INFORMATICA PER LA COMUNICAZIONE

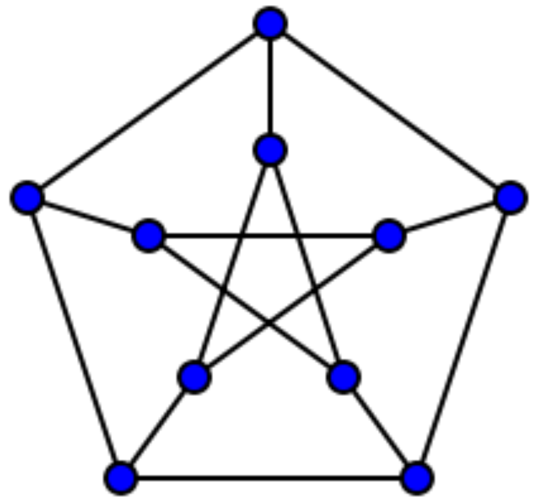
Reti e grafi

Reti casuali

Reti e grafi

Grafi regolari

- Fino agli studi di Erdos e Renyi la teoria dei grafi si concentra sulle proprietà di **grafi regolari**



Modelli di grafi

- I grafi vengono utilizzati per descrivere diversi **sistemi complessi**
- Quali **caratteristiche** hanno?
- Come è fatta la **struttura** generale di un grafo che rappresenta tali sistemi?
- Esistono **differenze** tra sistemi complessi differenti?

Modelli di reti

- Per spiegare le caratteristiche di tali sistemi: modelli **matematici**
- Definiscono le regole con cui **le reti si formano**
  - Quando si forma un legame (creazione di un arco)

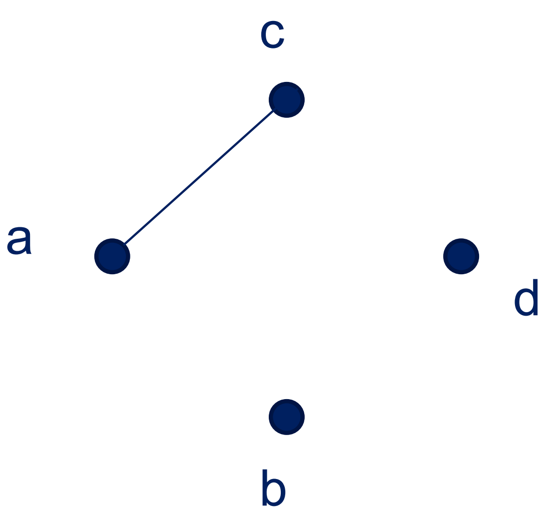
Grafi casuali

- Primo modello di riferimento per spiegare la **formazione** di grafi
- Introdotto nel 1959 da **Erdos** e **Renyi**
- Gli archi che collegano i vertici sono scelti **casualmente**

Grafi casuali

- Grafo con *N* nodi
- Gli archi che collegano i vertici sono scelti **casualmente** con una probabilità *p*:
  - Con **probabilità** *p* si **forma un arco** tra due nodi
  - In alternativa (probabilità *1-p*) **non si forma un arco** tra i due nodi

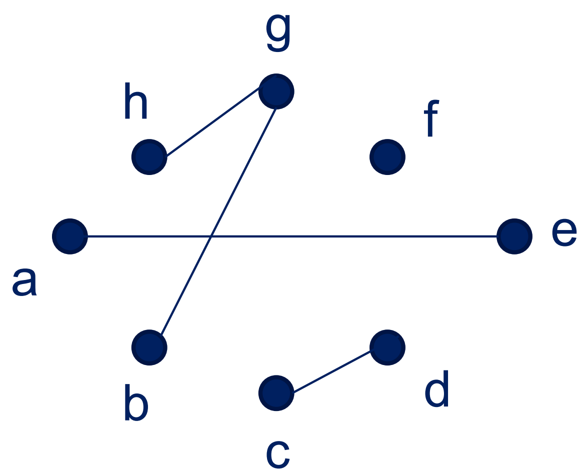
Grafi casuali



Per ogni coppia di nodi:

1. Con probabilità **p**: ho un **arco**
2. Altrimenti l'arco **non è presente**

Grafi casuali



Per ogni coppia di nodi:

1. Genero un **numero casuale**
2. Se è **maggiore di *p***: ho un **arco**
3. Altrimenti l'arco **non è presente**

Numero di archi

- Calcoliamo il numero **medio (valore atteso)** di archi

$$\bar{M} = p \, \frac{N(N-1)}{2}$$

Numero di archi

Qual è la probabilità che un grafo abbia *M* archi?

$$p_M$$

Calcoliamo il **valore atteso**, sommando per *M=0,…, N (N-1)/2*

$$\bar{M} = \sum_M M\, p_M = p\, \frac{N(N-1)}{2}$$

Cosa succede se p è molto piccolo o molto grande?

Numero di archi

Qual è la probabilità che un grafo abbia *M* archi?

$$p_M = \binom{N(N-1)/2}{M} p^M \, (1-p)^{\frac{N(N-1)}{2} - M}$$

Calcoliamo il valore atteso, sommando per M=0,…, N (N-1)/2

$$\bar{M} = \sum_M M \, p_M = p \, \frac{N(N-1)}{2}$$

Numero di archi

Essendo

$$\bar{M} = p \, \frac{N(N-1)}{2}$$

Grado medio $\bar{k} = \frac{2\,E}{N} = \frac{2\bar{M}}{N}$

Il grado medio di un nodo in una rete casuale è

$$\bar{k} = \frac{2\,\bar{M}}{N} = p\,(N-1)$$

Il **grado medio** (la **densità**) aumenta al **crescere di *p***

Numero di archi

Essendo

$$\bar{M} = \sum_M M\, p_M = p\, \frac{N(N-1)}{2} \quad \text{e} \quad \bar{k} = \frac{2\,E}{N}$$

Il grado medio di un nodo è

$$\bar{k} = \frac{2\,\bar{M}}{N} = p\,(N-1)$$

La **densità** aumenta al **crescere di *p***

Distribuzione del grado

- Qual è la **distribuzione del grado** di un grafo casuale?
- **Probabilità** che un nodo abbia grado *k*:
  - Probabilità che un nodo abbia *k* link: p^k
  - Probabilità che gli altri *N-1-k* link siano assenti: $(1-p)^{N-1-k}$
  - Numero di scelte di *k* connessioni su *N-1*: $\binom{N-1}{k}$

Distribuzione del grado

- **Probabilità** che un nodo abbia grado *k*:

$$p_k = \binom{N-1}{k} p^k (1-p)^{N-1-k}$$

- Possiamo approssimare la distribuzione del grado (per N $\gg$ $\bar{k}$) con la distribuzione di Poisson:

$$p_k = e^{-k} \frac{\bar{k}}{k!}$$

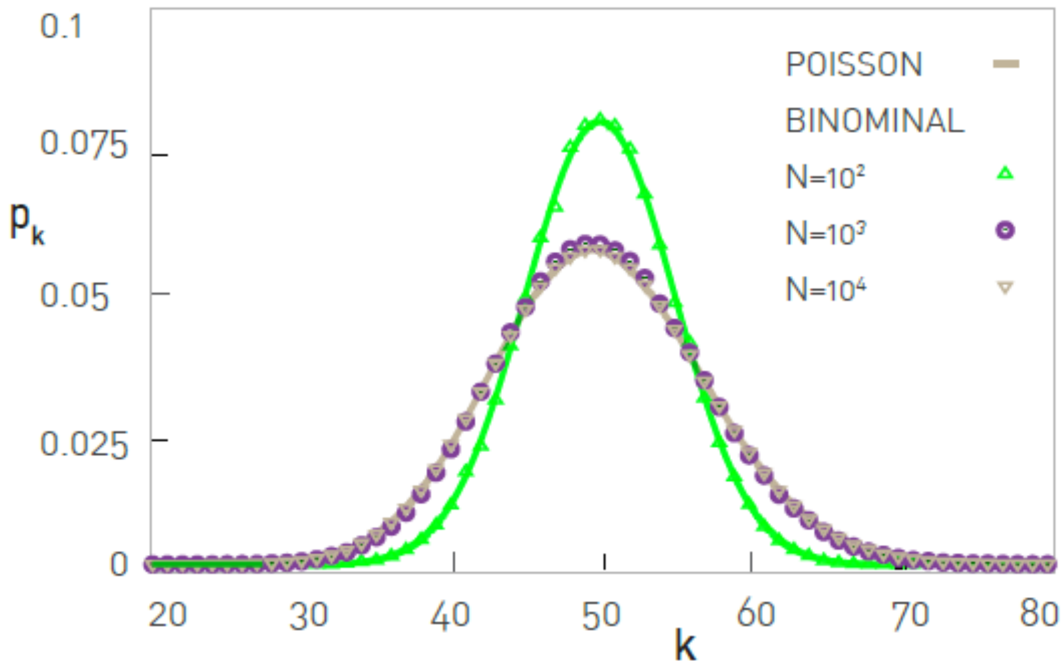
Distribuzione del grado

- Possiamo approssimare la distribuzione del grado (per $N \gg \bar{k}$) con la distribuzione di Poisson:

$$p_k = e^{-k} \frac{\bar{k}}{k!}$$

Distribuzione del grado

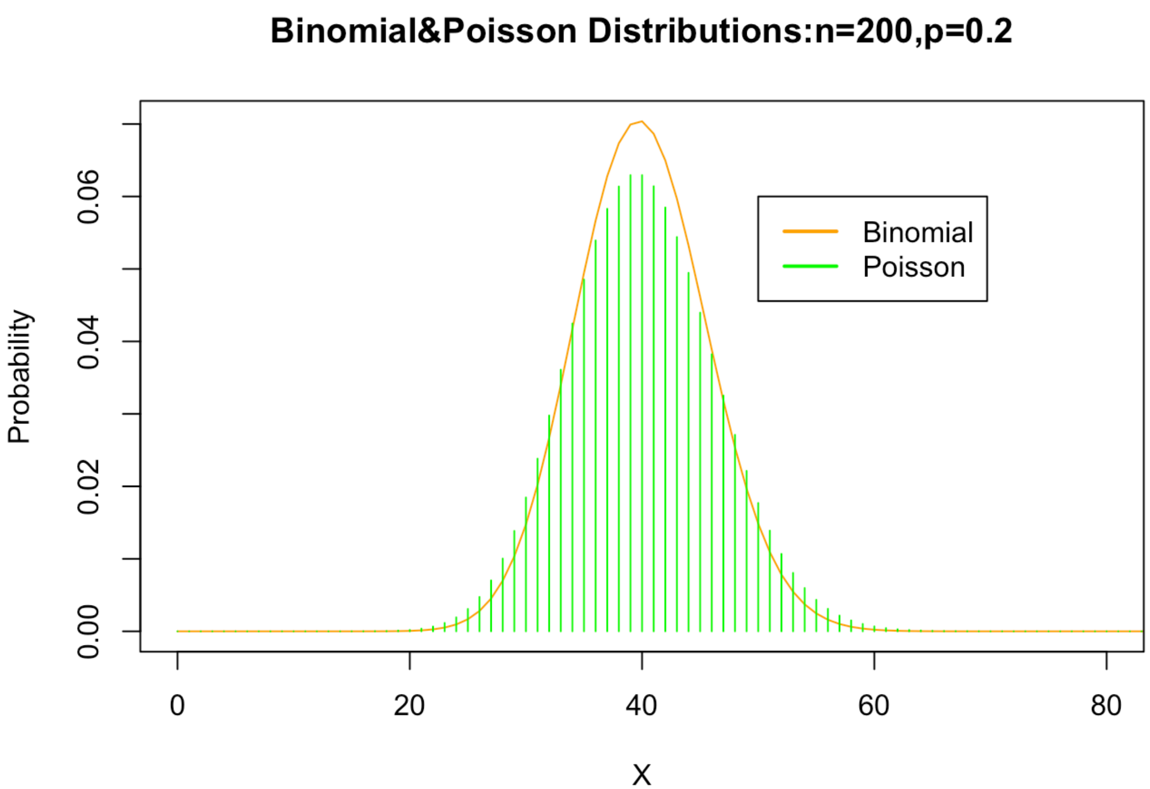
Distribuzione binomiale e di Poisson



Da A. L. Barabasi, *Network Science*

Distribuzione del grado

Distribuzione binomiale e di Poisson



Da https://rpubs.com/mcolwell/316763

Evoluzione dei grafi casuali

- Cosa accade ai grafi casuali nella loro formazione?
- Come cambiano le connessioni in base al parametro *p*?
- Quando si formano componenti connesse?

- Consideriamo il rapporto N_G/N, dove N_G è il numero di nodi **della più grande componente connessa**

Grafi casuali

Proprietà dei grafi casuali (Erdos e Renyi):

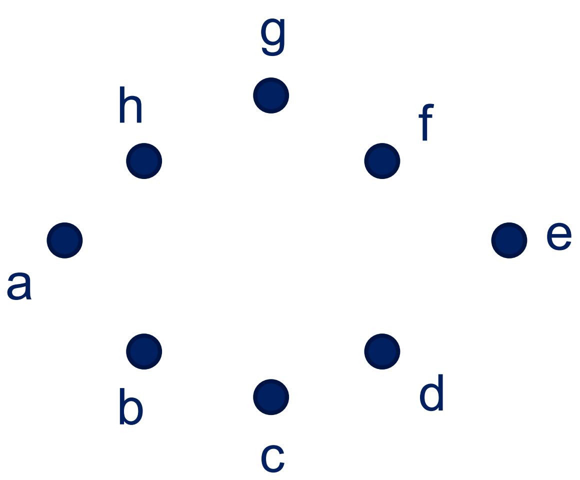
- Se in media ogni nodo ha grado **uno**: componente con molti nodi $\longrightarrow$ **componente gigante**
- Se in media i nodi hanno grado **inferiore a uno**: il grafo non è connesso

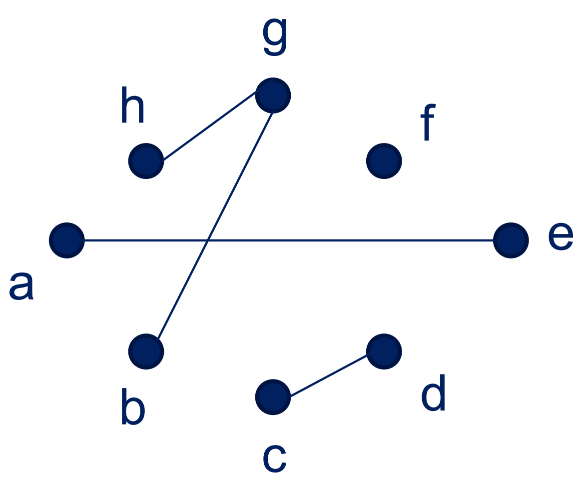
Evoluzione dei grafi casuali

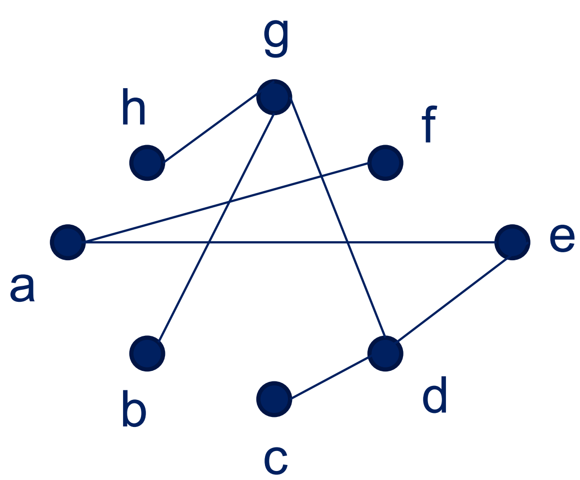
Proprietà dei grafi casuali (Erdos e Renyi):

- Pochi archi nel grafo (grado medio basso) $\longrightarrow$ componenti **non connesse**; N_G/N è vicino a zero
- Se vengono inseriti più archi (il grado medio viene incrementato): le **componenti si connettono**
- Oltre un **valore critico**: una sola componente che include quasi tutti i nodi

Grafi casuali







Evoluzione dei grafi casuali

- Nell’evoluzione delle reti casuali, posso individuare quattro fasi:
- **Regime subcritico**
- **Punto critico**
- **Regime supercritico**
- **Regime connesso**

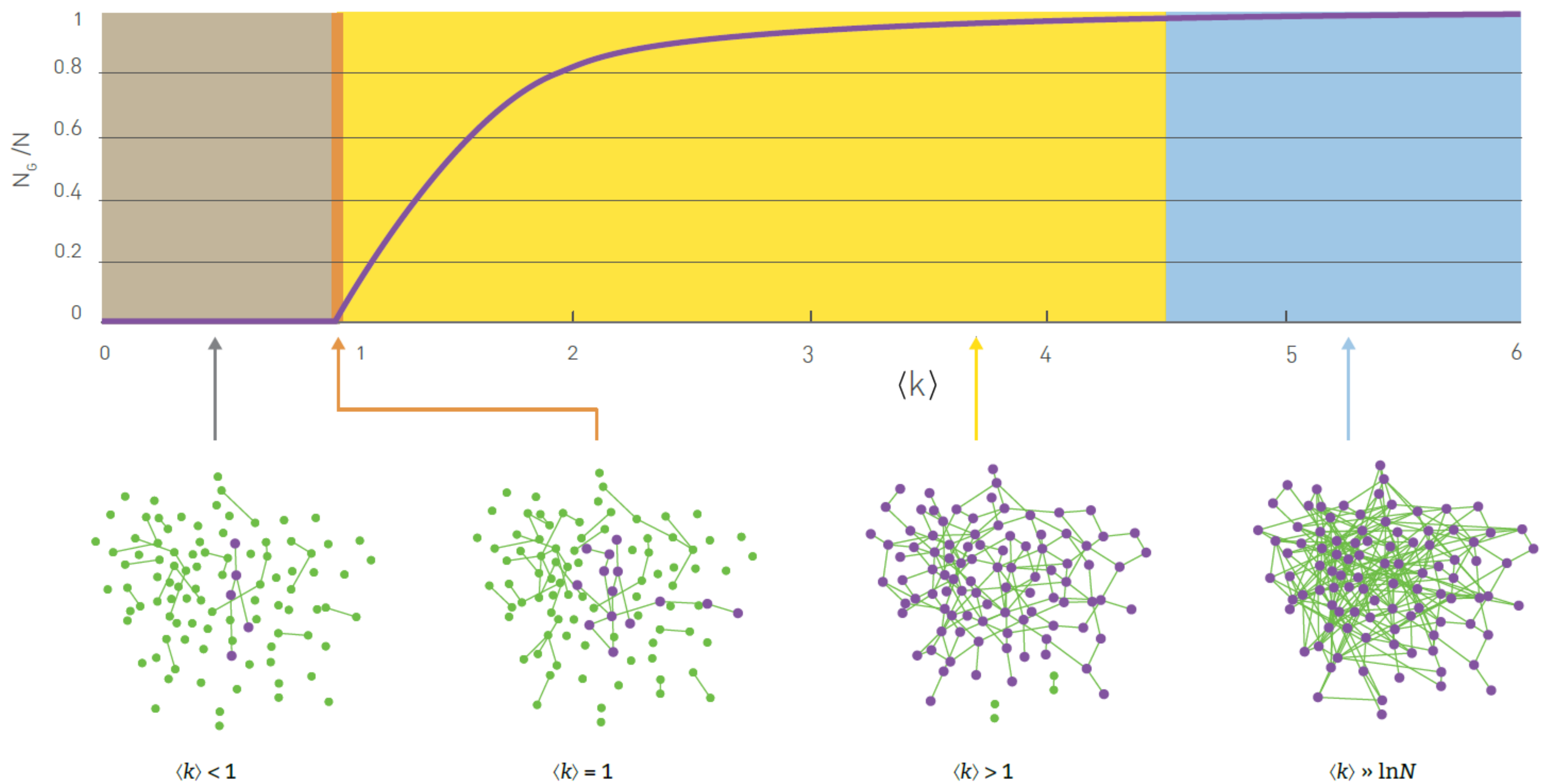
Evoluzione dei grafi casuali

- **Regime subcritico**: grado medio < 1
  - Piccole **componenti connesse**
  - Grafo sostanzialmente **disconnesso**
- **Punto critico**: grado medio = 1
  - Inizio della formazione della **componente gigante**

Evoluzione dei grafi casuali

- **Regime supercritico**: grado medio > 1
- Formazione di una **componente gigante** che contiene una **parte molto significativa** della rete
- **Regime connesso**: grado medio > *ln N*
  - Componente connessa che contiene la totalità dei nodi

Evoluzione dei grafi casuali



Da A. L. Barabasi, *Network Science*

Grafi casuali e reti reali

Proprietà dei grafi casuali di particolare importanza per le reti reali:

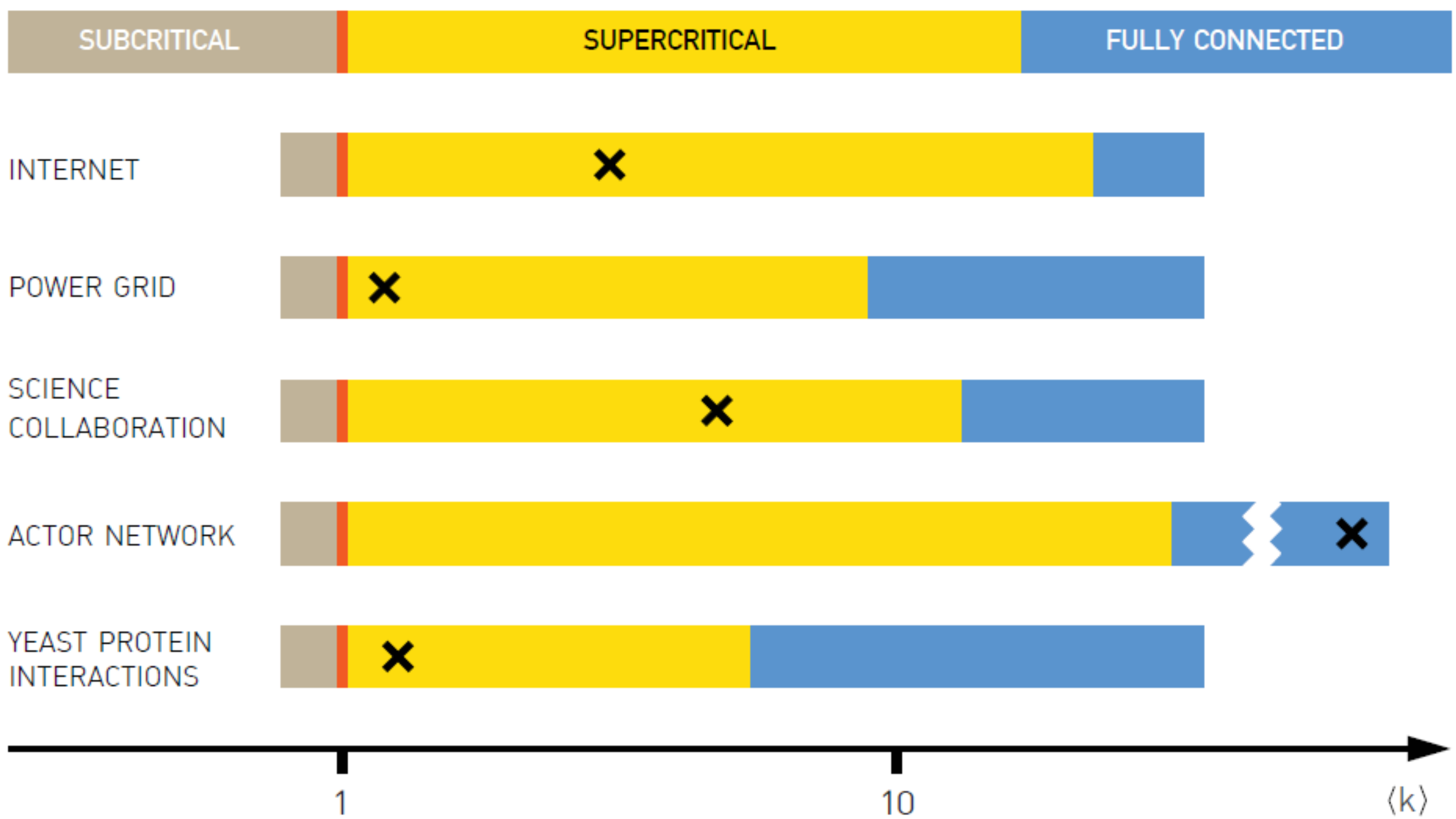
- **Connessione** progressiva dei nodi con grado medio >= 1
- Esistenza di un'unica **componente gigante** per grado medio > *ln N*

Grafi casuali e reti reali

NETWORK	N	E	$\bar{k}$	lnN
Internet	192,244	609,066	6.34	12.17
Power Grid	4,941	6,594	2.67	8.51
Science Collaboration	23,133	94,439	8.08	10.05
Actor Network	702,388	29,397,908	83.71	13.46
Protein Interactions	2,018	2,930	2.90	7.61

Da A. L. Barabasi, *Network Science*

Grafi casuali e reti reali



Da A. L. Barabasi, *Network Science*

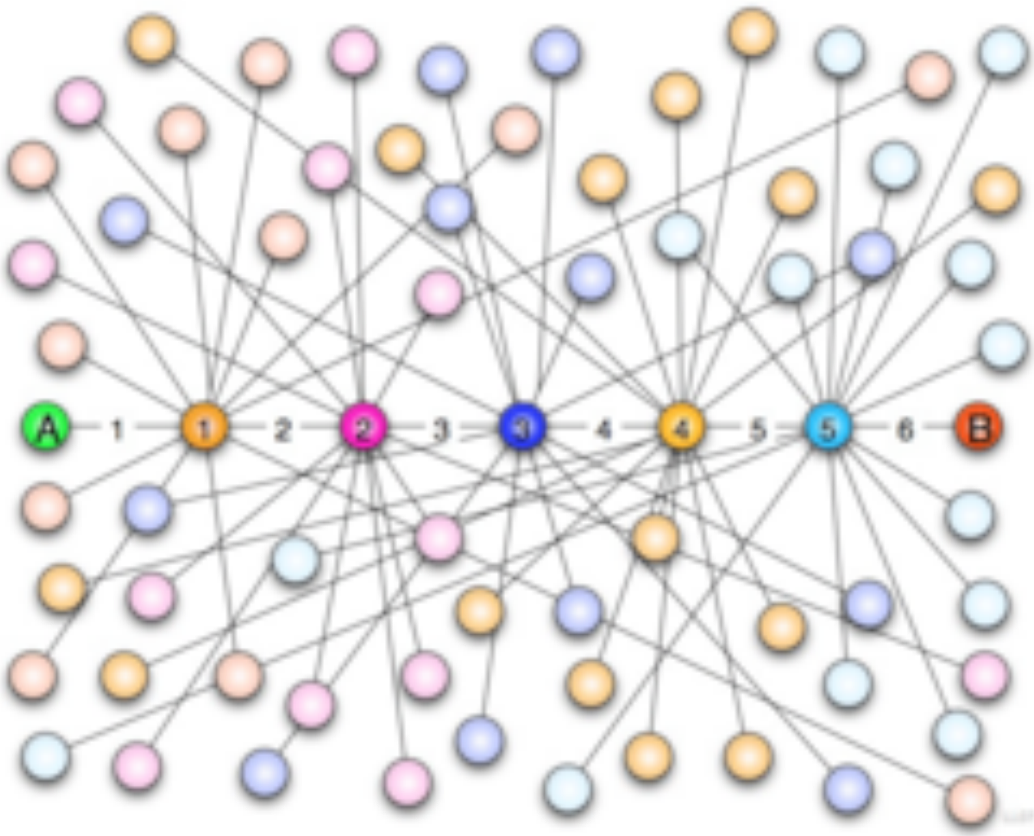
Gradi di separazione

Reti sociali e grafi

Proprietà delle reti

- **Caratteristica fondamentale** di alcune reti: **diametro** di un grafo
- Nelle reti casuali e in molti casi concreti: **diametro limitato**
- Nel 1928 prospettato da uno scrittore, Karinthy
- Noto come **sei gradi di separazione**

Reti– sei gradi di separazione



Reti– sei gradi di separazione

- **Ricerca** del sociologo Milgram (anni ’60)
- Invio di messaggi tra due città degli Stati Uniti

Esperimento di Milgram

- L'esperimento di Milgram prevede che ogni persona:
  - Consegni il **messaggio al destinatario, se è un conoscente**
  - Invii un messaggio **a un conoscente** che abbia maggiore probabilità di consegnare il messaggio al destinatario

Esperimento di Milgram

- In totale
  - 296 lettere inviate
  - 64 lettere arrivarono a destinazione
- Quanti passaggi necessari?
  - **Numero medio** di passaggi tra intermediari: **5.2**
  - **Sei gradi di separazione**

Reti– sei gradi di separazione

Esempio: **Rete persone - conoscenze**

- Per collegare uno studente Unibg e Presidente degli Stati Uniti

1. Studente → Rappresentante degli studenti
2. Rappresentante degli studenti → Rettore
3. Rettore → Ministro Università
4. Ministro Università → Presidente del Consiglio
5. Presidente del Consiglio → Presidente degli Stati Uniti

Reti– sei gradi di separazione

Possibile spiegazione

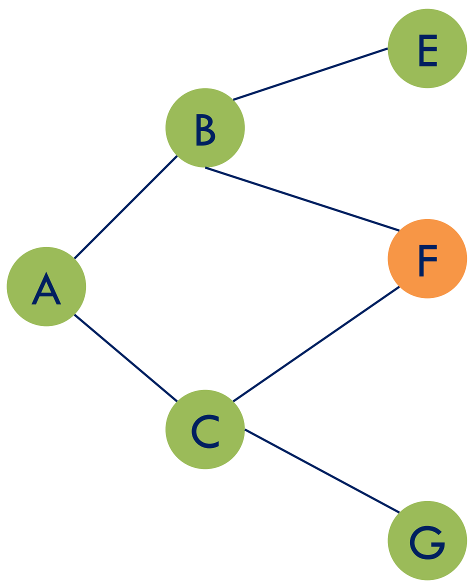
- **Ipotesi**: una persona conosce in media 100 persone

Persone raggiungibili

- 1 passaggio → 100
- 2 passaggi→ 100*100 = 10.000
- 3 passaggi → 100*10.000 = 1.000.000
- 4 passaggi → 100*1.000.000 = 100.000.000
- 5 passaggi → 100*100.000.000 = 10.000.000.000

Reti– sei gradi di separazione

- Stima **non esatta**
- Persone considerate più di una volta



Se A, B, C avessero 2 nuovi vicini distinti: 2 * 2 nodi raggiungibili da A con 1 livello intermedio

Grado di separazione del Web

- Nel 1998 Barabasi analizzò il **Web**: rete formata da link ipertestuali
- Quanti passaggi sono necessari per passare da una pagina a un’altra?
- **Problemi di analisi** del Web
  - Come fare a individuare le pagine?

Grado di separazione del Web

- Analisi preliminare del dominio **nd.edu**
- Costituito da circa **100.000 pagine**
- Grado di separazione: **11**
- Nel Web (al 1999), considerando il numero di pagine, il grado di separazione dovrebbe essere **33.000**?

Grado di separazione del Web

- Basandosi su risultati sperimentali su insiemi sempre più vasti: **stima** del grado di separazione del Web
- Risultati dell'analisi
  - **800.000.000** di nodi
  - Grado di separazione: 18,59 → **19 gradi di separazione**

Reti – sei gradi di separazione

- **Caratteristiche** di molte reti reali:
  - Numero notevole di nodi
  - **Bassa distanza media** tra i nodi
- Analisi della rete di Facebook (2011) e della sua evoluzione
  - Analisi della rete globale e di reti regionali

Rete di Facebook

- **Caratteristiche** rete Facebook (2007 - 2011)
- Utenti **attivi**: accesso ultimi 28 giorni
- Considerati utenti con al più 5000 amici
- Nel 2011:
  - **721** milioni di utenti/nodi
  - **69** miliardi di archi
  - **Grado di separazione: 4**

Rete di Facebook

- Nel 2007 grafo **disconnesso**
- Dal 2009 grafo **connesso**, dati consolidati
- Nel 2011
  - Grafi **regionali**: grado di separazione circa 3
  - **Grado medio di separazione**: 3,74
  - 92% nodi con grado di separazione <=4

Facebook, Web - Milgram

Differenze tra le analisi di Facebook/Web e l'esperimento di Milgram

- Nelle analisi di Facebook/Web: utilizzo **cammini minimi**
- Nell'esperimento di Milgram: i cammini sono minimi?

Facebook, Web - Milgram

Altre **differenze** tra le analisi di Facebook e l'esperimento di Milgram

- Amicizie reali/amicizie in Facebook
- Milgram: **22%** catene completate
- Facebook: 721 milioni di utenti; **92% nodi con grado di separazione ≤ 4**

Reti– sei gradi di separazione

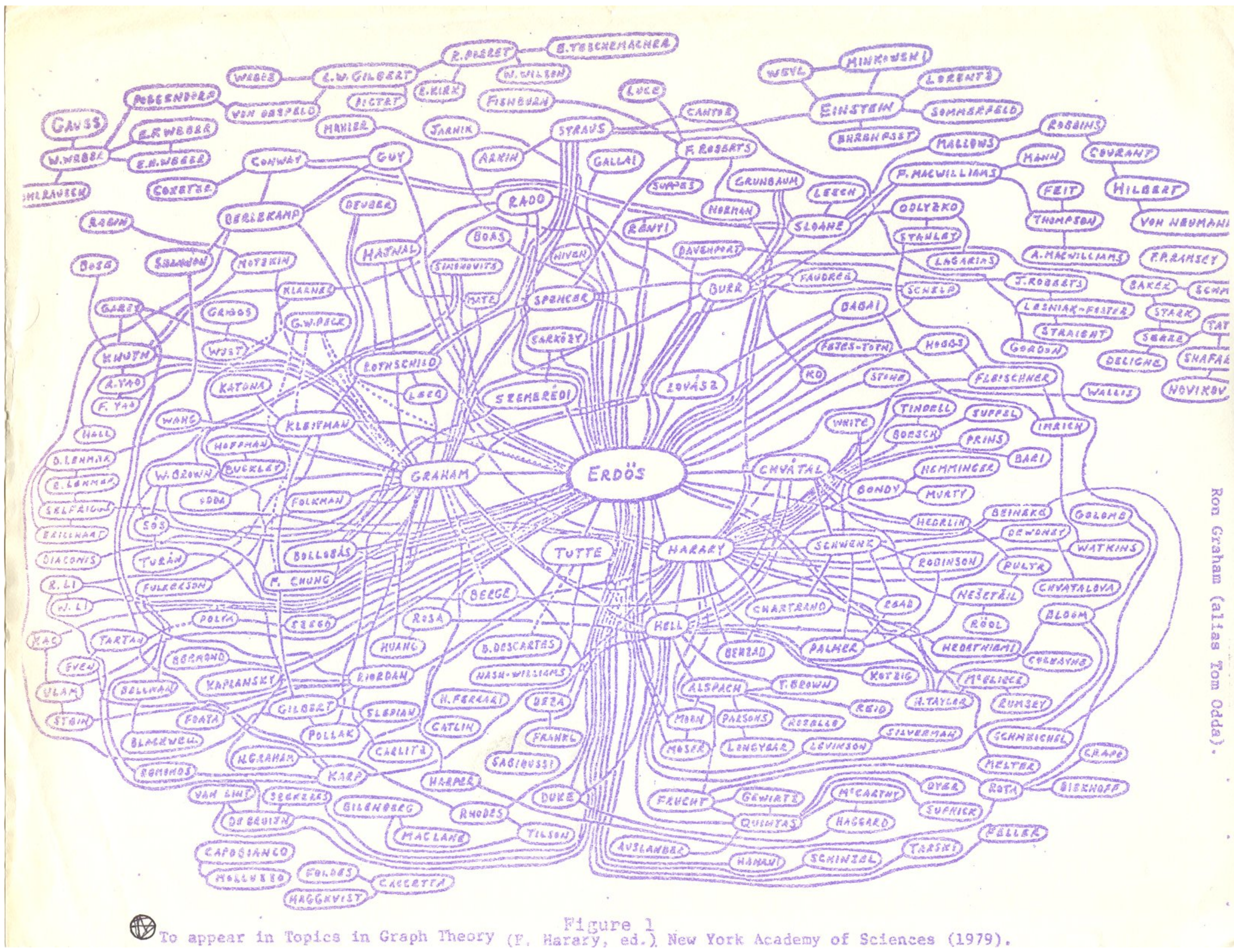
Esempio di altre **reti sociali**

- **Collaborazioni scientifiche**: esempio di relazione tra individui
- Collegamento tra due autori: **coautori** di una pubblicazione

Reti– sei gradi di separazione

- Collaborazioni matematiche-informatiche: **numero di Erdos**
  - *Paul Erdos* (1913-1996): matematico ungherese
  - Rete di collaborazioni molto ampia

Reti– sei gradi di separazione

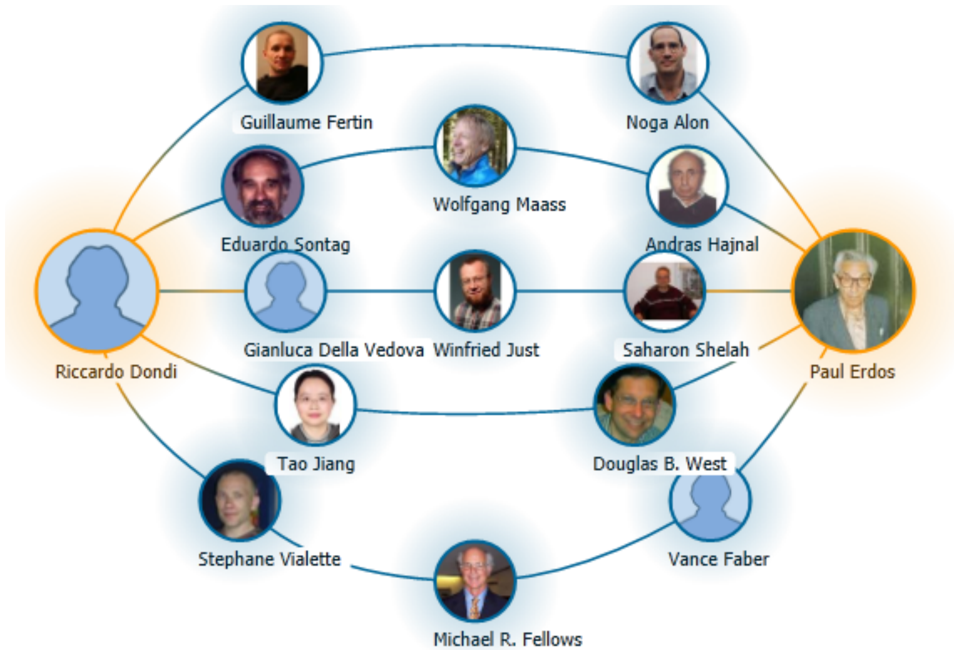


Reti– sei gradi di separazione

- **Numero di Erdos** di un matematico/informatico: lunghezza del più breve percorso che porta da Erdos all'autore
- Numero associato a Erdos → 0

Reti– sei gradi di separazione

- **Stima**: *90% dei matematici hanno un numero di Erdos non superiore a 8*
- Calcolo numero di Erdos



Reti sociali – sei gradi di separazione

- La rete di Erdos è in **continua evoluzione**
  - Aumentano i nodi
  - Nuove collaborazioni: nuovi archi
    - Tra nodi già presenti
    - Collegamento di nuovi nodi

Reti sociali – sei gradi di separazione

- Il **numero di Bacon** misura la distanza nella rete degli attori da Kevin Bacon
- Misurazione della distanza da Bacon

Reti – sei gradi di separazione

- Fissiamo un nodo *x* della rete
- I nodi a distanza *d* da *x* sono:

$$N(d) = 1 + \bar{k} + \bar{k}^2 + \ldots + \bar{k}^d = \frac{\bar{k}^{d+1} - 1}{\bar{k} - 1}$$

Se *d* è la distanza massima

$$N(d) = N$$

Reti – sei gradi di separazione

- Il numero di nodi a distanza al più *d* da un nodo ($\bar{k}$ è il grado medio):

$$N(d) = 1 + \bar{k} + \bar{k}^2 + \ldots + \bar{k}^d = \frac{\bar{k}^{d+1} - 1}{\bar{k} - 1} = \mathrm{N}$$

- Visto che *d* è il diametro *N(d)=N* (il numero totale di nodi)
- Ignorando i due valori *-1*, abbiamo

$$\bar{k}^d = \mathrm{N}$$

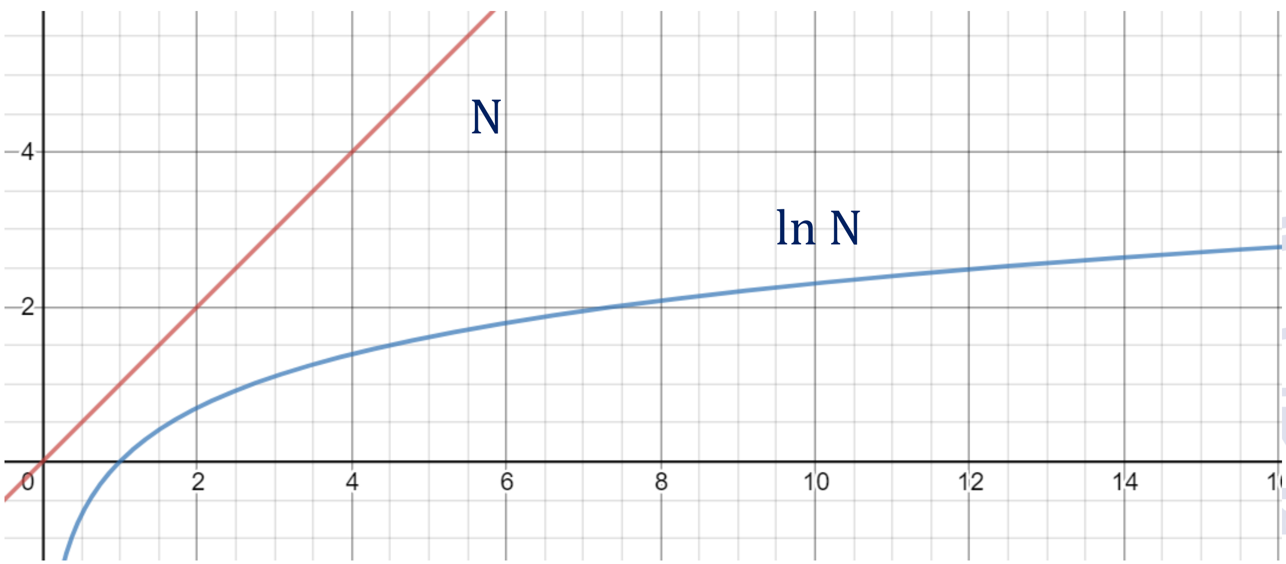
- Segue

$$d = \frac{\ln N}{\ln \bar{k}}$$

Reti – sei gradi di separazione

Distanza al più *d* tra i nodi ($\bar{k}$ è il grado medio):

$$d = \frac{\ln N}{\ln \bar{k}}$$



Reti – sei gradi di separazione

Distanza al più *d* tra nodi:

$$d = \frac{\ln N}{\ln \bar{k}}$$

- Il termine $\ln N$ sta ad indicare che il **diametro** è sensibilmente più **piccolo** della dimensione della rete: **mondo piccolo**
- **Reti a piccolo mondo**

Reti – sei gradi di separazione

Distanza al più *d* tra nodi:

$$d = \frac{\ln N}{\ln \bar{k}}$$

- Termine $1/\ln \bar{k}$: influenza del grado medio
- **Reti più dense** hanno **diametro inferiore**

Reti – sei gradi di separazione

- In molti casi il diametro è influenzato da alcune coppie con distanza maggiore della media
- Empiricamente possiamo considerare:

$$\bar{d} = \frac{\ln N}{\ln \bar{k}}$$

Reti sociali – sei gradi di separazione

NETWORK	N	L	$\langle k \rangle$	$\langle d \rangle$	d_{max}	$\frac{\ln N}{\ln \langle k \rangle}$
Internet	192,244	609,066	6.34	6.98	26	6.58
WWW	325,729	1,497,134	4.60	11.27	93	8.31
Power Grid	4,941	6,594	2.67	18.99	46	8.66
Mobile Phone Calls	36,595	91,826	2.51	11.72	39	11.42
Email	57,194	103,731	1.81	5.88	18	18.4
Science Collaboration	23,133	93,439	8.08	5.35	15	4.81
Actor Network	702,388	29,397,908	83,71	3,91	14	3,04
Citation Network	449,673	4,707,958	10.43	11,21	42	5.55
E. Coli Metabolism	1,039	5,802	5.58	2.98	8	4.04
Protein Interactions	2,018	2,930	2.90	5.61	14	7.14

Da A. L. Barabasi, *Network Science*

Reti e clusterizzazione

Reti sociali e grafi

Limiti delle reti casuali

- Il modello delle **rete casuali non rappresenta le caratteristiche** di alcune reti reali
  - Reti sociali
  - Web e Internet
  - Reti biologiche
  - …

Legami di una rete

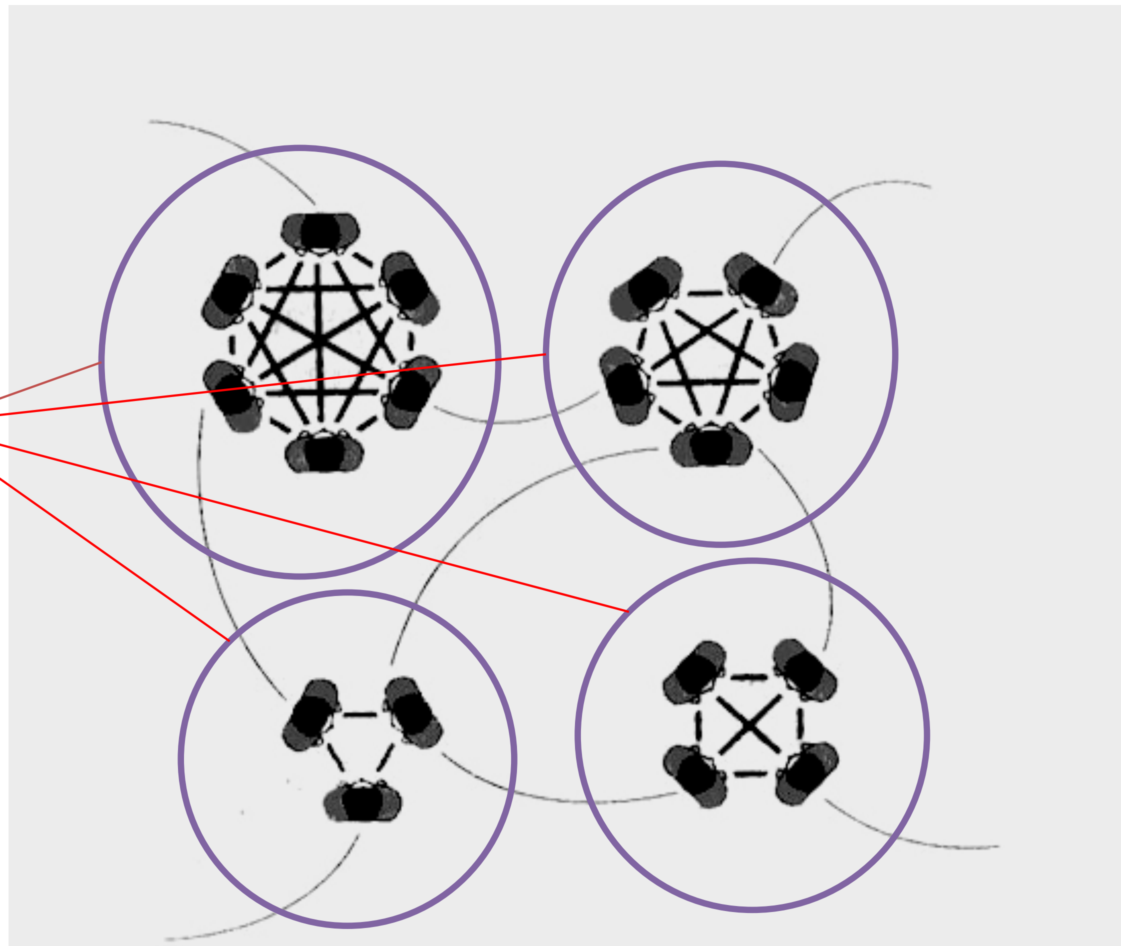
- Nel 1969-1973 Granovetter analizzò le reti interpersonali (nella ricerca di lavoro)
- I suoi studi individuarono
  - **Struttura** della rete
  - **Diversi tipi di legami** (archi)

Struttura delle reti

- **Struttura della rete**: presenza di un numero significativo di **cluster** (grafi completi o quasi completi)
- Presenza di alcuni **collegamenti tra cluster**

Legami di una rete

Cluster



Legami di una rete

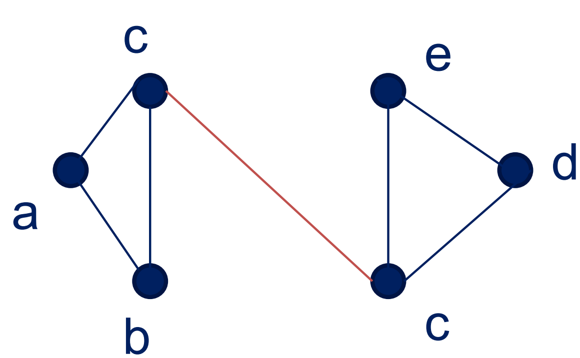
I legami possono essere di due tipi

- **Forti**: legami tra elementi di uno stesso cluster
- **Deboli**: legami tra elementi di cluster differenti

Legami deboli

I **legami deboli** hanno una grande importanza:

- **Mettono in comunicazione diversi gruppi**
- **Rimozione** → **scollegamento** della rete



Clusterizzazione

- Ulteriori studi (Strogatz e Watts , 1998) introdussero un indice per misurare la presenza di ***clusterizzazione*** (**coefficiente di clustering**)
- Il coefficiente di clustering misura quanto un grafo è **costituito da cluster**

Coefficiente di clustering locale

- Dato un nodo *i*, **clustering locale** di *i* misura quanto i nodi vicini ad *i* sono tra loro collegati
  - $C_i = 0$ nessuno dei vicini è collegato
  - $C_i = 1$ se inducono un grafo completo

Coefficiente di clustering locale

- Dato un nodo *i*, con grado k_i, il **coefficiente di clustering locale** di *i* è:

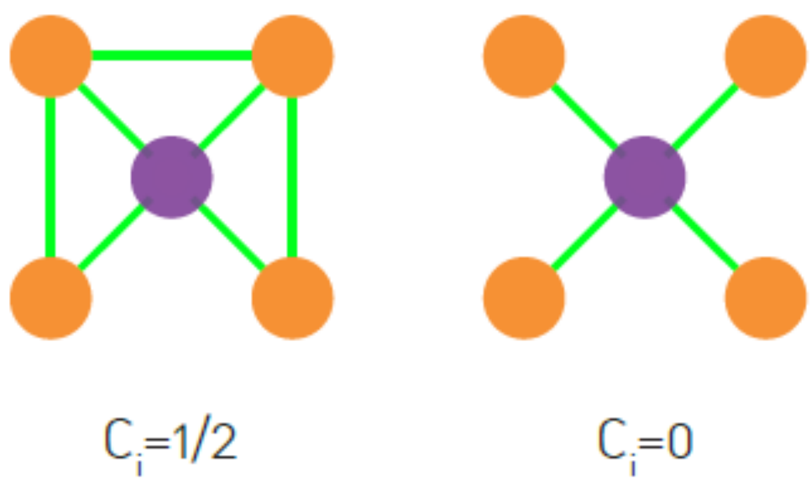
$$C_i = \frac{2E_i}{\mathrm{k_i}(\mathrm{k}_i - 1)}$$

dove $\boldsymbol{E_i}$: numero di archi tra vicini di *i*

- $C_i = 0$ se i vicini di *i* sono scollegati
- $C_i = 1$ se inducono un grafo completo

Coefficiente di clustering

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



Da A. L. Barabasi, *Network Science*

Coefficiente di clustering globale

- Il **coefficiente di clustering medio** di un grafo G è:

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

dove C_i rappresenta il coefficiente di clustering locale del nodo *i*

Coefficiente di clustering globale

- **Esercizio:** abbiamo un grafo con 4 nodi, dove alcuni nodi sono connessi tra loro:
- Nodo 1 ha un coefficiente di clustering locale di 0.5
- Nodo 2 ha un coefficiente di clustering locale di 0.3
- Nodo 3 ha un coefficiente di clustering locale di 0.0
- Nodo 4 ha un coefficiente di clustering locale di 1.0
- Il **coefficiente di clustering medio?**

Clusterizzazione

- Il modello osservato da Granovetter implica
  - Un **alto coefficiente di clustering**: presenza di un numero significativo di cluster
  - Legami **distribuiti non casualmente**: se due individui hanno un amico in comune è probabile che si conoscano

Clustering e reti casuali

- Consideriamo il coefficiente di clustering medio di una **rete casuale**
- In una rete casuale, con probabilità *p*, il numero di archi tra i vicini del nodo *i* è:

$$E_i = p\,\frac{k_i(k_i-1)}{2}$$

Clustering e reti casuali

- Numero di archi tra i vicini del nodo *i* è:

$$E_i = p\,\frac{k_i(k_i - 1)}{2}$$

- Il coefficiente di clustering (locale) è:

$$C_i = \frac{2E_i}{k_i(k_i - 1)} = p = \frac{\bar{k}}{N}$$

Numero di archi

Essendo

$$\bar{k} = p\,(N-1)$$

Il **grado medio** (la **densità**) aumenta al **crescere di *p***

Clustering e reti casuali

- Il **coefficiente di clustering** (locale) in una **rete casuale** è:

$$C_i = p = \frac{\bar{k}}{N-1}$$

- Possiamo dedurre:
  - A parità di grado medio, **più è vasta la rete e minore è il coefficiente di clustering**
  - Il coefficiente di clustering è **indipendente dal grado** di un nodo specifico

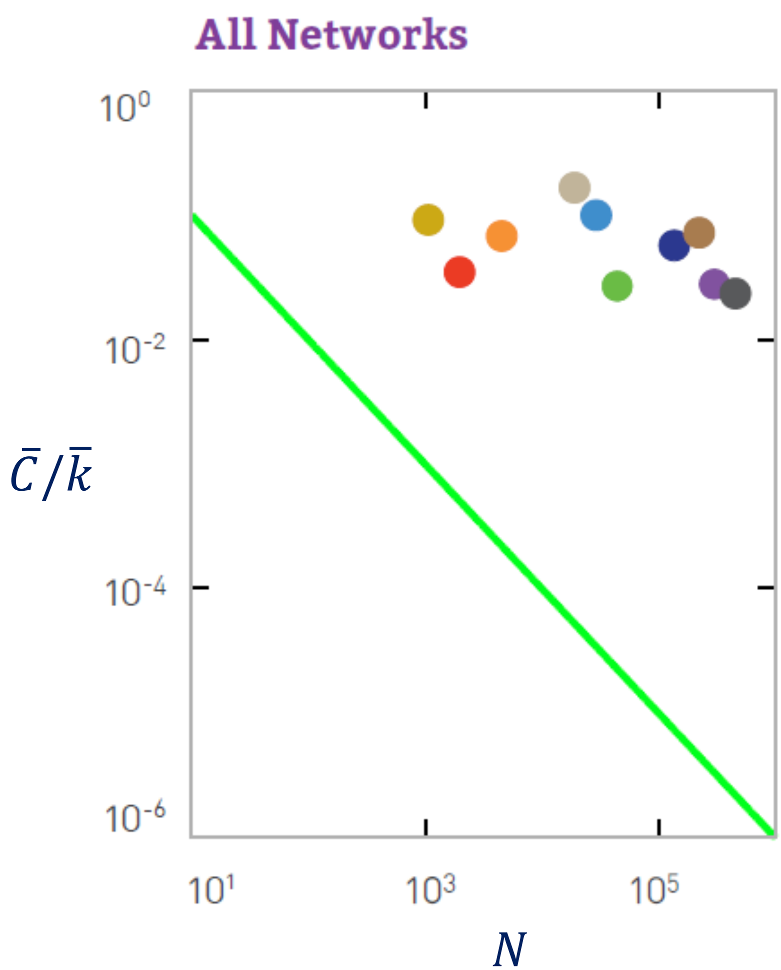
Clustering e reti casuali

- Il coefficiente di clustering (locale) è:

$$C_i = \frac{2E_i}{k_i(k_i - 1)} = p = \frac{\bar{k}}{N - 1}$$

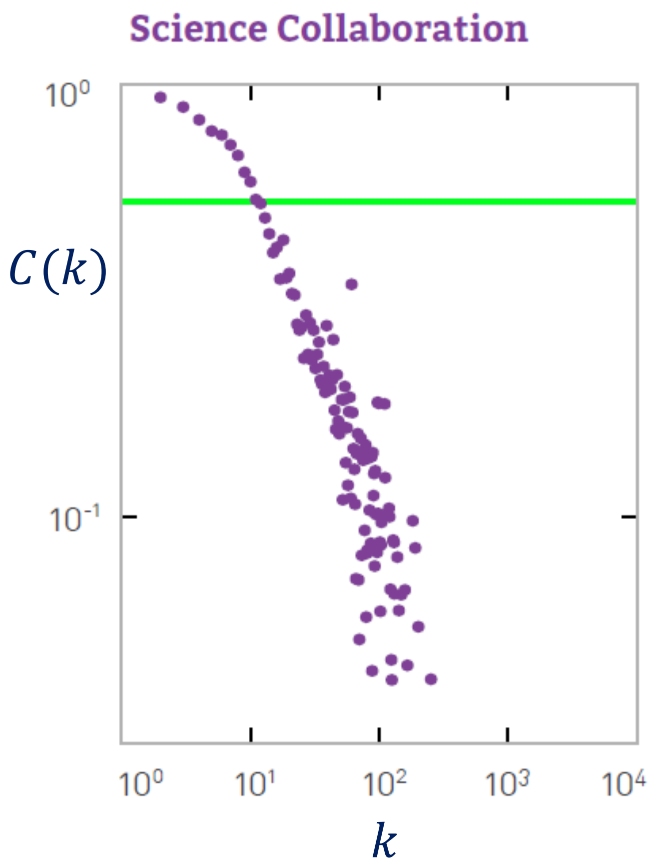
- Possiamo dedurre:
  - A parità di grado medio, **più è vasta la rete e minore è il coefficiente di clustering**
  - Il coefficiente di clustering è **indipendente dal grado** di un nodo specifico

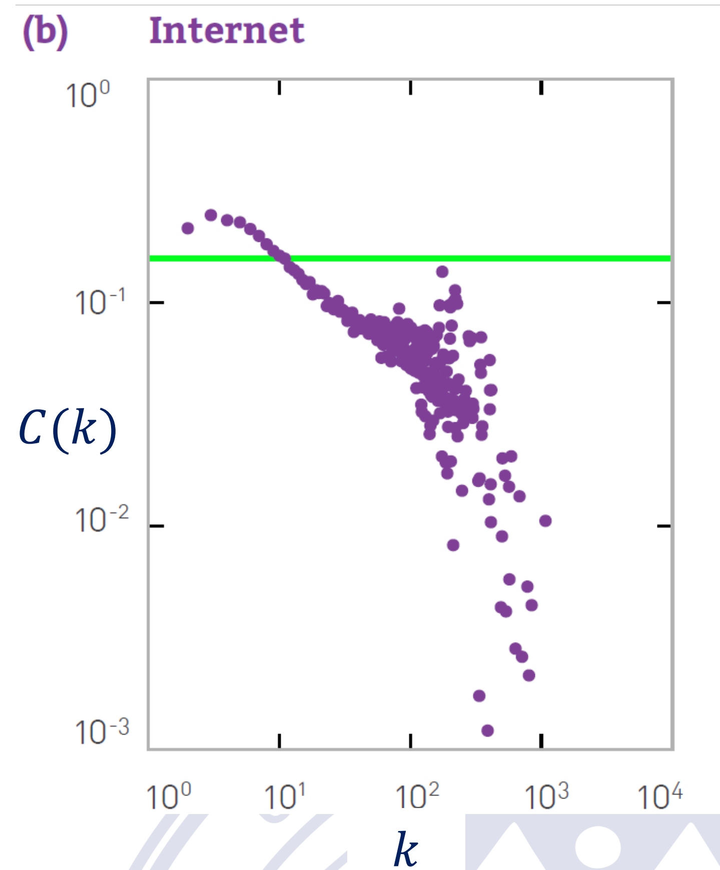
Clustering: reti reali



Da A. L. Barabasi, *Network Science*

Clustering: reti reali





Dipendenza dal grado

Da A. L. Barabasi, *Network Science*

Clustering

Diverse reti hanno un **alto grado di clustering**

- Reti di collaborazioni scientifiche
- Reti biologiche *(es. C. elegans)*
- World Wide Web
- Internet