Lezione n.11
ANALISI DI RETI COMPLESSE
21/4/2009

Libro di Testo, Capitolo 3

RIASSUNTO DELLE PROSSIME LEZIONI

- Introduzione : Reti complesse
- Small-Worlds
  - Social Networks
  - L'esperimento di Milgram
- Grafi Random
- Il modello Watts-Strogatz, il modello di Kleinberg
- Scale-Free Networks
  - Risultati sperimentali
  - Il modello di Barabasi-Albert
- Applicazioni ai sistemi P2P

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 i Medici divennero signori di Firenze
  - social networks online: Linkedin, MySpace, Facebook
  - small world: l'esperimento di Milgram

- 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
- Reti costruite per la distribuzione di qualche risorsa (rete elettrica, reti stradali,...)

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

- Necessità di nuovi modelli

PROPRIETA' DELLE RETI COMPLESSE

- Proprietà interessanti caratteristiche di molte reti complesse
  - Small World
  - Clustering
  - Distribuzione dei gradi dei nodi
  - Network Resilience

- I grafi random possiedono solo alcune di queste proprietà

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

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

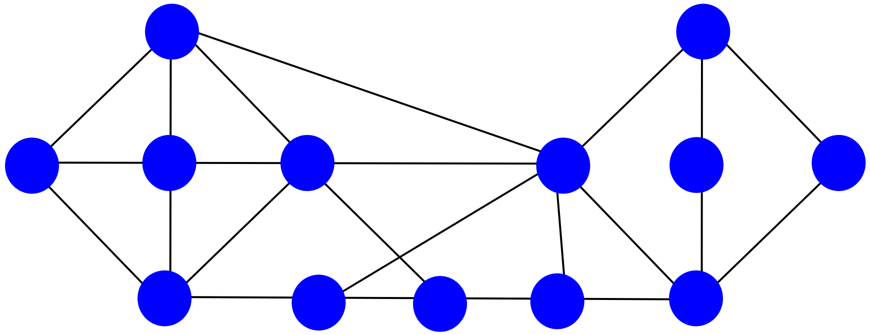
SMALL WORLD SOCIAL NETWORKS

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 Nebraska 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"

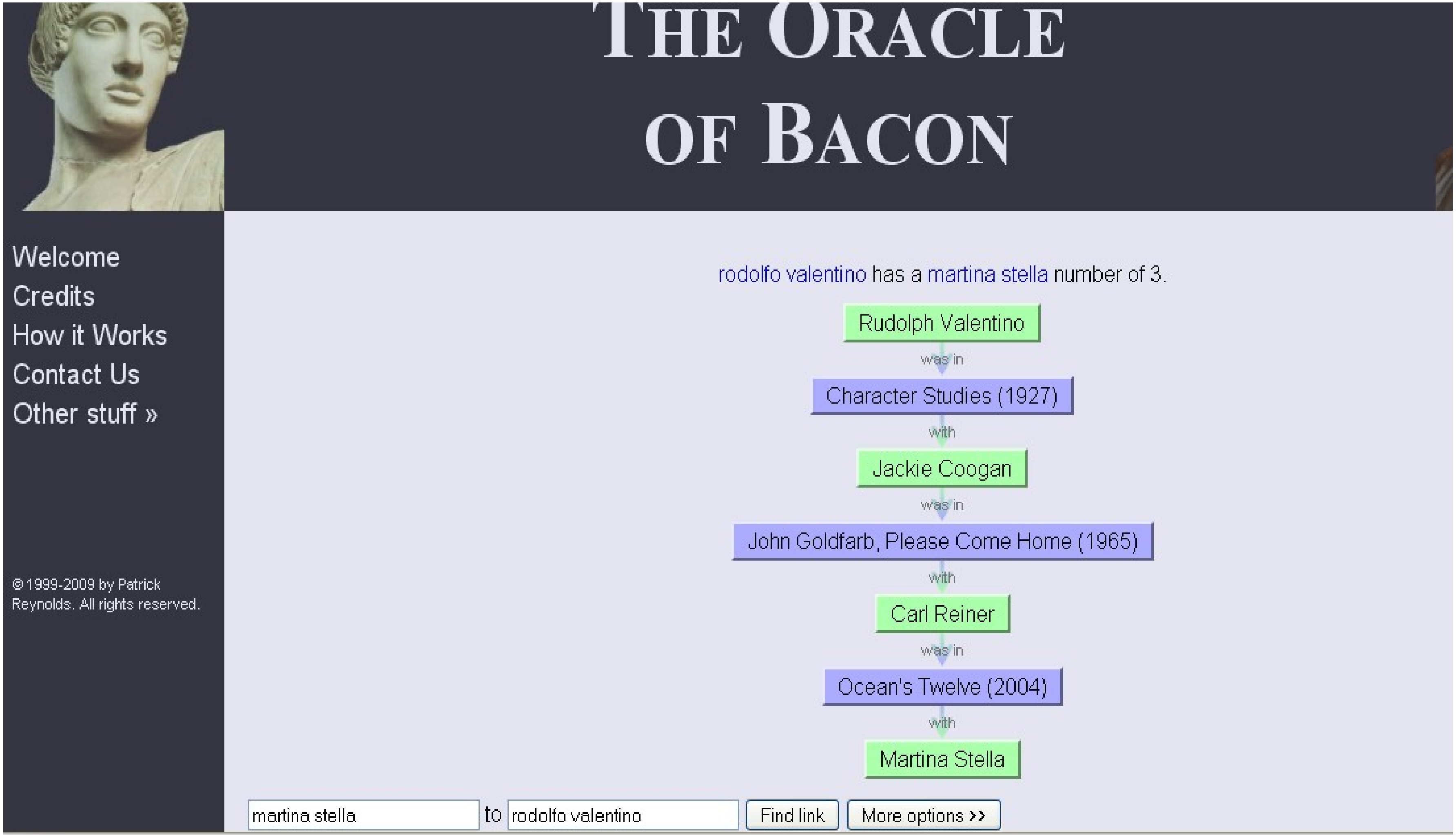
- L'esperimento dimostrò che il diametro della social network analizzata risultava molto piccolo, nonostante la località della rete

- 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



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

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

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 L cresce in modo logaritmico (o inferiore) in funzione di n, dove n è il numero di nodi della rete. Il grado dei nodi del grafo ha un valore medio prefissato

- E' 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 è una rete small world

SMALL WORLDS: IMPLICAZIONI

Implicazioni sulla dinamica di processi che avvengono su small worlds networks:

- Diffusione rapida dell'informazione su una small world network
- Esempio: diffusione di *gossips*
- Diffusione di un pacchetto su una rete
- Diffusione di un virus nella popolazione
- ....

CLUSTERING

- Una proprietà comune di molte reti complesse, specialmente delle social networks è l'alta clusterizzazione dei nodi
- Esempio: i miei amici sono amici tra di loro, le persone che io conosco, si conoscono, etc...
- Per misurare questa caratteristica: Coefficiente di clusterizzazione
- Si consideri un nodo n con k vicini. Coefficiente di clusterizzazione C di n

$$c = CE/CT$$

  - CE = connessioni esistenti tra i k vicini di n
  - CT= k(k-1)/2 = numero totale di connessioni che possono essere definite tra i k vicini di n
- Alto coefficiente di clusterizzazione osservato in diverse reti reali
- I grafi random invece presentano un basso coefficiente di clusterizzazione
  - non sono un modello valido per molte reti reali

DEGREE DISTRIBUTION

- Grado di un nodo = numero di archi incidenti in un nodo
- Analisi della distribuzione dei gradi dei nodi di una rete
  - P(k) = probabilità che un nodo scelto in modo casuale abbia esattamente grado k
- Grafi Random
  - gli archi tra i nodi sono scelti in modo casuale
  - la maggior parte dei nodi ha lo stesso grado k, che è approssimativamente il grado medio dei nodi della rete
  - La probabilità di avere nodi con grado significativamente diverso da k decresce esponenzialmente
  - La distribuzione di probabilità è poissoniana

SCALE FREE NETWORKS

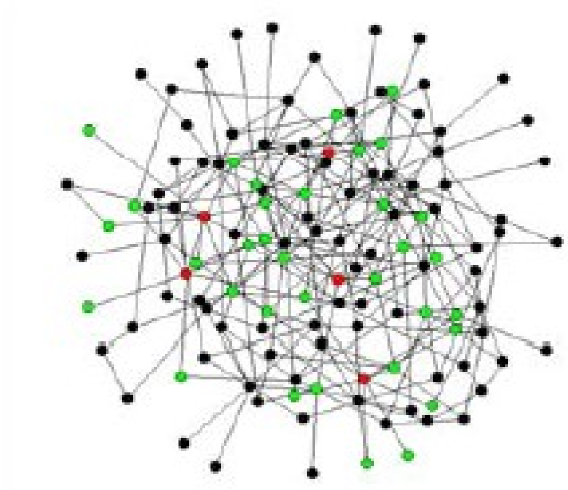
- L'analisi sperimentale di diverse reti reali ha mostrato che la distribuzione dei gradi dei nodi della rete si discosta parecchio dalla distribuzione di Poisson
- Es: distribuzione dei gradi dei nodi per WEB, Internet, segue una power law

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

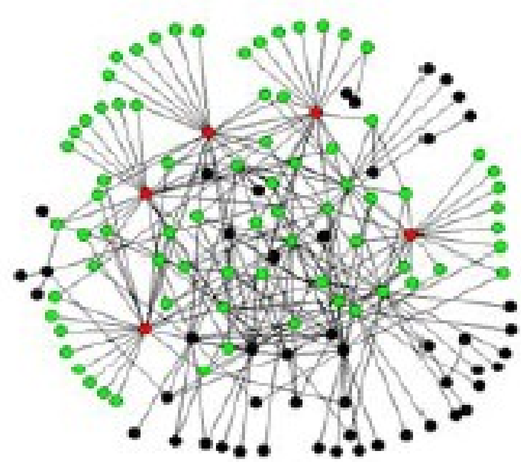
  maggiore è il grado considerato, minore è la probabilità di individare nodi con quel grado all'interno della rete
- Scale Free Networks (Barabasi Albert)
  - si considera che la rete che si accresce continuamente mediante l'aggiunta di nuovi nodi
  - la probabilità che un nuovo nodo 'si connetta' ad un nodo esistente dipende dal grado del nodo
  - nuovi nodi si connettono in maniera preferenziale a nodi esistenti caratterizzati da un alto grado

SCALE FREE NETWORKS:TOPOLOGIA

Random Graph



Scale Free Network



- le reti rappresentate hanno lo stesso numero di nodi e di archi
- in rosso i nodi con il maggior numero di links, in verde i loro vicini

- tutti i nodi hanno approssimamente lo stesso numero di vicini
- solo il 27% dei nodi della rete sono raggiunti direttamente dai nodi rossi

- molti nodi con pochi vicini, pochi nodi con un alto numero di vicini
- il 60% dei nodi della rete possono essere raggiunti direttamente dai nodi rossi

RETI COMPLESSE: ALTRE PROPRIETA'

Altre proprietà delle reti complesse

- Grado massimo, grado minimo,....
- Network resilience
  - Comportamento della rete in conseguenza di rimozione di nodi (es: vaccinazione di individui, attacchi ad Internet,...)
- Mixing patterns
  - Classi di vertici connessi
- Community Structures
  - Gruppi di nodi connessi da molti archi, con pochi archi tra i gruppi
- .....vedere materiale consigliato

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

- I grafi random rappresentano il modello più semplice per reti complesse
  - Assunzioni semplici
  - Proprietà analizzabili con strumenti statistici

- Idea Base: Dato un numero fisso n di vertici, gli archi vengono creati in modo casuale, secondo una certa distribuzione di probabilità

Modelli proposti

- Erdós-Renyi
- Gilbert

Si ricavano proprietà statistiche. Una proprietà si verifica con una certa probabilità.

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

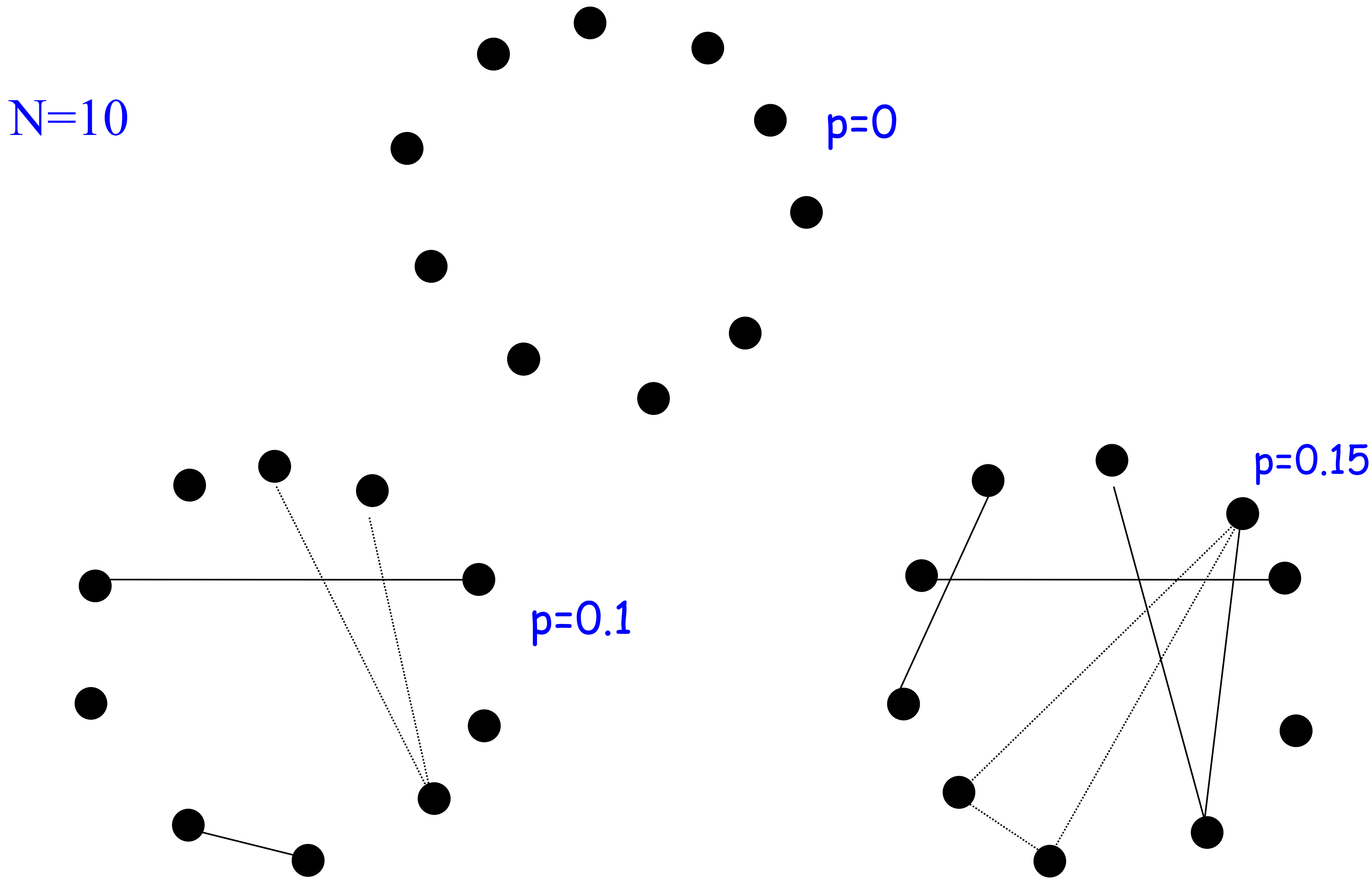
- Costruzione del grafo. Esempio: $p = \frac{1}{4} = 0.25$

  si considera una coppia di nodi (u,v), si genera un numero casuale x compreso tra 0 ed 1, se $x < p$, si definisce un arco tra u e v, altrimenti i nodi non vengono connessi

- Distribuzione binomiale: la presenza/assenza di un arco è indipendente da quella degli altri archi

- Numero medio n di archi per un grafo di N nodi $E(n)=p[N(N-1)/2]$

GRAFI RANDOM: IL MODELLO DI ERDOS RENYI



GRAFI RANDOM: IL MODELLO DI ERDOS RENYI

- Obiettivo dell'analisi dei grafi random: determinare la probabilità $p(N)$ in corrispondenza della quale il grafo ottenuto gode di una certa proprietà PR.

- Esempio: in corrispondenza di quale valore di $p(N)$ il grafo contiene sottografi con certe caratteristiche (ad esempio contiene triangoli,...)

- Si dimostra che data una proprietà PR esiste un valore critico $p_c(N)$ tale che
  - se $p(N)$ cresce meno velocemente di $p_c(N)$ per $N \to \infty$ allora ogni grafo corrispondente a p (N) non possiede la proprietà PR
  - altrimenti, se $p(N)$ cresce più velocemente di $p_c(N)$ per $N \to \infty$ allora ogni grafo corrispondente a $p(N)$ è caratterizzato da PR

- Si dimostra che il passaggio da grafi che non verificano PR a quelli che la verificano risulta brusco in corrispondenza del valore $p_c(N)$

GRAFI RANDOM: DEGREE DISTRIBUTION

- Grado di un nodo= numero di archi incidenti in quel nodo
- Intuizione: poichè in un grafo random le connessioni vengono create in modo casuale, con probabilità p, è ragionevole aspettarsi che
  - La maggior parte dei nodi abbia lo stesso grado k
  - Il valore approssimato di k sia uguale al grado medio <k> della rete

$$\langle k \rangle \approx pN$$

- Più formalmente:
  - Con una buona approssimazione, la funzione di distribuzione dei gradi dei nodi di un grafo random può essere approssimata con una distribuzione binomiale
  - La distribuzione binomiale può essere approssimata, per valori alti di N, con una distribuzione di Poisson
- I grafi random non descrivono appropriatamente le scale free networks

GRAFI RANDOM: DEGREE DISTRIBUTION

- Distribuzione di Poisson. Probabilità che un nodo abbia grado r:

$$P(X = r) = e^{-\langle k \rangle} \langle k \rangle^{r} / r!$$

  dove <k> è il valore medio del grado di un nodo

- Invece le scale free networks presentano una distribuzione del tipo

$$P(X = r) \cong r^{-\gamma}$$

- I grafi random non modellano le scale free networks in modo adeguato

GRAFI RANDOM: IL DIAMETRO

- Sappiamo che la maggior parte dei nodi hanno grado K

- Dato un nodo n il numero medio di nodi a distanza d da n è K^d

- Quindi, per raggiungere 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

- I grafi random sono small worlds !

CLUSTERIZZAZIONE

Clusterizzazione:

- 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 è probabile che esista una connessione anche tra di loro
- proprietà di 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

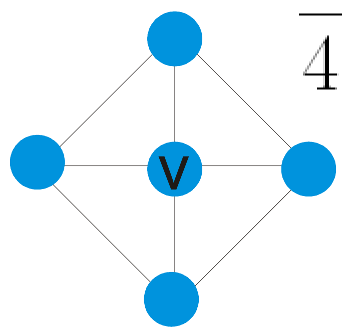
CLUSTERIZZAZIONE: DEFINIZIONE 1

Coefficente di clusterizzazione C(v) di un vertice v del grafo è calcolato come

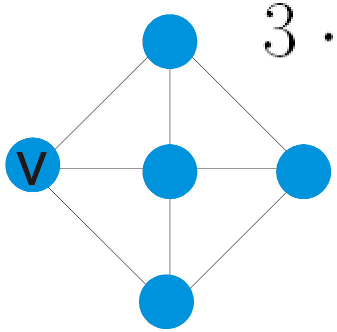
$$C(v) = \frac{e(v)}{deg(v)\,(deg(v)-1)\,/2}$$

e(v) denota il numero di connessioni esistenti tra i vicini di v

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

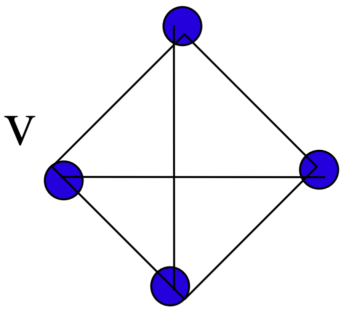


$$\frac{2 \cdot 2}{3 \cdot 2} = 2/3$$

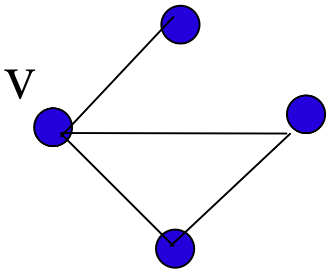


C(V) misura quanto i vicini di v formano strutture 'di tipo clique'

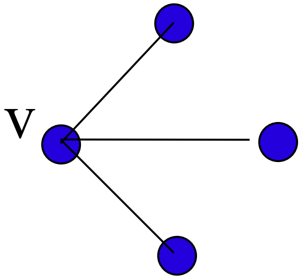
CLUSTERIZZAZIONE: DEFINIZIONE 1



C =1



C = 1/3



C = 0

CLUSTERIZZAZIONE: DEFINIZIONE 2

- 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

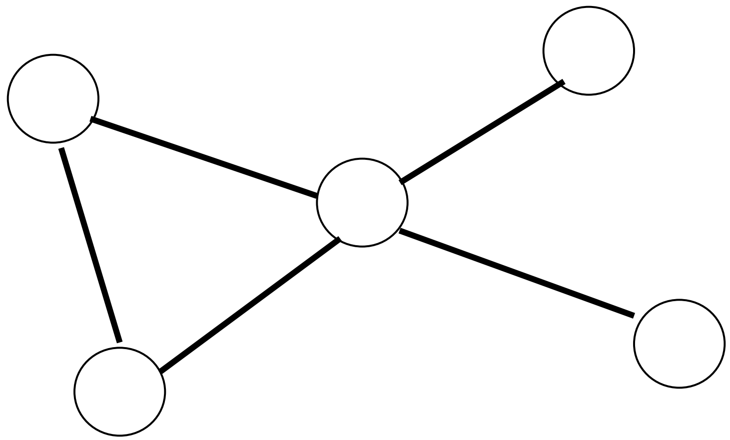
  fattore 3 misura il fatto che ogni triangolo contribuisce a 3 triple connesse

- 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

CLUSTERIZZAZIONE: DEFINIZIONE 2



La rete è caratterizzata da

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

GRAFI RANDOM: CLUSTERIZZAZIONE

- Clusterizzazione: se due nodi hanno un nodo vicino in comune è probabile che esista una connessione anche tra di essi
- In un grafo random la probabilità di connettere due nodi della rete è p, indipendentemente dal fatto che essi posseggano vicini a comune
- Il coefficente di clusterizzazione di un grafo random caratterizzato da una probabilità p è uguale a p.

Quindi i grafi casuali

- sono reti small worlds
- sono caratterizzati da un basso coefficiente di clusterizzazione e quindi hanno poca capacità di modellare l'aggregazione, caratteristica tipica di molte reti reali
- il basso coefficiente di aggregazione consente di ottenere bassi gradi di separazione tra i nodi

GRAFI RANDOM VS. RETI REALI

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

GRAFI RANDOM VS. RETI REALI

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

IL MODELLO DI WATTS-STROGATZ

- Watts-Strogatz considerano due caratteristiche fondamentali delle reti complesse:
  - il coefficente di clusterizzazione: misura la regolarità e la località della rete. Se tale coefficente è alto, gli archi collegano principalmente nodi vicini, piuttosto che vertici lontani.
  - la distanza tra i vertici.

- Se il coefficente di clusterizzazione è alto, la distanza media tra due nodi dovrebbe essere alta, perchè gli archi non sono "casuali" ma "locali"

- Ma..... la maggior parte delle reti osservate presenta un valore alto del coefficente di clusterizzazione (0.3-0.4) e un valore basso della distanza tra nodi

AL CONFINE TRA ORDINE E CAOS

Le idee fondamentali.

- Una rete regolare (ad esempio una griglia) presenta una forte aggregazione, ma non è una rete small world
- Un random graph è uno small world, ma non presenta aggregazione
- Si vuole definire un modello di rete che riunisca le due caratteristiche
  - sufficientemente regolare (alto coefficiente di clustering)
  - sufficientemente caotico per poter mantenere basso il grado di separazione tra i nodi (small worlds)

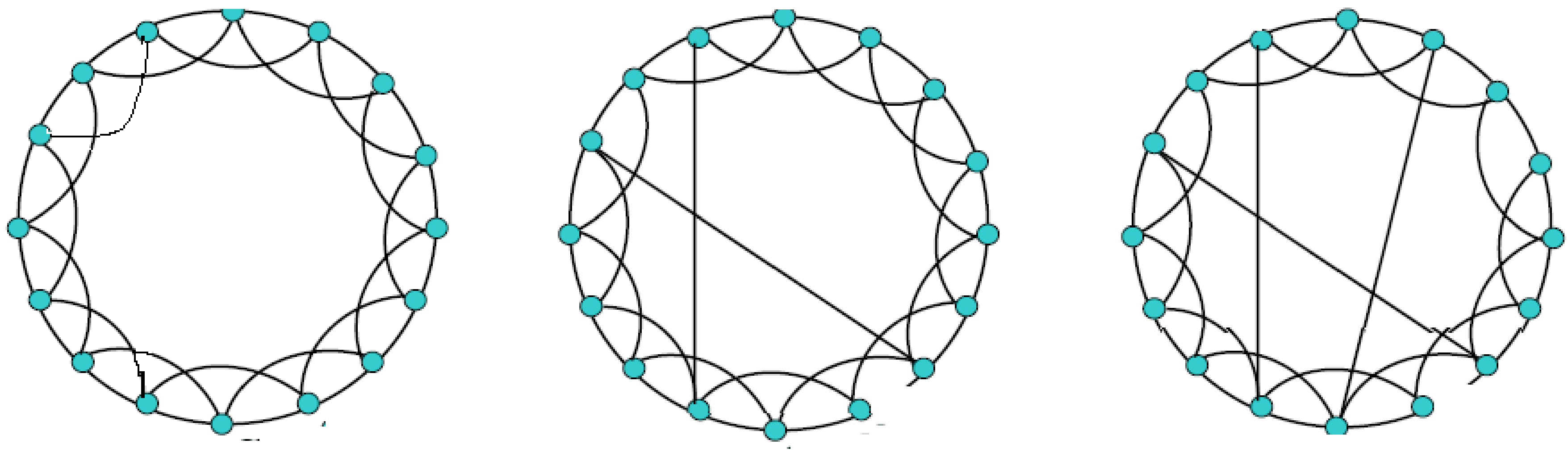
- Un modello che sia un compromesso tra ordine e caos....

IL MODELLO DI WATTS-STROGATZ

- Questo fenomeno non può essere spiegato mediante un modello grid-like, ne mediante un grafo random
  - una rete grid-like è caratterizzata da regolarità e località, ma da alto valore della distanza media
  - i grafi random hanno un coefficiente di clusterizzazione proporzionale a p.

- Watts-Strogatz propongono un modello ibrido
  - si parte da un anello di n vertici
  - si connette ogni vertice con i suoi k vicini sull'anello
  - si 'riavvolge' ogni vertice con probabilità uguale a p, si mantiene fisso uno dei vertici e si sceglie un nuovo target come altro vertice, tra tutti i vertici,in maniera casuale

IL MODELLO DI WATTS STROGATZ

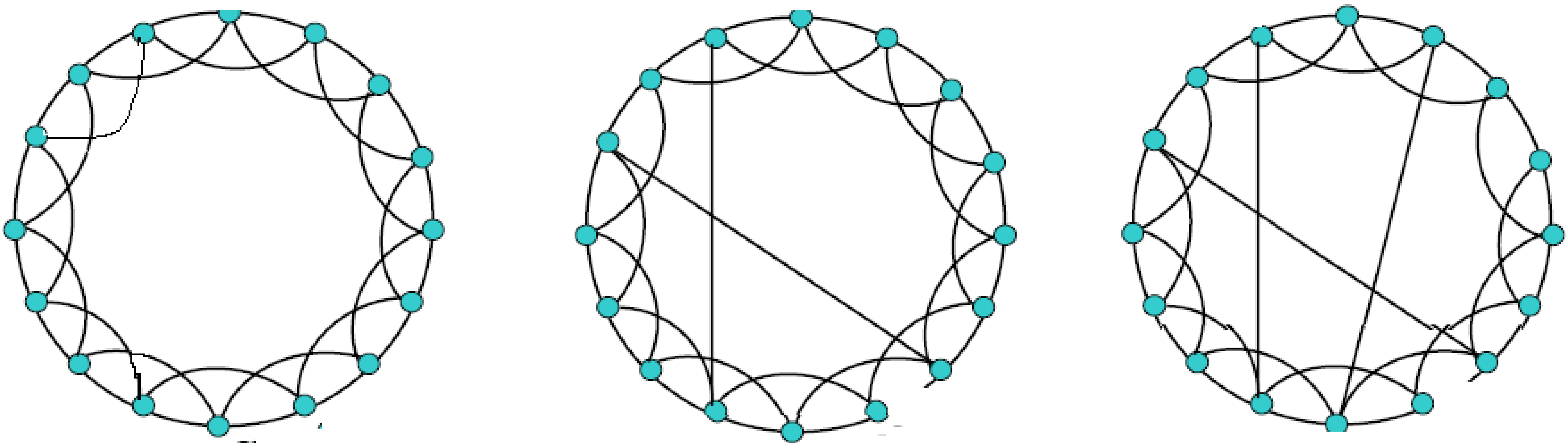


Procedura di costruzione del grafo

Si parte da un anello di n vertici e si connette ogni vertice con i suoi *k* vicini sull'anello

Si „riavvolge“ un arco (v,w) con probabilità uguale a p: si mantiene fisso v e si sceglie un nuovo vertice w', scelto tra tutti i vertici, in modo casuale come destinazione dell'arco

IL MODELLO DI WATTS STROGATZ



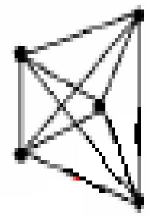
- Per p=0, la rete risultante è completamente regolare, con un coefficente di clusterizzazione di circa ¾ per valori grandi di k ed un diametro è O(n)

- Per p=1, la rete risultante si può considerare un grafo random con un coefficente di clusterizzazione basso ed un diametro O(log n)

- Casi interessanti: valori intermedi di p

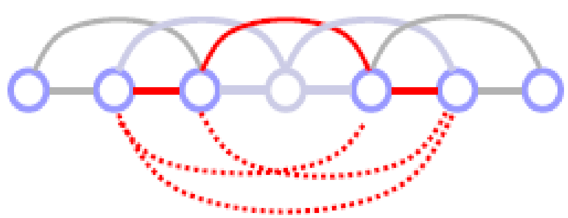
CLUSTERIZZAZIONE NELLA RETE REGOLARE

$$k = N-1,\ n = \frac{N(N-1)}{2}$$



Archi tra i vicini del nodo i

$$C_i \equiv \frac{n_i}{k_i(k_i-1)/2}, \quad k \neq 0,1$$



Clique: sottografo completamente connesso

Coefficiente di Clusterizzazione: misura quanto le connessioni tra i vicini di un nodo sono simili ad una clique

- k vicini possono avere al massimo $k*(k-1)/2$ connessioni
- solo alcune di queste connessioni sono presenti

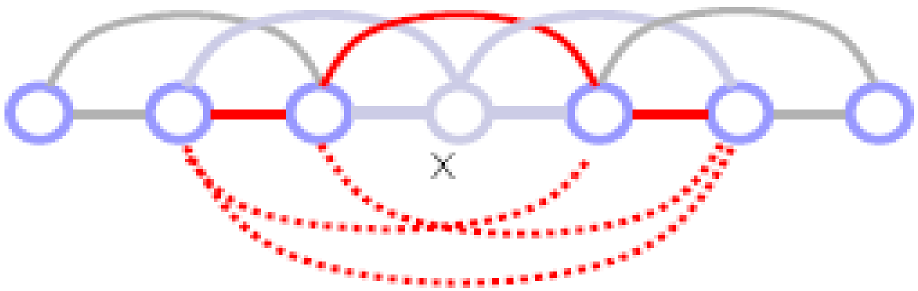
Nella figura k= 4, ed il coefficente è il seguente

$$C_i = 3/\ ((4*3)/2) = ½$$

CLUSTERIZZAZIONE NELLA RETE REGOLARE

- Coefficiente di clusterizzazione di una rete regolare, in cui ogni nodo è connesso a k vicini (in figura k=4)

$$C = 3(K-2) / 4(K-1)$$



- Consideriamo l'insieme N dei k/2 nodi presenti a destra di x

- Presi due nodi $\in$ N, esiste tra di loro un link, poichè essi distano meno di k /(al massimo distano k/2)

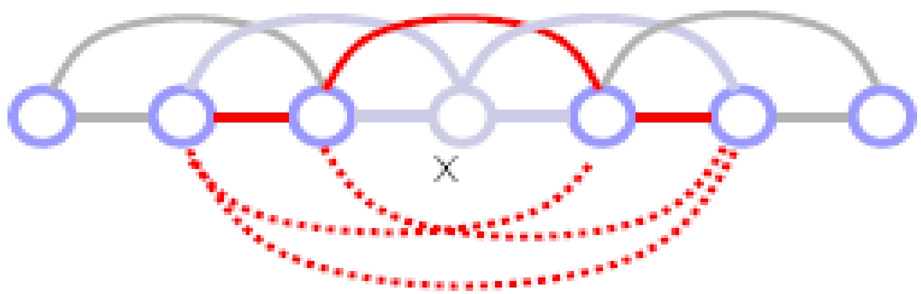
- Il numero totale di link tra nodi $\in$ N è (k/2) *((k/2)-1))/2
  (Nell'esempio (2*2)/2 = 1)

- Ragionamento analogo per i nodi a sinistra di x, per cui il numero totale di links tra vicini a destra + a sinistra = 2* ((k/2) *((k/2)-1))/2

CLUSTERIZZAZIONE NELLA RETE REGOLARE

Coefficiente di clusterizzazione di una rete regolare, in cui ogni nodo è connesso a k vicini

$$C = 3(K-2) / 4(K-1)$$



- inoltre ci sono i collegamenti tra i nodi a destra di X con quelli a sinistra di X
- questi collegamenti sono 0 per il nodo più distante da X, 1 per quello successivo, 2, ...,k/2 per il nodo più vicino ad X. In totale quindi

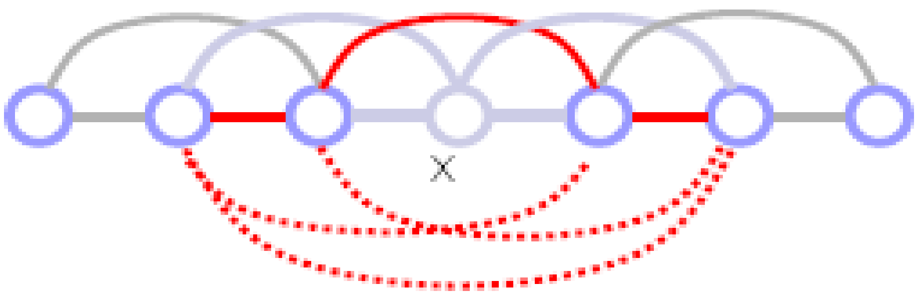
$$0+1+ \; .+ (k/2)= (k/2) *((k/2)-1))/2$$

- Nell'esempio (2*2)/2 = 1)

CLUSTERIZZAZIONE NELLA RETE REGOLARE

Coefficiente di clusterizzazione di una rete regolare, in cui ogni nodo è connesso a k vicini

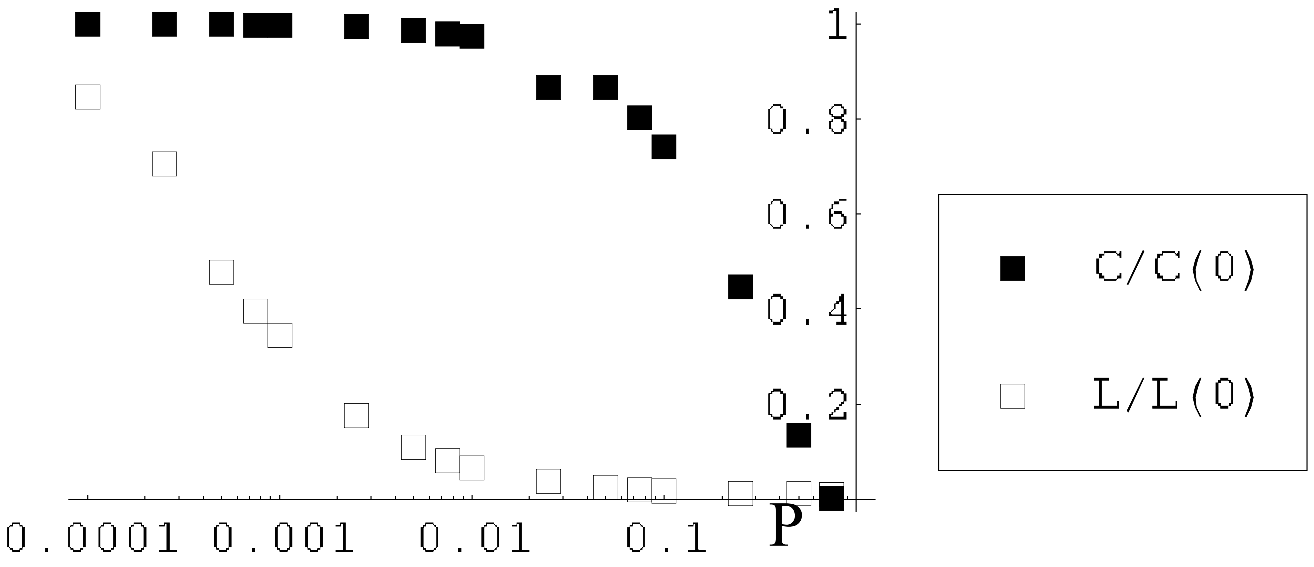
$$C = 3(K-2) / 4(K-1)$$



- in totale tra i vicini di X ci sono $3k/2((k/2)-1)/2$ links
- il numero totale di collegamenti possibili tra i vicini di X è $k(k-1)/2$
- da cui si ottiene

$$3k/2(k/2-1)/2 \; / \; k(k-1)/2 = 3(k-2)/4(k-1)$$

IL MODELLO DI WATTS-STROGATZ



- Il diagramma mostra il coefficente di clusterizzazione e la distanza media in funzione della probabilità p. (Normalizzata dai valori ottenuti per p=0.)
- Risultato: il coefficente di clusterizzazione è alto per valori piccoli di p ma la lunghezza media della distanza decresce rapidamente
- Small World networks combinano un alto coefficente di clusterizzazione con un valore piccolo della distanza media.

CLUSTERIZZAZIONE: DEFINIZIONE 3

- Definizione alternativa introdotta da Watts-Strogatz associa il coefficiente di clusterizzazione ad ogni nodo del grafo.

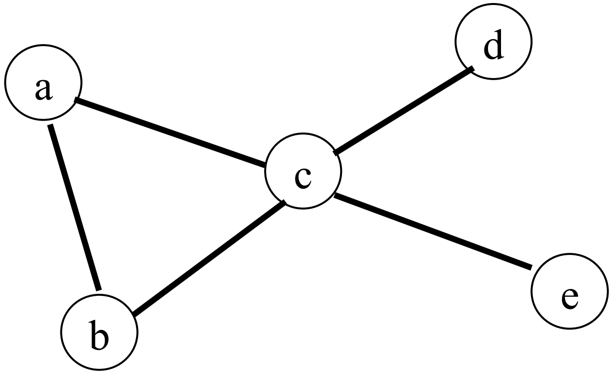
  C_i = numero di triangoli connessi ad i / numero di triple 'centrate' su i, per ogni vertice I del grafo

- C_i =0 per i vertici di grado 0 o di grado 1

- Coefficiente di clusterizzazione associato al grafo

$$C=(1/n) * \Sigma_i C_i$$

CLUSTERIZZAZIONE: DEFINIZIONE 3



Coefficenti di clusterizzazione associati ai nodi:

$Ca = 1$
$Cb = 1$
$Cc = 1/6$
$Cd = 0$
$Ce = 0$

Coefficiente di clusterizzazione del grafo:

$C = (1+1+1/6)\ /5 = 13/30$

CONCLUSIONI

- Esperimento di Milgram: evidenza in modo empirico il fenomeno "small world" in una rete sociale

- small world = esistono delle "catene di conoscenze" di lunghezza limitata che connettono una qualsiasi coppia di persone sconosciute

- Six Degree of Separation

- Caratterizzazione del fenomeno:
  - random networks
  - modello di Watts-Strogatz
  - modello di Kleinberg

CONCLUSIONI

Random Networks

- posseggono diametro limitato
- la lunghezza media dei cammini è $O(\log(n))$, dove n è il numero di nodi della rete
- adatte a modellare le "small world networks"
- caratterizzate da un basso coefficiente di clusterizzazione $\cong 1/n$: la probabilità che i vicini di un nodo siano essi stessi vicini è bassa
- situazione reale: i vicini dei miei amici sono i miei amici.....

MODELLO DI WATTS STROGATZ

- definisce di una rete regolare
  - griglia
  - anello con collegamenti con i nodi a distanza minore o uguale a k.

- "sovrappone" a questa struttura regolare un insieme limitato di collegamenti generati in modo casuale

- pochi link generati in modo casuale in un grafo clusterizzato producono dei cammini in media paragonabili a quelli di un grafo random

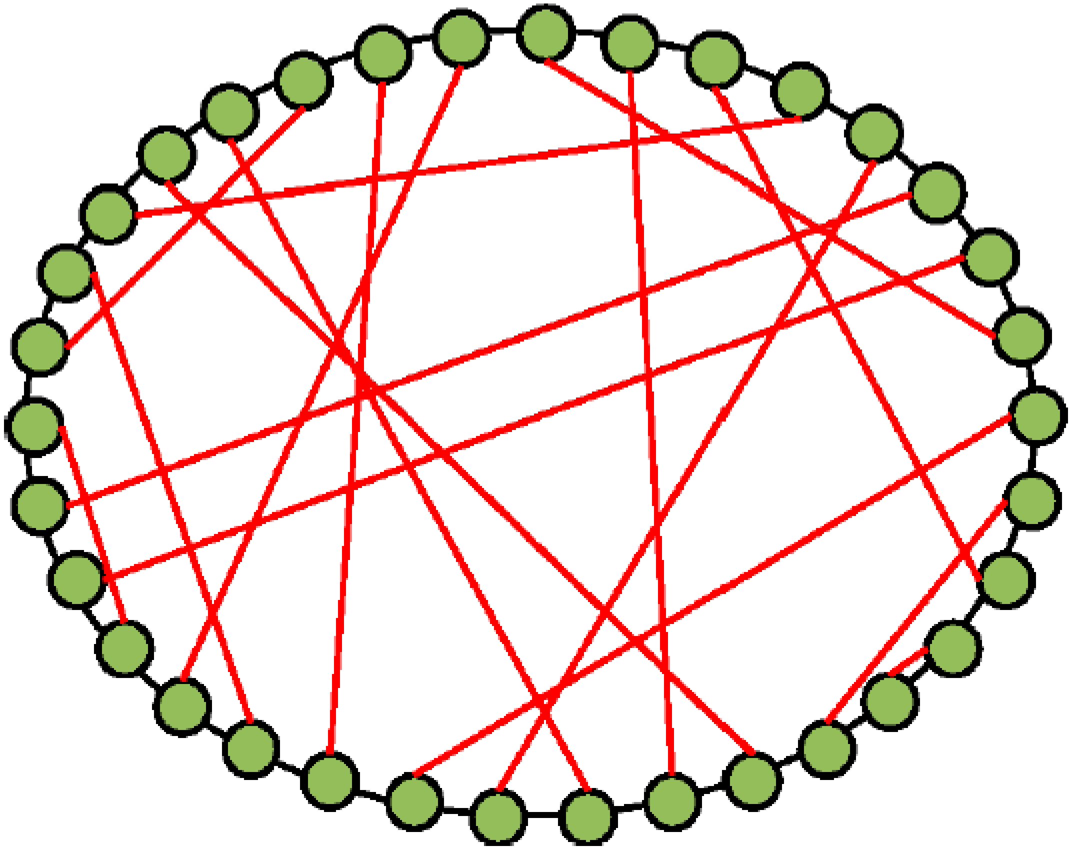
- la struttura regolare definisce un buon grado di clusterizzazione

- i collegamenti casuali garantiscono la proprietà di "small world"

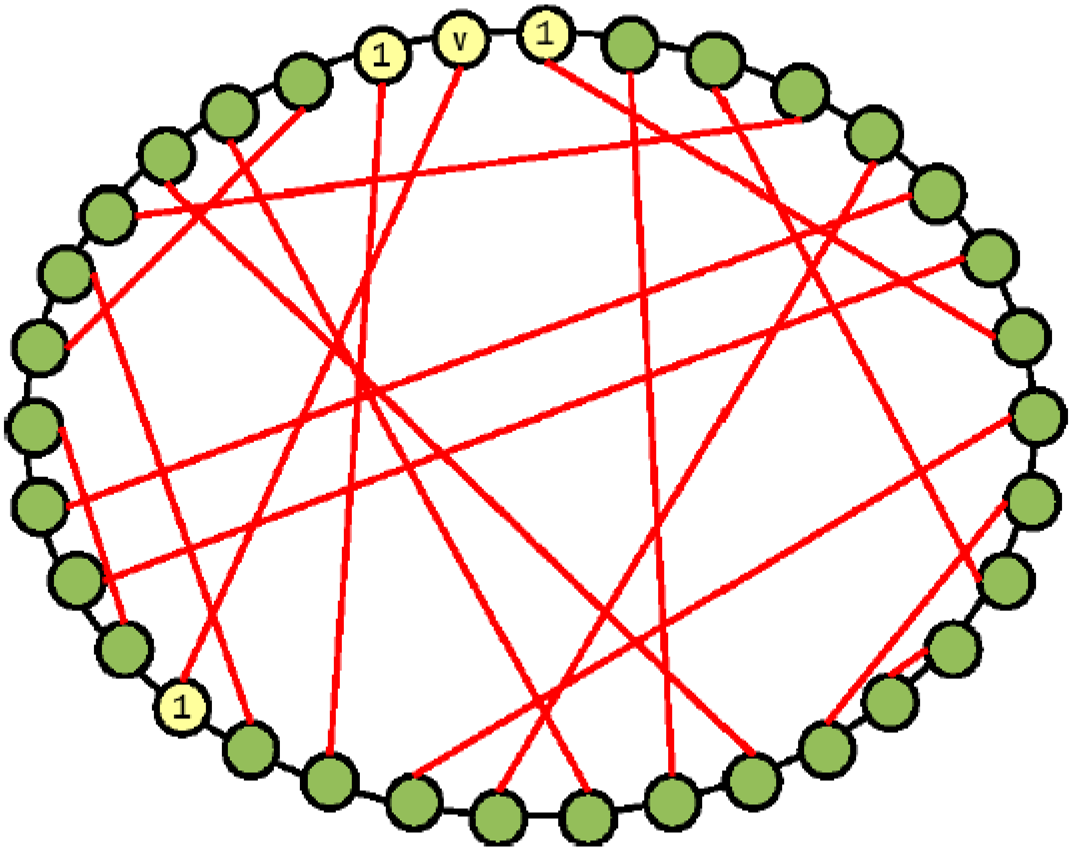
GRAFI REGOLARI:LUNGHEZZA MEDIA DEI CAMMINI

- In media un nodo si trova a distanza pari a metà dell'anello n/2
- Ad ogni passo, in una direzione, posso fare k/2 hops ad ogni passo, perchè ho a disposizione k vicini
- Lunghezza media dei cammini $\cong n/k \gg 1$

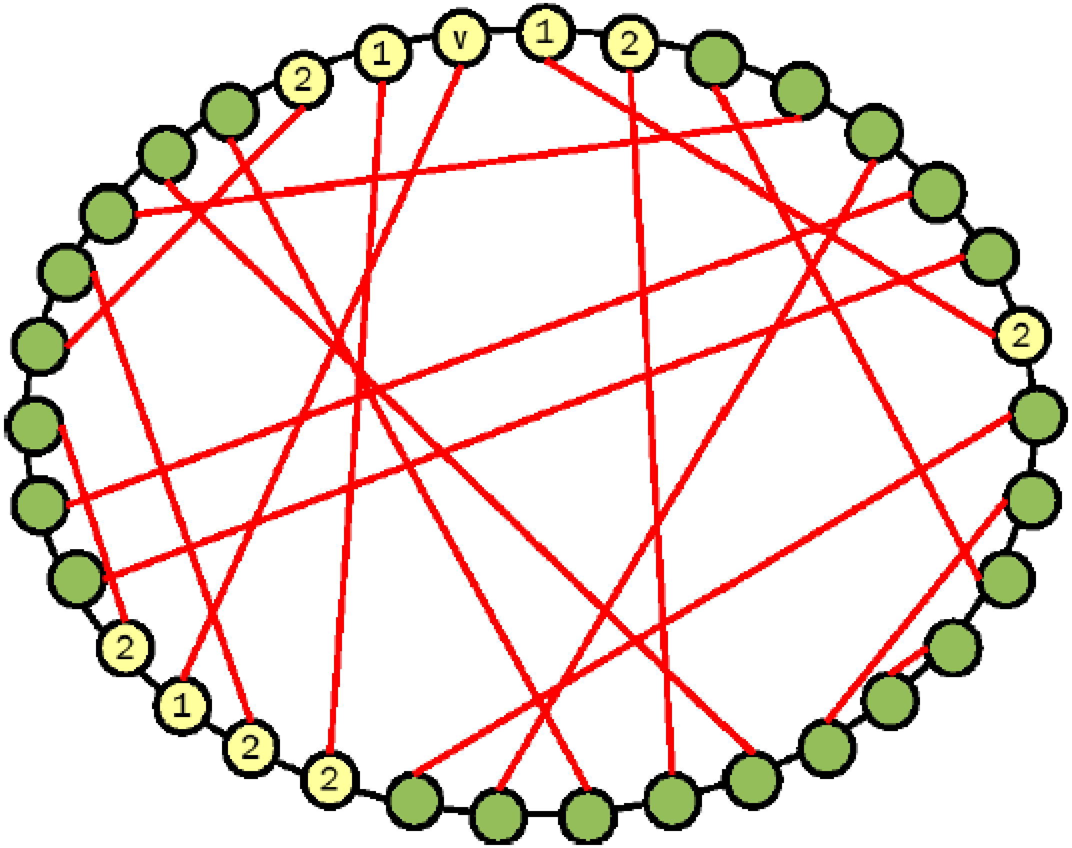
WATTS-STROGATZ: LUNGHEZZA DEI CAMMINI



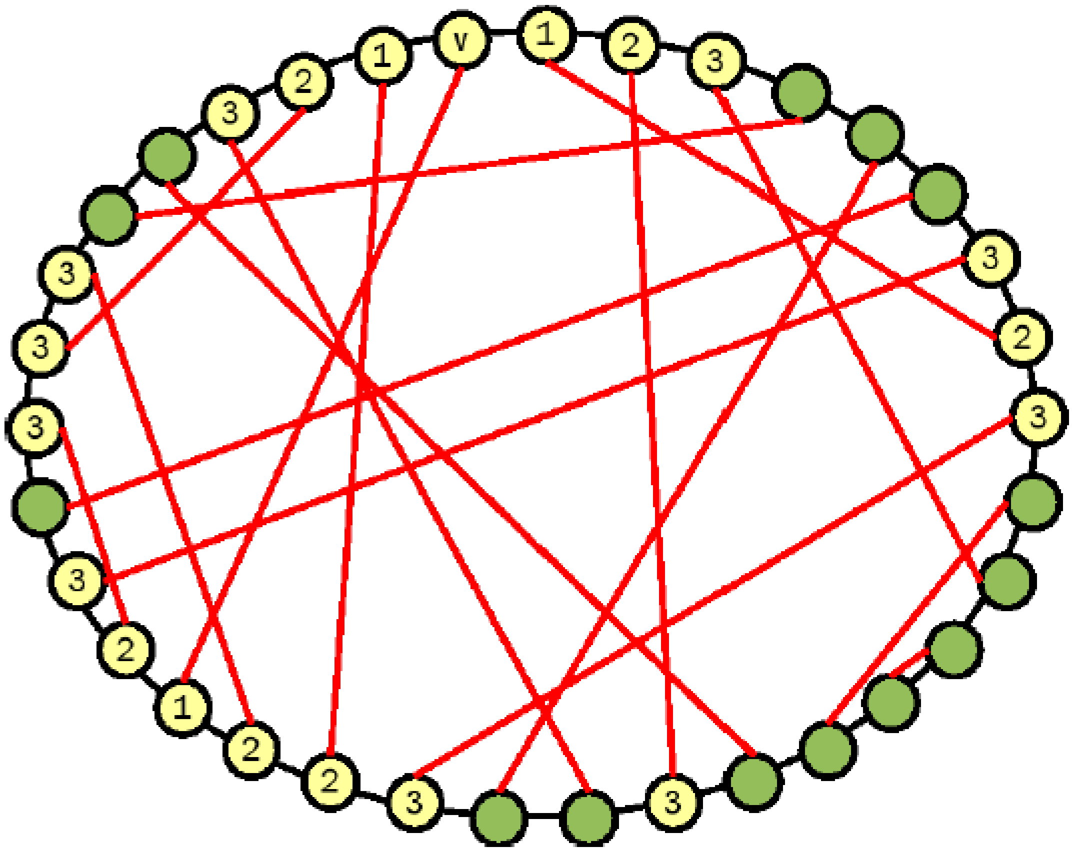
WATTS TROGATZ: LUNGHEZZA DEI CAMMINI



WATTS STROGATZ: LUNGHEZZA DEI CAMMINI



WATT STROGATZ: LUNGHEZZA DEI CAMMINI



GRAFI REGOLARI VS. GRAFI RANDOM

- Grafi Regolari: alto coefficiente di clusterizzazione, alto diametro
- Grafi Random: basso coefficiente di clusterizzazione, basso diametro

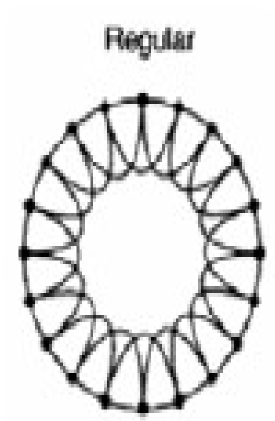
Regular Graph (k=4)

Long paths

- $L \sim n/(2k)$

Highly clustered

- $C \sim 3/4$



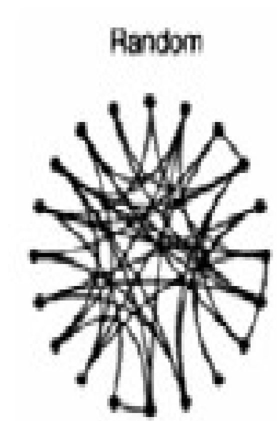
Random Graph (k=4)

Short path length

- $L \sim \log_k N$

Almost no clustering

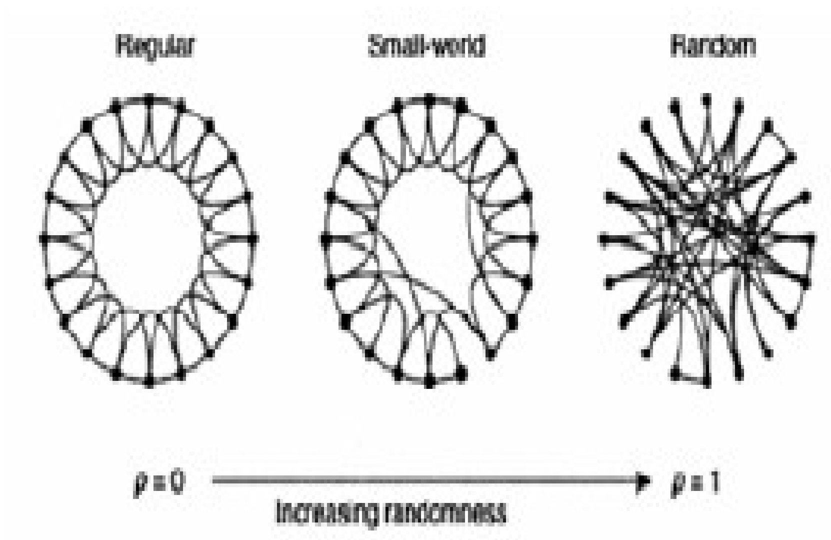
- $C \sim k/n$



MODELLO DI WATTS STROGATZ

Una rete generata secondo il modello di Watts Strogatz unisce le proprietà dei grafi regolari e di quelli random

- alto coefficiente di clusterizzazione
- bassa lunghezza dei cammini



If $n >> k >> \ln n >> 1$ then

- for $p \approx 0$:
  $l \sim n/2k$ and $C \sim 3/4$
  Ordered lattice
- for $p \approx 1$:
  $l \sim \ln n / \ln k$ and $C \sim k/n$
  Random Network
- for $0.001 \leq p \leq 0.01$:
  $l \sim \ln n / \ln k$ and $C \sim 3/4$
  This is a **small world network**

SMALL WORLD: ROUTING

- Caratteristiche rilevate empiricamente dall'esperimento di Milgram
- esistono cammini di lunghezza limitata che connettono una qualsiasi coppia di individui
- gli individui sono in grado di scoprire questi cammini mediante una conoscenza parziale (locale) della rete
- perchè una qualsiasi coppia di individui è in grado di individuare in modo decentralizzato la catena limitata di conoscenza che li connettono?
- Quali caratteristiche della rete garantiscono l'esistenza di tale algoritmo di routing decentralizzato?

SMALL WORLD: ROUTING

Algoritmo di routing eseguito da ogni individuo:

- supponiamo che l'individuo sia posizionato nel vertice v di posizione Pos (v) all'interno di una griglia in uno spazio d-dimensionale $Pos(v) = (x_1, x_2, \ldots, x_d)$ dove tutti gli x_i sono interi. $x_i(v)$ è la posizione di v nella dimensione i.

- ogni individuo conosce la propria posizione, quella delle conoscenze dirette e quella del destinatario t (routing distribuito)

- il messaggio (la lettera) viene consegnata al conoscente w che è più vicino al destinatario. La misura della distanza $d_M(w,t)$ è la somma delle differenze in valore assoluto

  $|x_i(w)-x_i(t)|$ (Manhattan Distance).

- Gli individui sono in grado di utilizzare un routing distribuito per instradare il messaggio verso la destinazione

SMALL WORLD: ROUTING

- L'esperimento di Milgram suggerisce che la rete "incorpori" una conoscenza che consente di "guidare" il messaggio dalla sorgente alla destinazione, utilizzando ad ogni passo un insieme di conoscenze limitate.

- Milgram "Il messaggio si avvicina dal Nebraska al Massachussets in modo progressivo. Ogni volta che una persona si aggiunge alla catena, il messaggio si avvicina (geograficamente) al target".

Il modello di Watts Strogatz

- garantisce la presenza di cammini brevi tra coppie di nodi, ma non garantisce l'esistenza di un algoritmo di routing decentralizzato che individui tali cammini
- la rete generata da Watts Strogatz non "incorpora" la conoscenza necessaria per definire il routing decentralizzato

- Modello di Kleinberg: rappresenta un'evoluzione del modello di Watts-Strogatz