Frontiers of Network Science
Fall 2023

Class 3: Graph Theory & Random Networks
(Chapter 2-3 in Textbook)

Boleslaw Szymanski

based on slides by Albert-László Barabási & Roberta Sinatra

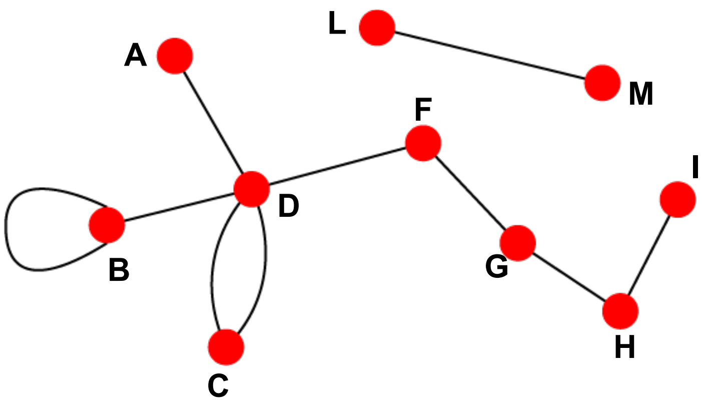
MEASURES

UNDIRECTED VS. DIRECTED NETWORKS

Undirected

Links: undirected (*symmetrical*)

Graph:



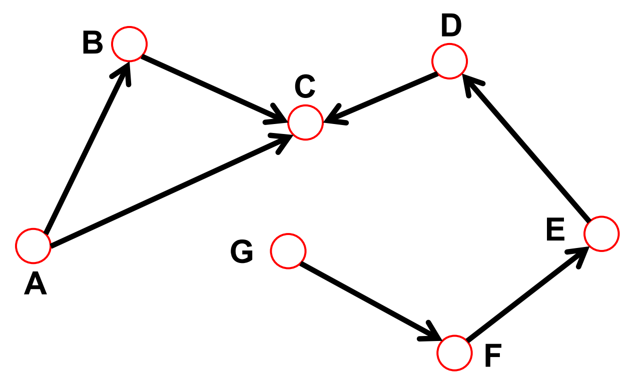
Undirected links :

coauthorship links
Actor network
protein interactions

Directed

Links: directed (*arcs*).

Digraph = directed graph:



An undirected link is the superposition of two opposite directed links.

Directed links :

URLs on the www
phone calls
metabolic reactions

Section 2.2 Reference Networks

NETWORK	NODES	LINKS	DIRECTED UNDIRECTED	N	L
Internet	Routers	Internet connections	Undirected	192,244	609,066
WWW	Webpages	Links	Directed	325,729	1,497,134
Power Grid	Power plants, transformers	Cables	Undirected	4,941	6,594
Mobile Phone Calls	Subscribers	Calls	Directed	36,595	91,826
Email	Email addresses	Emails	Directed	57,194	103,731
Science Collaboration	Scientists	Co-authorship	Undirected	23,133	93,439
Actor Network	Actors	Co-acting	Undirected	702,388	29,397,908
Citation Network	Paper	Citations	Directed	449,673	4,689,479
E. Coli Metabolism	Metabolites	Chemical reactions	Directed	1,039	5,802
Protein Interactions	Proteins	Binding interactions	Undirected	2,018	2,930

Degree, Average Degree and Degree Distribution

A BIT OF STATISTICS

BRIEF STATISTICS REVIEW

Four key quantities characterize a sample of N values $x_1, \ldots, x_N$:

Average (mean):

$$\langle x \rangle = \frac{x_1 + x_2 + \ldots + x_N}{N} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

The n^{th} moment:

$$\langle x^n \rangle = \frac{x_1^n + x_2^n + \ldots + x_N^n}{N} = \frac{1}{N}\sum_{i=1}^{N} x_i^n$$

Standard deviation:

$$\sigma_x = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(x_i - \langle x \rangle\right)^2}$$

Distribution of x:

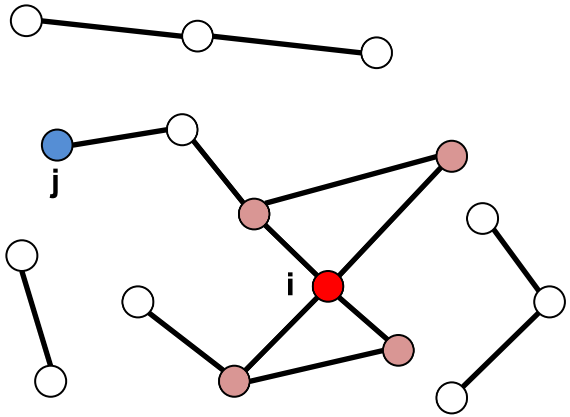
$$p_x = \frac{1}{N}\sum_i \delta_{x,x_i}$$

where p_x follows

$$\sum_i p_x = 1 \left(\int p_x \, dx = 1\right)$$

AVERAGE DEGREE

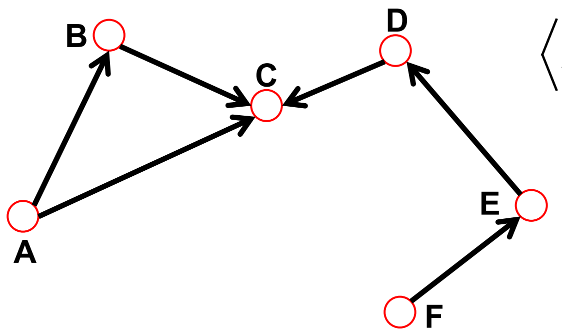
Undirected



$$\langle k \rangle \equiv \frac{1}{N}\sum_{i=1}^{N} k_i \qquad \langle k \rangle \equiv \frac{2L}{N}$$

N – the number of nodes in the graph

Directed

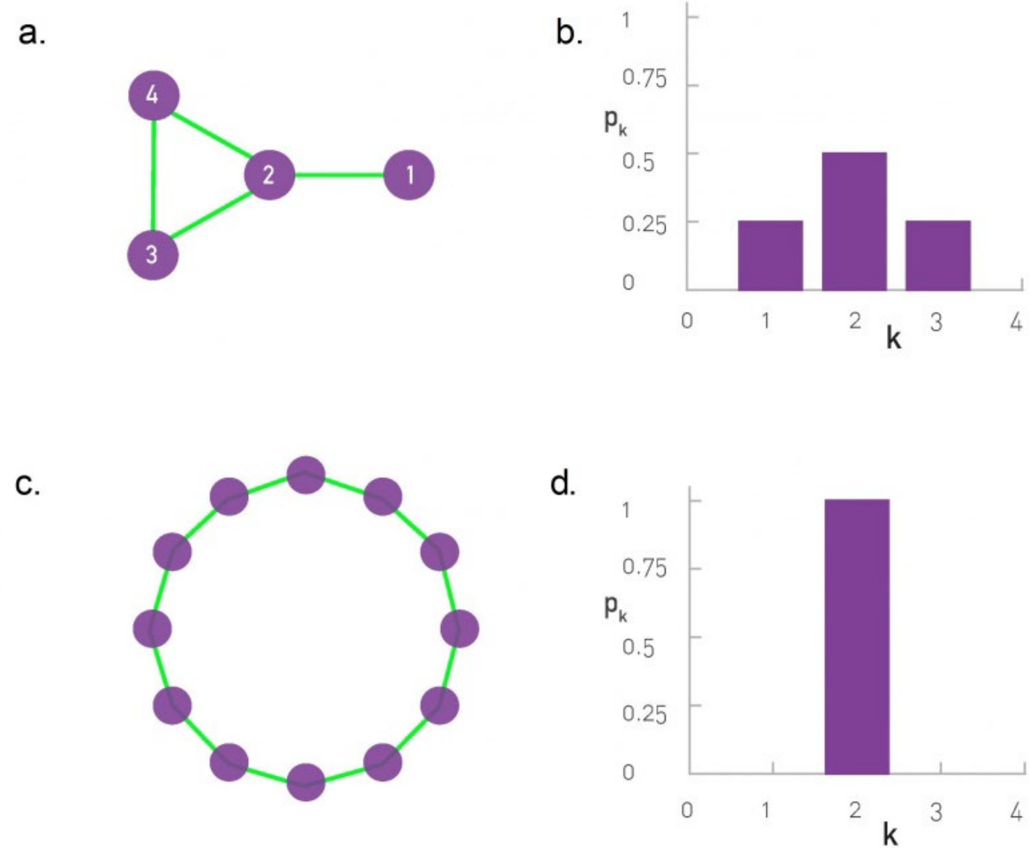


$$\langle k^{in} \rangle \equiv \frac{1}{N}\sum_{i=1}^{N} k_i^{in}, \quad \langle k^{out} \rangle \equiv \frac{1}{N}\sum_{i=1}^{N} k_i^{out}, \quad \langle k^{in} \rangle = \langle k^{out} \rangle$$

$$\langle k \rangle \equiv \frac{L}{N}$$

Degree distribution

P(k): probability that a randomly chosen node has degree *k*



N_k = # nodes with degree k

$P(k) = N_k / N$ ❾ plot

DEGREE DISTRIBUTION

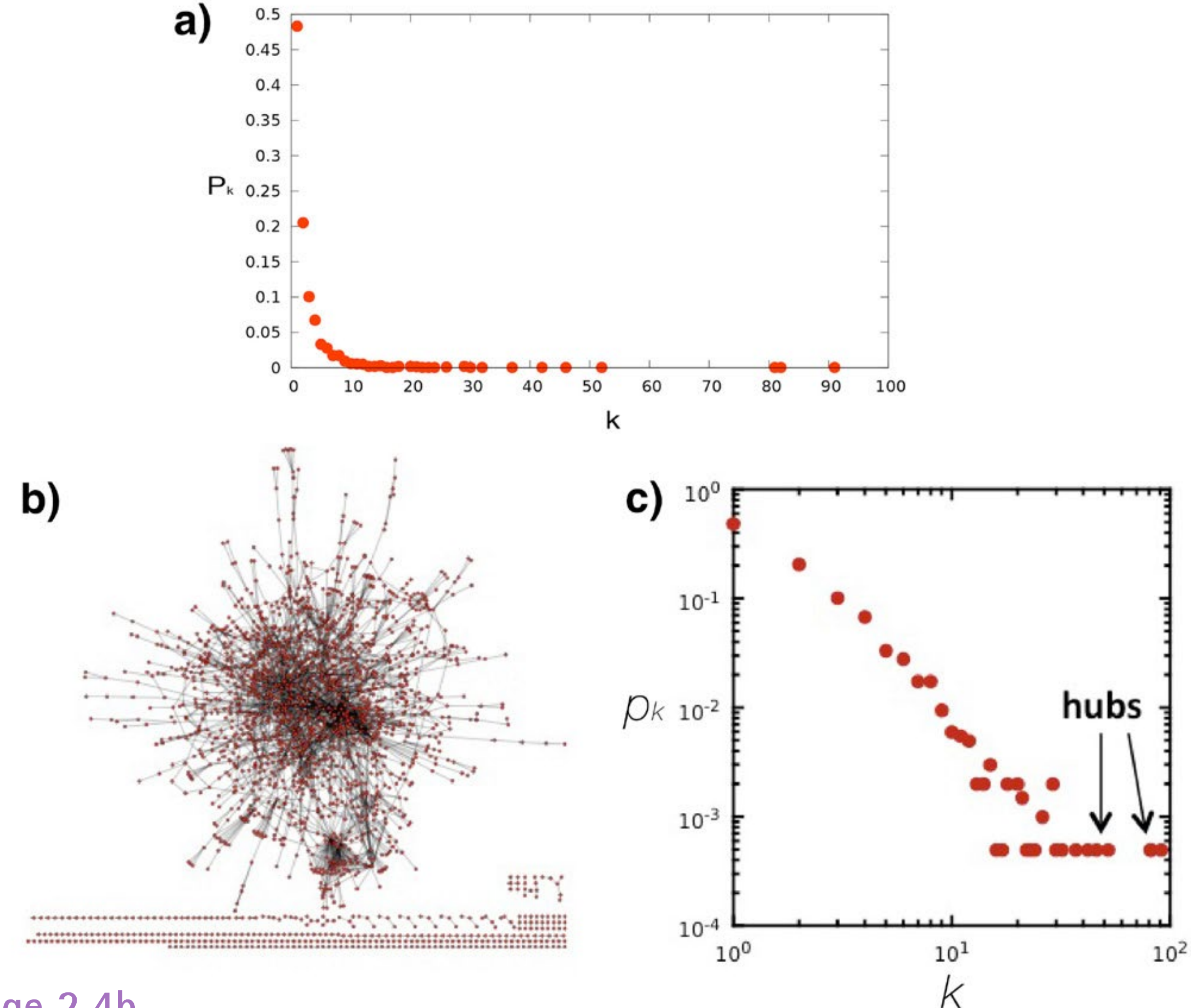


Image 2.4b

DEGREE DISTRIBUTION

Discrete Representation: $\mathbf{p_k}$ is the probability that a node has degree **k**.

Continuum Description: **p(k)** is the pdf of the degrees, where

$$\int_{k_1}^{k_2} p(k)dk$$

represents the probability that a node's degree is between $\mathbf{k_1}$ and $\mathbf{k_2}$.

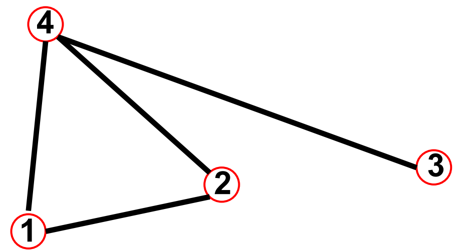
Normalization condition:

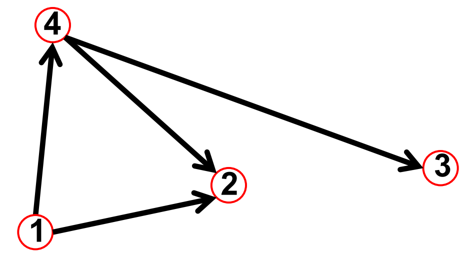
$$\sum_{0}^{\infty} p_k = 1 \qquad \int_{K_{min}}^{\infty} p(k)dk = 1$$

where K_{min} is the minimal degree in the network.

Adjacency matrix

ADJACENCY MATRIX





A_{ij}=1 if there is a link between node *i* and *j*

A_{ij}=0 if nodes *i* and *j* are not connected to each other.

$$A_{ij} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \qquad A_{ij} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

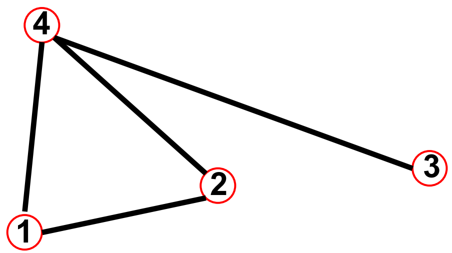
Note that for a directed graph (right) the matrix is not symmetric.

$A_{ij} = 1$ if there is a link pointing from node *j* and *i*

$A_{ij} = 0$ if there is no link pointing from *j* to *i*.

ADJACENCY MATRIX AND NODE DEGREES

Undirected



$$A_{ij} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

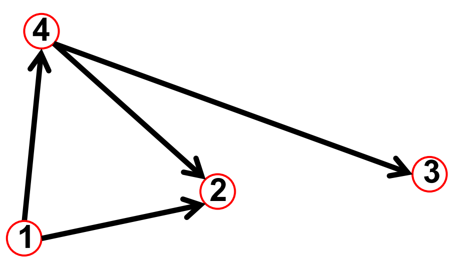
$$A_{ij} = A_{ji}$$
$$A_{ii} = 0$$

$$k_i = \sum_{j=1}^{N} A_{ij}$$

$$k_j = \sum_{i=1}^{N} A_{ij}$$

$$L = \frac{1}{2}\sum_{i=1}^{N} k_i = \frac{1}{2}\sum_{ij}^{N} A_{ij}$$

Directed



$$A_{ij} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

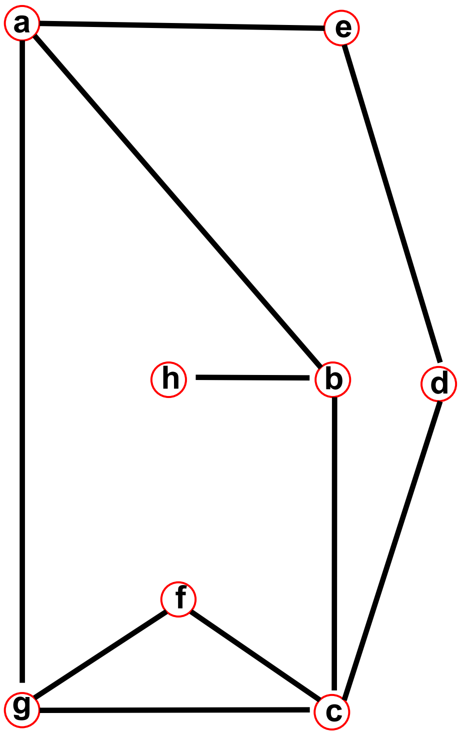
$$A_{ij} \neq A_{ji}$$
$$A_{ii} = 0$$

$$k_i^{in} = \sum_{j=1}^{N} A_{ij}$$

$$k_j^{out} = \sum_{i=1}^{N} A_{ij}$$

$$L = \sum_{i=1}^{N} k_i^{in} = \sum_{j=1}^{N} k_j^{out} = \sum_{i,j}^{N} A_{ij}$$

ADJACENCY MATRIX

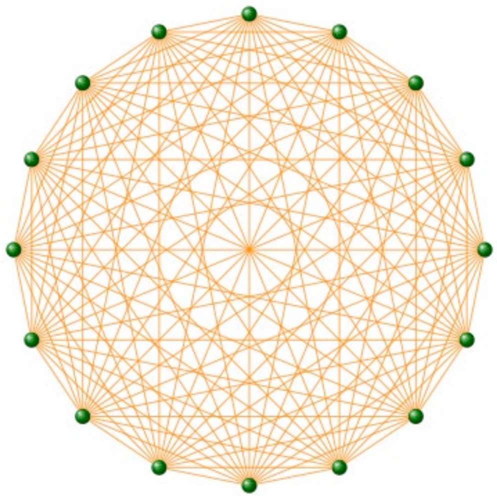


	a	b	c	d	e	f	g	h
a	0	1	0	0	1	0	1	0
b	1	0	1	0	0	0	0	1
c	0	1	0	1	0	1	1	0
d	0	0	1	0	1	0	0	0
e	1	0	0	1	0	0	0	0
f	0	0	1	0	0	0	1	0
g	1	0	1	0	0	0	0	0
h	0	1	0	0	0	0	0	0

Real networks are sparse

COMPLETE GRAPH

The maximum number of links a network of N nodes can have is: $L_{max} = \binom{N}{2} = \frac{N(N-1)}{2}$



A graph with degree $L=L_{max}$ is called a complete graph, and its average degree is **$\langle k \rangle = N-1$**

REAL NETWORKS ARE SPARSE

Most networks observed in real systems are sparse:

$$L \ll L_{max}$$

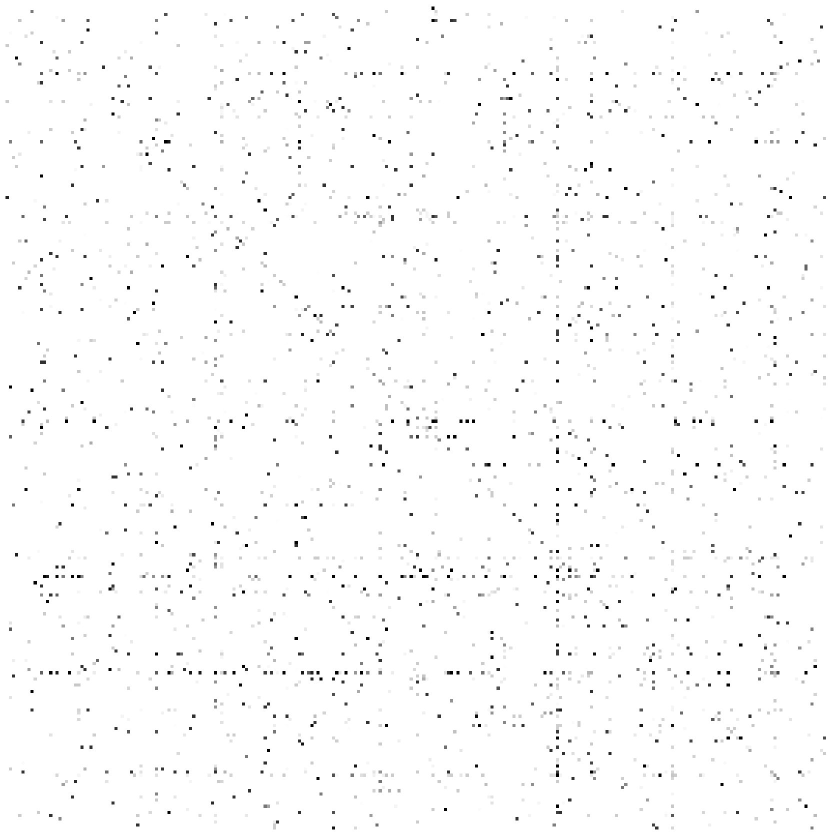
or

$$\langle k \rangle \ll N-1.$$

WWW (ND Sample):	N=325,729;	$L=1.4\ 10^6$	$L_{max}=10^{12}$	$\langle k \rangle=4.51$
Protein (*S. Cerevisiae*):	N= 1,870;	L=4,470	$L_{max}=10^7$	$\langle k \rangle=2.39$
Coauthorship (Math):	N= 70,975;	$L=2\ 10^5$	$L_{max}=3\ 10^{10}$	$\langle k \rangle=3.9$
Movie Actors:	N=212,250;	$L=6\ 10^6$	$L_{max}=1.8\ 10^{13}$	$\langle k \rangle=28.78$

(Source: Albert, Barabasi, RMP2002)

ADJACENCY MATRICES ARE SPARSE



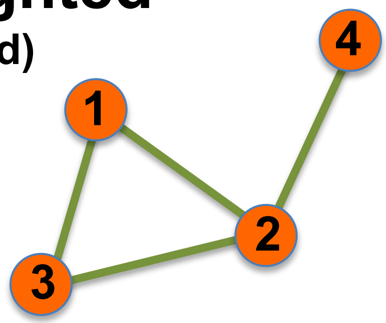
WEIGHTED AND UNWEIGHTED NETWORKS

WEIGHTED AND UNWEIGHTED NETWORKS

$$\boldsymbol{A_{ij} = w_{ij}}$$

Unweighted

(undirected)



$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

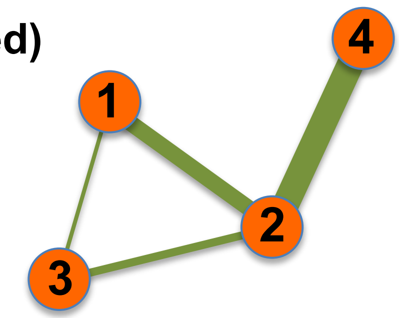
$$A_{ii} = 0 \qquad A_{ij} = A_{ji}$$

$$L = \frac{1}{2}\sum_{i,j=1}^{N} A_{ij} \qquad <k>= \frac{2L}{N}$$

protein-protein interactions, www

Weighted

(undirected)



$$A_{ij} = \begin{pmatrix} 0 & 2 & 0.5 & 0 \\ 2 & 0 & 1 & 4 \\ 0.5 & 1 & 0 & 0 \\ 0 & 4 & 0 & 0 \end{pmatrix}$$

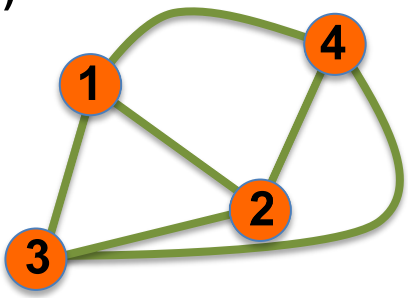
$$A_{ii} = 0 \qquad A_{ij} = A_{ji}$$

$$L = \frac{1}{2}\sum_{i,j=1}^{N} nonzero(A_{ij}) \qquad <k>= \frac{2L}{N}$$

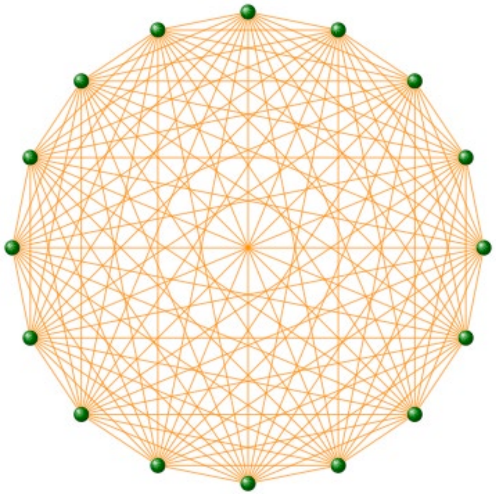
Call Graph, metabolic networks

Complete Graph

(undirected)



$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

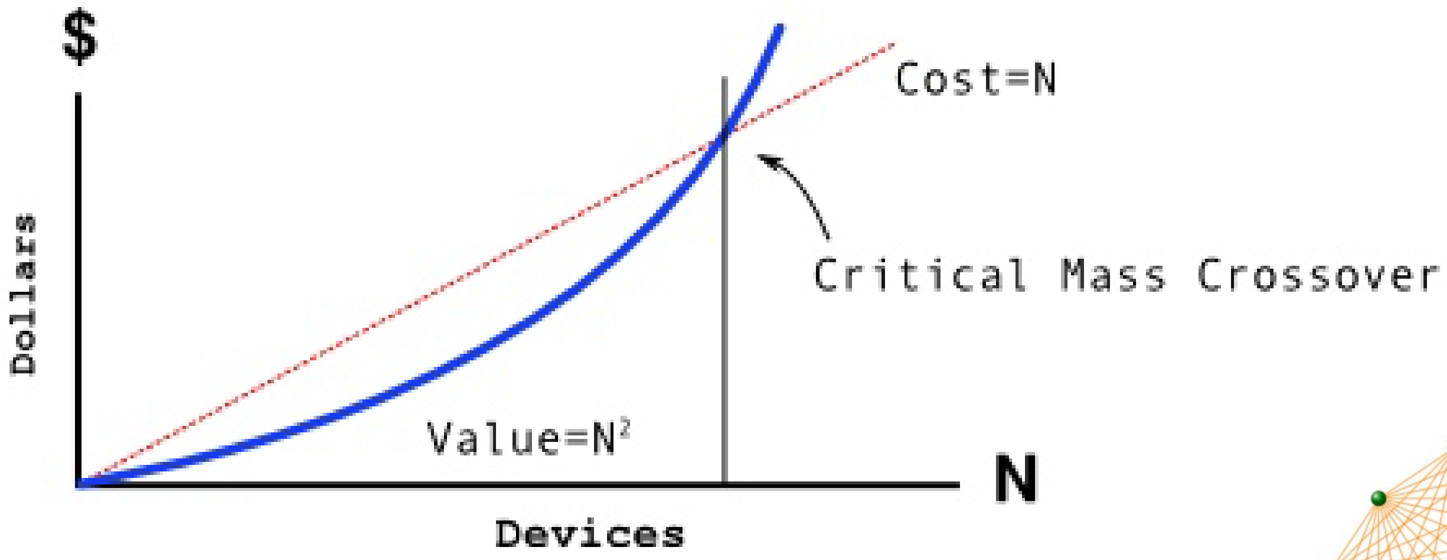


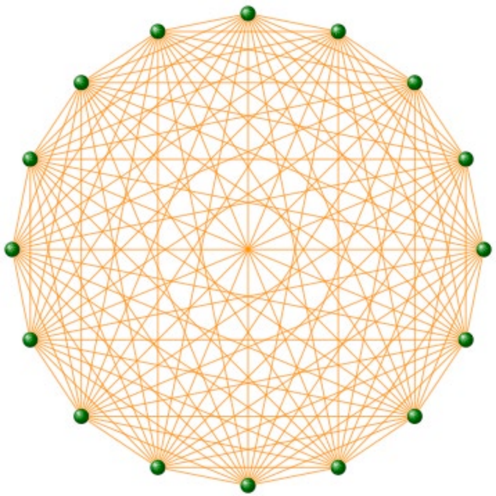
$$A_{ii} = 0 \qquad A_{i \neq j} = 1$$

$$L = L_{\max} = \frac{N(N-1)}{2} \qquad < k >= N-1$$

Actor network, protein-protein interactions

METCALFE'S LAW



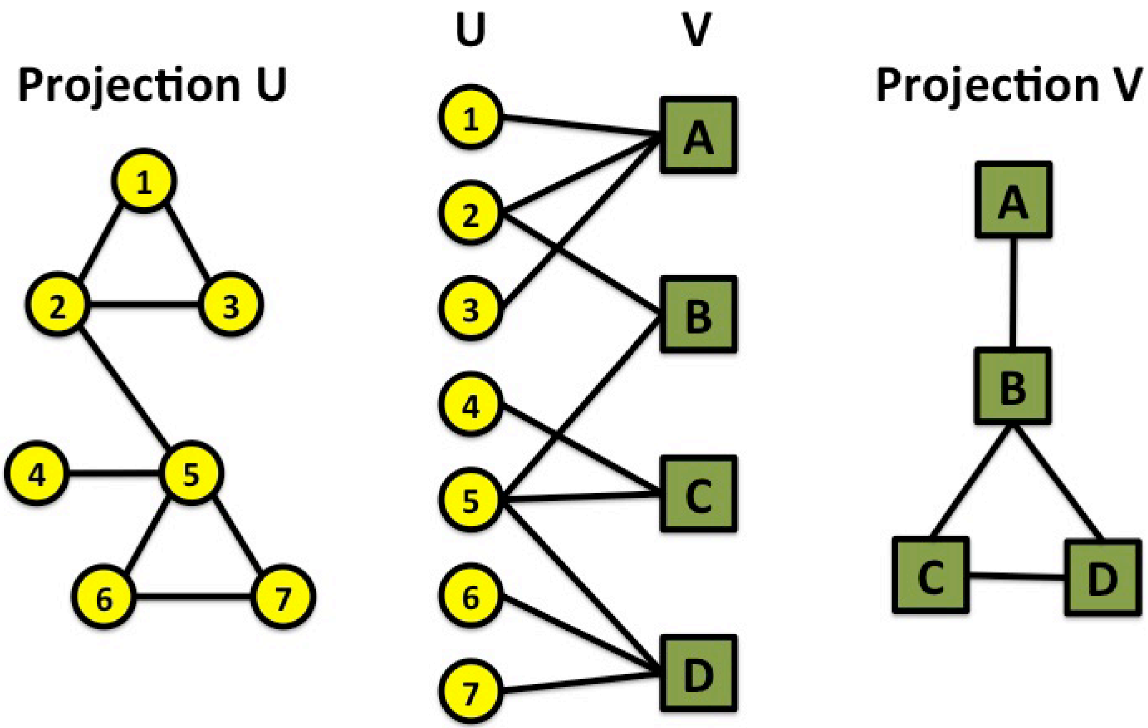


The maximum number of links a network of N nodes can have is: $L_{\max} = \binom{N}{2} = \frac{N(N-1)}{2}$

BIPARTITE NETWORKS

BIPARTITE GRAPHS

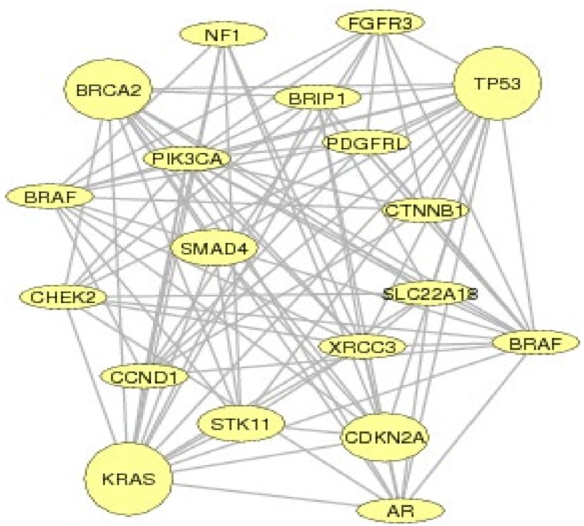
bipartite graph (or **bigraph**) is a graph whose nodes can be divided into two disjoint sets *U* and *V* such that every link connects a node in *U* to one in *V*; that is, *U* and *V* are independent sets.



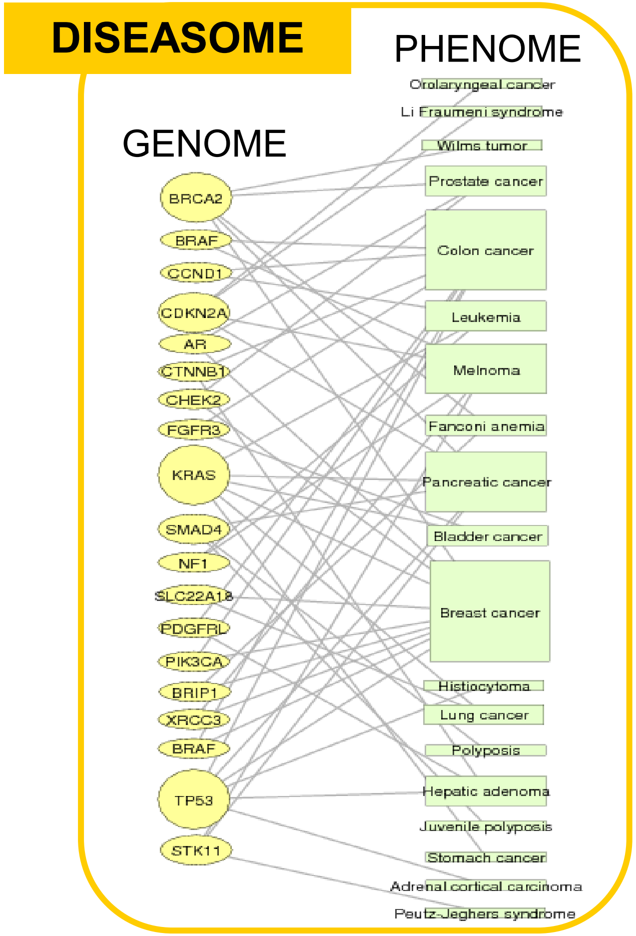
Examples:

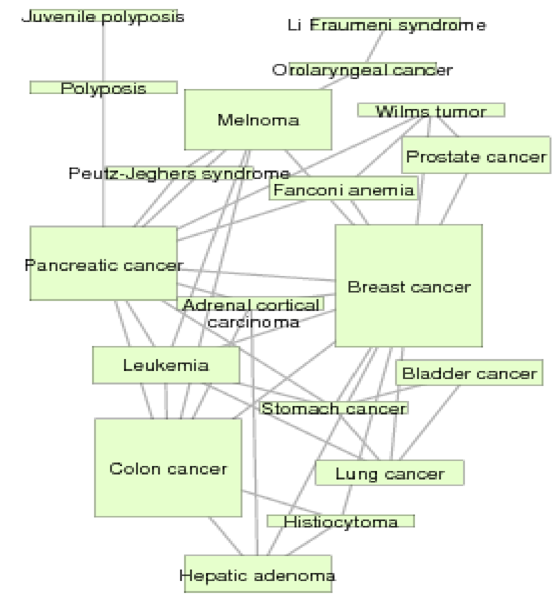
Hollywood actor network
Collaboration networks
Disease network (diseasome)

GENE NETWORK – DISEASE NETWORK



Gene network

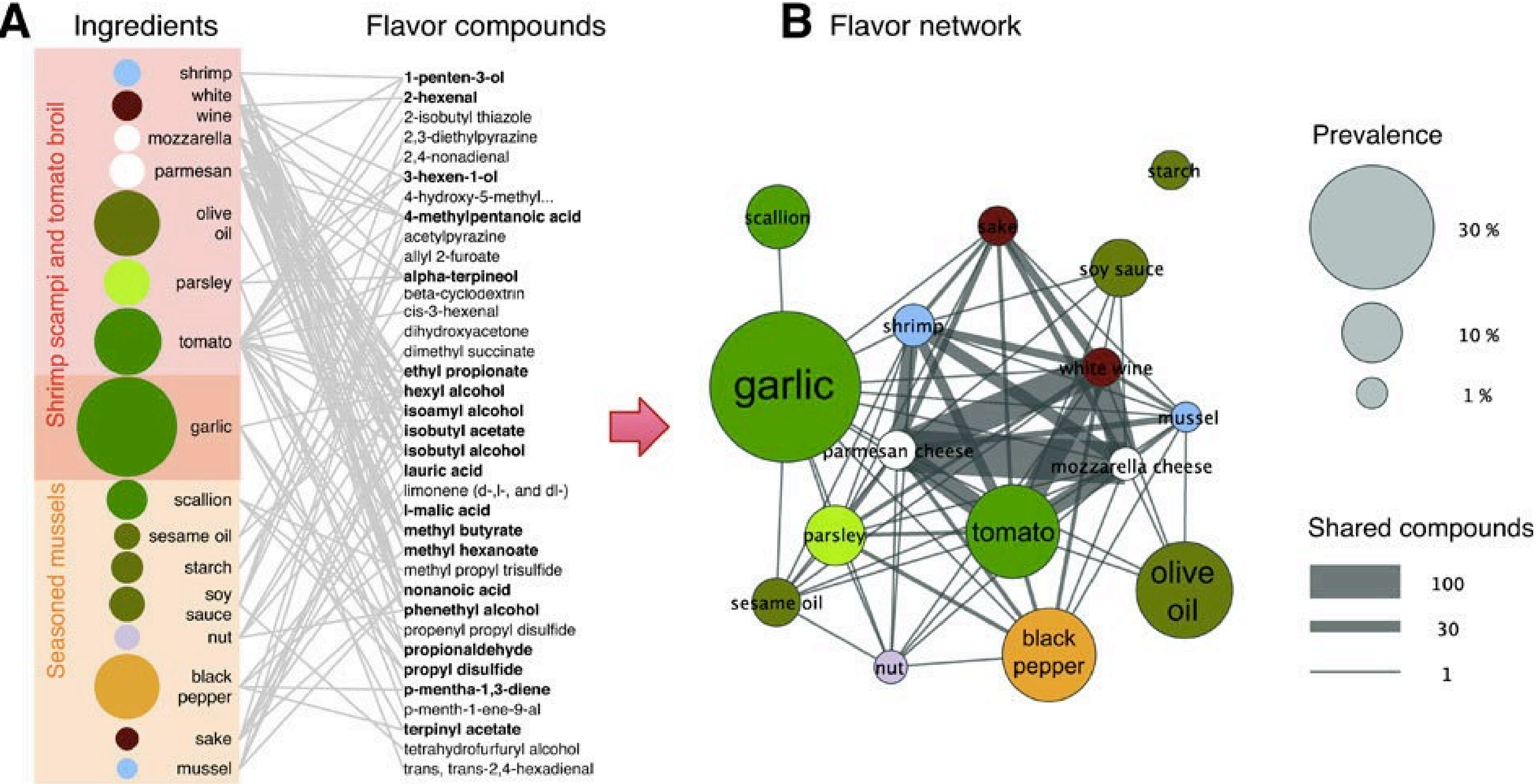




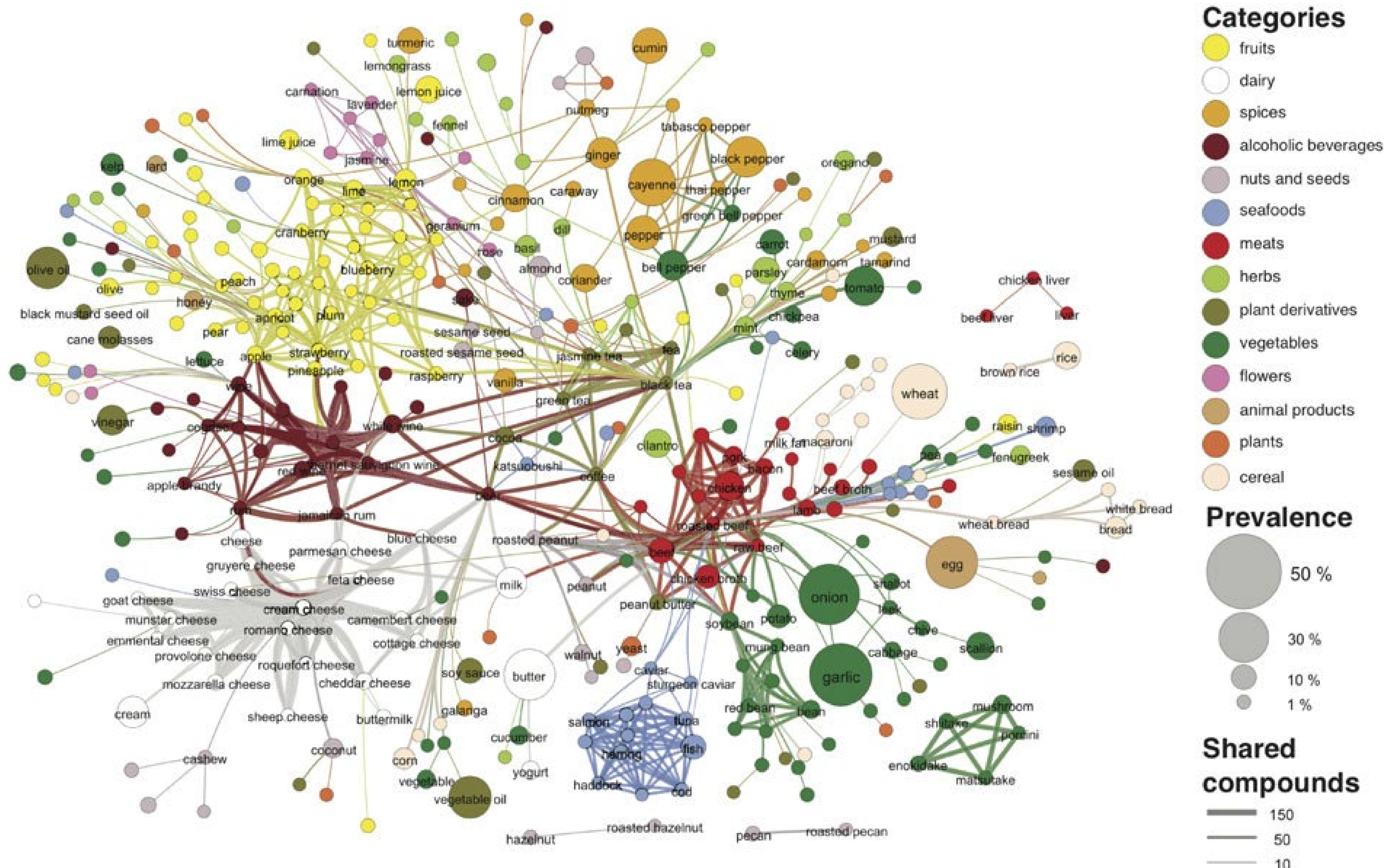
Disease network

Goh, Cusick, Valle, Childs, Vidal & Barabási, PNAS (2007)

Ingredient-Flavor Bipartite Network



Y.-Y. Ahn, S. E. Ahnert, J. P. Bagrow, A.-L. Barabási Flavor network and the principles of food pairing , Scientific Reports 196, (2011).



Categories
fruits
dairy
spices
alcoholic beverages
nuts and seeds
seafoods
meats
herbs
plant derivatives
vegetables
flowers
animal products
plants
cereal
Prevalence
50 %
30 %
10 %
1 %
Shared compounds
150
50
10
turmeric
cumin
lemongrass
carnation
lavender
lemon juice
nutmeg
fennel
tabasco pepper
lime juice
jasmine
ginger
black pepper
kelp
lard
orange
lime
lemon
cayenne
caraway
thai pepper
oregano
cinnamon
green bell pepper
cranberry
geranium
dill
pepper
mustard
rose
basil
carrot
olive oil
almond
coriander
bell pepper
cardamom
tamarind
blueberry
parsley
peach
olive
tomato
chicken liver
thyme
honey
sake
black mustard seed oil
apricot
plum
chickpea
beef liver
liver
sesame seed
mint
cane molasses
pear
strawberry
roasted sesame seed
jasmine tea
tea
celery
lettuce
apple
pineapple
raspberry
brown rice
rice
vanilla
black tea
wheat
green tea
vinegar
white wine
cocoa
cilantro
milk fat
macaroni
raisin
shrimp
pork
bacon
pea
fenugreek
katsuobushi
sesame oil
apple brandy
coffee
chicken
beef broth
rum
jamaican rum
lamb
white bread
roasted beef
wheat bread
bread
cheese
blue cheese
roasted peanut
parmesan cheese
beef
raw beef
egg
gruyere cheese
feta cheese
chicken broth
milk
peanut
shallot
swiss cheese
onion
goat cheese
cream cheese
peanut butter
leek
munster cheese
camembert cheese
soybean
potato
chive
emmental cheese
romano cheese
cottage cheese
walnut
yeast
mung bean
cabbage
scallion
provolone cheese
roquefort cheese
caviar
garlic
mozzarella cheese
cheddar cheese
soy sauce
butter
sturgeon caviar
mushroom
cream
sheep cheese
buttermilk
galanga
red bean
bean
shiitake
salmon
tuna
porcini
cucumber
fish
cashew
coconut
corn
herring
enokidake
yogurt
vegetable
matsutake
haddock
cod
vegetable oil
roasted hazelnut
hazelnut
pecan
roasted pecan

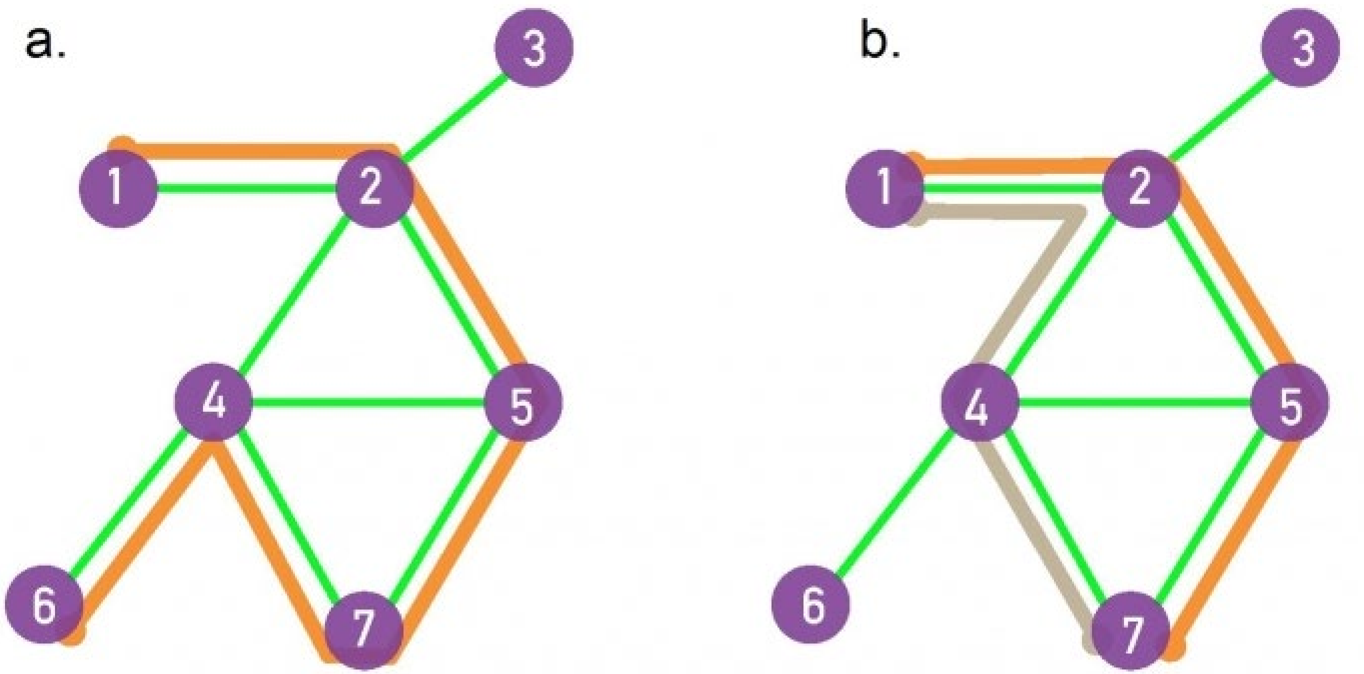
PATHOLOGY

PATHS

A *path is* a sequence of nodes in which each node is adjacent to the next one

$P_{i0,in}$ of length *n* between nodes i_0 and i_n is an ordered collection of *n+1* nodes and *n* links

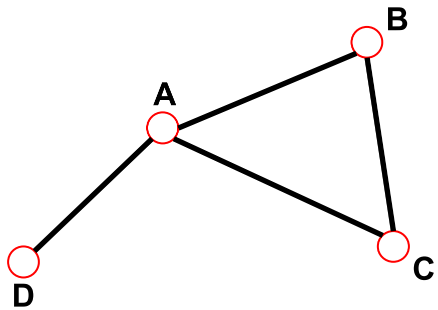
$$P_n = \{i_0, i_1, i_2, \ldots, i_n\} \qquad P_n = \{(i_0, i_1), (i_1, i_2), (i_2, i_3), \ldots, (i_{n-1}, i_n)\}$$



• In a directed network, the path can follow only the direction of an arrow.

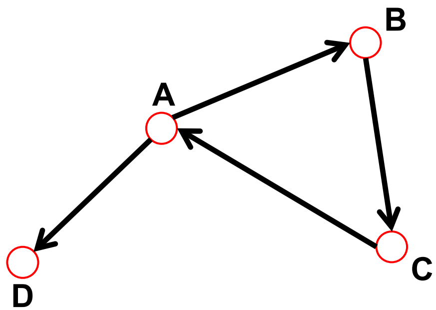
DISTANCE IN A GRAPH

Shortest Path, Geodesic Path



The *distance (shortest path, geodesic path)* between two nodes is defined as the number of edges along the shortest path connecting them.

*If the two nodes are disconnected, the distance is infinity.



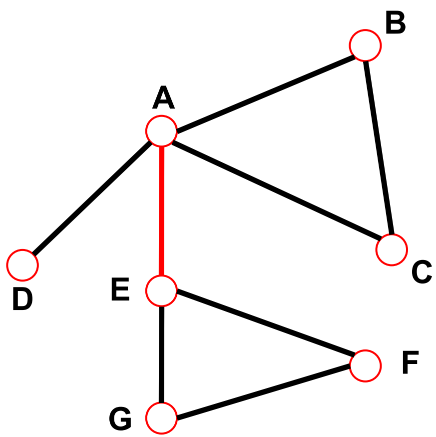
In directed graphs each path needs to follow the direction of the arrows.

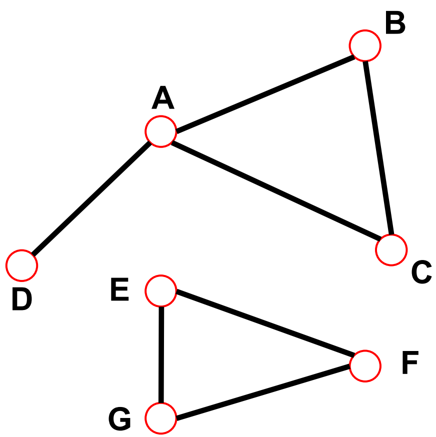
Thus in a digraph the distance from node A to B (on an AB path) is generally different from the distance from node B to A (on a BCA path).

CONNECTEDNESS

CONNECTIVITY OF UNDIRECTED GRAPHS

Connected (undirected) graph: any two vertices can be joined by a path.
A disconnected graph is made up by two or more connected components.



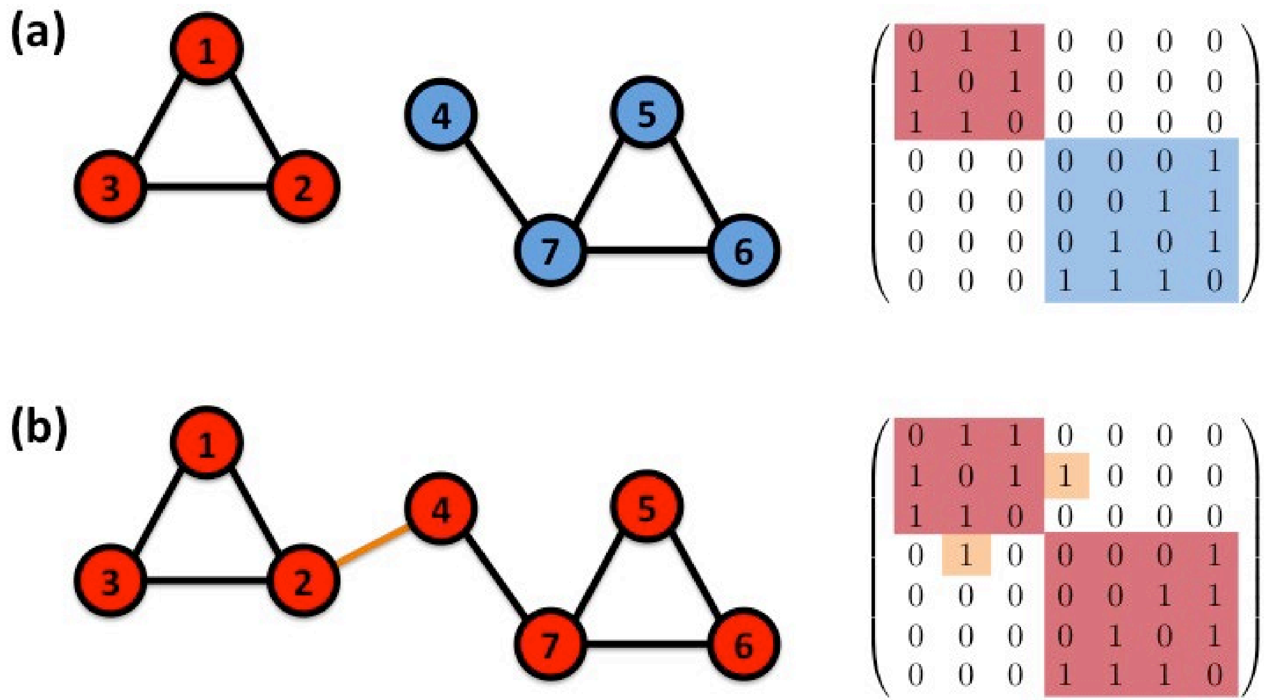


Largest Component:
Giant Component

The rest: **Isolates**

Bridge: if we erase it, the graph becomes disconnected.

The adjacency matrix of a network with several components can be written in a block-diagonal form, so that nonzero elements are confined to squares, with all other elements being zero:

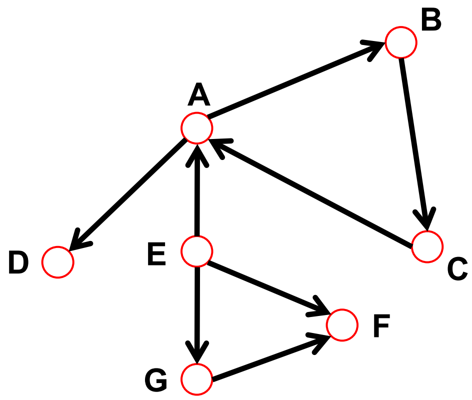


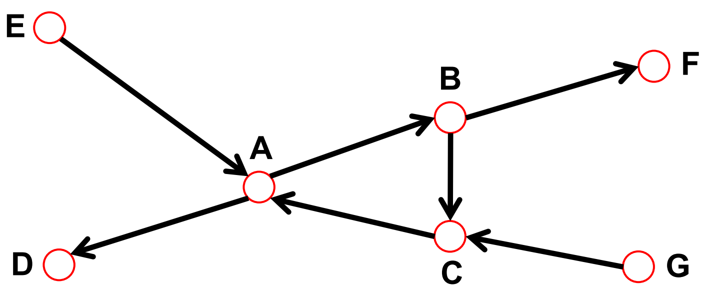
CONNECTIVITY OF DIRECTED GRAPHS

Strongly connected directed graph: has a path from each node to every other node and vice versa (e.g. AB path and BA path).

Weakly connected directed graph: it is connected if we disregard the edge directions.

Strongly connected components can be identified, but not every node is part of a nontrivial strongly connected component.





In-component: nodes that can reach the scc,

Out-component: nodes that can be reached from the scc.

Finding the Connected Components of a Network

1. Start from a randomly chosen node *i* and perform a BFS (BOX 2.5). Label all nodes reached this way with $n = 1$.
2. If the total number of labeled nodes equals *N*, then the network is connected. If the number of labeled nodes is smaller than *N*, the network consists of several components. To identify them, proceed to step 3.
3. Increase the label $n \rightarrow n + 1$. Choose an unmarked node *j*, label it with *n*. Use BFS to find all nodes reachable from *j*, label them all with *n*. Return to step 2.

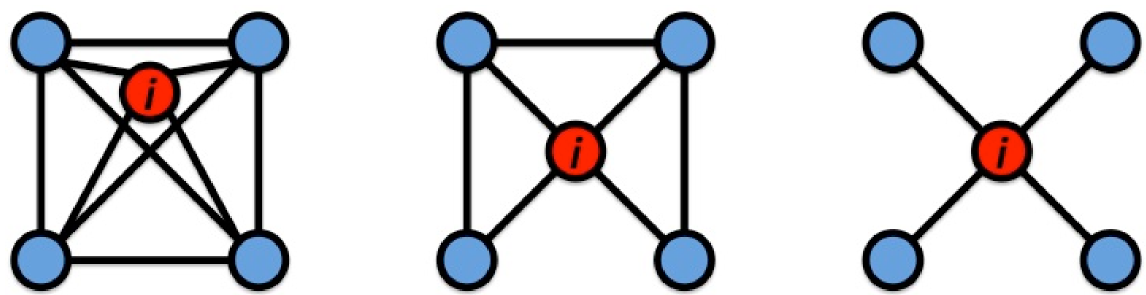
Clustering coefficient

Clustering coefficient:

what fraction of your neighbors are connected?

- Node i with degree k_i
- C_i in [0,1]

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



$C_i = 1$ $C_i = 1/2$ $C_i = 0$

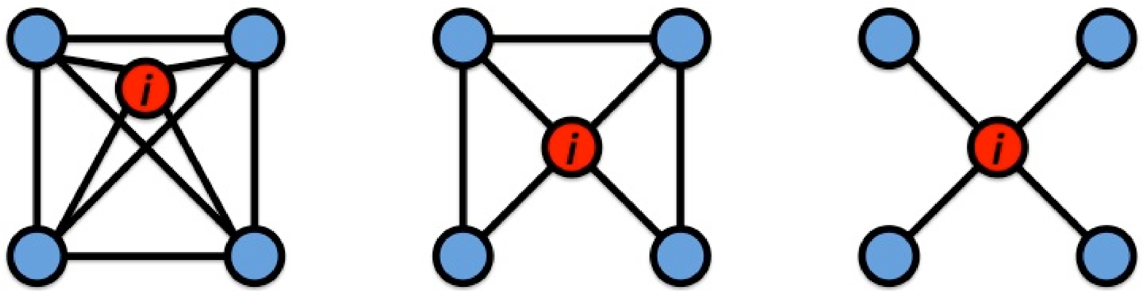
Watts & Strogatz, Nature 1998.

Clustering coefficient:

what fraction of your neighbors are connected?

- Node i with degree k_i
- C_i in [0,1]

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



$C_i = 1$ $C_i = 1/2$ $C_i = 0$

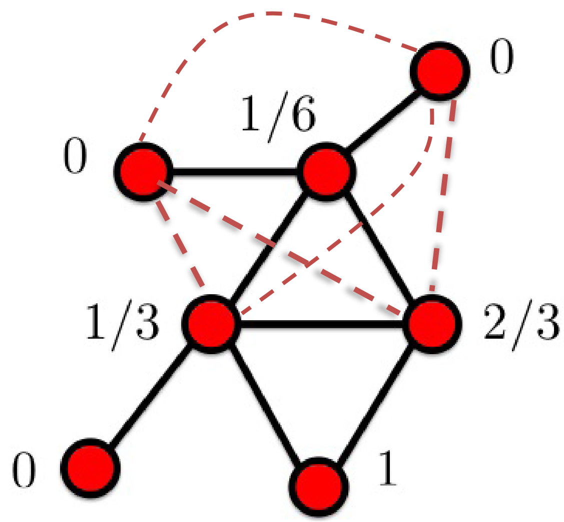
Watts & Strogatz, Nature 1998.

Clustering coefficient and Global clustering coefficient

what fraction of your neighbors are connected?

- Node i with degree k_i
- C_i in [0,1]

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



$$\langle C \rangle = \frac{13}{42} \approx 0.310$$

$$C_\Delta = \frac{3}{8} = 0.375$$

$$C_\Delta = \frac{3 * \Delta}{<}$$

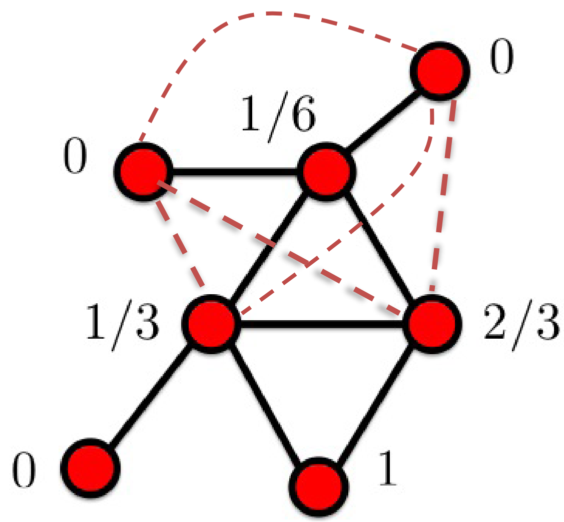
Watts & Strogatz, Nature 1998.

Clustering coefficient and Global clustering coefficient

what fraction of your neighbors are connected?

- Node i with degree k_i
- C_i in [0,1]

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



$$\langle C \rangle = \frac{13}{42} \approx 0.310$$

$$C_\Delta = \frac{3}{8} = 0.375$$

$$C_\Delta = \frac{3 * \Delta}{<}$$

Watts & Strogatz, Nature 1998.

summary

THREE CENTRAL QUANTITIES IN NETWORK SCIENCE

Degree distribution: **P(k)**

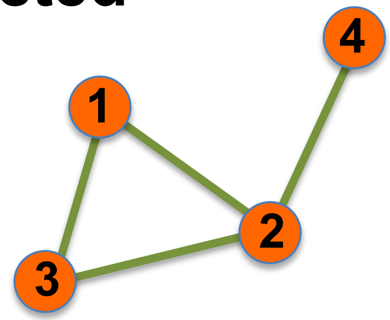
Path length: **<*d*>**

Clustering coefficient:

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

GRAPHOLOGY 1

Undirected



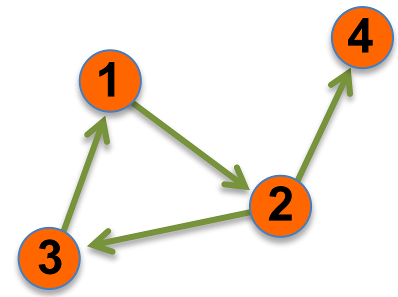
$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0 \qquad A_{ij} = A_{ji}$$

$$L = \frac{1}{2}\sum_{i,j=1}^{N} A_{ij} \qquad <k> = \frac{2L}{N}$$

Actor network, protein-protein interactions

Directed



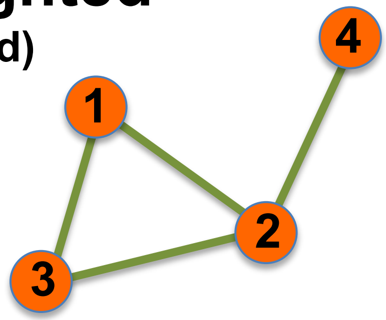
$$A_{ij} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0 \qquad A_{ij} \neq A_{ji}$$

$$L = \sum_{i,j=1}^{N} A_{ij} \qquad <k> = \frac{L}{N}$$

WWW, citation networks

Unweighted
(undirected)



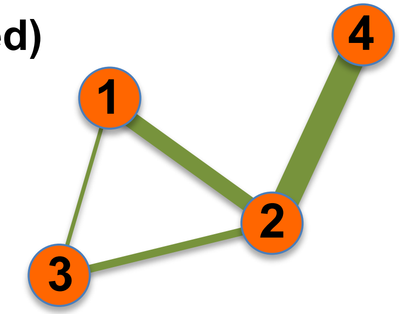
$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0 \qquad A_{ij} = A_{ji}$$

$$L = \frac{1}{2}\sum_{i,j=1}^{N} A_{ij} \qquad <k> = \frac{2L}{N}$$

protein-protein interactions, www

Weighted
(undirected)



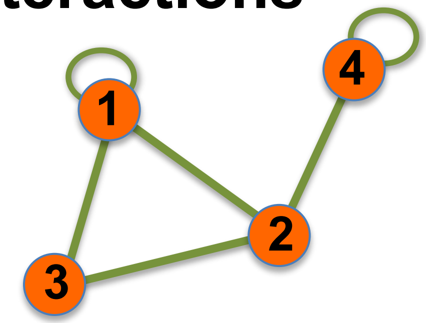
$$A_{ij} = \begin{pmatrix} 0 & 2 & 0.5 & 0 \\ 2 & 0 & 1 & 4 \\ 0.5 & 1 & 0 & 0 \\ 0 & 4 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0 \qquad A_{ij} = A_{ji}$$

$$L = \frac{1}{2}\sum_{i,j=1}^{N} nonzero(A_{ij}) \qquad <k> = \frac{2L}{N}$$

Call Graph, metabolic networks

Self-interactions



$$A_{ij} = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

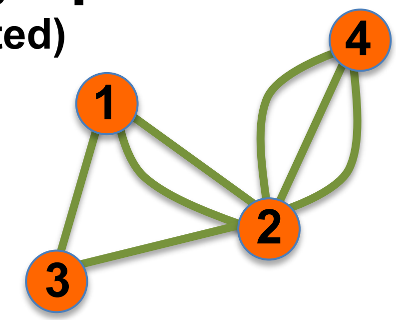
$A_{ii} \neq 0$ $\quad$ $A_{ij} = A_{ji}$

$$L = \frac{1}{2}\sum_{i,j=1,i\neq j}^{N} A_{ij} + \sum_{i=1}^{N} A_{ii} \qquad ?$$

Protein interaction network, www

Multigraph

(undirected)



$$A_{ij} = \begin{pmatrix} 0 & 2 & 1 & 0 \\ 2 & 0 & 1 & 3 \\ 1 & 1 & 0 & 0 \\ 0 & 3 & 0 & 0 \end{pmatrix}$$

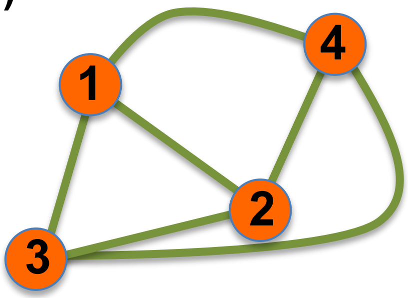
$A_{ii} = 0$ $\quad$ $A_{ij} = A_{ji}$

$$L = \frac{1}{2}\sum_{i,j=1}^{N} nonzero(A_{ij}) \qquad < k >= \frac{2L}{N}$$

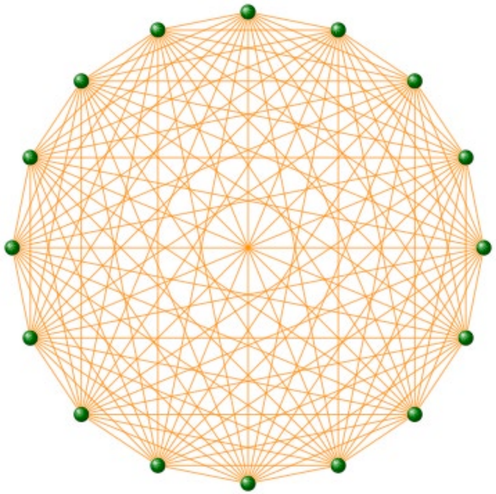
Social networks, collaboration networks

Complete Graph

(undirected)



$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$



$$A_{ii} = 0 \qquad A_{i \neq j} = 1$$

$$L = L_{\max} = \frac{N(N-1)}{2} \qquad \langle k \rangle = N - 1$$

Actor network, protein-protein interactions

GRAPHOLOGY: Real networks can have multiple characteristics

WWW > **d**irected multigraph with self-interactions

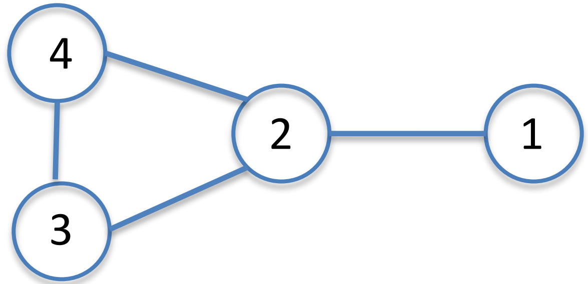
Protein Interactions > **u**ndirected unweighted with self-interactions

Collaboration network > **u**ndirected multigraph or weighted.

Mobile phone calls > directed, weighted.

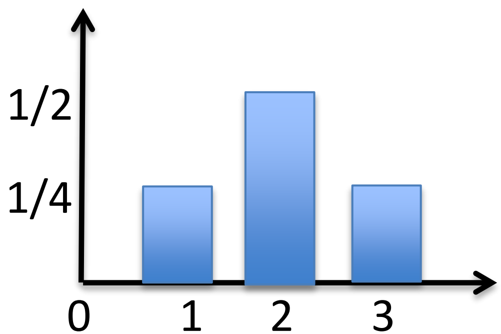
Facebook Friendship links > **u**ndirected,
unweighted.

THREE CENTRAL QUANTITIES IN NETWORK SCIENCE



$d_{1,2}=d_{2,3}=d_{2,4}=d_{3,4}=1$
$d_{1,3}=d_{1,4}=2 = d_{max}$
$\langle d \rangle=(4*1+2*2)/6=1.33$

A)



B) $\langle d \rangle = 1.33$
$d_{max} = 2$

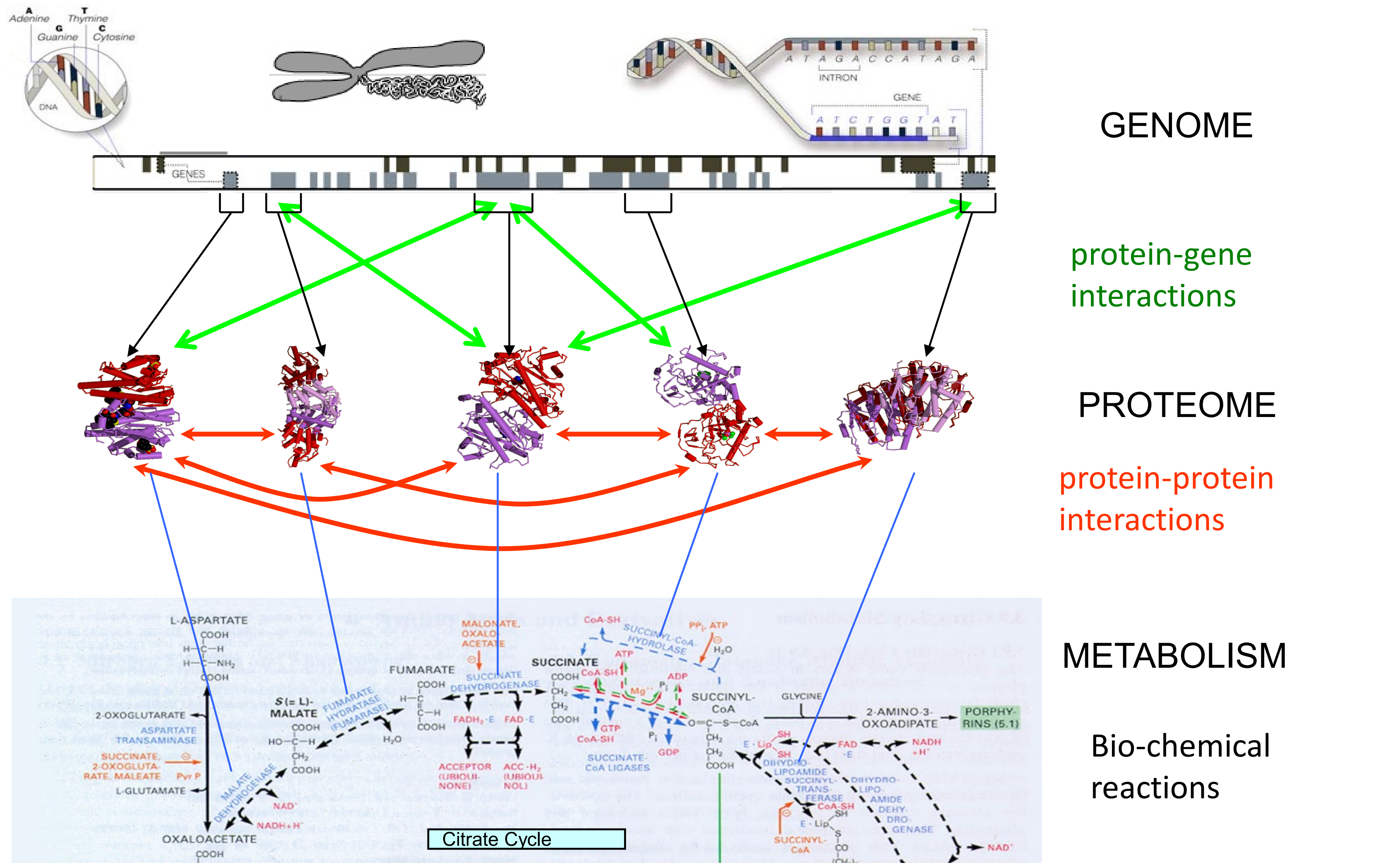
C) $C_1 = 0; C_2 = 1/3 = 0.33$
$C_3 = C_4 = 1/1 = 1$
$\langle C \rangle = 1.33/4=0.33$

A. Degree distribution: $p_k = \{0.25; 0.5; 0.25\}$

B. Path length: $\langle d \rangle = 1.33$

C. Clustering coefficient:

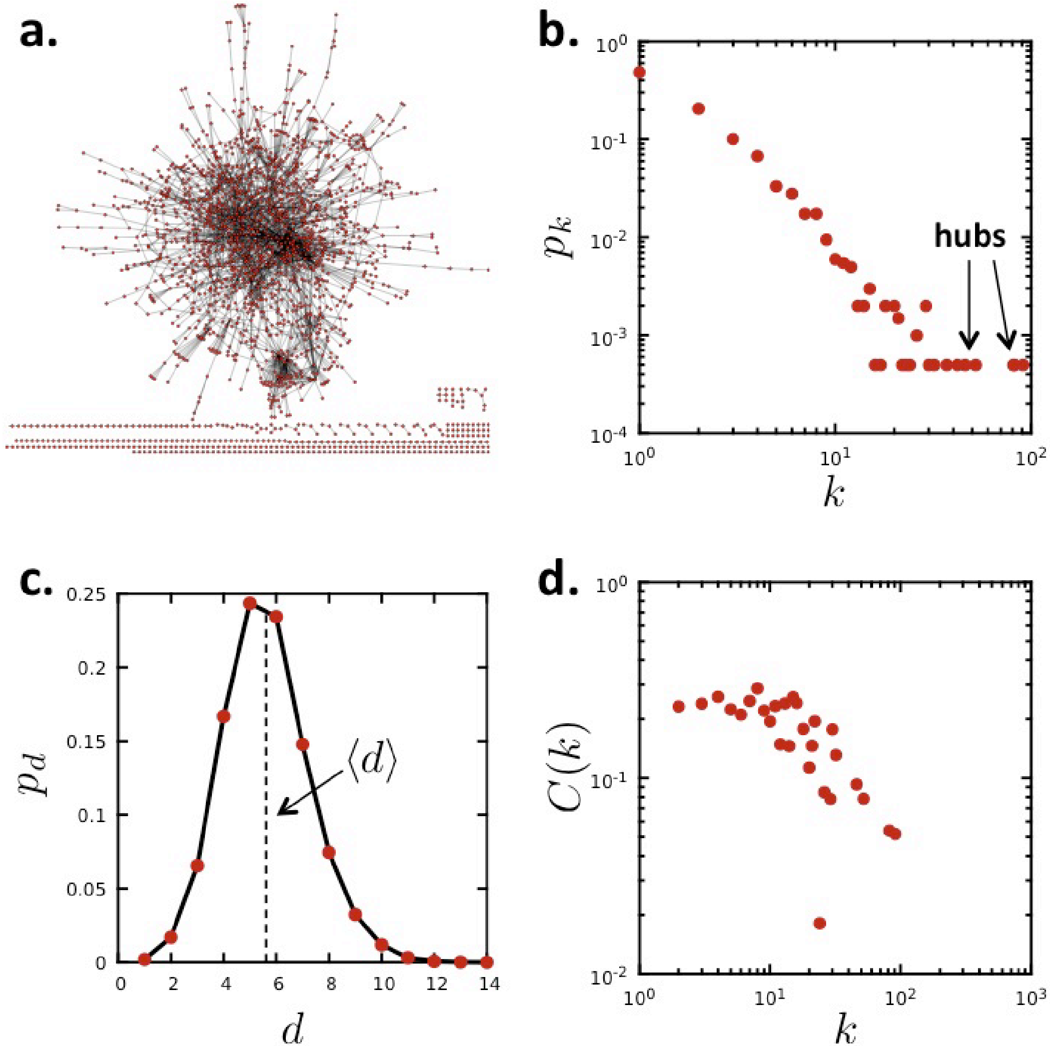
$$C_i = \frac{2e_i}{k_i(k_i-1)}$$



GENOME
protein-gene interactions
PROTEOME
protein-protein interactions
METABOLISM
Bio-chemical reactions
Adenine
Thymine
Guanine
Cytosine
DNA
INTRON
GENE
GENES
A T A G A C C A T A G A
A T C T G G T A T
Citrate Cycle
L-ASPARTATE
2-OXOGLUTARATE
ASPARTATE TRANSAMINASE
L-GLUTAMATE
OXALOACETATE
S(= L)-MALATE
MALATE DEHYDROGENASE
FUMARATE HYDRATASE (FUMARASE)
FUMARATE
SUCCINATE DEHYDROGENASE
MALONATE, OXALO-ACETATE
SUCCINATE
SUCCINYL-CoA HYDROLASE
SUCCINATE-CoA LIGASES
SUCCINYL-CoA
GLYCINE
2-AMINO-3-OXOADIPATE
PORPHY-RINS (5.1)
DIHYDROLIPOAMIDE SUCCINYL-TRANSFERASE
DIHYDROLIPOAMIDE DEHYDROGENASE
ACCEPTOR (UBIQUINONE)

Metabolic Network

A CASE STUDY: PROTEIN-PROTEIN INTERACTION NETWORK



Undirected network

N=2,018 proteins as nodes

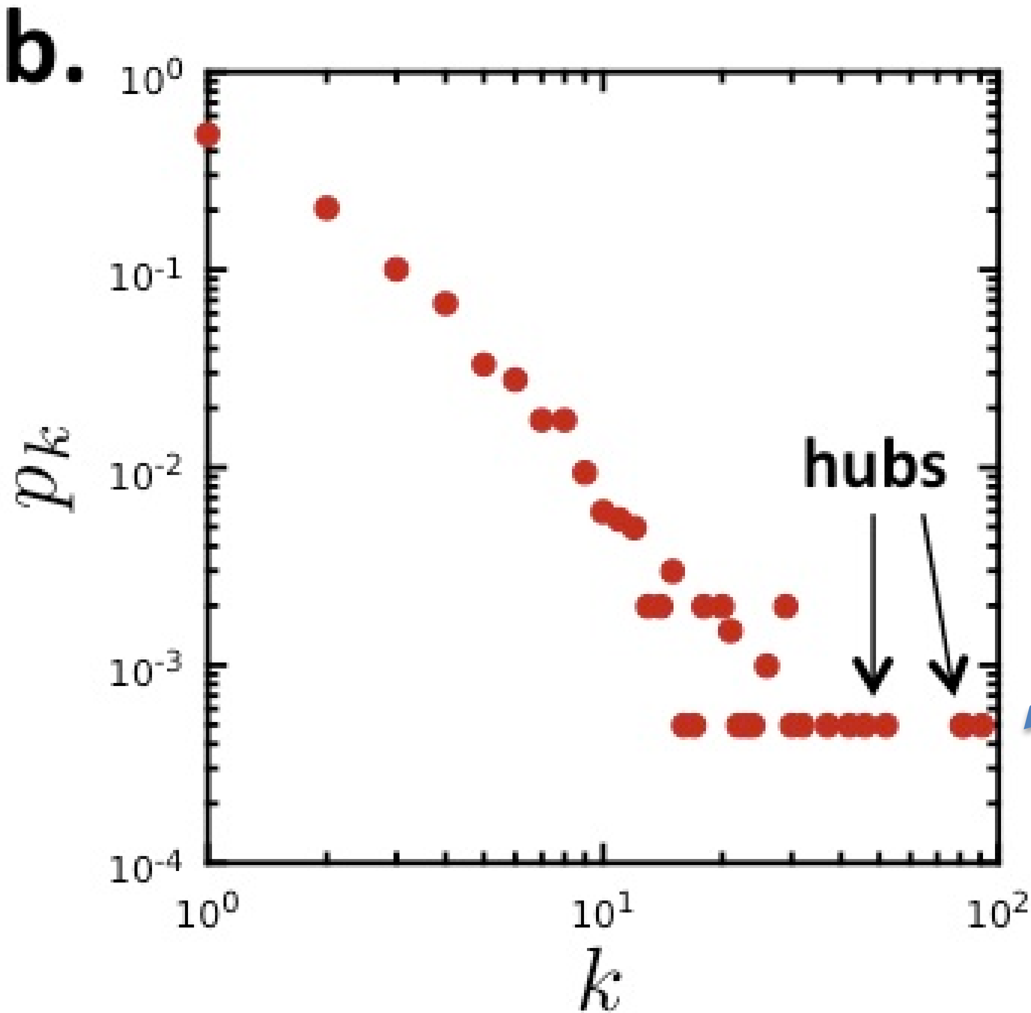
L=2,930 binding interactions as links.

Average degree <k>=2.90.

Not connected: 185 components

the largest (giant component) 1,647 nodes

A CASE STUDY: PROTEIN-PROTEIN INTERACTION NETWORK



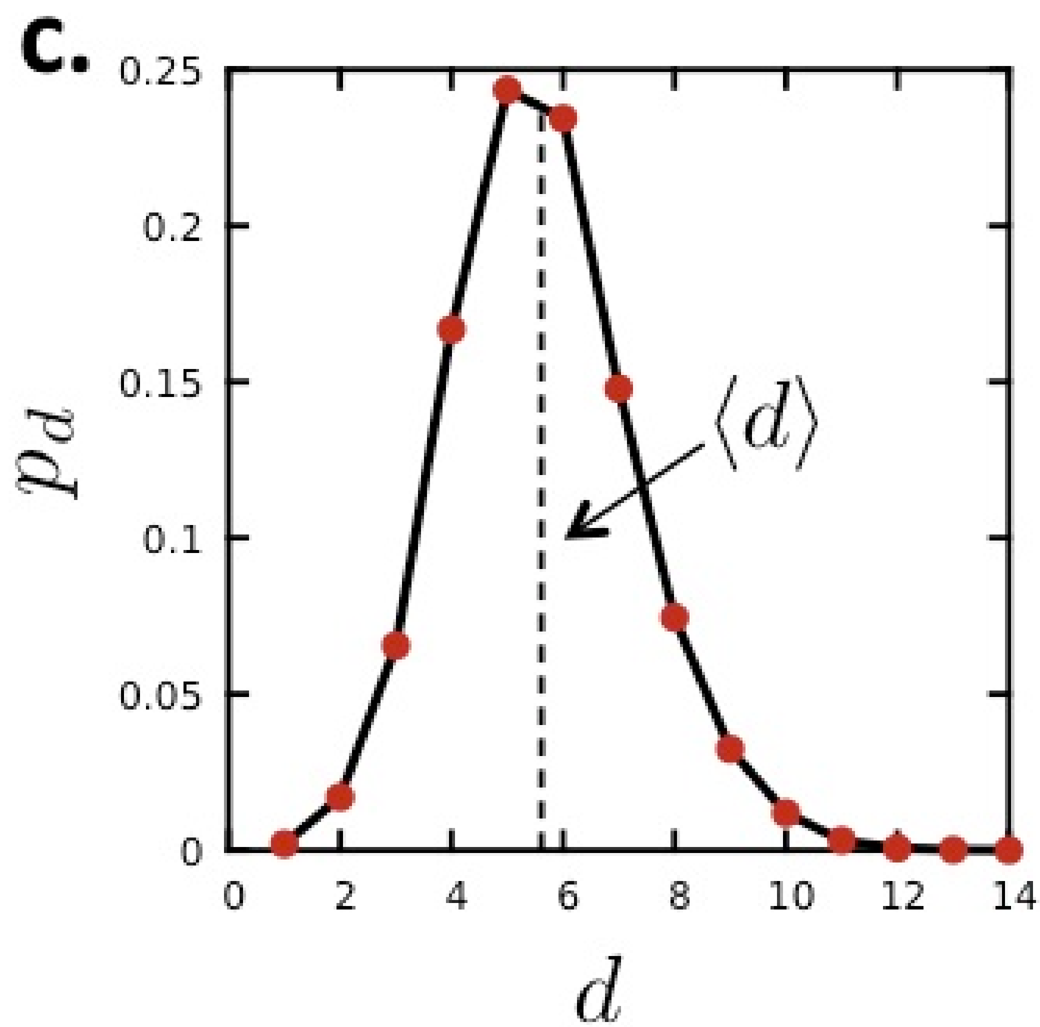
$\mathbf{p_k}$ is the probability that a node has degree **k**.

N_k = # nodes with degree k

$N_k = N^* p_k$

There is the same number of hubs with the increasing node degrees

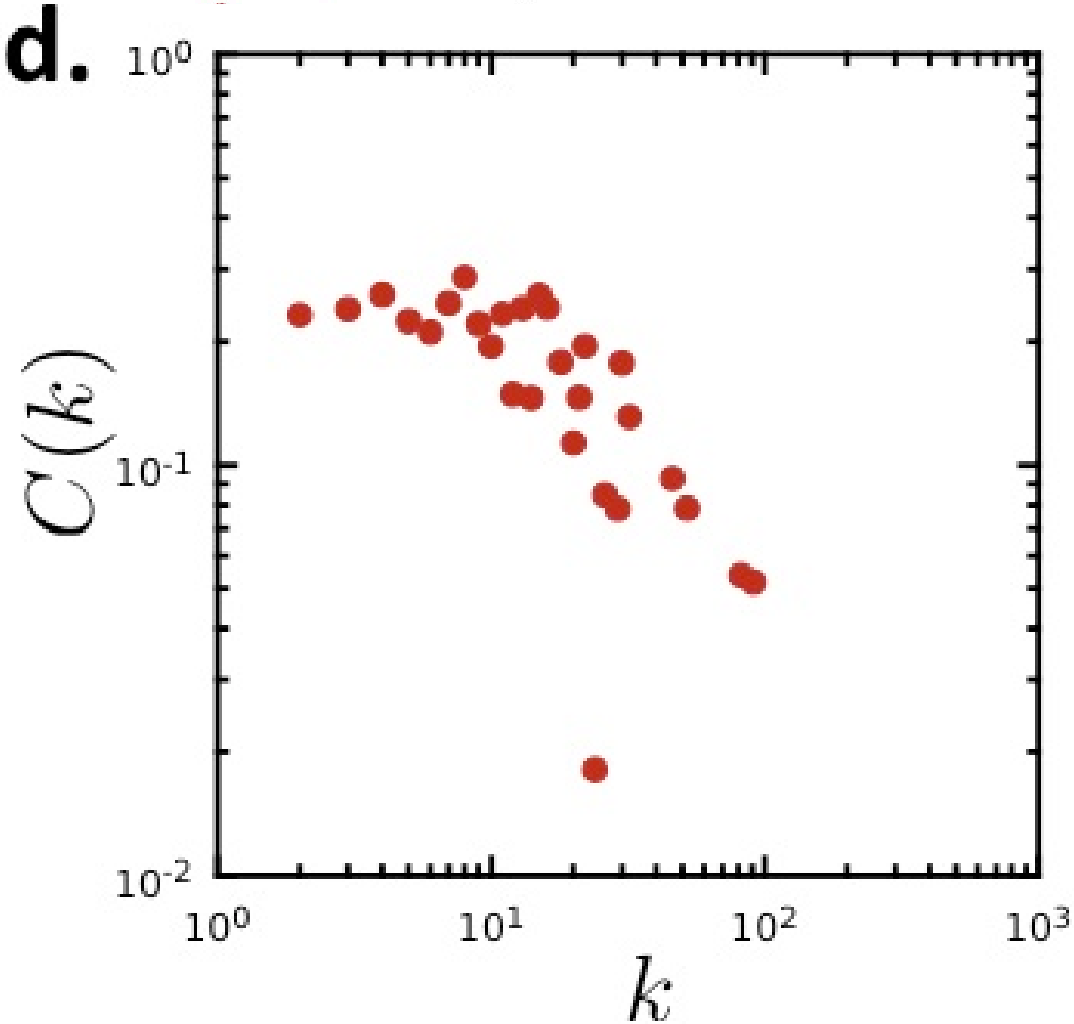
A CASE STUDY: PROTEIN-PROTEIN INTERACTION NETWORK

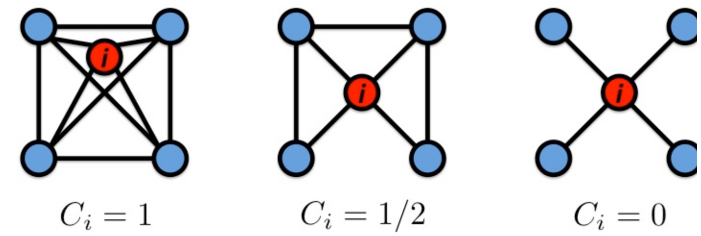


d_{max}=14

<d>=5.61

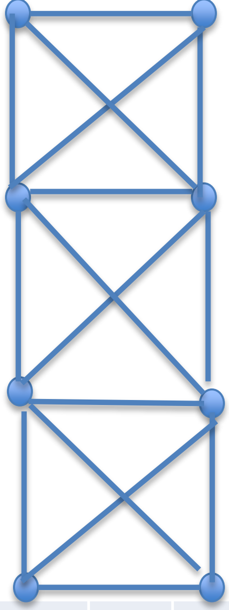
A CASE STUDY: PROTEIN-PROTEIN INTERACTION NETWORK



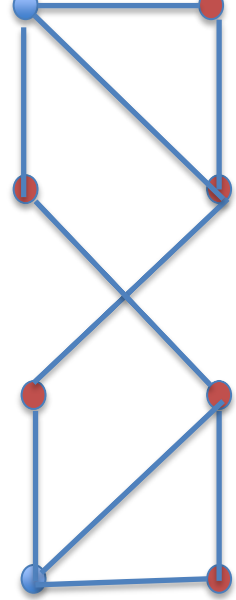


$$C_i = \frac{2e_i}{k_i(k_i - 1)}$$

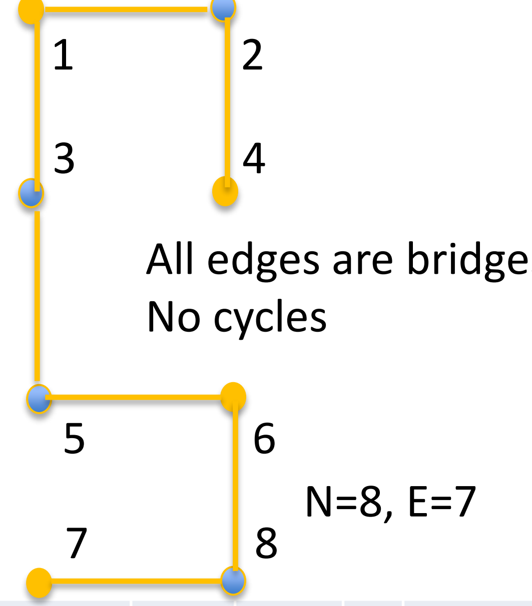
<C>=0.12



N=8, E=16



N=8, E=10

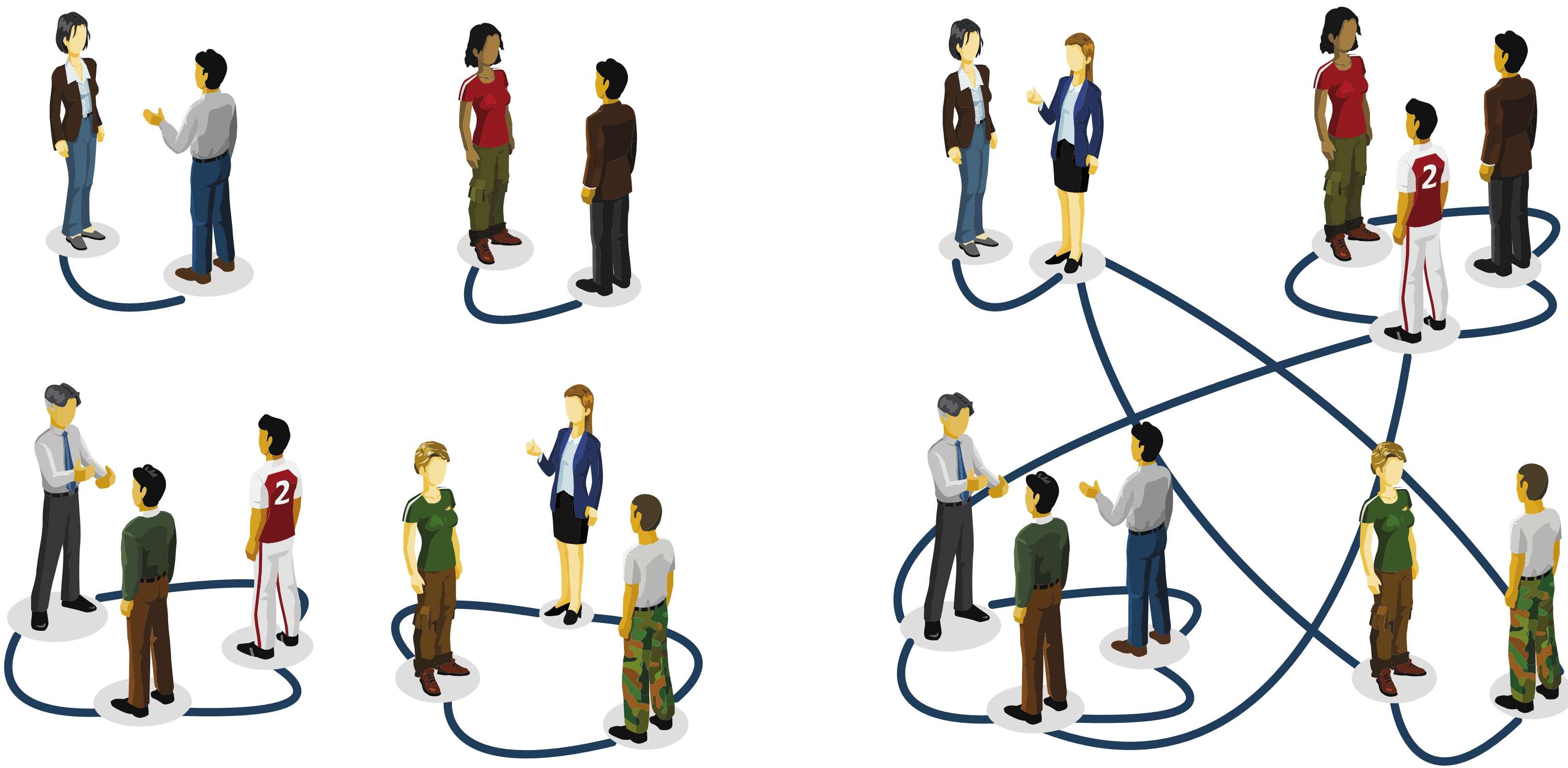


N=8, E=7

Node	deg	cc	paths		deg	cc	paths		deg	cc	paths
1	3	1	3*1+2*2+2*3=13		3	0.33	3*1+2*2+2*3=13		2	0	2*1+2*2+3+4+5=18
2	3	1	3*1+2*2+2*3=13		2	1	2*1+2*2+2*3+4=16		2	0	2*1+2+3+4+5+6=22
3	5	0.6	5*1+2*2=9		2	0	2*1+4*2+3=13		2	0	2*1+2*2+2*3+4=16
4	5	0.6	5*1+2*2=9		3	0.33	3*1+2*2+2*3=13		1	0	1+2+3+4+5+6+7=28
5	5	0.6	5*1+2*2=9		2	0	2*1+4*2+3=13		2	0	2*1+2*2+2*3+4=16
6	5	0.6	5*1+2*2=9		3	0.33	3*1+2*2+2*3=13		2	0	2*1+2*2+3+4+5=18
7	3	1	3*1+2*2+2*3=13		3	0.33	3*1+2*2+2*3=13		1	0	1+2+3+4+5+6+7=28
8	3	1	3*1+2*2+2*3=13		2	1	2*1+2*2+2*3+4=16		2	0	2*1+2+3+4+5+6=22
ave	4	0.8	1.571428571		2.5	0.33	1.928571429		1.75	0	3
max	5	1	3		3	1	4		2	0	7

Introduction

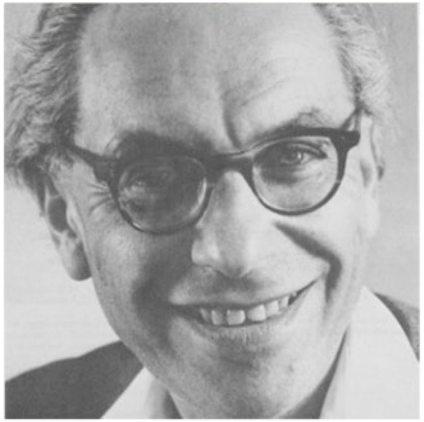
RANDOM NETWORK MODEL



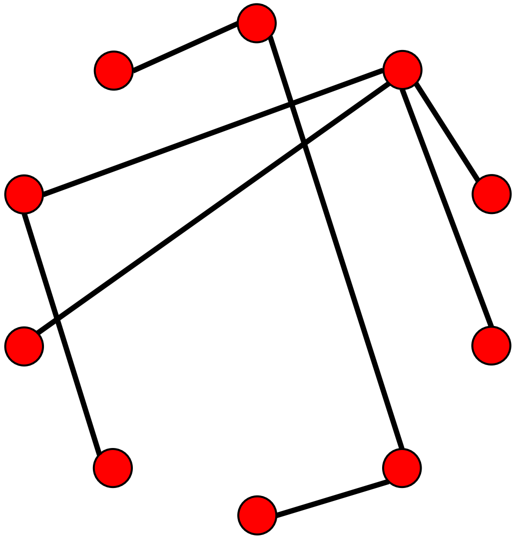
The random network model

RANDOM NETWORK MODEL

Pál Erdös
(1913-1996)



Alfréd Rényi
(1921-1970)



Erdös-Rényi model (1960)

Connect with probability p

p=1/6 N=10

<k> ~ 1.5

RANDOM NETWORK MODEL

Definition:

A **random graph** is a graph of N nodes where each pair of nodes is connected by probability **p**.

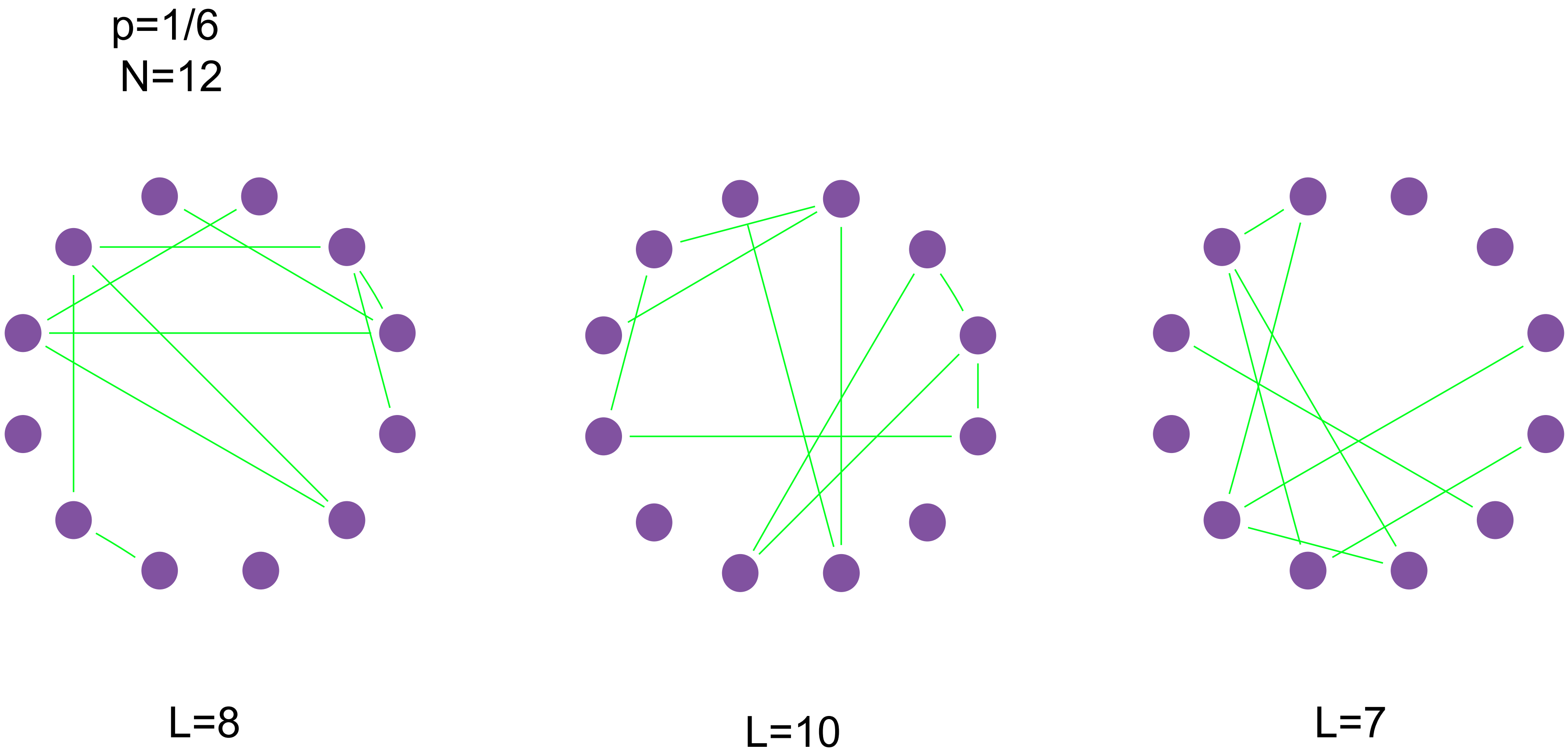
$G(N, L)$ Model

N labeled nodes are connected with L randomly placed links. Erdős and Rényi used this definition in their string of papers on random networks [2-9].

$G(N, p)$ Model

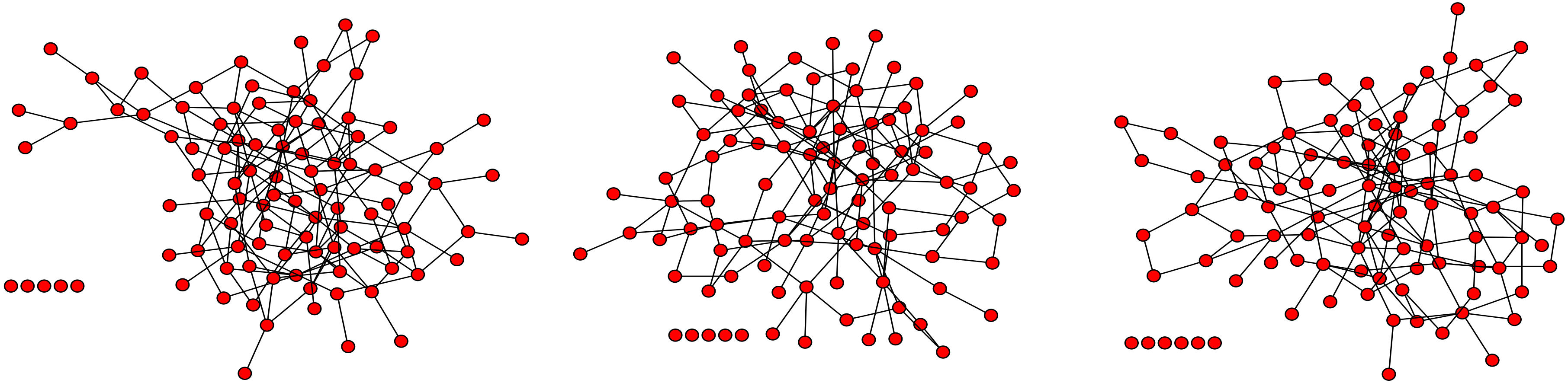
Each pair of N labeled nodes is connected with probability p, a model introduced by Gilbert [10].

RANDOM NETWORK MODEL



RANDOM NETWORK MODEL

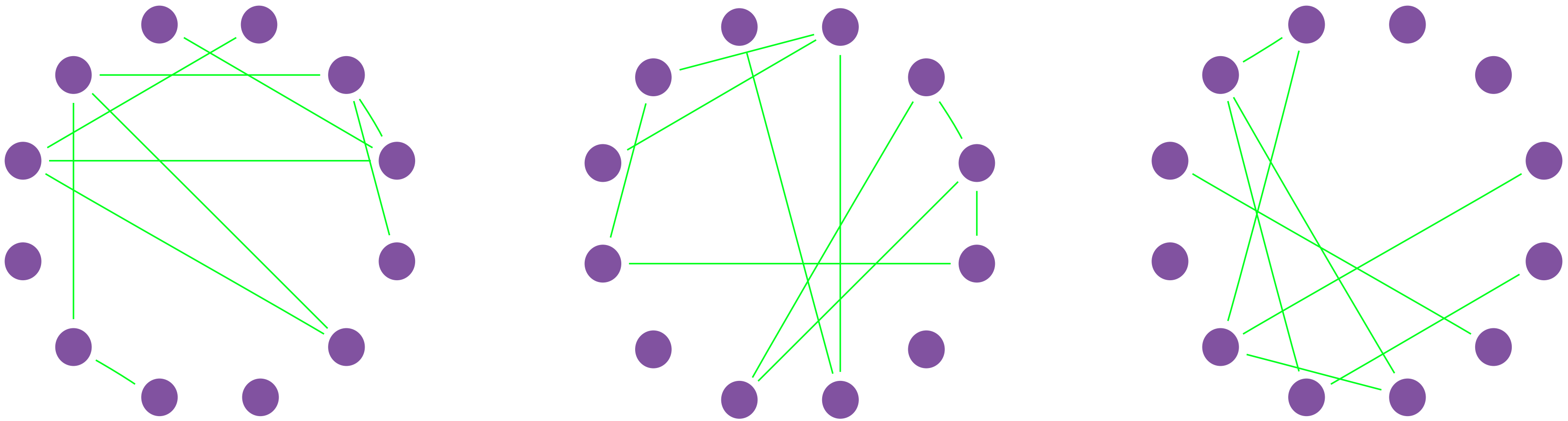
p=0.03
N=100



The number of links is variable

RANDOM NETWORK MODEL

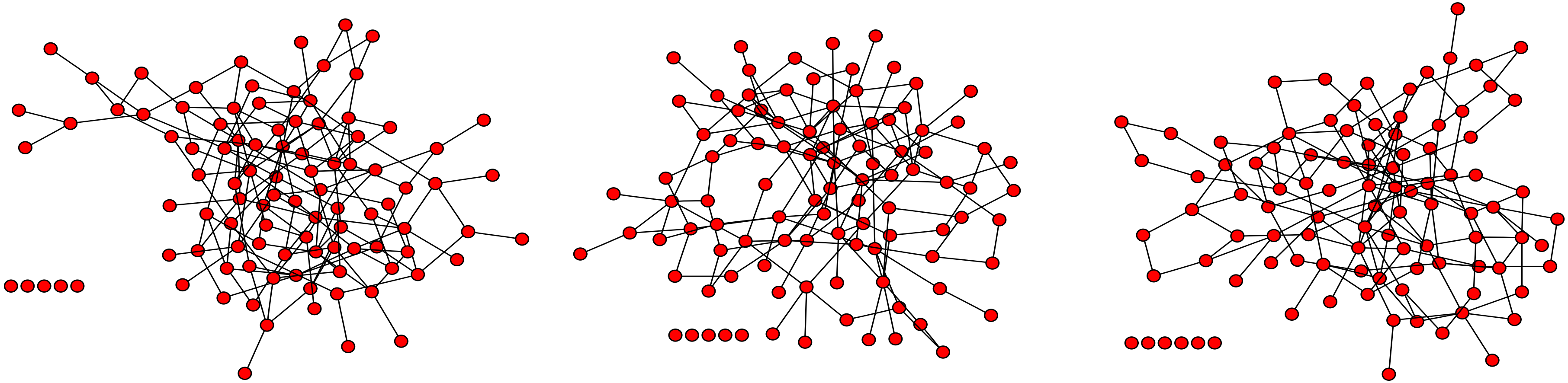
p=1/6
N=12



L=8 L=10 L=7

RANDOM NETWORK MODEL

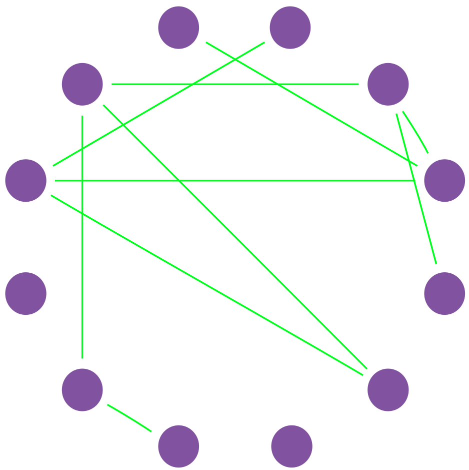
p=0.03
N=100



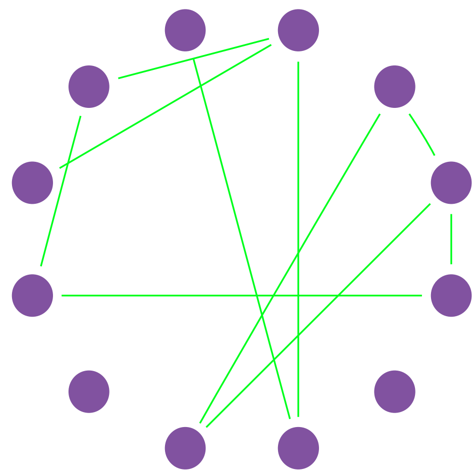
The number of links is variable

RANDOM NETWORK MODEL

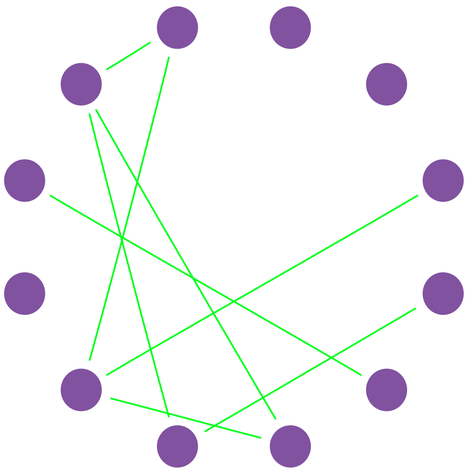
p=1/6
N=12



L=8



L=10



L=7

Number of links in a random network

P(L): the probability to have exactly *L* links in a network of *N* nodes and probability *p*:

The maximum number of links in a network of N nodes.

$$P(L) = \binom{\binom{N}{2}}{L} p^L (1-p)^{\frac{N(N-1)}{2} - L}$$

Number of different ways we can choose L links among all potential links.

Binomial distribution...

MATH TUTORIAL Binomial Distribution: The bottom line

$$P(x) = \binom{N}{x} p^x (1-p)^{N-x}$$

$$< x >= Np$$

$$< x^2 >= p(1-p)N + p^2 N^2$$

$$\sigma_x = (< k^2 > - < k >^2)^{1/2} = [p(1-p)N]^{1/2}$$

http://keral2008.blogspot.com/2008/10/derivation-of-mean-and-variance-of.html

RANDOM NETWORK MODEL

P(L): the probability to have a network of exactly *L* links

$$P(L) = \binom{\binom{N}{2}}{L} p^{L} (1-p)^{\frac{N(N-1)}{2}-L}$$

•The average number of links *<L>* in a random graph

$$< L >= \sum_{L=0}^{\frac{N(N-1)}{2}} LP(L) = p\frac{N(N-1)}{2}$$

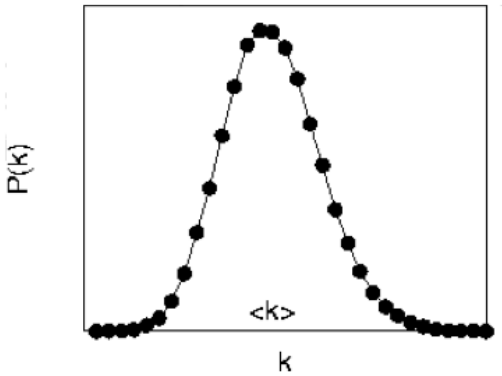
$$< k >= 2L/N = p(N-1)$$

•The standard deviation

$$\sigma^2 = p(1-p)\frac{N(N-1)}{2}$$

Degree distribution

DEGREE DISTRIBUTION OF A RANDOM GRAPH



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

Select k nodes from N-1

probability of having *k* edges

probability of missing N-1-k edges

$$<k> = p(N-1) \qquad \sigma_k^2 = p(1-p)(N-1)$$

$$\frac{\sigma_k}{<k>} = \left[\frac{1-p}{p}\frac{1}{(N-1)}\right]^{1/2} \approx \frac{1}{(N-1)^{1/2}}$$

As the network size increases, the distribution becomes increasingly narrow—we are increasingly confident that the degree of a node is in the vicinity of <k>.

DEGREE DISTRIBUTION OF A RANDOM GRAPH

$$P(k) = \binom{N-1}{k} p^k (1-p)^{(N-1)-k} \qquad <k> = p(N-1) \qquad p = \frac{<k>}{(N-1)}$$

For large *N* and small *k*, we can use the following approximations:

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

$$\ln[(1-p)^{(N-1)-k}] = (N-1-k)\ln(1-\frac{<k>}{N-1}) = -(N-1-k)\frac{<k>}{N-1} = -<k>(1-\frac{k}{N-1}) \cong -<k>$$

$$(1-p)^{(N-1)-k} = e^{-<k>} \qquad \ln(1+x) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} x^n = x - \frac{x^2}{2} + \frac{x^3}{3} - \ldots \quad \text{for} \quad |x| \le 1$$

$$P(k) = \binom{N-1}{k} p^k (1-p)^{(N-1)-k} = \frac{(N-1)^k}{k!} p^k e^{-<k>} = \frac{(N-1)^k}{k!} \left(\frac{<k>}{N-1}\right)^k e^{-<k>} = e^{-<k>} \frac{<k>^k}{k!}$$

POISSON DEGREE DISTRIBUTION

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

$$<k> = p(N-1)$$

$$p = \frac{<k>}{(N-1)}$$

For large *N* and small *k*, we arrive at the Poisson distribution:

$$P(k) = e^{-<k>} \frac{<k>^k}{k!}$$

DEGREE DISTRIBUTION OF A RANDOM GRAPH

<k>=50

$$P(k) = e^{-<k>} \frac{<k>^k}{k!}$$