Grafi aleatori e reti nel mondo reale

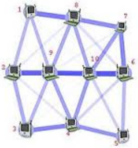
Federico Bassetti

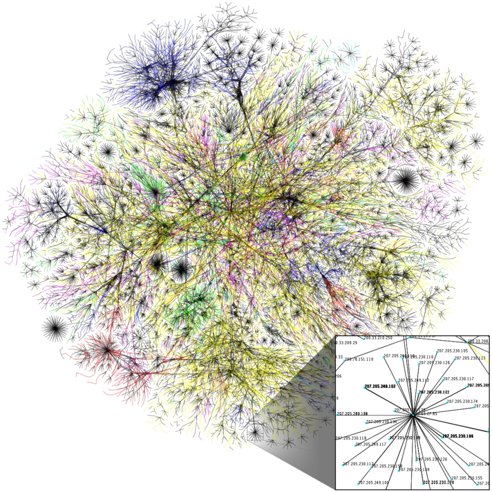
Università degli Studi di Pavia

Pavia, 15 Dicembre 2016

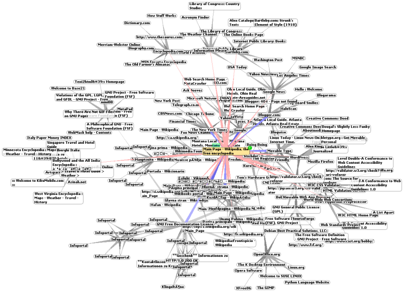
1. Reti (Networks) nel mondo reale.

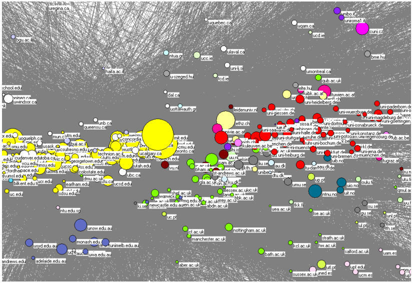
WWW = World Wide Web





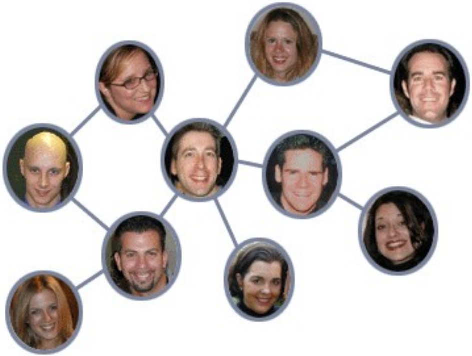
WWW = World Wide Web





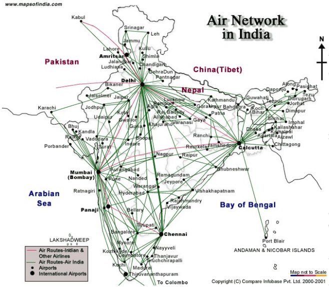
Reti di relazione

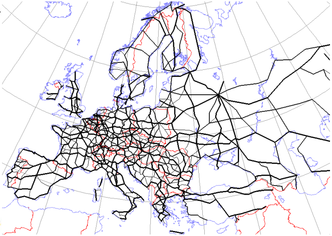
(reti di amicizia, reti sociali, reti di collaborazione scientifica,...)



Reti di comunicazione telematica o reale

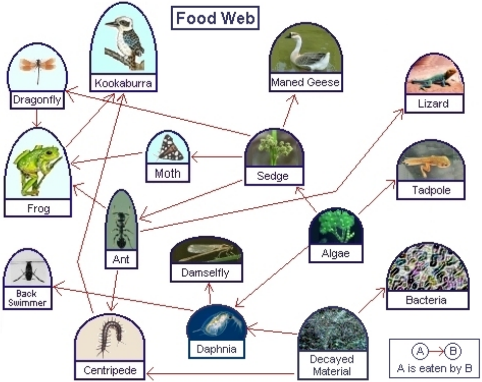
(reti telefoniche, reti stradali, reti ferroviarie, rotte aeree)





Rete delle linee aeree in India / autostrade EU

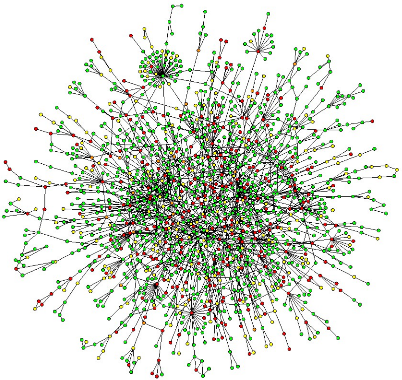
Reti che descrivono sistemi ecologici



Food web (catena alimentare)

Reti biologiche

(reti neurali, reti di trascrizione genetica, reti metaboliche...)



Rete di interazione delle proteine in un fungo

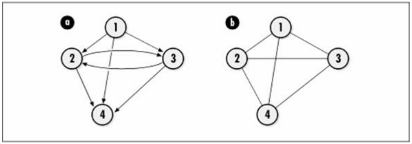
2. Le reti in matematica: i grafi.

Cos'è un grafo?

In matematica una rete si chiama grafo. Un grafo è una coppia

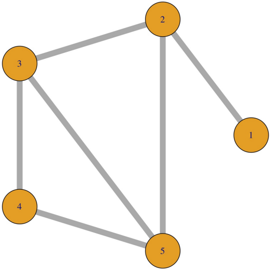
$$G = (\mathcal{N}, \mathcal{A})$$

dove $\mathcal{N}$ è un insieme di nodi e $\mathcal{A}$ è un insieme di archi (link) che li collegano.
Gli archi possono essere diretti o non diretti. Nel caso di archi diretti si parla di grafo diretto.



Grafo non diretto

Ecco un grafo non diretto

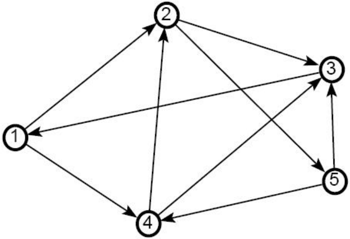


$$\mathcal{N} = \{1, 2, 3, 4, 5\}$$

$$\mathcal{A} = \{(1, 2); (2, 3); (2, 5); (3, 4); (3; 5); (4, 5)\}$$

Grafo diretto

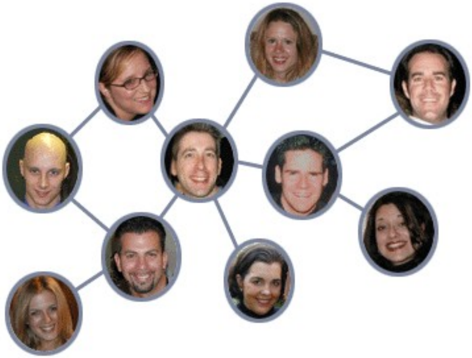
ed ecco un grafo diretto



$$\mathcal{N} = \{1, 2, 3, 4, 5\}$$

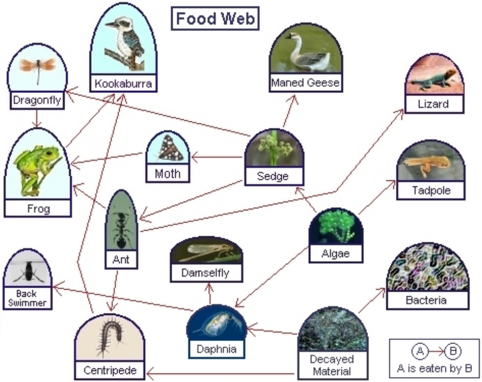
$$\mathcal{A} = \{(1, 2); (1, 4); (2, 3); (2, 5); (3, 1); (4; 3); (5; 3); (5; 4)\}$$

la rete di amicizie è non diretta



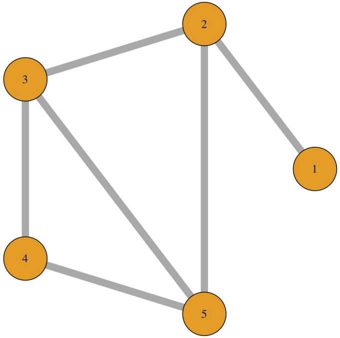
(... si spera...)

la catena alimentare è diretta



Grado di un nodo

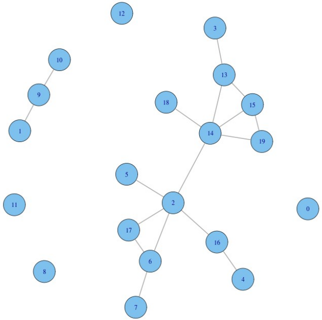
Dato un nodo in un grafo non diretto il suo grado è il numero di archi (link) che possiede.



$$deg(1) = 1,\ deg(2) = 3,\ deg(3) = 3,\ deg(4) = 2,\ deg(5) = 3$$

Componenti connesse

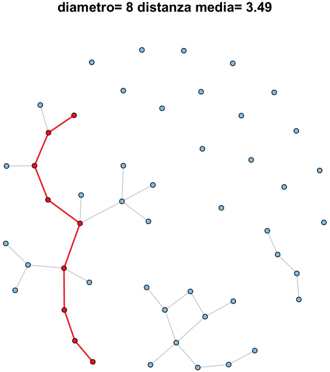
In un grafo non diretto la componente connessa di un nodo v è il sottografo costituito dai nodi raggiungibili da v.



Il 99.91 % della rete Facebook è un'unica componente connessa!

Diametro e distanza fra due nodi

Dato un grafo non diretto la distanza fra due punti a e b (nella stessa componente connessa) è il numero di nodi nel percorso più breve che collega a e b. Il diametro è la distanza massima fra le coppie di nodi contenuti nella stessa componente connessa.



Small-world (mondo piccolo)

In molti grafi reali la distanza fra due nodi è molto bassa.

Sei gradi di separazione

Distanza media fra due utenti Facebook

anno	distanza media
2008	5.28
2011	4.74
2016	3.57

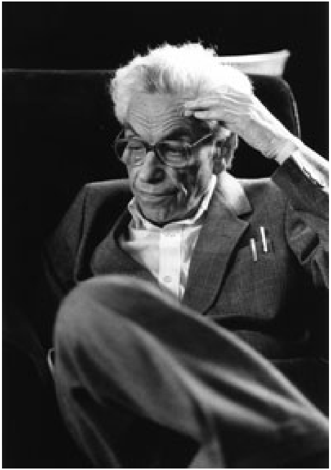
Hub

Un hub (mozzo, elemento centrale) è un nodo con molti più link degli altri...

L'aeroporto JFK di New York è un HUB!

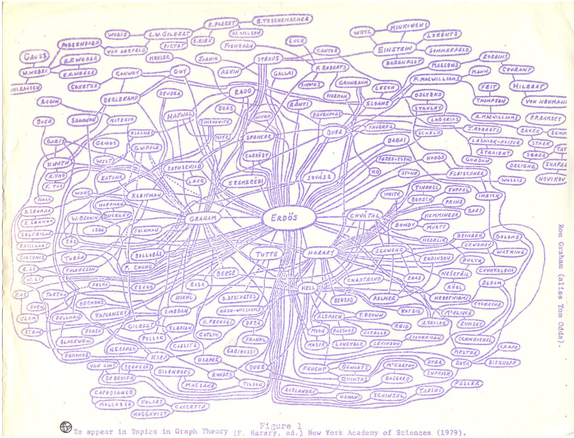
Hub

Anche Paul Erdös è un hub...



Paul Erdös (Budapest, 26 marzo 1913– Varsavia, 20 settembre 1996)

rete di collaborazioni (vicino a Erdös)



3. Reti (grafi) aleatori.

Perché un grafo aleatorio?

Molti grafi non nascono in modo deterministico.

Vogliamo confrontarli con ipotetiche realizzazioni di grafi casuali per comprendere meglio con che dinamiche si formano, per prevederene (con una certa probabilità) il comportamento...

Perché un grafo aleatorio?

Molti grafi non nascono in modo deterministico.

Vogliamo confrontarli con ipotetiche realizzazioni di grafi casuali per comprendere meglio con che dinamiche si formano, per prevederene (con una certa probabilità) il comportamento...

Cos'è un grafo aleatorio?

casuale, aleatorio = non deterministico
= non completamente prevedibile
$\Downarrow$
probabilità

Cos'è un grafo aleatorio?

casuale, aleatorio = non deterministico

= non completamente prevedibile

$\Downarrow$

probabilità

Modelli di grafi aleatori

Erdös-Rény (1959)

...

Molti modelli sono nati per descrivere caratteristiche specifiche osservate in grafi reali.

...

Watts and Strogatz (1998) (small word)

Albert and Barabási (1999,2002) (preferential attachment)

...

Modelli di grafi aleatori

Erdös-Rény (1959)

...

Molti modelli sono nati per descrivere caratteristiche specifiche osservate in grafi reali.

...

Watts and Strogatz (1998) (small word)

Albert and Barabási (1999,2002) (preferential attachment)

...

Cos'è un grafo aleatorio?

Per descrivere una rete aleatoria è importante specificare come genero tale rete, ossia con quali regole probabilistiche la costruisco.

In termini matematici quale sia la sua distribuzione di probabilità.

$$Prob(\gamma) \qquad \text{per ogni } \gamma \text{ in } \Gamma$$

dove Γ è l'universo dei grafi γ che voglio considerare.

Cos'è un grafo aleatorio?

Per descrivere una rete aleatoria è importante specificare come genero tale rete, ossia con quali regole probabilistiche la costruisco.

◇◇

In termini matematici quale sia la sua distribuzione di probabilità.

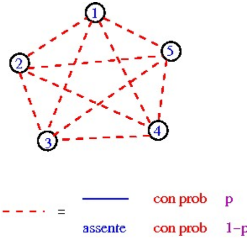
$$Prob(\gamma) \qquad \text{per ogni } \gamma \text{ in } \Gamma$$

dove Γ è l'universo dei grafi γ che voglio considerare.

Grafo alla Erdös-Rényi $\mathcal{G}(n, p)$

$\mathcal{G}(n, p)$

- Fisso un numero n nodi.
- Considero tutti gli archi possibili, li accendo (in modo indipendente) con probabilità p (ossia, tiro una moneta e se esce testa metto il link mentre se esce croce non lo metto).



Distribuzione dei link in Erdös-Rényi

Come calcolo la probabilità che $deg(g) = k$, ossia che un nodo g abbia k links?

Guardo tutti i possibili archi che collegano g agli altri $n-1$ nodi, ossia in tutto $n-1$ archi. Ricordo che nel grafo aleatorio in questione un arco viene acceso con probabilità p e non viene acceso con probabilità $1-p$.

Dunque

$$P\{\text{un nodo abbia } k \text{ links}\} = P\{\text{ottenere } k \text{ successi su } n-1 \text{ tiri}\}$$

Distribuzione dei link in Erdös-Rényi

Come calcolo la probabilità che $deg(g) = k$, ossia che un nodo g abbia k links?

Guardo tutti i possibili archi che collegano g agli altri $n - 1$ nodi, ossia in tutto $n - 1$ archi. Ricordo che nel grafo aleatorio in questione un arco viene acceso con probabilità p e non viene acceso con probabilità $1 - p$.

Dunque

$$P\{\text{un nodo abbia } k \text{ links}\} = P\{\text{ottenere } k \text{ successi su } n-1 \text{ tiri}\}$$

Distribuzione dei link in Erdös-Rényi

Come calcolo la probabilità che $deg(g) = k$, ossia che un nodo g abbia k links?

Guardo tutti i possibili archi che collegano g agli altri $n-1$ nodi, ossia in tutto $n-1$ archi. Ricordo che nel grafo aleatorio in questione un arco viene acceso con probabilità p e non viene acceso con probabilità $1-p$.

Dunque

$$P\{\text{un nodo abbia } k \text{ links}\} = P\{\text{ottenere } k \text{ successi su } n-1 \text{ tiri}\}$$

Distribuzione dei links in Erdös-Rényi

In altri termini:

◇◇ Distribuzione binomiale dei links

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

dove

$$\binom{n-1}{k} = \frac{(n-k)(n-k+1)\dots(n-1)}{2 \cdot 3 \dots k}$$

$p =$ probabilità di accensione

Numero medio di link in un nodo.

◇◇

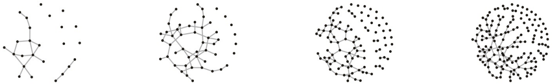
Il numero medio di link di un nodo è

$$\langle deg(g) \rangle = p(n-1).$$

Si ricordi che il numero medio di una v.a. binomiale $B(n-1,p)$ è $(n-1)p$.

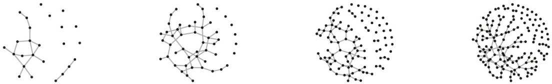
Cosa succede quando n è grande?

Tipicamente i grafi reali hanno molti nodi, ossia n è molto grande.
Si studia la rete aleatoria per n tendente all'infinito (i.e., molto grande).



Cosa succede quando n è grande?

Tipicamente i grafi reali hanno molti nodi, ossia n è molto grande.
Si studia la rete aleatoria per n tendente all'infinito (i.e., molto grande).



Cosa succede quando n è grande?

Facciamo crescere n ma rendiamo piccolo p, mantenendo fisso il numero medio di link

$$\lambda = p(n-1).$$

◇◇

$$p = p_n = \frac{\lambda}{n-1}.$$

Con questa scelta il numero medio di link di un nodo è costante ed uguale a λ.

Cosa succede quando n è grande?

Facciamo crescere n ma rendiamo piccolo p, mantenendo fisso il numero medio di link
$$\lambda = p(n-1).$$

◇◇

$$p = p_n = \frac{\lambda}{n-1}.$$

Con questa scelta il numero medio di link di un nodo è costante ed uguale a λ.

Approsimazione di Poisson.

◇◇ Ricordando che $p = \lambda/(n-1)$,

$$P\{\text{un nodo abbia } k \text{ links}\} = \binom{n-1}{k}\left(\frac{\lambda}{n-1}\right)^k\left(1-\frac{\lambda}{n-1}\right)^{n-1-k}$$

◇◇ Per n molto grande

$$P\{deg(g) = k\} \sim \frac{\lambda^k}{k!}e^{-\lambda}$$

dove $k! = 2 \cdot 3 \dots k$.

La distribuzione di probabilità

$$\frac{\lambda^k}{k!}e^{-\lambda} \qquad k = 0, 1, 2, \dots$$

è detta distribuzione di Poisson.

Approsimazione di Poisson.

◇◇ Ricordando che $p = \lambda/(n-1)$,

$$P\{\text{un nodo abbia } k \text{ links}\} = \binom{n-1}{k}\left(\frac{\lambda}{n-1}\right)^k\left(1-\frac{\lambda}{n-1}\right)^{n-1-k}$$

◇◇ Per n molto grande

$$P\{deg(g) = k\} \sim \frac{\lambda^k}{k!}e^{-\lambda}$$

dove $k! = 2 \cdot 3 \ldots k$.

La distribuzione di probabilità

$$\frac{\lambda^k}{k!}e^{-\lambda} \qquad k = 0, 1, 2, \ldots$$

è detta distribuzione di Poisson.

Approsimazione di Poisson.

◇◇ Ricordando che $p = \lambda/(n-1)$,

$$P\{\text{un nodo abbia } k \text{ links}\} = \binom{n-1}{k}\left(\frac{\lambda}{n-1}\right)^k\left(1-\frac{\lambda}{n-1}\right)^{n-1-k}$$

◇◇ Per n molto grande

$$P\{deg(g) = k\} \sim \frac{\lambda^k}{k!}e^{-\lambda}$$

dove $k! = 2 \cdot 3 \ldots k$.

La distribuzione di probabilità

$$\frac{\lambda^k}{k!}e^{-\lambda} \qquad k = 0, 1, 2, \ldots$$

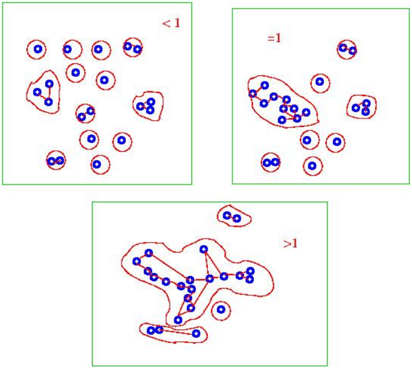
è detta distribuzione di Poisson.

simulazioni...

...è ora di fare un po' di simulazioni...

Transizione di fase nelle componenti connesse.

Quando n diverge (diventa grande) osserviamo una transizione di fase dipendente dal parametro λ.



Transizione di fase nelle componenti connesse.

Con probabilità uno, quando $n \to +\infty$

◇◇

- se $\lambda < 1$, allora ogni componente connnessa C contiene un numero di nodi $\sim k_\lambda \log(n)$;
- se $\lambda > 1$, allora esiste un'unica componente connessa (componente gigante) comprendente un numero di nodi $\sim z_\lambda n$ mentre tutte le altre componenti connesse contengono $\sim \log(n)$;
- se $\lambda = 1$, la componente gigante ha dimensione $cn^{2/3}$;

◇◇ per n grande

$$\log(n) << n^{2/3} << n.$$

Transizione di fase nelle componenti connesse.

Con probabilità uno, quando $n \to +\infty$

◇◇

- se $\lambda < 1$, allora ogni componente connnessa C contiene un numero di nodi $\sim k_\lambda \log(n)$;
- se $\lambda > 1$, allora esiste un'unica componente connessa (componente gigante) comprendente un numero di nodi $\sim z_\lambda n$ mentre tutte le altre componenti connesse contengono $\sim \log(n)$;
- se $\lambda = 1$, la componente gigante ha dimensione $cn^{2/3}$;

◇◇ per n grande

$$\log(n) << n^{2/3} << n.$$

Transizione di fase nelle componenti connesse.

Con probabilità uno, quando $n \to +\infty$

◇◇

- se $\lambda < 1$, allora ogni componente connnessa C contiene un numero di nodi $\sim k_\lambda \log(n)$;
- se $\lambda > 1$, allora esiste un'unica componente connessa (componente gigante) comprendente un numero di nodi $\sim z_\lambda n$ mentre tutte le altre componenti connesse contengono $\sim \log(n)$;
- se $\lambda = 1$, la componente gigante ha dimensione $cn^{2/3}$;

◇◇ per n grande

$$\log(n) << n^{2/3} << n.$$

Transizione di fase nelle componenti connesse.

Con probabilità uno, quando $n \to +\infty$

◇◇

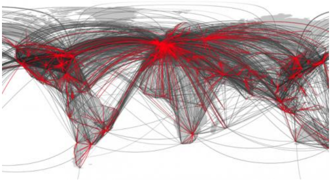
- se $\lambda < 1$, allora ogni componente connnessa C contiene un numero di nodi $\sim k_\lambda \log(n)$;
- se $\lambda > 1$, allora esiste un'unica componente connessa (componente gigante) comprendente un numero di nodi $\sim z_\lambda n$ mentre tutte le altre componenti connesse contengono $\sim \log(n)$;
- se $\lambda = 1$, la componente gigante ha dimensione $cn^{2/3}$;

◇◇ per n grande

$$\log(n) << n^{2/3} << n.$$

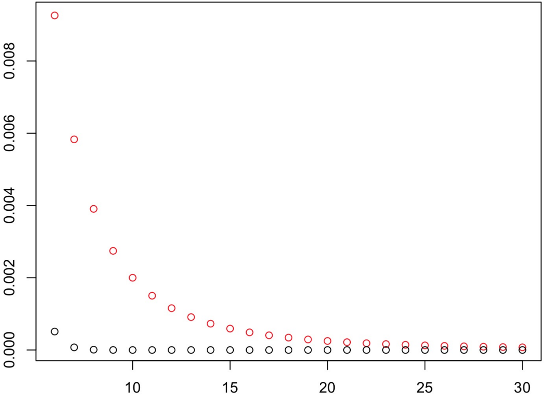
... tornando alla distribuzione dei link:

Come sono fatti i grafi reali? Che distribuzione di link hanno?



code pesanti vs. code leggere

decresce velocemente= code leggere
decresce lentamente= code pesanti



code pesanti vs. code leggere

Più formalmente:

◇◇

- code leggere:

$$Prob(\text{k links}) \sim e^{-k};$$

- code pesanti (legge di potenza di esponente γ):

$$Prob(\text{k links}) \sim \frac{1}{k^{\gamma}}$$

... assenza/presenza di hub.

Se ho code leggere è poco probabile che si formi un hub, la quasi totalità dei nodi ha un numero di link simile ed è molto imporbabile che nasca un nodo con distribuzione che si discosti significativamente dalla media.

Se ho code pesanti è probabile si formi un hub, la maggior parte dei nodi avrà un numero relativamente basso di link, ma qualche nodo avrà molti link.

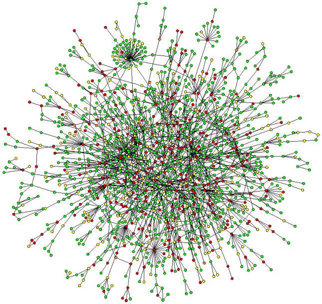
... assenza/presenza di hub.

Se ho code leggere è poco probabile che si formi un hub, la quasi totalità dei nodi ha un numero di link simile ed è molto imporbabile che nasca un nodo con distribuzione che si discosti significativamente dalla media.

Se ho code pesanti è probabile si formi un hub, la maggior parte dei nodi avrà un numero relativamente basso di link, ma qualche nodo avrà molti link.

grafi reali

In molti grafi reali si osserva la presenza di hub. Inoltre, studiando la distribuzione empirica (la frequenza) dei link dei vari nodi, si osserva tipicamente un comportamento di tipo code pesanti.



grafi reali

Code pesanti:

$$Prob(\text{k link}) \sim \frac{1}{k^\gamma}.$$

Ecco alcuni esempi

- rete reazioni metaboliche: n.nodi $500 - 800$, n.link $1500 - 300$, $\gamma = 2.2$;
- rete chiamate telefoniche: n.nodi 46×10^6, n.link 8×10^8, $\gamma_{in} = 2.1$;
- rete di collaborazione fra gli attori nel cinema: n.nodi $212,000$, n.link 61×10^6, $\gamma = 2.3$;
- rete di collaborazione fra matematici: n.nodi $70,000$, n.link $132,000$, $\gamma = 2.1$;
- ...

modelli più recenti

Molti grafi reali non sono dunque descritti bene dal modello alla Erdös-Rényi...

preferentail attachment

... in real networks linking is very often *preferential*... (Barabási-Albert)
... nei grafi reali i collegamenti sono spesso di tipo preferenziale ...

idea:
Se ho molti amici, in futuro ne avrò ancora di più...

modelli più recenti

Molti grafi reali non sono dunque descritti bene dal modello alla Erdös-Rényi...

preferentail attachment

... in real networks linking is very often *preferential*...
(Barabási-Albert)
... nei grafi reali i collegamenti sono spesso di tipo preferenziale ...

idea:

Se ho molti amici, in futuro ne avrò ancora di più...

modelli più recenti

Molti grafi reali non sono dunque descritti bene dal modello alla Erdös-Rényi...

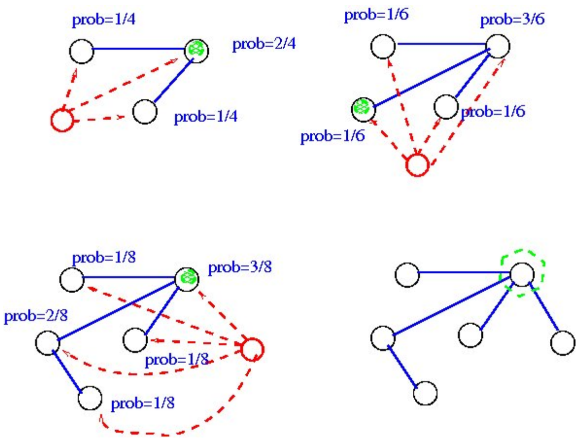
preferentail attachment

... in real networks linking is very often *preferential*...
(Barabási-Albert)
... nei grafi reali i collegamenti sono spesso di tipo preferenziale ...

idea:

Se ho molti amici, in futuro ne avrò ancora di più...

Albert and Barabási (1999,2002)



code pensanti in Albert and Barabási

◇◇ code pesanti

In questo modello si ha che

$$P(\text{un nodo } v \text{ ha } k \text{ link}) \propto \frac{1}{k^3}.$$

... altri problemi...

- connettività, sottoalberi, diametro massimo, distanza minima...
- diffusione su una rete (e.g., malattie)
- vulnerabilità di una rete
- previsione statistica

... altri problemi...

- connettività, sottoalberi, diametro massimo, distanza minima...
- diffusione su una rete (e.g., malattie)
- vulnerabilità di una rete
- previsione statistica

... altri problemi...

- connettività, sottoalberi, diametro massimo, distanza minima...
- diffusione su una rete (e.g., malattie)
- vulnerabilità di una rete
- previsione statistica

... altri problemi...

- connettività, sottoalberi, diametro massimo, distanza minima...
- diffusione su una rete (e.g., malattie)
- vulnerabilità di una rete
- previsione statistica

... altri problemi...

- connettività, sottoalberi, diametro massimo, distanza minima...
- diffusione su una rete (e.g., malattie)
- vulnerabilità di una rete
- previsione statistica

grazie per l'attenzione!