

## Lezione n.14

# Random Graphs, Small-Worlds, Scale-Free Networks

Materiale didattico:  
Peer-to-Peer Systems  
and Applications  
Capitolo 6

# RIASSUNTO

- 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
- Modelli e reti peer to peer
  - Come modellare le reti P2P?

# INTRODUZIONE

- Diversi fenomeni possono essere modellati mediante reti complesse, caratterizzate da un alto numero di nodi/vertici
- Esempi:
  - Social networks
  - Information Networks
  - Technological Networks
  - Biological Networks
  - .....
- Gli strumenti della teoria dei grafi classica non sono idonei per definire modelli opportuni per questo tipo di fenomeni
- Questo tipo di reti può essere descritto da un insieme di **proprietà statistiche**

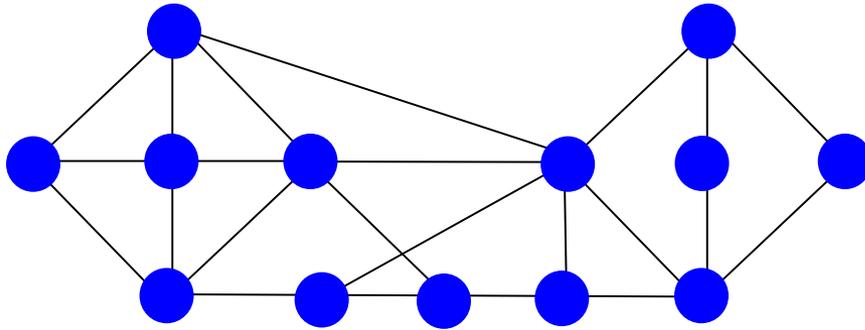
# SOCIAL NETWORKS

Social Networks: gli archi del grafo descrivono le interazioni sociali tra un insieme di individui (vertici del grafo)

- relazioni di amicizia
- relazioni commerciali tra compagnie
- chiamate telefoniche
- collaboration networks:
  - film collaboration network: descrive gli attori che hanno recitato insieme in almeno un film (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“



Sotto questa assunzione, il diametro di una social network cresce come  $O(\sqrt{n})$ .

# SOCIAL NETWORKS: L'ESPERIMENTO DI MILGRAM

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

- ad alcune persone scelte in modo casuale in Kansas e 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

# SOCIAL NETWORKS: L'ESPERIMENTO DI MILGRAM

## COMMUNICATIONS PROJECT

322 ENGLISH HALL HARVARD UNIVERSITY CAMBRIDGE, MASSACHUSETTS 02138

We need your help in an unusual scientific study carried out at Harvard University. We are studying the nature of social contact in American society. Could you, as an active American, contact another American citizen regardless of his work or life? If the name of an American citizen were picked out of a hat, could you get to know that person using only your network of friends and acquaintances? Just how open is our "open society"? To answer these questions, which are very important to our research, we ask for your help.

TARGET PERSON

Name, address, and distance (up about 1000 MILES FROM LA PLEASE) (none)

You will notice that this letter has come to you from a friend. He has asked this study by sending this letter on to you. We hope that you will aid the study by forwarding this letter to someone else. The name of the person who sent you this letter is typed on the Router at the bottom of this sheet.

In the box on the right you will find the name and address of an American citizen who has agreed to serve as the "target person" in this study. The idea of the study is to transmit this letter to the target person using only a chain of friends and acquaintances.

HOW TO TAKE PART IN THIS STUDY

<p style="font-size: 2em; font-weight: bold; text-align: center;">1</p> <p style="font-size: x-small;">ADD YOUR NAME TO THE ROSTER AT THE BOTTOM OF THIS SHEET, so that the other person who receives this letter will know who it came from.</p>	<p style="font-size: 2em; font-weight: bold; text-align: center;">3</p> <p style="font-size: x-small;">IF YOU KNOW THE TARGET PERSON ON A PERSONAL BASIS, MAIL THIS FOLDER DIRECTLY TO HIM (HER). Do this only if you have personally met the target person and know each other on a first name basis.</p>
<p style="font-size: 2em; font-weight: bold; text-align: center;">2</p> <p style="font-size: x-small;">DETACH ONE POSTCARD, FILL IT OUT AND RETURN IT TO HARVARD UNIVERSITY. No stamp is needed. The postcard is very important. It allows us to keep track of the progress of the letter as it moves toward the target person.</p>	<p style="font-size: 2em; font-weight: bold; text-align: center;">4</p> <p style="font-size: x-small;">IF YOU DO NOT KNOW THE TARGET PERSON ON A PERSONAL BASIS, DO NOT TRY TO CONTACT HIM DIRECTLY. INSTEAD, MAIL THIS FOLDER (POST CARDS AND ALL) TO A PERSONAL ACQUAINTANCE WHO IS MORE LIKELY THAN YOU TO KNOW THE TARGET PERSON. You may send the folder on to a friend, relative, or acquaintance, but not to the someone you know on a first name basis.</p>

Remember, the aim is to move this letter toward the target person using only a chain of friends and acquaintances. Do first thought you may feel you do not know anyone who is acquainted with the target person. This is correct, but at least you can start in the right direction! Who among your acquaintances might conceivably have in the same social circles as the target person? The real challenge is to identify among your friends and acquaintances a person who can advance the folder toward the target person. It may take several steps beyond your friend to get to the target person, but what counts most is to start the folder on its way! The person who receives this folder will then repeat the process until the folder is received by the target person. May we ask you to begin!

Every person who participates in this study and returns the post card to us will receive a certificate of appreciation from the Communications Project. All participants are entitled to a report describing the results of the study.

Please transmit this folder within 24 hours. Your help is greatly appreciated.

Very sincerely,  
  
 Stanley Milgram, Ph.D.  
 Director, Communications Project

<p style="text-align: center; font-weight: bold; font-size: small;">ROSTER</p> <p style="font-size: x-small;">1 _____</p> <p style="font-size: x-small;">2 _____</p> <p style="font-size: x-small;">3 _____</p> <p style="font-size: x-small;">4 _____</p> <p style="font-size: x-small;">5 _____</p> <p style="font-size: x-small;">6 _____</p> <p style="font-size: x-small;">7 _____</p> <p style="font-size: x-small;">8 _____</p> <p style="font-size: x-small;">9 _____</p> <p style="font-size: x-small;">10 _____</p> <p style="font-size: x-small;">11 _____</p> <p style="font-size: x-small;">12 _____</p> <p style="font-size: x-small;">13 _____</p> <p style="font-size: x-small;">14 _____</p> <p style="font-size: x-small;">15 _____</p> <p style="font-size: x-small;">16 _____</p> <p style="font-size: x-small;">17 _____</p> <p style="font-size: x-small;">18 _____</p> <p style="font-size: x-small;">19 _____</p> <p style="font-size: x-small;">20 _____</p> <p style="font-size: x-small; text-align: center;">SIGN YOUR NAME HERE.</p>	<p style="font-size: x-small; text-align: center;">PLEASE FILL IN THE INFORMATION ABOUT YOURSELF</p> <p style="font-size: x-small;">MR NAME _____</p> <p style="font-size: x-small;">MR ADDRESS _____</p> <p style="font-size: x-small;">MR OCCUPATION _____</p> <p style="font-size: x-small;">MR SEX _____</p> <p style="font-size: x-small; text-align: center;">DETACH ONE POSTCARD. FILL IT OUT AND RETURN IT TO HARVARD UNIVERSITY.</p>	<p style="font-size: x-small; text-align: center;">PLEASE FILL IN THE FOLLOWING INFORMATION ABOUT THE PERSON TO WHOM YOU ARE SENDING THE FOLDER</p> <p style="font-size: x-small;">MR NAME _____</p> <p style="font-size: x-small;">MR ADDRESS _____</p> <p style="font-size: x-small;">MR OCCUPATION _____</p> <p style="font-size: x-small;">APPROXIMATE AGE _____ SEX _____</p> <p style="font-size: x-small;">NATURE OF HIS RELATIONSHIP TO YOU _____</p> <p style="font-size: x-small; text-align: center;">PLEASE EXPLAIN FACTORS</p> <p style="font-size: x-small;">HE IS A FRIEND, ACQUAINTANCE, RELATIVE, ETC.</p>
---	---	---

# L'ESPERIMENTO DI MILGRAM: I RISULTATI

- Milgram calcolò il numero medio di "passaggi di mano,, di ogni lettera che aveva raggiunto il destinatario
- Ogni lettera arrivata a destinazione aveva richiesto non più di 6 passaggi (6 degree of separation)
- 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'ESPERIMENTO DI MILGRAM

- 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 venir incontro alla necessità di modelli per 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
  - Watts-Strogatz
  - Kleinberg
  - Barabasi-Albert

# ALTRI ESEMPI DI RETI COMPLESSE

## Information Networks (Knowledge Networks) (Eugene Garfield):

- Rete definita dai riferimenti (citazioni) presenti tra pubblicazioni accademiche: nodi= articoli pubblicati, archi= citazioni
- World Wide Web: nodi: pagine web, archi= iperlinks che collegano le pagine
- P2P Networks

## Technological Networks

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

## Biological Networks

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

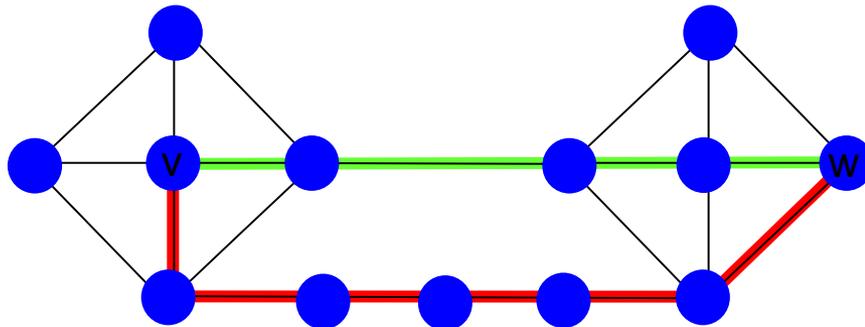
# GRAFI: DEFINIZIONI

Dato un grafo  $G = (V, E)$

- outdegree  $\deg_+(v)$  of un nodo  $v$  = numero di vertici  $w$  connessi a  $v$  da un arco  $(v, w)$
- indegree  $\deg_-(v)$  of  $v$  = numero di vertici  $w$  connessi a  $v$  da un arco  $(w, v)$
- $\text{degree}(v) = \deg_+(v) + \deg_-(v)$
- Clique= sottografo in cui ogni nodo è connesso ad ogni altro nodo

# GRAFI: DEFINIZIONI

- cammino  $P(v,w)$  è un insieme di vertici  $\{v=v_0, v_1, \dots, v_k=w\}$  con  $(v_i, v_{i+1})$  appartenenti ad  $E$   $0 \leq i \leq k$
- la lunghezza del cammino  $|P(v,w)|$  è definita come il numero di archi in  $P$  (si considerano archi non pesati)
- *distanza*  $d(v,w)$ : la lunghezza del cammino minimo tra  $v$  e  $w$
- diametro  $D(G)$  di un grafo  $G$ : massima distanza esistente tra due nodi qualsiasi del grafo



— Un cammino di lunghezza 6

— Il cammino minimo tra  $v$  e  $w$  ha lunghezza 4

La distanza tra  $v$  e  $w$  è 4

# RETI COMPLESSE: SMALL WORLDS

- **Small World Network** = una qualsiasi coppia di nodi è collegata da un cammino caratterizzato da un numero molto limitato di hops
- Si considera  $l$ , la distanza media tra coppie di nodi del grafo

$$l = 1/(n(n-1)/2) \sum_{i > j} d_{ij}$$

dove  $d_{ij}$  è la distanza tra il vertice  $i$  ed il vertice  $j$

- 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

# RETI COMPLESSE: RISULTATI SPERIMENTALI

attori	non orientata	449913	25516482	3.48
coautori	non orientata	52909	245300	6.19
food web	orientata	135	598	2.05
WWW	orientata	203549046	2130000000	16.18
internet	non orientata	10697	31992	3.31
P2P	non orientata	880	1296	4.28

rete

tipo

nodi

archi

$\lambda$

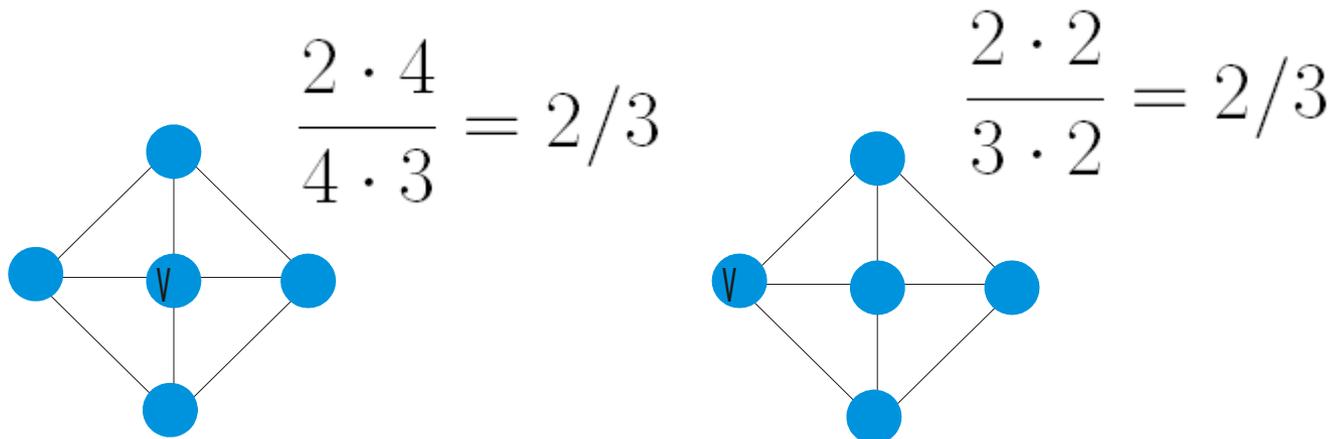


# CLUSTERIZZAZIONE: DEFINIZIONE 1

Coefficiente 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  $v$

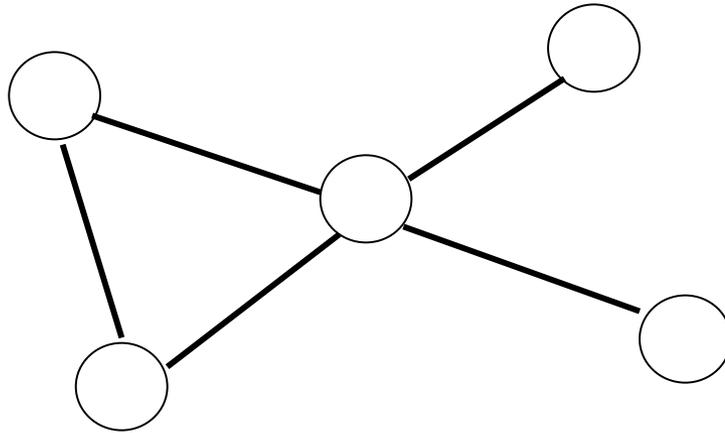


$C(v)$  misura 'quanto i vicini di  $v$  formano strutture 'di tipo clique'

# 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  
 $C = (3 \cdot \text{numero di triangoli nella rete}) / \text{numero di triple connesse}$   
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  $C = 3 * 1/8 = 3/8$

# CLUSTERIZZAZIONE: DEFINIZIONE 3

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

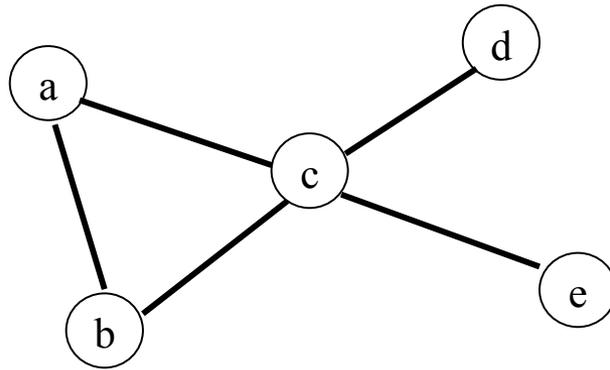
$C_i$  = numero di triangoli connessi ad  $i$  / numero di triple 'centrate' su  $i$

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

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

# CLUSTERIZZAZIONE: DEFINIZIONE 3



Coefficienti di clusterizzazione associati ai nodi:

$$C_a = 1$$

$$C_b = 1$$

$$C_c = 1/6$$

$$C_d = 0$$

$$C_e = 0$$

Coefficiente di clusterizzazione del grafo:

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

# RETI COMPLESSE: RISULTATI SPERIMENTALI

attori	non orientata	449913	25516482	3.48	0.20	0.78
coautori	non orientata	52909	245300	6.19	0.45	0.56
Food web	orientata	135	598	2.05	0.16	0.23
WWW	orientata	203549046	2130000000	16.18		
internet	non orientata	10697	31992	3.31	0.035	0.39
P2P	non orientata	880	1296	4.28	0.012	0.011

rete

tipo

nodi

archi

|

coeff.clust.



# RETI COMPLESSE: PROPRIETA'

## Altre proprietà delle reti complesse

- Distribuzione del grado dei nodi  
Grado dei nodi (scale free networks, power laws,...)
- Network resilience
- Community Structures
- ....vedere materiale consigliato

# 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

- *Erdős-Renyi*
- *Gilbert*

Si ricavano *proprietà* statistiche. Una proprietà si verifica con una certa probabilità. Una proprietà si verifica con alta probabilità, se la probabilità che non si verifichi è  $O(1/n)$ .

# GRAFI RANDOM: IL MODELLO DI ERDOS-RENYI

- Modello estremamente semplice: costruisce un grafo  $G_{n,p}$  di  $n$  vertici mediante la seguente procedura
  - 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  $n$  compreso tra 0 ed 1, se  $n < p$ , si definisce un arco tra  $u$  e  $v$
- Distribuzione binomiale: la presenza/assenza di un arco è indipendente da quella degli altri archi
- Poisson Random Graph

# IL MODELLO DI ERDOS-RENYI: PROPRIETA'

- Consideriamo un grafo  $G(n,m)$  con  $n$  vertici ed  $m$  archi
- $G(n,m)$  è con alta probabilità un grafo connesso se  $m$  è maggiore di  $\log n$ .
- Il diametro di  $G(n,m)$  è  $O(\log n)$  con alta probabilità se il grafo è connesso  
se  $m$  è maggiore di  $\log n$   
un random graph riproduce l'effetto 'small world'
- Coefficiente di clusterizzazione  $C=p$   
la probabilità di connettere due vertici è  $p$ , ed è indipendente dal fatto che i due vertici abbiano un vicino in comune
- Distribuzione del grado dei nodi = distribuzione di Poisson (binomiale), non corrisponde alle distribuzioni rilevate nelle reti reali

# IL MODELLO DI WATTS-STROGATZ

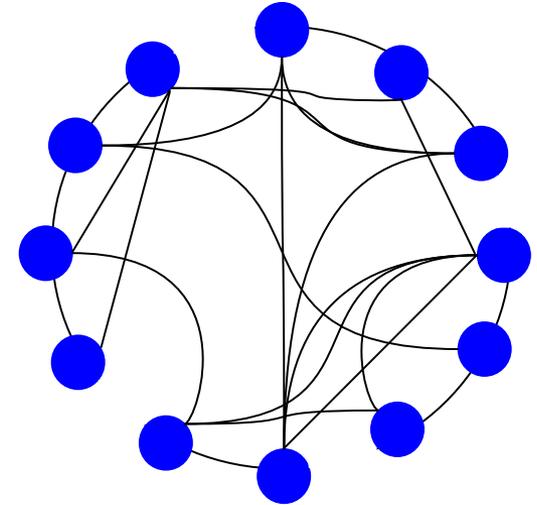
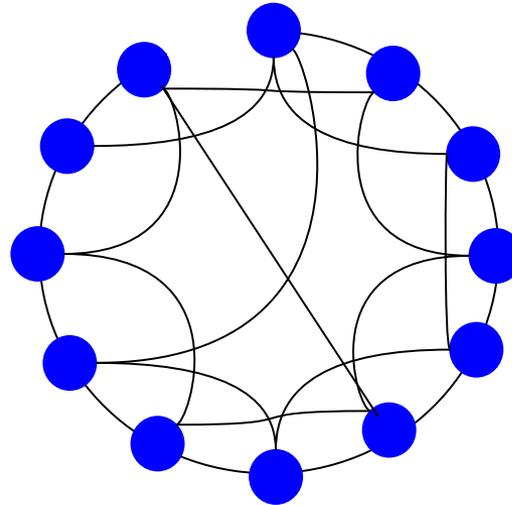
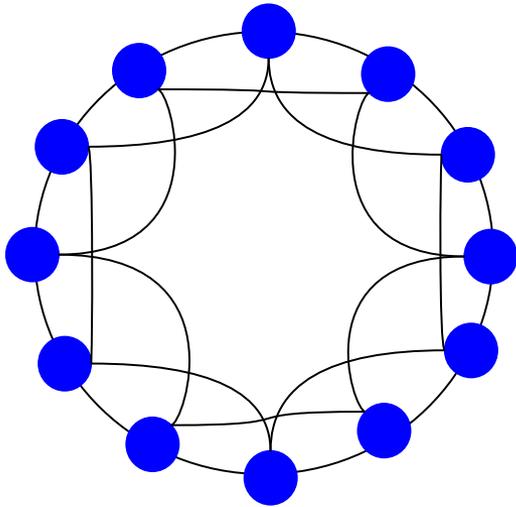
- Watts-Strogatz considerano due caratteristiche fondamentali delle reti complesse:
  - il coefficiente di clusterizzazione: misura la regolarità e la località della rete. Se tale coefficiente è alto, gli archi collegano principalmente nodi vicini, piuttosto che vertici lontani.
  - la distanza tra i vertici.
- Se il coefficiente 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 coefficiente di clusterizzazione (0.3-0.4) e un valore basso della distanza tra nodi

# 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

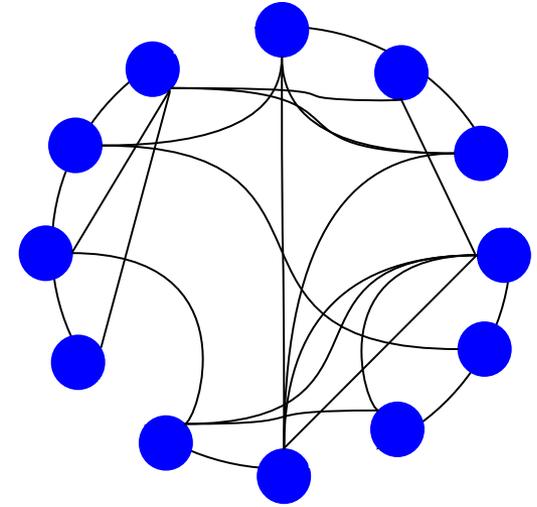
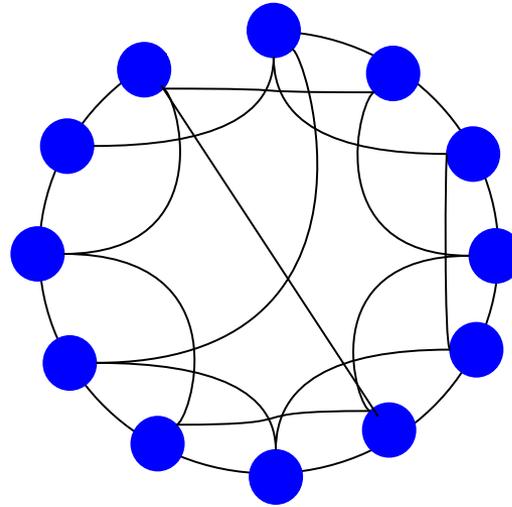
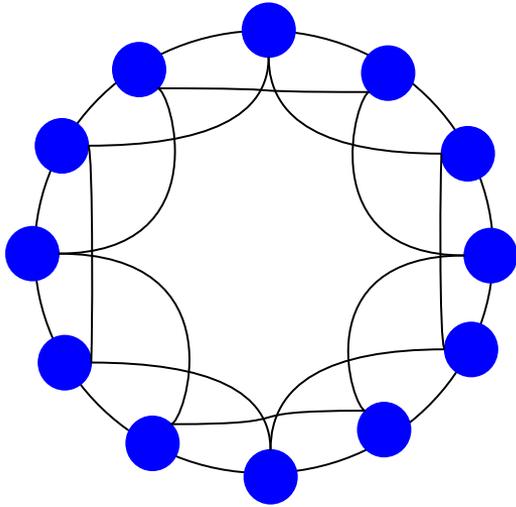
# IL MODELLO DI WATTS-STROGATZ



## Procedura di costruzione del grafo

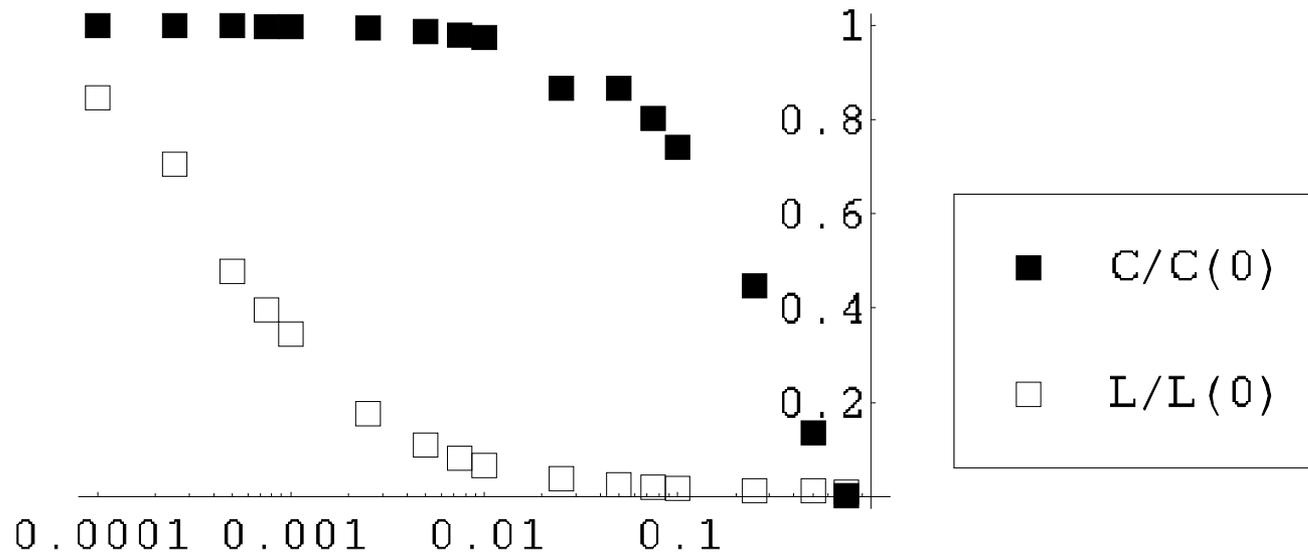
- Si parte da un anello di  $n$  vertici e si connette ogni vertice con i suoi  $k$  vicini sull'anello
- Si „riavvolge“ un arco  $(v,w)$  con probabilità uguale a  $p$ : si mantiene fisso  $v$  e si sceglie un nuovo vertice  $w'$ , scelto tra tutti i vertici, in modo casuale

# IL MODELLO DI WATTS-STROGATZ



- Per  $p=0$ , la rete risultante è completamente regolare, con un coefficiente di clusterizzazione di circa  $\frac{3}{4}$  per valori grandi di  $k$  ed un diametro è  $O(n)$
- Per  $p=1$ , la rete risultante si può considerare un garfo random regolare con un coefficiente di clusterizzazione pari a  $p$  ed un diametro  $O(\log n)$
- Casi interessanti: valori intermedi di  $p$

# IL MODELLO DI WATTS-STROGATZ



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