Lezione n.15
ANALISI
DI RETI COMPLESSE
Materiale Didattico

Van Steen, GRAPH THEORY ANDCOMPLEX NETWORKS
Cap 1,2,6,7,8,9

Laura Ricci

19/04/2013

RETI COMPLESSE: ESEMPI

- Diversi fenomeni possono essere modellati mediante reti complesse, caratterizzate da un altissimo numero di nodi

- Social networks
  - relazioni di amicizia tra individui
  - relazioni d'affari tra compagnie
  - studio dei legami matrimoniali tra 16 famiglie fiorentine del XV secolo con lo scopo di analizzare attraverso quali canali relazionali e come i Medici divennero signori di Firenze
  - social networks online: Linkedin, MySpace, Facebook

- Information Networks
  - Rete (aciclica) di citazioni in articoli accademici
  - World Wide Web (rete ciclica)
  - .....

RETI COMPLESSE: ESEMPI

Biological Networks

- Food web: i nodi corrispondono alle specie di un ecosistema, gli archi descrivono le relazioni predatore-preda
- Neural networks

Technological Networks

- Internet: nodi corrispondono ai routers, gli archi alle connessioni fisiche esistenti tra i routers
- Reti P2P: nodi corrispondono ai peer, gli archi sono quelli dell'overlay network, oppure esprimono relazioni sociali tra i peer
- Reti costruite per la distribuzione di qualche risorsa (rete elettrica, reti stradali,…)

ANALISI DI RETI COMPLESSE P2P

Come utilizzeremo l'analisi delle rete complesse in questo corso?

- analisi di overlay P2P non strutturati (es. Gnutella) e strutturati (es: Chord): ricavare proprietà dell'overlay (grado medio dei nodi, diametro della rete,...) dall' analisi della rete
- reti sociali ricavate dalle interazioni di utenti in una rete P2P:
  - Interest Graphs:
    - si considerano gli utenti (peer) e le risorse accedute (files,...).
    - un arco connette due utenti se e solo se l'intersezione tra gli insiemi di risorse accedute da essi in un certo intervallo di tempo supera una certa soglia
    - analisi di Kazaa, BitTorrent
  - Exchange Graphs
    - nodi= utenti (peer).
    - connessioni tra due utenti se uno dei due ha fornito risorse all'altro
    - analisi di eMule, Gnutella

ANALISI DI RETI COMPLESSE P2P

- La topologia di overlay strutturati (es: DHT) può essere studiata ceonsiderando i vincoli imposti nella costruzione della DHT

- La struttura della rete sociale può essere ricavata generando una traccia degli accessi alle risorse da parte dei peer
  - Exchange Graph può essere ricavato
    - da un server eMule che tiene traccia delle query sottomesse e dei peer che possono fornire i file richiesti dalla query
    - da un superpeer di Kazaa che tiene traccia delle query inviate dai peer gestiti/delle risposte ricevute

  - Interest Graph può essere ricavato da un Internet Service Provider tiene traccia del traffico generato da Kazaa/BitTorrent
    - per ogni query inviata sulla rete dagli utenti dell'ISP, si traccia il peer che ha prodotto la query e le risorse richieste

RETI COMPLESSE: ANALISI

- Reti complesse: contengono milioni di nodi. La definizione di modelli appropriati per queste reti è necessaria per descriverne proprietà topologiche

- L'analisi "classica" di reti di piccole dimensioni non risulta più utile nel caso di reti complesse

- Esempio:
  - rete di piccola dimensione, proprietà interessante: "esiste un vertice che è indispensabile per mantenere la connettività della rete" ?
  - reti complesse, proprietà interessante: "che percentuale di nodi devo rimuovere per modificare in qualche misura la connettività della rete"?

- L'analisi di reti complesse richiede strumenti basati sull'analisi statistica delle proprietà della rete

RETI COMPLESSE: MODELLI

- Strumenti di base per analisi di reti complesse: Grafi Random (Erdos, Renyi, anni '50)
  - considerare un grafo di n nodi
  - connettere ogni coppia di nodi con probabilità p.
  - ogni connessione è indipendente dalle altre (distribuzione binomiale o di Poisson).

- Anni '90: La capacità di calcolo dei sistemi ha permesso un'analisi sperimentale della struttura di reti complesse reali con milioni di nodi

- Da questa analisi è risultato che il modello dei random graphs non ripecchia completamente la struttura di molti reti complesse, ma è una base per il loro studio

- Necessità di nuovi modelli

RETI COMPLESSE:PROPRIETA'

- Proprietà che caratterizzano una rete complessa
  - Lunghezza media dei cammini
  - Clustering
  - Distribuzione dei gradi dei nodi
  - Network Resilience

- Reti sociali caratterizzate da basso diametro ed alta clusterizzazione

- I grafi random possiedono solo alcune di queste proprietà

- Nuovi modelli proposti recentemente sono in grado di descrivere tutte le proprietà precedenti

ANALISI DI RETI COMPLESSE

- Analisi di reti complesse costituite da milioni di nodi e milioni di archi
  - overlay P2P, Reti sociali, Internet, Web,....
- In tutti i casi precedenti la rete può essere talmente grande e complessa che può diventare difficile scoprire le sue proprietà mediante la sua visualizzazione
- E' richiesto l'utilizzo di strumenti e metodi matematici
- Metriche base per l'analisi di una rete complessa
  - Considera la distribuzione dei gradi dei nodi: quanti sono i vertici con un alto grado rispetto a quelli di grado basso?
  - Analisi statistica delle distanze tra nodi: come sono posizionati i vertici nella rete: quanto è distante un vetice dagli altri, se occupa una posizione centrale nella rete,....
  - Clustering: fino a che punto i miei vicini sono adiacenti?
  - Centralità: esistono vertici che sono più importanti di altri?
  - Betweeness

MODELLI PER RETI COMPLESSE

- Problema: ricerca di un modello che riesca a descrivere in modo fedele il comportamento di una rete sociale

- Anni '90: il problema è stato affrontato approfonditamente anche per modellare reti complessi quali Internet, WWW e poi P2P

- Alcuni dei modelli proposti si sono dimostrati utili sia nel campo delle scienze sociali che in quello dell'informatica

- Modelli proposti
  - **Random Graphs:** modello semplice, ma non riesce a descrivere alcune proprietà interessanti, ad esempio le reti con alto fattore di clustering
  - **Watts-Strogatz:** small worlds + clusterizzazione
  - **Kleinberg:** utilizzato per definire overlays per reti P2P
  - **Barabasi-Albert:** scale free networks

GRAFI RANDOM

- Erdos-Renyi, nel 1950, iniziarono lo studio di grafi costruiti mediante l'aggiunta casuale di archi tra nodi
- Idea Base: si considera un grafo semplice, connesso, in cui, dato un numero fisso n di vertici, gli archi vengono creati in modo casuale, secondo una certa distribuzione di probabilità
- Si studiano proprietà statistiche, cioè proprietà che si verificano con una certa probabilità.
- Molti sistemi reali possono essere modellati mediante grafi random/varianti di grafi random
  - Alcune Reti P2P
  - Collaboration Networks
    - Rete di attori: due attori collegati se hanno recitato nello stesso film
  - Citation networks: Che citazioni ha un articolo?
  - Food webs: Chi mangia chi?
  - Sistemi Ferroviari,Aerei, Reti di computers : la probabilità di avere un arco tra due nodi dipende dalla distanza tra i due nodi.

GRAFI RANDOM: IL MODELLO DI ERDOS-RENYI

Modello di Erdos-Renyi:

- grafo non orientato $G_{n,p}$ **(ER(n,p))**, di n vertici.
- costruzione del grafo:
  - si considerano gli n vertici
  - i modi possibili di collegare gli n vertici tra di loro sono $\binom{n}{2}$
  - si considera ogni possibile arco (u,v) , (u≠v) tra i possibili $\binom{n}{2}$
    - in un ordine qualsiasi
    - in modo indipendente
    - si aggiunge l'arco considerato al grafo con probabilità p.

Osservazioni:

- ER(n,p) è un grafo semplice: non ci sono cicli e ci può essere al massimo un arco tra due vertici distinti

- In seguito faremo riferimento a questa definizione anche se ne esiste una alternativa, descritta successivamente

GRAFI RANDOM: IL MODELLO DI ERDOS-RENYI

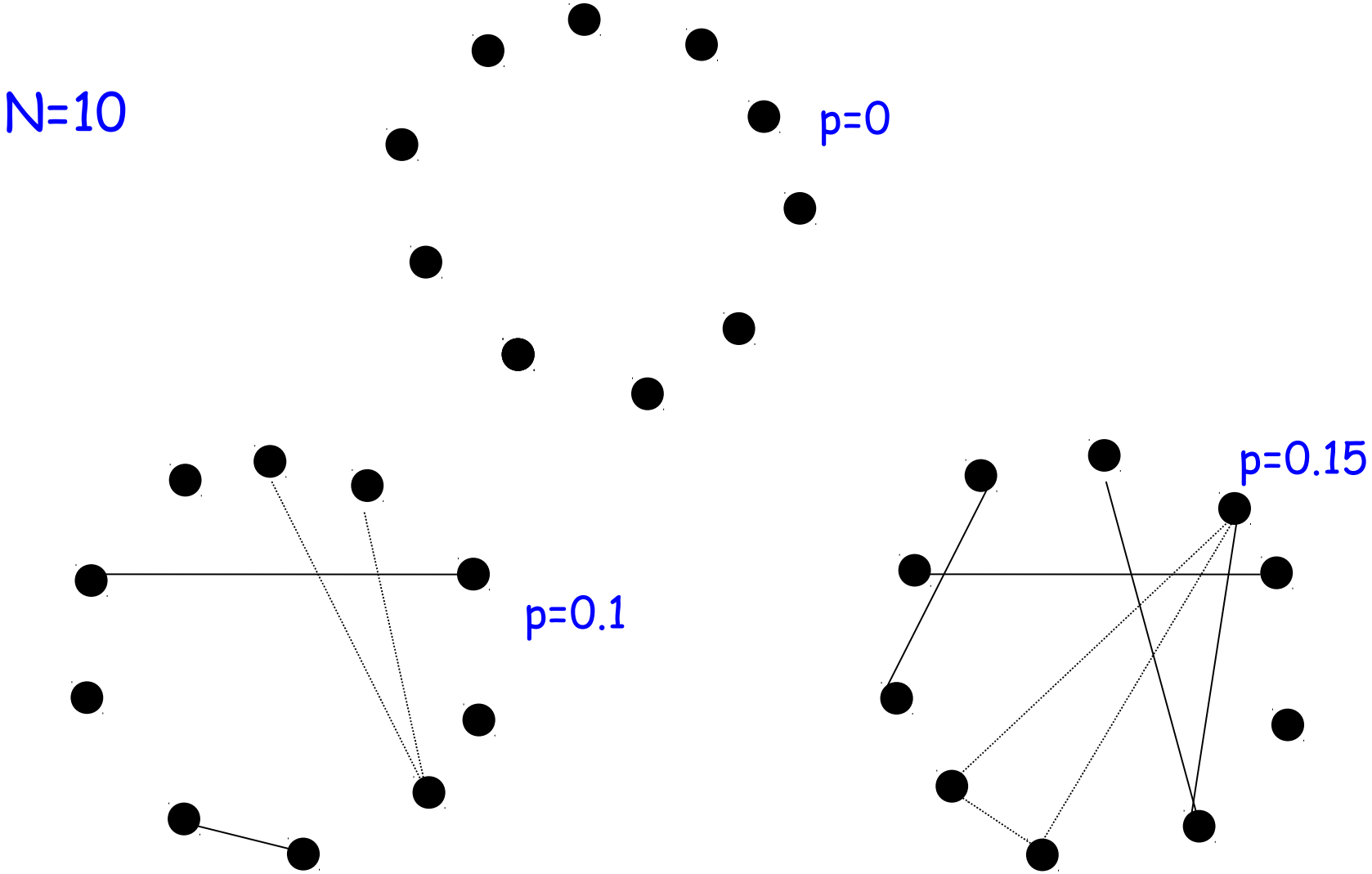
- Modello estremamente semplice:
  - si considerano n nodi
  - ogni coppia di nodi viene connessa da un arco con probabilità p

- Esempio di costruzione del grafo. p = $\frac{1}{4}$ = 0.25
  - si considera una coppia di nodi (u,v), si genera un numero casuale x, $0 \leq x \leq 1$,
  - se $x \leq p$, si definisce un arco tra u e v
  - altrimenti i nodi non vengono connessi

- La presenza/assenza di un arco è indipendente da quella degli altri archi

- L'esperimento rispetta lo schema di Bernoulli:
  - l' esecuzione ripetura di una data prova (definire o meno un arco) che può produrre due risultati (successo od insuccesso, definire o meno un certo arco).
  - le diverse prove sono indipendenti.

GRAFI RANDOM: IL MODELLO DI ERDOS RENYI



GRAFI RANDOM: IL MODELLO DI ERDOS-RENYI

Definizione Alternativa del Modello :

- $G_{n,N}(ER(n,N))$ è un grafo non orientato con n vertici ed esattamente N archi
- processo di costruzione del grafo:
  - inizio con un grafo che non contiene alcun arco
  - scelgo in modo uniforme uno dei possibili $\binom{n}{2}$ archi e lo aggiungo al grafo
  - termino quando ho aggiunto N archi

Notare la similarità tra grafi random e modello balls e bins:

- "scegliere" un arco nel modello $G_{n,N}$ è come lanciare due palle contemporaneamente in due bins

GRAFI RANDOM: ANALISI DELLA RETE

- Per ogni modello proposto, interessa studiare alcune proprietà dei grafi generati mediante quel modello
  - distribuzione dei gradi dei nodi
  - lunghezza media dei cammini sul grafo
  - coefficente di clusterizzazione
  - creazione di giant components
  - betweeness
  - centralità
  - .....

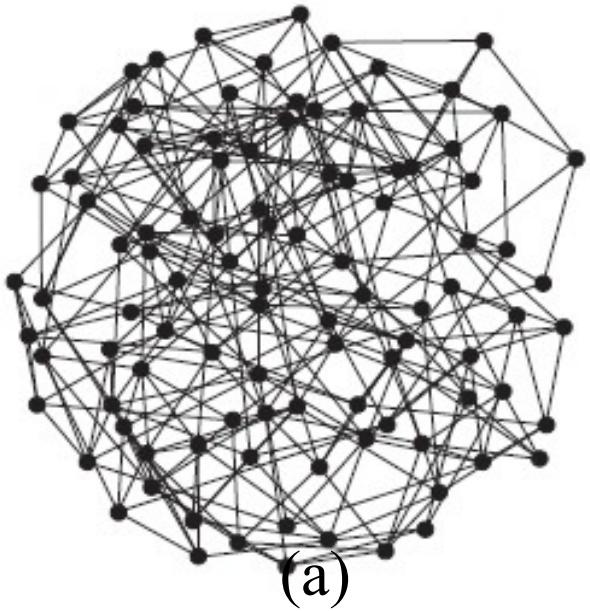
DISTRIBUZIONE DEI GRADI DEI NODI

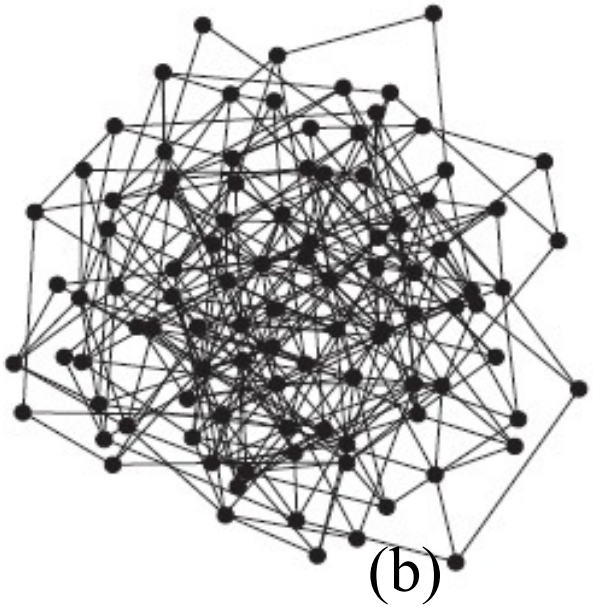
- Grado(degree) di v, $\delta(v)$: Numero di archi incidenti in v. Un ciclo viene contato due volte. Indegree, Outdegree per archi orientati

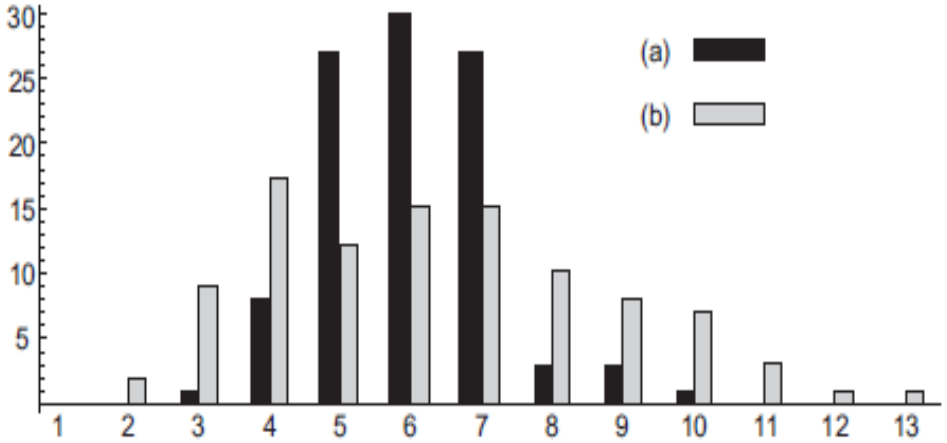
- L'analisi della distribuzione dei gradi dei nodi di un grafo può essere utilizzata per ottenere informazioni sulla struttura della rete, ad esempio:
  - se la maggior parte dei vertici ha lo stesso grado (distribuzione uniforme), i vertici hanno ruoli simili nella rete
  - se solo alcuni nodi sono caratterizzati da un grado alto, questi nodi svolgono la funzione di "hub". La loro rimozione può implicare il partizionamento della rete in diverse componenti

- Strumenti per l'analisi dei gradi dei nodi
  - Istogrammi – visualizzano il numero di vertici che hanno un certo grado
  - Degree sequences
  - Degree correlation

DISTRIBUZIONE DEI GRADI: ISTOGRAMMI







Numero di vertici

di (a) e di (b) = 100

GRAFI RANDOM: DISTRIBUZIONE DEI GRADI

- Consideriamo un grafo **ER(n,p)**. Per ogni vertice del grafo ci sono al più n-1 altri vertici con cui quel vertice può essere connesso
- Valutiamo la probabilità che il grado di un vertice sia k
  - ci sono massimo n-1 altri vertici che possono essere adiacenti ad u
  - dobbiamo scegliere tra gli n-1 vertici, k vertici da connettere ad u
  - ci sono $\binom{n-1}{k}$ modi di scegliere k vertici diversi da n-1 vertici
  - per ogni combinazione di k vertici, la probabilità di creare un arco con ognuno di essi è la seguente

$$p^k(1-p)^{n-1-k}$$

- Variabile casuale "grado di un nodo", $\delta(u)$: distribuzione bernoulliana
- la probabiltà che un nodo abbia grado k $P[\delta(u)=k]$ è

$$\mathbb{P}[\delta(u)=k]=\binom{n-1}{k}p^k(1-p)^{n-1-k}$$

GRAFI RANDOM: DISTRIBUZIONE DEI GRADI

- Conclusione: La variabile δ che rappresenta il grado di nodi del grafo presenta una distribuzione binomiale

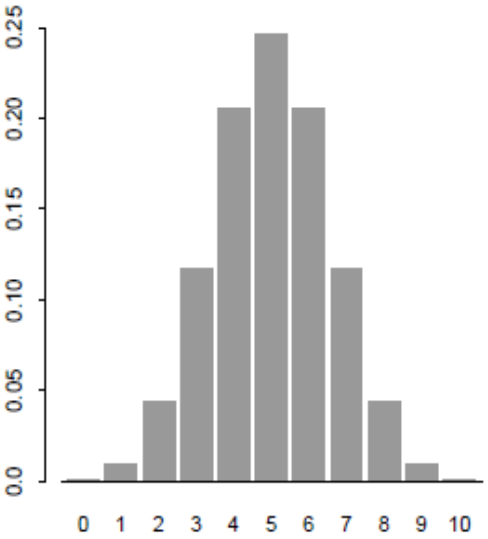
- Il grado medio dei nodi di un grafo ER(n,p) è uguale a p(n-1) (valore atteso di una variabile binomiale)
  - ad esempio, il grado medio dei vertici di un grafo ER(100,$\frac{1}{4}$) è 25

- Come si distribuiscono i valori della variabile casuale intorno alla media?

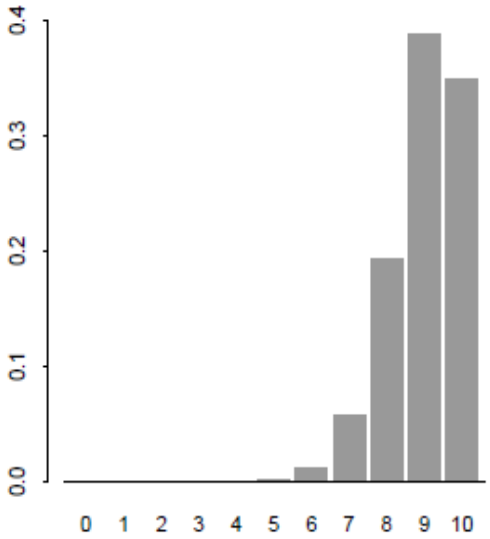
- La "forma" della distribuzione binomiale dipende dai valori di n e di p
  - per valori di n piccoli distribuzione 'simmetrica' per valori di p vicini a 0.5, distribuzione molto 'distorta' per valori di p vicini a 0 oppure ad 1
  - per valori di n grandi la distribuzione diventa sempre 'più simmetrica' forma a campana) e per apprezzare una piccola distorzione si devono considerare probabilità molto vicine allo 0 oppure all'1.

DISTRIBUZIONE BINOMIALE

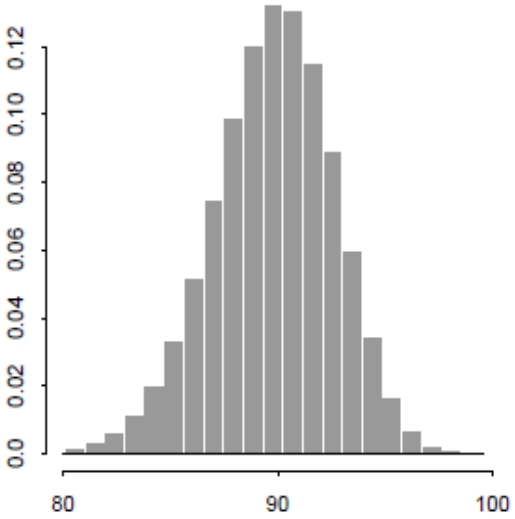
$n = 10, p = 0.5$



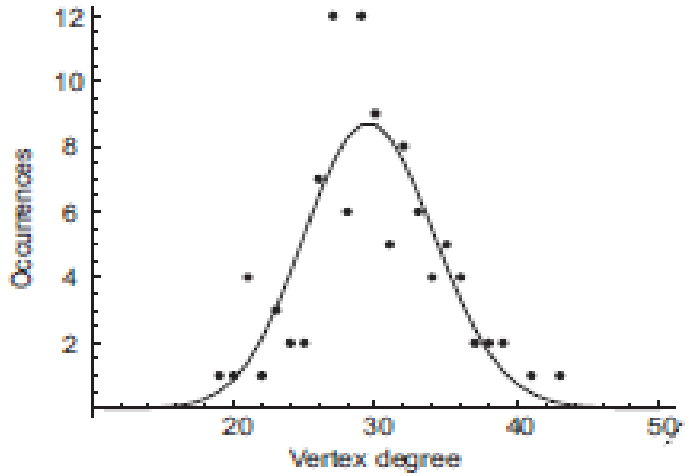
$n = 10, p = 0.9$



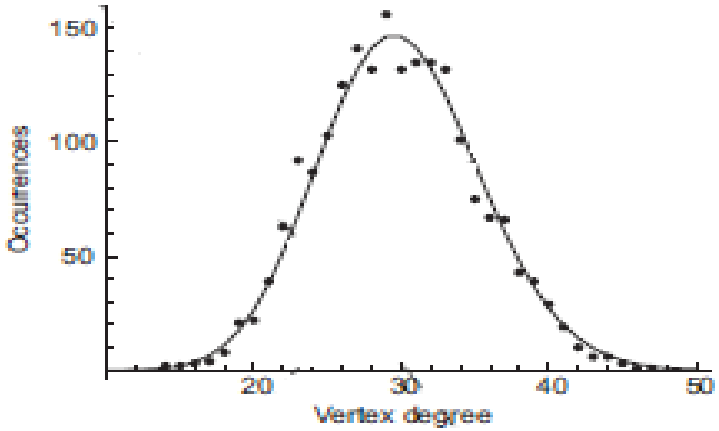
$n = 100, p = 0.9$



GRAFI RANDOM: DISTRIBUZIONE DEI GRADI



Distribuzione dei gradi dei nodi per ER(100, 0.3), grado medio =30



Distribuzione dei gradi dei nodi per ER(2000,0.015), grado medio=30

Se si considera un numero maggiore di vertici, la distribuzione diventa sempre più 'simmetrica', intorno alla media

A noi interessano reti con altissimo numero di nodi: in questo caso la varianza risulta molto bassa

DISTRIBUZIONE DI POISSON

- se consideriamo valori di n tendenti ad infinito e manteniamo il grado medio costante (p(n-1)) si ottiene la distribuzione di Poisson

- distribuzione di Poisson:
  - eventi indipendenti che accadono in un intervallo di tempo fissato o in uno spazio fissato
  - la frequenza media λ degli eventi in un dato intervallo di tempo è costante

- esempio: si consideri il numero annuale di incidenti su strada che avvengono in un anno in Nuova Zelanda
  - gli incidenti sono indipendenti
  - la frequenza media degli incidenti è costante per ogni anno

- conoscendo il valore della media, la distribuzione di Poisson indica la probabilità di scostarsi dal valor medio.

DISTRIBUZIONE POISSONIANA

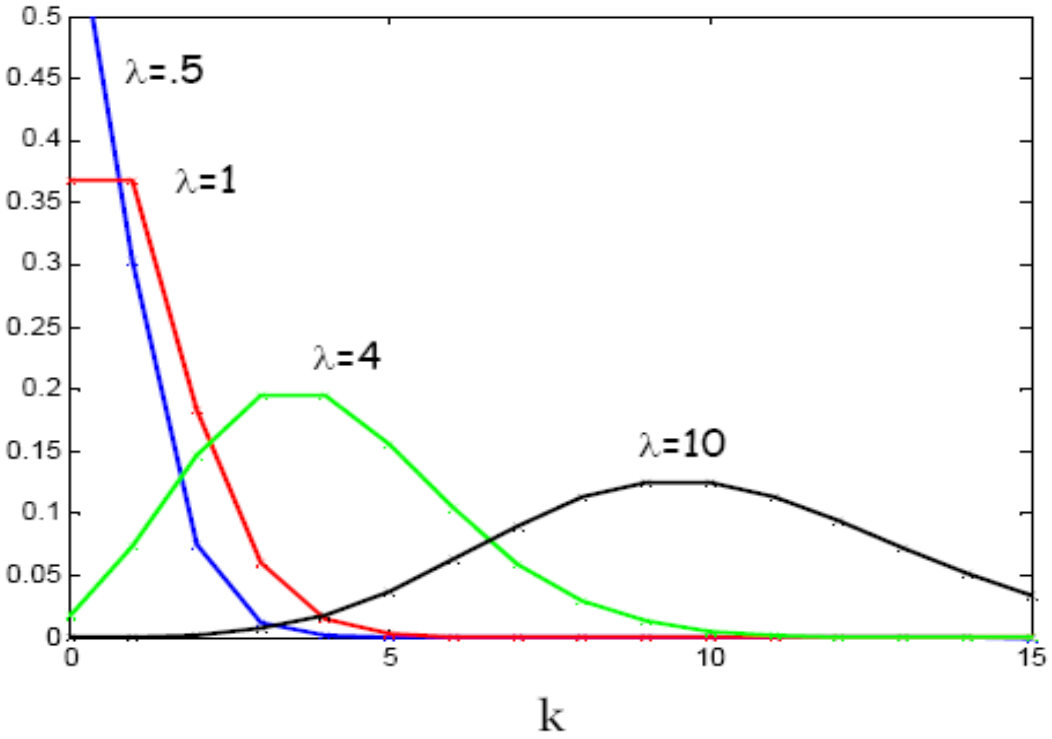
$$X \sim \text{Poisson}(\lambda)$$

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

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

- 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

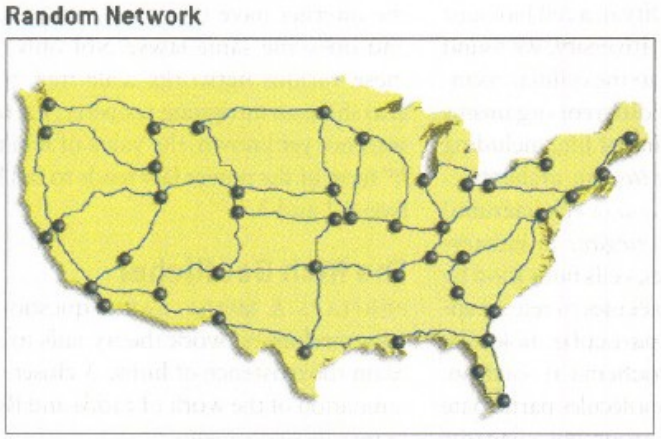
DISTRIBUZIONE POISSONIANA

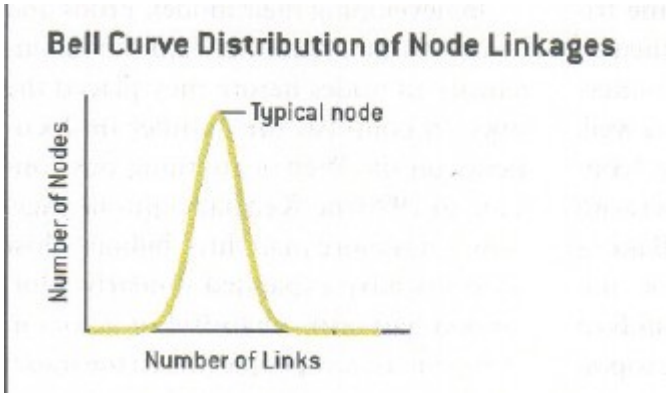


La forma della distribuzione dipende dal valore di λ
Al crescere di λ la distribuzione diventa sempre più simmetrica fino ad avere una forma 'a campana' per valori molto alti di λ

DISTRIBUZIONE POISSONIANA

- Random Graphs (Erdos-Renyi):
  - 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 con distribuzione dei gradi dei nodi poissoniano: sistema autostradale americano





DISTANZA TRA NODI

Si consideri un grafo connesso,

- d(u,v), distanza tra i nodi d ed u del grafo = lunghezza del cammino minimo tra u e v
- Eccentricità di un nodo u $\varepsilon(u)$: $\max \{d(u,v) | v \in V(G)\}$

  indica quanto è distante da u il vertice più distante esistente nella rete
- Raggio del grafo rad(G) : $\min \{\varepsilon(u) | u \in V(G)\}$

  indicazione di quanto sono sparsi i nodi nella rete
- Diametro del grafo diam(G) : $\max\{d(u,v) | u,v \in V(G)\}$

  indica la massima distanza all'interno della rete

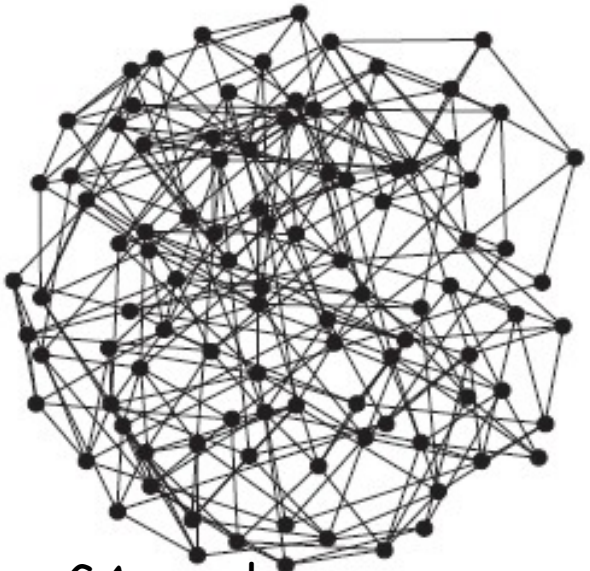
$$\bar{d}(u) = \frac{1}{|V|-1} \sum_{v \in V, v \neq u} d(u,v)$$

Lunghezza media dei cammini minimi da u ad ogni altro vertice v in G

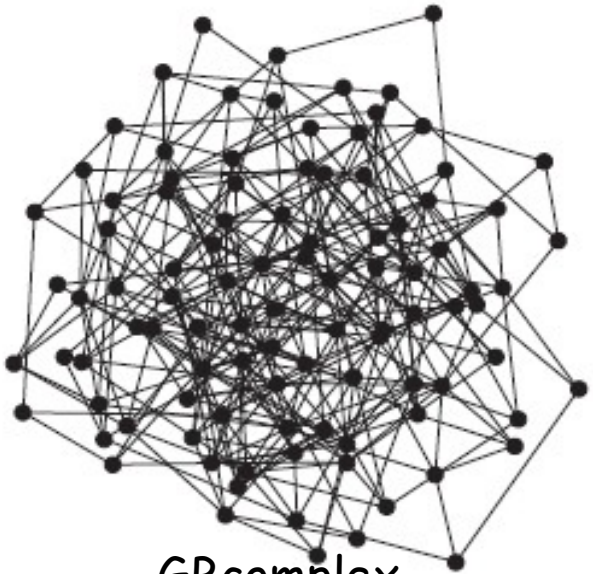
$$\bar{d}(G) = \frac{1}{|V|} \sum_{u \in V} \bar{d}(u) = \frac{1}{|V|^2 - |V|} \sum_{u,v \in V, u \neq v} d(u,v)$$

Lunghezza media dei cammini del grafo

DISTANZA TRA NODI



GAcomplex



GBcomplex

Metric	$GA_{complex}$	$GB_{complex}$
Average eccentricity	4.59	4.09
Radius	4	3
Diameter	6	5
Average path length	2.96	2.67
Characteristic path length	2.95	2.63

GRAFI RANDOM: IL DIAMETRO

- Sappiamo che il grado medio dei nodi è K
- Dato un nodo u il numero medio di nodi a distanza d da u è K^d
- Quindi, per raggiungere da u un qualsiasi nodo di una rete di n nodi saranno necessari L passi, dove $K^L = n$
- Quindi il diametro della rete può essere calcolato come segue
  $\log K^L = \log n \Rightarrow L * \log K = \log n \Rightarrow L = \log n / \log K$
- Il "grado di separazione" di un grafo random cresce in modo logaritmico con il numero di nodi
  - Grafi random caratterizzati da un diametro molto basso

GRAFI RANDOM

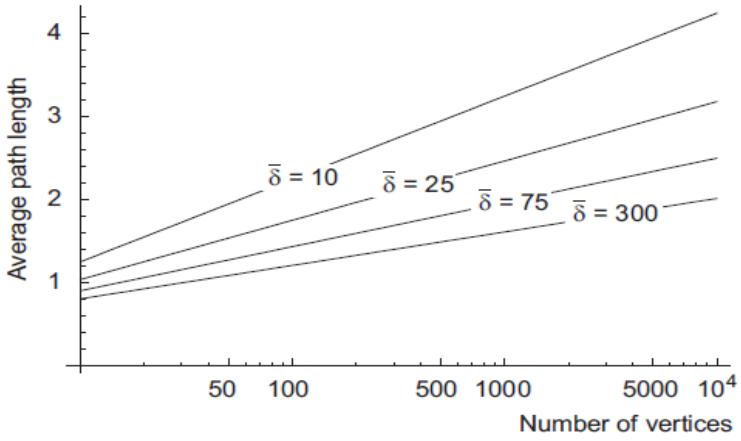
- Per ogni H ∈ ER(n,p) la lunghezza media dei cammini è

$$\bar{d}(H) = \frac{\ln(n) - \gamma}{\ln(pn)} + 0.5$$

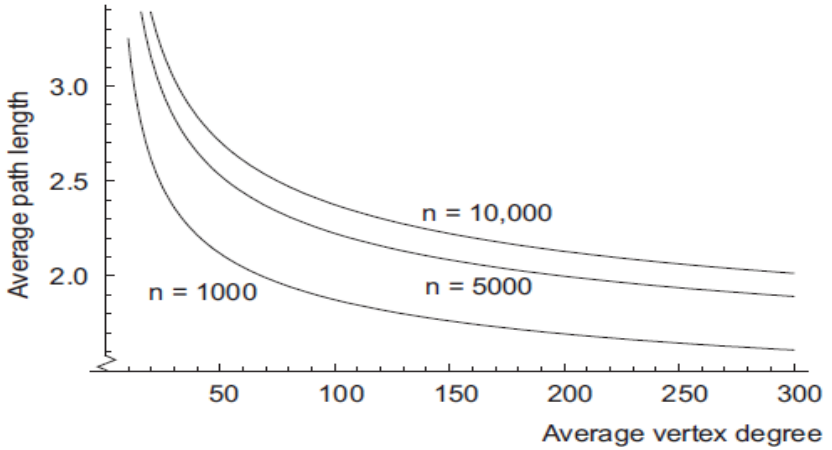
dove γ è la costante di Eulero ($g \approx 0.57772$)

- Poichè δ=p(n-1), si ottiene

$$\bar{d}(H) \approx \frac{\ln(n) - \gamma}{\ln(\bar{\delta})} + 0.5$$



Grado δ fisso, dimensione variabile



Dimensione fissa, grado variabile

COEFFICENTE DI CLUSTERIZZAZIONE

- Coefficente di clusterizzazione indica, per un certo vertice v, fino a che punto i vicini di v sono anche vicini tra di loro
  - i miei amici sono amici tra di loro
  - l'amico del mio amico è mio amico

- In termini del grafo:
  - se un nodo ha due vicini esista una connessione anche tra di loro
  - transitività
    - se A è connesso a B e B è connesso a C, esiste un'alta probabilità che A risulti connesso a C

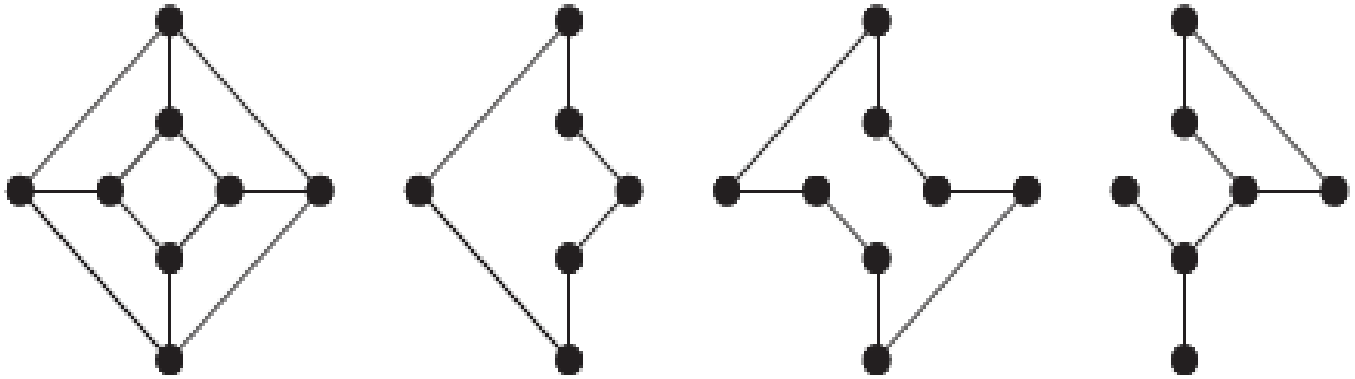
- la topologia della rete è caratterizzata dalla presenza di molti triangoli

- diverse definizioni di coefficente di clusterizzazione, definite secondo
  - la visione locale di un vertice
  - la visione globale dell'intero grafo

COEFFICENTE DI CLUSTERIZZAZIONE

- La reti reali sono spesso rappresentate come un insieme di nodi interconnessi ed è spesso possibile individuare comunità di nodi
- diversi links tra i nodi membri e pochi links tra le diverse comunità
- l'esistenza delle comunità può essere misurata mediante il coefficente di clusterizzazione
- Il coefficente di clusterizzazione può essere utilizzato anche per misurare la velocità con cui si diffonde l'informazione in una rete
  - una misura utile per lo sviluppo di algoritmi di diffusione dell'informazione in reti P2P
    - algoritmi di *gossiping*/ disseminazione epidemica dell'informazione, basati sulla selezione casuale dei vicine
    - definizione di protocolli di disseminazione dell'informazione in teri altamente clusterizzate

RICHIAMI DI TEORIA DEI GRAFI



- H è un sottografo di G se $V(H) \subseteq V(G)$ e $E(H) \subseteq E(G)$ tale che $\forall e \in E(H)$, $e = \langle u,v \rangle$, : $u,v \in V(H)$
- Dato un grafo G ed un insieme di vertici $V^* \subseteq V(G)$, il sottografo indotto da V^* include l'insieme di vertici V^* ed l'insieme di archi E^* tali che

$$E^* = \{e \in E(G) \mid e = \langle u,v \rangle , \ u,v \in V^*\}$$

- Dato un grafo G ed un insieme di archi $E^* \subseteq E(G)$, il sottografo indotto da E^* ha l'insieme di archi E^* ed un insieme di vertici V^* definito come

$$V^* = \{\mathbf{u,v} \in V(G) \mid \exists\, e = E^* : e = \langle u,v \rangle \}$$

COEFFICENTE DI CLUSTERIZZAZIONE LOCALE

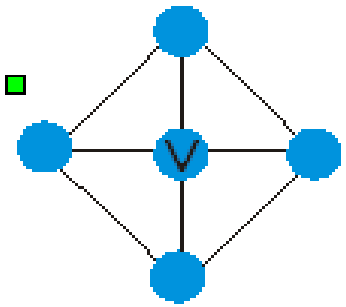
- Si consideri un grafo G semplice, connesso, non orientato. Sia
  - v un vertice $\in$ G
  - N(v) l'insieme di vicini di v.
  - $n_v = |N(v)|$
  - il massimo numero di archi esistenti tra i vicini di v è quindi $\binom{n_v}{2}$
  - sia m_v il numero di archi del sottografo indotto da N(v): $m_v = |E(G(N(v)))|$
- Il coefficente di clusterizzazione cc(v) del vertice v è definito da:

$$cc(v) = \begin{cases} m_v / \binom{n_v}{2} = \frac{2 \cdot m_v}{n_v(n_v-1)} & \text{if } \delta(v) > 1 \\ \text{undefined} & \text{otherwise} \end{cases}$$

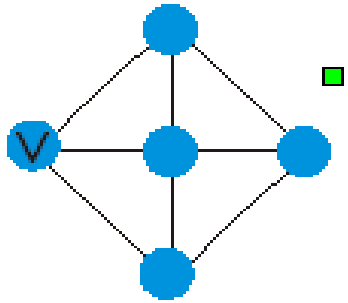
- Sia V* ={v$\in$G |δ(v)>1}. Il coefficente di clusterizzazione CC(G) è definito come:

$$CC(G) = \frac{1}{|V^*|} \sum_{v \in V^*} cc(v)$$

COEFFICENTE DI CLUSTERIZZAZIONE LOCALE



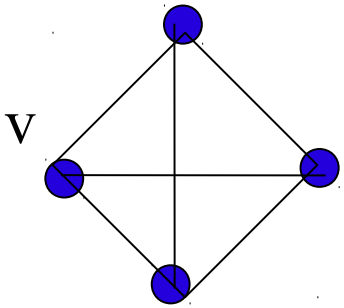
$$\frac{2 \cdot 4}{4 \cdot 3} = 2/3$$



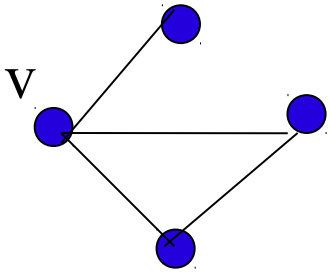
$$\frac{2 \cdot 2}{3 \cdot 2} = 2/3$$

- C(v) misura quanto i vicini di v formano strutture 'di tipo clique'
- Clusterizzazione massima: tutti i miei vicini sono adiacenti tra di loro: i vicini formano un grafo completo

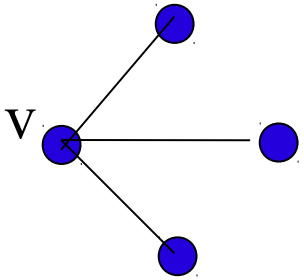
COEFFICENTE DI CLUSTERIZZAZIONE LOCALE



$C = 1$



$C = 1/3$



$C = 0$

COEFFICENTE DI CLUSTERIZZAZIONE GLOBALE

Dato un grafo *G* semplice e non orientato ed un vertice $v \in V(G)$,

- un triangolo in v è un sottografo completo di *G* che include esattamente tre vertici, che includono v.
- una tripla centrata in v è un sottografo di esattamente tre vertici e due archi, dove i due archi incidono in v
- Notazione:
  - $n_\Lambda(v)$ numero di triple in v
  - $n_\Delta(v)$ numero di triangoli in v
  - $n_\Lambda(G)$ numero totale di triple nel grafo *G*
  - $n_\Delta(G)$ numero totale di triangoli nel grafo *G*
- Coefficente di clusterizzazione o Network Transitivity $\tau(G)$: Dato un grafo *G* semplice, connesso con $n_\Delta(v)$ triangoli distinti e $n_\Lambda(v)$ triple distinte,

$$\tau(G) = 3 * n_\Delta(G) / n_\Lambda(G)$$

- Talvolta viene omesso il fattore 3

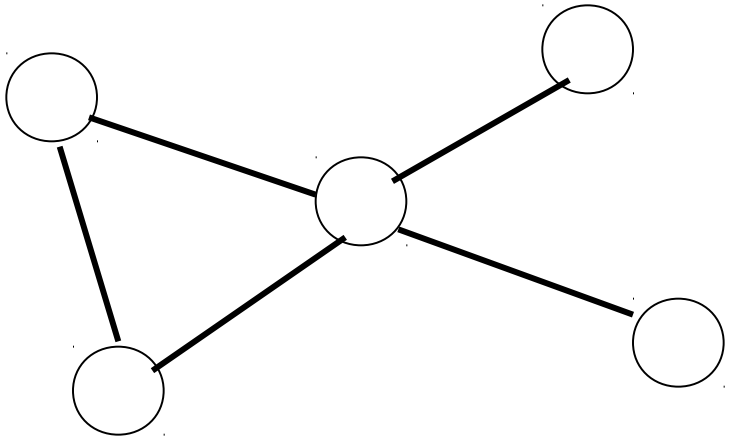
COEFFICENTE DI CLUSTERIZZAZIONE GLOBALE

- La definizione precedente viene talvolta ristretta a cliques contenenti tre vertici (triangolo)
- Tripla connessa ad un vertice v = v + due vertici ad esso connessi
- Coefficiente di clusterizzazione di un grafo

  C= (3*numero di triangoli nella rete) / numero di triple connesse

  fattore 3 misura il fatto che ogni triangolo contribuisce a 3 triple connesse
- C ($0 \leq C \leq 1$) misura la frazione di triple connesse che formano triangoli
- Un alto livello di clusterizzazione implica la presenza di molti "triangoli" sulla rete
- Clustering= transitività, se un vertice A è connesso ad un vertice B ed il vertice B è connesso al vertice C, è probabile che A risulti connesso a C

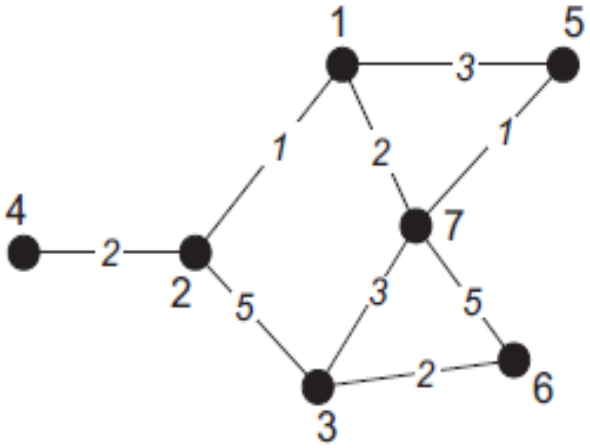
COEFFICENTE DI CLUSTERIZZAZIONE GLOBALE



La rete è caratterizzata da

- un unico triangolo
- 8 triple
- coefficiente di clusterizzazione $C = 3 * 1/8 = 3/8$

COEFFICENTE DI CLUSTERIZZAZIONE GLOBALE VS. LOCALE

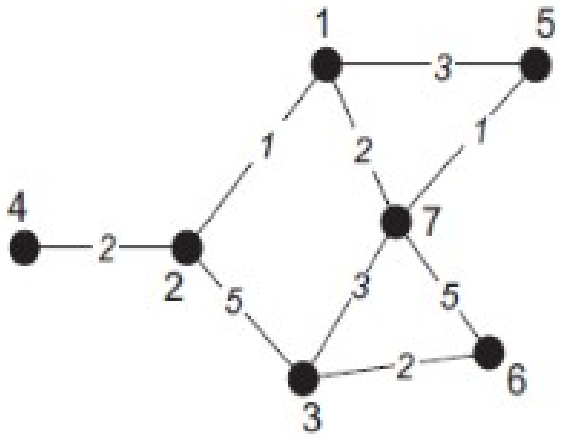


Vertex:	1	2	3	4	5	6	7
cc:	1/3	0	1/3	*undefined*	1	1	1/3
n_Λ:	3	3	3	0	1	1	6

Vertex 1 $N(1) = \{2,5,7\}; E(G[N(1)]) = \langle 5,7 \rangle \Rightarrow cc(1) = \frac{1}{3}$

Triples at 1: $G[\{2,1,5\}], G[\{2,1,7\}], G[\{5,1,7\}]$

COEFFICENTE DI CLUSTERIZZAZIONE GLOBALE VS. LOCALE



Vertex:	1	2	3	4	5	6	7
cc:	1/3	0	1/3	*undefined*	1	1	1/3
n_Λ:	3	3	3	0	1	1	6

$CC_{loc}(G) = 3/6$

$\tau(G) = 2/17$

GRAFI RANDOM: CLUSTERIZZAZIONE

Coefficente di clusterizzazione: rapporto tra il numero di archi esistenti tra i vicini ed il numero massimo di archi tra i vicini

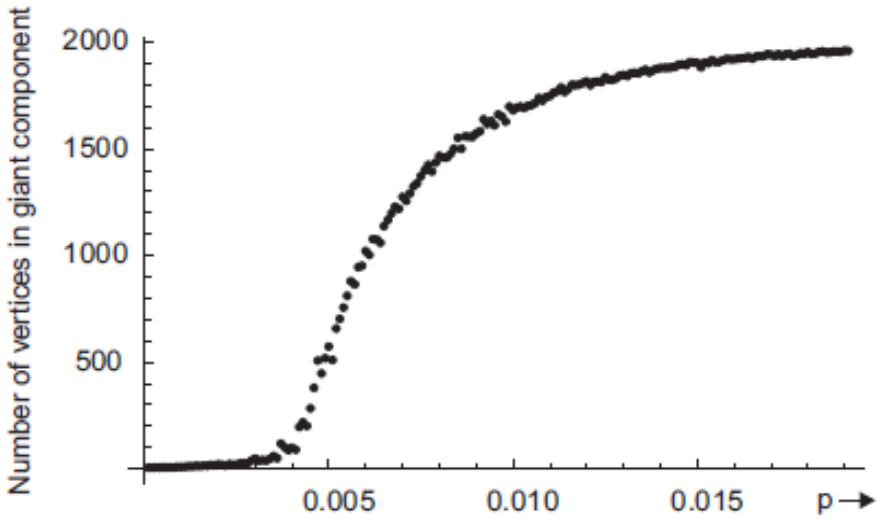
- Ogni coppia di vicini ha una probabilità p di essere connessa da un arco

- Numero medio di archi tra i k vicini $\binom{k}{2}p$ massimo numero di archi tra i k vicini $\binom{k}{2}$
- Il coefficente di clusterizzazione è dato dal rapporto tra $\binom{k}{2}p$ e $\binom{k}{2}$

Il coefficente di clusterizzazione di un grafo random caratterizzato da una probabilità p è uguale a p

Coefficiente di clusterizzazione basso: la probabilità di connettere due vertici è p indipendemente dal fatto chei due vertici possiedano un vicino comune

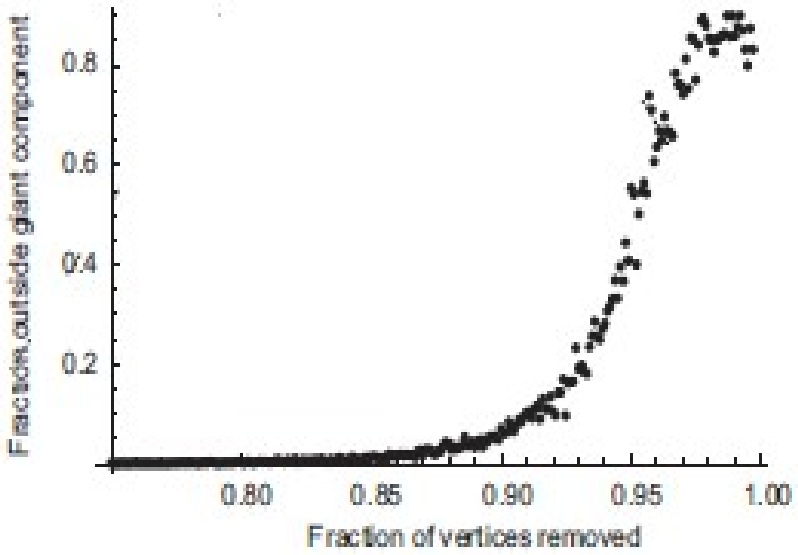
GRAFI RANDOM: CONNETTIVITA'



ER(2000,p)

- Analisi della connettività del grafo: quante componenti connesse contiene il grafo?
- all'aumentare della probabilità p
  - aumenta la densità del grafo
  - la maggior parte dei vertici sono contenuti in una unica componente, detta giant component,
  - i vertici restanti sono sparsi in un numero basso di componenti

GRAFI RANDOM: CONNETTIVITA'



ER(2000,0.015)

- Asse X: numero di vertici rimossi dal grafo
- Asse Y: numero di vertici non appartenenti alla 'giant component'
- Conclusioni: se rimuovo il 95% dei vertici, metà dei vertici del grafo rimangono comunque all'interno della 'giant component'

GRAFI RANDOM: CLUSTERIZZAZIONE

Conclusioni: in un random graph

- Il grado medio di un vertice dipende dalla probabilità p e dal numero di vertici del grafo.
- In media tutti i nodi hanno lo stesso grado
- Il diametro della rete è basso
- Il coefficiente di clusterizzazione è basso e quindi hanno poca capacità di modellare l'aggregazione, caratteristica tipica di molte reti reali
- il basso coefficiente di clusterizzazione consente di ottenere bassi gradi di separazione tra i nodi

LE RETI REALI POSSONO ESSERE MODELLATE COME GRAFI RANDOM?

Network	Size	$\langle k \rangle$	ℓ	ℓ_{rand}	C	C_{rand}	Reference	Nr.
WWW, site level, undir.	153 127	35.21	3.1	3.35	0.1078	0.00023	Adamic, 1999	1
Internet, domain level	3015–6209	3.52–4.11	3.7–3.76	6.36–6.18	0.18–0.3	0.001	Yook *et al.*, 2001a, Pastor-Satorras *et al.*, 2001	2
Movie actors	225 226	61	3.65	2.99	0.79	0.00027	Watts and Strogatz, 1998	3
LANL co-authorship	52 909	9.7	5.9	4.79	0.43	1.8×10^{-4}	Newman, 2001a, 2001b, 2001c	4
MEDLINE co-authorship	1 520 251	18.1	4.6	4.91	0.066	1.1×10^{-5}	Newman, 2001a, 2001b, 2001c	5
SPIRES co-authorship	56 627	173	4.0	2.12	0.726	0.003	Newman, 2001a, 2001b, 2001c	6
NCSTRL co-authorship	11 994	3.59	9.7	7.34	0.496	3×10^{-4}	Newman, 2001a, 2001b, 2001c	7
Math. co-authorship	70 975	3.9	9.5	8.2	0.59	5.4×10^{-5}	Barabási *et al.*, 2001	8
Neurosci. co-authorship	209 293	11.5	6	5.01	0.76	5.5×10^{-5}	Barabási *et al.*, 2001	9
E. coli, substrate graph	282	7.35	2.9	3.04	0.32	0.026	Wagner and Fell, 2000	10
E. coli, reaction graph	315	28.3	2.62	1.98	0.59	0.09	Wagner and Fell, 2000	11
Ythan estuary food web	134	8.7	2.43	2.26	0.22	0.06	Montoya and Solé, 2000	12
Silwood Park food web	154	4.75	3.40	3.23	0.15	0.03	Montoya and Solé, 2000	13
Words, co-occurrence	460.902	70.13	2.67	3.03	0.437	0.0001	Ferrer i Cancho and Solé, 2001	14
Words, synonyms	22 311	13.48	4.5	3.84	0.7	0.0006	Yook *et al.*, 2001b	15
Power grid	4941	2.67	18.7	12.4	0.08	0.005	Watts and Strogatz, 1998	16
C. Elegans	282	14	2.65	2.25	0.28	0.05	Watts and Strogatz, 1998	17

LE RETI REALI POSSONO ESSERE MODELLATE COME GRAFI RANDOM?

Nella figura presentata nel lucido precedente

- size = dimensione della rete
- <k> = grado medio dei nodi
- l = lunghezza media del cammino tra due nodi
- C = coefficiente di clusterizzazione
- l_{rand} = lunghezza media del cammino tra due nodi di un grafo random della stessa dimensione, in cui il grado medio dei nodi è lo stesso
- C_{rand} = coefficiente di clusterizzazione medio di un grafo random della stessa dimensione, in cui il grado medio dei nodi è lo stesso

RETI COMPLESSE: SMALL WORLD

- Small Worlds. Cercare un fondamento scientifico per situazioni come la seguente, che riporta il dialogo tra due amici ed evidenzia una proprietà delle social networks
  - *ciao, mi sono trasferito a Lucca*
  - *Ah sì, ma allora forse conosci Mario Rossi?*
  - *Si, lo conosco. Certo che il mondo è piccolo....*

Small World:

- definiamo la distanza tra due nodi come il numero di archi che appartengono al cammino più breve che li collega
- nella maggior parte delle reti si osserva che la distanza tra due nodi qualsiasi delle rete è relativamente piccola rispetto alle dimensioni della rete

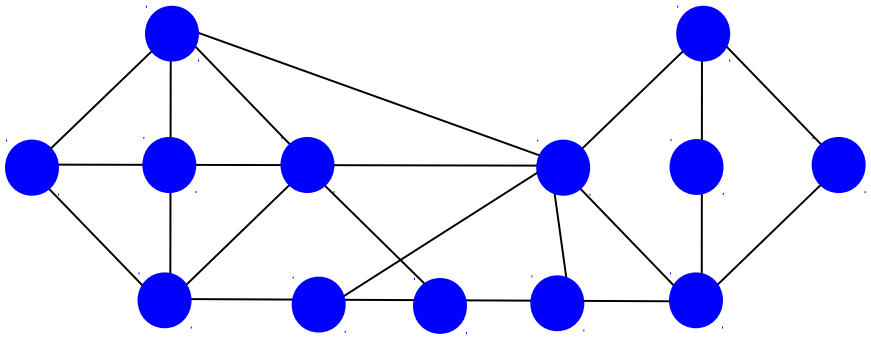
SMALL WORLD SOCIAL NETWORKS

La proprietà di "small world" è stata osservata inizialmente nelle rete sociali

- relazioni di amicizia
- relazioni commerciali tra compagnie
- chiamate telefoniche
- collaboration networks:
  - film collaboration network: descrive gli attori che hanno recitato insieme in almeno un file (internet movie database)
  - coauthorship network: vertici, autori di pubblicazioni scientifiche, archi congiungono due individui se e solo se sono stati coautori in qualche lavoro

SOCIAL NETWORKS: STRUTTURA

- Come può essere descritta la struttura di una social network?
- Intuizione: ogni persona ha un insieme di conoscenze generalmente riguardanti le persone vicine, es: i vicini di casa, i colleghi, i membri della squadra in cui gioca
- La rete risultante dovrebbe avere una struttura "a griglia"



- Se il modello descrive correttamente la realtà, il diametro di una social network di n nodi cresce come $O(\sqrt{n})$.
- I risultati sperimentali dimostrano che questa intuizione non è corretta

SOCIAL NETWORKS: L'ESPERIMENTO DI MILGRAM

Nel 1960 il sociologo Stanley Milgram (Harvald) condusse una serie di esperimenti per analizzare la struttura di una social network

- ad alcune persone (circa 160) scelte in modo casuale in Omaha, Nebraska & Wichita, Kansas fu chiesto di consegnare una lettera ad un operatore di borsa a Boston, di cui era noto il nome.
- ogni persona conosceva solamente queste informazioni sul destinatario della lettera
- fu chiesto ad ognuno di non usare l'indirizzo del destinatario, ma di consegnare la lettera solo a conoscenti diretti
- ogni persona doveva consegnare la lettera solo a persone direttamente conosciute che riteneva avere qualche punto di contatto con il destinatario della lettera

L'ESPERIMENTO DI MILGRAM: I RISULTATI

- Milgram calcolò il numero medio di "passaggi di mano" di ogni lettera che aveva raggiunto il destinatario
- 6 degree of separation: ogni lettera arrivata a destinazione aveva richiesto non più di 6 passaggi
- Questo valore risulta inferiore al valore $O(\sqrt{n})$ che misura il diametro di una social network con struttura "grid like"
- L'esperimento dimostrò che il diametro della social network analizzata risultava molto piccolo, nonostante la località della rete (l'alto grado di clusterizzazione della rete)
- Conclusione: il modello "a griglia" non è sufficiente a descrivere la struttura di quella rete sociale

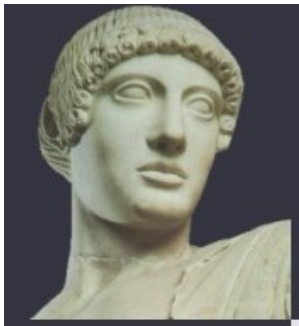
L'ORACOLO DI KEVIN BACON

- film collaboration network: descrive gli attori che hanno recitato insieme in almeno un file (internet movie database)
- Una "demo" che questa rete è una small world network è accessibile alla URL http://oracleofbacon.org/, mediante il gioco di Kevin Bacon
- Gioco di Kevin Bacon:
  - pensa ad un attore *A*
  - se *A* ha recitato in un film con Bacon, *A* ha numero di Bacon = 1
  - se *A* non ha mai recitato personalmente con Bacon, ma ha recitato con qualcuno che a sua volta ha recitato con Bacon, *A* ha numero di Bacon =2
  - e così via....

L'ORACOLO DI KEVIN BACON

- Pensa ad un attore X
  - Se questo attore ha recitato con Kevin Bacon, la sua Bacon Distance è 1
  - Se questo attore ha recitato con Y ed Y ha recitato con Kevin Beacon, la distanza e' 2
  - etc. etc.
- Esempi:
  - Marcello Mastroianni: Bacon Distance 2
    - Marcello Mastroianni in Poppies Are Also Flowers (1966) con Eli Wallach
    - Eli Wallach in Mystic River (2003) con Kevin Bacon
  - Brad Pitt: Bacon Distance 1
    - Brad Pitt was in Sleepers (1996) with Kevin Bacon
- Elvis Presley: Bacon Distance 2
  - Elvis Presley has recitsto in Live a Little, Love a Little (1968) con John Wheeler
- John Wheeler ha recitato in Apollo 13 (1995) with Kevin Bacon

L'ORACOLO DI KEVIN BACON



THE ORACLE OF BACON

Welcome
Credits
How it Works
Contact Us
Other stuff »

How good a center is | Kevin Bacon | ?

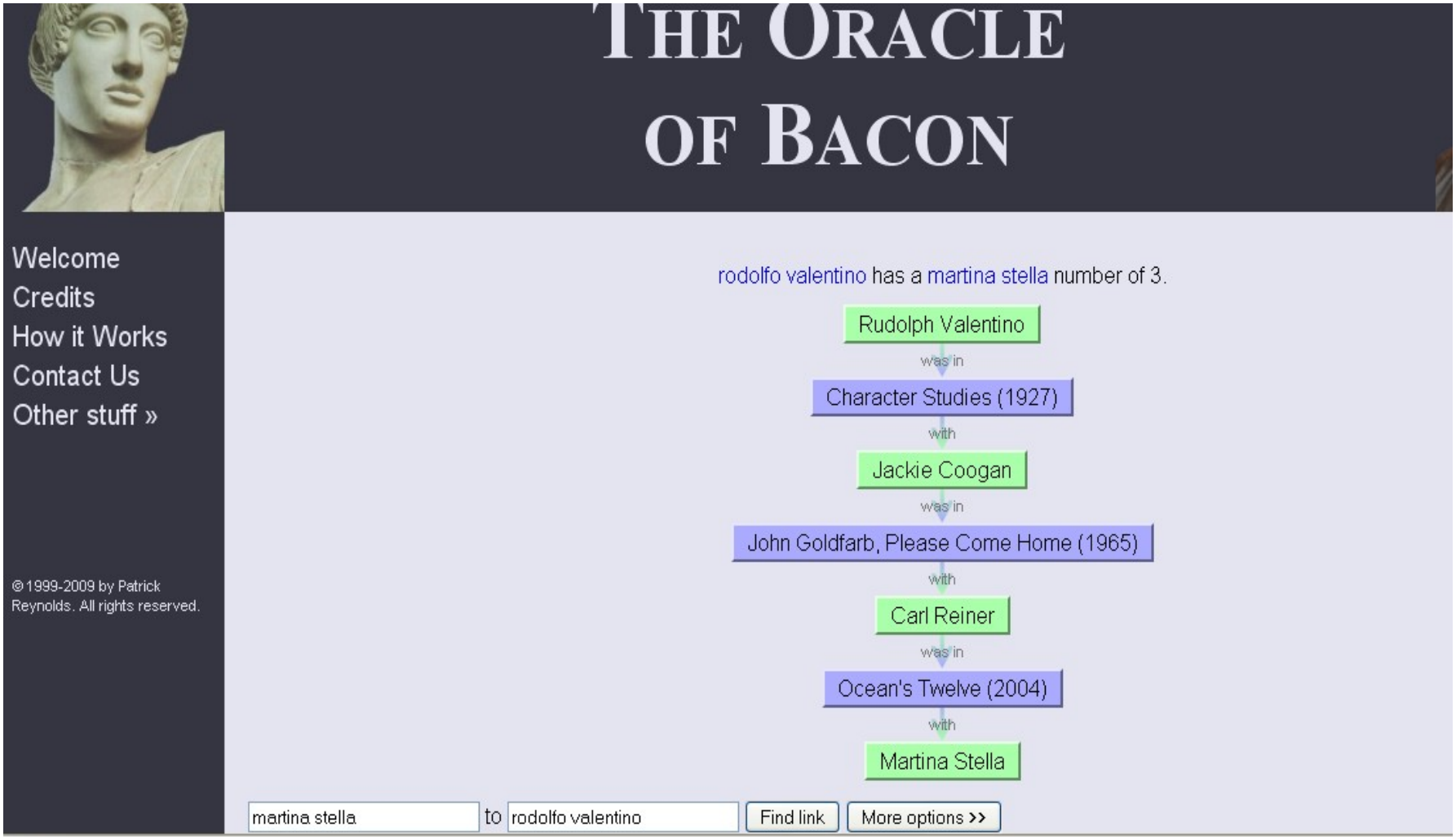
Kevin Bacon Number	# of People
0	1
1	2349
2	223940
3	666941
4	153220
5	9662
6	877
7	134
8	15

Total number of linkable actors: 1057139
Weighted total of linkable actors: 3118562
Average Kevin Bacon number: 2.950

L'ORACOLO DI KEVIN BACON

- Tutti i dati sugli attori e sui film provengono dall'Internet Movie Database dell'Università della Virginia
- Ogni attore è separato da Kevin Bacon da pochi link
- La media dei Bacon Number è 2.78
- Kevin Bacon, un attore non di primo piano, sembra essere al centro della rete di collaborazione tra gli attori, ma.... le cose non stanno realmente così
- Kevin Bacon posside un numero relativamente limitato di links con altri attori, ha girato un numero relativamente limitato di films ma la distanza media di un attore da lui (2.78) è relativamente alta

L'ORACOLO ESTESO A TUTTI GLI ATTORI



CENTRALITA' DEGLI ATTORI (212250 ATTORI)

Rank	Name	Average distance	#of movies	# of links
1	Rod Steiger	2.537527	112	2562
2	Donald Pleasence	2.542376	180	2874
3	Martin Sheen	2.551210	136	3501
4	Christopher Lee	2.552497	201	2993
5	Robert Mitchum	2.557181	136	2905
6	Charlton Heston	2.566284	104	2552
7	Eddie Albert	2.567036	112	3333
8	Robert Vaughn	2.570193	126	2761
9	Donald Sutherland	2.577880	107	2865
10	John Gielgud	2.578980	122	2942
11	Anthony Quinn	2.579750	146	2978
12	James Earl Jones	2.584440	112	3787
...				
876	Kevin Bacon	2.786981	46	1811

ALTRI ESEMPI DI SMALL WORLDS

- La topologia di Internet Topology (routers)
  - Grado medio di separazione 6
  - Rete di 100.000 nodes

- La rete degli aereoporti
  - Grado medio di seprazione tra due aereoporti e' circa 3.5

- Ed inoltre....
  - La rete delle collaborazioni industriali
  - La rete delle colleaborazioni scientifiche
  - etc....

SMALL WORLDS NETWORKS

- Conclusione: la rete degli attori è una small world

- Infine alcune curiosità. Six Degrees of Separation:
  - Six Degrees of Separation: nel 1991 una commedia teatrale di John Guare, da cui è stato tratto nel 1993 un film di Fred Schepisi.

- ***Citazione dal film : "I read somewhere that everybody on this planet is separated by only six other people. Six degrees of separation between us and everyone else on this planet. The President of the United States, a gondolier in Venice, just fill in the names. I find that extremely comforting, that we're so close, but I also find it like Chinese water torture that we're so close because you have to find the right six people to make the connection. It's not just big names -- it's anyone. A native in a rain forest, a Tierra del Fuegan, an Eskimo. I am bound -- you are bound -- to everyone on this planet by a trail of six people..... How everyone is a new door, opening into other worlds."***

SMALL WORLDS NETWORKS

- Small World Network = una qualsiasi coppia di nodi è collegata da un cammino caratterizzato da un numero molto limitato di hops, anche se la rete presenta un numero di nodi molto elevato

- Una rete presenta un comportamento di tipo small world se e solo se
  - Il diametro della rete cresce in modo logaritmico (o inferiore) in funzione di n, dove n è il numero di nodi della rete.
  - ma...la rete è caratterizzata da un alto coefficente di clusterizzazione

- Implicazioni sulla dinamica di processi che avvengono su small worlds networks:
  - diffusione rapida dell'informazione su una small world network

    esempio: diffusione di gossip
  - diffusione di un pacchetto su una rete
  - diffusione di un virus nella popolazione

SMALL WORLDS VS. RANDOM GRAPHS

- Osservazione generale: molte reti reali presentano un comportamento di tipo 'small world'

- Abbiamo visto che è possibile dimostrare che la distanza tra una qualsiasi coppia di nodi appartenenti ad un grafo random cresce come il logaritmo del numero di nodi della rete

- Un grafo random possiede un basso diametro, tuttavia possiede un coefficente di clusterizzazione basso

- Domanda: E' possibile costruire modelli più realistici di reti reali? E' possibile costruire modelli che siano caratterizzati contemporaneamente dalla proprietà di small world e da un alto coefficente di clusterizzazione?
  - Watts Strogatz
  - Kleinberg