Computational Social Science
Social Network Analysis

Fariba Karimi

Prof.

Computer Science Department, TU Graz

e-mail: karimi@tugraz.at

web: www.networkinequality.com

TU Graz, 22.10.2024

Recap

History of Social psychology

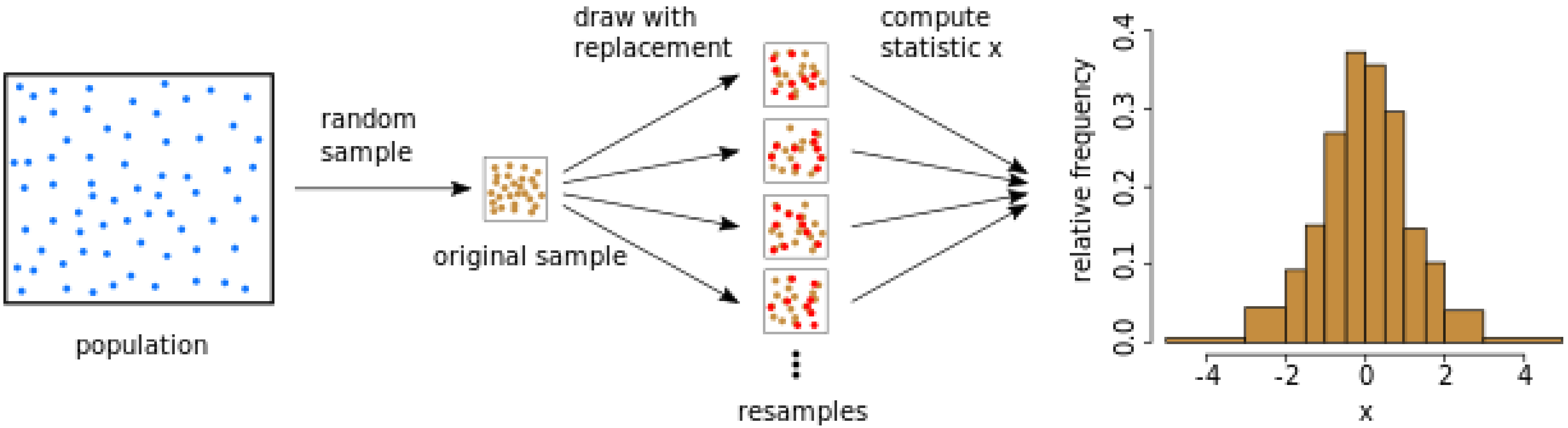
Social Impact Theory

and magnetism) operating in a social force field or social structure. As an example of what I mean by a social force field, Figure 1 depicts the plight of a hapless striped target beset by a variety of spotted sources, all having some impact.

Principle 1: Social Forces, I = f(*SIN*)

As a first principle, I suggest that when some number of social sources are acting on a target individual, the amount of impact experienced by the target should be a multiplicative function of the

Recap
Bootstrapping



Source: Wikipedia

Network analysis

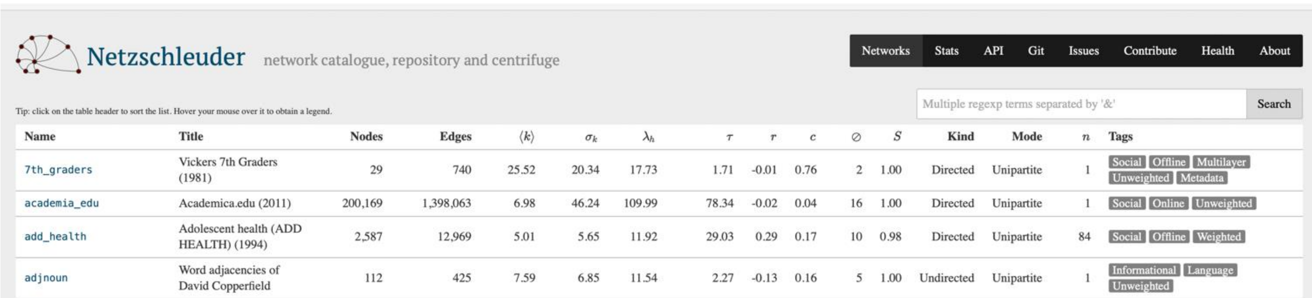
Network science & Graph theory have common roots

- Network elements: Nodes (vertex, actor), Links (edges, connections)

Different ways we use network analysis:

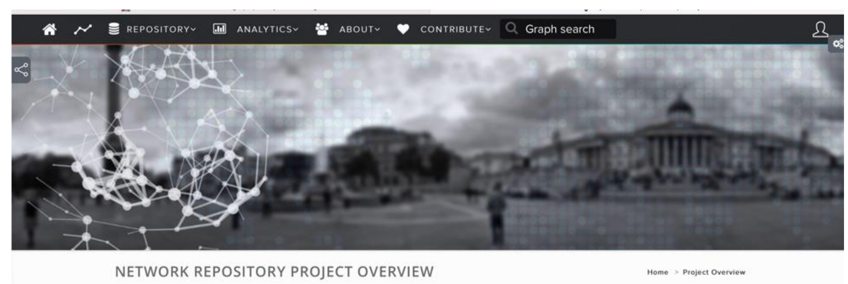
- Mapping empirical data into a network
- Generating/modeling synthetic networks with different mechanism
- Inferring mechanisms from real data : network inference

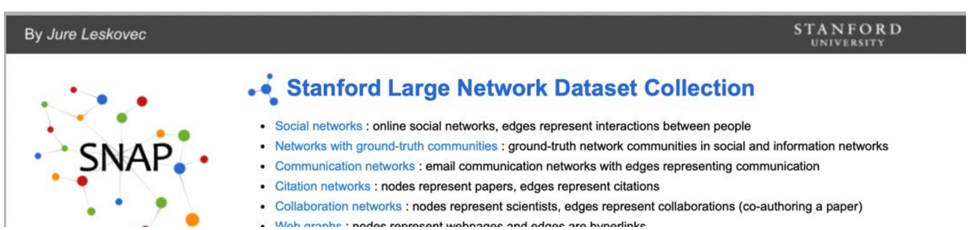
Network Datasets



Name	Title	Nodes	Edges	$\langle k\rangle$	σ_k	λ_h	τ	r	c	$\oslash$	S	Kind	Mode	n	Tags
7th_graders	Vickers 7th Graders (1981)	29	740	25.52	20.34	17.73	1.71	-0.01	0.76	2	1.00	Directed	Unipartite	1	Social Offline Multilayer Unweighted Metadata
academia_edu	Academica.edu (2011)	200,169	1,398,063	6.98	46.24	109.99	78.34	-0.02	0.04	16	1.00	Directed	Unipartite	1	Social Online Unweighted
add_health	Adolescent health (ADD HEALTH) (1994)	2,587	12,969	5.01	5.65	11.92	29.03	0.29	0.17	10	0.98	Directed	Unipartite	84	Social Offline Weighted
adjnoun	Word adjacencies of David Copperfield	112	425	7.59	6.85	11.54	2.27	-0.13	0.16	5	1.00	Undirected	Unipartite	1	Informational Language Unweighted

https://networks.skewed.de





https://networkrepository.com/platform.php

https://snap.stanford.edu/data/

Generating synthetic networks (network modeling)

- Understand the logic of algorithms
- **Write** your own code!!!
- Double-check your own code with the source code.

Example of a source code:

- https://networkx.org/documentation/stable/_modules/networkx/generators/random_graphs.html#gnp_random_graph

Overview of today‘s lecture

Theme: Properties of social networks and why they are important to study them.

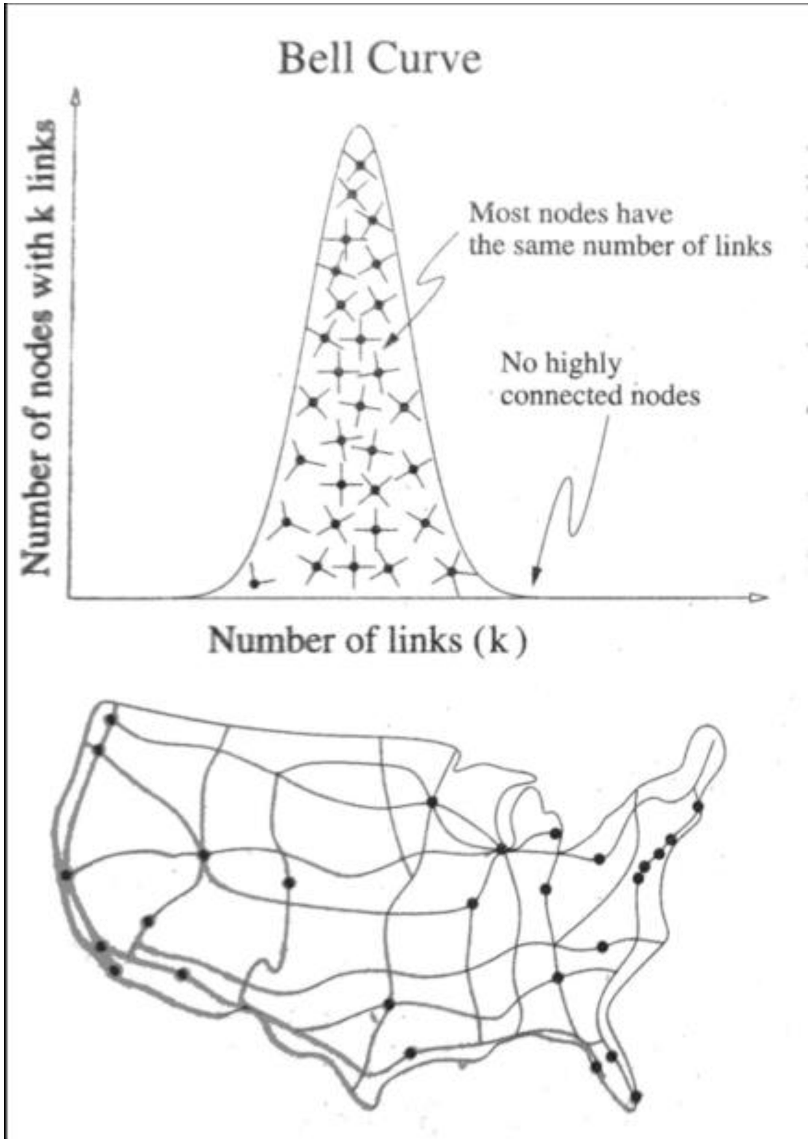
- Small world
- Popularity in social networks
- Many clusters
- Assortative

Part 1

SHAPE OF OUR SOCIAL NETWORKS: NON-RANDOMNESS

POPULARITY OR RICH GET RICHER EFFECT

Classical view to networks



Classical view to social network assumed that connections follow a normal distribution.

In this view, people in the network on average have a certain number of connections.

Researchers used random graph models as models of social networks. Examples of widely used random graph models are Erdős–Rényi model or ERGM.

Contemporary network science

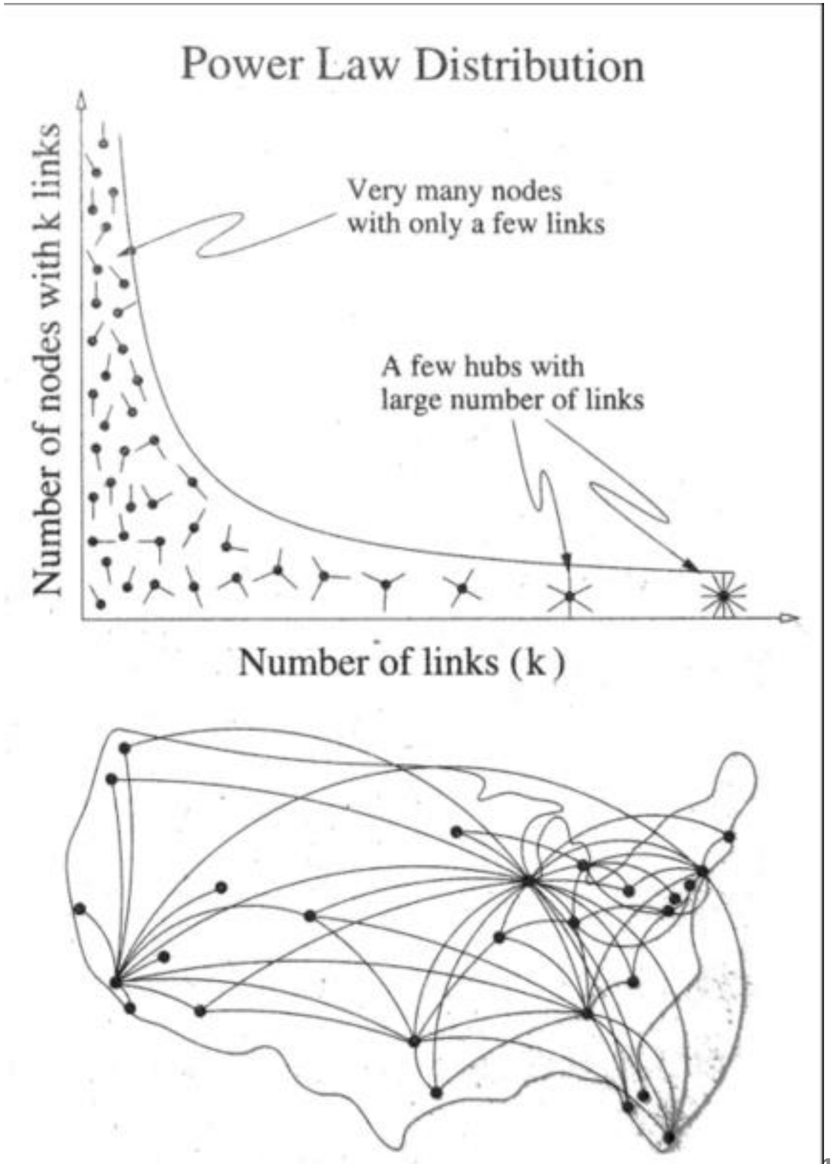
In reality, many large-scale social and technical systems follow a different kinds of distribution.

Many social and technical networks follow a power-law degree distribution.

Power-law networks are more heterogeneous.

We can no longer talk about an "average man".

Average degree is not well-defined. We need to revisit the mathematical tools.

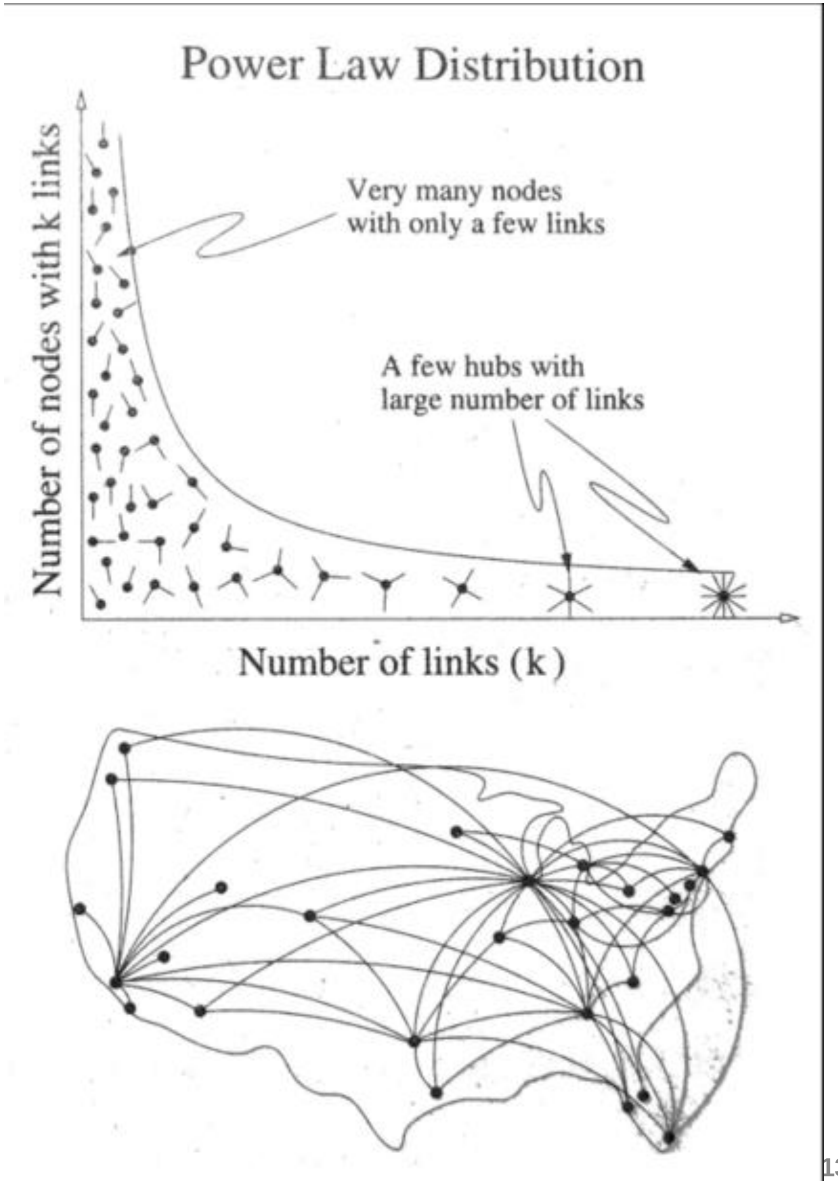


Contemporary network science

Plotting power-law networks

$P(k) \sim k^{-a} \sim 1/\, k^{a}$

$\text{Log }(p(k)) = ?$

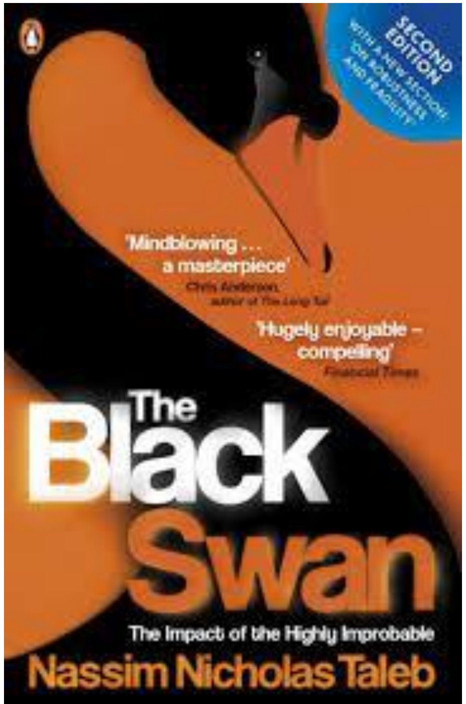


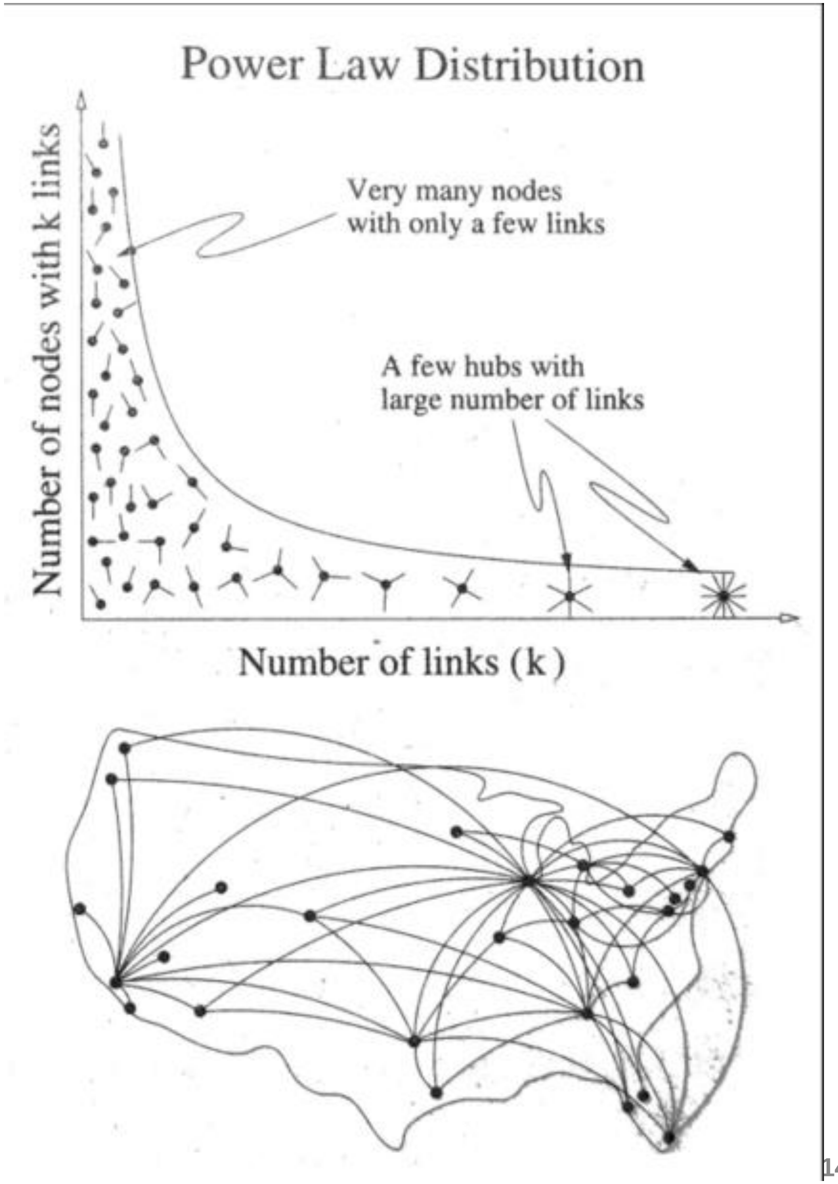
Contemporary network science

Plotting power-law (scale-free) networks

$$P(k) \sim k^{-a} \sim 1/\, k^{a}$$

$$\log (p(k)) = -a \log(k)$$





Power law distributions are observed in many socio-economic systems

Wikipedia
The Free Encyclopedia

Search Wikipedia | Search

Power law

22 languages

Article Talk | Read Edit View history Tools

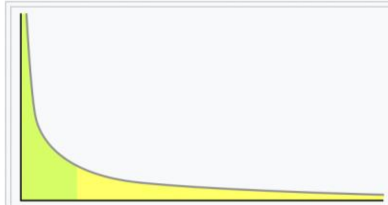
Contents hide

From Wikipedia, the free encyclopedia

Not to be confused with Force (law). For other uses, see Power (disambiguation).
"Scaling law" redirects here. For statistical laws of scaling deep learning models, see Neural scaling law.

In statistics, a **power law** is a functional relationship between two quantities, where a relative change in one quantity results in a relative change in the other quantity proportional to the change raised to a constant exponent: one quantity varies as a power of another. The change is independent of the initial size of those quantities.

For instance, the area of a square has a power law relationship with the length of its side, since if the length is doubled, the area is multiplied by 2^2, while if the length is tripled, the area is multiplied by 3^2, and so on.[1]



An example power-law graph that demonstrates ranking of popularity. To the right is the long tail, and to the left are the few that dominate (also known as the 80–20 rule).

Empirical examples [edit]

The distributions of a wide variety of physical, biological, and human-made phenomena approximately follow a power law over a wide range of magnitudes: these include the sizes of craters on the moon and of solar flares,[2] cloud sizes,[3] the foraging pattern of various species,[4] the sizes of activity patterns of neuronal populations,[5] the frequencies of words in most languages, frequencies of family names, the species richness in clades of organisms,[6] the sizes of power outages, volcanic eruptions,[7] human judgments of stimulus intensity[8][9] and many other quantities.[10] Empirical distributions can only fit a power law for a limited range of values, because a pure power law would allow for arbitrarily large or small values. Acoustic attenuation follows frequency power-laws within wide frequency bands for many complex media. Allometric scaling laws for relationships between biological variables are among the best known power-law functions in nature.

Question:

What mechanism generates power law networks, and how do we model it?

Preferential Attachment as a mechanism for generating scale-free networks [Barabasi 1999]

„The rich get richer“ effect or Matthew effect in sociology.

For whoever has will be given more, and they will have an abundance. Whoever does not have, even what they have will be taken from them. (Matthew 25:29)

In the sociology of science, "Matthew effect" was a term coined by Robert K. Merton to describe how, among other things, eminent scientists will often get more credit than a comparatively unknown researcher, even if their work is similar; it also means that credit will usually be given to researchers who are already famous

Preferential Attachment as a mechanism for generating scale-free networks [Barabasi 1999]

Barabasi and Albert defined a similar mechanism in netowrks:

Preferential Attachment refers to the high probability of a new vertex to connect to a vertex that already has a large number of connections

Example:

1. a new website linking to more established ones
2. a new individual linking to well-known individuals in a social network

Preferential Attachment Example

Which node has the highest probability of being linked by a new node in a network that exhibits traits of preferential attachment?

Example

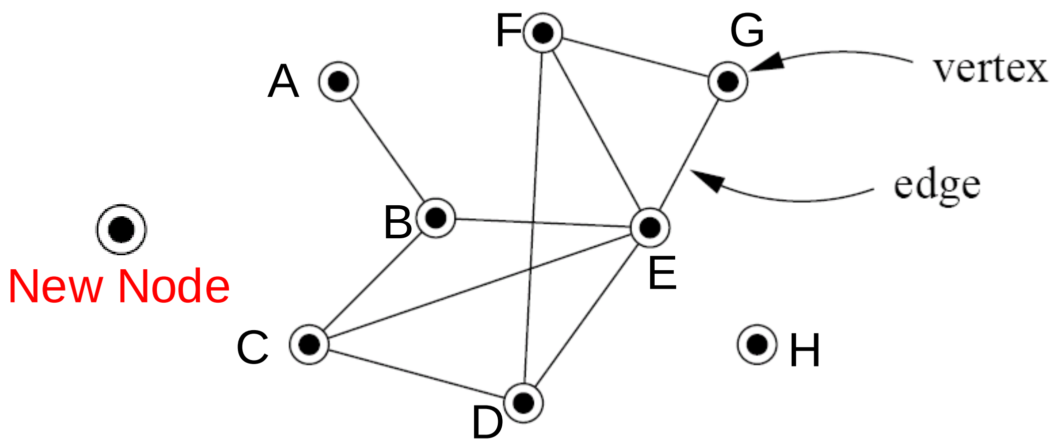


FIG. 1 A small example network with eight vertices and ten edges.

[Newman 2003]

Preferential Attachment Example

Which node has the highest probability of being linked by a new node in a network that exhibits traits of preferential attachment?

Example

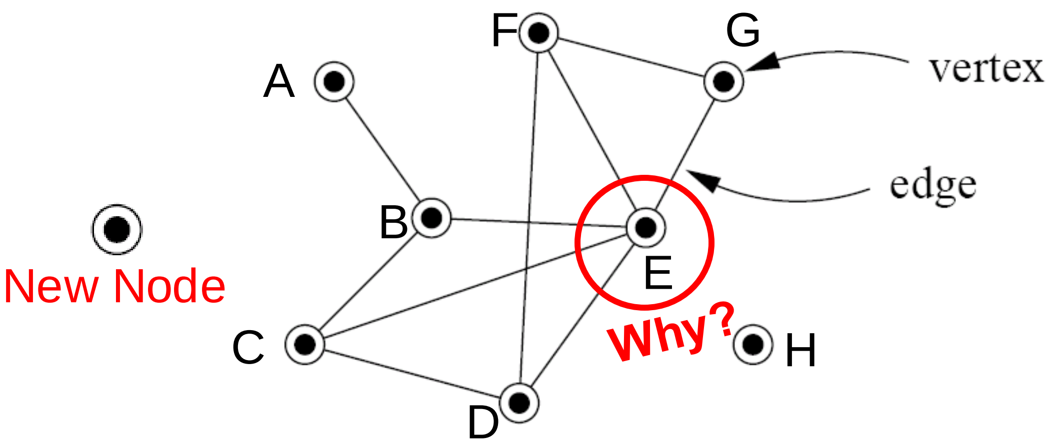


FIG. 1 A small example network with eight vertices and ten edges.

[Newman 2003]

Preferential attachment model

Demo:
http://estebanmoro.org/2012/11/preferential-attachment-be-first/

Preferential attachment mechanism produce the scale-free property in networks.

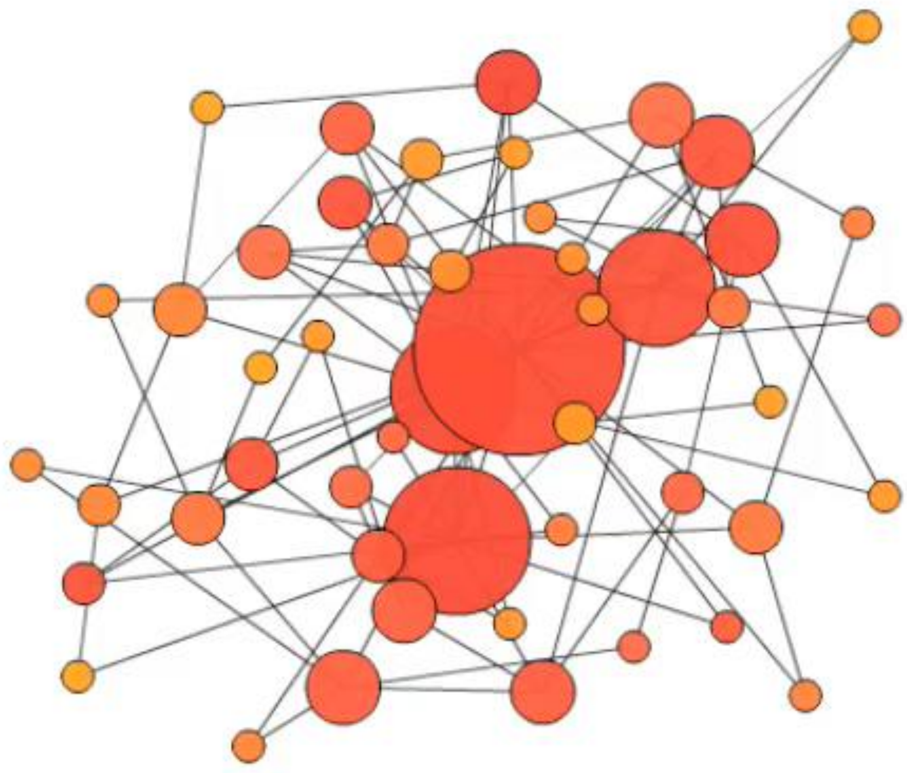
Barabasi&Albert.
Science 1999.

Why should we care about the shape of our social networks?

The structure of networks impacts its **resilience** to attack, the capacity to spread ideas and information, and other dynamical processes

Network resilience

Scale-free networks are robust against random attacks



Networksciencebook.com

23

Network resilience

Scale-free networks are vulnerable against targeted attacks

Demo: http://networksciencebook.com/images/ch-08/video-8-2.webm

Networksciencebook.com

Part 2

SHAPE OF OUR SOCIAL WORLD: SMALL WORLD

The Bacon Number

www.oracleofbacon.org

The Bacon Number
[Watts 2002]

TABLE 3.1 DISTRUBUTION OF ACTORS ACCORDING TO BACON NUMBER

BACON NUMBER	NUMBER OF ACTORS	CUMULATIVE TOTAL NUMBER OF ACTORS
0	1	1
1	1,550	1,551
2	121,661	123,212
3	310,365	433,577
4	71,516	504,733
5	5,314	510,047
6	652	510,699
7	90	510,789
8	38	510,827
9	1	510,828
10	1	510,829

The Erdös Number

Paul Edrös was a famous Hungarian Mathematician, 1913-1996.

Erdös posed and solved problems in number theory and other areas and founded the field of discrete mathematics.

- 511 co-authors (Erdös number 1)
- ~ 1500 Publications

The Erdös Number

The Erdös Number:
Through how many research collaboration links is an arbitrary scientist connected to Paul Erdös?

More generally, how many handshakes away are two researchers? There is a link if they write a paper together (co-authorship link)

What is Erdös Number of your favorite professor to Tim Berners-Lee? https://www.csauthors.net/distance/

me -> -> P. Erdös ?

Stanley Milgram coined the term Small World

- A social psychologist
- Yale and Harvard University

- Study on the Small World Problem,
  beyond well defined communities
  and relations
  (such as actors, scientists, …)

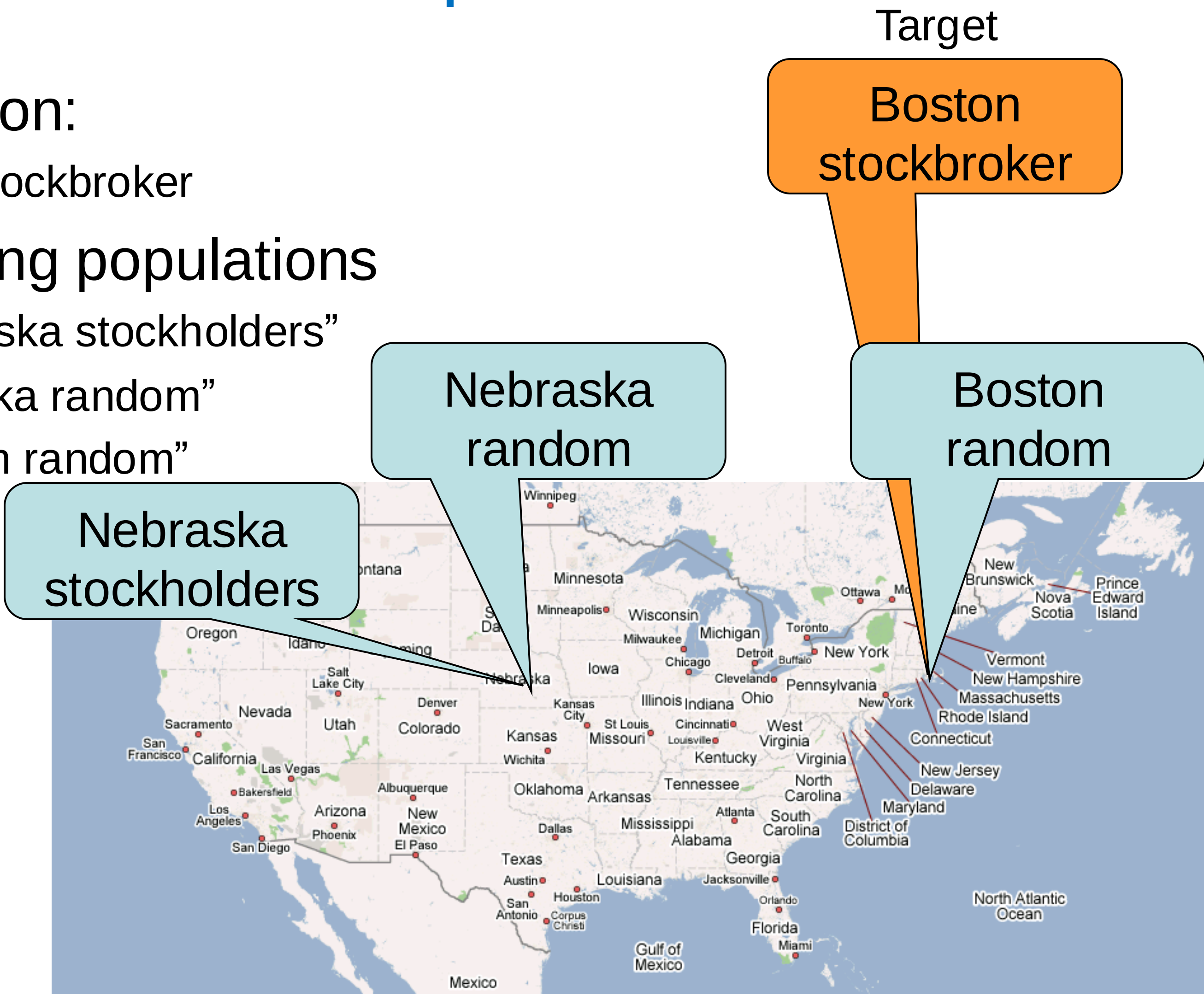
1933-1984

- Milgram is famous for the obedience Study

- What we will discuss today:
  „An Experimental Study of the Small World Problem"

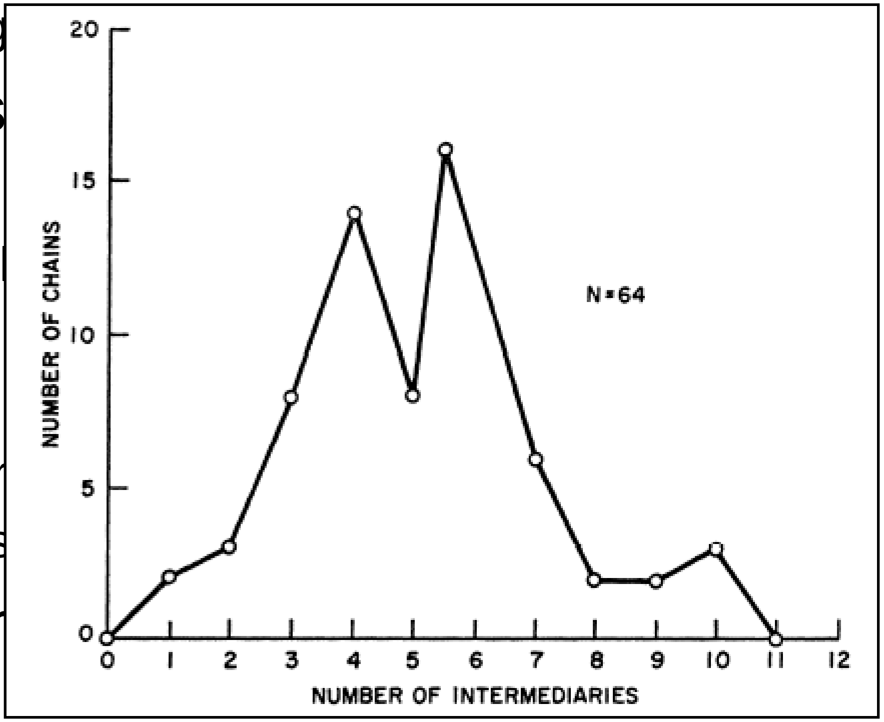
Set Up

- Target person:
  - A Boston stockbroker
- Three starting populations
  - 100 “Nebraska stockholders”
  - 96 “Nebraska random”
  - 100 “Boston random”



Results I

- How many of the starters would be able to establish contact with the target?
  - 64 out of 296 reached the targ
- How many intermediaries
  starters with the target?
  - Well, that depends: the overall
  - Through hometown: 6.1 links
  - Through business: 4.6 links
  - Boston group faster than Nebr
  - Nebraska stockholders not fas
- What form would the distr
  take?



What does it mean?

- Imagine you know on average 100 people very well with their first name. everyone on average know 100 persons very well. How many people are you connected to with 3 handshakes?

Small Worlds

http://www.infosci.cornell.edu/courses/info204/2007sp/

- Every pair of nodes in a graph is connected by a path with an extremely small number of steps
  Social networks have a low diameter.

- Question 1: Why social networks are small worlds? What **social mechanisms** create such small-world phenomena?

Small Worlds

http://www.infosci.cornell.edu/courses/info204/2007sp/

- Every pair of nodes in a graph is connected by a path with an extremely small number of steps
(low diameter)

- Question 1: Why social networks are small worlds? What **social mechanisms** create such small world phenomena?
- Question 2: **Why is it important** to understand the Small World properties of our social networks? What does it mean?

Formalizing the Small World Problem

[Watts and Strogatz 1998]

The small-world phenomenon is assumed to be present when local clustering is high and average path lengths are relatively small

average path length: *L*

clustering coefficient: *C*

Examples for Small World Networks

[Watts and Strogatz 1998]

Table 1 Empirical examples of small-world networks

$L > L_{random}$ but $C >> C_{random}$	L_{actual}	L_{random}	C_{actual}	C_{random}
Film actors	3.65	2.99	0.79	0.00027
Power grid	18.7	12.4	0.080	0.005
C. elegans	2.65	2.25	0.28	0.05

Characteristic path length L and clustering coefficient C for three real networks, compared to random graphs with the same number of vertices (n) and average number of edges per vertex (k). (Actors: $n = 225{,}226$, $k = 61$. Power grid: $n = 4{,}941$, $k = 2.67$. *C. elegans*: $n = 282$, $k = 14$.) The graphs are defined as follows. Two actors are joined by an edge if they have acted in a film together. We restrict attention to the giant connected component[16] of this graph, which includes ~90% of all actors listed in the Internet Movie Database (available at http://us.imdb.com), as of April 1997. For the power grid, vertices represent generators, transformers and substations, and edges represent high-voltage transmission lines between them. For *C. elegans*, an edge joins two neurons if they are connected by either a synapse or a gap junction. We treat all edges as undirected and unweighted, and all vertices as identical, recognizing that these are crude approximations. All three networks show the small-world phenomenon: $L \gtrsim L_{random}$ but $C \gg C_{random}$.

Formalizing the Small World Problem

[Watts and Strogatz 1998]

The small-world phenomenon is assumed to be present when

$L \gtrsim L_{\text{random}}$ but $C \gg C_{\text{random}}$

Or in other words: We are looking for networks where local clustering is high and global path lengths are small

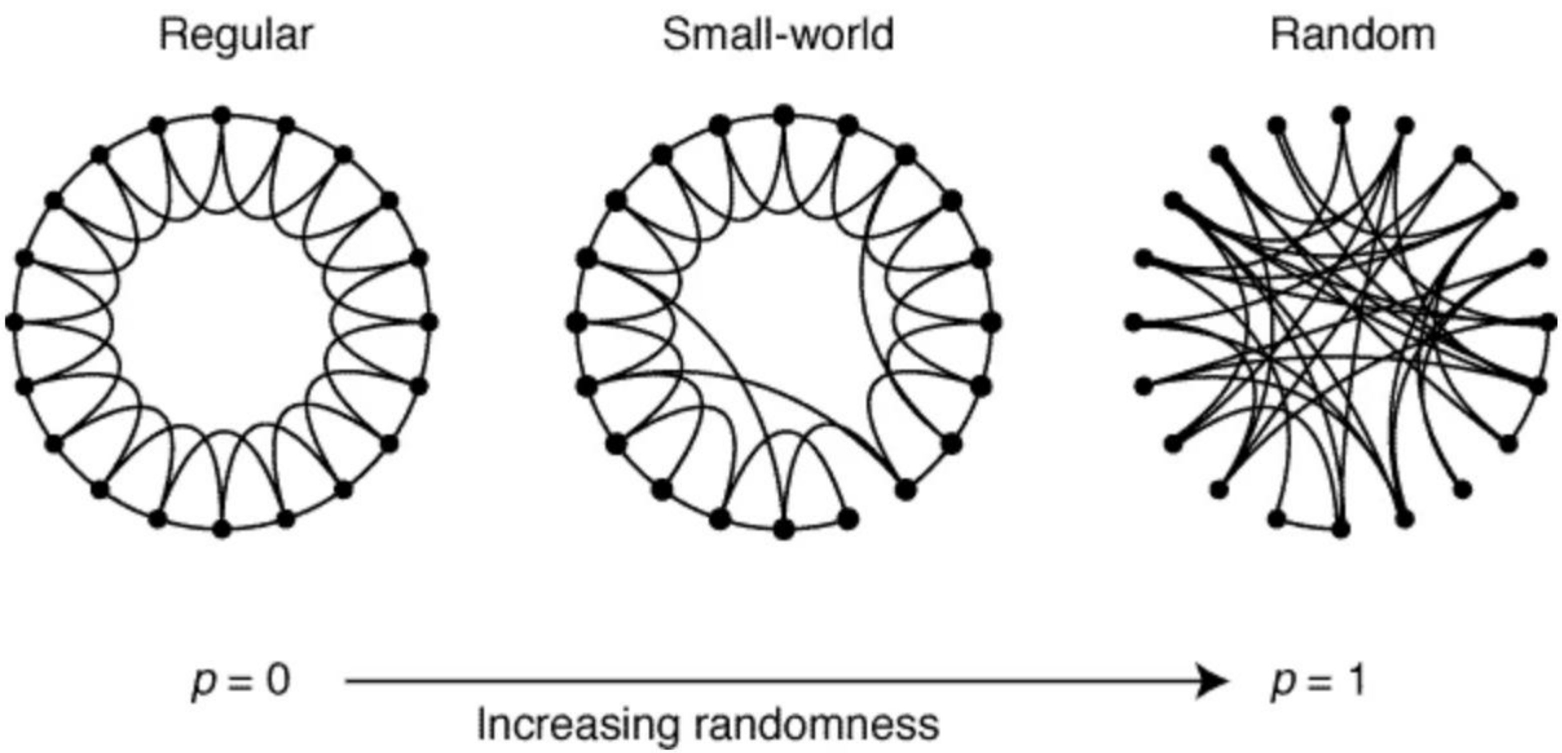
What's the rationale for the above formalism?

One potential answer:

Cavemen and Solaris Worlds

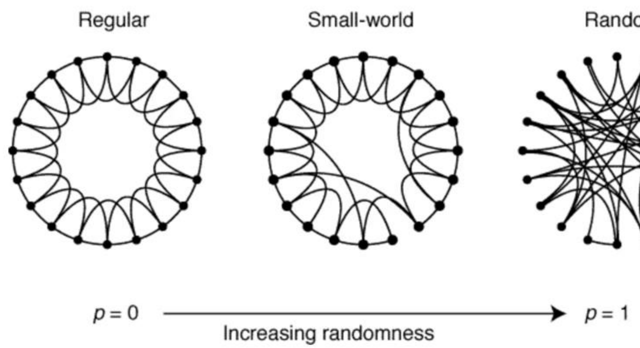
Formalizing the Small World Problem

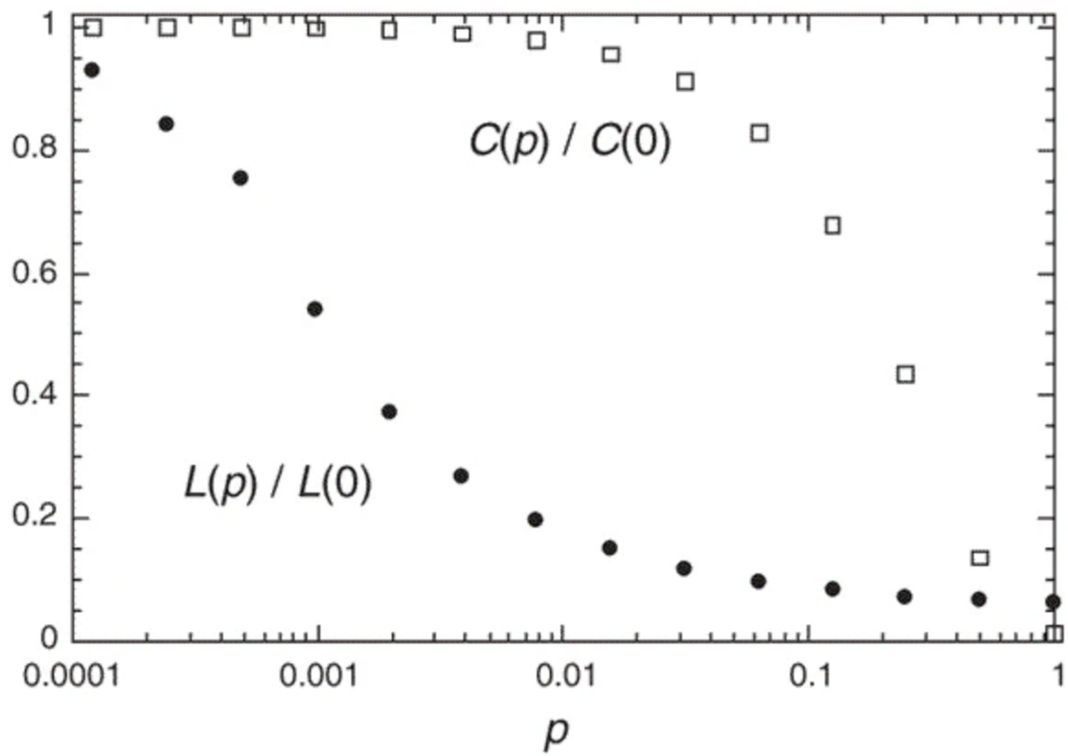
[Watts and Strogatz, Nature 1998]



Formalizing the Small World Problem

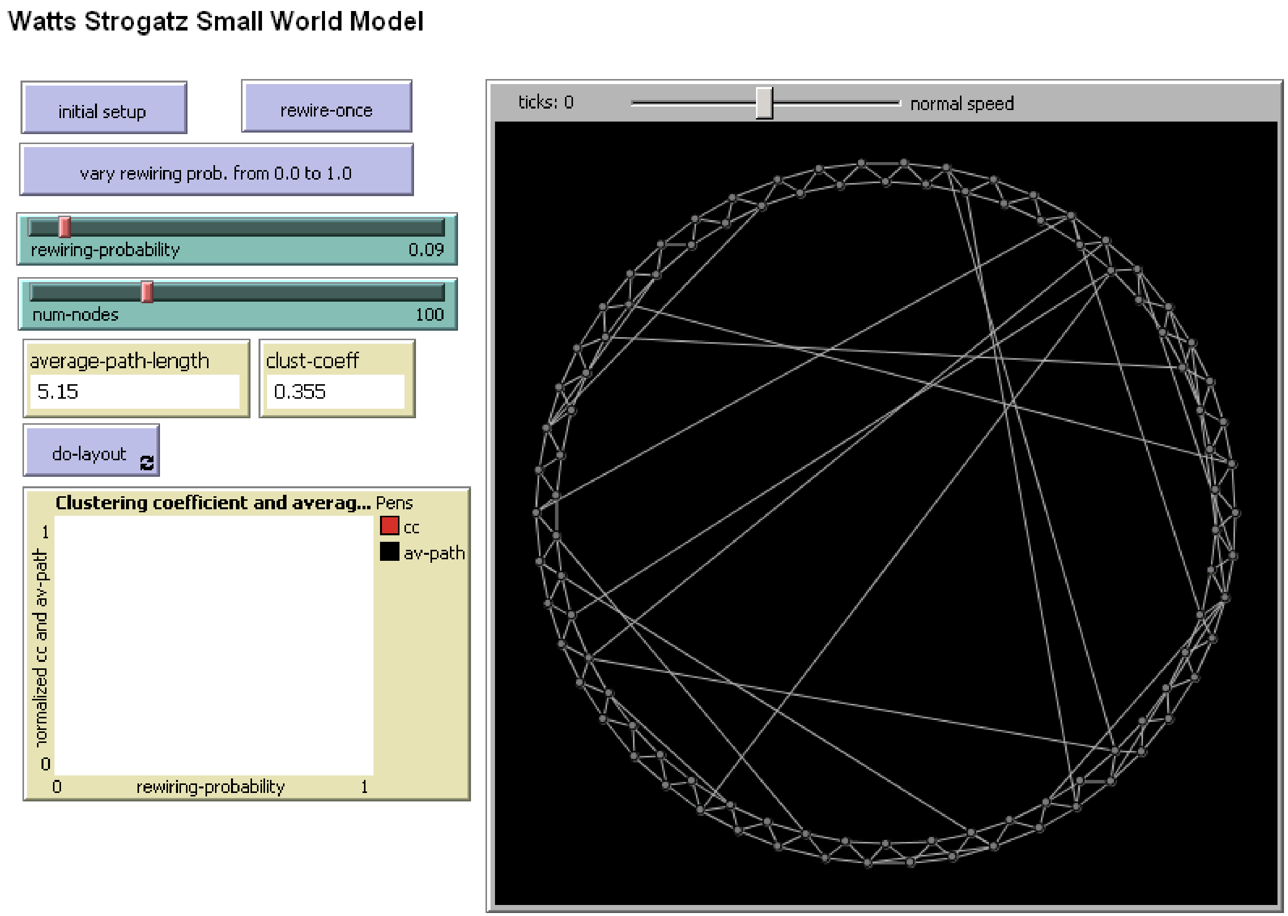
[Watts and Strogatz, Nature 1998]





Demo – Small Worlds

http://www.netlogoweb.org/launch



Contemporary Software

- Where does the small-world phenomenon come into play in contemporary software, in organizations, ..?

- Xing, LinkedIn, Facebook, …
- Business Processes, Information and Knowledge Flow

Part 3

SHAPE OF OUR SOCIAL WORLD: MANY CLUSTERS

Fundamental Concepts in SNA
[Wassermann and Faust 1994]

- Triad
  - Def: A subgroup of three actors and the possible ties among them

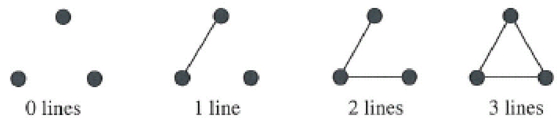


Fig. 4.3. Four possible triadic states in a graph

  - **Transitivity**
    - If actor i „likes“ j, and j „likes“ k, then i also „likes“ k
  - **Balance**
    - If actor i and j like each other, they should be similar in their evaluation of some k
    - If actor i and j dislike each other, they should evaluate k differently

Fundamental Concepts in SNA

[Wassermann and Faust 1994]

- Triad
  - Def: A subgroup of three actors and the possible ties among them

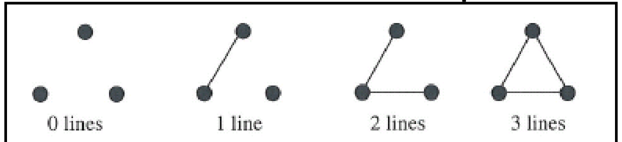
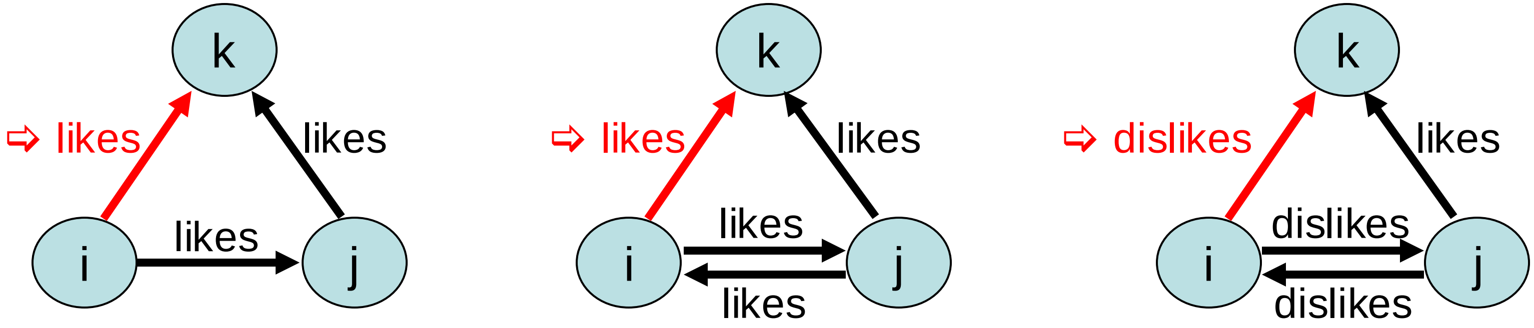


Fig. 4.3. Four possible triadic states in a graph

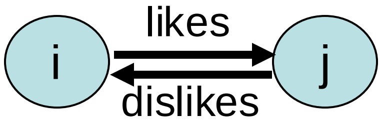
  - **Transitivity**
    - If actor i „likes“ j, and j „likes“ k, then i also „likes“ k
  - **Balance**
    - If actor i and j like each other, they should be similar in their evaluation of some k
    - If actor i and j dislike each other, they should evaluate k differently



Example 1: Transitivity
Example 2: Balance
Example 3: Balance

Fundamental Concepts in SNA
[Wassermann and Faust 1994]

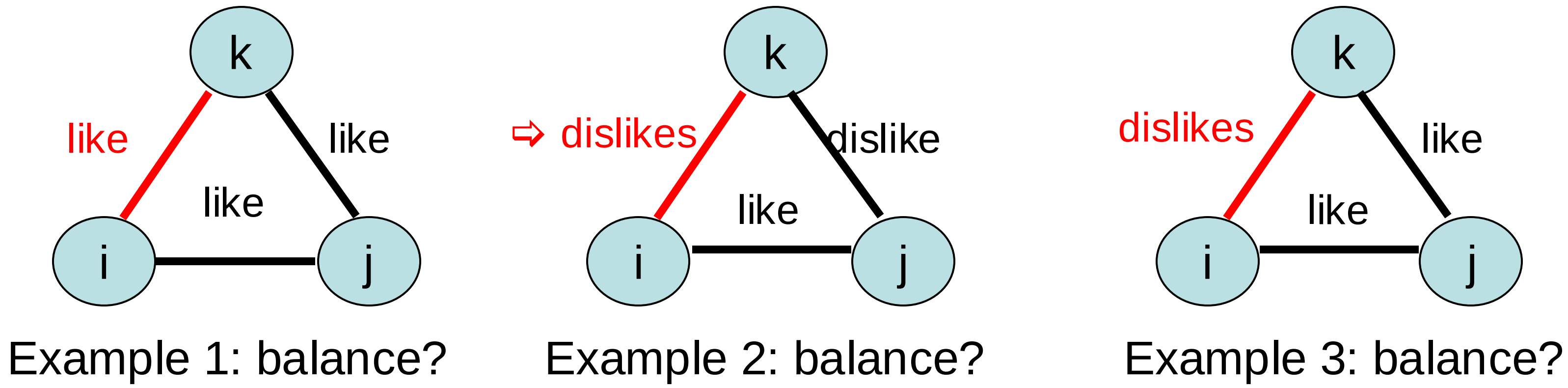
– **Maintaining social ties**



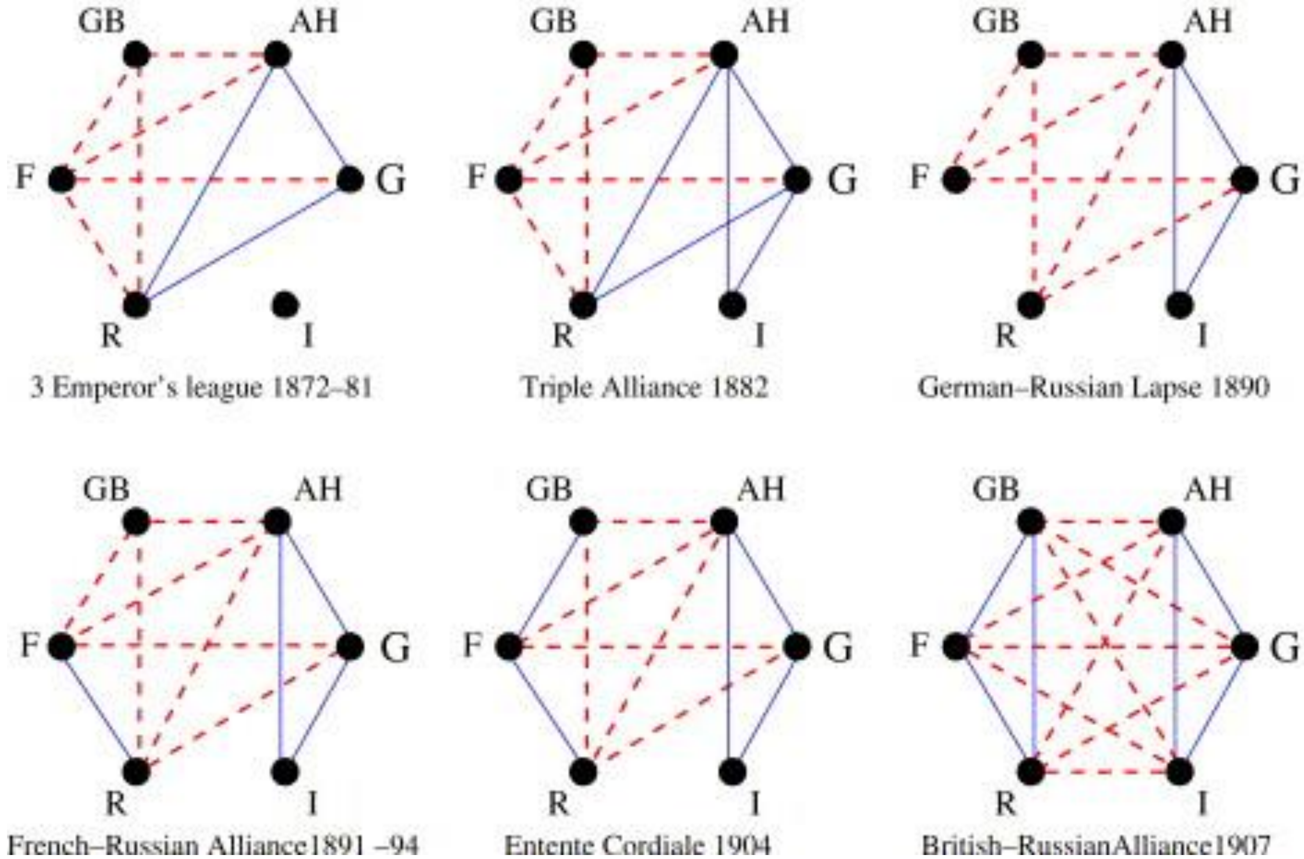
What happens to this relationship?

Fundamental Concepts in SNA
[Wassermann and Faust 1994]

- **What happens to these relationships?**
- **Balance theory**
  - A social triangle is balanced when the cognitive capacity to maintain it is low



Balance theory and conflict



Evolution of the major relationship changes between the protagonists of World War I from 1872–1907. Here GB=Great Britain, AH=Austria–Hungary, G=Germany, I=Italy, R=Russia, and F=France. Antal et al, Physica D (2006).

Part 4

SHAPE OF OUR SOCIAL NETWORKS: ASSORTATIVE

Assortative Mixing
[Newman 2003]

Assortative Mixing refers to selective linking of nodes to other nodes who share some common property

- Topological example: Degree assortativity
  high degree nodes in a network associate preferentially with other high-degree nodes
- Social example: Homophily
  nodes of a certain type tend to associate with the same type of nodes (e.g. by race)

Homophily and friendship in high school [Moody 2001]

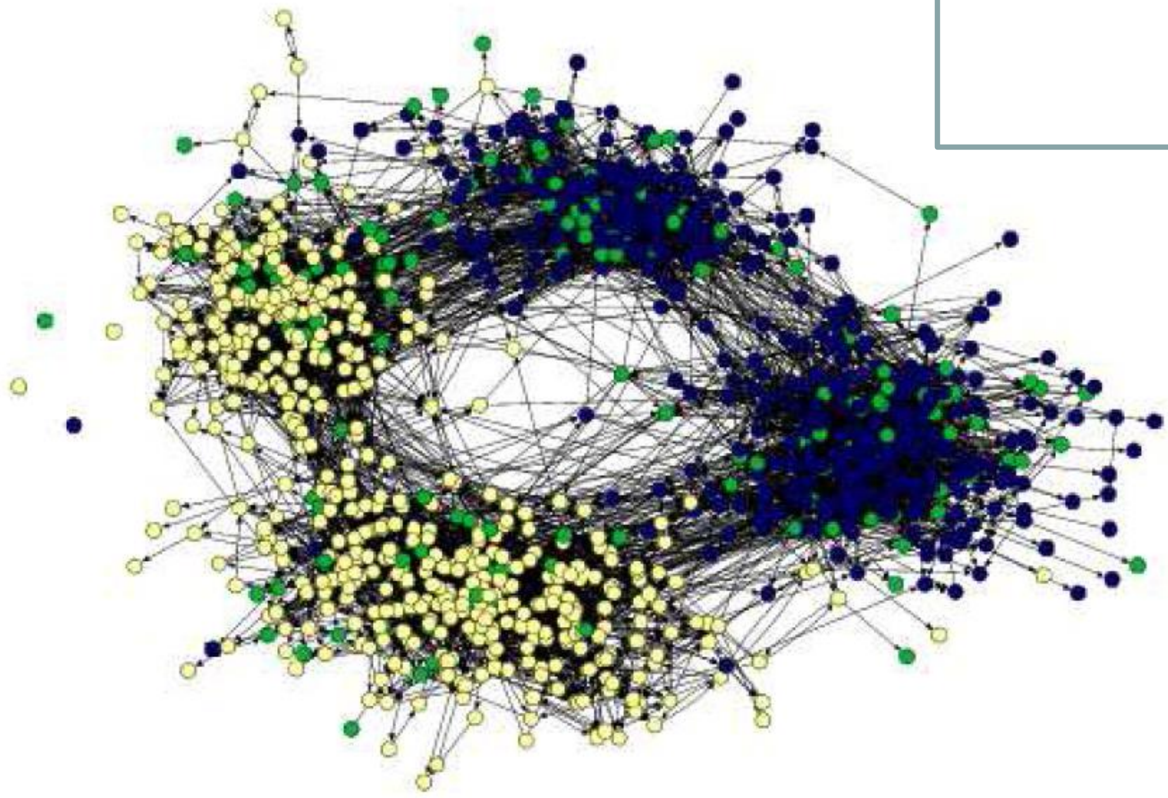


FIG. 8 Friendship network of children in a US school. Friendships are determined by asking the participants, and hence are directed, since A may say that B is their friend but not *vice versa*. Vertices are color coded according to race, as marked, and the split from left to right in the figure is clearly primarily along lines of race. The split from top to bottom is between middle school and high school, i.e., between younger and older children. Picture courtesy of James Moody.

Why undrestanding assortativity/homophily is important in social networks?

- Degree assortativity can affect the spreading dynamic
- Homophily affects the adoption of norms and culture

Part 5

MECHANISTIC MODELS: BA-HOMOPHILY MODEL

EXAMPLE

Why modeling networks?

Why simple rules in modeling?

Occam's razor

is the problem-solving principle that recommends searching for explanations constructed with the smallest possible set of elements. It is also known as the **principle of parsimony**.

Mechanistic models of Networks

Quantitative social science is not only about regression analysis or, in general, data inference. Computer simulations of social mechanisms have an over 60 years long history. They have been used for many different purposes—to test scenarios, to test the consistency of descriptive theories (proof-of-concept models), to explore emergent phenomena, for forecasting, etc… (Holme and Liljeros. "Mechanistic models in computational social science." *Frontiers in Physics* (2015))

Mechanistic models of Networks

Quantitative social science is not only about regression analysis or, in general, data inference. Computer simulations of social mechanisms have an over 60 years long history. They have been used for many different purposes—to test scenarios, to test the consistency of descriptive theories (proof-of-concept models), to explore emergent phenomena, for forecasting, etc… (Holme and Liljeros. "Mechanistic models in computational social science." *Frontiers in Physics* (2015))

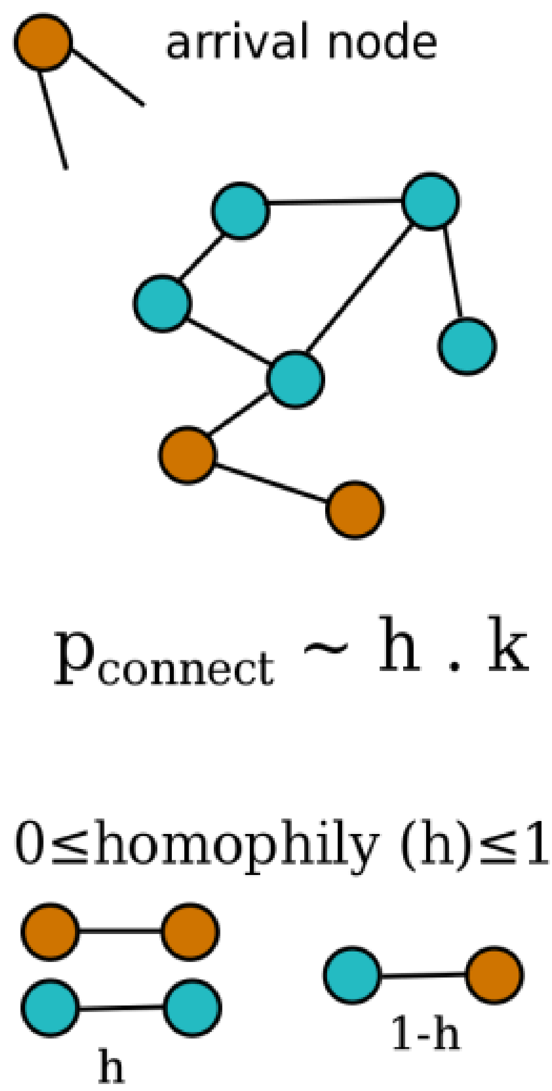
Examples of Mechanistic models in social networks:
Preferential attachment models, fitness model, homophily models are mechanistic models in which they use mechanisms to generate networks in order to:

- Understand causality,
- micro to macro behaviour,
- analytically tractable

Network growth model with Preferential Attachment (BA) and tunable homophily

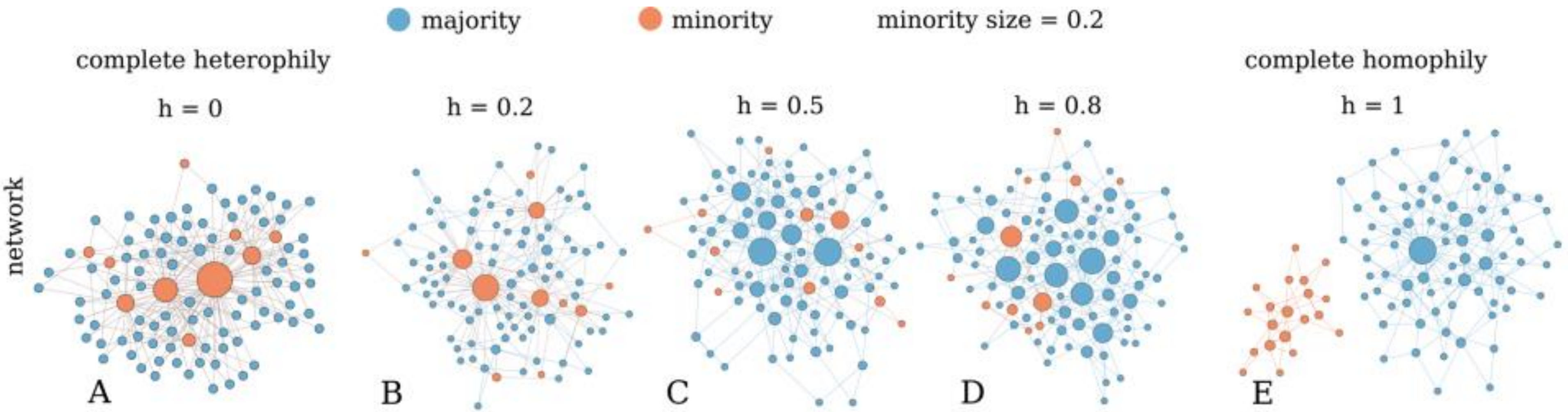
- 2 group of nodes with unequal size

- Arrival node connects to existing nodes based on preferential attachment homophily

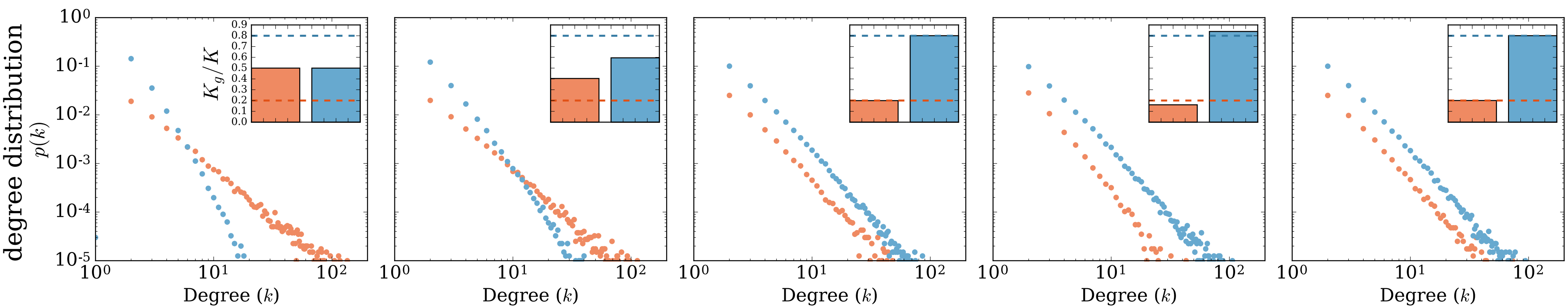


Karimi et al, Scientific Reports (2018)

BA-Homophily network model
Karimi (2018)



Degree distribution



Part 6

SOCIAL TIES AND THEIR MEANINGS

In reality …

Isn‘t all of this an over simplification of the world of social systems?

- Ties/relationships vary in intensity
- People who have strong ties tend to share a similiar set of acquaintances
- Ties change over time
- Nodes (people) have different characteristics, and they are *actors*
- *…*

The Strength of Weak Ties [Granovetter 1973]

The strength of an interpersonal tie is a

- (probably linear) combination of the amount of time
- The emotional intensity
- The intimacy
- The reciprocal services which characterize the tie

Mark Granovetter,

Stanford University

Can you give examples of strong / weak ties?

The Strength of Weak Ties and Mutual Acquaintances [Granovetter 1973]

Consider:

Two arbitrarily selected individuals A and B and

The set S = [C,D,E, …] of all persons with ties to either or both of them

Hypothesis:

The stronger the tie between A and B, what happens to the set S?

The Strength of Weak Ties and Mutual Acquaintances [Granovetter 1973]

Consider:

Two arbitrarily selected individuals A and B and

The set S = [C,D,E, …] of all persons with ties to either or both of them

Hypothesis:

The stronger the tie between A and B, the larger the proportion of individuals in S to whom they will both be tied.

Theoretical corroboration:

Stronger ties involve larger time commitments – probability of B meeting with some friend of A (who B does not know yet) is increased

The stronger a tie connecting two individuals, the more similar they are

The Strength of Weak Ties [Granovetter 1973]

The forbidden triad

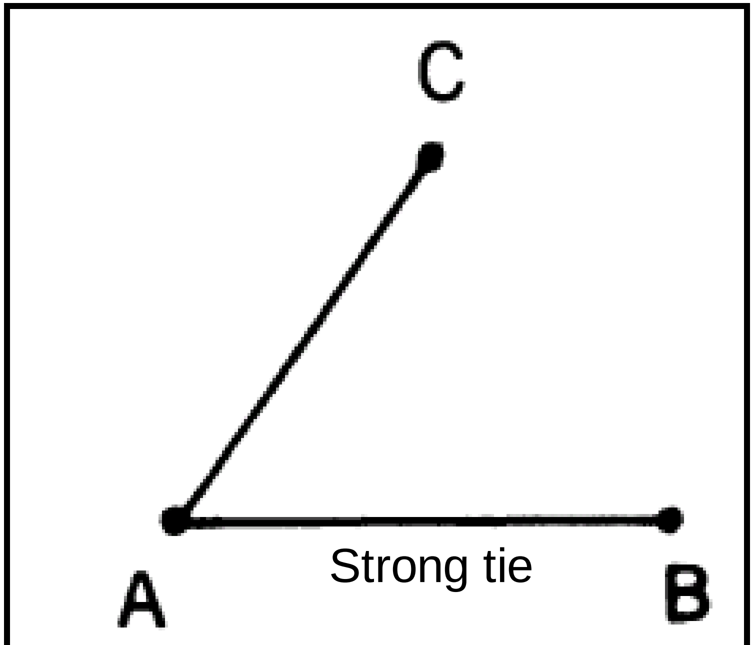


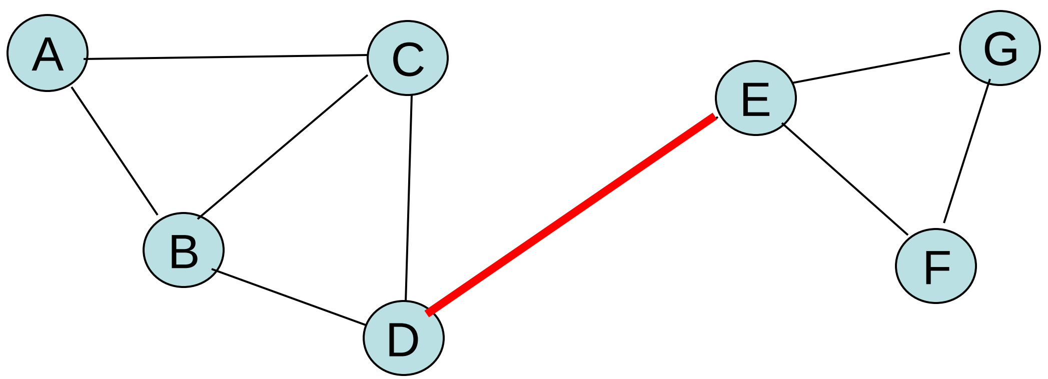
FIG. 1.—Forbidden triad

Why is it called the forbidden triad?

Bridges
[Granovetter 1973]

A bridge is a line in a network which provides **the only path** between two points.

In social networks, a bridge between A and B provides the only route along which information or influence can flow from any contact of A to any contact of B



Which edge represents a bridge? Why?

Bridges and Strong Ties [Granovetter 1973]

Example:

1. Imagine the strong tie between A and B
2. Imagine the strong tie between A and C
3. Then, the forbidden triad **implies** that a tie **exists** between C and B (it forbids that a tie between C and B does not exist)
1. From that follows, that A-B is not a bridge (because there is another path A-B that goes through C)

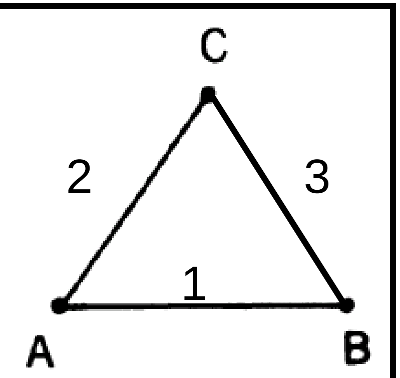


FIG. 1.—Forbidden triad

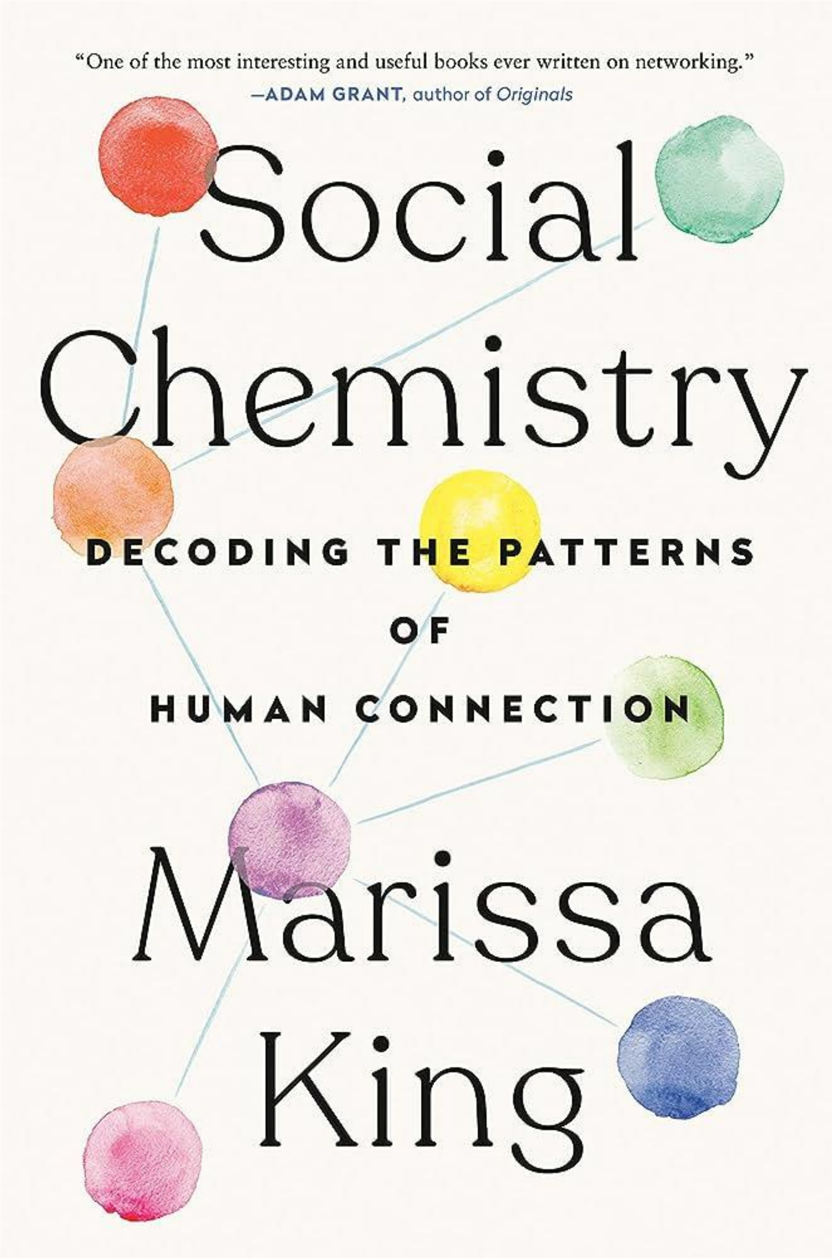
Why is this interesting?

⇨Strong ties can be a bridge ONLY IF neither party to it has any other strong ties

⇨Highly unlikely in a social network of any size

⇨Weak ties suffer no such restriction, though they are not automatically bridges

⇨But, all bridges are weak ties


"One of the most interesting and useful books ever written on networking."
—ADAM GRANT, author of Originals
Social Chemistry
DECODING THE PATTERNS OF HUMAN CONNECTION
Marissa King

In Reality ….
[Granovetter 1973]

it probably happens only rarely, that a specific tie provides the only path between two points

Local bridges: the shortest path between its two points (other than itself)

- Bridges are efficient paths
- Alternatives are more costly
- Local bridges of degree n
- A local bridge is more significant as its degree increases

What‘s the degree of a bridge in an absolute sense?

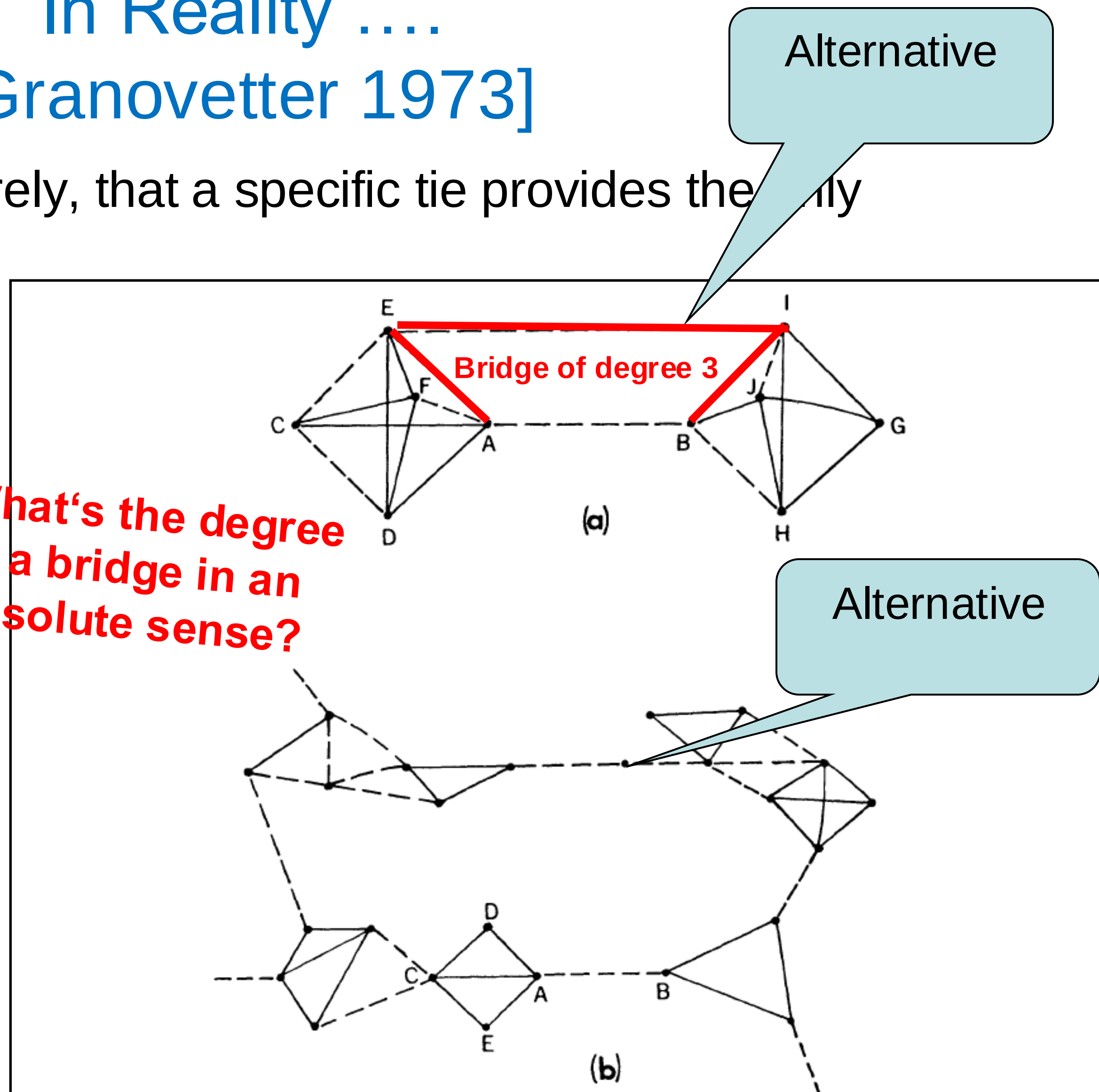


FIG. 2.—Local bridges. *a*, Degree 3; *b*, Degree 13. ——— = strong tie; — — — = weak tie.

In Reality …

Strong ties can represent *local* bridges BUT
They are weak (i.e. they have a low degree)

Why?

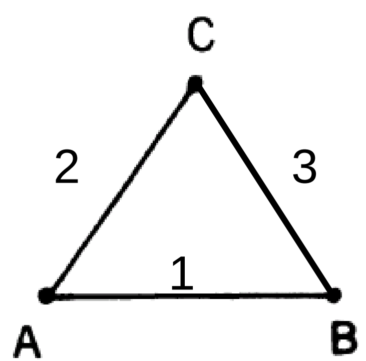


FIG. 1.—Forbidden triad

What's the degree of the local bridge A-B?

Implications of Weak Ties [Granovetter 1973]

- Those weak ties, that are local bridges, create more, and shorter paths.
- The removal of the average weak tie would do more damage to transmission probabilities than would that of the average strong one
- **Paradox**: While *weak ties* have been denounced as generative of alienation, *strong ties*, breeding local cohesion, lead to overall fragmentation

What are sources of weak ties/bridges?

Can you identify some implications for social networks on the web / for search in these networks?

How does this relate to Milgram's experiment?

Completion rates in Milgram's experiment were reported higher for acquaintance than friend relationships [Granovetter 1973]

Implications of Weak Ties [Granovetter 1973]

- Example: Spread of information/rumors in social networks
  - Studies have shown that people rarely act on mass-media information unless it is also transmitted through personal ties [Granovetter 2003, p 1274]
  - Information/rumors moving through strong ties is much more likely to be limited to a few cliques than that going via weak ones, bridges will not be crossed

How does information spread through weak ties?

Any questions?

See you next week!