

Complex Networks

Vincent Gauthier

Associate Professor

 @vincentgauthier

 <http://complex.luxbulb.org>

Telecom SudParis-Institut Mines-Telecom/CNRS SAMOVAR

Software

R

Matlab

Python

Networkx [Python]

Graph-tool [Python+C++]

GraphLab [Python+C++]

Graph Editor

Gephi

GraphViz

Pajek

Datasets

Mark Newman's network data set

Stanford Network Analysis Project

Carnegie Mellon CASOS data sets

NCEAS food web data sets

UCI NET data sets

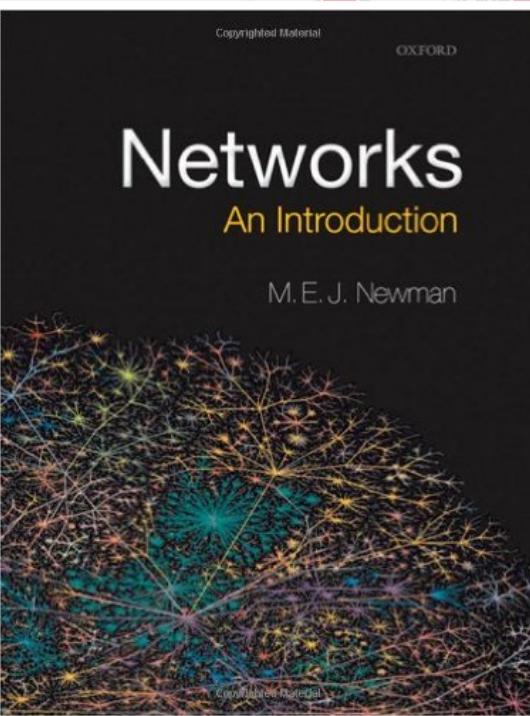
Pajek data sets

Linkgroup's list of network data sets

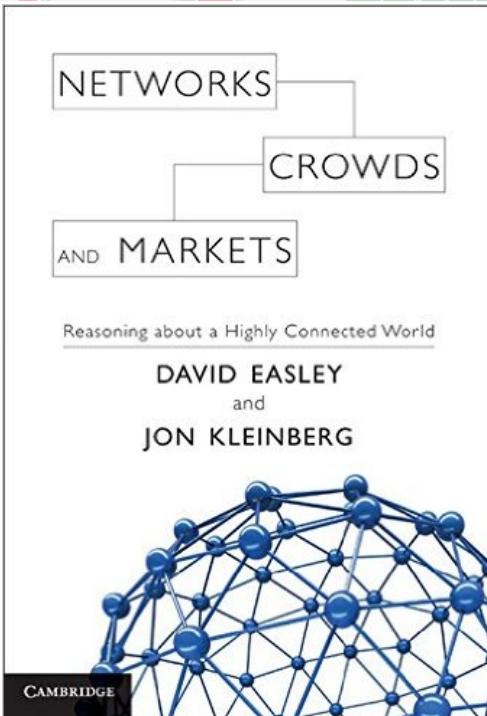
Barabasi lab data sets

Jake Hofman's online network data sets

Alex Arenas's data sets

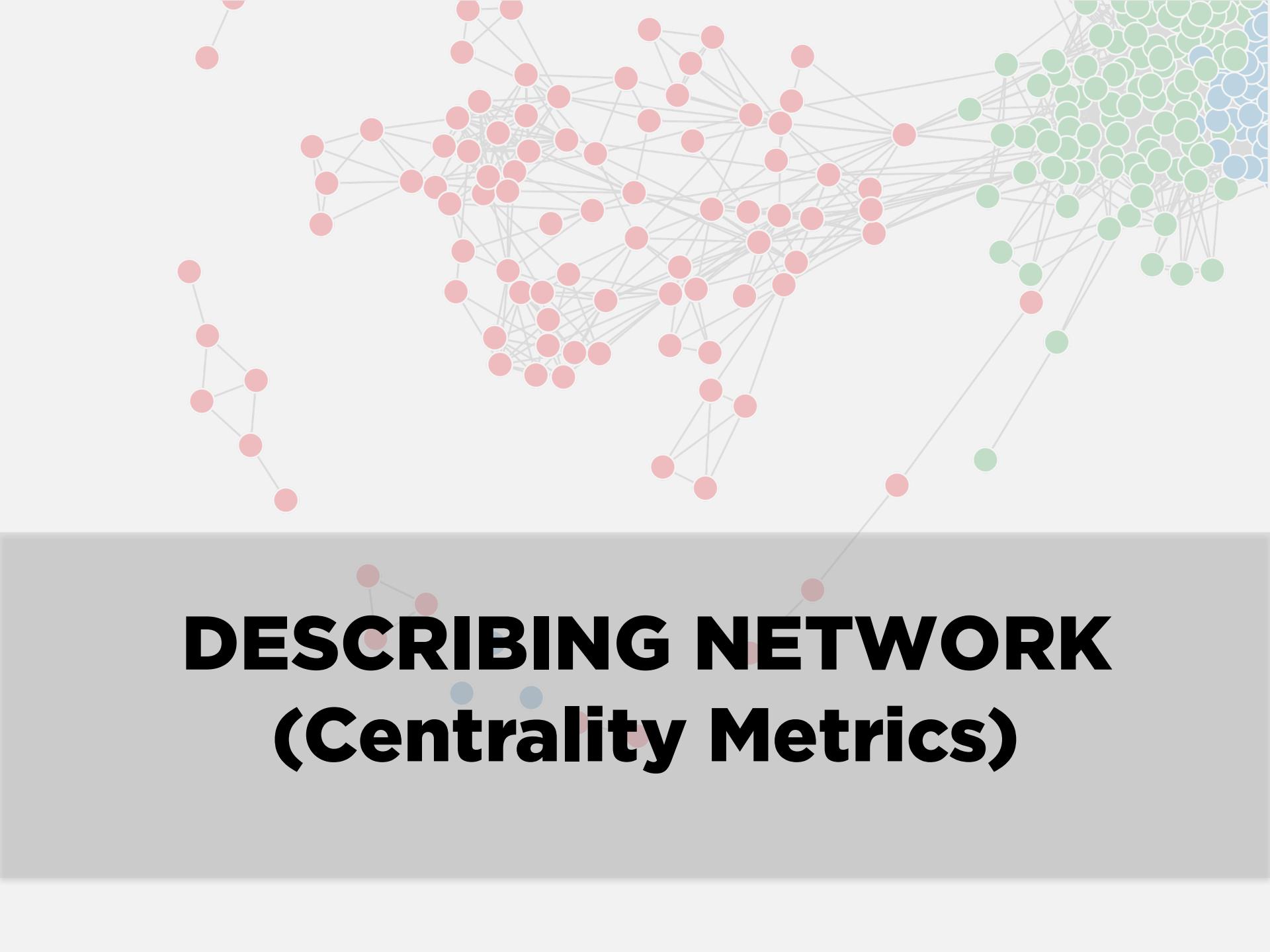


Networks An Introduction
Mark Newman, 2010.



Networks, Crowds, and Markets
Reasoning About a Highly Connected World
David Easley, Jon Kleinberg, 2010.

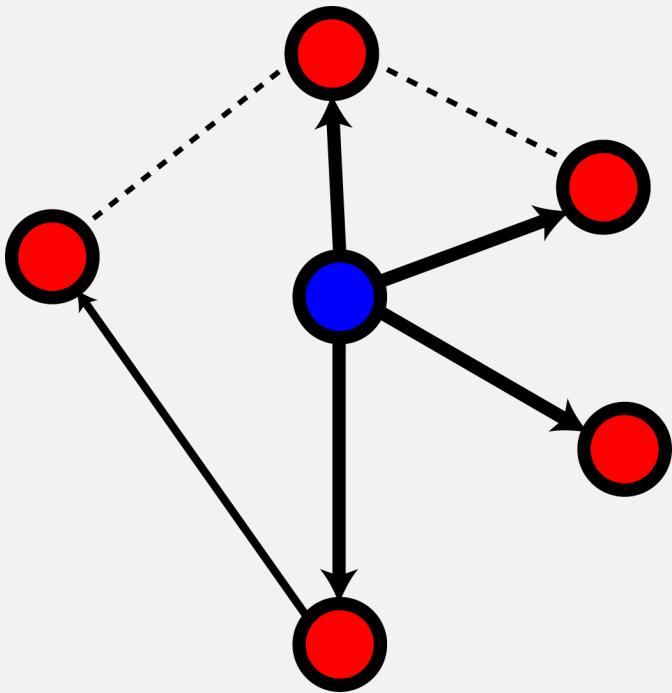
1. Definition of a network
2. Basics of graph theory
3. Describing a network
4. Network centrality
5. Eigen centrality, PageRank
6. Small world network (Watts-Strogatz)
7. Network Growth (Barabasi-Albert)
8. Social Network Analysis
9. Diffusion on network



A network graph visualization showing connections between nodes. The nodes are colored into several distinct communities: a large central cluster of pink nodes, a smaller pink cluster on the left, a green cluster on the right, and a few isolated blue nodes at the top right. The connections are represented by thin grey lines.

DESCRIBING NETWORK (Centrality Metrics)

Describing Networks



Position=Centrality

Measure of positional “importance”

geometric
Connectivity

Harmonic centrality

Closeness centrality

Betweenness centrality

Degree centrality

Eigenvector centrality

PageRank

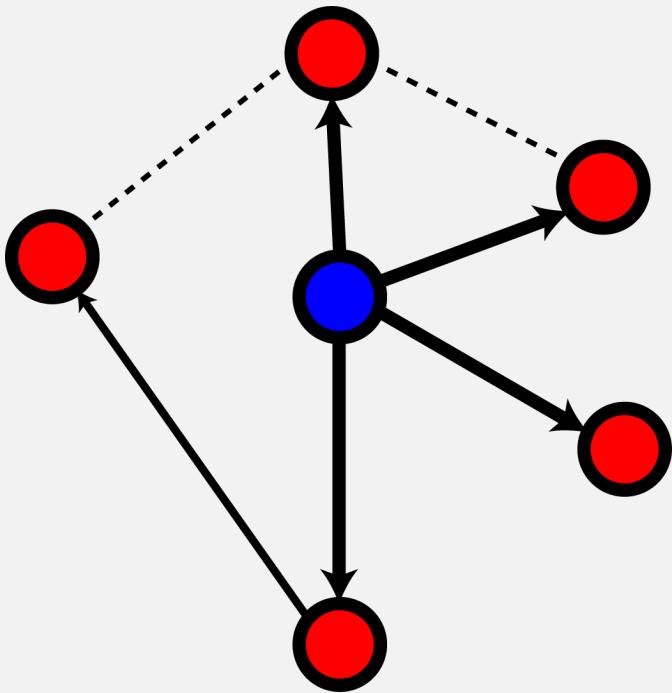
Katz centrality

Many many more

■ Axioms for Centrality
Paolo Boldi & Sebastiano Vigna
Internet Mathematics Vol. 10, Iss. 3-4, 2014

■ Centrality and network flow
SP Borgatti
Social networks 27 (1), 55-71, 2005.

Describing Networks



Position=Centrality:

Harmonic, closeness centrality

Importance of being in the
“center” of the network

$$\text{Harmonic: } c_i = \frac{1}{n-1} \sum_{j \neq i} \frac{1}{d_{ij}}$$

Length of shortest path

$$\text{Distance: } d_{ij} = \begin{cases} l_{ij} & \text{if } j \text{ reachable from } i \\ \infty & \text{otherwise} \end{cases}$$

❑ Axioms for Centrality
Paolo Boldi & Sebastiano Vigna
Internet Mathematics Vol. 10, Iss. 3-4, 2014

❑ Centrality and network flow
SP Borgatti
Social networks 27 (1), 55-71, 2005.

Network position: example

Robust Action and the Rise of the Medici

John F .Padgett and Christopher K. Ansell

1993



Duomo

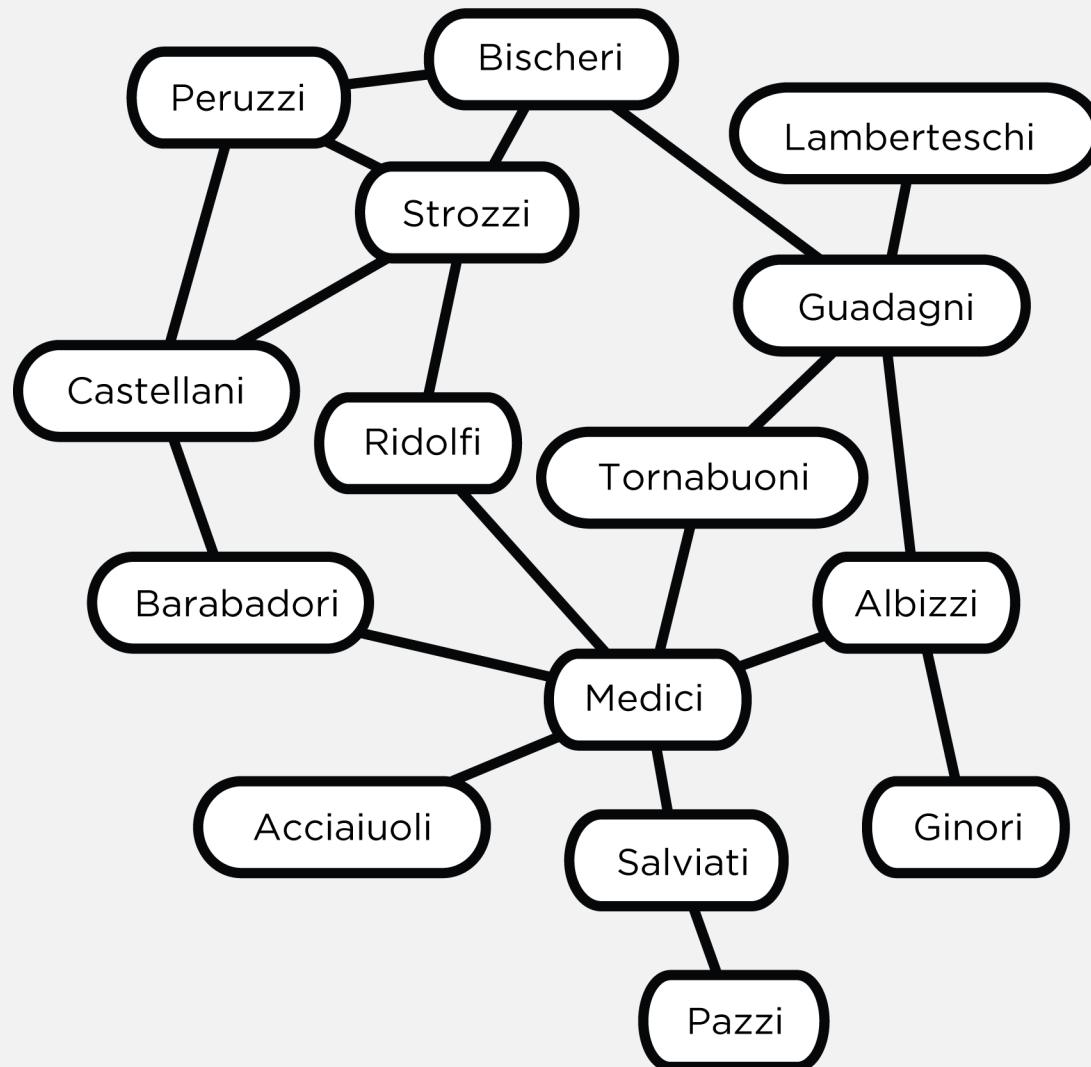


Plazzo Medici



Giovanni de Medici

Network position: closeness

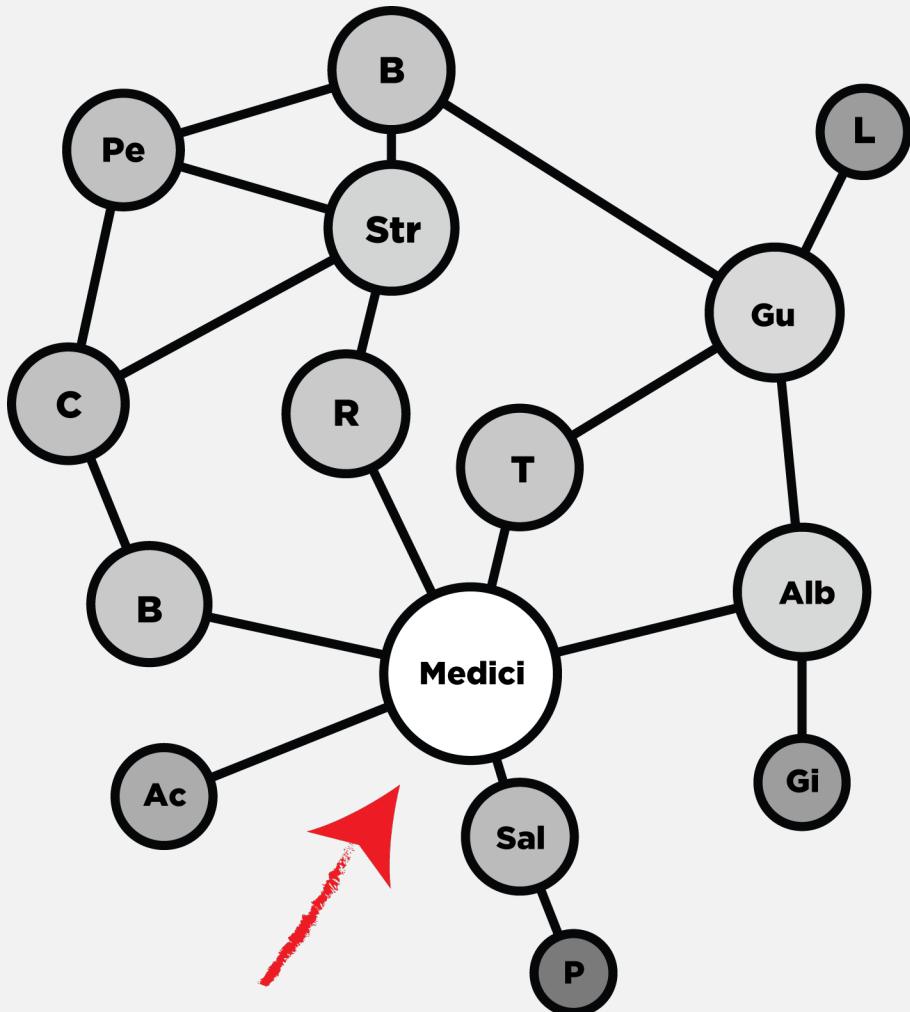


Node: Florence families

Edges: inter-family marriages

Which family is most central ?

Network position: closeness



Medici	9.5
Guadagni	7.92
Albizzi	7.83
Strozzi	7.67
Ridolfi	7.25
Bischeri	7.2
Tornabuoni	7.17
Barbadori	7.08
Peruzzi	6.87
Castellani	6.87
Salviati	6.58
Acciaiuoli	5.92
Ginori	5.33
Lamberteschi	5.28
Pazzi	4.77

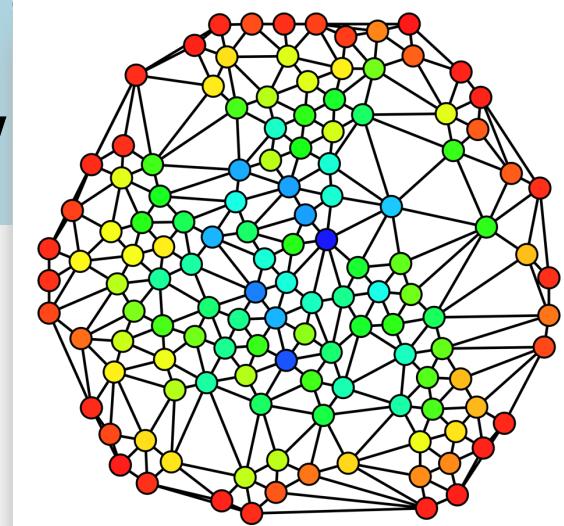


Betweenness Centrality

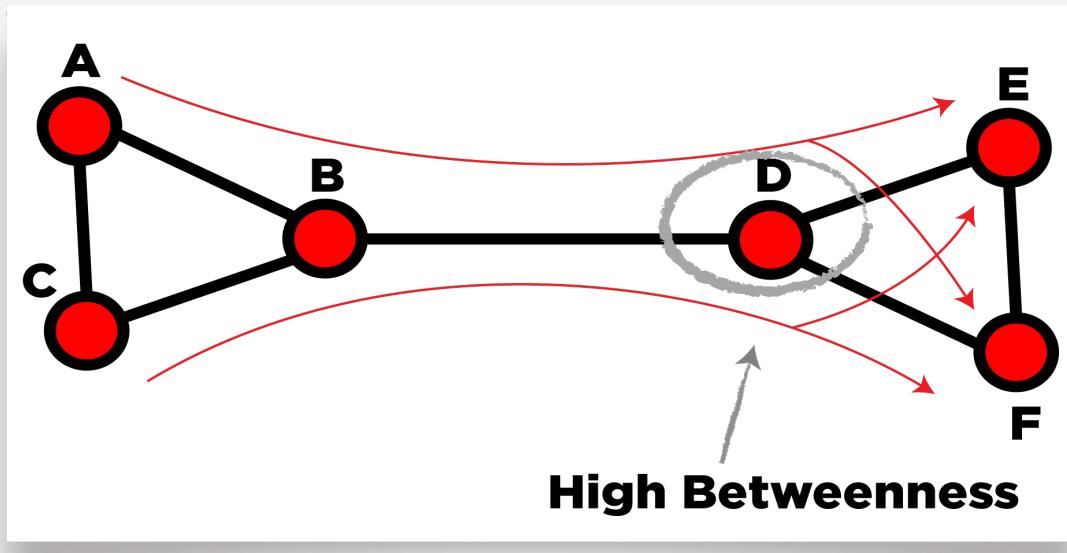
$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Number of shortest path
between node s and t
passing through node v

Number of shortest path
between node s and t



Position=Centrality:
Betweenness centrality

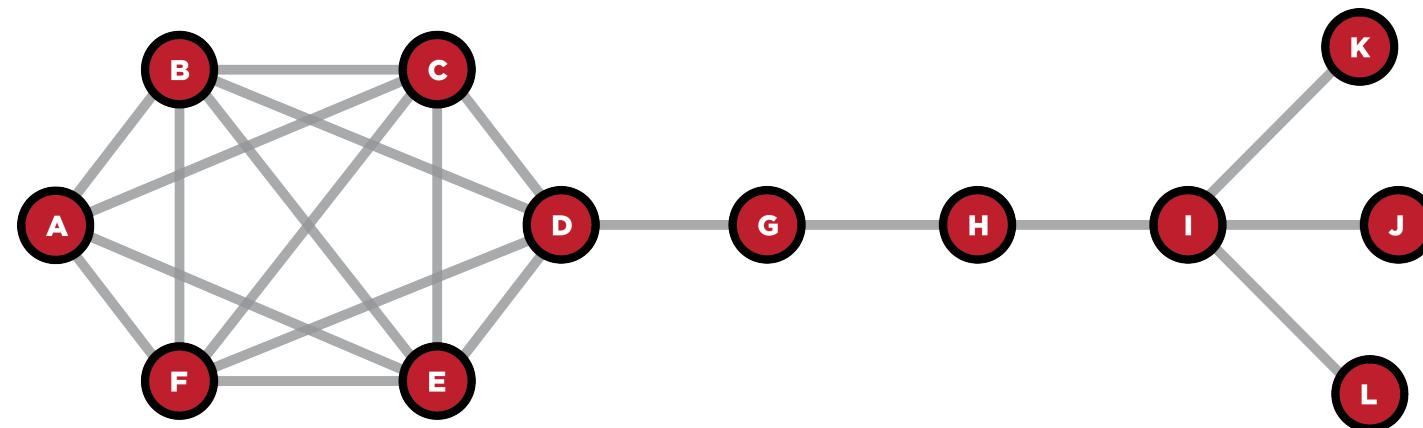


Importance of being a
“bridge” in the network

intuition: how many pairs of individuals would have to go through you in order to reach one another in the minimum number of hops?

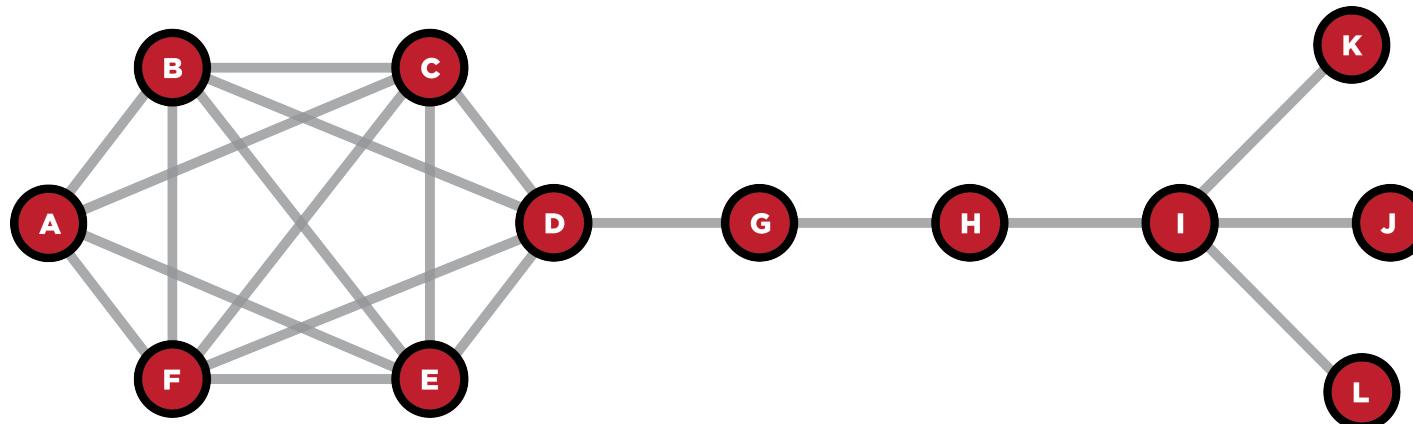
Quiz:

Find a node that has high betweenness but low degree

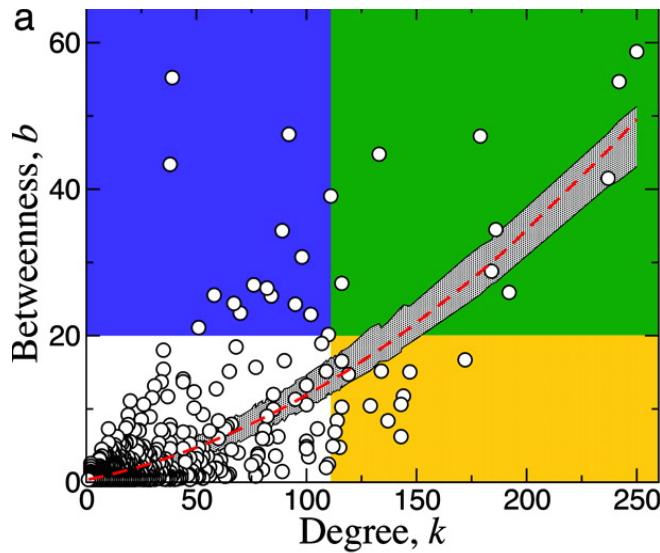


Quiz:

Find a node that has low betweenness but high degree



Centrality: Betweenness example



Rank	City	b	b/b_{ran}	Degree
1	Paris	58.8	1.2	250
2	Anchorage	55.2	16.7	39
3	London	54.7	1.2	242
4	Singapore	47.5	4.3	92
5	New York	47.2	1.6	179
6	Los Angeles	44.8	2.3	133
7	Port Moresby	43.4	13.6	38



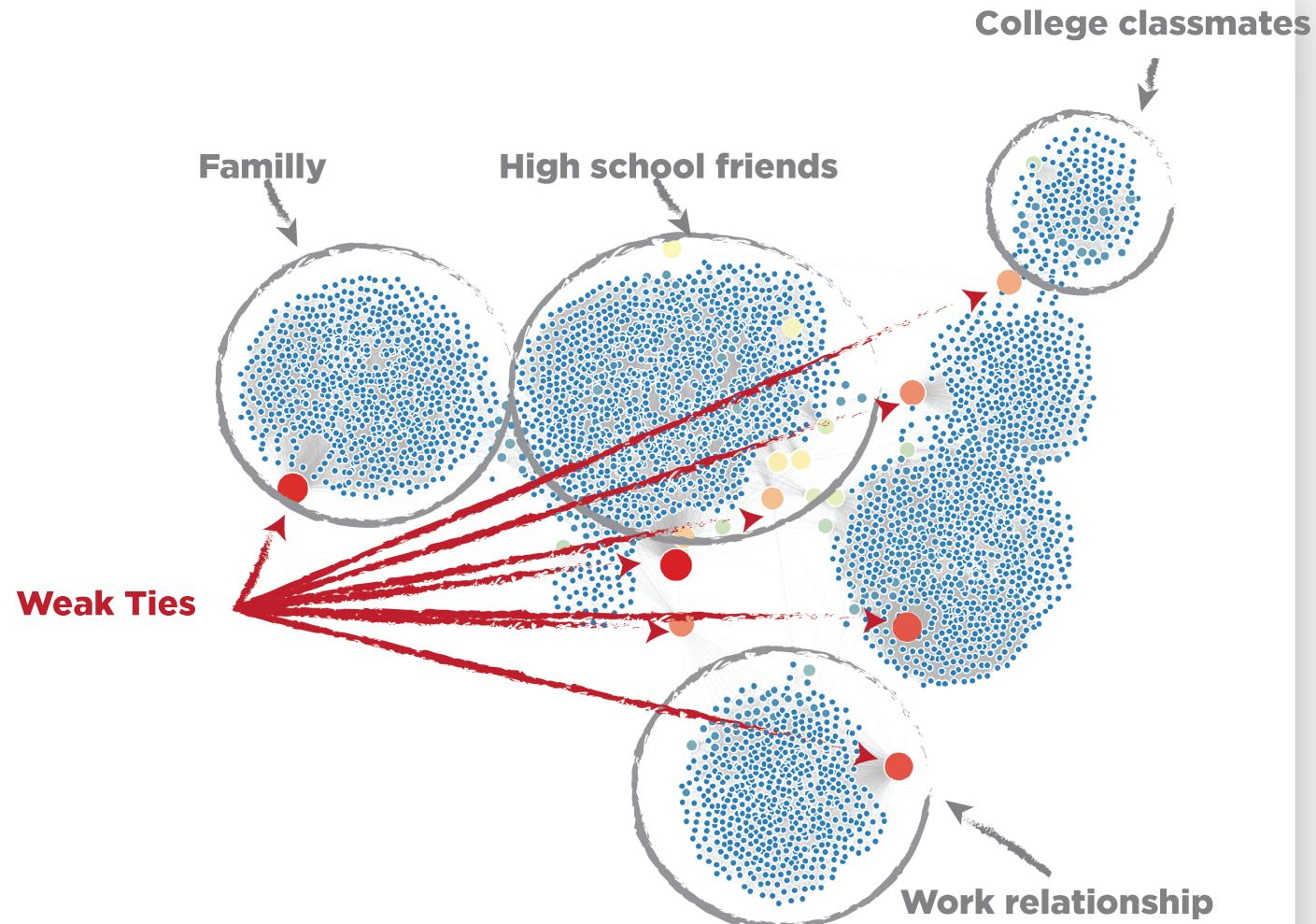
The worldwide air transportation network: Anomalous centrality, community structure, and cities' global
R. Guimerà, S. Mossa, A. Turtschi, and L. A. N. Amaral
PNAS, 2005.

Betweenness centrality in social network

Example:
Facebook
network:

nodes are sized
and colored by
betweenness.

Meaning:
High betweenness
is a sign of
person is a bridge
between different
group of people

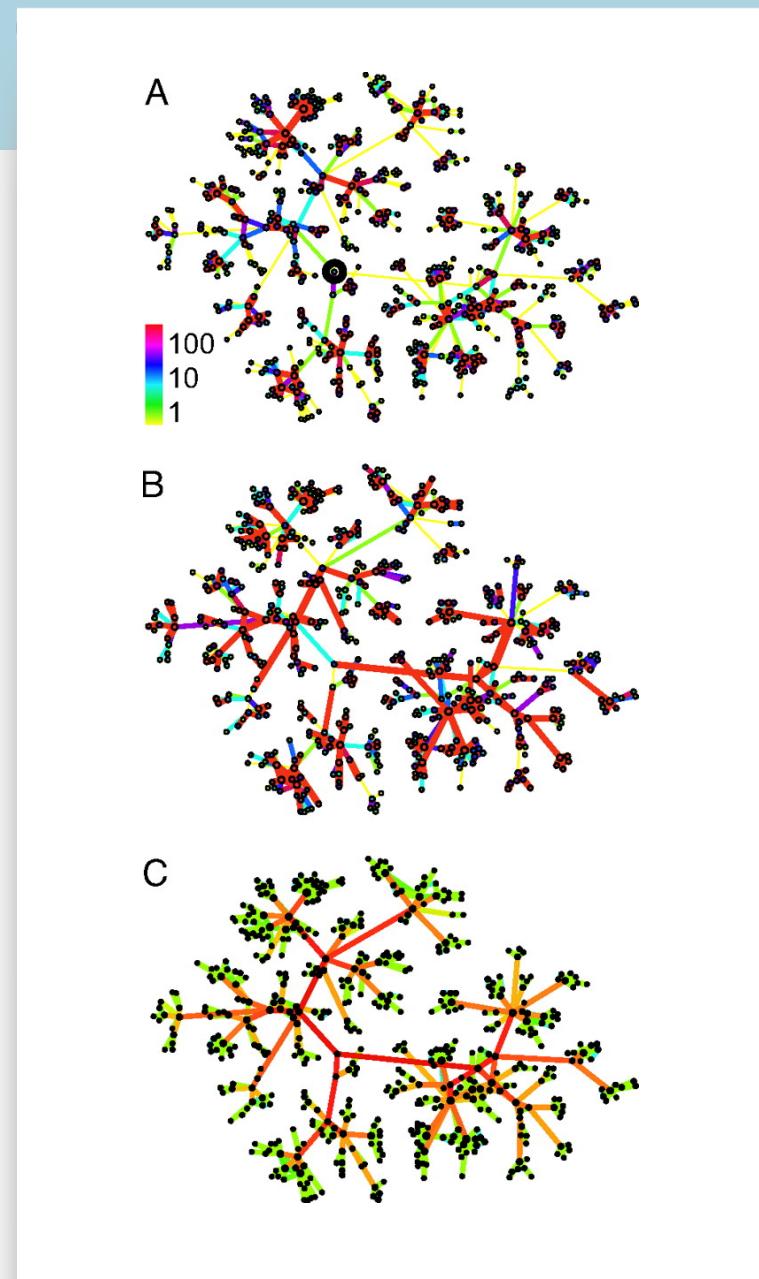


Betweenness centrality

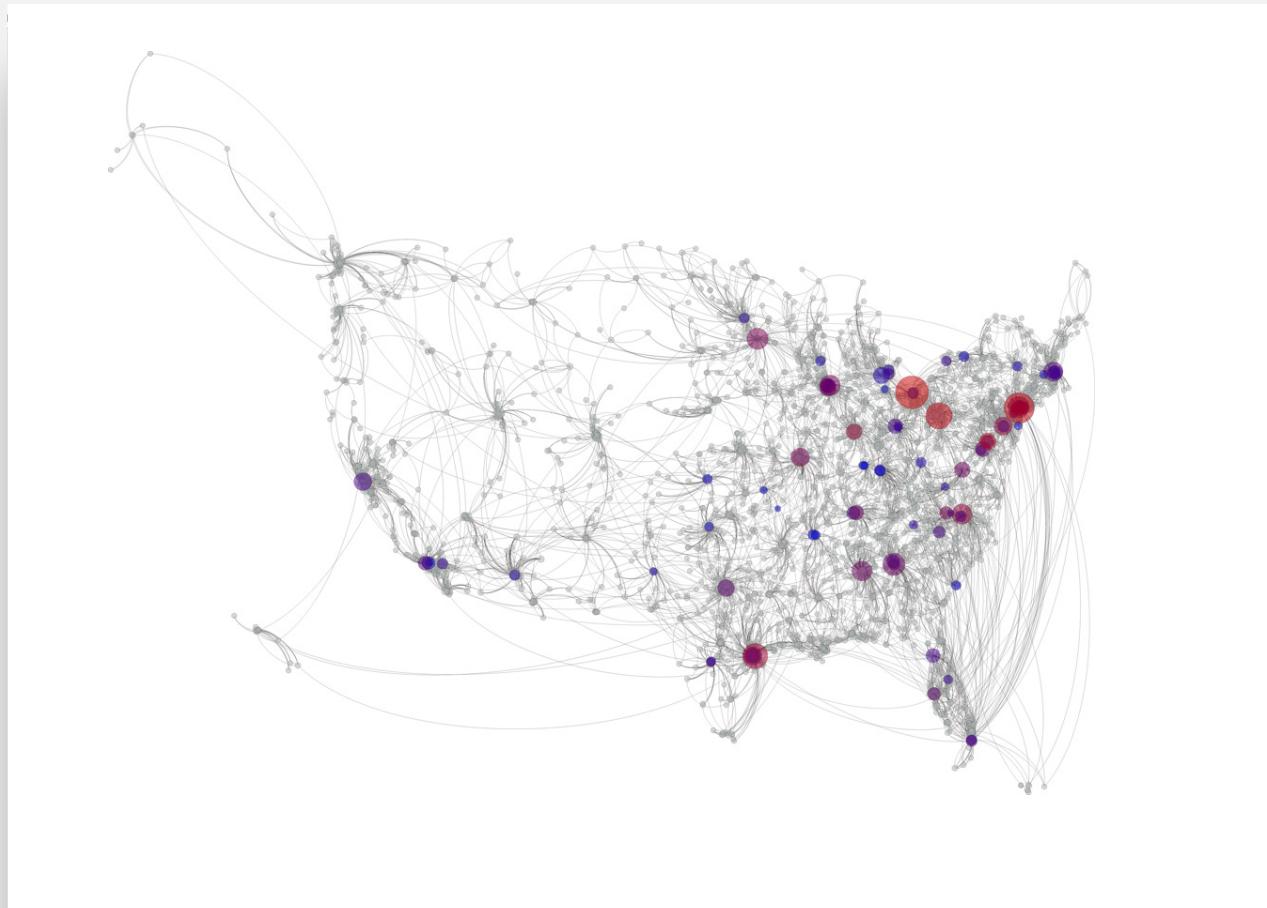
Example: mobile network

The structure of the Mobile Call Graph around a randomly chosen individual. Each link represents mutual calls between the two users, and all nodes are shown that are at distance less than six from the selected user, marked by a circle in the center.

- (A) The real tie strengths, observed in the call logs, defined as the aggregate call duration in minutes (see color bar).
- (B) The dyadic hypothesis suggests that the tie strength depends only on the relationship between the two individuals. To illustrate the tie strength distribution in this case, we randomly permuted tie strengths for the sample in A.
- (C) The weight of the links assigned on the basis of their betweenness centrality b_{ij} values for the sample in A as suggested by the global efficiency principle. In this case, the links connecting communities have high b_{ij} values (red), whereas the links within the communities have low b_{ij} values (green).



Hospital patient transfer

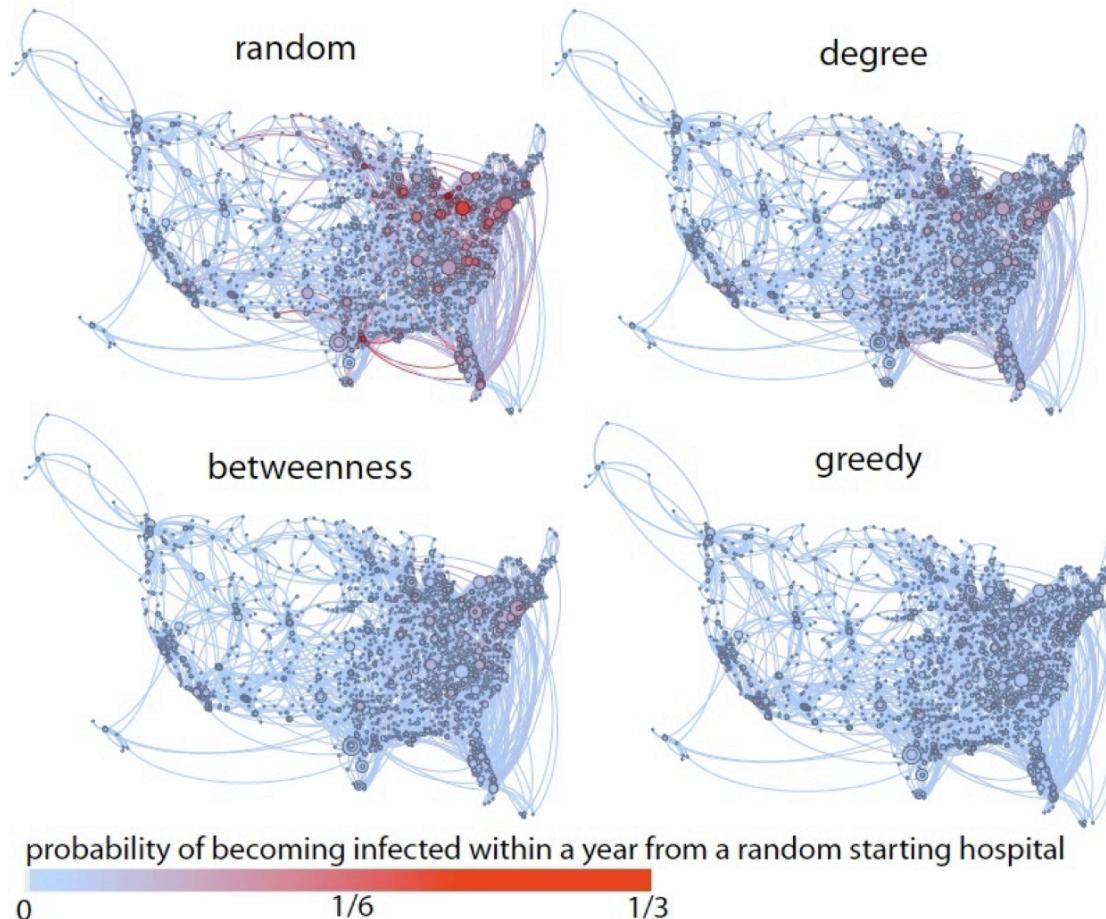


■ Limiting the spread of highly resistant hospital-acquired microorganisms via critical care transfers: a simulation study

Karkada UH, Adamic LA, Kahn JM, Iwashyna TJ
Intensive Care Med. 2011

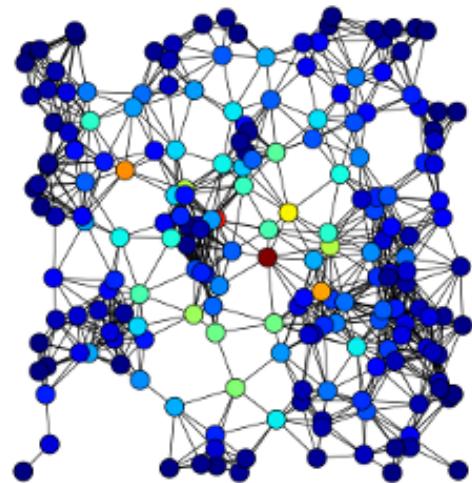
Hospital patient transfer

Infection prevention strategies in a hospital patient transfer network

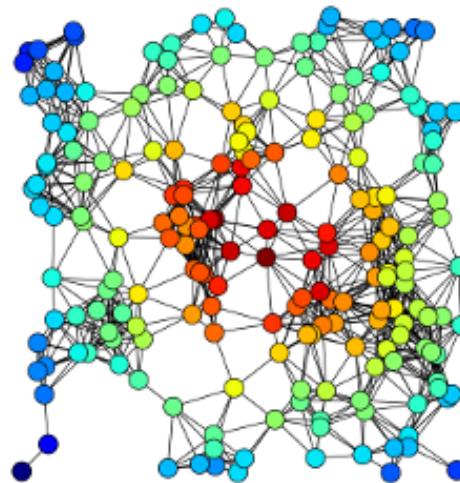


Centrality continued

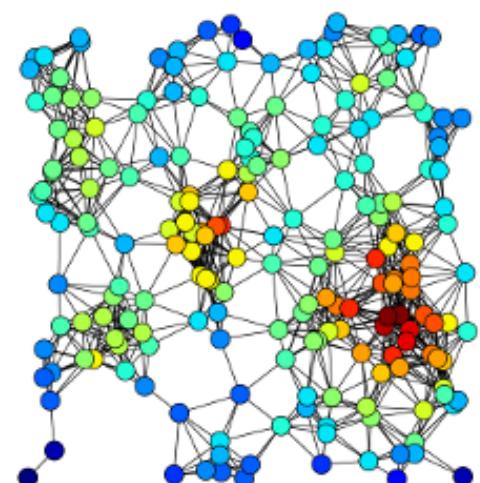
- ▶ Understanding whom are popular
- ▶ Understanding whom are at the center
- ▶ Understanding whom are receiving more in-flux



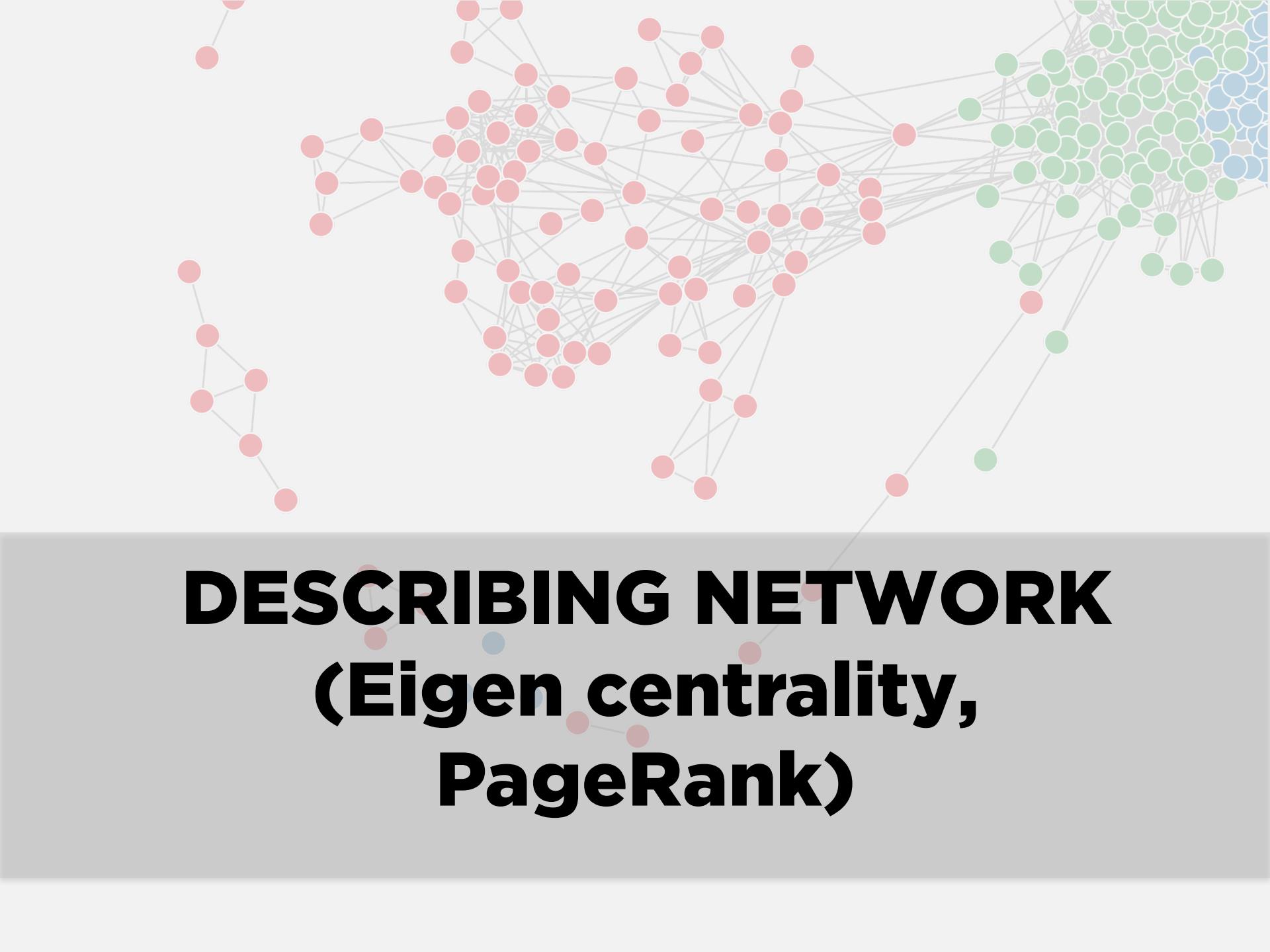
Betweenness



Closeness



Degree



DESCRIBING NETWORK

(Eigen centrality, PageRank)

Eigen centrality, PageRank

The PageRank Citation Ranking: Bringing Order to the Web

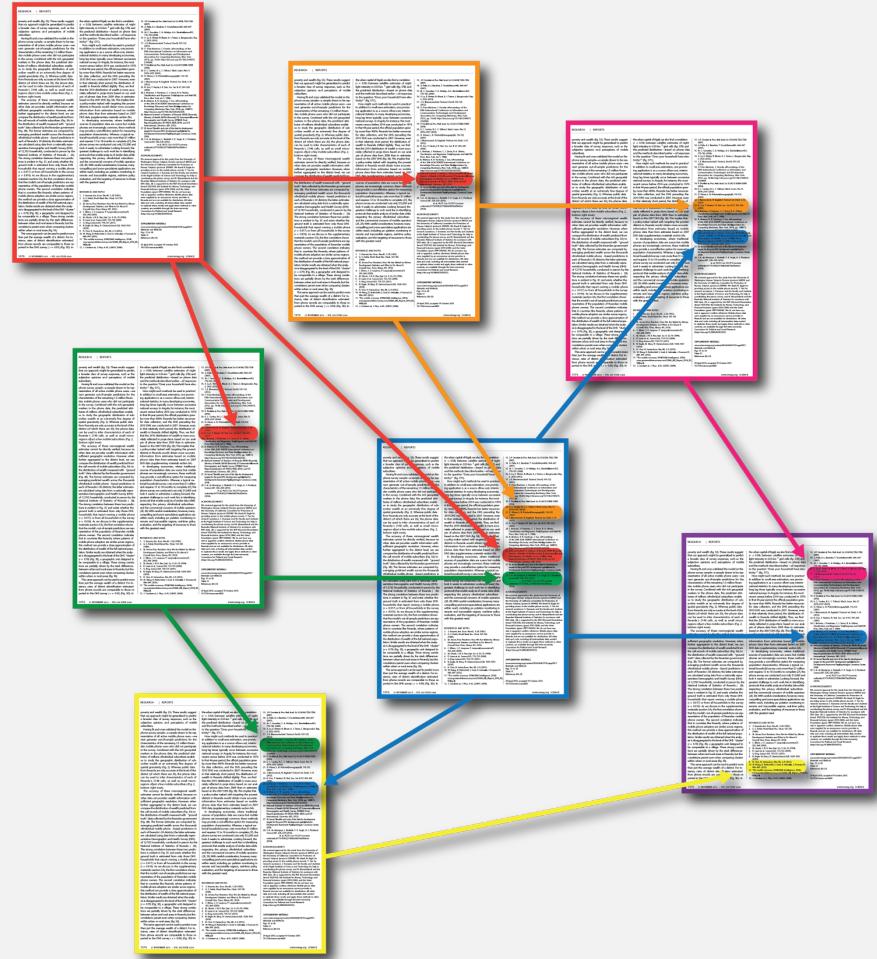
Larry Page and Sergey Brin

1999

The problem:

How to provide a good algorithm to **rank** web page according to their importance in a given topics.

Same problem **rank** research paper according to their importance.

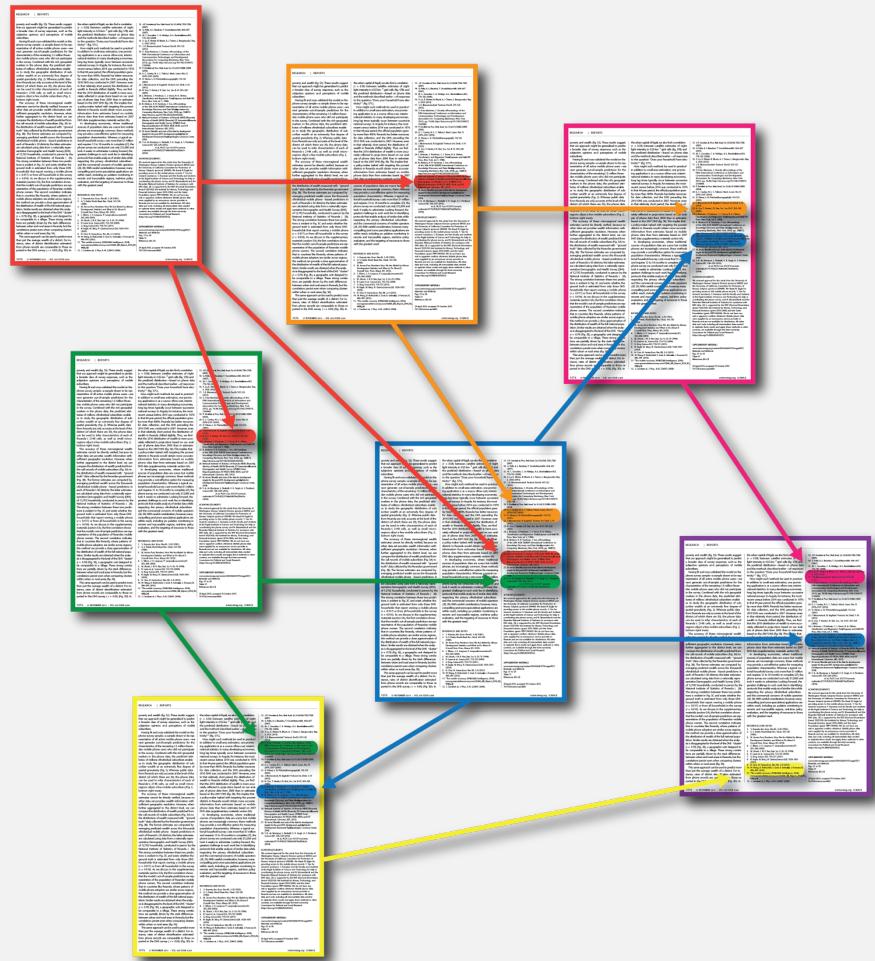


Eigen centrality, PageRank

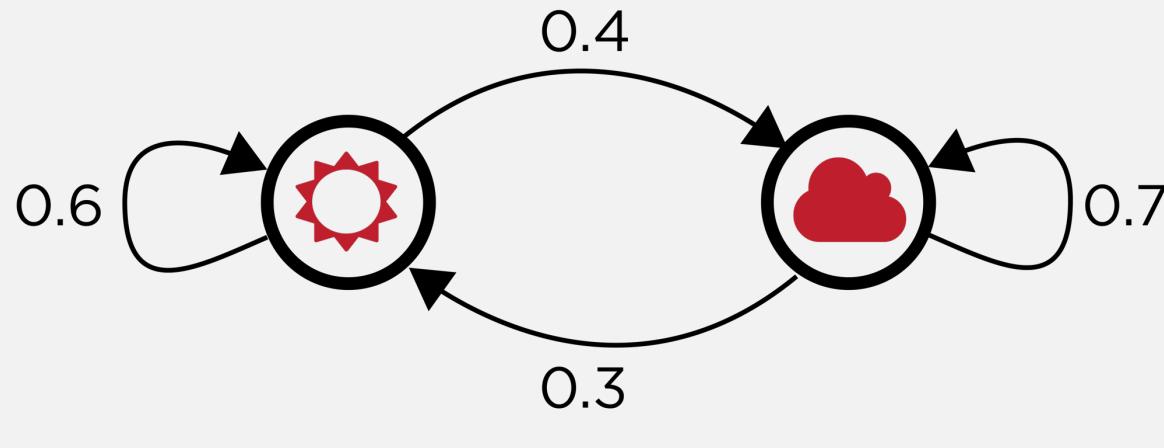
The Idea behind Eigen Centrality:

- ▶ Define π_i as the importance of the node i
- ▶ π_i depend on π_j if j is a neighbor of i
- ▶ The importance is transmitted through the network

$$\pi_i \propto \sum_j a_{ji} \pi_j$$



Remember the Markov Chains



Weather Tomorrow

Weather Prediction

Weather Today

$$\begin{bmatrix} \text{Sun} \\ \text{Cloud} \end{bmatrix} = \begin{bmatrix} \text{Sun} & \text{Cloud} \\ \text{Cloud} & \text{Cloud} \end{bmatrix}^T \times \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Probability of a sunny weather tomorrow given that today it is cloudy

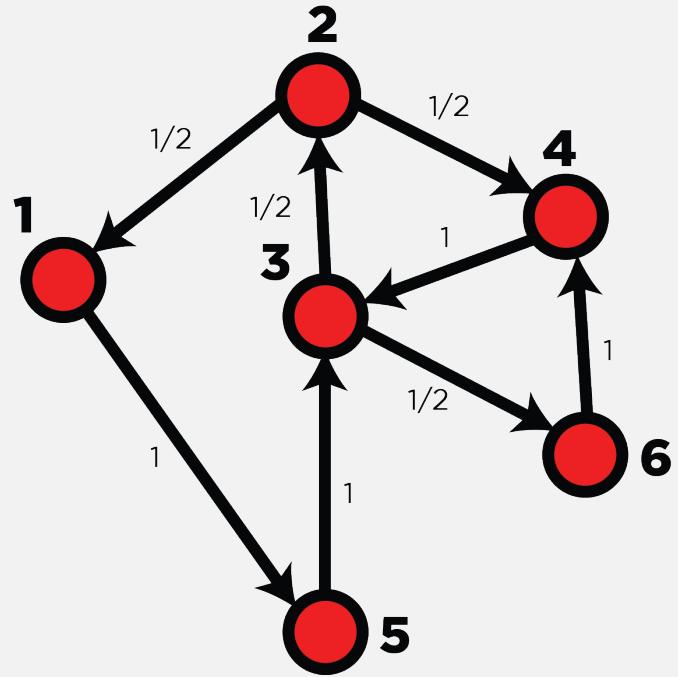
Weather of the day after Tomorrow

Weather Prediction

Weather Today

$$\begin{bmatrix} \text{Sun} \\ \text{Cloud} \end{bmatrix} = \begin{bmatrix} \text{Sun} & \text{Cloud} \\ \text{Cloud} & \text{Cloud} \end{bmatrix}^T \times \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Eigen centrality

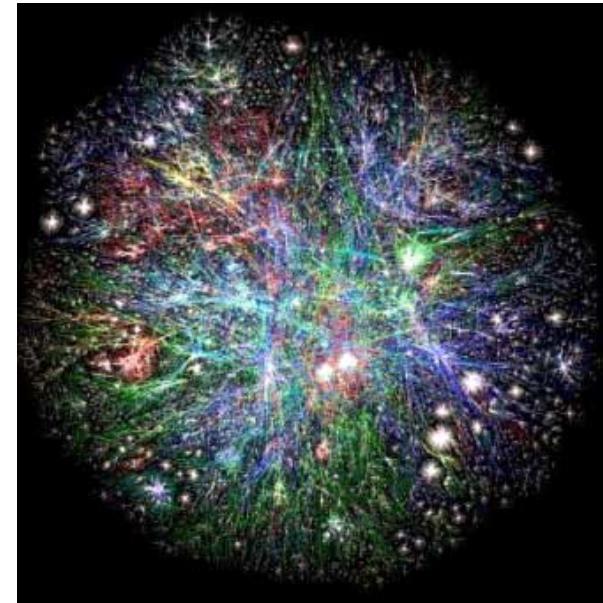


$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 1/2 & 0 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 & 1/2 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

Row stochastic matrix

$$\pi_0 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Eigen centrality



$$\begin{aligned} \pi_1 &= \mathbf{A}^T \pi_0 \\ \pi_2 &= \mathbf{A}^T \pi_1 \\ &\vdots \\ \pi_k &= \mathbf{A}^T \pi_{k-1} \\ \pi_k &= (\mathbf{A}^T)^k \pi_0 \end{aligned}$$

π_k

Steady state

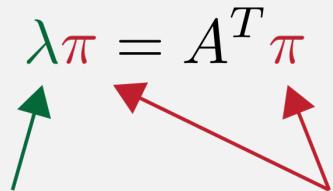
Eigen centrality

continued

$$\pi_{k+1} = A^T \pi_k$$

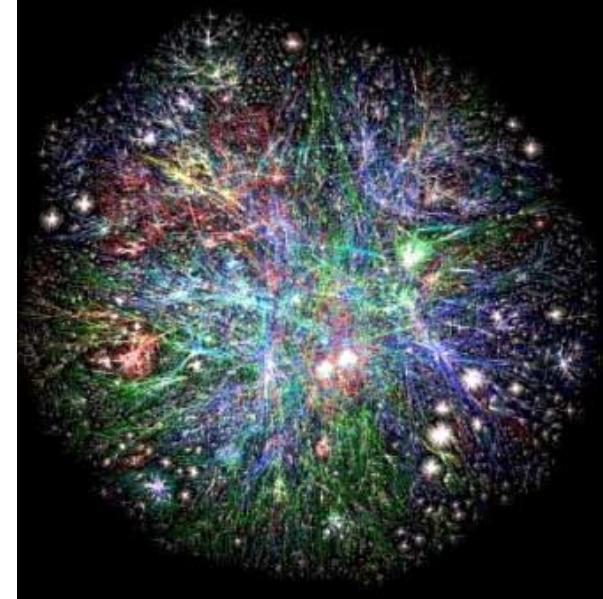
$$\underbrace{\pi}_{\text{Stationary state}} = A^T \pi \text{ when } k \rightarrow \infty$$

Stationary state

$$\lambda \pi = A^T \pi$$


Eigen value = 1

Eigen vector

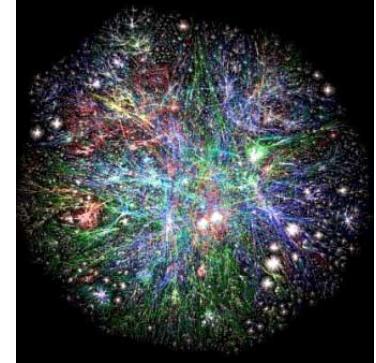


Perron-Frobenius theorem

- ▶ If A is a positive matrix and primitive
 - ✓ A has a real eigenvalue
 - ✓ There exist a eigenvalue $\lambda_{max} = 1$ (if A is stochastic row vector)

Eigen centrality

continued

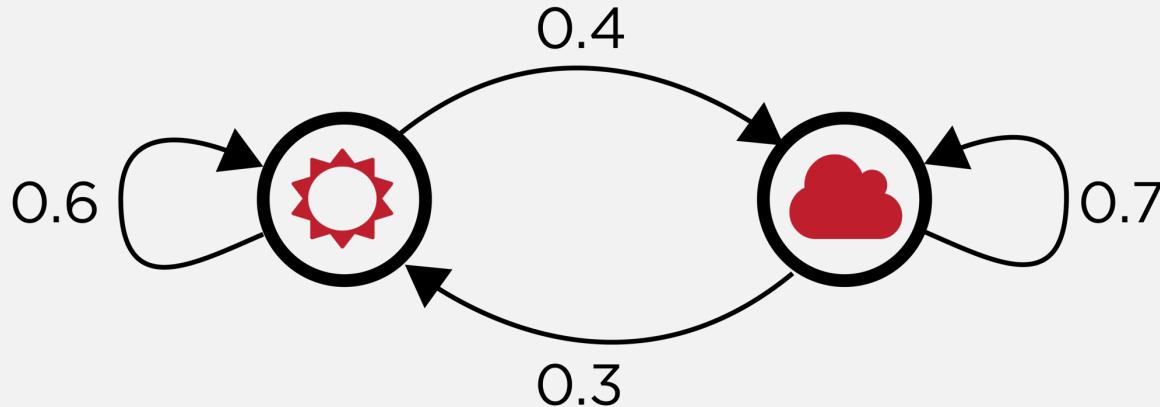


How to understand Perron-Frobenius theorem:

- ▶ The network is irreducible: meaning there is only one component: “all the node are reachable”
- ▶ The network is aperiodic: meaning there is no sink state in the network

The Eigen value problem

finding the eigen value



$$\mathbf{A}^T \pi = \lambda \pi$$

$$\mathbf{A}^T \pi - \lambda \pi = \vec{0}$$

$$[\mathbf{A}^T - \lambda \mathbf{I}] \pi = \vec{0}$$

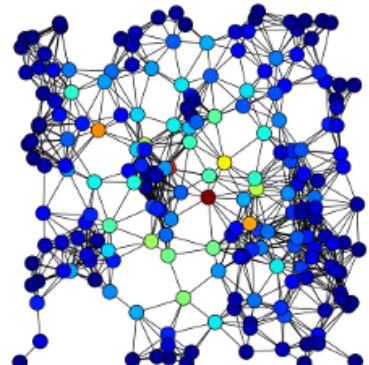
$$\det [\mathbf{A}^T - \lambda \mathbf{I}] = 0$$

Find the eigenvalue λ by solving the equation

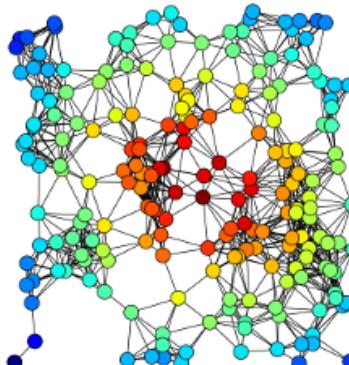
solve
$$\begin{vmatrix} 0.6 - \lambda & 0.3 \\ 0.4 & 0.7\lambda \end{vmatrix} = 0$$

Centrality continued

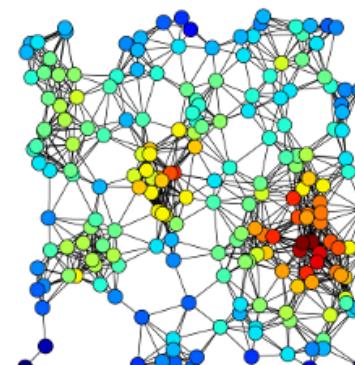
- ▶ Understanding whom are popular
- ▶ Understanding whom are at the center
- ▶ Understanding whom are receiving more in-flux
- ▶ Receiving more Trust



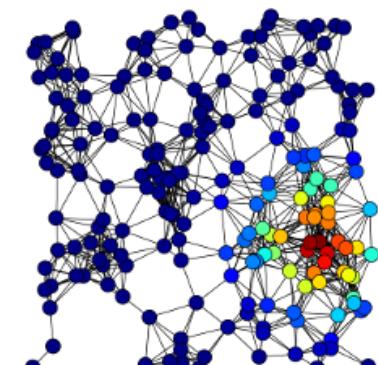
Betweenness



Closeness



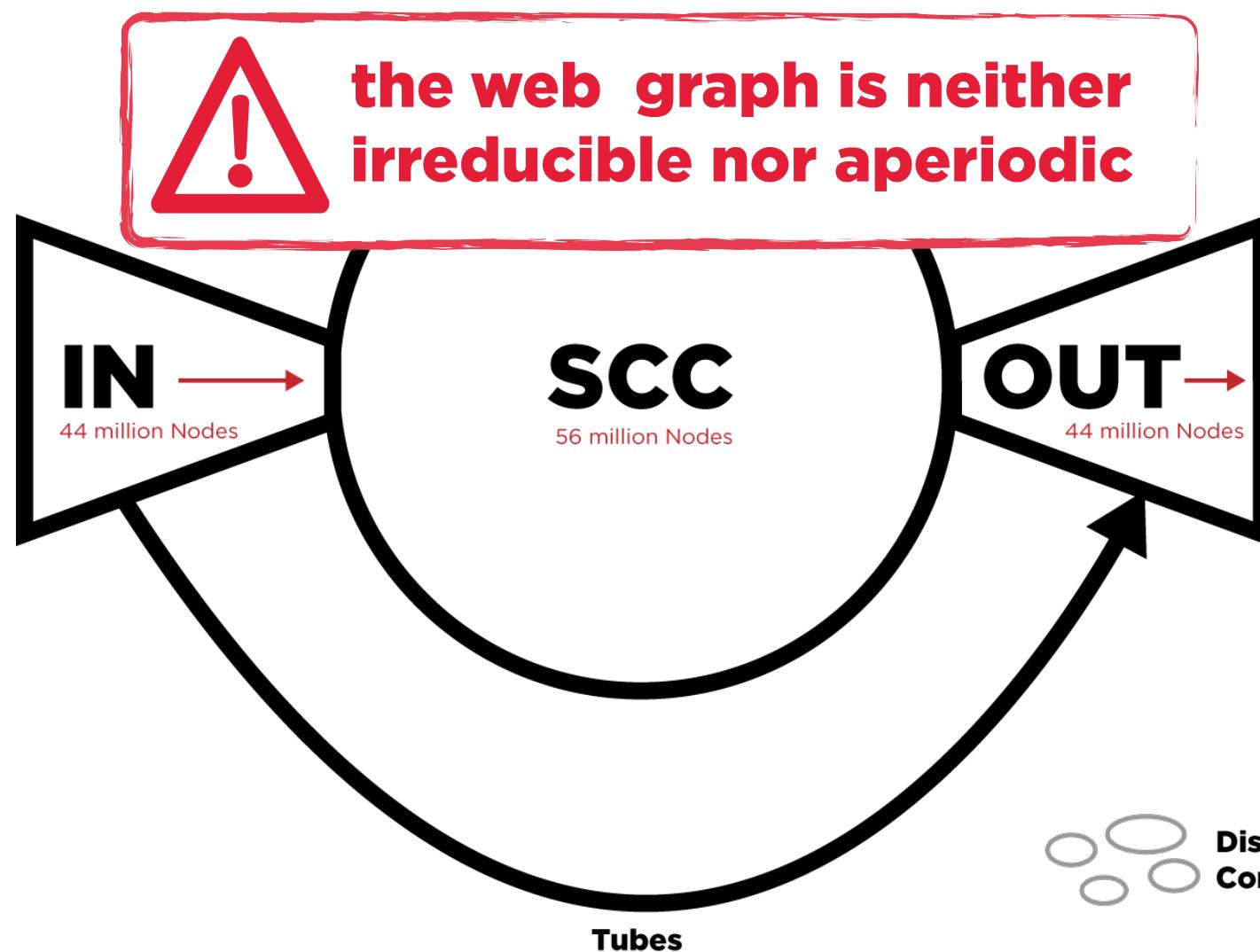
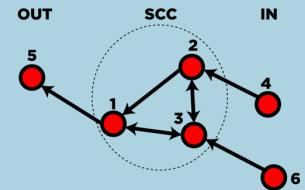
Degree



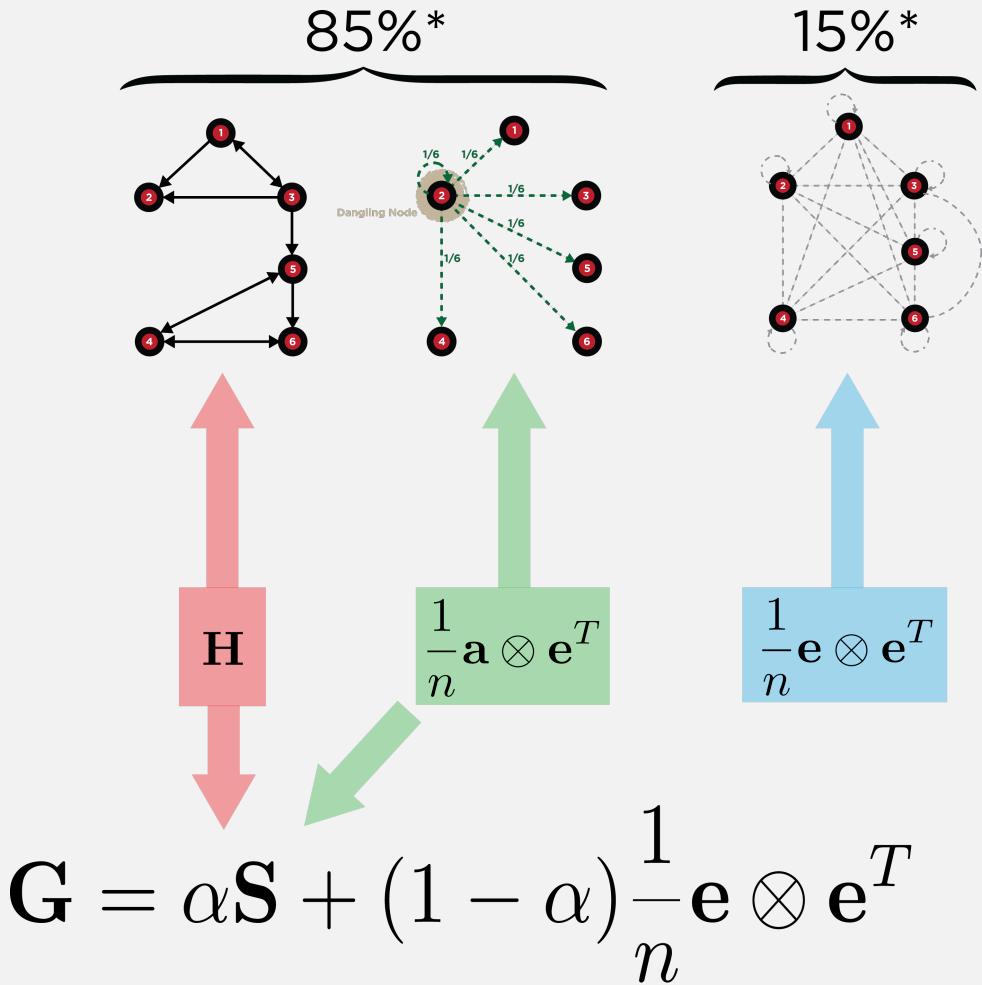
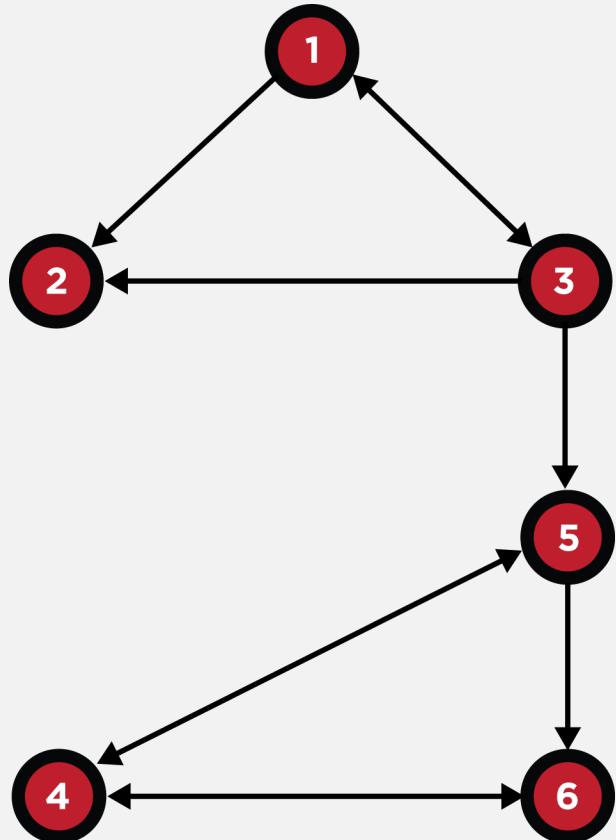
Eigen centrality

Wait ... the SCC

of the Web Graph

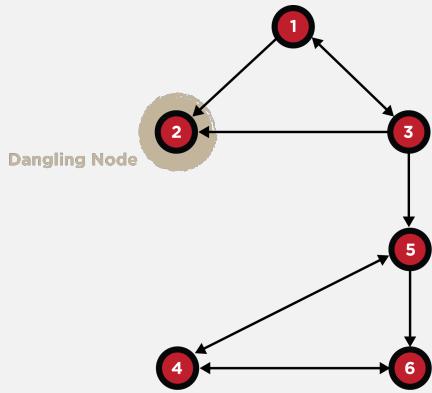


PageRank



*Pour $\alpha = 0.85$

Build the Google matrix



$$\mathbf{a} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\mathbf{G} = \alpha S + (1 - \alpha) \frac{1}{n} \mathbf{e} \otimes \mathbf{e}^T$$

$$\mathbf{S} = \mathbf{H} + \frac{1}{n} \mathbf{a} \otimes \mathbf{e}^T$$

$$\frac{1}{n} \mathbf{e} \otimes \mathbf{e}^T = \begin{bmatrix} 1/n & 1/n & 1/n & 1/n & 1/n & 1/n \\ 1/n & 1/n & 1/n & 1/n & 1/n & 1/n \\ 1/n & 1/n & 1/n & 1/n & 1/n & 1/n \\ 1/n & 1/n & 1/n & 1/n & 1/n & 1/n \\ 1/n & 1/n & 1/n & 1/n & 1/n & 1/n \\ 1/n & 1/n & 1/n & 1/n & 1/n & 1/n \end{bmatrix}$$

Power iteration to solve the PR

Data: Google Matrice \mathbf{G} , Nombre d'itération $MaxIter$, Tolérance ϵ

Réultat: Vecteur \mathbf{x}

$\mathbf{x} := 1/n$

$\mathbf{xold} := 0$

for i in $range(MaxIter)$ **do**

$\mathbf{xold}^T := \mathbf{x}^T$

$\mathbf{x}^T := \mathbf{x}^T \mathbf{G}$

if $\epsilon > n \|\mathbf{x} - \mathbf{xold}\|$ **then**

return \mathbf{x}^T

end

end

$$\frac{1}{n} \mathbf{a} \otimes \mathbf{e}^T = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1/n & 1/n & 1/n & 1/n & 1/n & 1/n \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Google search engine



Google general overview:

1. Compute the PageRank for the web graph
2. For a request filter by keywords
3. Display result as function of their PageRank

Filter by keywords

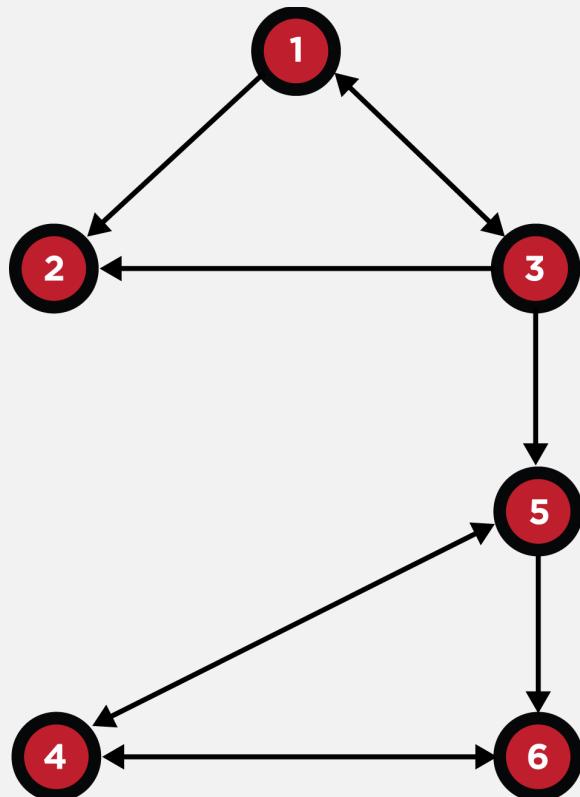
Display result as function of their PR

The screenshot shows a Google search results page for the query "graph theory". The search bar at the top contains the query. Below it, there are tabs for "Web", "Images", "Videos", "Books", "Maps", and "More". A message about cookie usage is displayed. The search results list several links:

- Graph theory - Wikipedia, the free encyclopedia**
en.wikipedia.org/wiki/Graph_theory
In mathematics and computer science, graph theory is the study of graphs, which are mathematical structures used to model pairwise relations between objects.
- Graph Theory Tutorials**
www.utm.edu/departments/math/graph/ -
This is the home page for a series of short interactive tutorials introducing the basic concepts of graph theory. There is not a great deal of theory here, we will just ...
- Graph Theory - graph theory textbooks and resources**
www.graphtheory.com/ -
Requiring only a foundation in discrete mathematics, it can serve as the textbook in a combinatorial methods course or in a combined graph theory and ...
- [PDF] GRAPH THEORY**
math.tut.fi/~ruohonen/GT_English.pdf -
an introduction to basic concepts and results in graph theory, with a special ... One of the usages of graph theory is to give a unified formalism for many very ...

On the right side of the results, there are several small images related to graph theory, such as a network graph, a pentagonal prism, a map, and a cycle graph. Below the images, there is a "More Images" button. At the bottom of the page, there is a section titled "Graph theory" with a brief description and a "Wikipedia" link. There are also sections for "Related topics" and "Explore: Graph coloring".

Google search engine



Search for: **Algorithms**

Page Id	PR	Keywords
x_1	0.5	Flower, Garden, Pool
x_2	0.2	News, Computer Science
x_3	0.1	Computer Science, Algorithms
x_4	0.1	Algorithms, Graph, Dijkstra
x_5	0.05	Algorithms, Graph, Dijkstra, C++
x_6	0.05	Network Science, Graphs, Algorithms

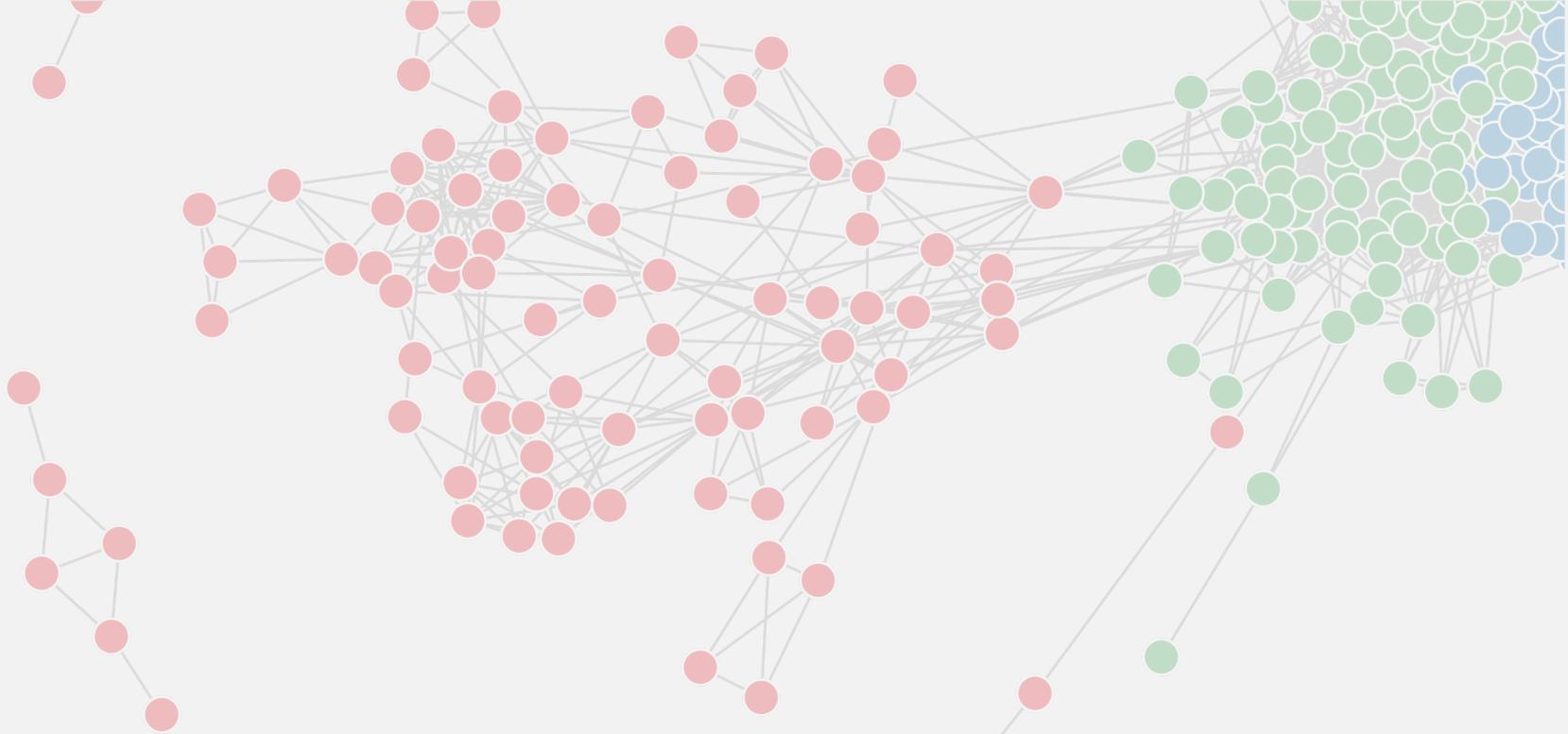
Ordered by their PR

**Return
Results**

Many other applications

Ranking anything

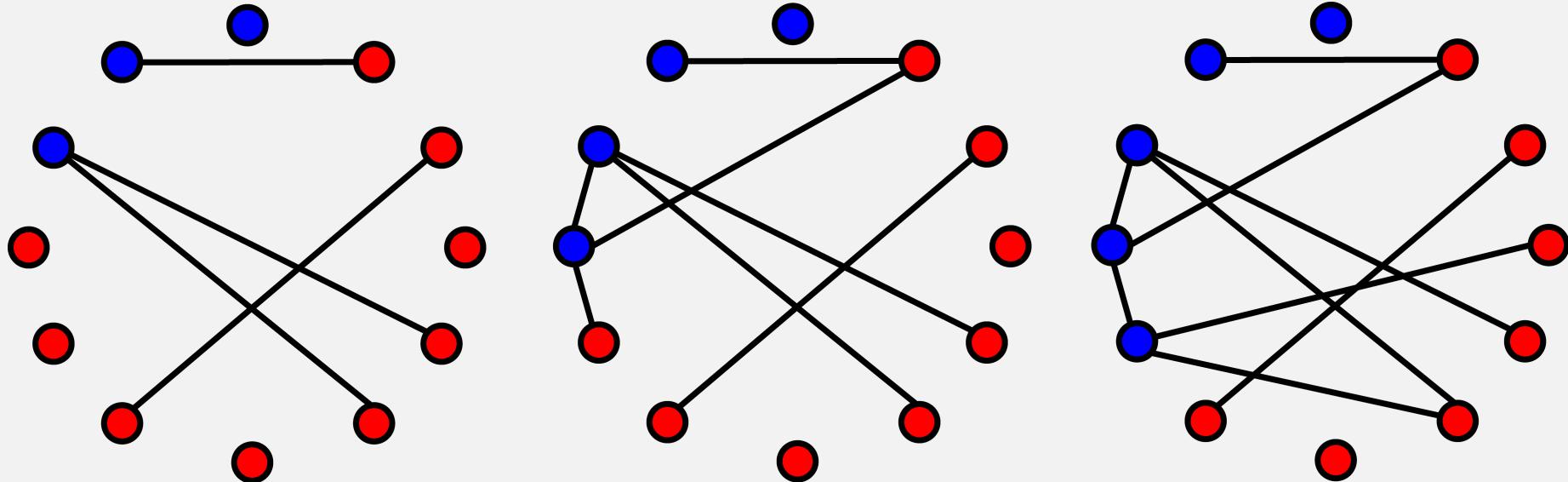
- ▶ Language 
- ▶ Social Network (klout) 
- ▶ Bitcoin
- ▶ College Football



RANDOM NETWORK (ERDOS-RENY)

Random Network Models

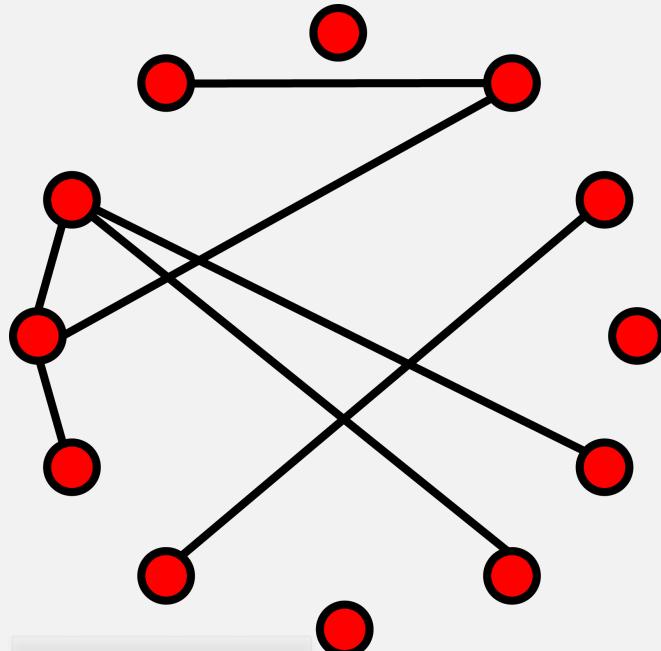
Erdos-Reny Model



The **NULL** model

Erdos-Reny Model (1960)

Connect with probability p



$$N = 12, p = 1/10$$
$$\langle k \rangle = 1.1$$



Pál Erdős (1913-1996)



Alfréd Rényi (1921-1970)

Definition:

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

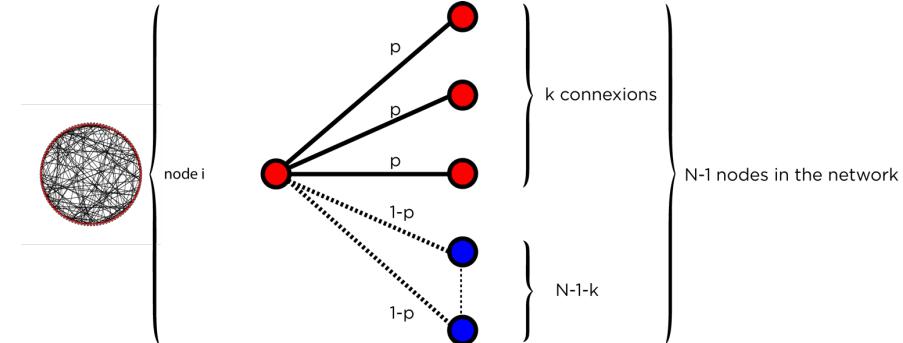
We will call it $G(N, p)$.

Degree distribution

Each node has $(N - 1)$ tries to get edges

Each try is a success with probability p

The binomial distribution gives us the probability that a node has degree k



Probability of having exactly k edges:

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

G(N,p)-model:

For each potential edge we flip a biased coin:

with probability p we add the edge
with probability $(1-p)$ we don't

How many edges per node?

$$\langle k \rangle = (N-1)p$$

$$\sigma_k \approx \frac{1}{(N-1)^{1/2}}$$

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 $\langle k \rangle$.

Reminder on Binomial Distribution

$$\begin{aligned}\langle k \rangle &= E(K) = \sum_{k=0}^{N-1} k P_k = \sum_{k=0}^{N-1} k \binom{N-1}{k} p^k (1-p)^{N-1-k} \\&= \sum_{k=1}^{N-1} k \frac{(N-1)!}{k!(N-1-k)!} p^k (1-p)^{N-1-k} \\&= (N-1)p \sum_{k=1}^{N-1} \frac{(N-2)!}{(k-1)!(N-1-k)!} p^{k-1} (1-p)^{N-1-k}, \quad a = k-1, b = N-2, N-1 = b+1, k = a+1 \\&= (N-1)p \sum_{a=0}^b \frac{b!}{a!(b-a)!} p^a (1-p)^{b-a} \\&= (N-1)p \sum_{a=0}^b \binom{b}{a} p^a (1-p)^{b-a}\end{aligned}$$

Average Degree of G(N,p)

$$\langle k \rangle = (N-1)p$$

Approximations

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

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

$$p_k = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(k-z)^2}{2\sigma^2}}$$

Binomial

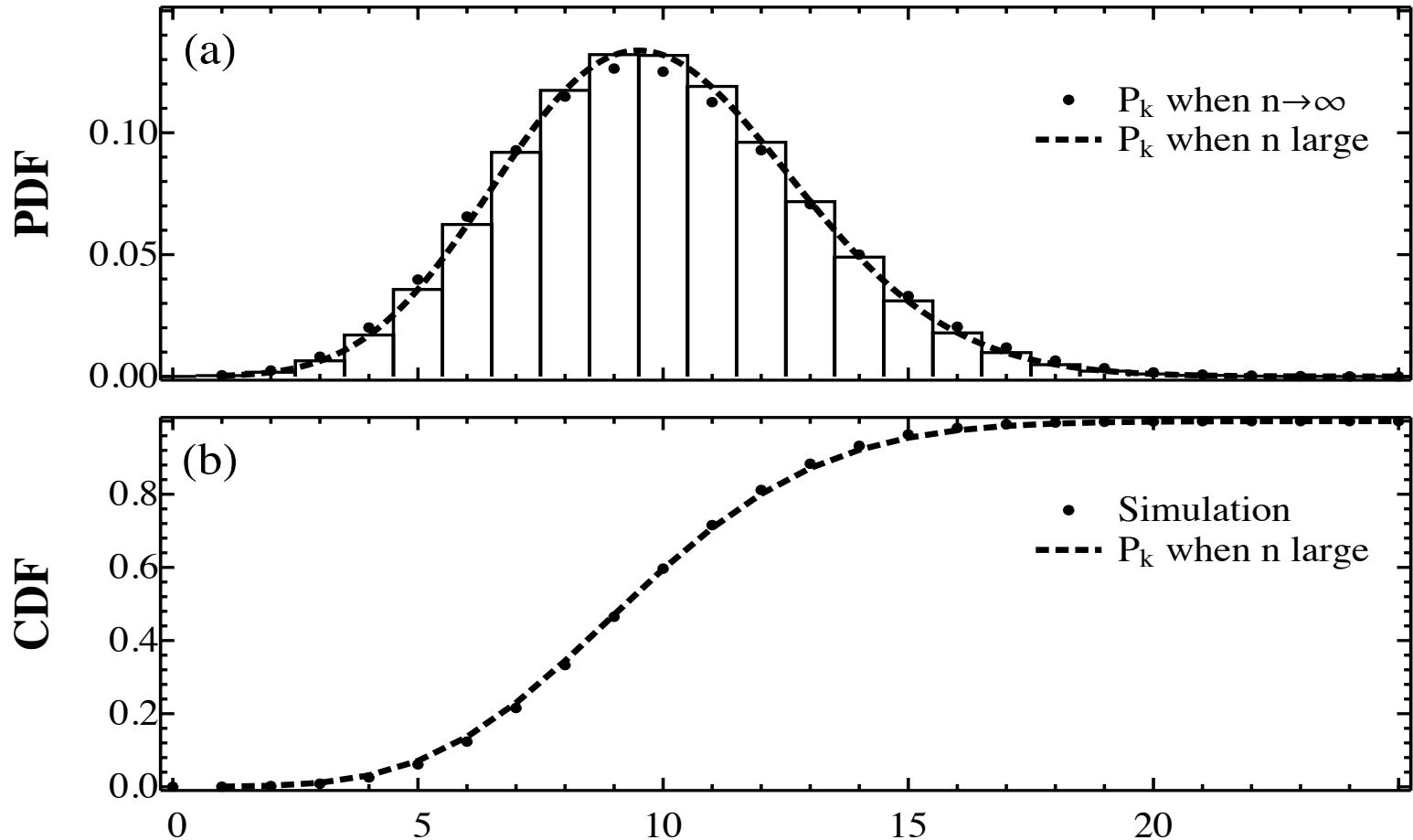
limit p small

Poisson

limit large n

Normal

Degree Distribution



Quiz Q1

The maximum degree of a node in a simple (no multiple edges between the same two nodes) N node graph is:

- a) N
- b) $N - 1$
- c) $N / 2$

Quiz Q2

What is the probability that a specific node in a N nodes graph has exactly 4 edges (when N=100 and p=0.1):

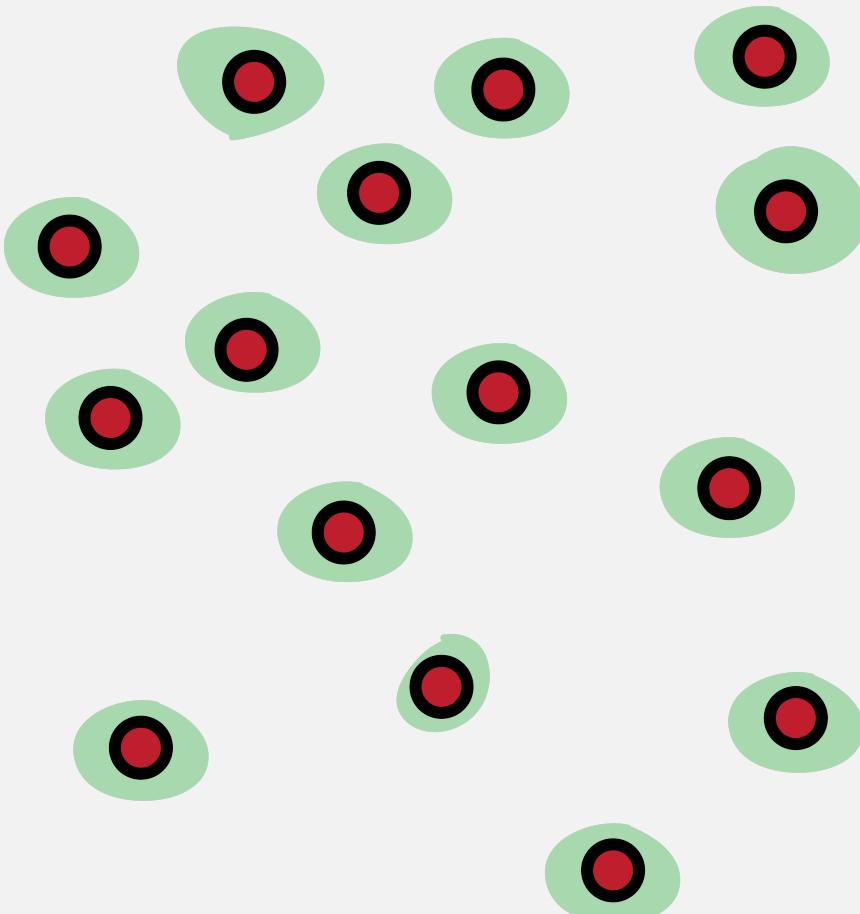
$$P_4 = \frac{99!}{4!95!} 0.1^4 (1 - 0.1)^{95}$$

Quiz Q3

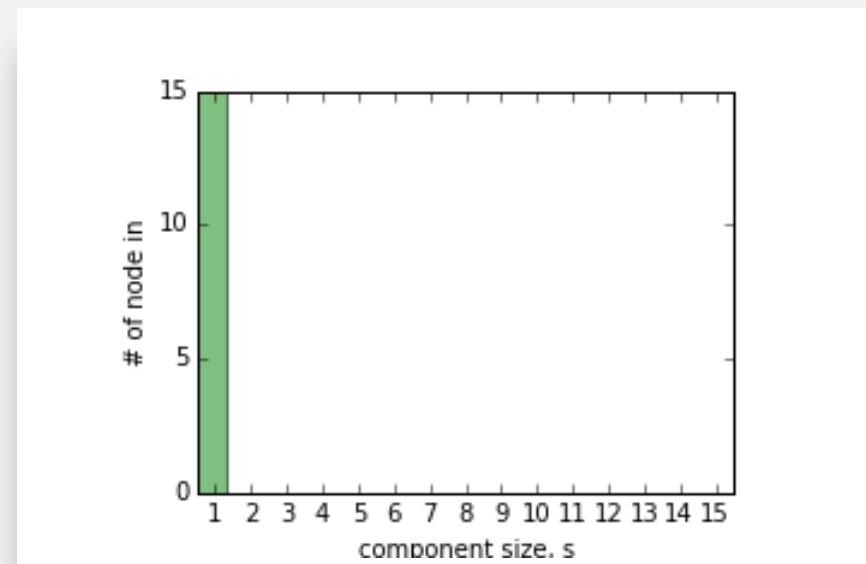
What is the average degree a node in a network $G(N,p)$ where $N=20$, $p=0.1$:

- a) 1.0
- b) 1.9
- c) 5.4

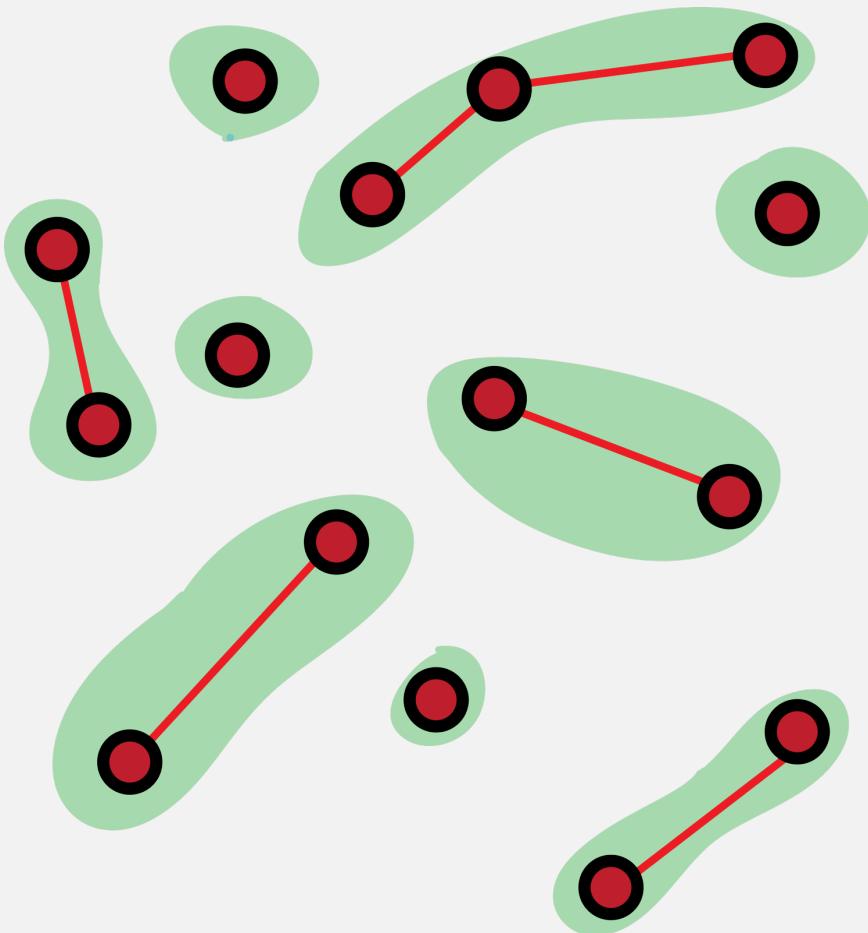
Percolation



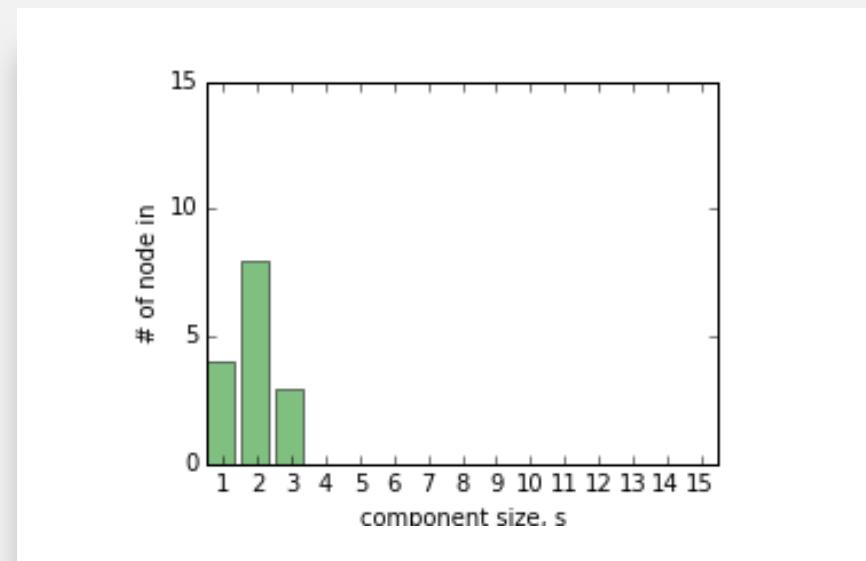
The percolation game:
choose X random pairs
connect them repeat (count #
of components)



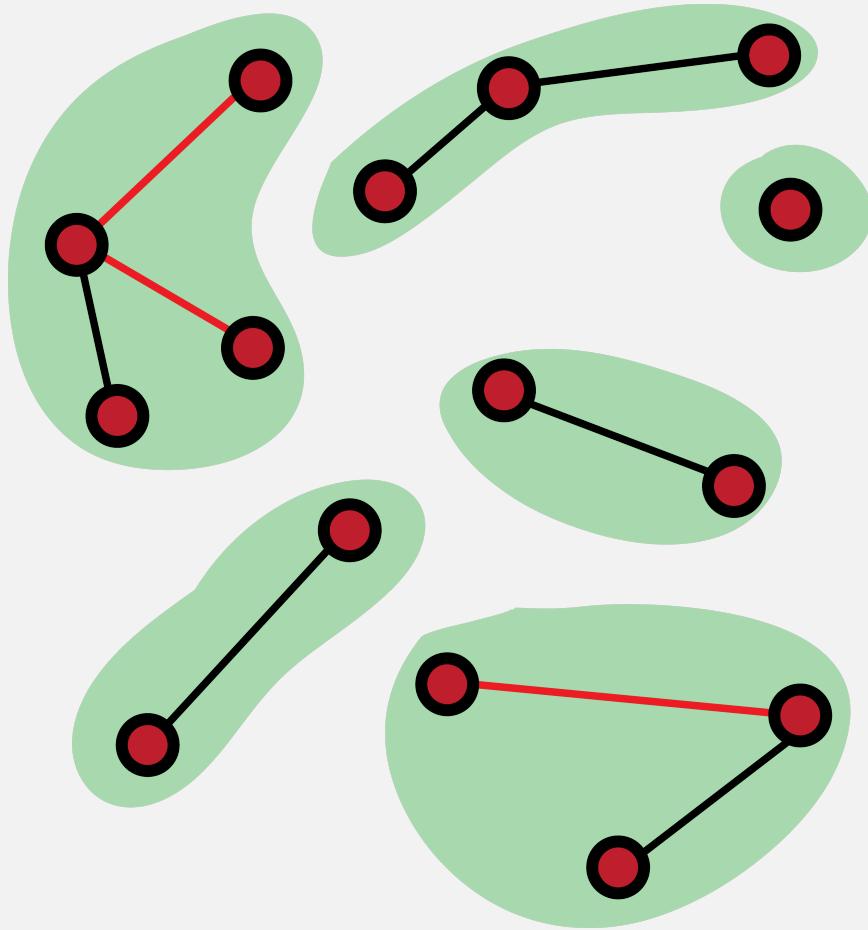
Percolation



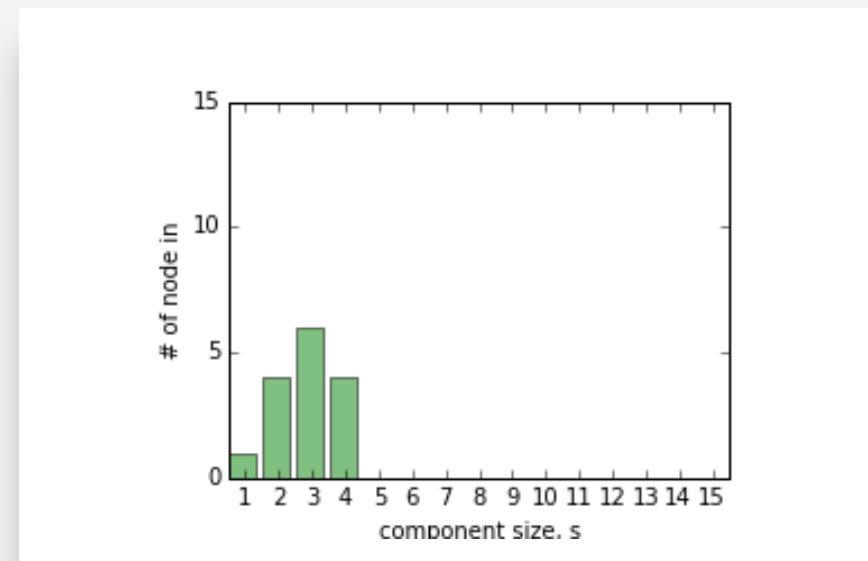
The percolation game:
choose X random pairs
connect them repeat (count #
of components)



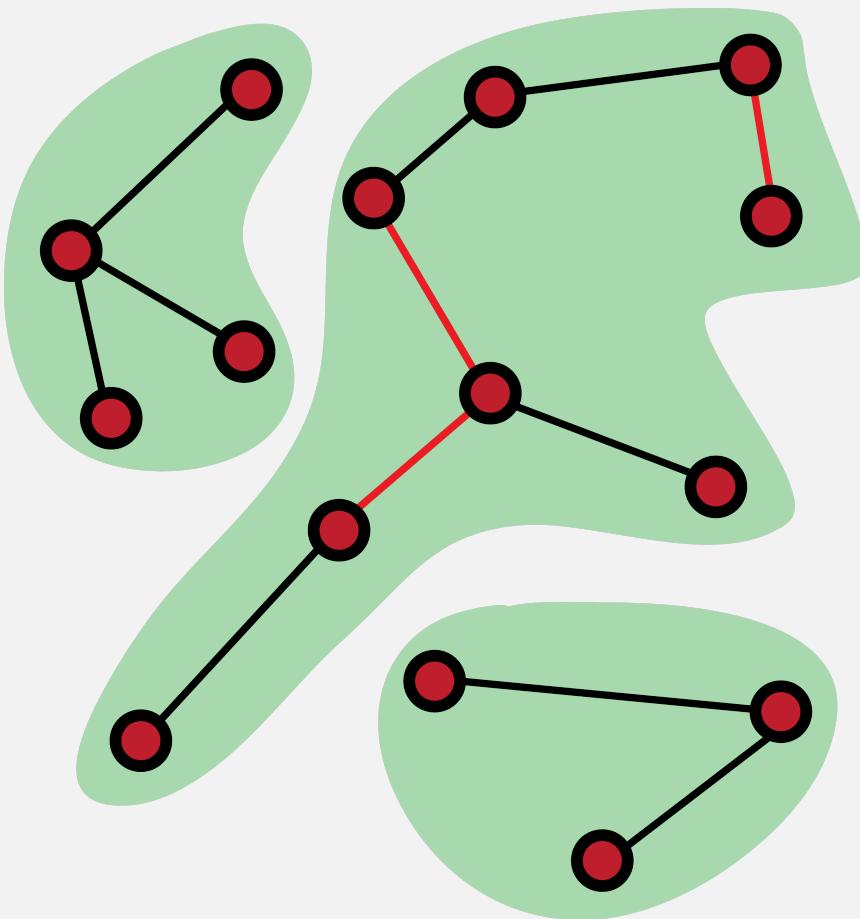
Percolation



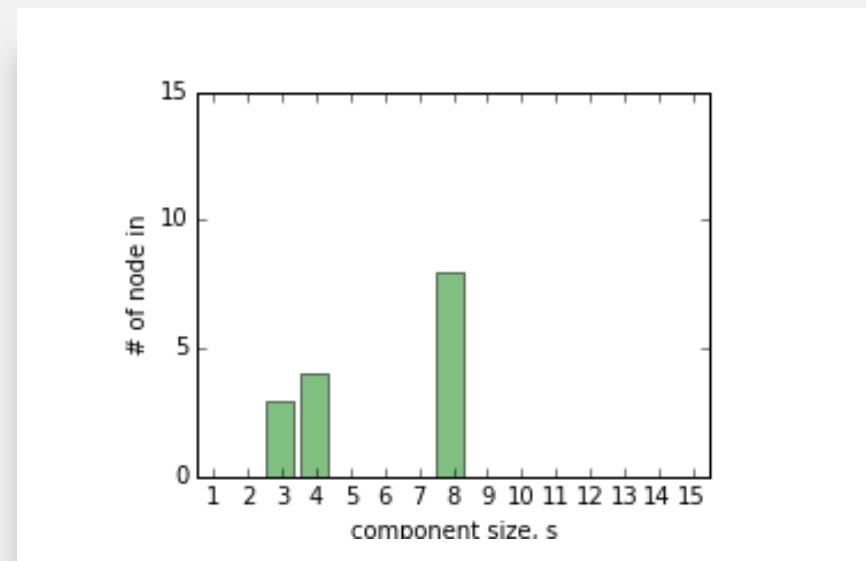
The percolation game:
choose X random pairs
connect them repeat (count #
of components)



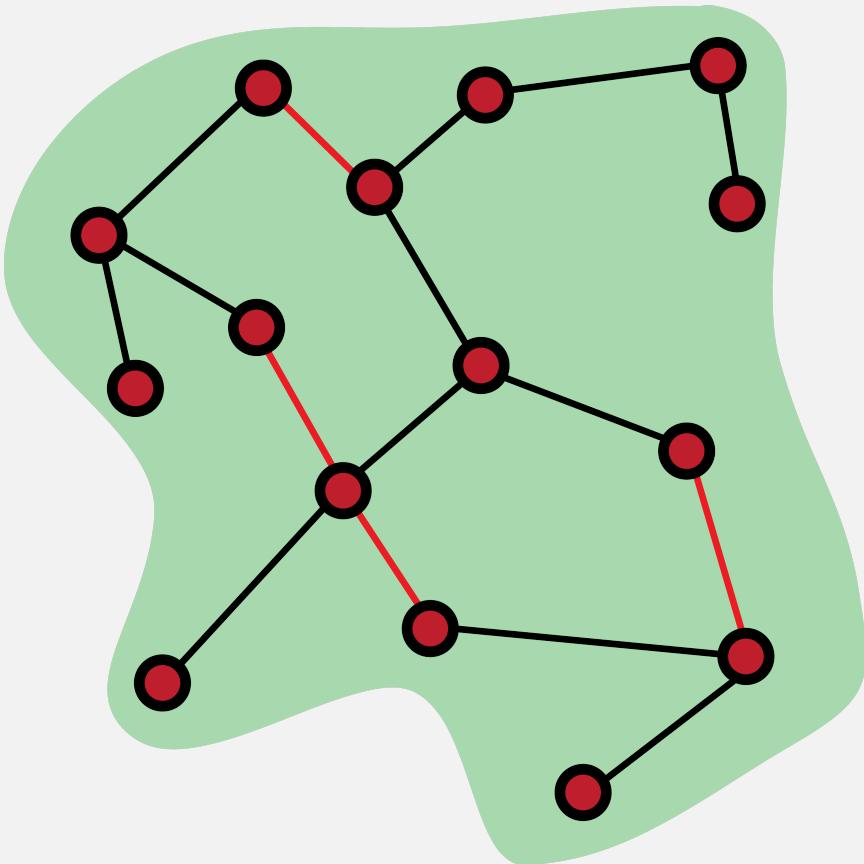
Percolation



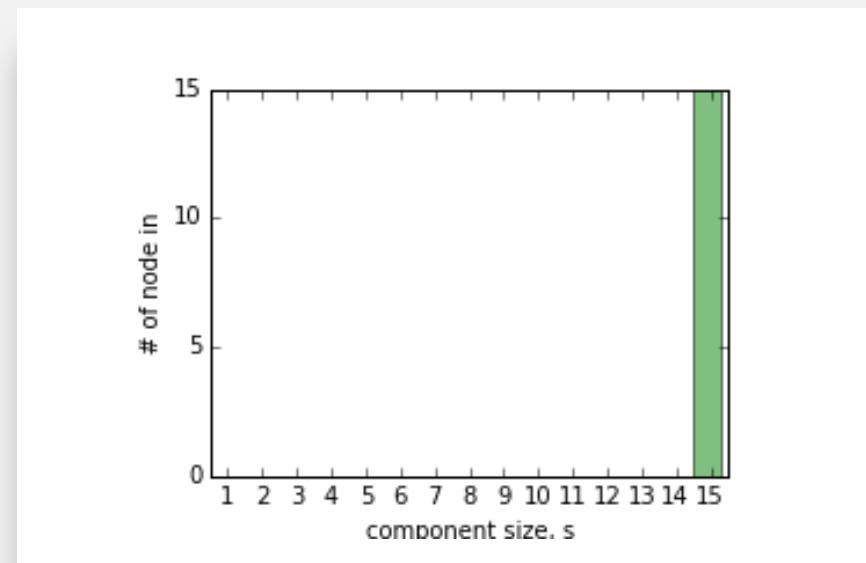
The percolation game:
choose X random pairs
connect them repeat (count #
of components)



Percolation



The percolation game:
choose X random pairs
connect them repeat (count #
of components)

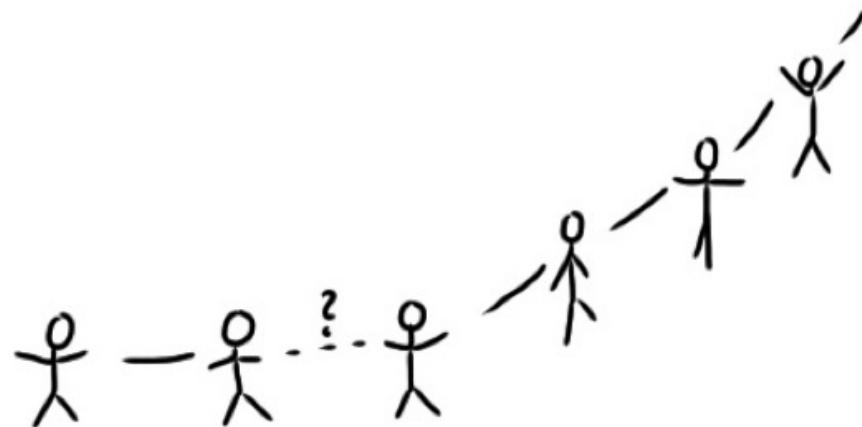


Quiz Q5:

What is the average degree $\langle k \rangle$ at which the giant component start to emerge:

- a) 5
- b) 3
- c) 1

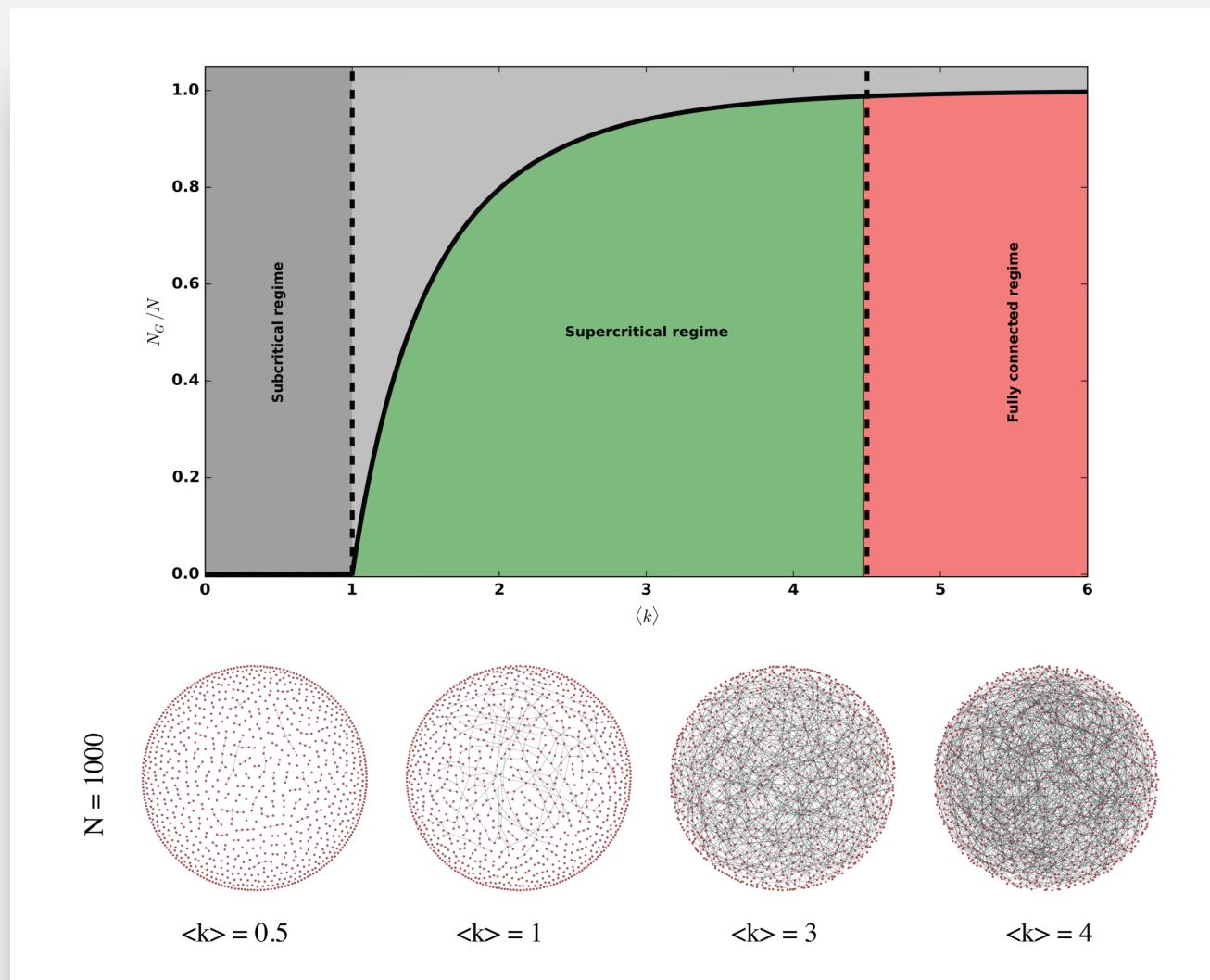
Giant component – another angle



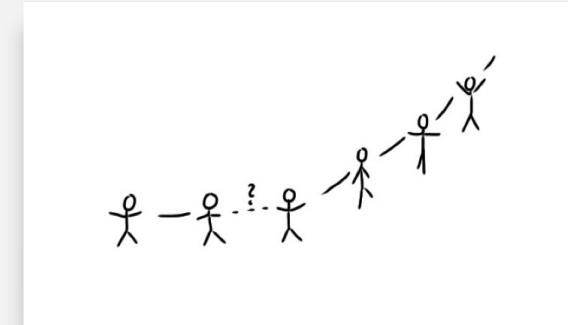
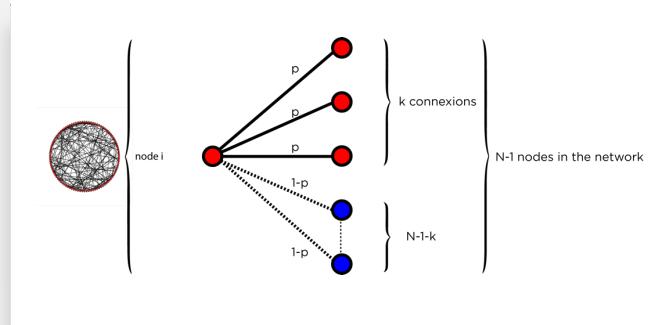
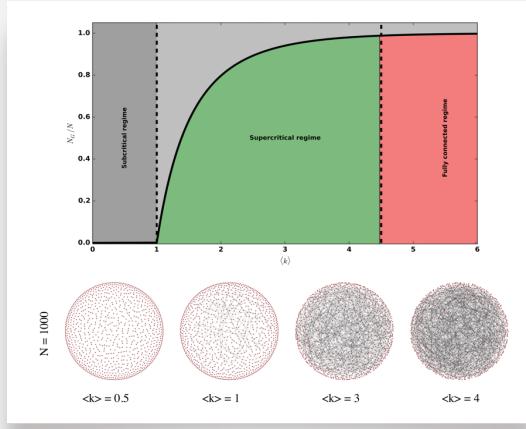
By property of degree distribution:

the average degree of your friends, you excluded, is $\langle k \rangle$ so at $\langle k \rangle = 1$, each of your friends is expected to have another friend, who in turn have another friend, etc. the giant component emerges

Critical Regimes

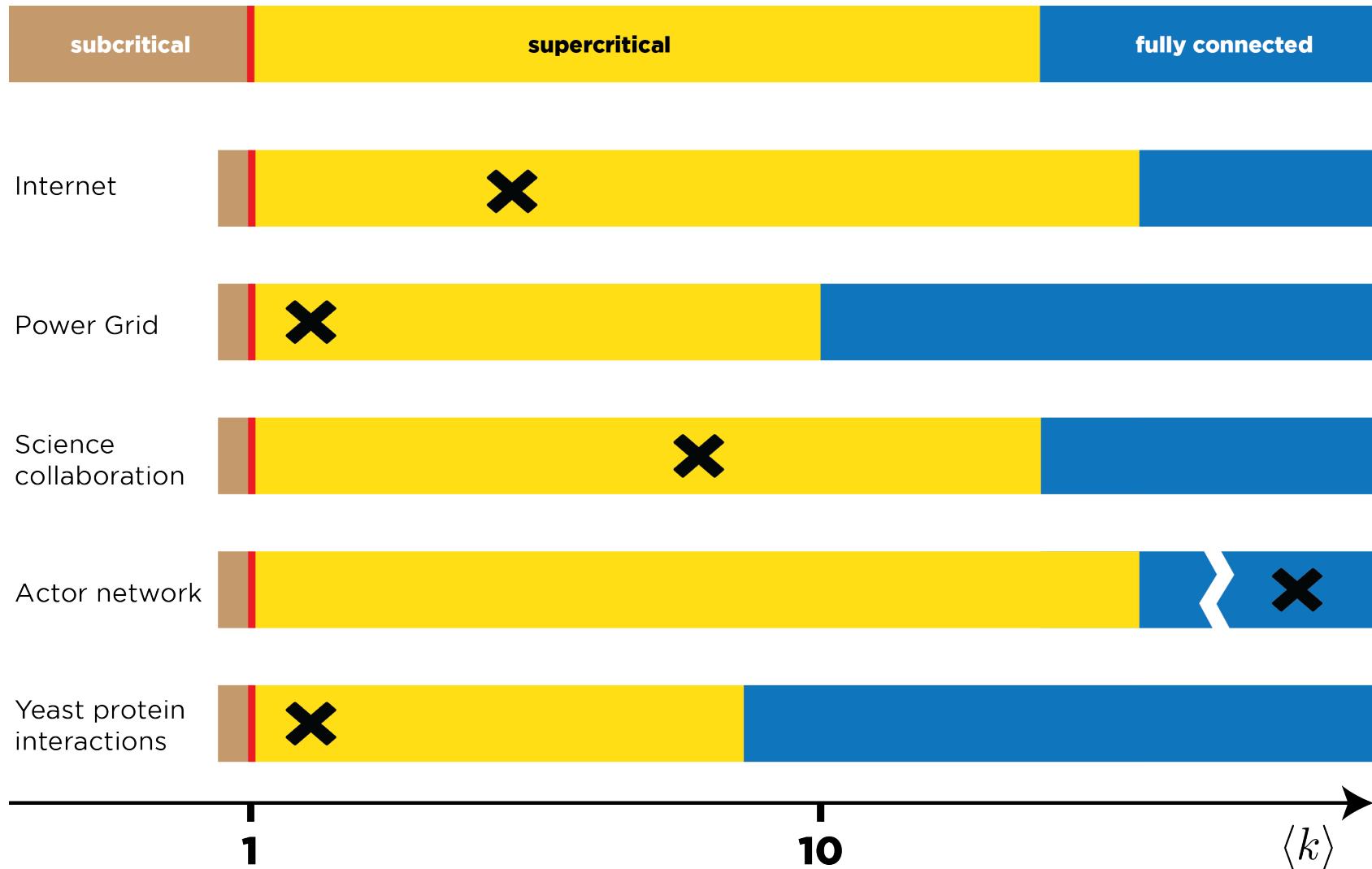


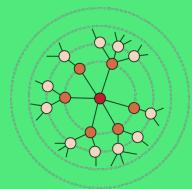
Critical regimes



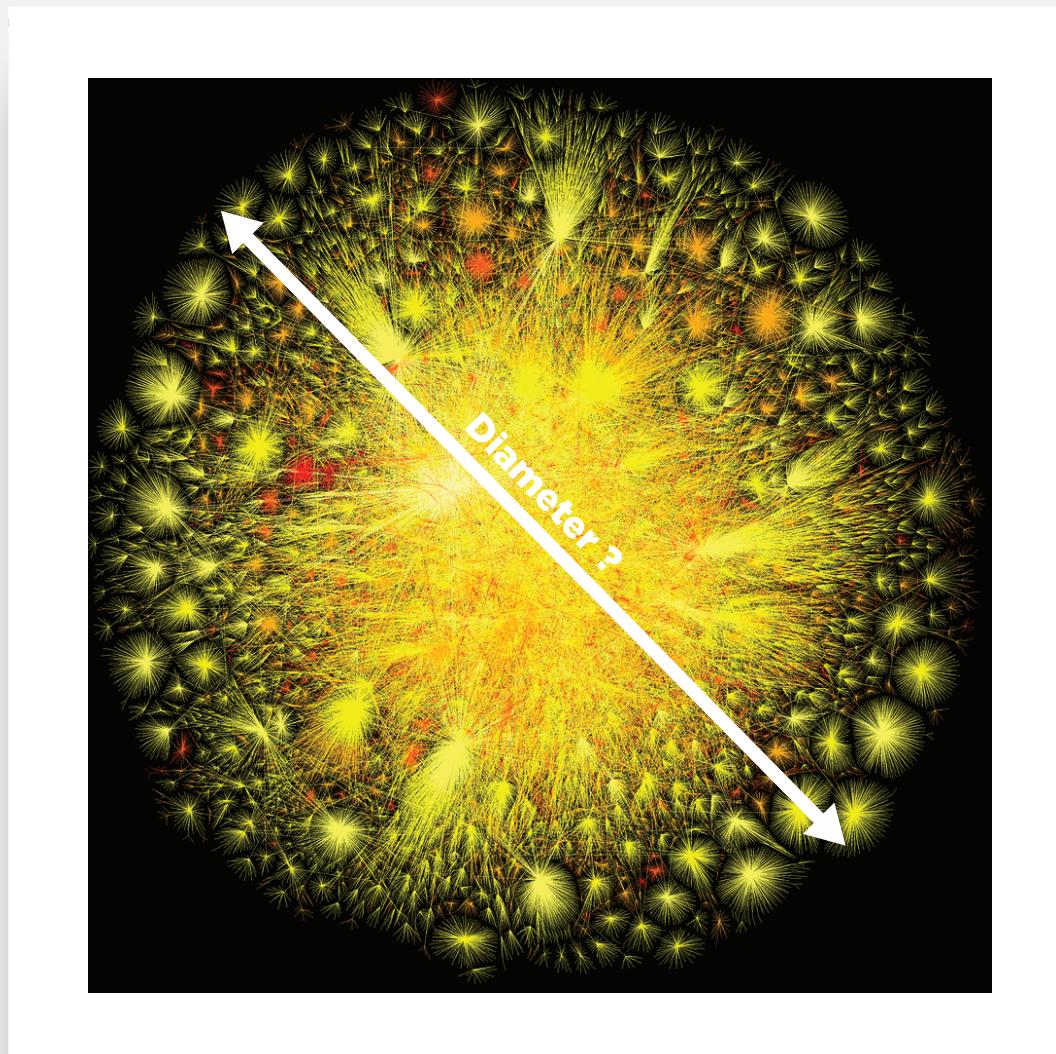
- 1) Once the average degree exceeds $\langle k \rangle = 1$, a giant component should emerge that contains a finite fraction of all nodes. Hence only for $\langle k \rangle > 1$ the nodes organize themselves into a recognizable network.
- 2) For $\langle k \rangle > \ln(N)$ all components are absorbed by the giant component, resulting in a single fully connected network.

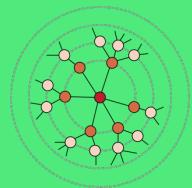
Most real networks are supercritical



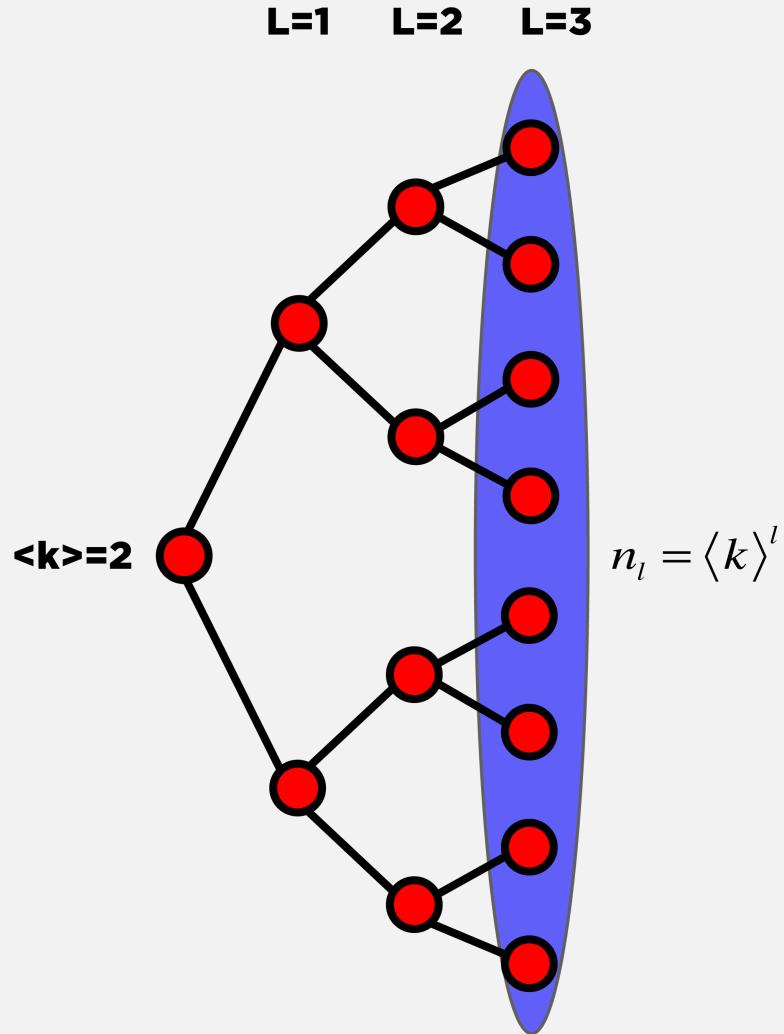


Average shortest path





Average shortest path

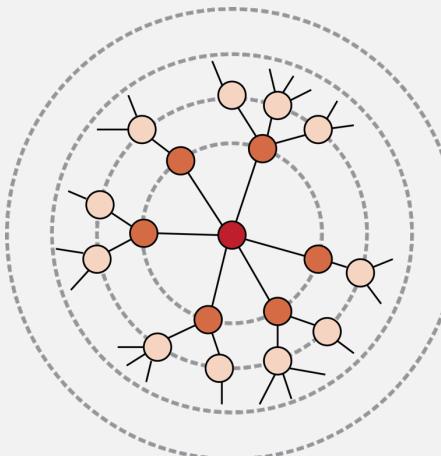


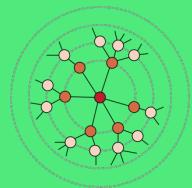
Scaling of the average shortest path:

$$l_{avg} \sim \frac{\log N}{\log \langle k \rangle} \quad \text{Eq. (1)}$$

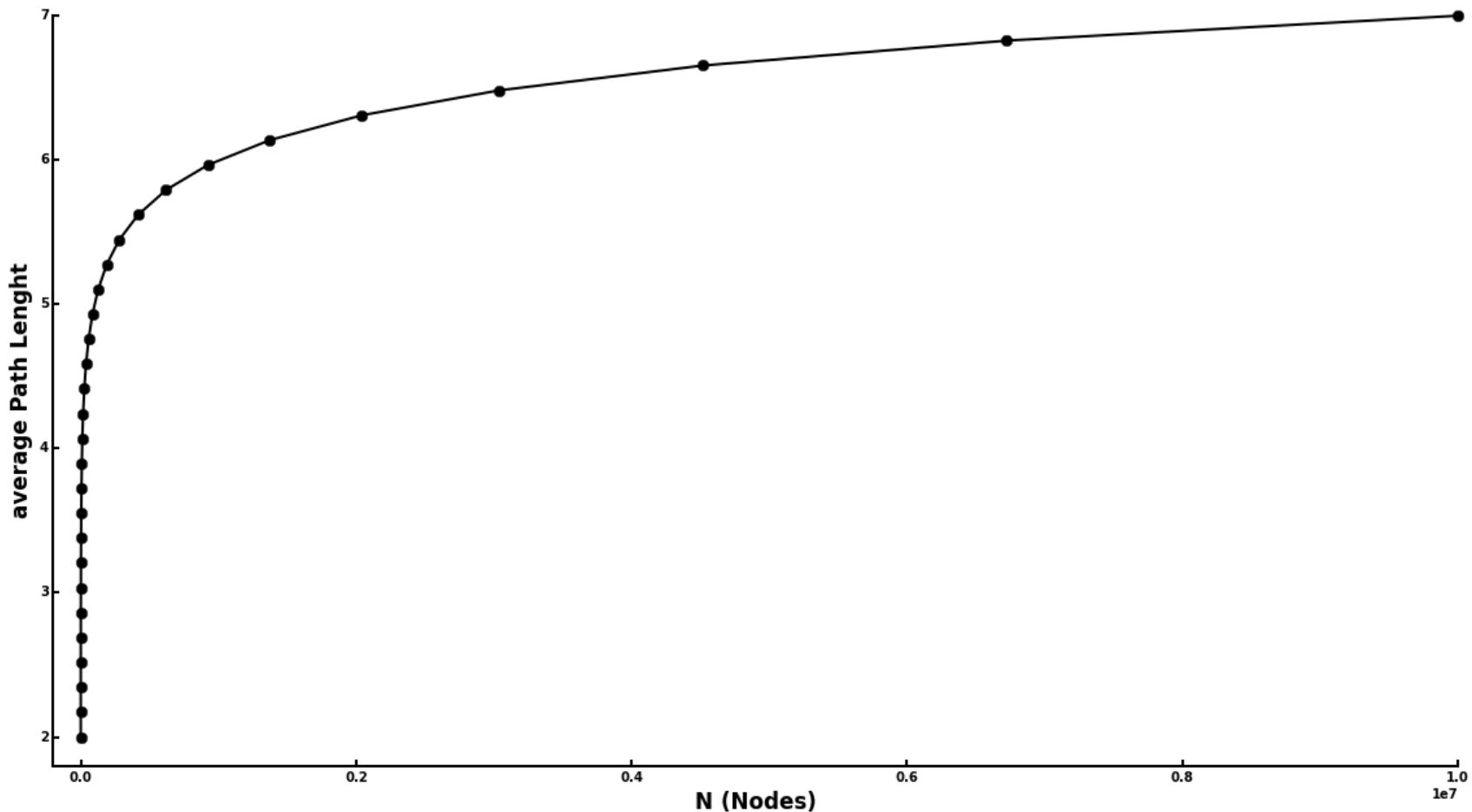
a rough argument:

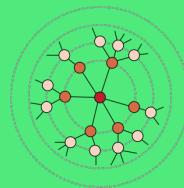
- ▶ ER graph is locally tree-like (no loops; low clustering coefficient)





Average Path Length



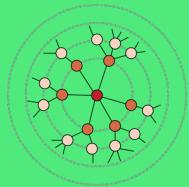


Real Network

Network	N	L	$\langle k \rangle$	$\langle d \rangle$	d_{\max}	$\ln N / \ln \langle k \rangle$
Internet	192,244	609,066	6.34	6.98	26	6.58
WWW	325,729	1,497,134	4.60	11.27	93	8.31
Power Grid	4,941	6,594	2.67	18.99	46	8.66
Mobile Phone Calls	36,595	91,826	2.51	11.72	39	11.42
Email	57,194	103,731	1.81	5.88	18	18.4
Science Collaboration	23,133	93,439	8.08	5.35	15	4.81
Actor Network	702,388	29,397,908	83.71	3.91	14	3.04
Citation Network	449,673	4,707,958	10.43	11.21	42	5.55
E. Coli Metabolism	1,039	5,802	5.58	2.98	8	4.04
Protein Interactions	2,018	2,930	2.90	5.61	14	7.14

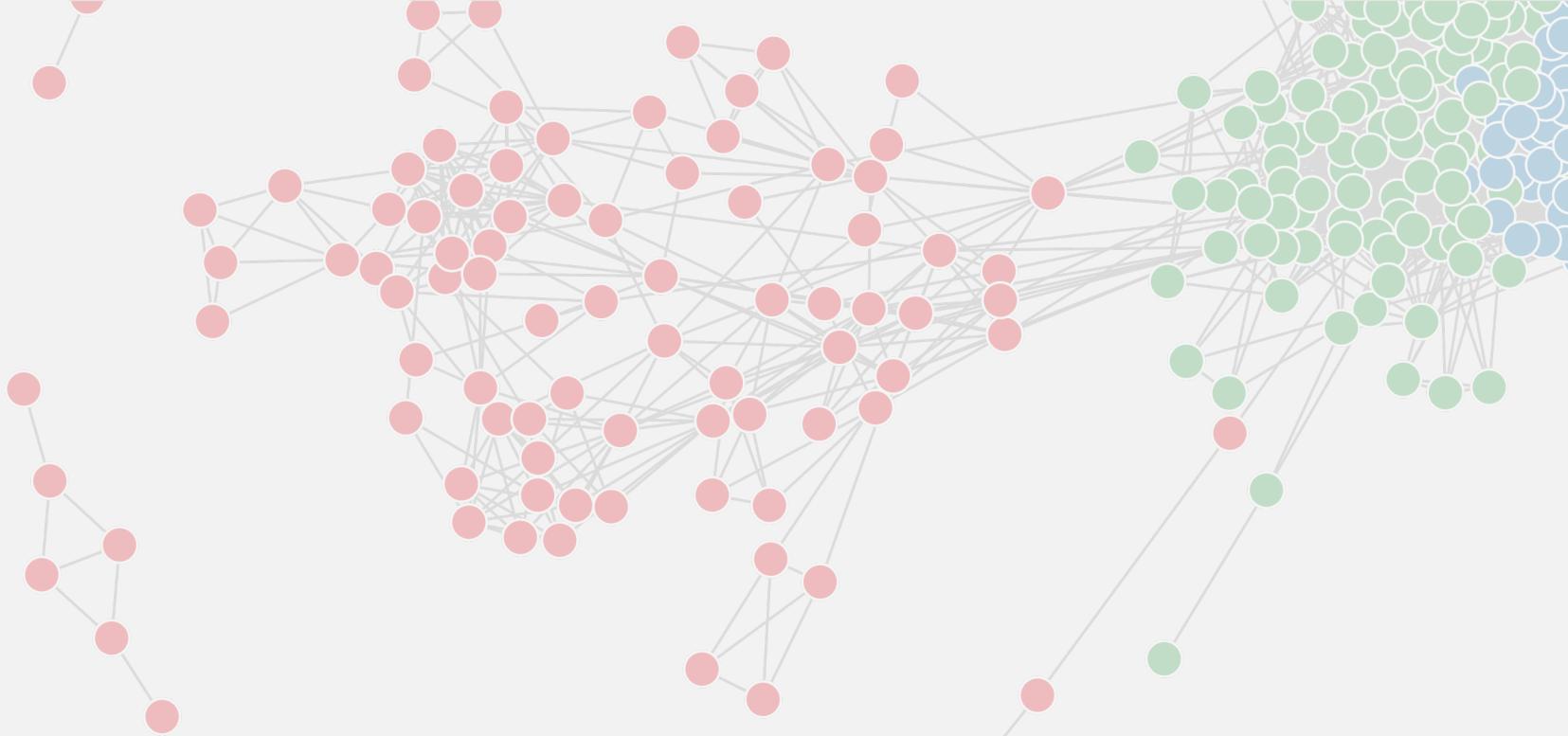
Table

The average distance and the maximum distance d_{\max} for the ten reference networks. The last column provides predicted by (eq.1 slide 33), indicating that it offers a reasonable approximation to the measured . Yet, the agreement is not perfect.



Quiz Q5:

Ignore for the time being the fact that many of your friends' friends are your friends as well. If everyone has 500 friends, the average person would have how many friends of friends?



SMALL WORLD NETWORK (WATTS-STROGATZ MODEL)

Milgram experiment



Instructions:

Given a target individual (stockbroker in Boston), pass the message to a person you correspond with who is “closest” to the target.

Outcome:

20% of initiated chains reached target
average chain length = 6.5

“Six degrees of separation”

Milgram's experiment repeated

Email experiment:

- ▶ 18 targets
- ▶ 13 different countries
- ▶ 60,000+ participants
- ▶ 24,163 message chains
- ▶ 384 reached their targets
- ▶ average path length 4.0

Is 6 is a surprising number?

In the 1960s? Today? Why?

Pool and Kochen in (1978 established
that the average person has between
500 and 1500 acquaintances)



Interpreting Milgram's experiment

► Is 6 is a surprising number?

In the 1960s? Today? Why?

► If social networks were random... ? Pool and Kochen (1978) ~500-1500 acquaintances / person:

500 choices 1st link

$500^2 = 250,000$ potential 2nd degree neighbors

$500^3 = 125,000,000$ potential 3rd degree neighbors

► If networks are completely cliquish?

all my friends' friends are my friends

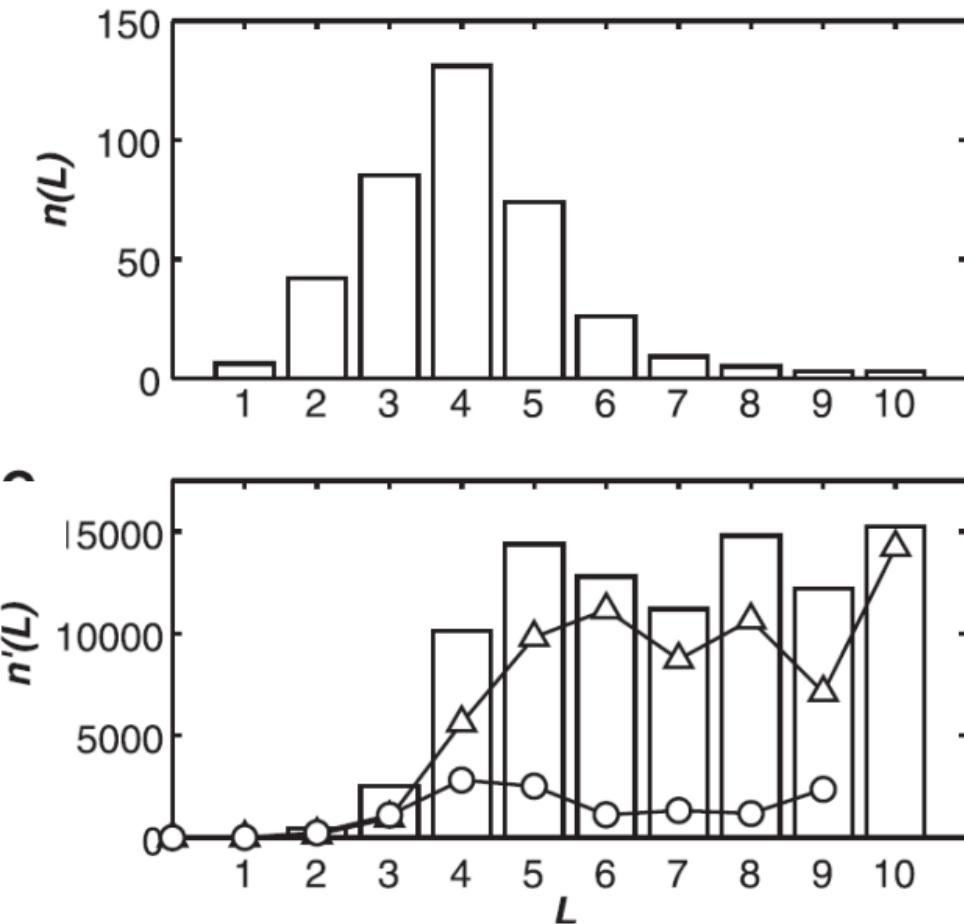
what would happen?

Quiz:

If the network were completely cliquish, that is all of your friends of friends were also directly your friends, what would be true:

- a) I am in an isolated group
- b) Your shortest path length to your friend's friends would be 2
- c) None of your friendship edges would be a part of a triangle (closed triad)

Estimating the true distance



observed chain lengths

recovered histogram of path length

 **inter-country**
 **intra-country**

Navigability and search strategy: Small world experiment @ Columbia

Successful chains disproportionately used

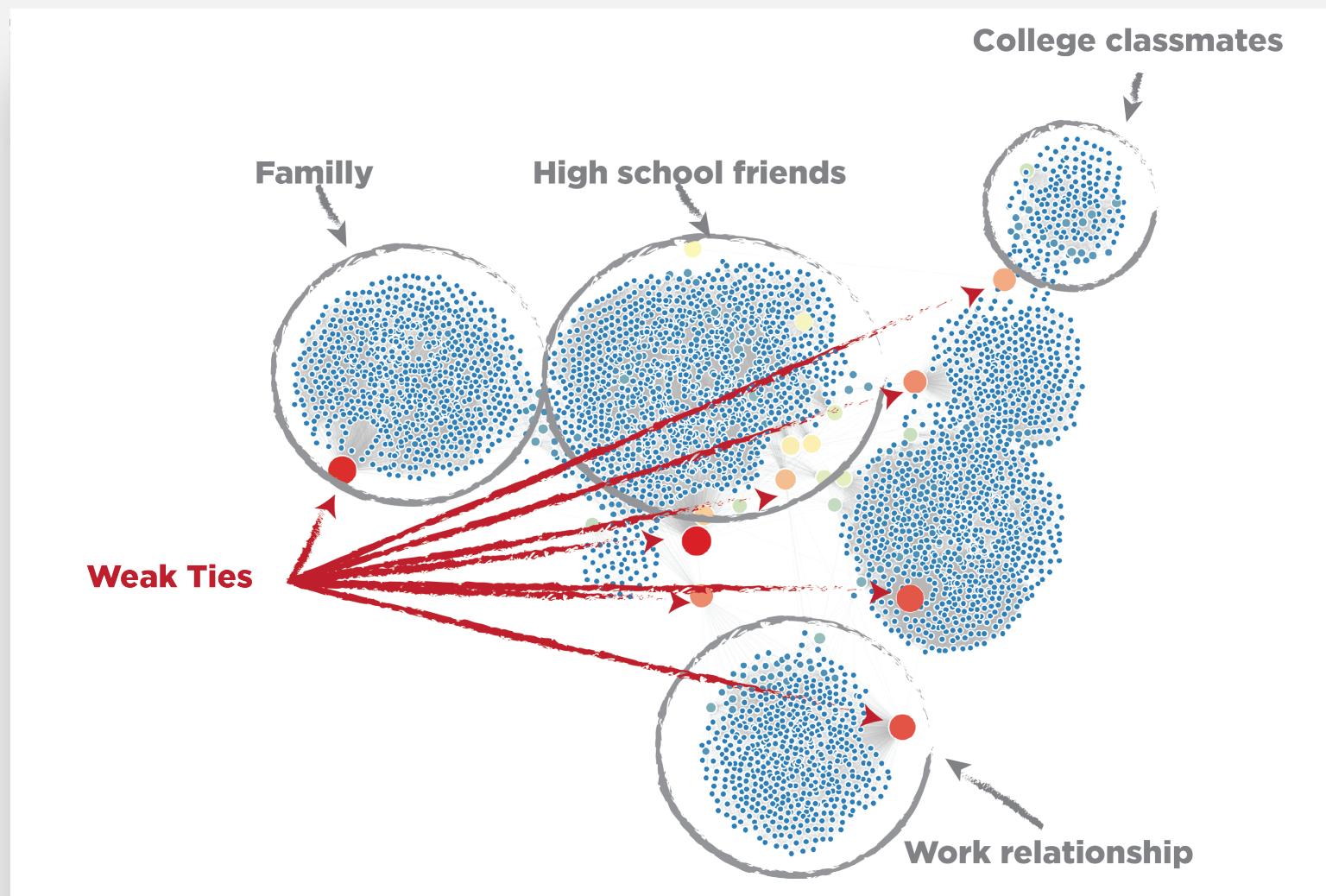
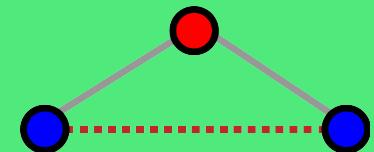
- ▶ weak ties (Granovetter)
- ▶ professional ties (34% vs. 13%)
- ▶ ties originating at work/college
- ▶ target's work (65% vs. 40%)

... and disproportionately avoided

- ▶ hubs (8% vs. 1%) (+ no evidence of funnels)
- ▶ family/friendship ties (60% vs. 83%)

Strategy: Geography → Work

Strength of weak ties



Remember the Clustering

How many of my friends are also firends ?

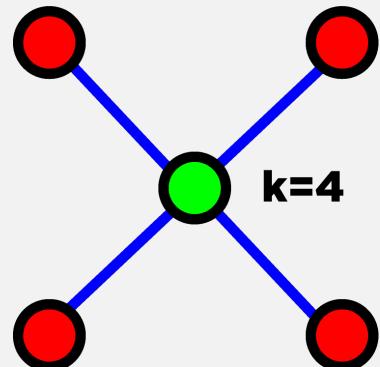
Local Clustering Coeficient

$$C(i) = \frac{\# \text{ of links between neighbors}}{\frac{k(k-1)}{2}}$$

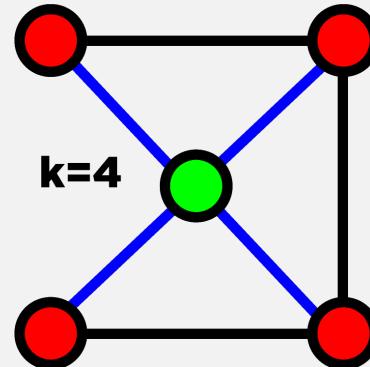
Average Clustering Coeficient

$$\bar{C} = \frac{1}{N} \sum_{i=1}^N C(i)$$

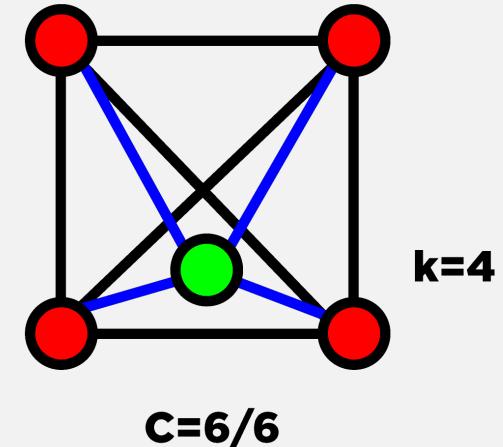
C=0



C=0.5



C=1



Strong tie and embeded structure

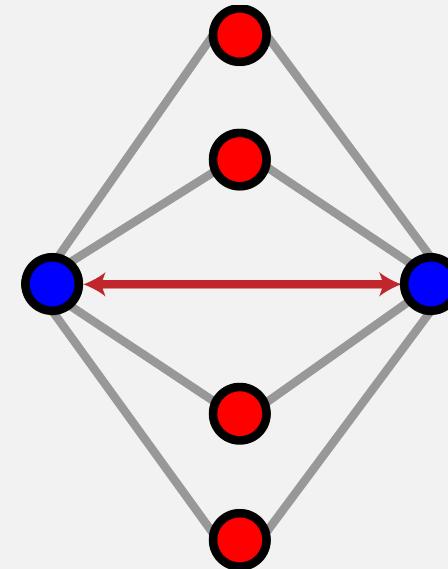
embeddeness: number of common neighbors the two endpoints have

► A strong tie:

frequent contact

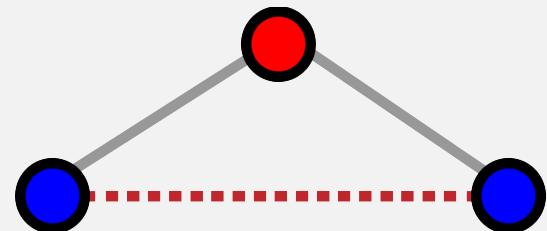
affinity

many mutual contacts

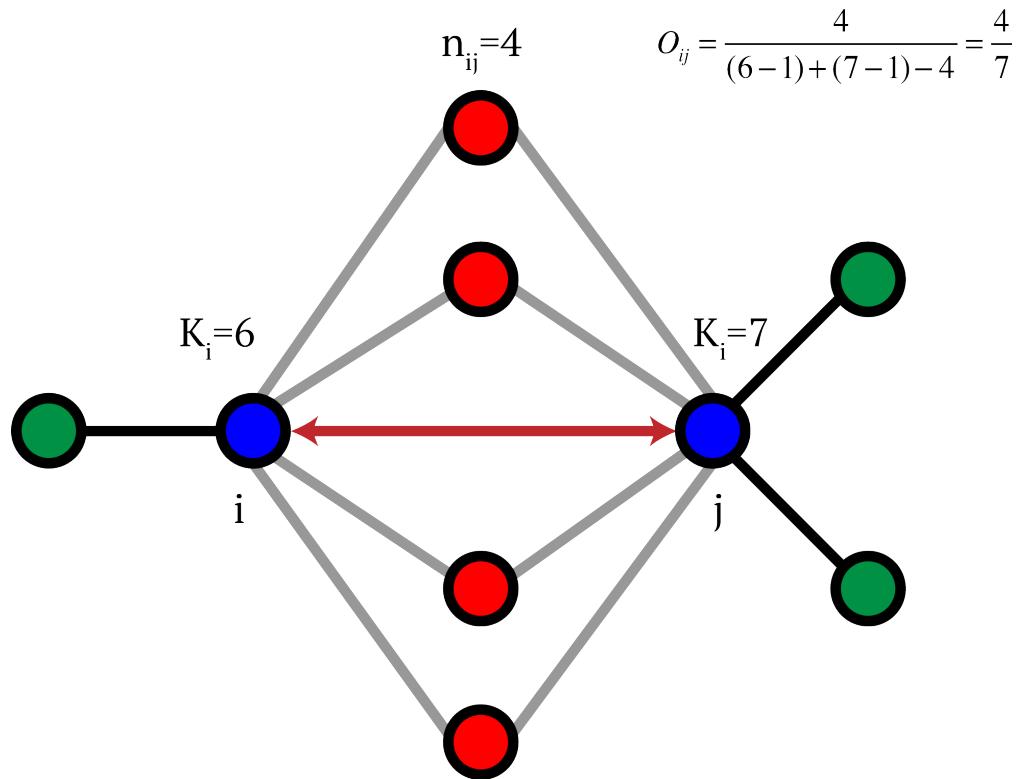


► Forbidden triad:

strong ties are likely to “close”



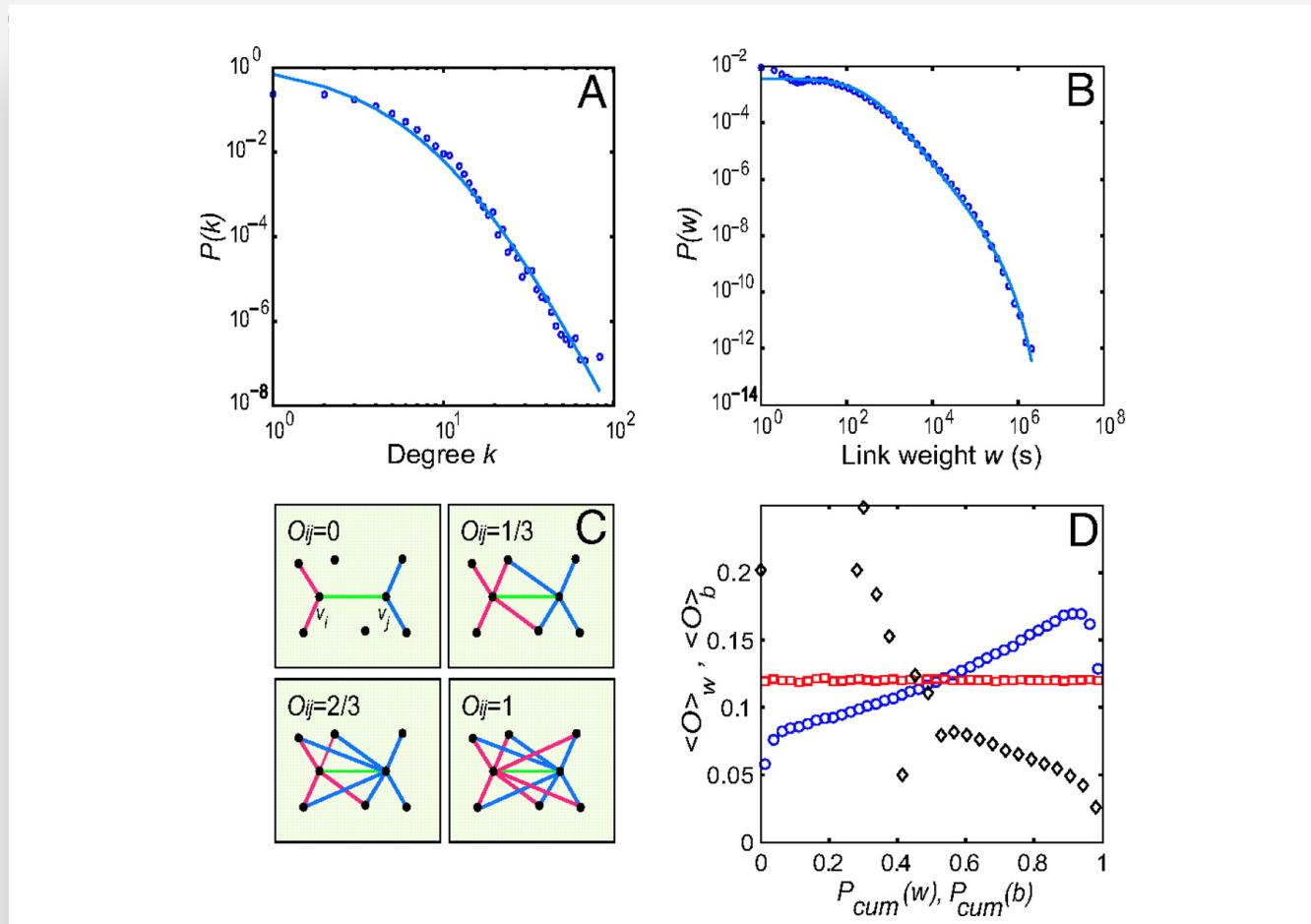
Embeddeness



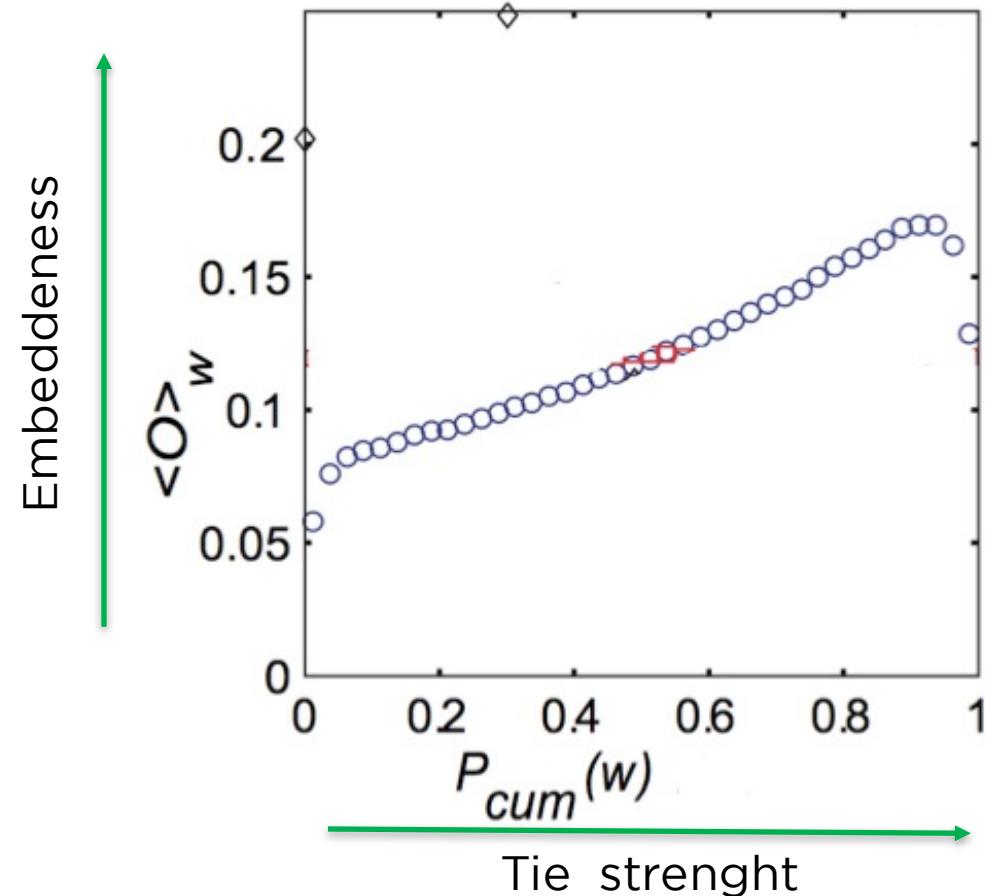
How many common neighbors
a pair of node has

$$O_{ij} = \frac{n_{ij}}{(k_i - 1) + (k_j - 1) - n_{ij}}$$

Characterizing the large-scale structure and the tie strengths of the mobile call graph



Edge neighborhood overlap as a function of tie strength

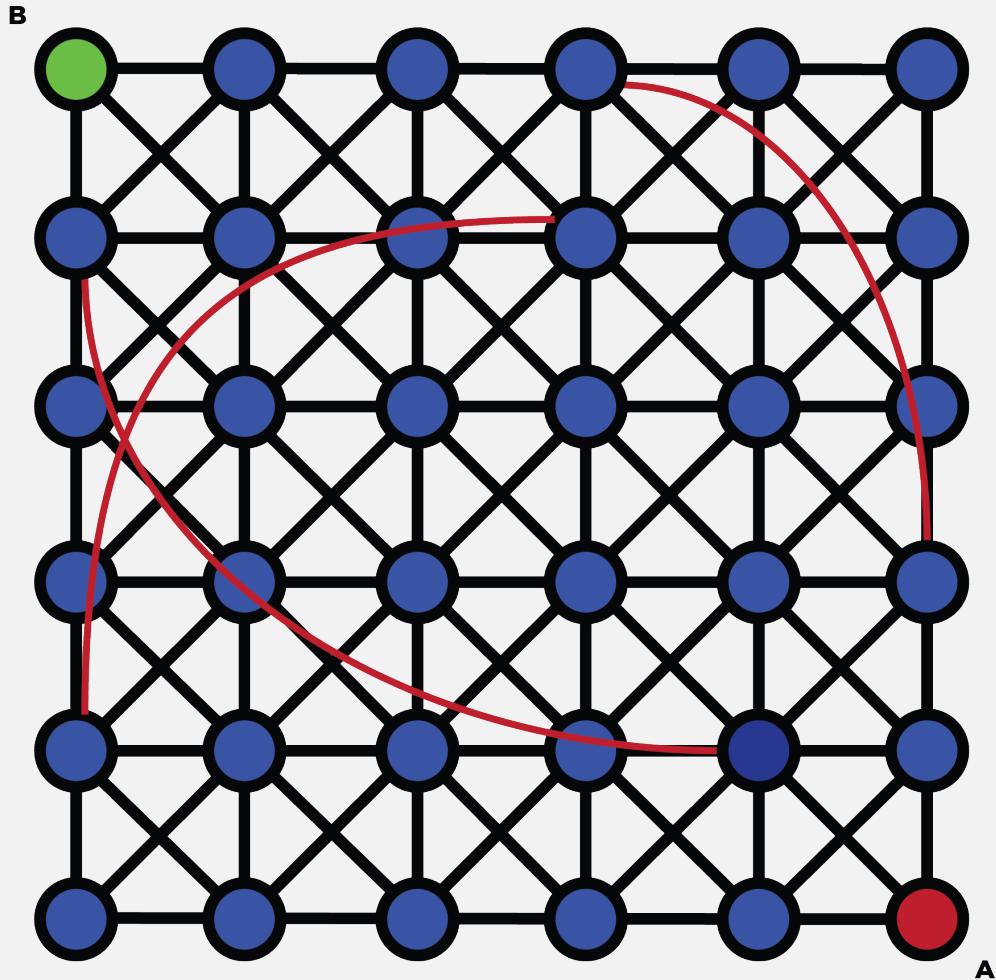


Randomness + regularity

- ▶ Need “clustering” (your friends are likely to know each other)
- ▶ **Regularities bring clustering**
- ▶ Connected random networks have **short average path lengths**

$$d_{ab} = 6$$

$$d_{ab} = 3$$



With the **red edge** the overall $\langle d \rangle$ decrease

Watts-Strogatz model

Generating small world graphs

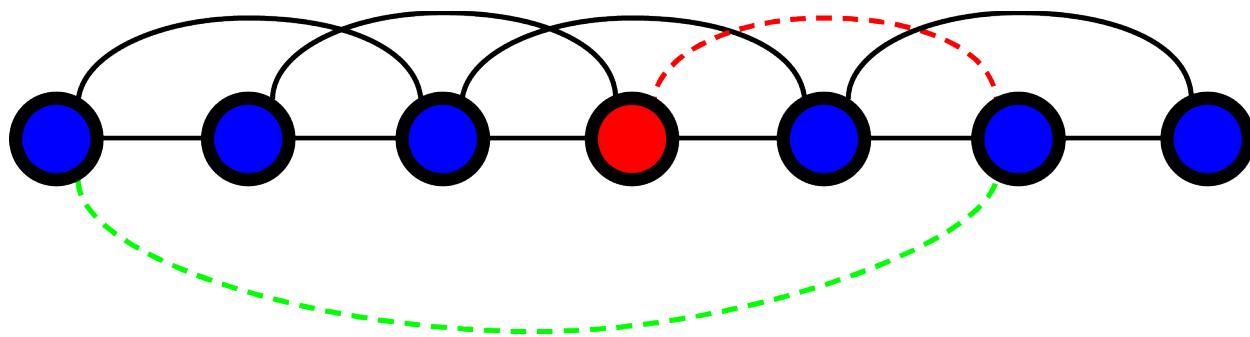
- ▶ Each node has $K \geq 4$ nearest neighbors
- ▶ tunable: vary the probability p of rewiring any given edge
- ▶ small p : regular lattice
- ▶ large p : classical random graph

Collective dynamics of “small-world” networks

Duncan J. Watts & Steven H. Strogatz

Nature

1998



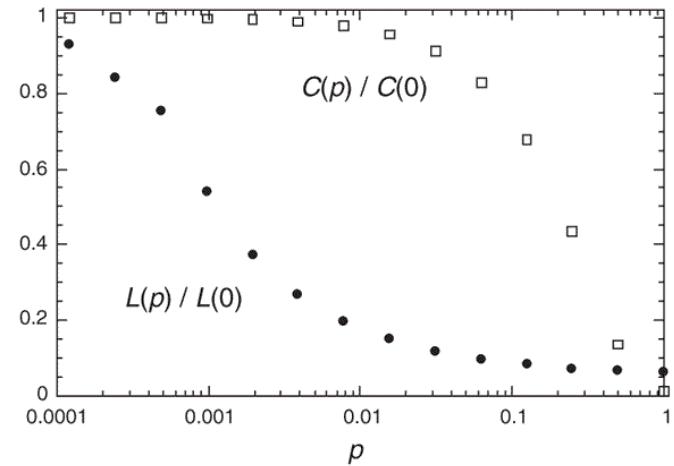
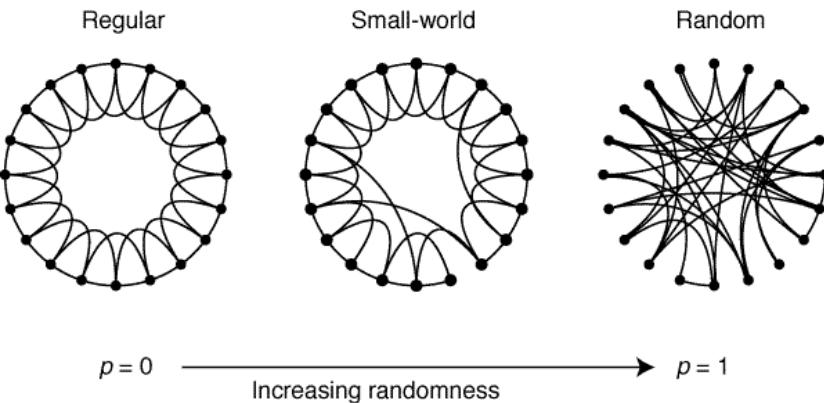
It is a small world after all !

Network	L_{actual}	L_{random}	C_{actual}	C_{random}
Film actors	3.65	2.99	0.79	0.00027
Power grid	18.7	12.40	0.08	0.005
C. elegans	2.65	2.25	0.28	0.05

Small-world networks were found everywhere:

- neural network of *C. elegans*,
- semantic networks of languages,
- actor collaboration graph,
- food webs,
- social networks of comic book characters.

Bret Victor's Scientific Communication As Sequential Art



Collective dynamics of 'small-world' networks
Duncan J. Watts and Steven H. Strogatz
Nature 393, 440-442, 1998.

Small World (Watts-Strogatz model)

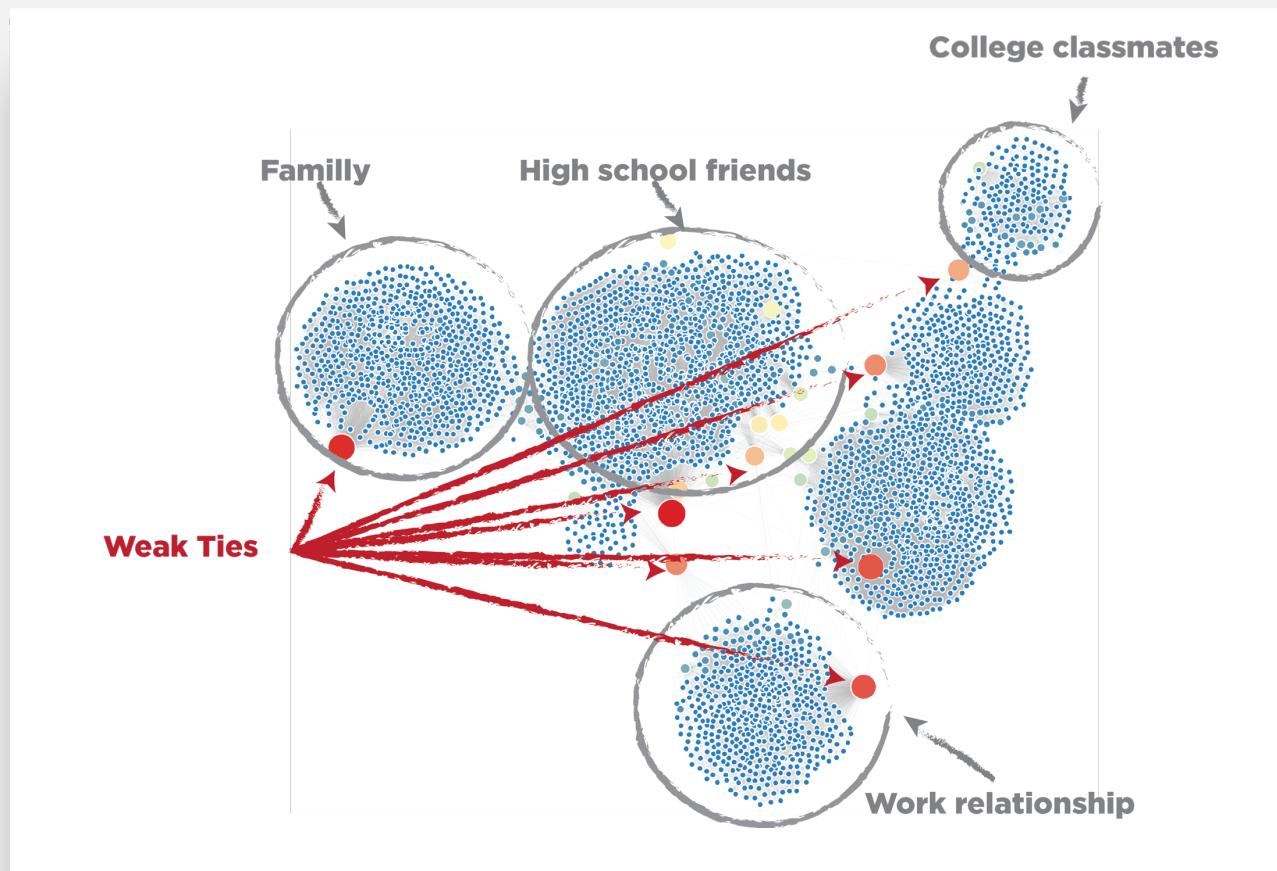
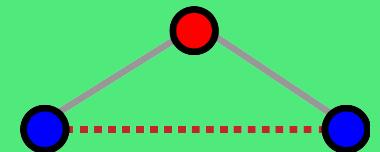
► **Small World Property**

In real networks the average distance between two nodes depends logarithmically on N , rather than following a polynomial expected for regular lattices

► **High Clustering**

The average clustering coefficient of real networks is much higher than expected for a random network of similar number of node and diameter (or average path length)

Social network



- ▶ Characteristics of social Network
 - ▶ High Clustering
 - ▶ Few Bridge Node

Ties and geography

“ The geographic movement of the [message] from Wyoming to Massachusetts is striking. There is a progressive closing in on the target area as each new person is added to the chain ”

S. Milgram “The small world problem”

Psychology Today 1, 61, 1967

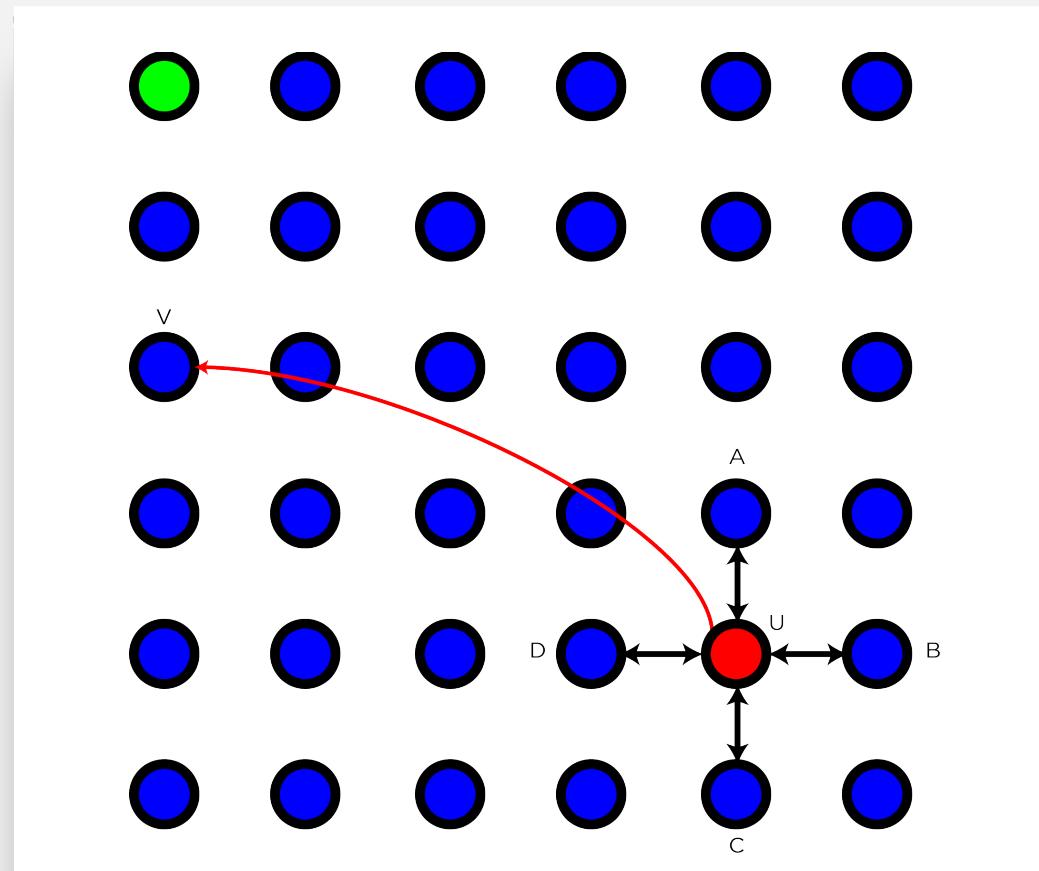
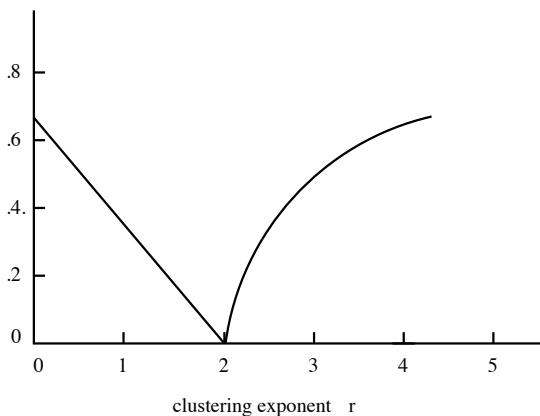


Kleinberg's geographical small world model

nodes are placed on a lattice and connect to nearest neighbors

additional links placed with:

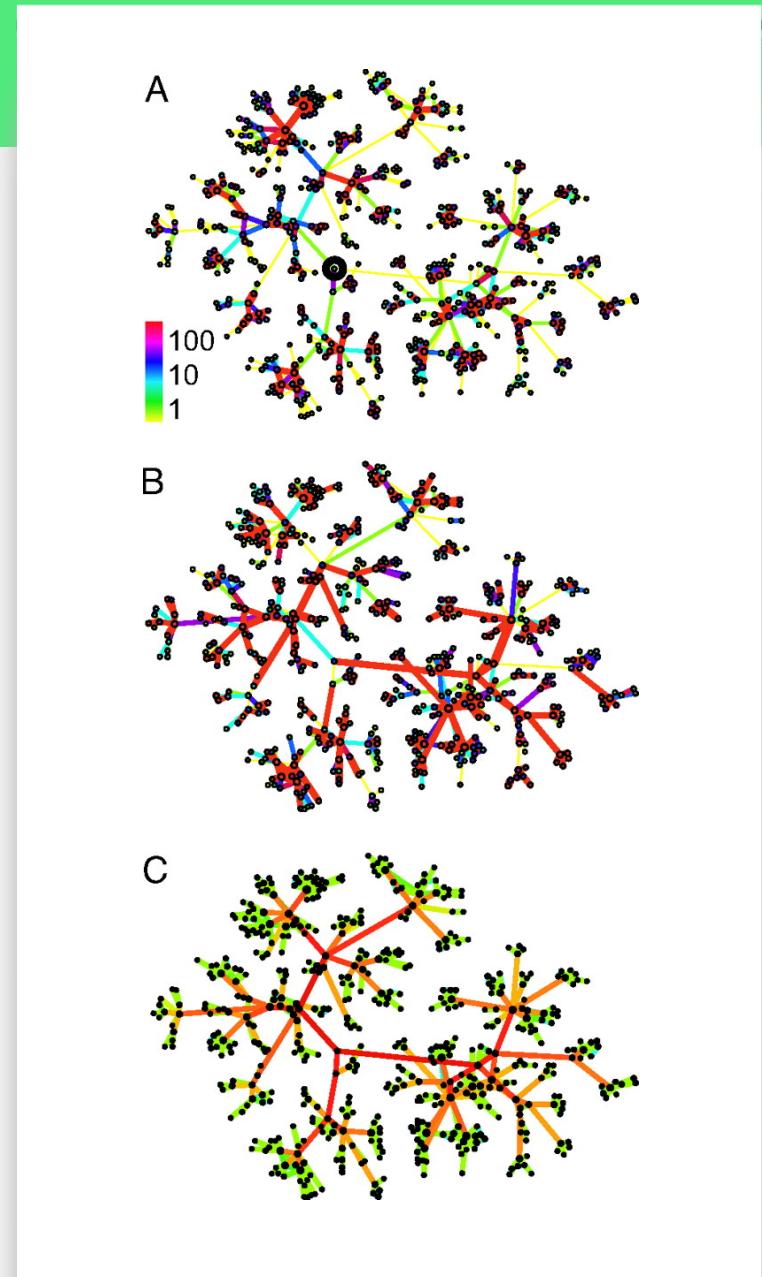
$$P_{link}(u,v) = d(u,v)^r$$



Conclusion

- ▶ clustering and embeddeness play a important role especially in social network
- ▶ there is an interplay between network diameter and the clustering coefficient
- ▶ in social network node tend to form group (or community)
- ▶ strong tie and weak tie both play an important role

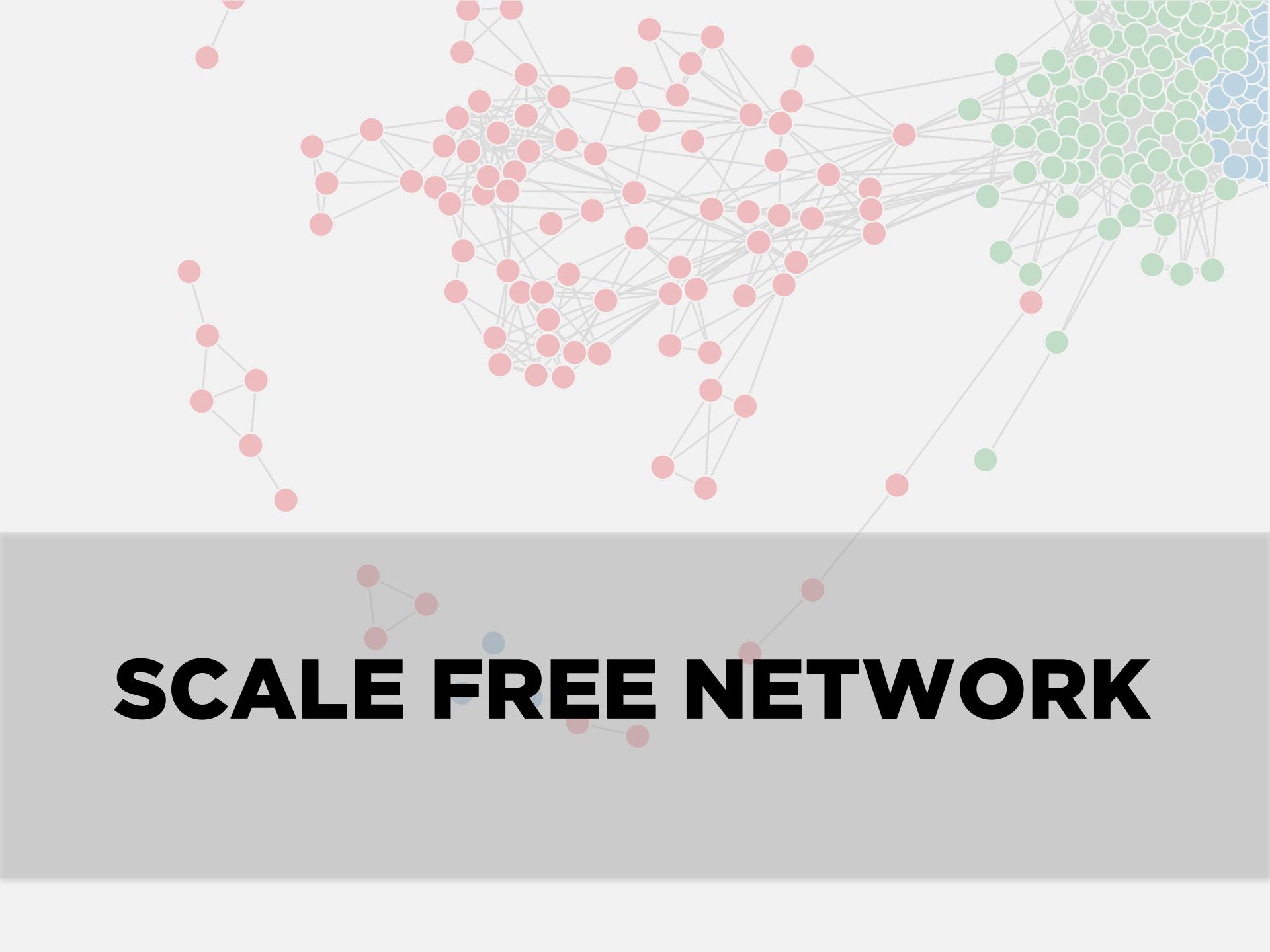
We need a better understanding of the link between clustering coefficient and the network diameter



What we learned

1. Most of graph in nature are **supercritical**
2. Once the **average degree** exceeds $\langle k \rangle = 1$, a giant component should emerge that contains a finite fraction of all nodes
3. **Clustering** shape the way network a designed
4. In real networks the **average distance** between two nodes depends logarithmically on N

SCALE FREE NETWORK



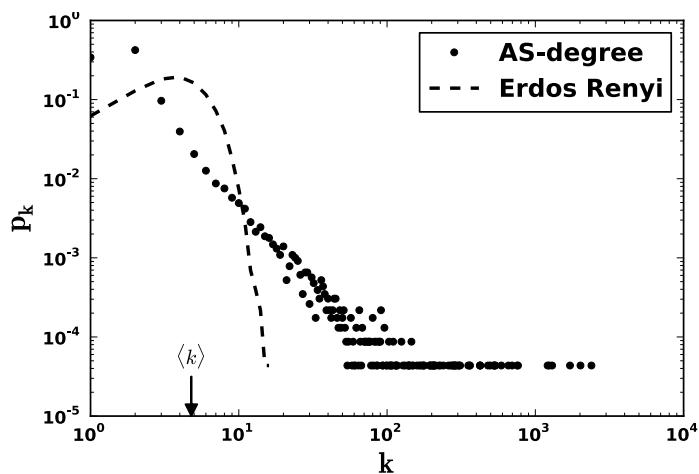
How real networks behave ?

Emergence of scaling in random networks

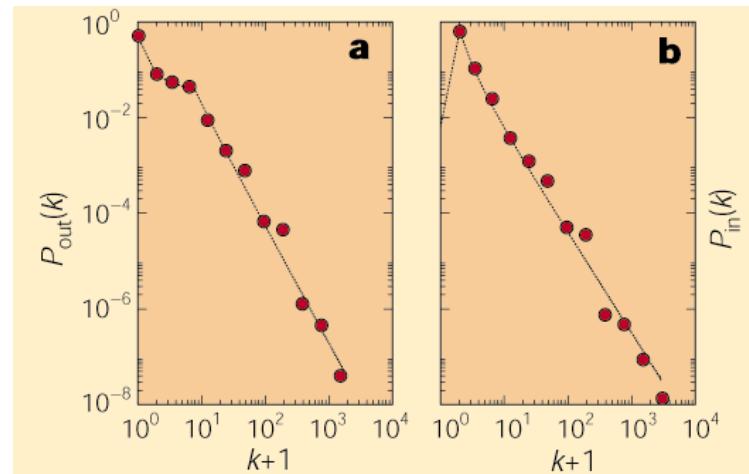
1999

AL Barabási, R Albert

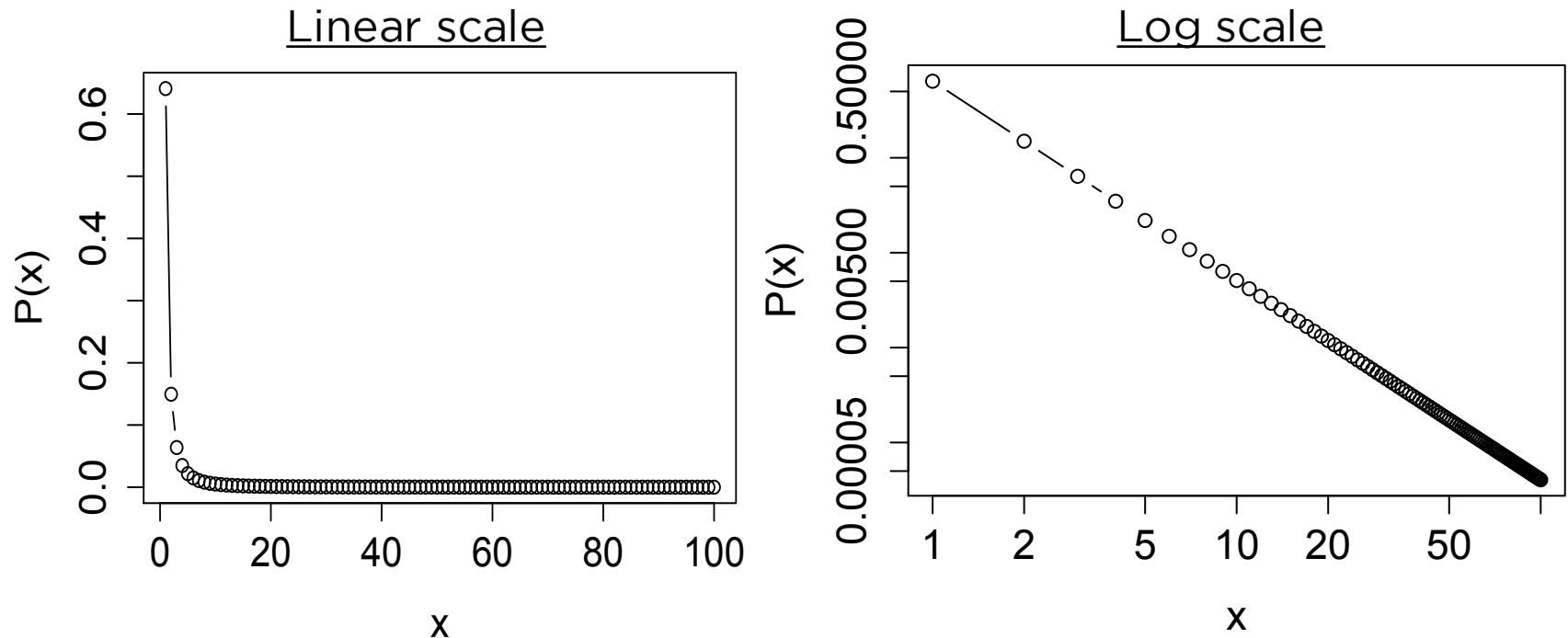
Internet AS



WWW



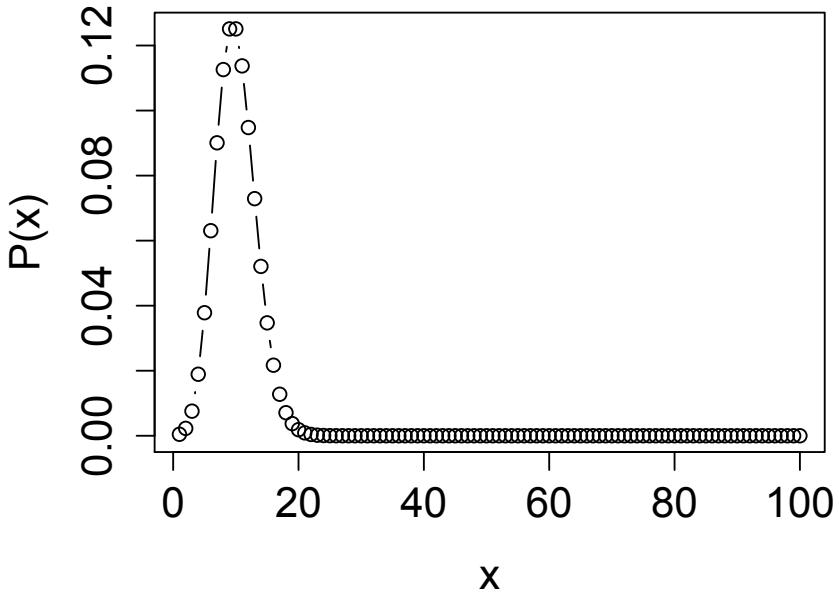
Power-law distribution



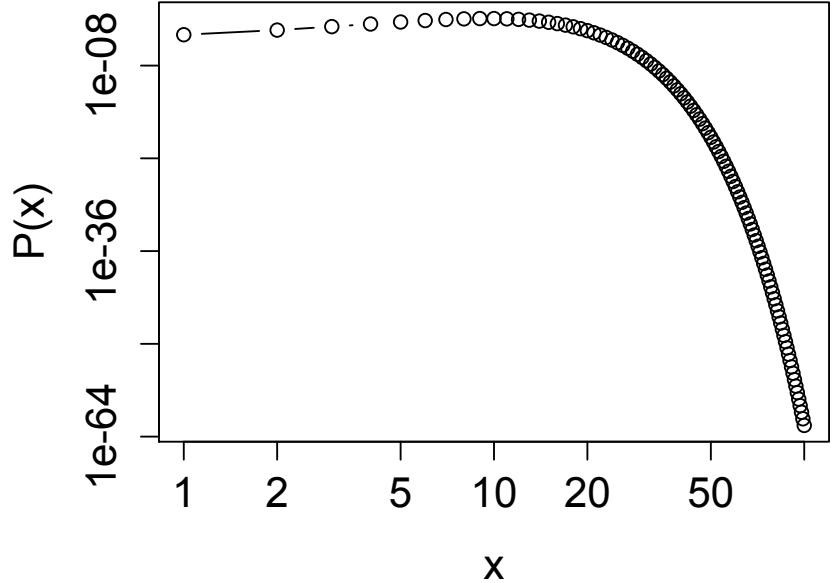
- ▶ high skew (asymmetry)
- ▶ straight line on a log-log plot

Poisson distribution

Linear scale



Log scale



- ▶ little skew (asymmetry)
- ▶ curved on a log-log plot

Power law distribution

Straight line on a log-log plot

$$\ln(p(k)) = c - \alpha \ln(k)$$

Exponentiate both sides to get that $p(k)$, the probability of observing a node of degree k is given by:

$$p(k) = C k^{-\alpha}$$

power law exponent

normalization

constant (probabilities over
all k must sum to 1)

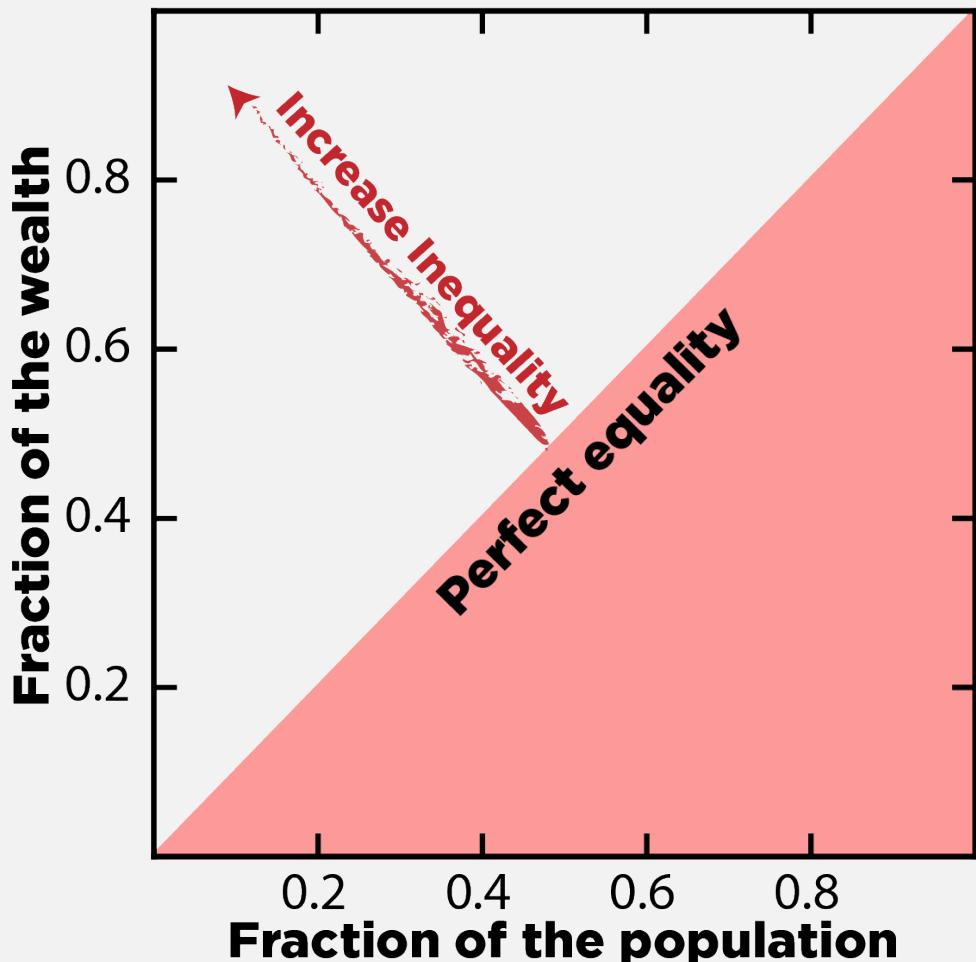
Degree distribution

power law distribution what does it mean ?

Degree “wealth”

What fraction of total wealth W is owned by richest fraction P

Lorentz curve



Degree distribution

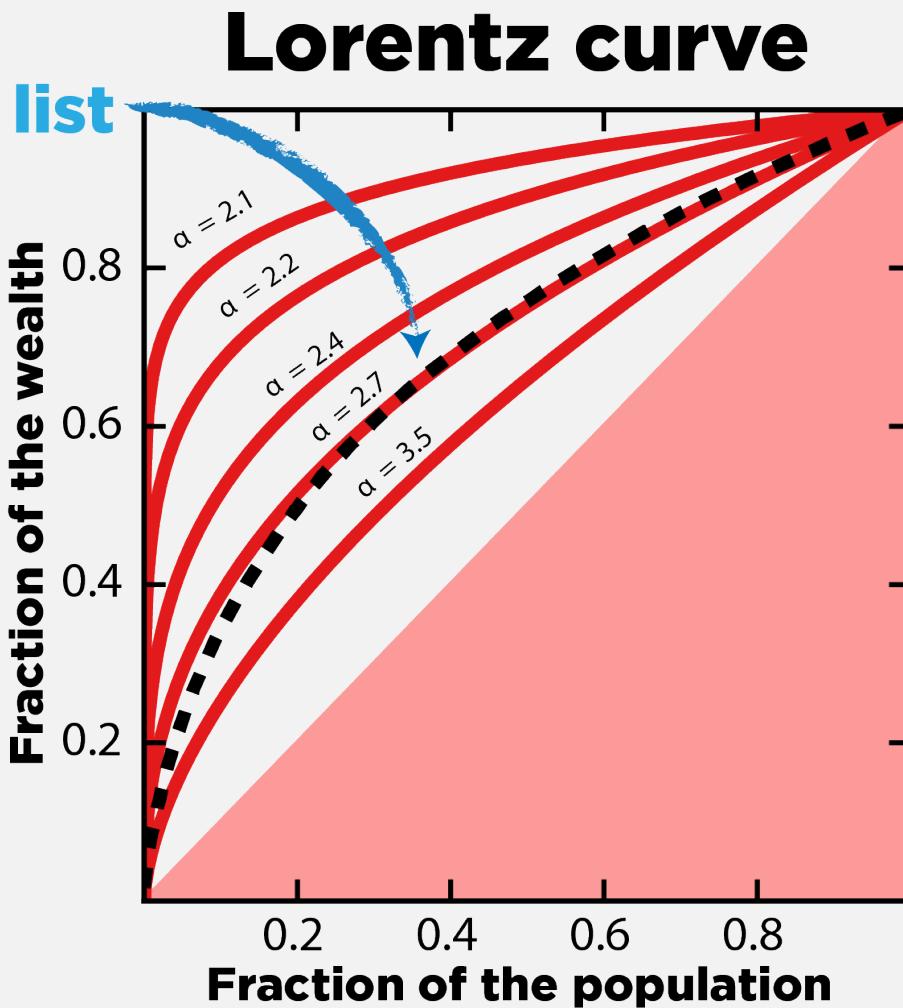
power law probability distribution what does it mean ?

Degree
“wealth” **Forbes 500 list**

What fraction of total wealth W is owned by richest fraction P

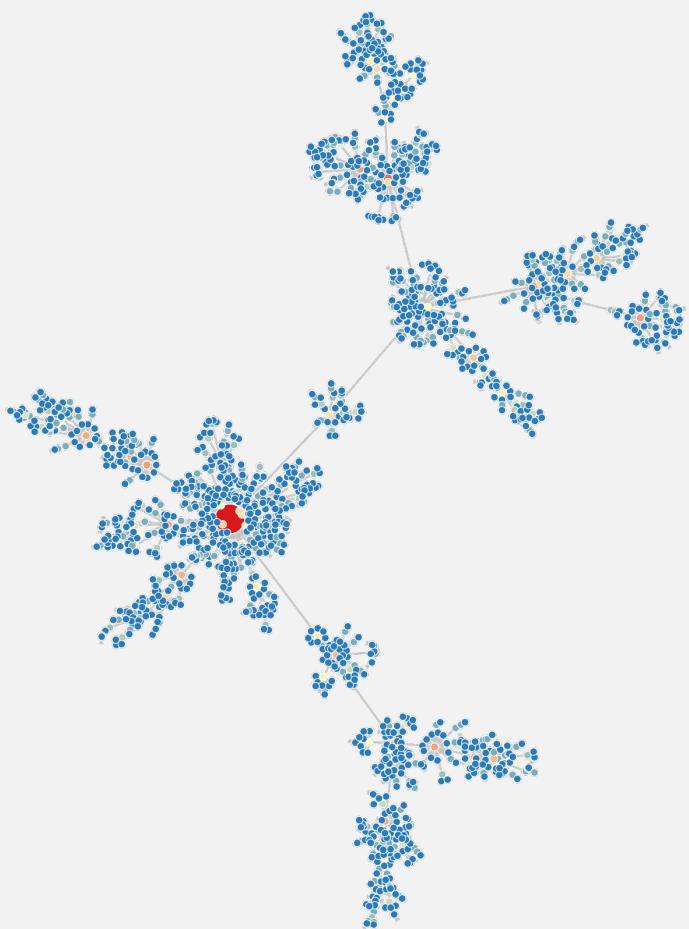
$$\Pr(k) \propto k^{-\alpha}$$

power-law distribution
80/20 rule

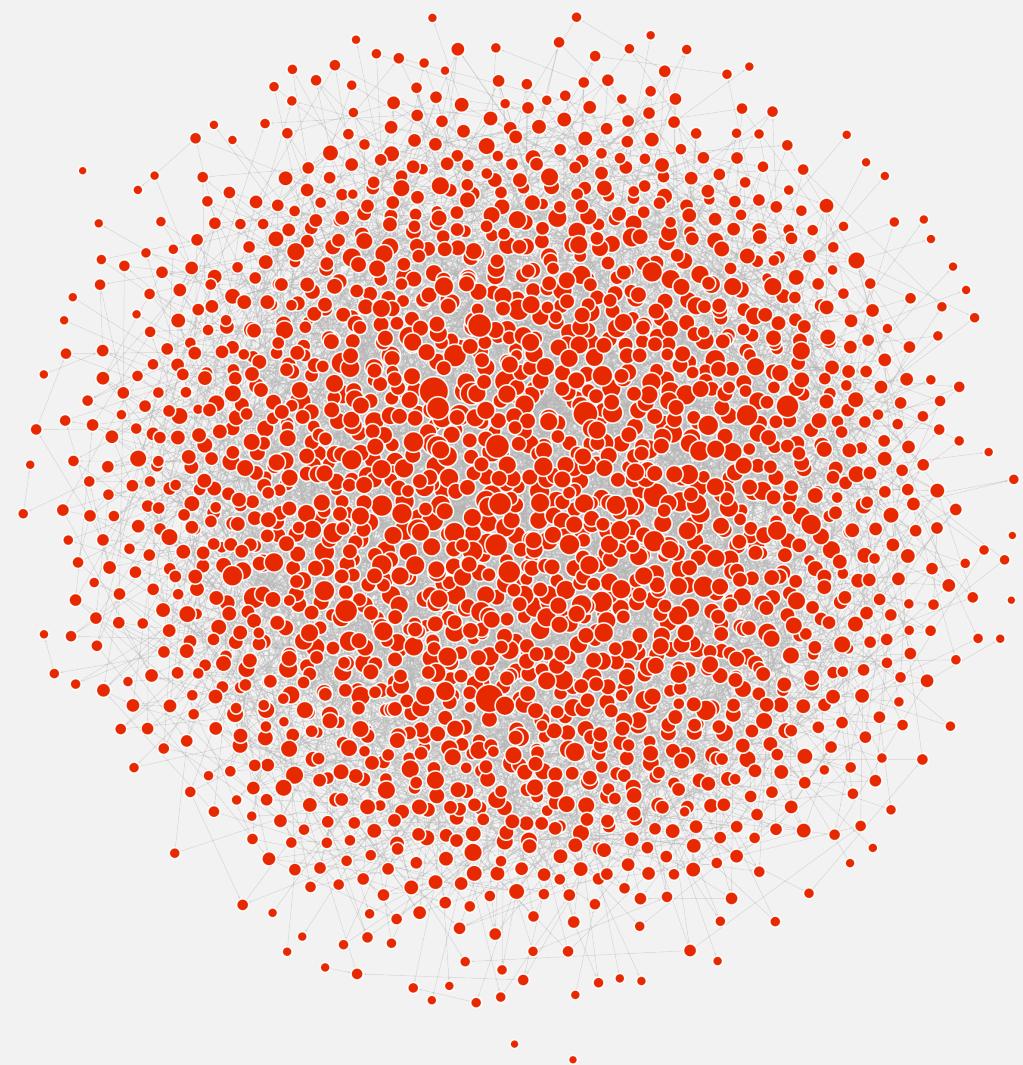


Scale free compare to Erdos-Reny

Scale free network



Random network



Quiz:

As the exponent α increases, the downward slope of the line on a log-log plot:

- a) stays the same
- b) becomes milder
- c) becomes steeper

Quiz:

As the exponent α^a decreases, the difference between lowest degree node and the highest degree node :

- a) stays the same
- b) becomes milder
- c) becomes steeper

a: for $\alpha > 1$

2nd ingredient: preferential attachment

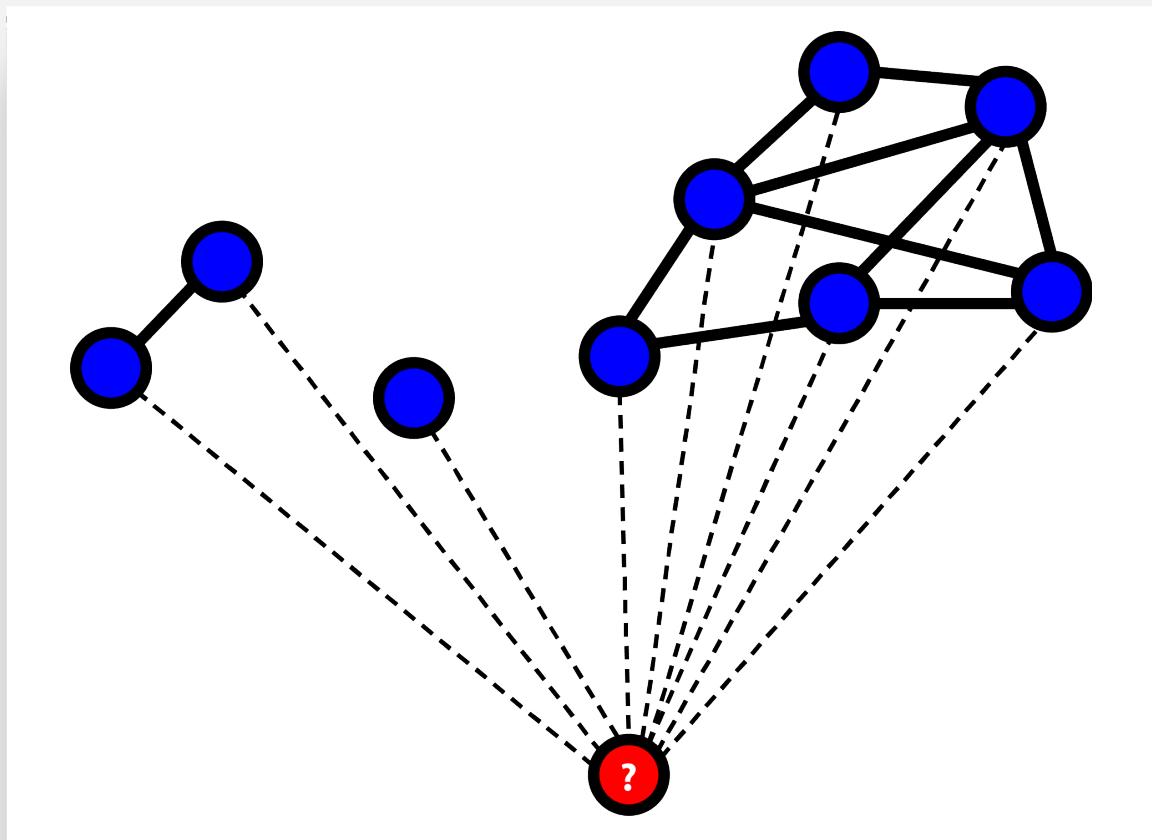
Preferential attachment:

- ▶ new nodes prefer to attach to well-connected nodes over less-well connected nodes

Process also known as

- ▶ cumulative advantage
- ▶ rich-get-richer
- ▶ Matthew effect

Generating power-law networks



Nodes prefer to attach to nodes with many connections
(preferential attachment, cumulative advantage)

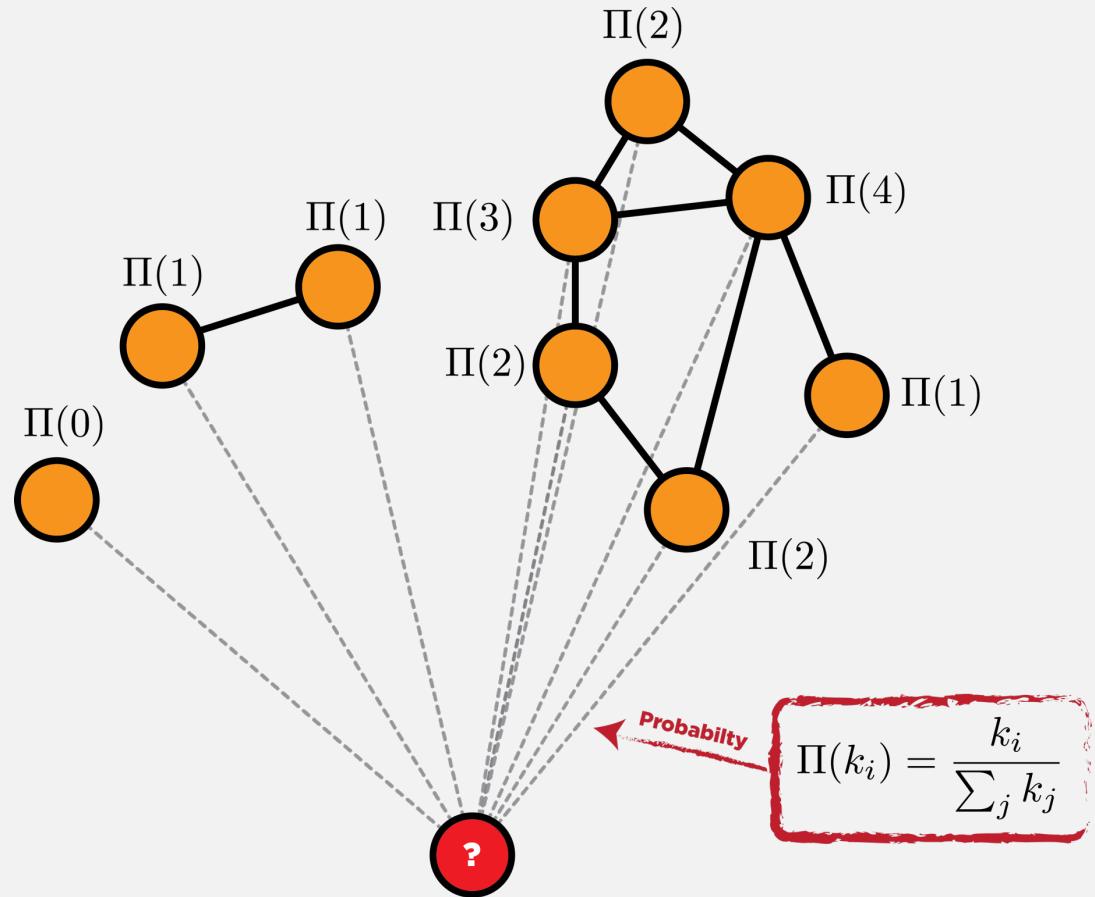
Barabasi-Albert model

First used to describe skewed degree distribution of the World Wide Web.

Growth + Preferential attachment

Each node connects to other nodes with probability proportional to their degree

- ▶ Step 1: Start with m_0 disconnected nodes
- ▶ Step 2:
 1. Growth a node appears at each time step
 2. Each node make m connections to nodes already present
 3. Preferential attachment probability of connecting the i th node is $\propto k_i$
- ▶ Results in power-law with exponent $\alpha = 3$



Quiz Q:

How could one make the growth model more realistic for social networks?

- a) old nodes die
- b) some nodes are more sociable
- c) friendships vane over time
- d) all of the above

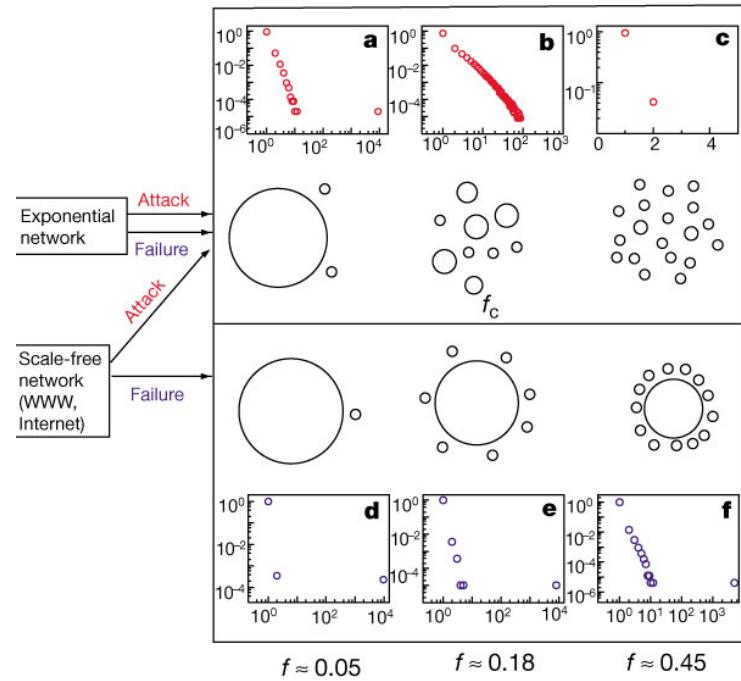
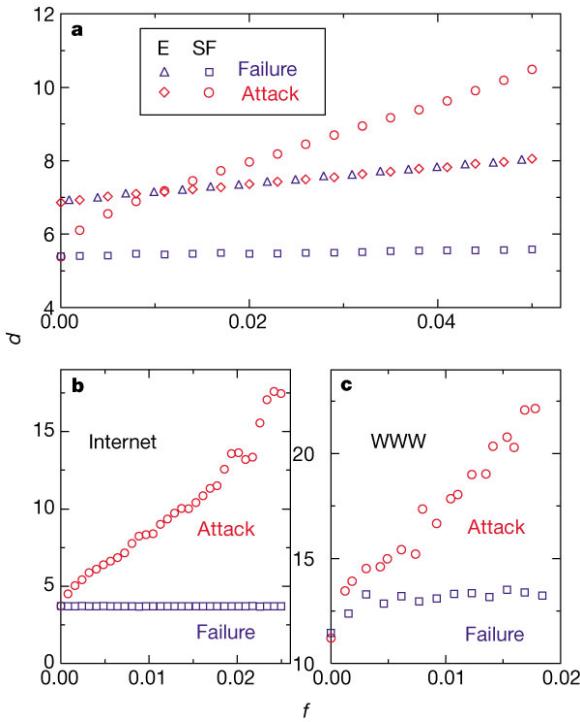
Quiz Q:

The degree distribution for a growth model where new nodes attach to old nodes at random will be

- a) a curved line on a log-log plot
- b) a straight line on a log-log plot

Robust yet fragile

- ▶ Plots of network diameter as a function of fraction of nodes removed
- ▶ Erdős-Rényi versus scale-free networks
- ▶ blue symbols = random removal
▶ red symbols = targeted removal (most connected first)



Robust yet fragile

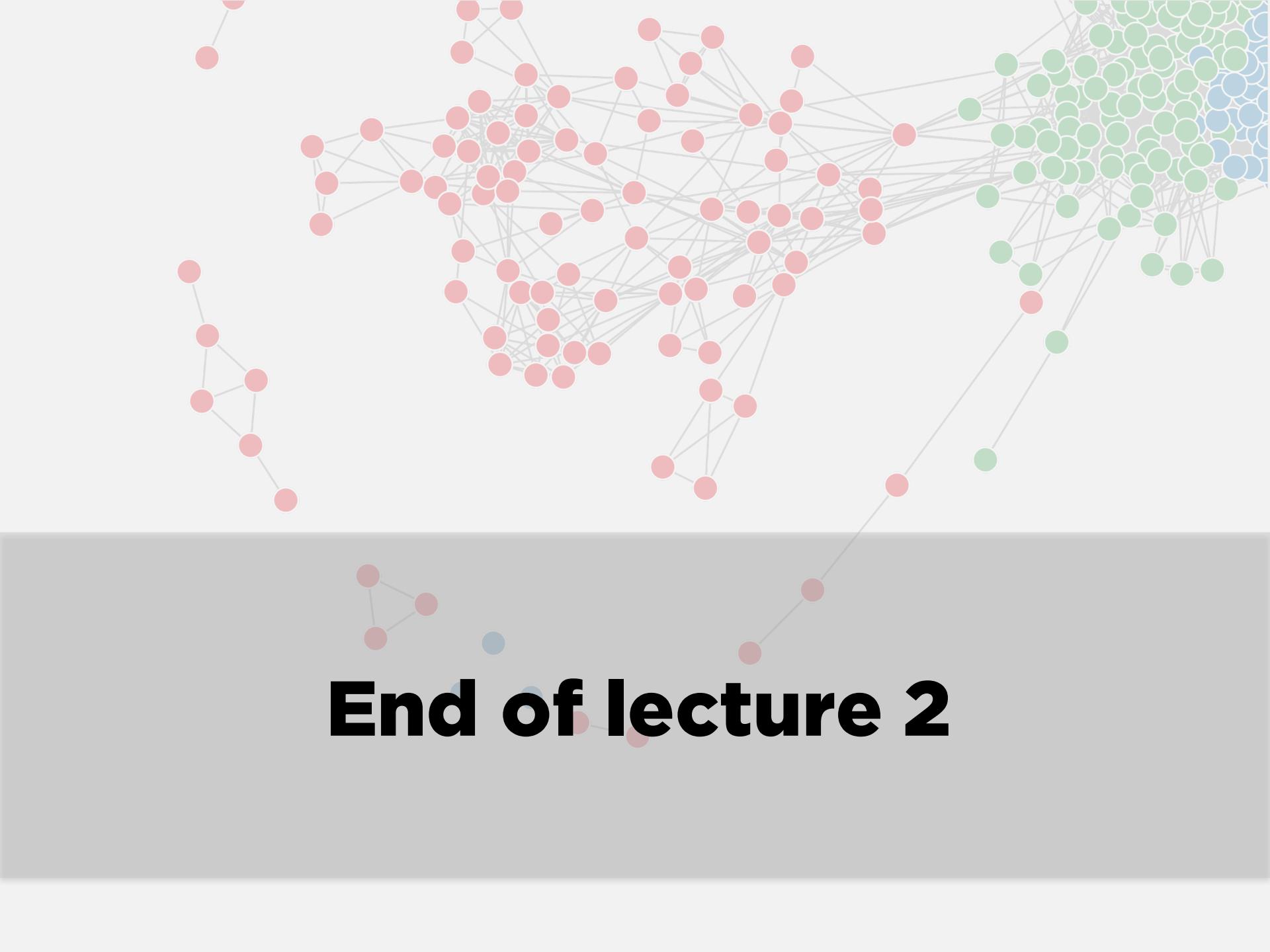
Scale-free networks are thus robust to random failures yet fragile to targeted ones

- ▶ All very reasonable: Hubs are a big deal.
- ▶ Representing all webpages as the same size node is obviously a stretch (e.g., google vs. a random person's webpage)
- ▶ Most connected nodes are either:
 1. Physically larger nodes that may be harder to 'target'
 2. or subnetworks of smaller, normal-sized nodes.
- ▶ Need to explore cost of various targeting schemes.

Quiz:

Which network is more diffusive:

- a) Scale Free Network
- b) Erdos-Reny Network



End of lecture 2

CREDITS

- 🎓 Complex Network Peter Dodds
- 🎓 Five Lectures on Networks, Aaron Clauset
- 🎓 Network Science course, Albert-László Barabási

- 📖 Networks An Introduction Mark Newman, 2010.
- 📖 Networks, Crowds, and Markets, D. Easley, J. Kleinberg, 2010.
- 📖 Network Science, Albert-László Barabási, 2016.