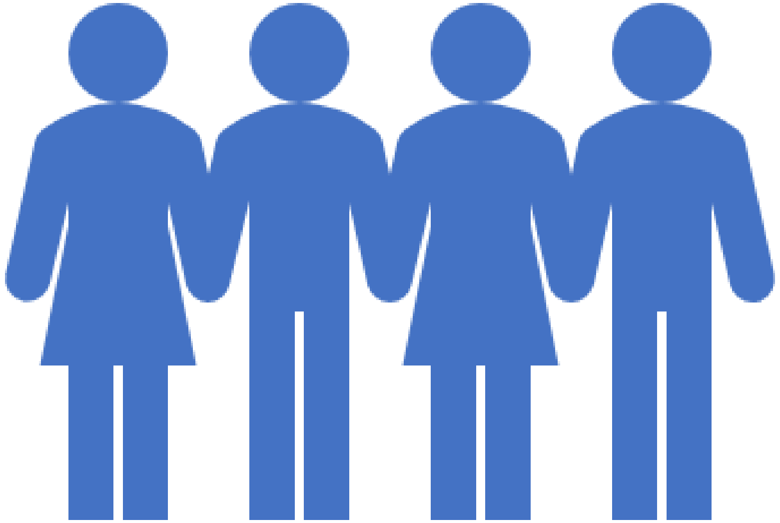
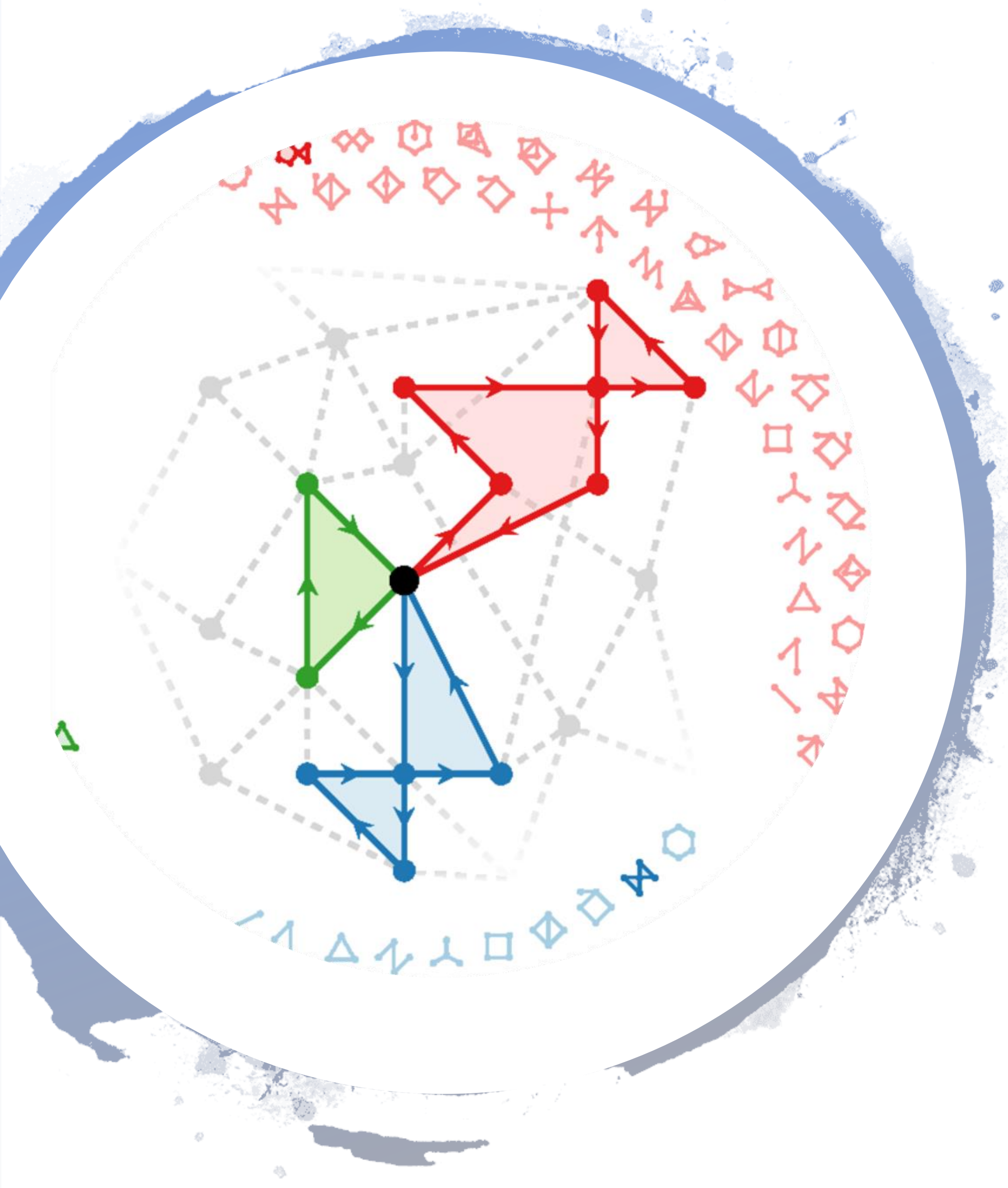
Network Analysis (INFOMNWA-2021)

Lecture 6: Network formation

Jiamin Ou



How individuals or complex systems are wired together? (by random chance or what?)

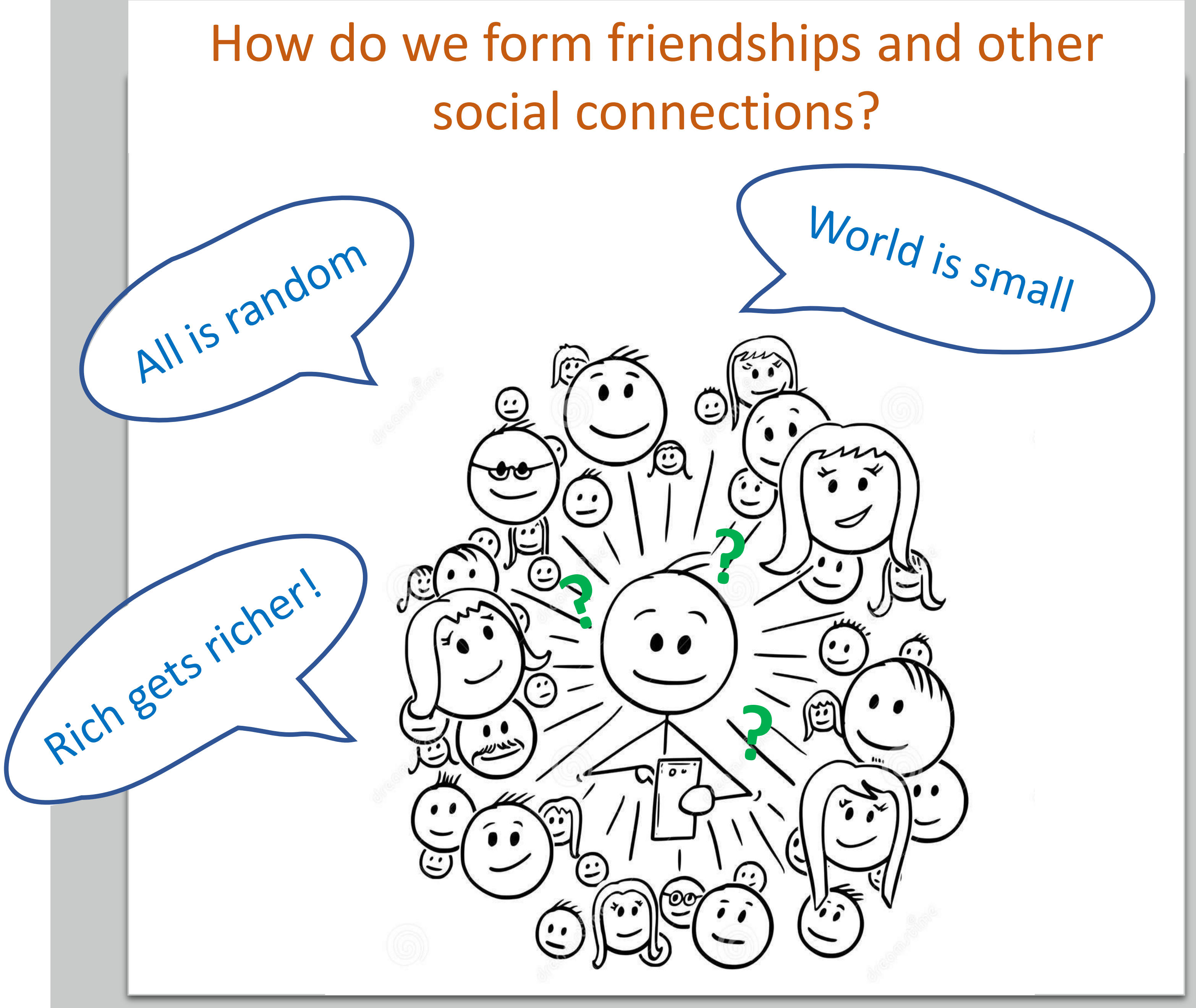


“Probably the most surprising discovery of modern network theory is the universality of the network topology: *Many real networks, from the cell to the Internet, independent of their age, function, and scope, converge to similar architectures*. It is this universality that allowed researchers from different disciplines to embrace network theory as a common paradigm.

------ Albert-László Barabási, 2009, Science, 325: 412-4.

Today's programe

- **Erdös–Rényi Random Graph Model**
- **Small-world Random Graph Model**
- **Barabasi-Albert (BA) model**

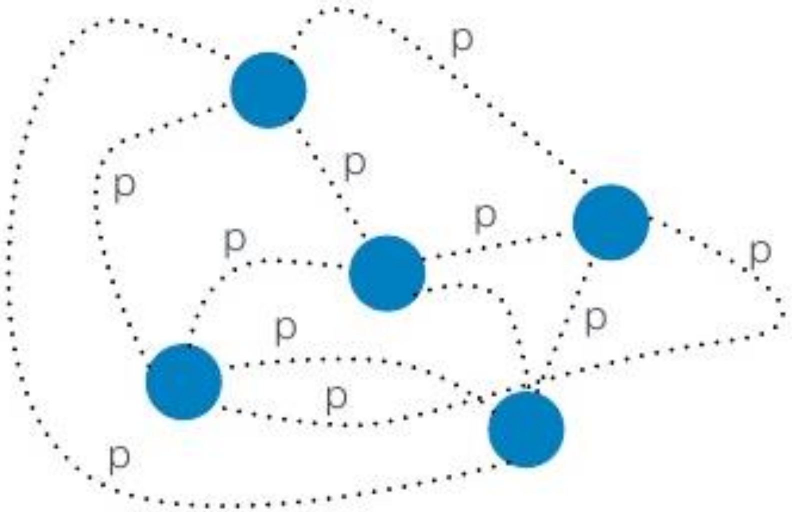


An evolution on our understanding on network formation

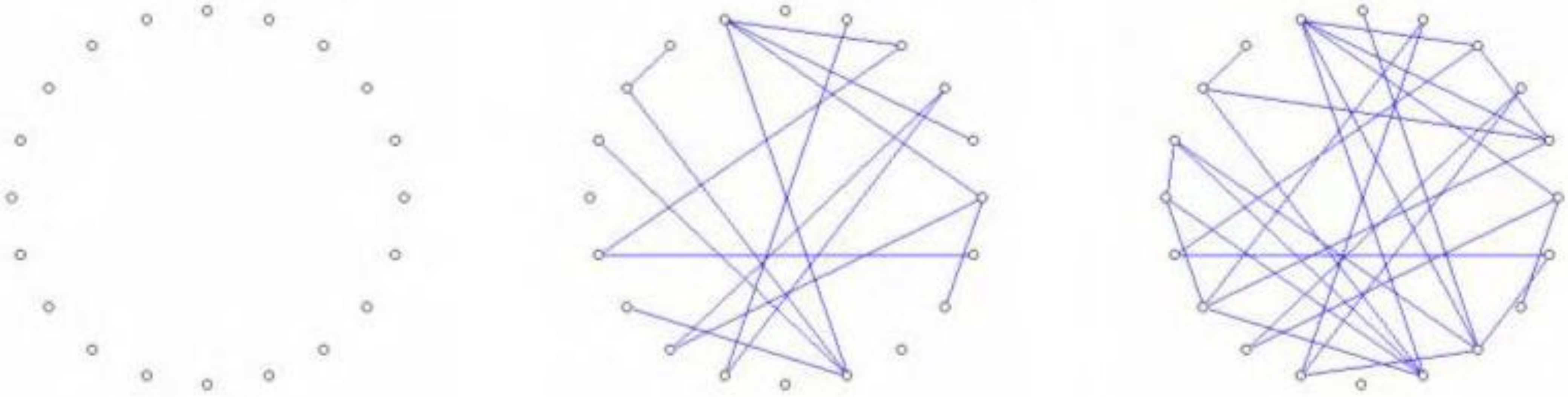
"i.i.d and random": Pál Erdös and Alfréd Rényi, 1959 (Erdös–Rényi (ER) **Random Graph Model**)

$G_{n,p}$: undirected graph on *n* nodes and each edge (u,v) appears i.i.d with probability *p* **(edges are independent)**

$G_{n,m}$: undirected graph with *n* nodes and *m* edges, where edges are picked uniformly at random **(each edge is equally likely/unweighted)**



As *p* increases….



$p = 0$

(a)

$p = 0.1$

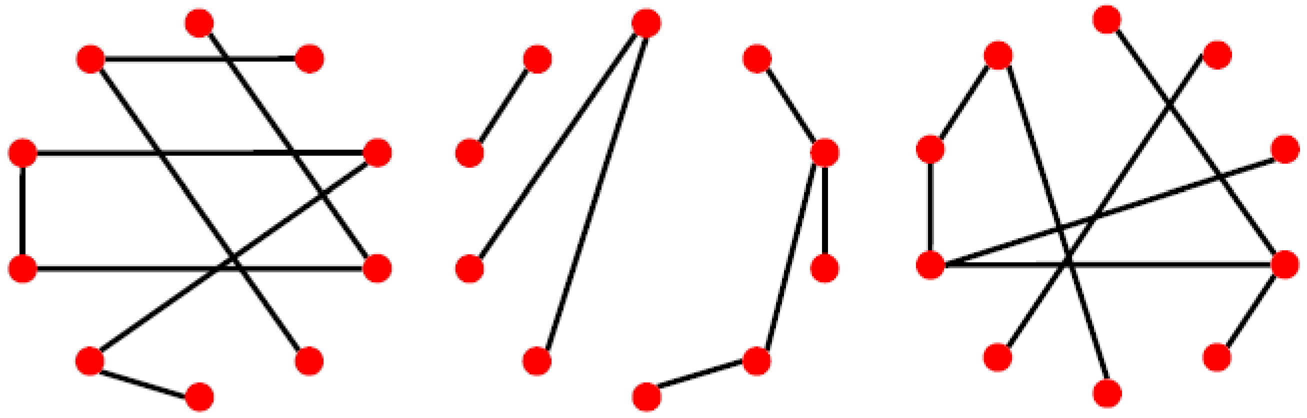
(b)

$p = 0.2$

(c)

For a set of n and p, do we have a unique graph?

- *n* and *p* do not uniquely determine the graph!
  - The graph is a result of a random process
- We can have many different realizations given the same *n* and *p*

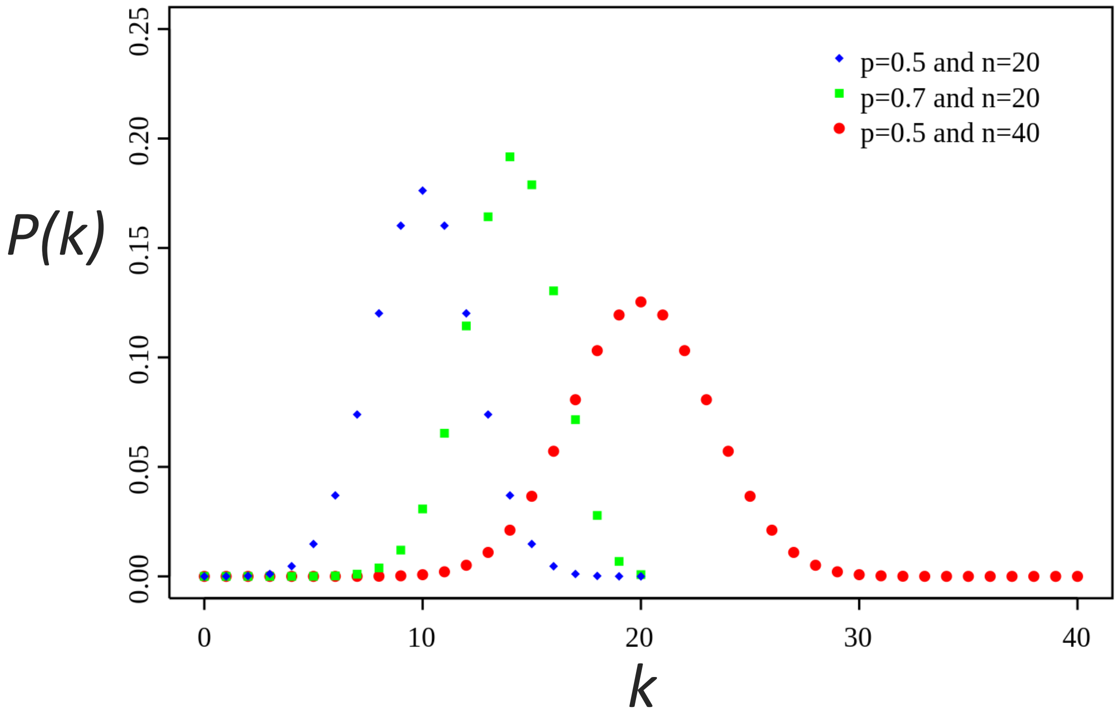


n=10 p=1/6

Key properties of random graph

1. Degree distribution: binomial

The degree by nodes (d_i) follow a distribution of $\{(p_{k(i)}, d_i), i \in N\}$



As the network size increases (n), the distribution becomes increasingly narrow- with increasingly confidence that the degree of a node is in the vicinity of k (as variance decreases)

Key properties of random graph

2. Clustering coefficient: small or large?

$$C = \frac{pk_i(k_i-1)}{k_i(k_i-1)} = p = \frac{\bar{k}}{n-1} \approx \frac{\bar{k}}{n}$$

For node i, clustering coefficient

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

e_i is the number of edges between i's neighbours

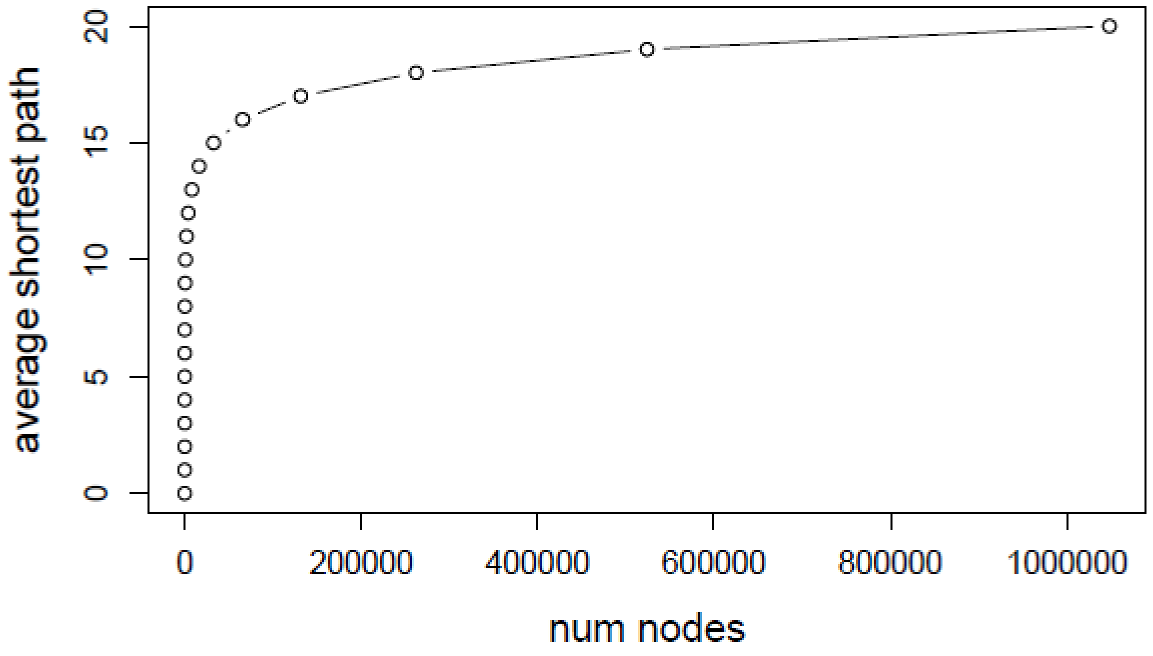
$$e_i = p\,\frac{k_i(k_i-1)}{2}$$

- Clustering coefficient of a random graph is *small*
- Bigger graphs with the same average degree k have lower clustering coefficient

Key properties of random graph

3. Average shortest path

Average path length = O (log n)



- Networks can grow to be very large but nodes will be just a few hops apart
- ER-graphs have a small average shortest path length

Our understanding on network formation

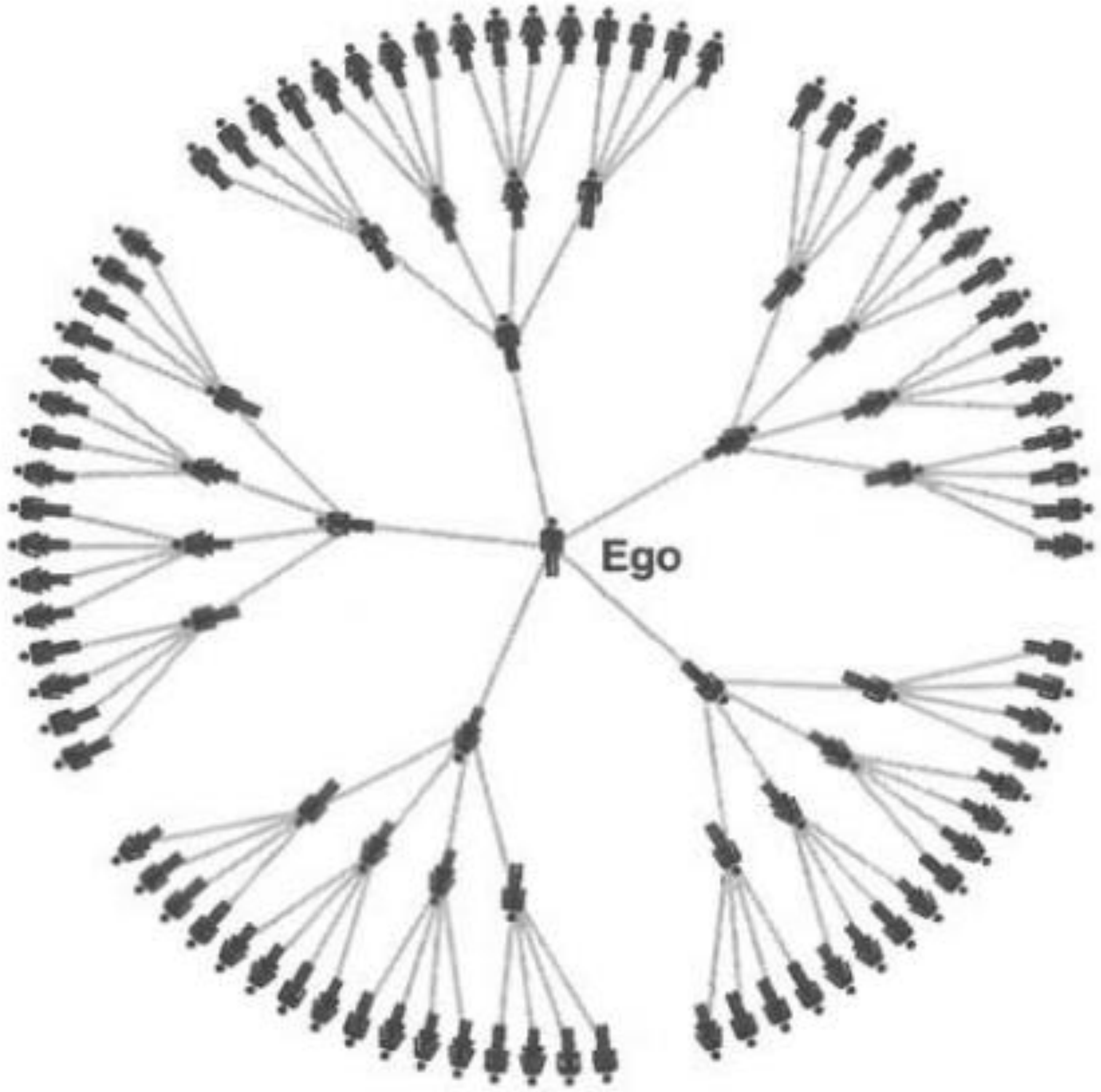
No local structure
(e.g., no triadic closure: friend of a friend is my friend)

E-R Random Graph Model
(Low clustering;
Small average shortest path length)

Small-world Random Graph Model
(High clustering;
Small average shortest path length)

Small world: Six degrees of separation

- Pick two people at random
- Try to measure their distance: A knows B knows C ...
- Experiment: Let Alice try to get a letter to Zach, whom she does not know.
- Strategy by Alice: choose Bob who she thinks has a better chance of reaching Zach.
- Result: On average 5.5 hops before letter reaches target.



- A pure branch network
- Ego knows only 5 people, but within two degrees of separation, ego can reach 25; within three degrees, 105 and so on.

Further testing of the small-world phenomenon

Erdős number (1969 till now): **the "collaborative distance"** between mathematician Paul Erdős and another person

Erdős wrote 1,500 mathematical articles in his lifetime and had 512 direct collaborators: the people with Erdős number 1.

The people who have collaborated with them (but not with Erdős himself) have an Erdős number of 2 (12,600 people as of 7 August, 2020),

Those who have collaborated with people who have an Erdős number of 2 (but not with Erdős or anyone with an Erdős number of 1) have an Erdős number of 3...

Now, **leading mathematicians tend to have particularly low Erdős numbers**:

– As of 2016, all Fields Medalists have a finite Erdős number, with values that range between 2 and 6, and a median of 3.

– The median Erdős number across all mathematicians

is **5**, with an extreme value of **13**.

Erdős spent a large portion of his later life living out of a suitcase, visiting his over 500 collaborators around the world

Further testing of the small-world phenomenon

The international film actor network: The Kevin Bacon Game, 1997

Bacon number: number of steps to **Kevin Bacon**

– Create a network of Hollywood actors

– Connect two actors if they co-appeared in the movie

– As of Dec 2007, the highest Bacon number reported is 8

Computer network in 2001 (Duncan Watts)

– An e-mail message as the “package” to be delivered to 48,000 senders and 19 destinations in 157 countries in 2001– The average number of intermediate steps was about six

Columbia Small World Project in 2003

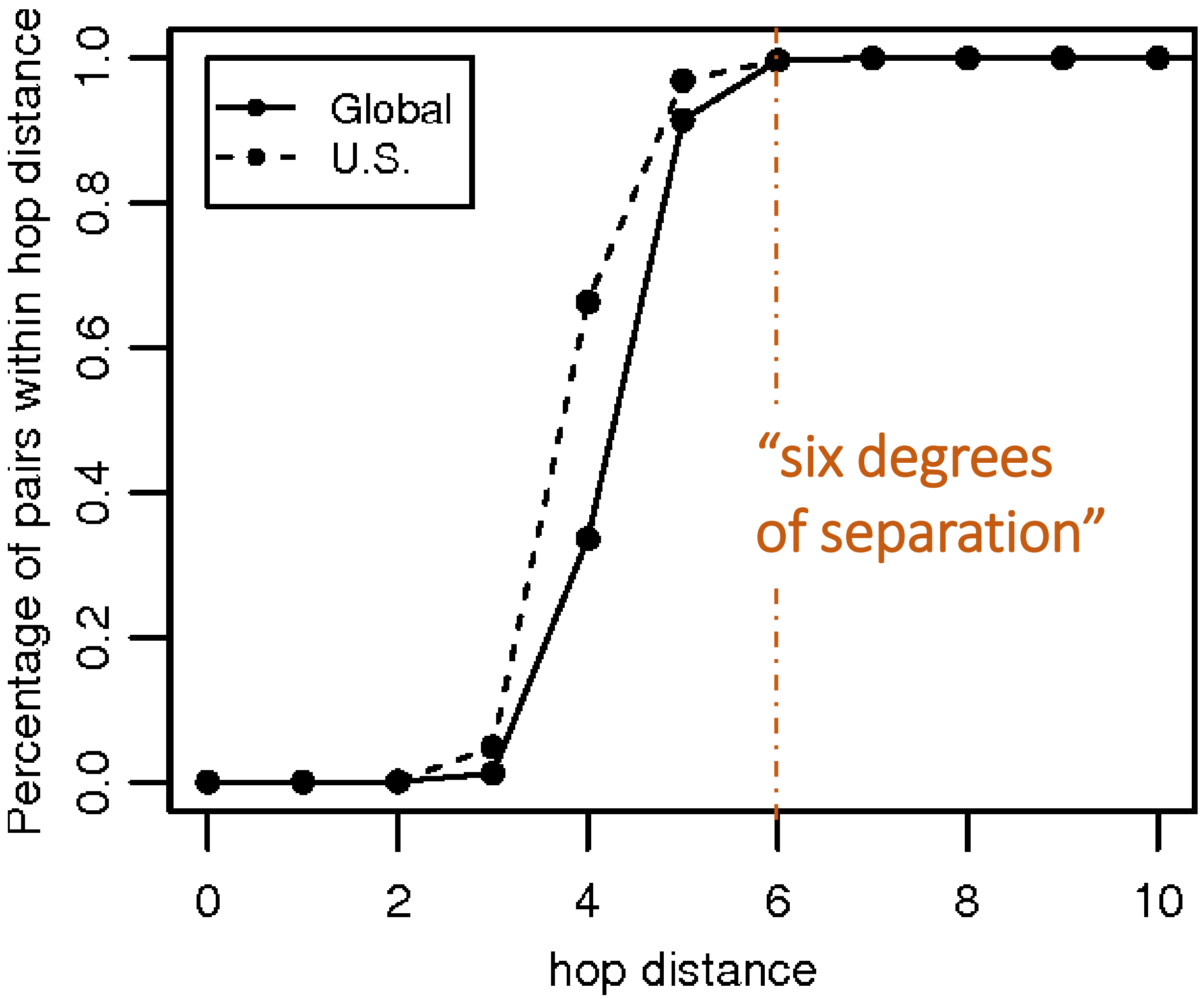
– 100,000 people registered

– Some only reached the goal after 7, 8, 9 or 10 steps (a limited selection of all internet users and high drop-out rate)

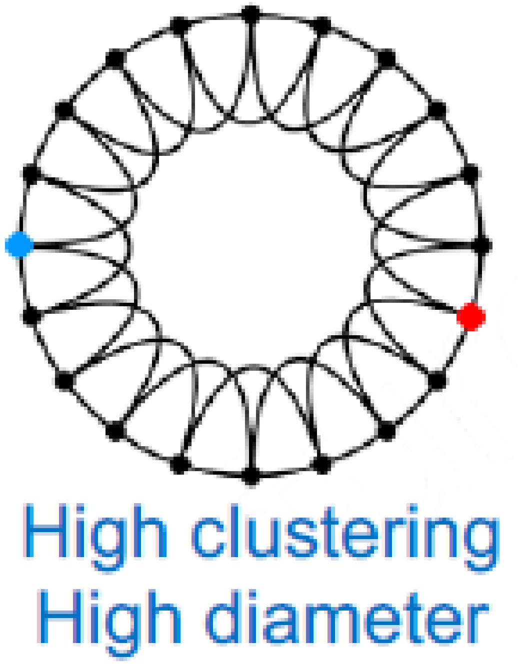
Renters in the same building (Haarlem)

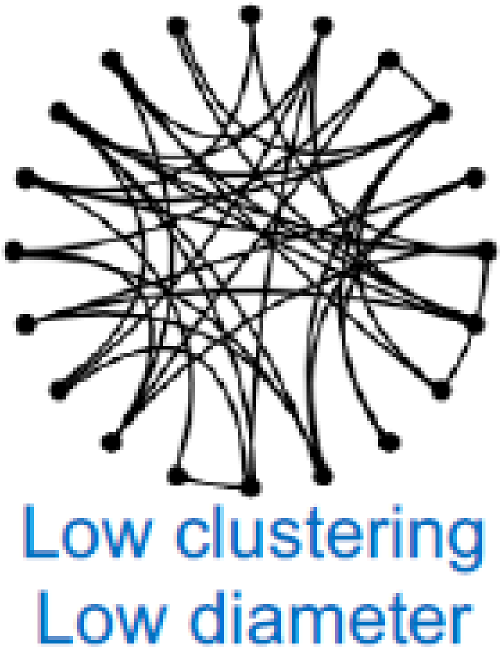
Facebook friendship

- The average distance between pairs of users was 4.7 for Facebook users and 4.3 for U.S. users.



Small-world: How?





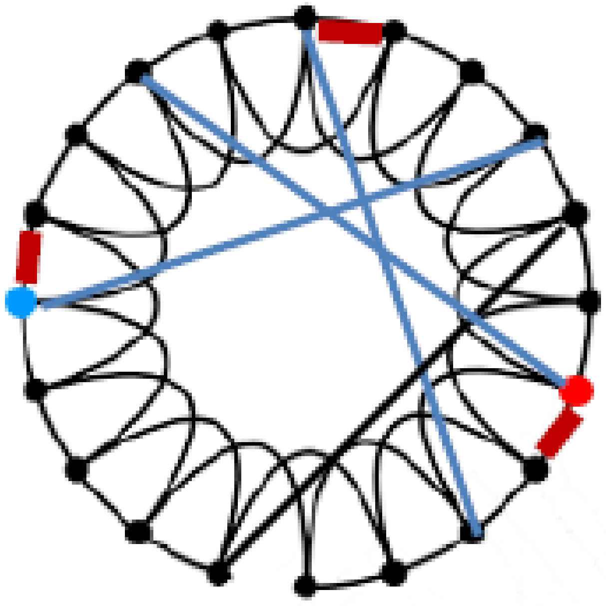
Can a network with high clustering be at the same time with small diameter?

The small-world model -Watts-Strogatz graphs

(1) Start with a low-dimensional regular lattice

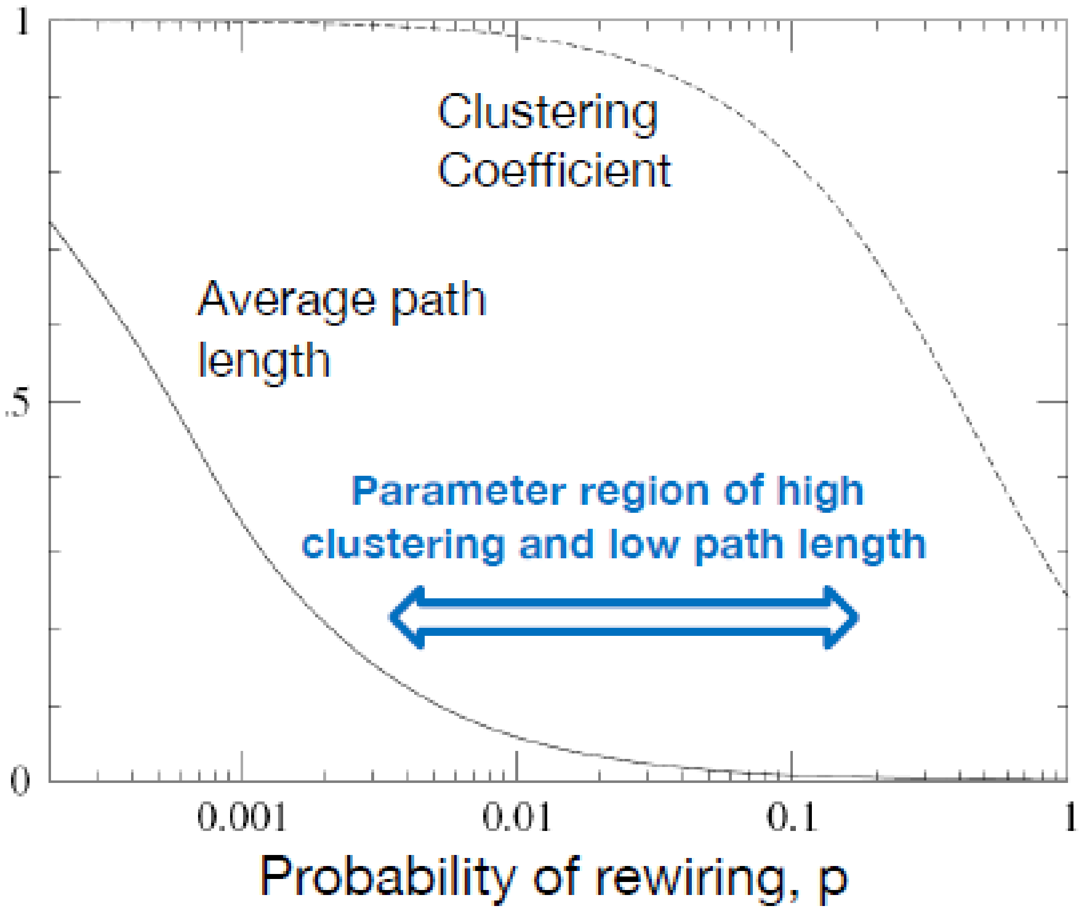
- Use a ring as a lattice
- Has both high clustering coefficient and high diameter

(2) Rewire: introduce shortcuts between clusters to reduce diameter



- Add/remove edges to creat shortcuts to join remote parts
- For each edge with a rewire probability p_r move the other end to another hop

How to get the rewire probability right



Srijan Kumar, Georgia Tech, CSE6240 Spring 2020: Web Search and Text Mining

Clustering is hard to break down, but not diameter

Our understanding on network formation

E-R Random Graph Model

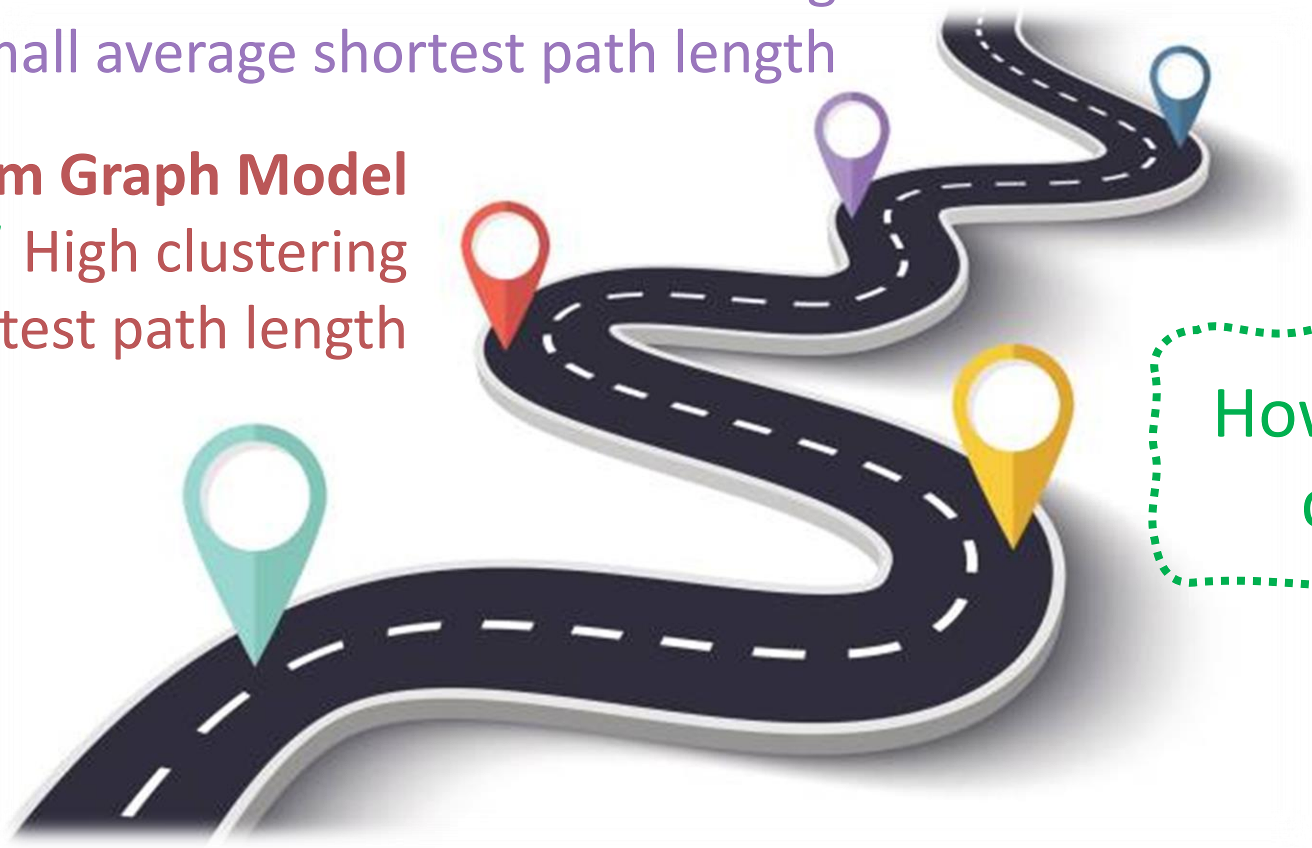
Low clustering

√ Small average shortest path length

Small-world Random Graph Model

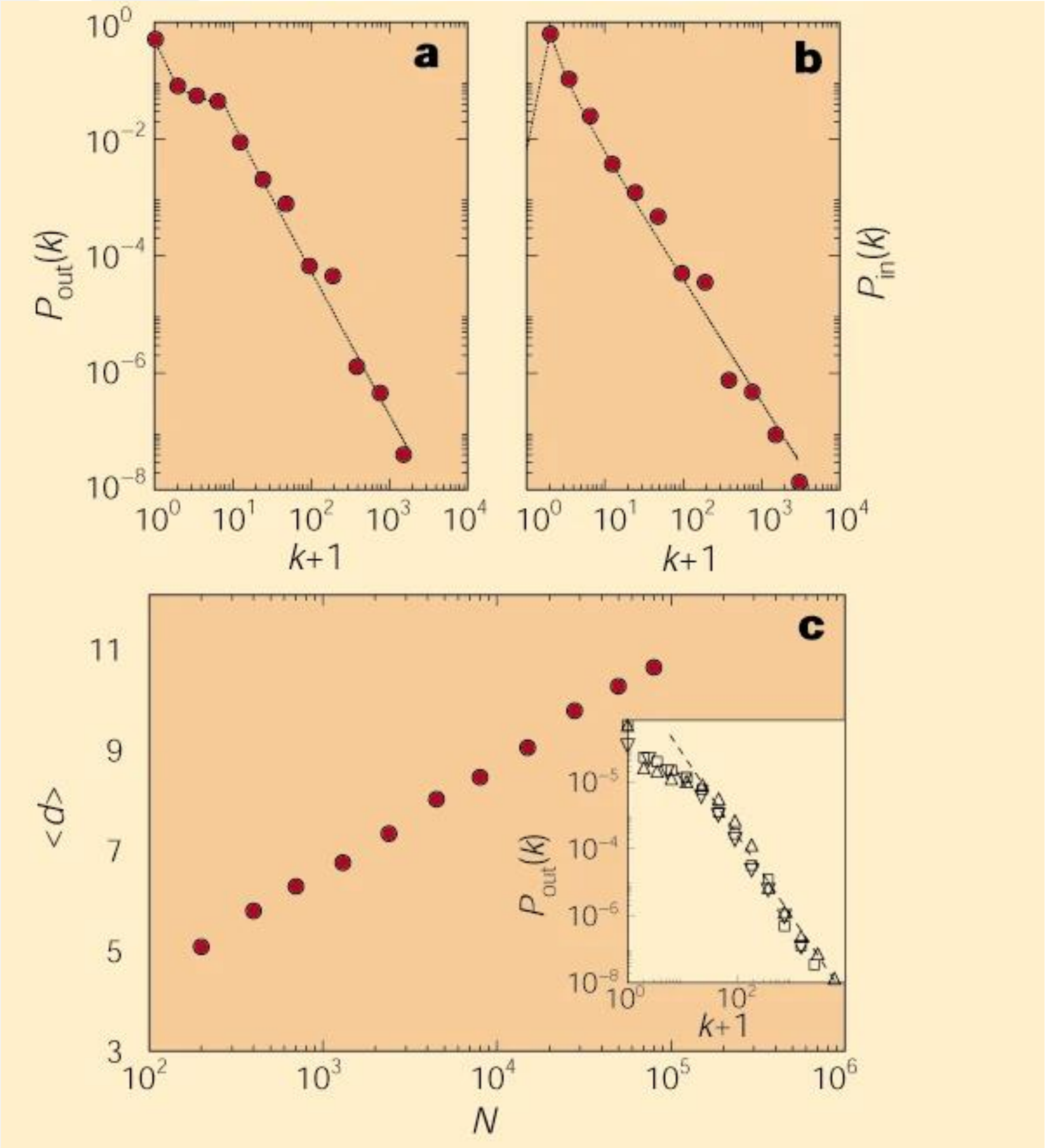
√ High clustering

√ Small average shortest path length

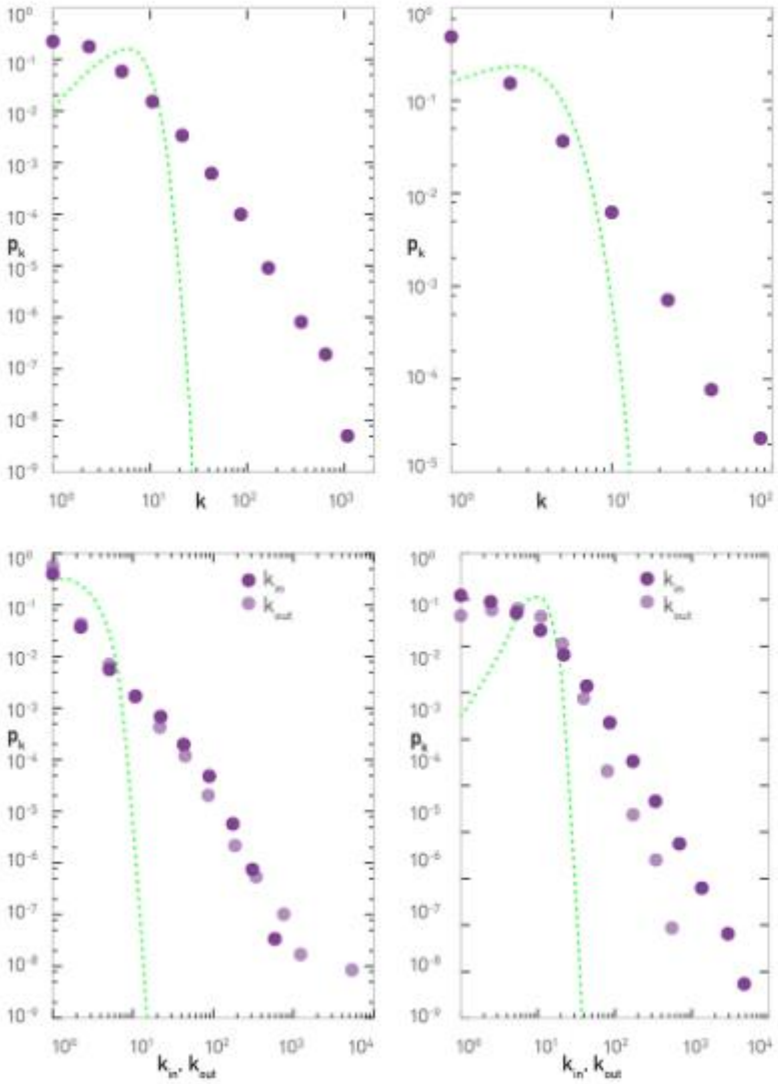


How about degree distribution?

Distribution of links on the World-Wide Web.



R. Albert, H. Jeong, A.-L. Barabási, Nature 401, 130 (1999).



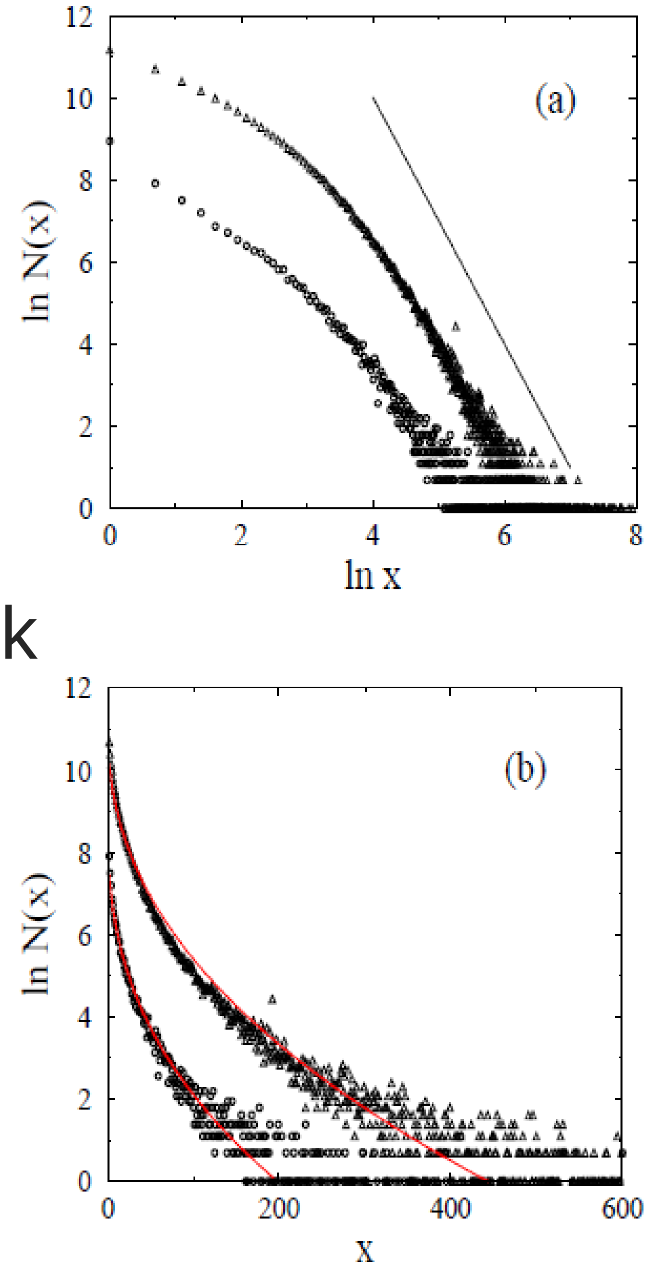
The probability that a Web page has exactly k links (in other words, degree *k*) follows a power law distribution $P(k) \sim k^{-\gamma}$

Green: ER graph distribution

With a large-citation tail of the citation distribution

Similar observations for the network of actors linked by movies and scientific papers linked by citations

Why??



S. Redner, Eur. Phys. J. B 4, 131 (1998).

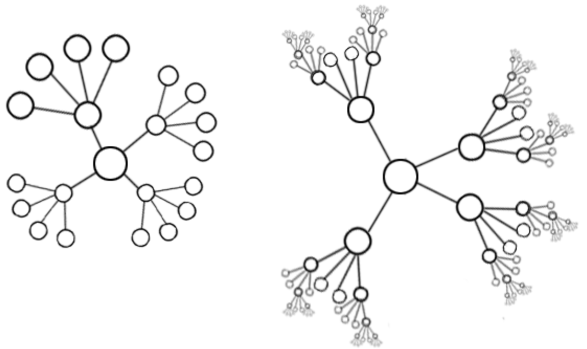
Preferential attachment

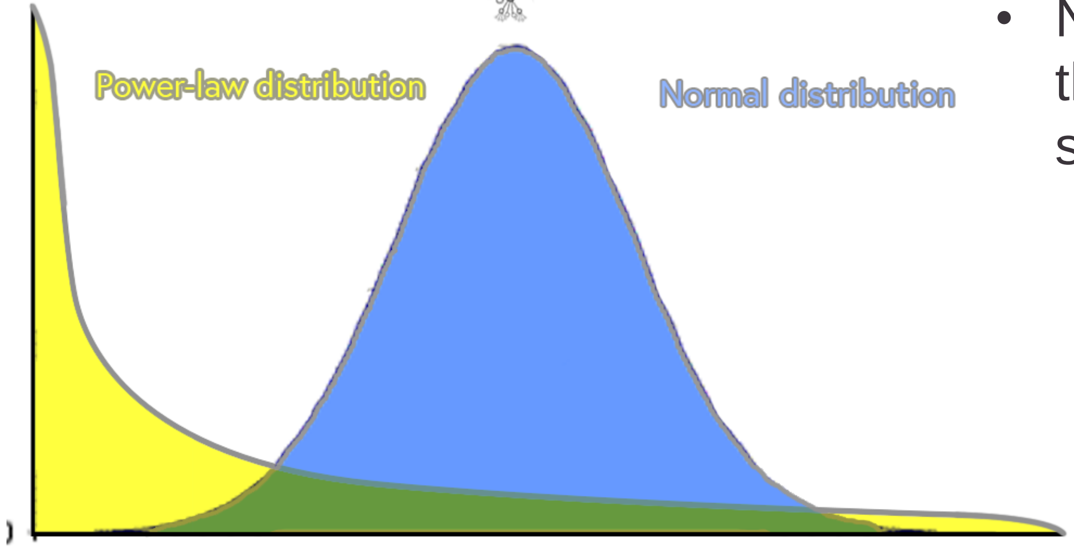
- Real-world network construction is a continues growth (path dependence)
- New nodes in most real networks prefer to link to the more connected nodes
- “Cumulative advantage", "the rich get richer" ("Matthew effect“)

Scale-free networks

Distribution of the number of edges of the nodes following a power law distribution;

Result from a growth process combined with preferential attachment.





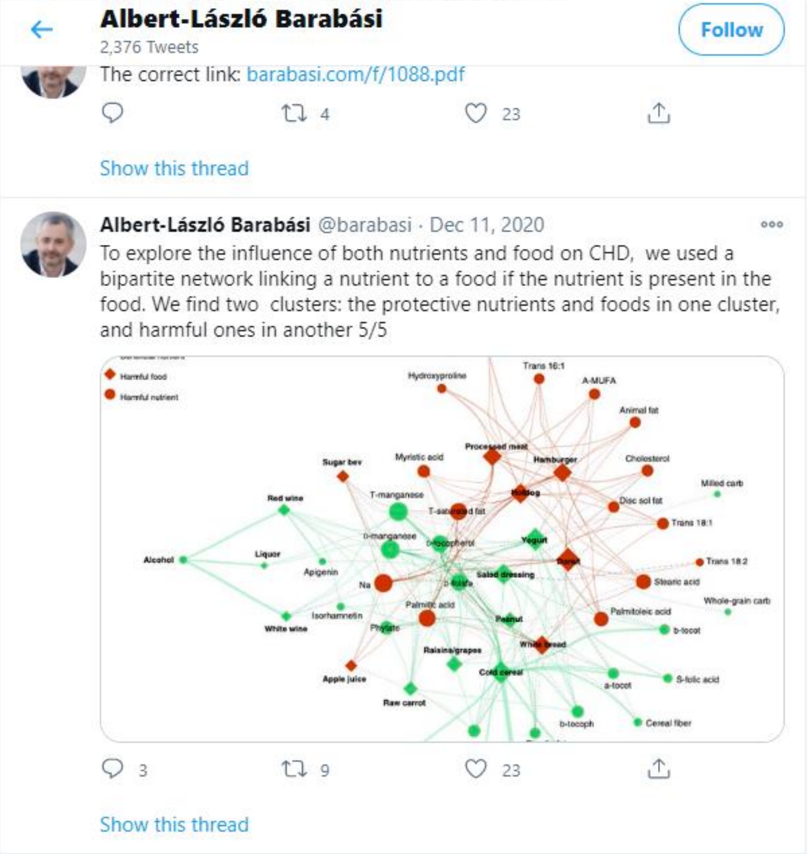
- Nodes with really high numbers of edges is much higher in the power-law distribution than in the normal distribution (a small number of very highly connected nodes).

Barabasi-Albert (BA) model

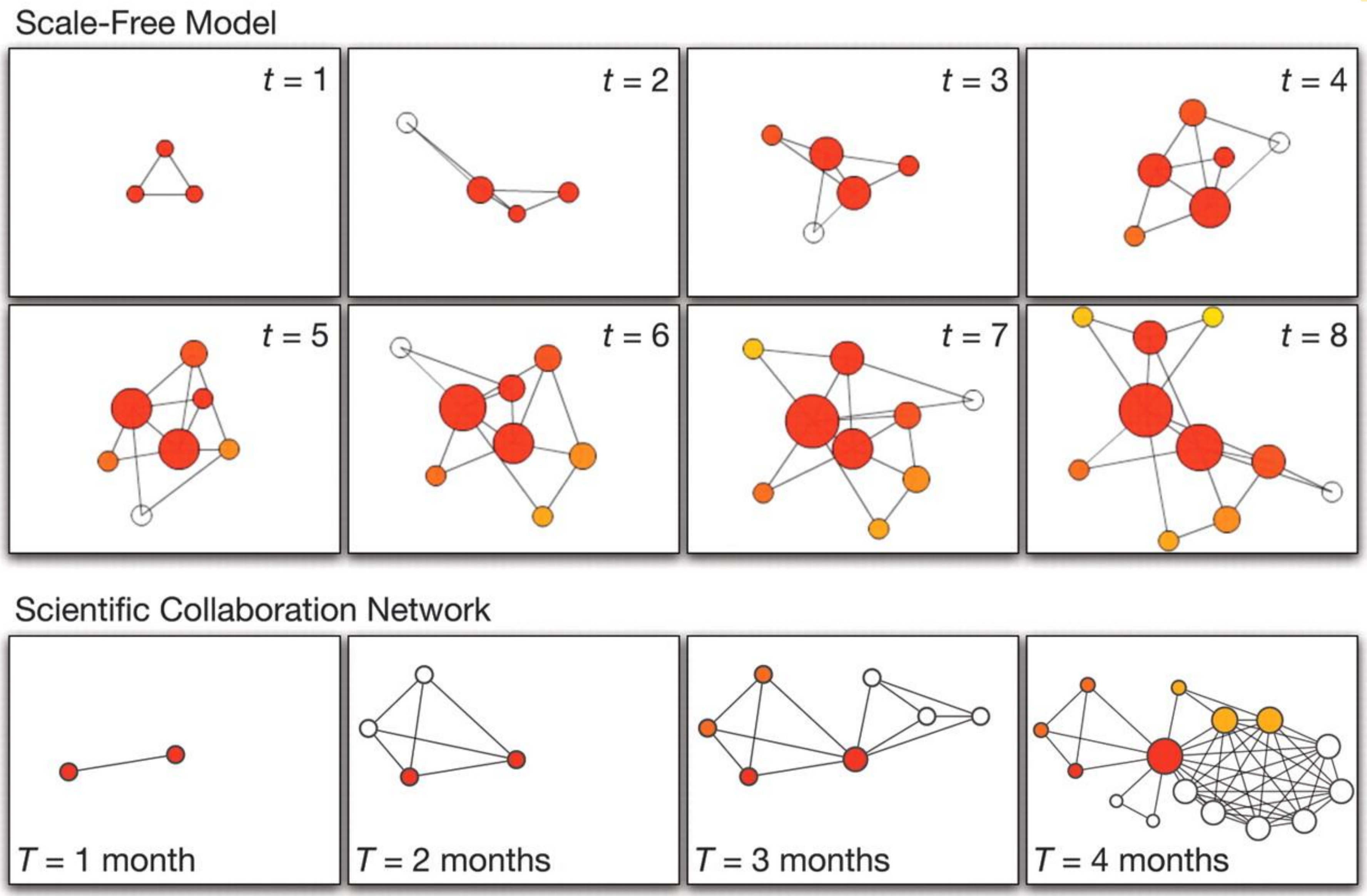
The probability ∏ (k_i) that a link of the new node connects to node i depends on the degree k_i,

$$\Pi\ (k_i) = \frac{k_i}{\sum_j k_j}$$

A probabilistic mechanism: A new node is free to connect to *any* node in the network, whether it is a hub or has a single link. However, that if a new node has a choice between a degree-two and a degree-four node, it is twice as likely that it connects to the degree-four node.

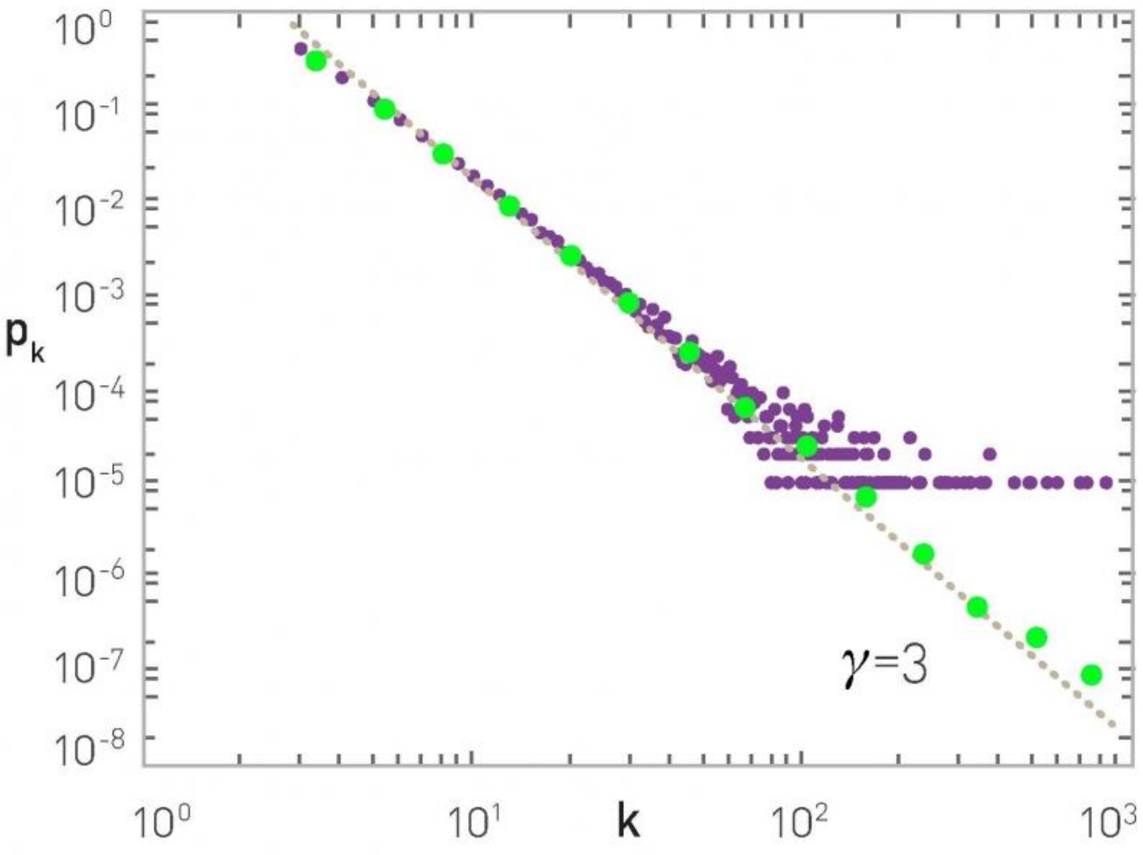


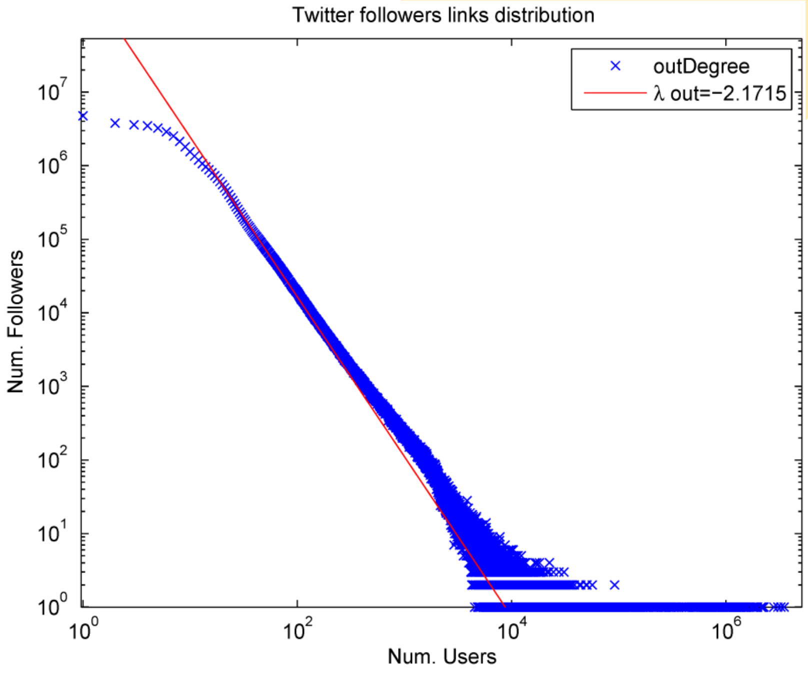
The birth of a scale-free network

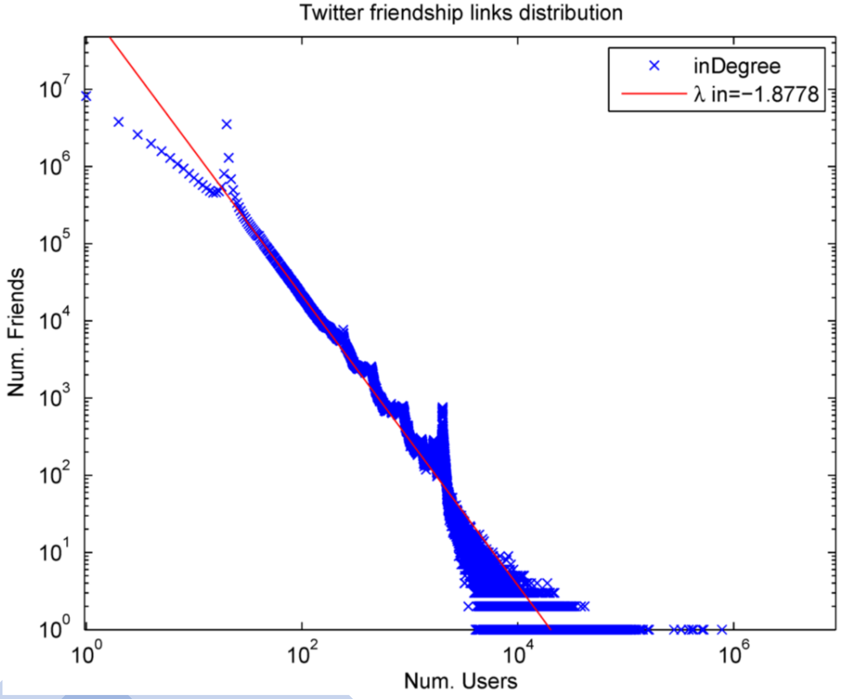


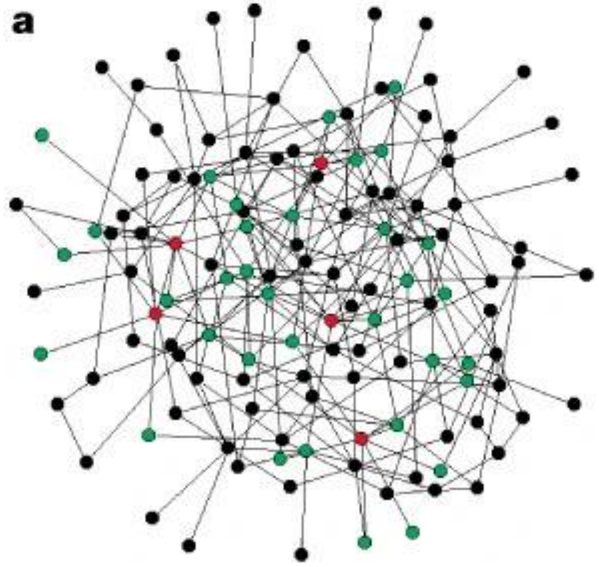
Degree distribution of Twitter followers and friendships
Many real-world networks: 2-8

A standard degree distribution from BA model

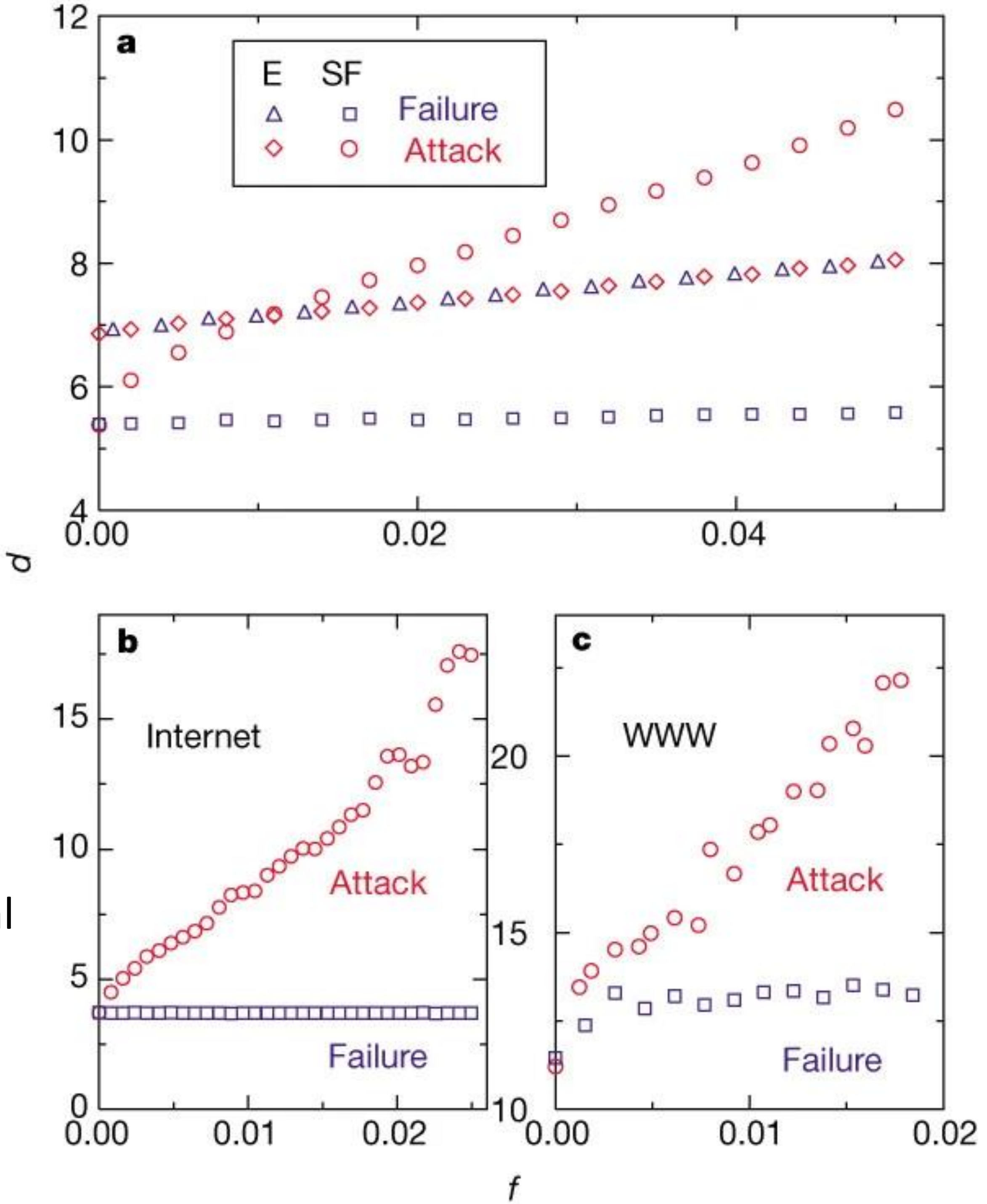








Erdös–Rényi (ER) network   Scale-free

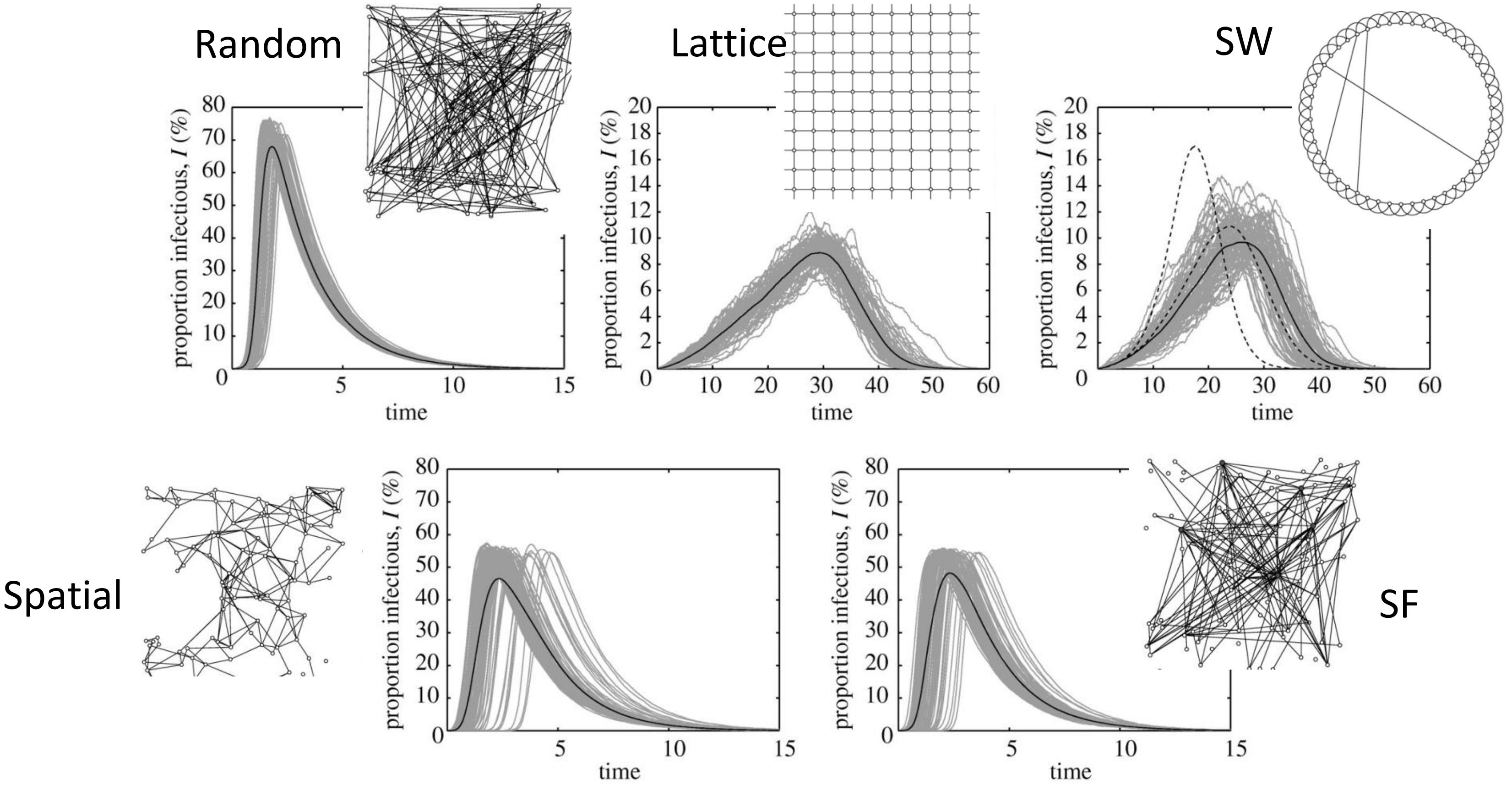


High error tolerance: the ability of their nodes to communicate being unaffected even by unrealistically high failure rates (random damage)

But extremely vulnerable to targeted attacks (that is, to the selection and removal of a few nodes that play a vital role in maintaining the network's connectivity).

Such error tolerance and attack vulnerability are **generic properties of communication networks**

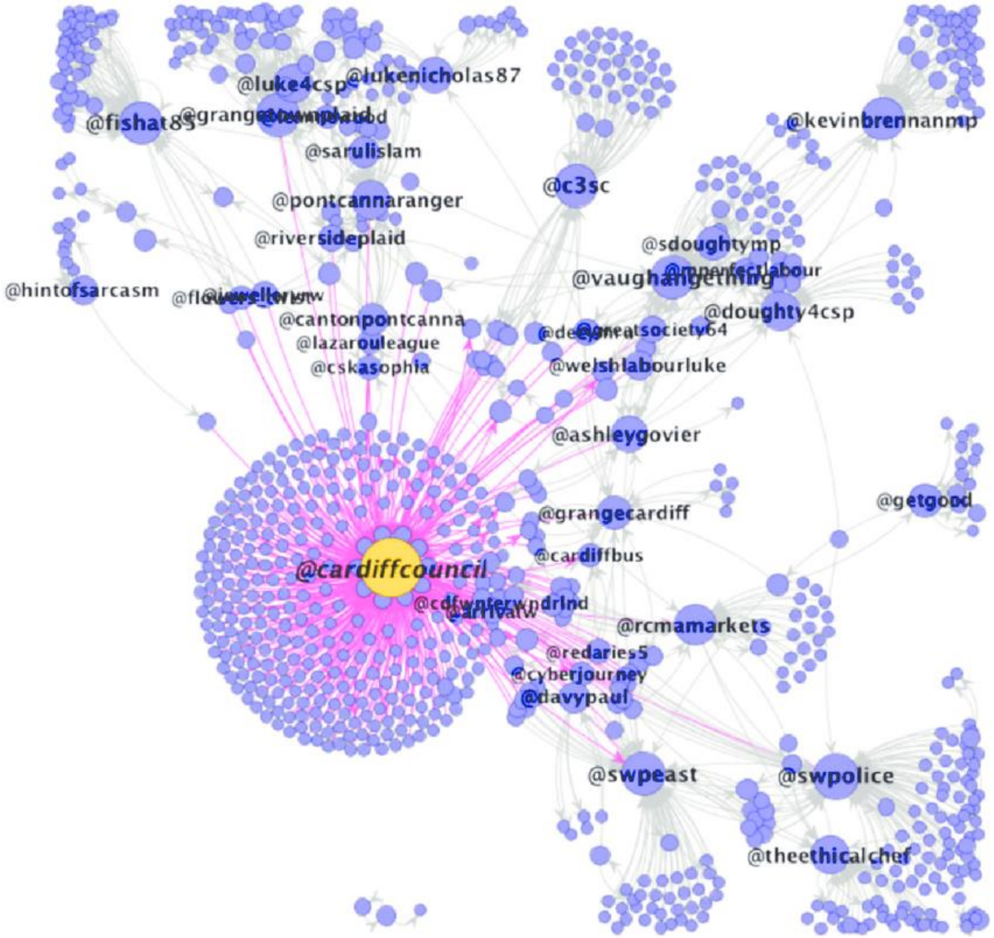
- Small number of highly connected hubs: **important for information and diseases spread** (e.g., epidemiology, public relations and marketing).
- Super-spreaders, core groups: at greatest risk of infections and, once infected, can transmit the disease to many other
- Random vaccination << targeted vaccination



Romualdo Pastor-Satorras and Alessandro Vespignani, Epidemic Spreading in Scale-Free Networks, 2001

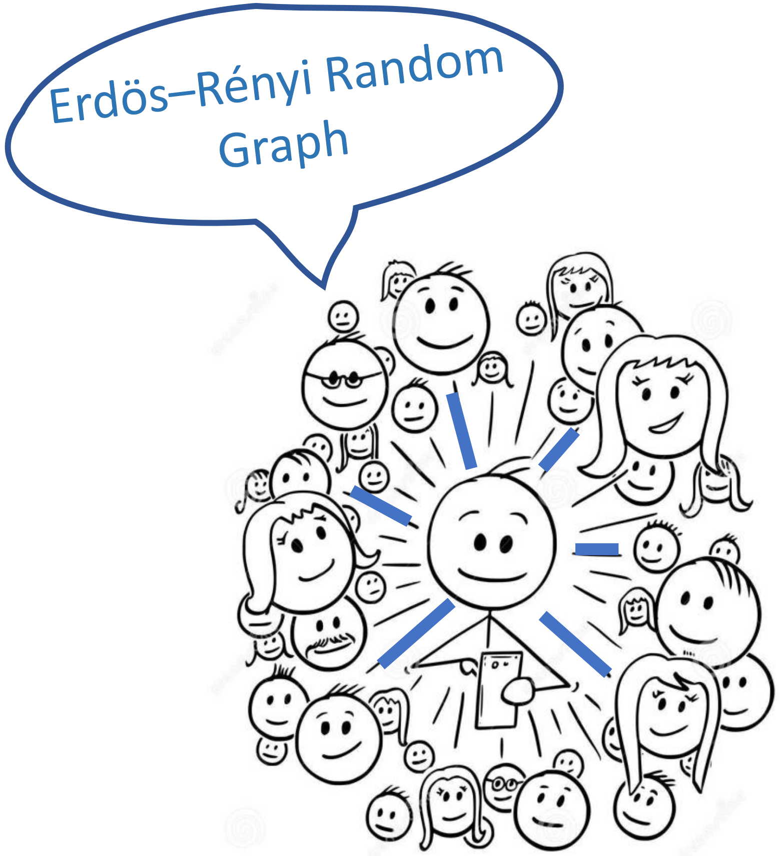
Scale-free networks in real-life

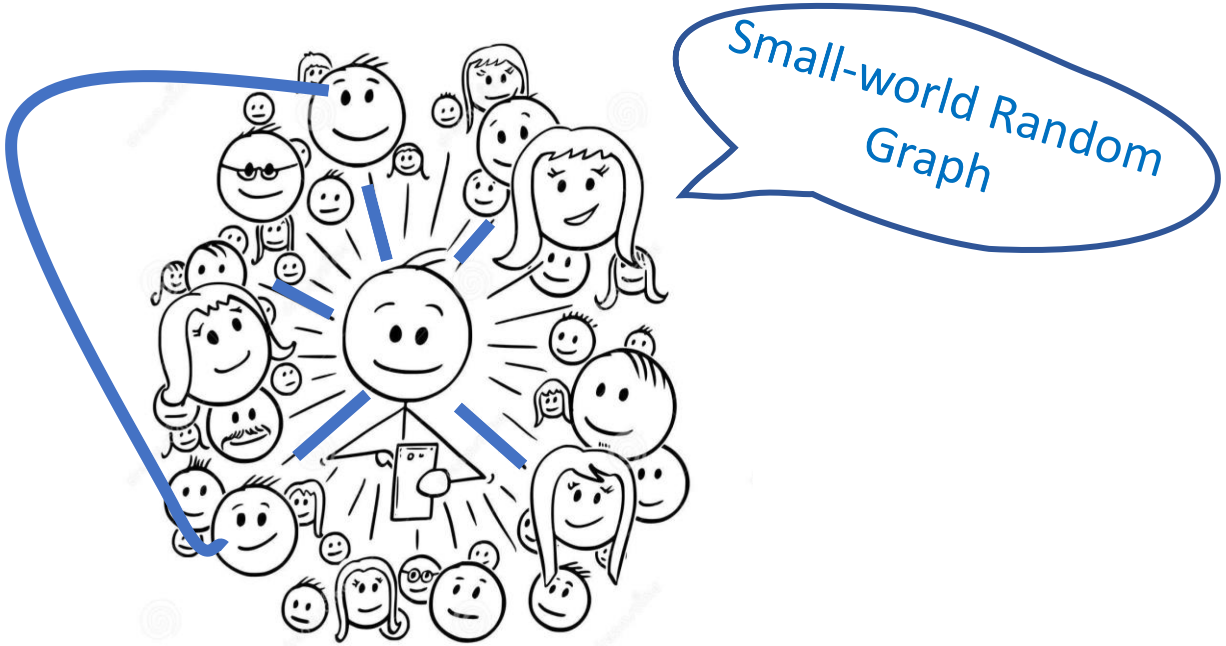
- Collaboration network (movie stars, scholars)
- The webgraph of the World Wide Web
- Modern online social networks (OSNs)
- Some financial networks such as interbank payment networks
- Airline network

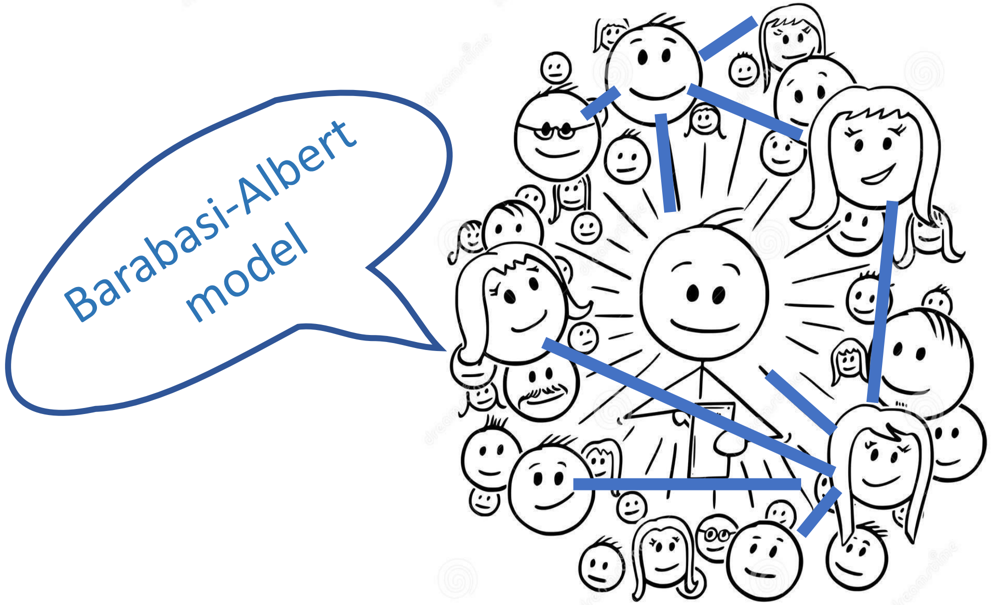


Visualization of the social network of Twitter communications in West Cardiff

Recap: How do we form friendships and other social connections?







Questions?