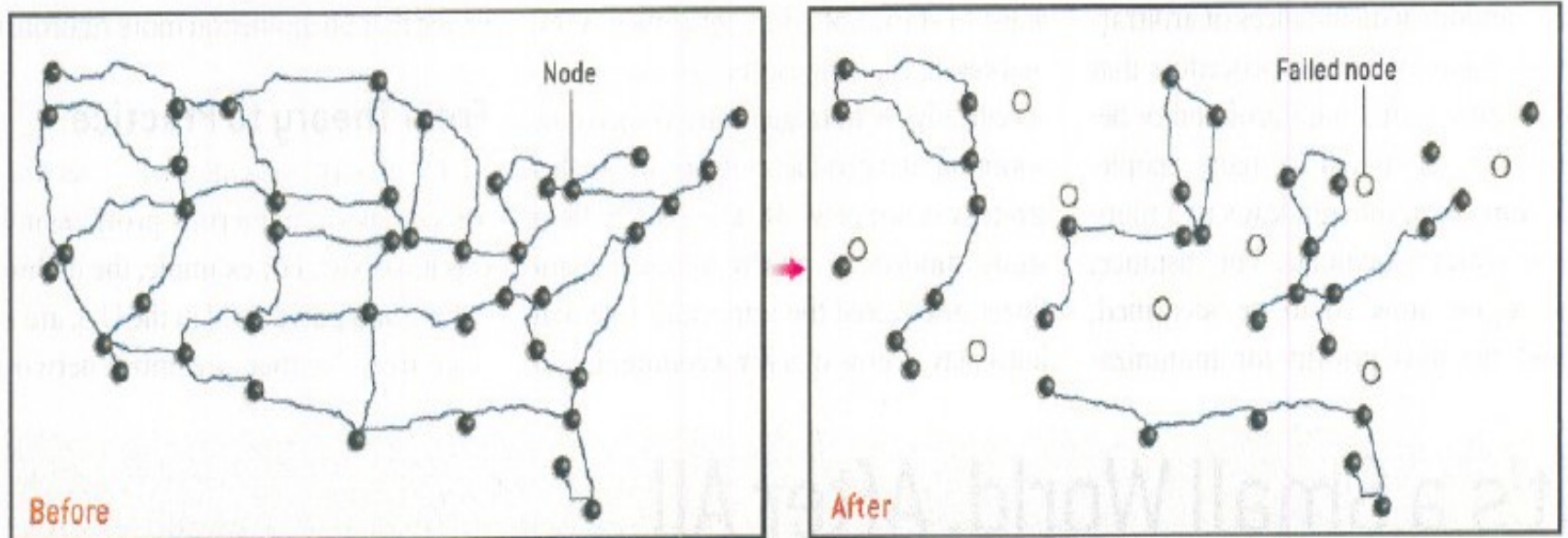
# SCALE FREE NETWORKS: SVANTAGGI

Svantaggio: vulnerabilità verso gli attacchi

- un attacco può colpire i vertici di grado più alto
- La rimozione di un numero limitato di hubs può provocare il frazionamento della rete
- **evento disastroso:** provoca il frazionamento della rete in tante componenti di piccola dimensione
- Attacco coordinato: rimuove prima l'hub di dimensione maggiore, poi quello successivo e così via...Dopo l'eliminazione di pochi hubs la rete si fraziona

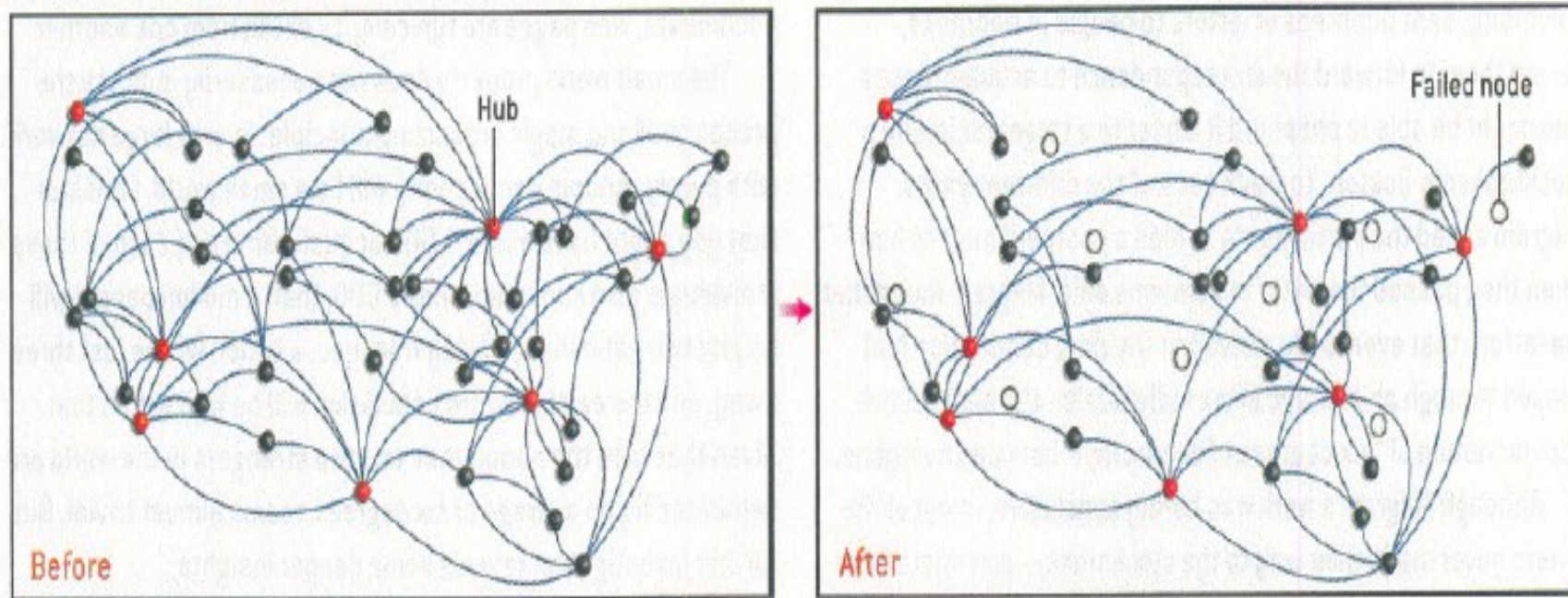
# RANDOM NETWORKS: FALLIMENTO ACCIDENTALE

Random Network, Accidental Node Failure



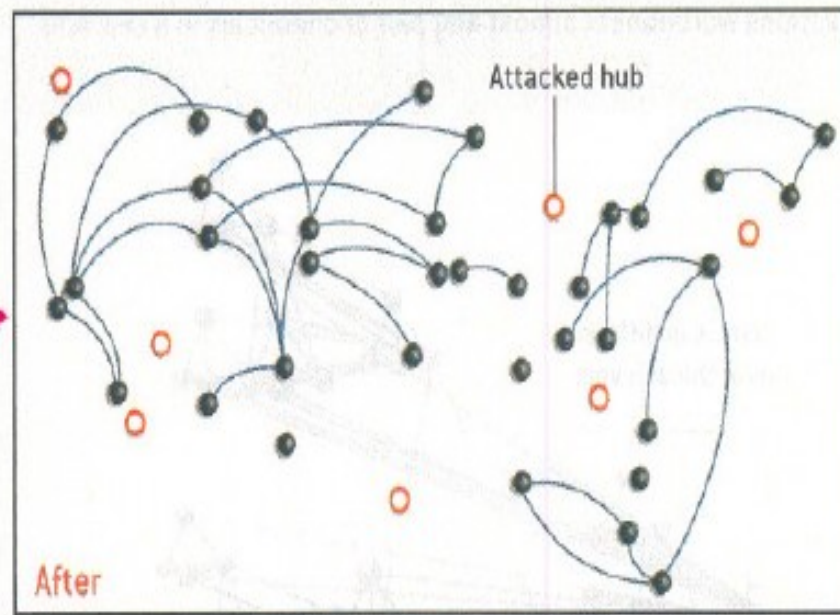
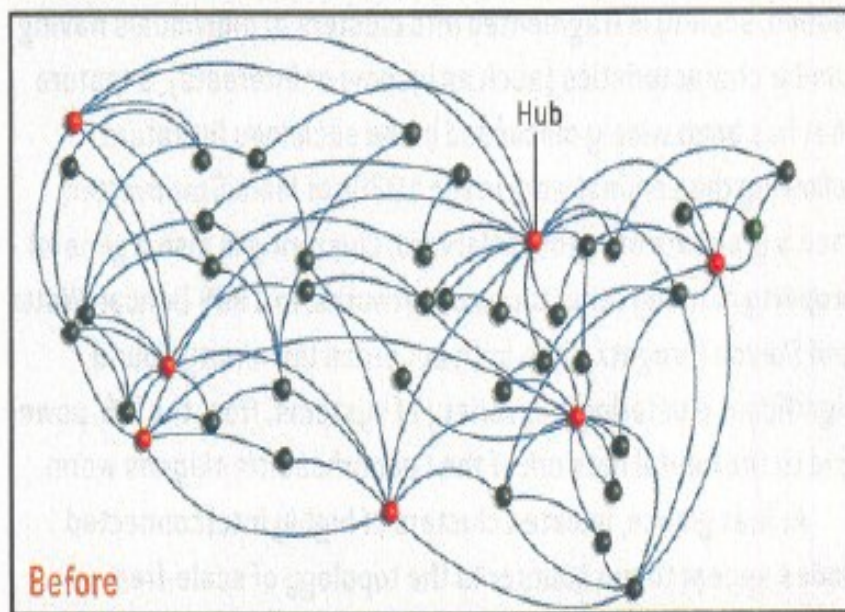
# SCALE FREE NETWORKS: FALLIMENTO ACCIDENTALE

Scale-Free Network, Accidental Node Failure



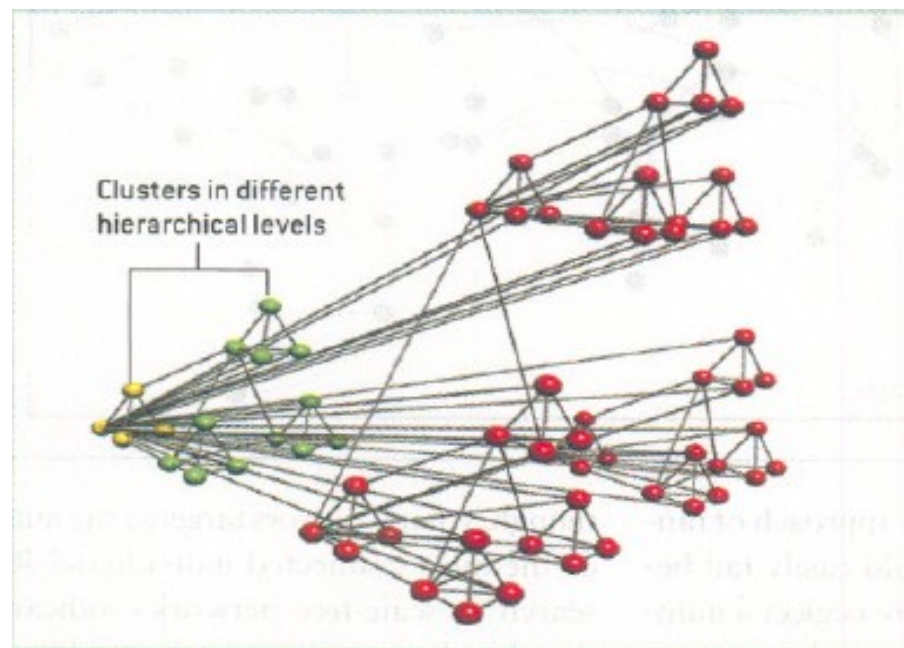
# SCALE FREE NETWORKS: ATTACCO AGLI HUBS

Scale-Free Network, Attack on Hubs



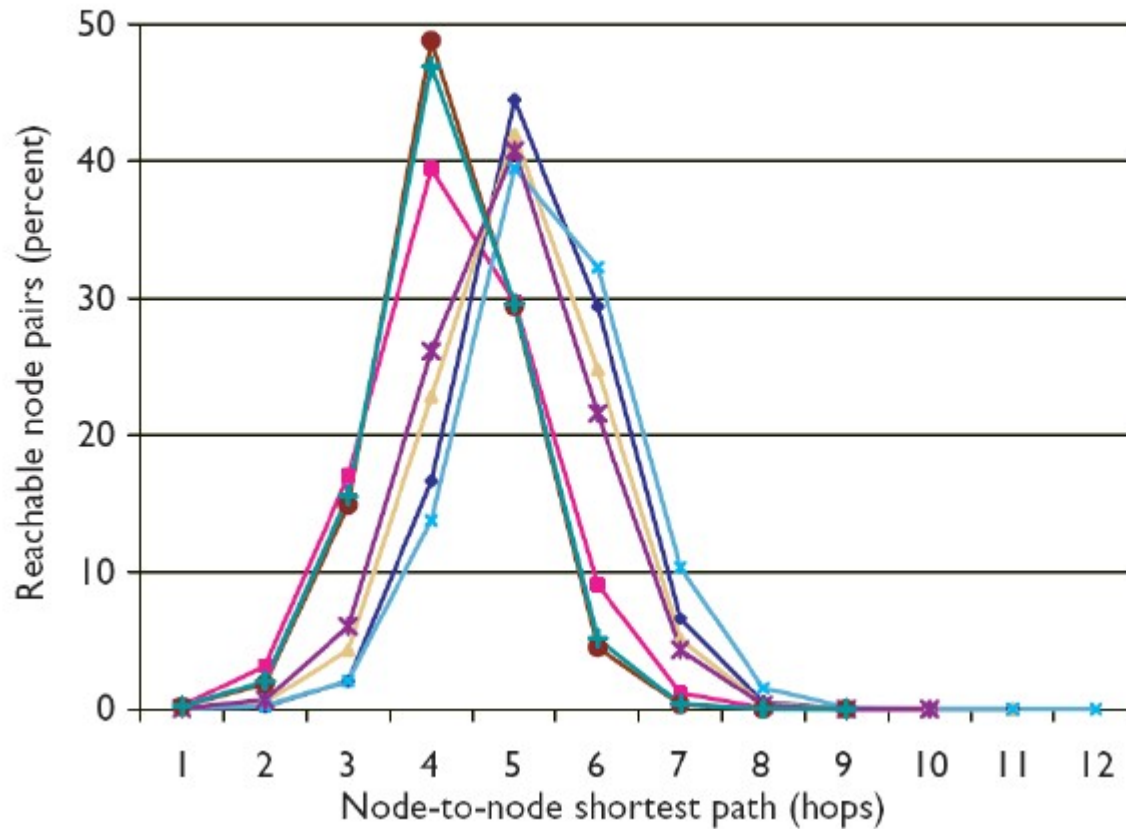
# SCALE FREE E CLUSTERING

- Relazione tra scale free networks e clusterizzazione
- Apparentemente in contrasto: la topologia delle scale-free networks in cui un numero limitato di hubs coonettono un grosso numero di nodi isolati sembra in contrasto con reti caratterizzate da un alto coefficiente di clusterizzazione
- Di recente proposto un modello che unisce le due caratteristiche

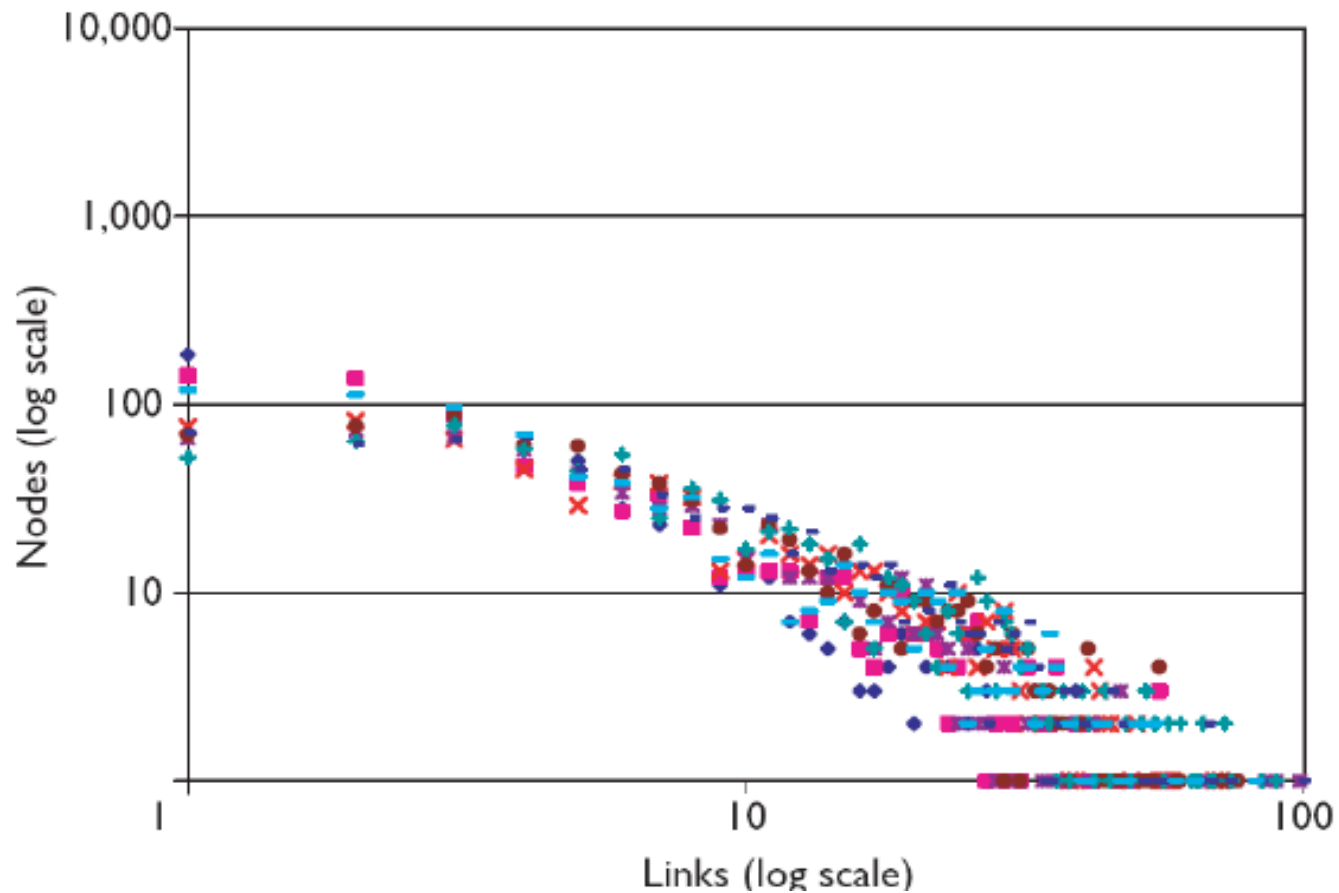


# GNUTELLA: CAMMINI (2001)

## Path lengths



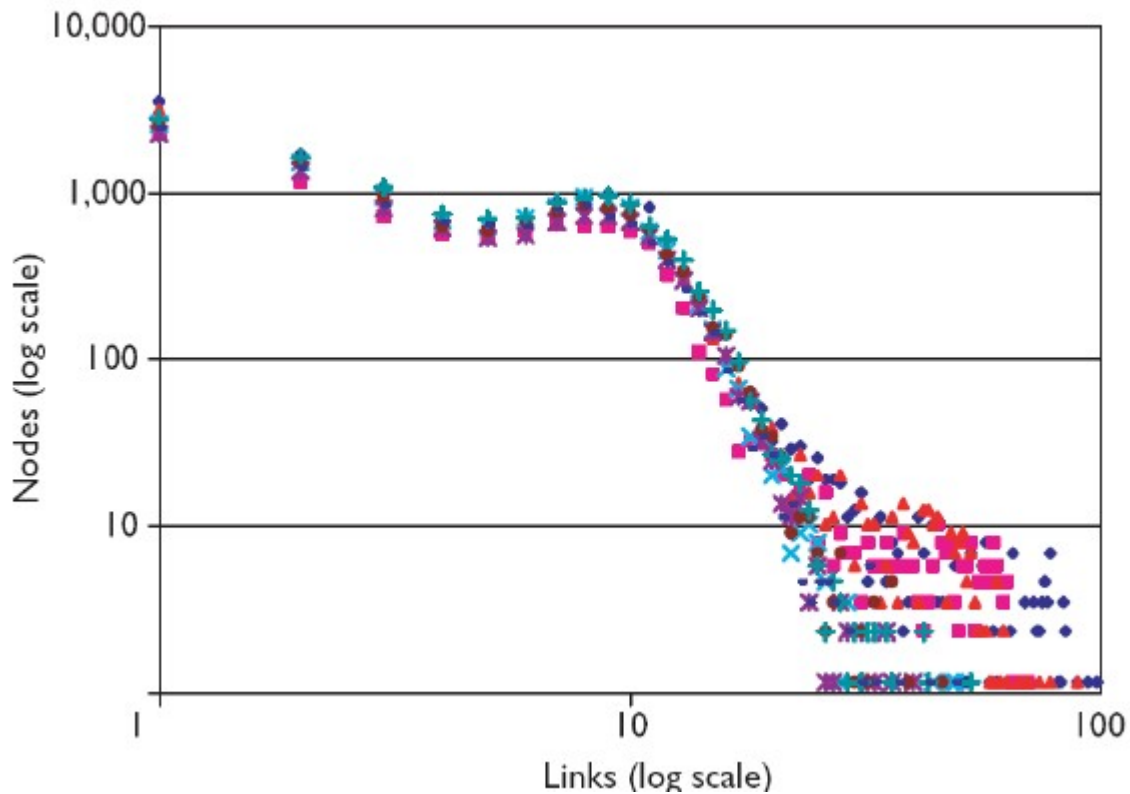
# GNUTELLA (DISTRIBUZIONE NODI)



Distribuzione rilevata nel **novembre 2000**



# GNUTELLA: DISTRIBUZIONE NODI



Distribuzione rilevata nel [maggio 2001](#)

# GNUTELLA: RANDOM WALKS

- La distribuzione dei gradi dei nodi di Gnutella segue una **power law**
- E' possibile utilizzare questa caratteristica per ottimizzare la ricerca in Gnutella?
- Strategia base: flooding limitato (TTL=7). Presenta evidenti problemi di scalabilità
- Alternativa: **Random Walk**
  - propagare la query ad un vicino scelto in **modo random** finchè non si è ottenuto un numero sufficiente di risposte
- Svantaggi del random walk
  - ricerca cieca: si propaga la query senza sapere se il vicino può risolverla
  - se si sceglie un vicino carico, la query viene memorizzata in una coda del vicino e il ritardo per ottenere la risposta può aumentare notevolmente

# GNUTELLA: POWER LAWS

- Alternativa ai random walks: sfruttare la struttura 'scale free' della rete
- Ogni nodo conosce i riferimenti ai files memorizzati nei propri vicini
- I nodi di alto grado possono memorizzare un 'indice' di una grossa porzione di dati memorizzati nell'overlay Gnutella
- Ad ogni passo
  - la query viene inoltrata ad un solo vicino
  - si sceglie il vicino  $V$  che presenta il grado maggiore
  - $V$  ha maggiore probabilità di risolvere la query perchè conosce un maggior numero di vicini
- Svantaggi di questo approccio: carico eccessivo associato ai nodi di alto grado
  - Definizione di meccanismi di controllo del flusso

# RETI P2P, SMALL WORLDS, SCALE FREE

Quali dovrebbero essere le caratteristiche di una rete peer to peer?

- se si vuole rendere la rete "navigabile" in modo decentralizzato si può definire una rete "small world" alla Kleinberg. In questo modo si può definire un **algoritmo di routing efficiente**
- se la rete P2P è definita in un contesto piccolo e sicuro (es una intranet di un'organizzazione) si può renderla una rete scale free, utilizzando un numero limitato di hubs, di alta banda
- se la rete P2P può essere soggetta ad attacchi si può definire una rete "small world", in cui la maggior parte dei vertici sono caratterizzati da un basso grado

# CONCLUSIONI

- La struttura di una rete peer-to-peer system influenza
  - la lunghezza media dei cammini,
  - la possibilità di definire algoritmi di routing greedy, decentralizzati
  - la stabilità della rete verso guasti
  - la sensibilità della rete verso gli attacchi
- Caratteristiche importanti di una rete P2P :
  - la lunghezza media dei cammini
  - il coefficiente di clusterizzazione
  - la distribuzione del grado dei nodi
  - ...
- E' importante stabilire le regole per la generazione degli archi della overlay network in modo tale che la struttura della rete definita garantisca le proprietà desiderate