Network Metrics

CE642: Social and Economic Networks
Maryam Ramezani
Sharif University of Technology
maryam.ramezani@sharif.edu

Introduction

- How can we characterize, model, and reason about the structure of social networks?
  - Triadic closure and "the strength of weak ties"
  - Power-laws and scale-free networks, "rich-get-richer" phenomena
  - Small-world phenomena
  - Hubs & Authorities; PageRank
  - Models of network structure

Triangles

- So far we've seen (a little about) how networks can be characterized by their connectivity patterns

- What more can we learn by looking at higher-order properties, such as relationships between triplets of nodes?

Motivation

- Q: Last time you found a job, was it through:
  - A complete stranger?
  - A close friend?
  - An acquaintance?

- A: Surprisingly, people often find jobs through acquaintances rather than through close friends (Granovetter, 1973)

Motivation

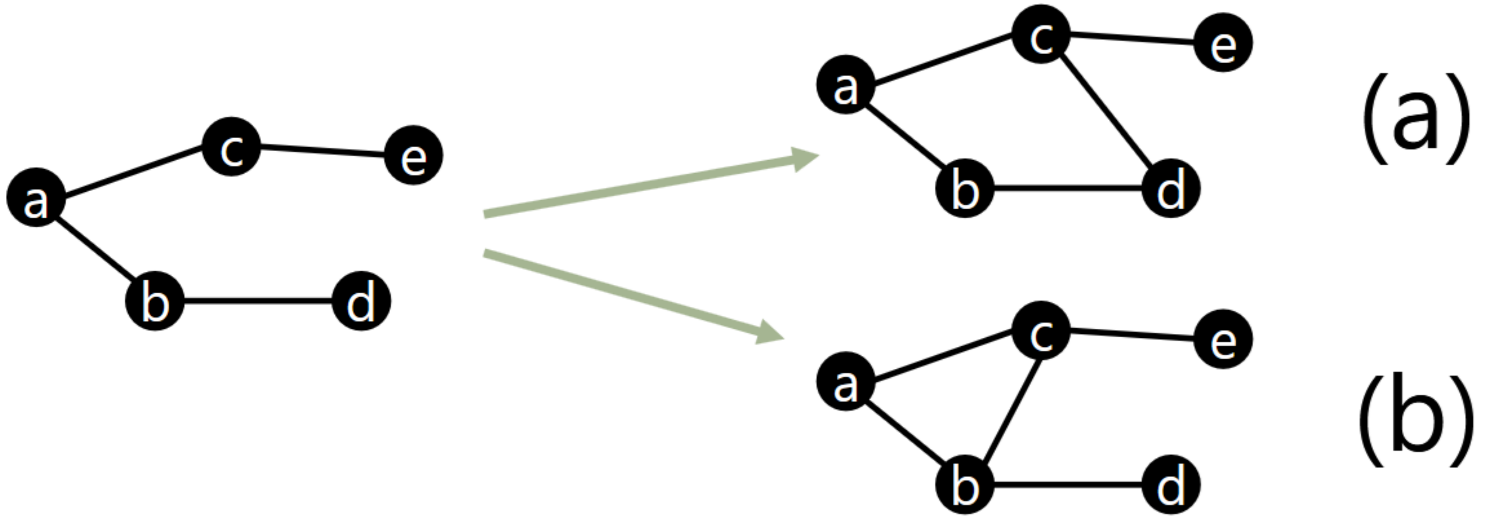
- Your friends (hopefully) would seem to have the greatest motivation to help you.
- But! Your closest friends have limited information that you don’t already know about.
- Alternately, acquaintances act as a “bridge” to a different part of the social network, and expose you to new information.
- This phenomenon is known as the **strength of weak ties.**

Motivation

- To make this concrete, we'd like to come up with some notion of "tie strength" in networks
- To do this, we need to go beyond just looking at edges in isolation, and looking at how an edge connects one part of a network to another

Triadic Closure

- Q: Which edge is most likely to form next in this (social) network?



- A: (b), because it creates a triad in the network

Triangles

- "If two people in a social network have a friend in common, then there is an increased likelihood that they will become friends themselves at some point in the future" (Ropoport, 1953)

- Three reasons (see Easley & Kleinberg):
  - Every mutual friend a between bob and chris gives them an opportunity to meet
  - If bob is friends with ashton, then knowing that chris is friends with ashton gives bob a reason to trust chris
  - If chris and bob don't become friends, this causes stress for ashton (having two friends who don't like each other), so there is an incentive for them to connect

01

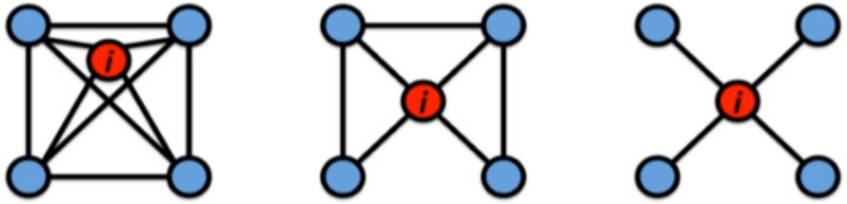
Clustering Coefficient

Clustering Coefficient

- **Clustering coefficient:**
  - What portion of i's neighbors are connected?
  - Node i with degree k_i
  - $C_i \in [0,1]$

    This ranges between 0 (none of my friends are friends with each other) and 1 (all of my friends are friends with each other)
  - $C_i = \frac{2e_i}{k_i(k_i-1)}$

    where e_i is the number of edges between the neighbors of node i



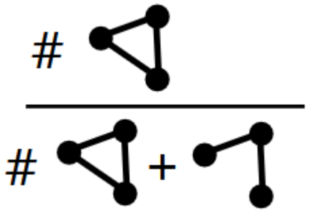
$C_i = 1$ $C_i = 1/2$ $C_i = 0$

- **Average clustering coefficient:** $C = \frac{1}{N}\sum_i^N C_i$

Clustering Coefficient

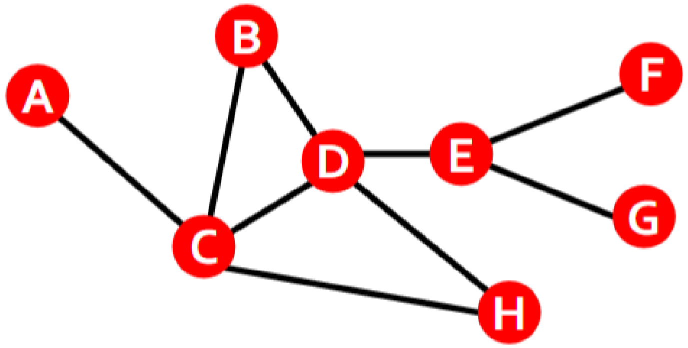
- Alternately it can be defined as the fraction of connected triplets in the graph that are closed (these do not evaluate to the same thing!):

$$C = \frac{\#\text{ of closed triplets}}{\#\text{ of connected triplets}}$$



Clustering Coefficient

- **Clustering coefficient:**
  - What portion of i's neighbors are connected?
  - Node i with degree k_i
  - $$C_i = \frac{2e_i}{k_i(k_i - 1)}$$ where e_i is the number of edges between the neighbors of node i



$k_B=2,\ e_B=1,\ C_B=2/2 = 1$

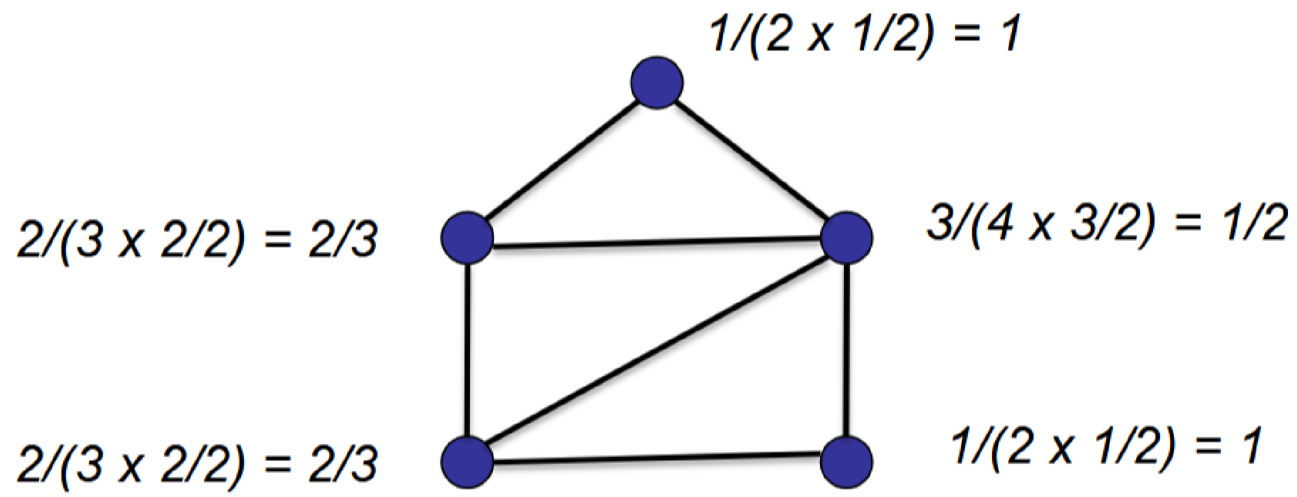
$k_D=4,\ e_D=2,\ C_D=4/12 = 1/3$

Avg. clustering: $C=0.33$

Clustering Coefficient

- Clustering coefficient of a graph G:
  CC(G) = average of c(u) over all vertices u in G
- What Do We Mean By "High" CC?
  - CC(G) measures how likely vertices with a common neighbor are to be neighbors themselves
  - Should be compared to how likely random pairs of vertices are to be neighbors
  - Let p be the edge density of network/graph G: $p = E/(N(N-1)/2)$
    - Here E = total number of edges in G
    - If we picked a pair of vertices at random in G, probability they are connected is exactly p
    - So, we will say clustering is high if CC(G) >> p

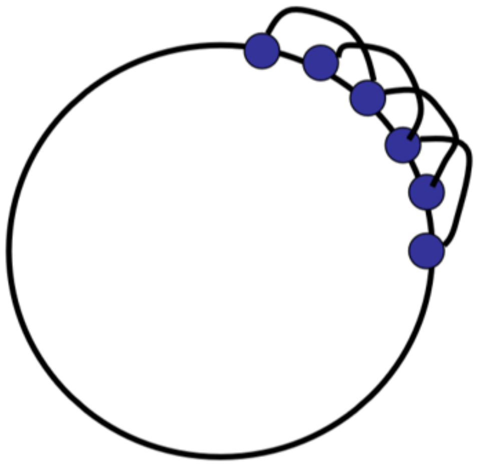
Example



C.C. = (1 + ½ + 1 + 2/3 + 2/3)/5 = 0.7666…
p = 7/(5 x 4/2) = 0.7
Not highly clustered

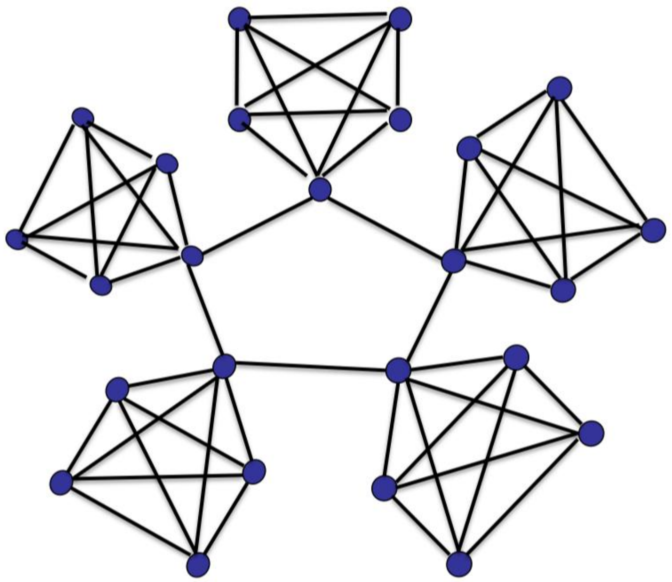
Example

- Network: simple cycle + edges to vertices 2 hops away on cycle
  - By symmetry, all vertices have the same clustering coefficient
  - Clustering coefficient of a vertex v:
    - Degree of v is 4, so the number of *possible* edges between pairs of neighbors of v is 4 x 3/2 = 6
    - How many pairs of v's neighbors actually *are* connected? 3 --- the two clockwise neighbors, the two counterclockwise, and the immediate cycle neighbors
    - So the c.c. of v is 3/6 = ½
  - Compare to overall edge density:
    - Total number of edges = 2N
    - Edge density p = 2N/(N(N-1)/2) ~ 4/N
    - As N becomes large, ½ >> 4/N
    - So this cyclical network is highly clustered



Example

- Divide N vertices into sqrt(N) groups of size sqrt(N) (here N = 25)
- Add all connections within each group (cliques), connect “leaders” in a cycle
- N – sqrt(N) non-leaders
  - CC of network as N becomes large?
  - Edge Density?



Add all connections within each group (cliques), connect “leaders” in a cycle
N – sqrt(N) non-leaders have C.C. = 1, so network C.C. → 1 as N becomes large
Edge density is p ~ 1/sqrt(N)

Research Study

Higher-order structures such as larger cliques are crucial to the structure and function of complex networks

Higher-order clustering in networks

Hao Yin[*]
Institute for Computational and Mathematical Engineering, Stanford University, Stanford, CA, 94305, USA

Austin R. Benson[†]
Department of Computer Science, Cornell University, Ithaca, NY, 14850, USA

Jure Leskovec[‡]
Computer Science Department, Stanford University, Stanford, CA, 94305, USA

(Dated: April 30, 2018)

A fundamental property of complex networks is the tendency for edges to cluster. The extent of the clustering is typically quantified by the clustering coefficient, which is the probability that a length-2 path is closed, i.e., induces a triangle in the network. However, higher-order cliques beyond triangles are crucial to understanding complex networks, and the clustering behavior with respect to such higher-order network structures is not well understood. Here we introduce higher-order clustering coefficients that measure the closure probability of higher-order network cliques and provide a more comprehensive view of how the edges of complex networks cluster. Our higher-order clustering coefficients are a natural generalization of the traditional clustering coefficient. We derive several properties about higher-order clustering coefficients and analyze them under common random graph models. Finally, we use higher-order clustering coefficients to gain new insights into the structure of real-world networks from several domains.

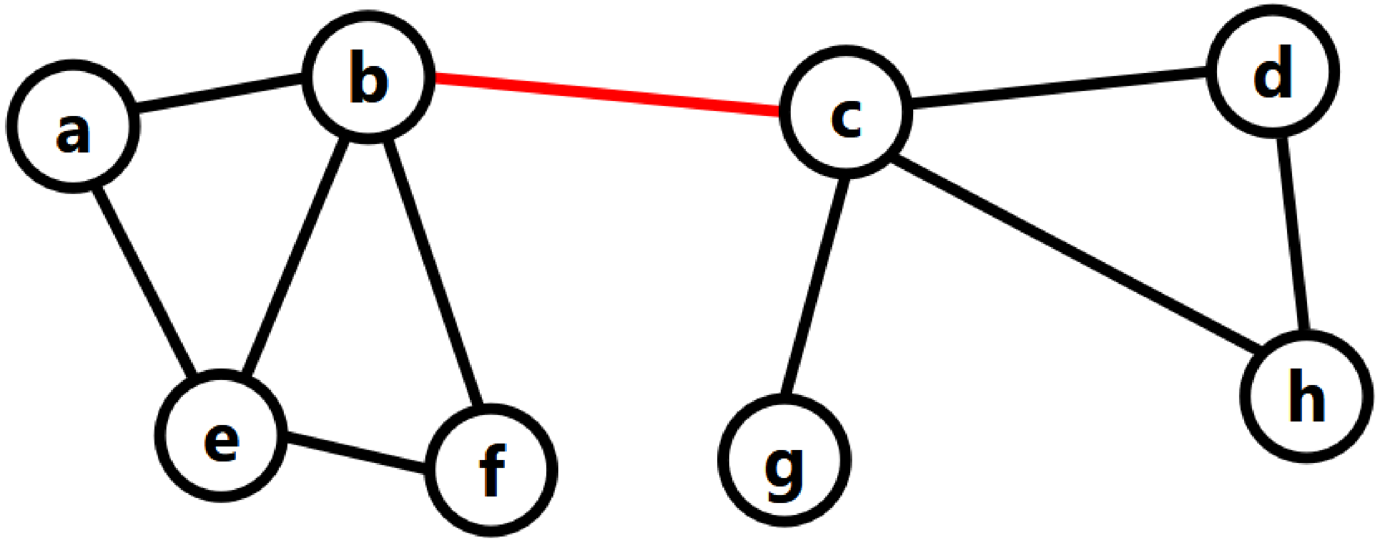
http://snap.stanford.edu/hocc/

02

Bridges

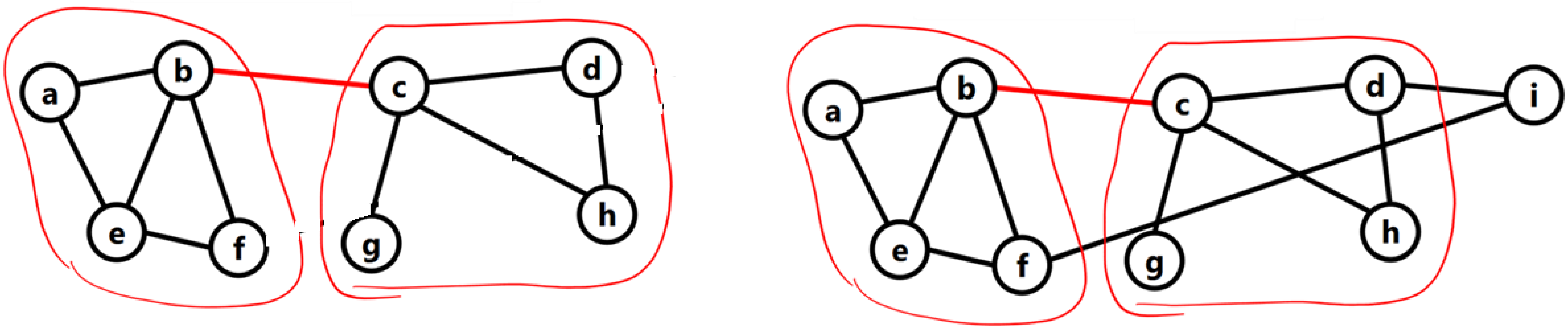
Bridge Edge

- An edge (b,c) is a bridge edge if removing it would leave no path between b and c in the resulting network



Local Bridge Edge

- An edge (b,c) is a local bridge if removing it would leave no edge between b's friends and c's friends (though there could be more distant connections)

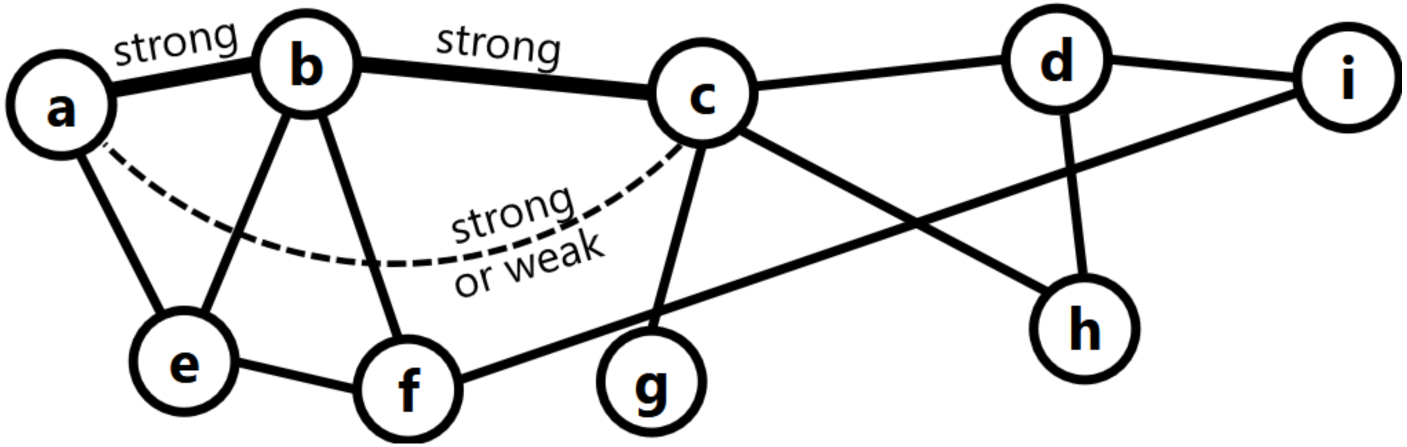


03

Strong & weak ties

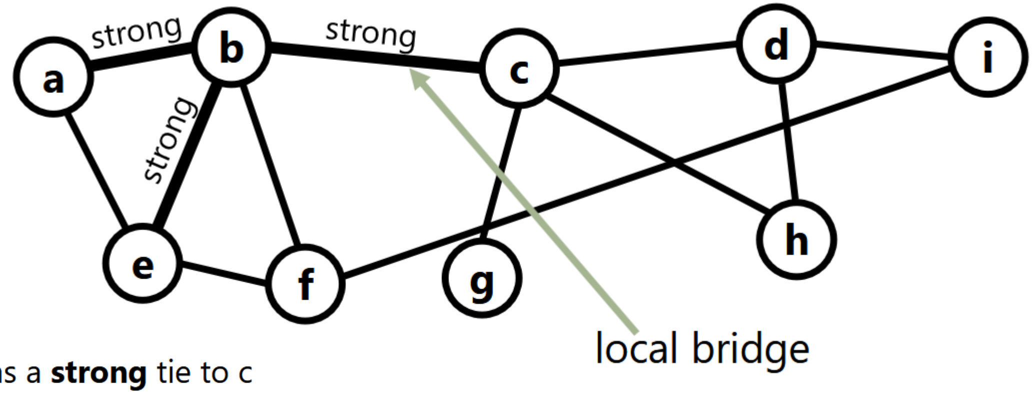
Strong Triadic Closure Property

- If (a,b) and (b,c) are connected by strong ties, there must be at least a weak tie between a and c.
  - Note: (a,c) can be weak or strong!



Granovetter's Theorem

- If the strong triadic closure property is satisfied for a node, and that node is involved in two strong ties, then any incident local bridge must be a weak tie.
  - Proof?



Proof (by contradiction): (1) b has two strong ties (to a and e); (2) suppose it has a **strong** tie to c via a local bridge; (3) but now a tie must exist between c and a (or c and e) due to strong triadic closure; (4) so b → c cannot be a bridge

Granovetter’s Theorem

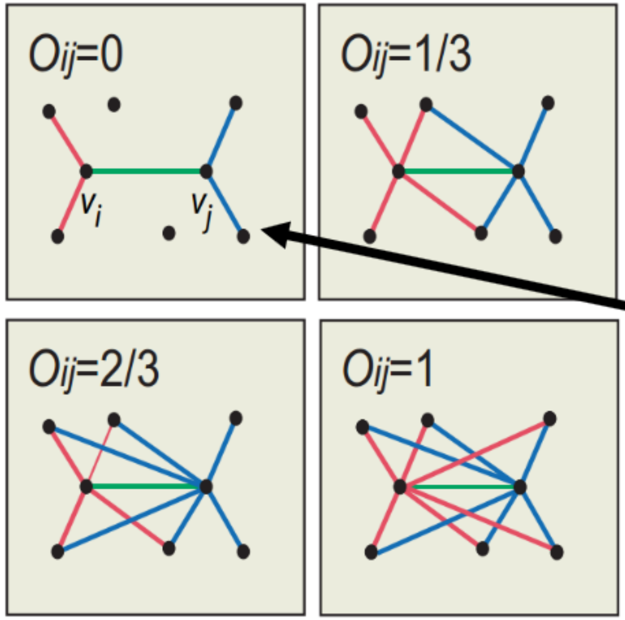
- So, if we’re receiving information from distant parts of the network (i.e., via “local bridges”) then we must be receiving it via weak ties.
- Q: How to test this theorem empirically on real data?
  - A: Onnela et al. 2007 studied networks of mobile phone calls.

Study Example #1

Defn. 1: Define the "overlap" between two nodes to be the Jaccard similarity between their connections

$$O_{i,j} = \frac{\Gamma(i) \cap \Gamma(j)}{\Gamma(i) \cup \Gamma(j)}$$

neighbours of i



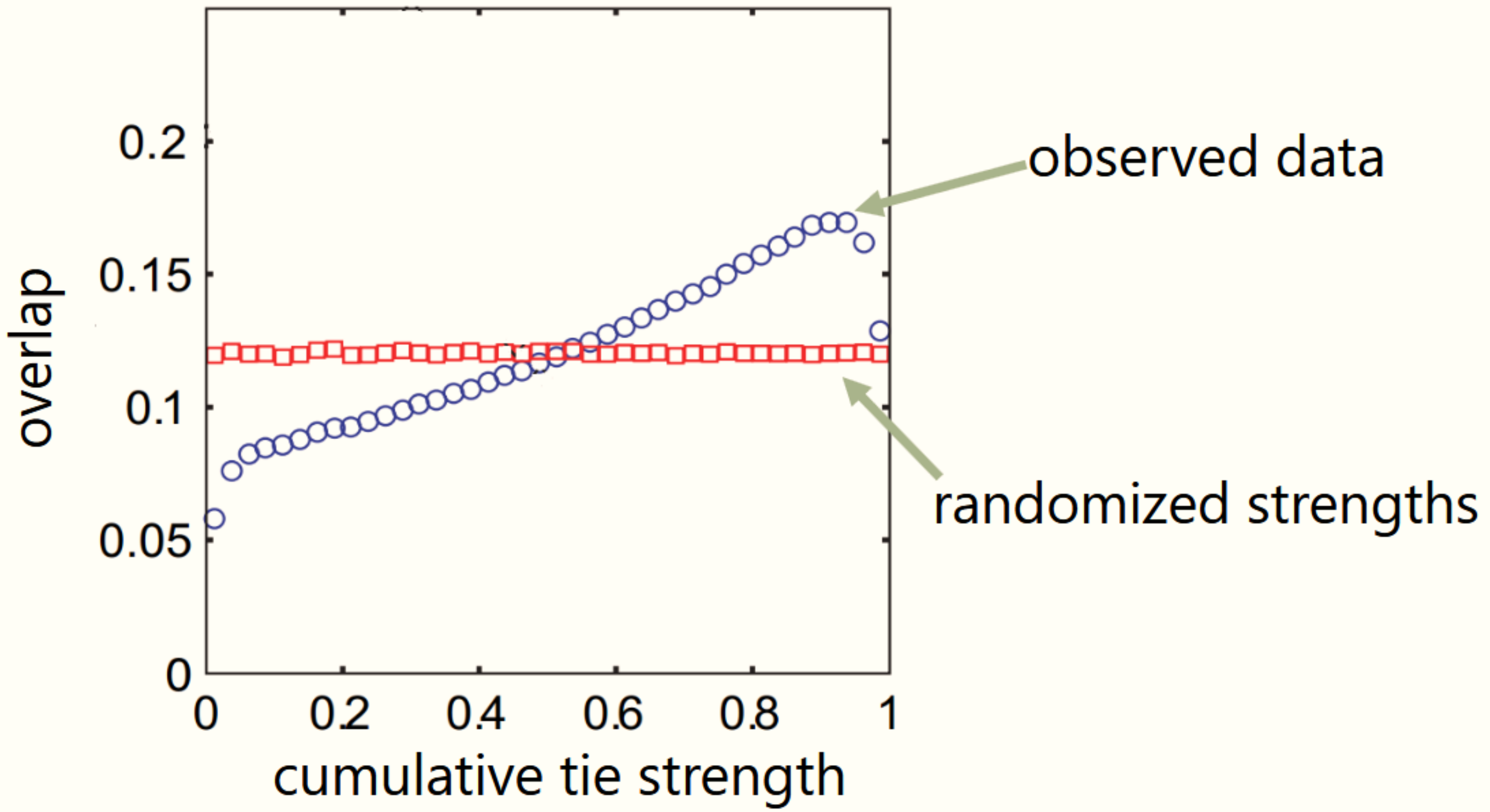
"local bridges" have overlap 0

(picture from Onnela et al., 2007)

Secondly, define the "strength" of a tie in terms of the number of phone calls between i and j

Study Example #1

finding: the “stronger” our tie, the more likely there are to be additional ties between our mutual friends



(picture from Onnela et al., 2007)

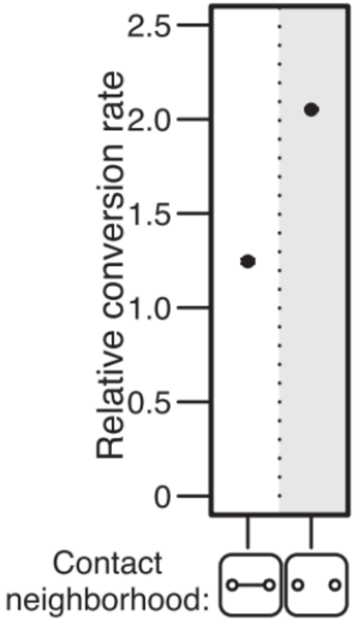
Study Example #2

- Suppose a user receives four e-mail invites to join facebook from users who are already on facebook. Under what conditions are we most likely to accept the invite (and join facebook)?
  - If those four invites are from four close friends?
  - If our invites are from four acquiantances?
  - If the invites are from a combination of friends, acquaintances, work colleagues, and family members?

hypothesis: the invitations are most likely to be adopted if they come from distinct groups of people in the network

Study Example #2

Let's consider the connectivity patterns amongst the people who tried to recruit us

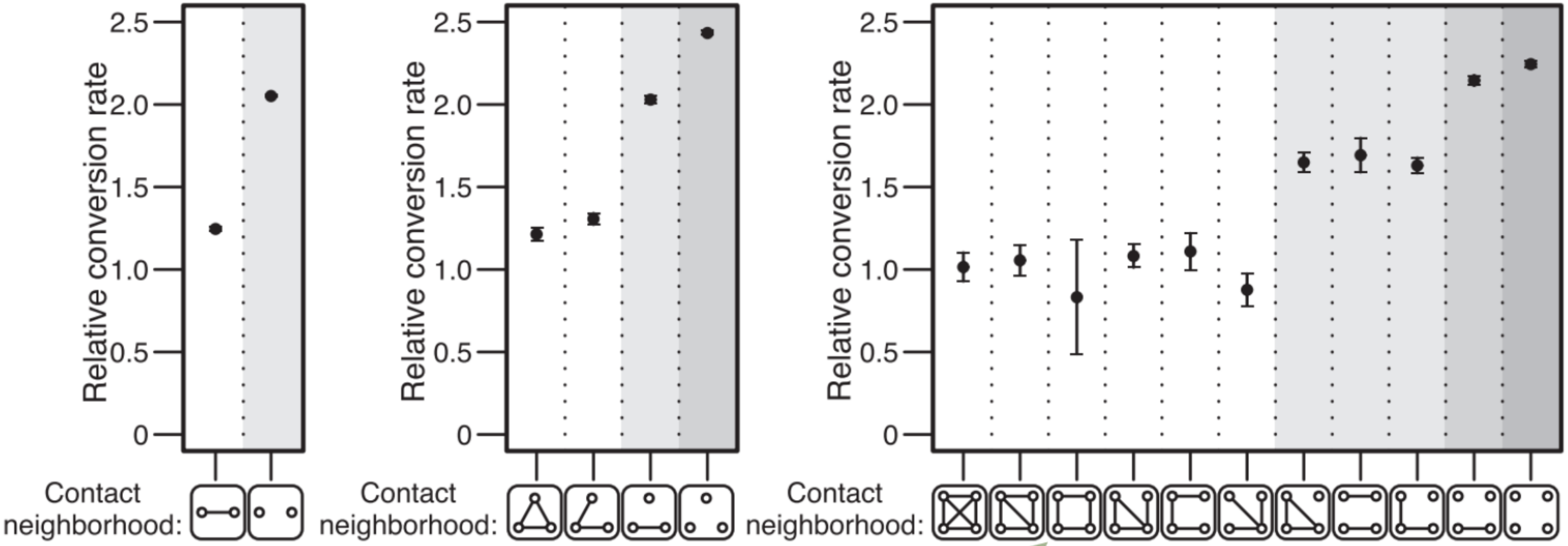


- **Case 1:** two users attempted to recruit
- **y-axis:** relative to recruitment by a single user
- **finding:** recruitments are **more likely to succeed** if they come from friends who are **not connected to each other**

(picture from Ugander et al., 2012)

Study Example #2

Let's consider the connectivity patterns amongst the people who tried to recruit us



error bars are high since this structure is very very rare

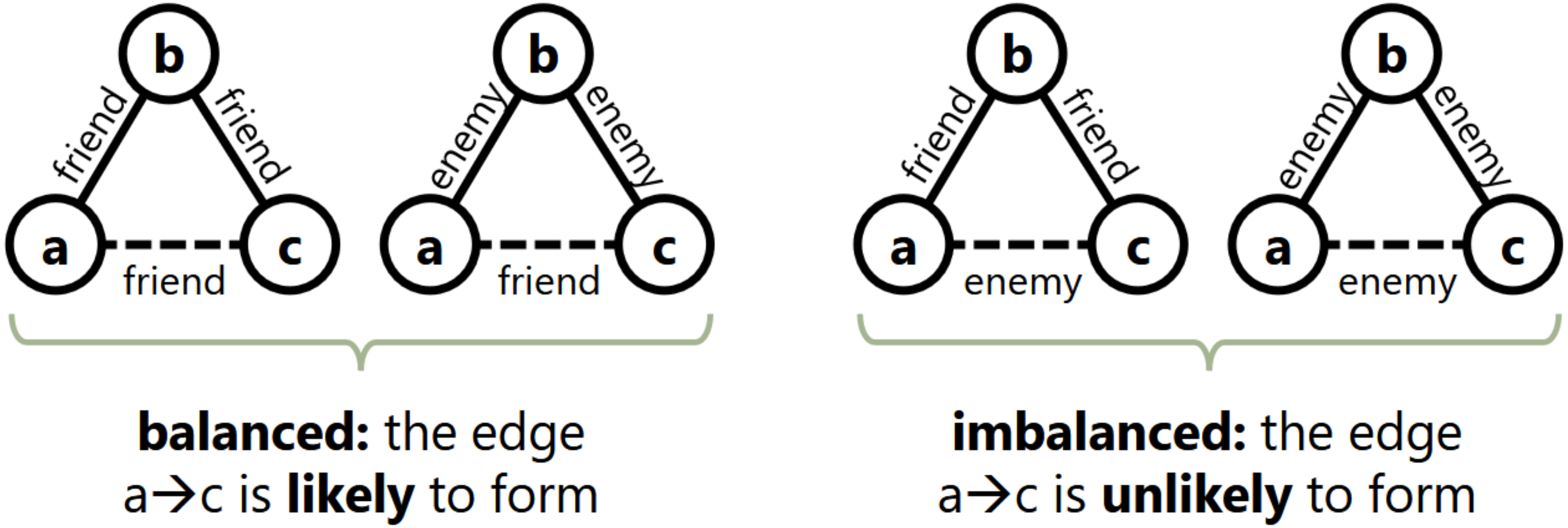
(picture from Ugander et al., 2012)

Conclusion

- Important aspects of network structure can be explained by the way an edge connects two parts of the network to each other:
  - Edges tend to close open triads (clustering coefficient etc.)
  - It can be argued that edges that bridge different parts of the network somehow correspond to “weak” connections (Granovetter; Onnela et al.)
  - Disconnected parts of the networks (or parts connected by local bridges) expose us to distinct sources of information (Granovettor; Ugander et al.)

Structural Balance

- Some of the assumptions that we've seen today may not hold if edges have signs associated with them

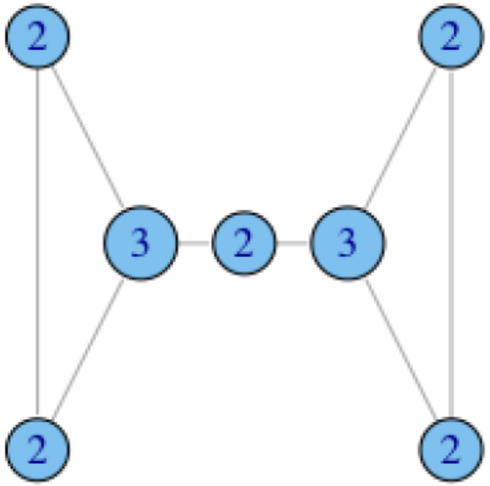


06

Degree

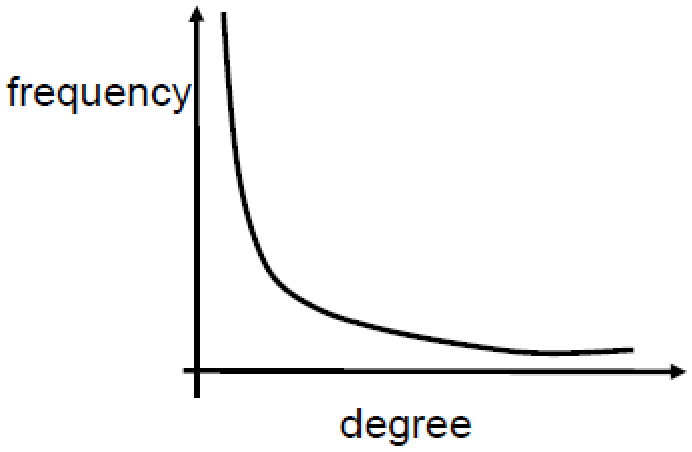
Is degree everything?

- Nodes with the same degree might have different properties
- In what ways does degree fail to capture centrality in the following graphs?
  - ability to broker between groups
  - likelihood that information originating anywhere in the network reaches you…



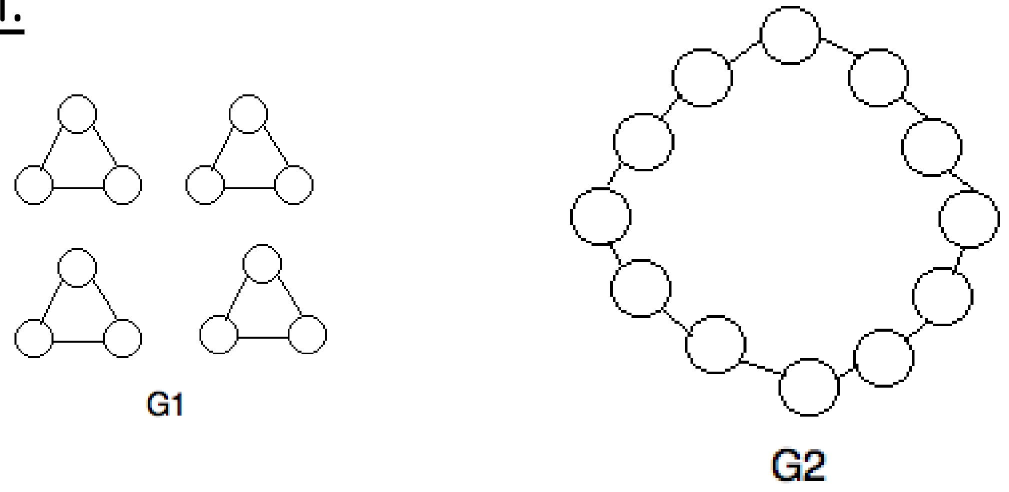
Degree and Degree Distribution

- Degree k_i of node i is a measure of its centrality
- Nodes with high degrees are called hubs
- Maximum degree $k_{max} = \max_i(k_i)$ is also an important measure
- The variance of node-degrees can be an indicator of network heterogeneity, i.e. the more the variance the more the heterogeneity.
- Degree distribution



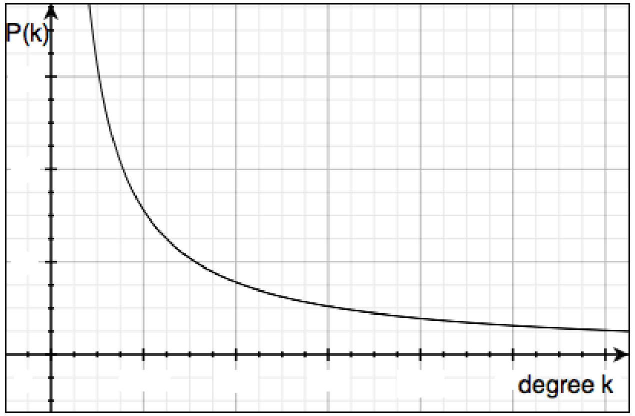
Degree Distribution

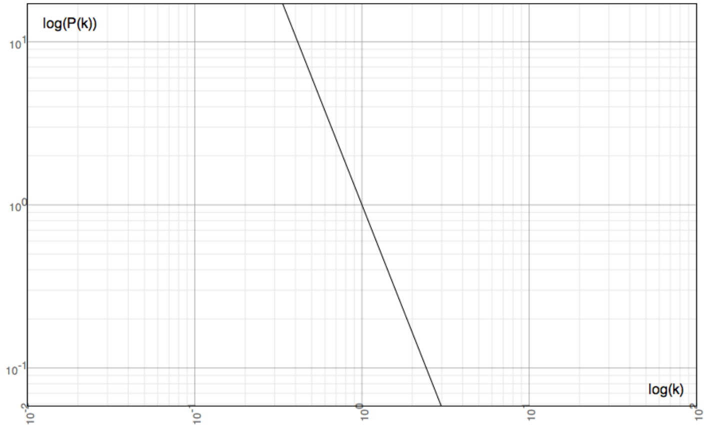
- However: degree distribution (and global properties in general) are weak predictors of network structure.
- Illustration:



- G_1 and G_2 are of the same size (i.e.,$|G_1|=|G_2|$ -- they have the same number of nodes and edges) and they have same degree distribution, but G_1 and G_2 have very different topologies (i.e., graph structure).

Degree Distribution



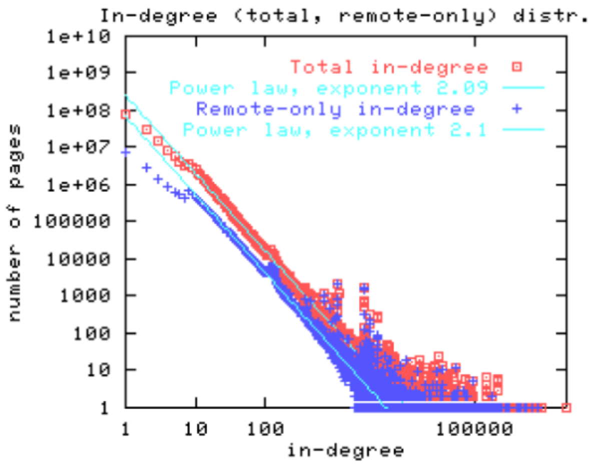


(log-log plot)

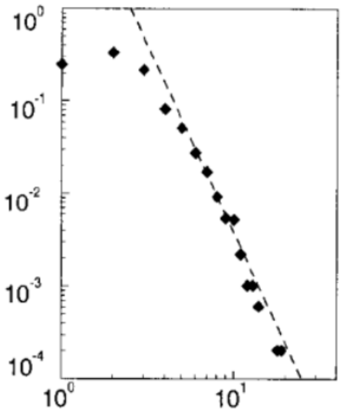
- Here $P(k) \sim k^{-\gamma}$, where often $2 \leq \gamma < 3$. This is a ***power-law***, heavy-tailed distribution.
- Networks with power-law degree distributions are called ***scale-free*** **networks**. In them, most of the nodes are of low degree, but there is a small number of highly-linked nodes (nodes of high degree) called “hubs.”

Powe Laws

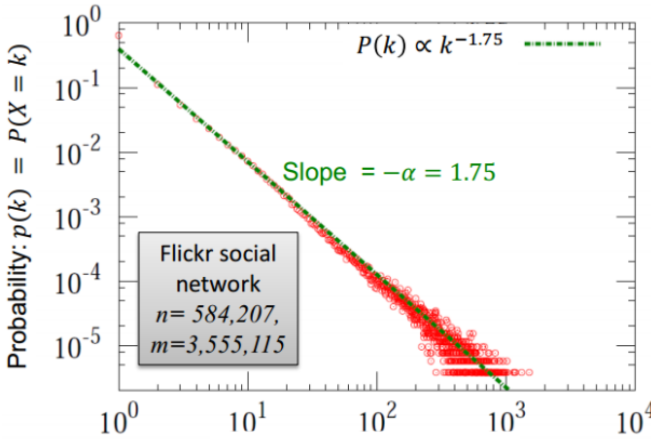
- Social and information networks often follow power laws, meaning that a few nodes have many of the edges, and many nodes have a few edges.



e.g. web graph (Broder et al.)



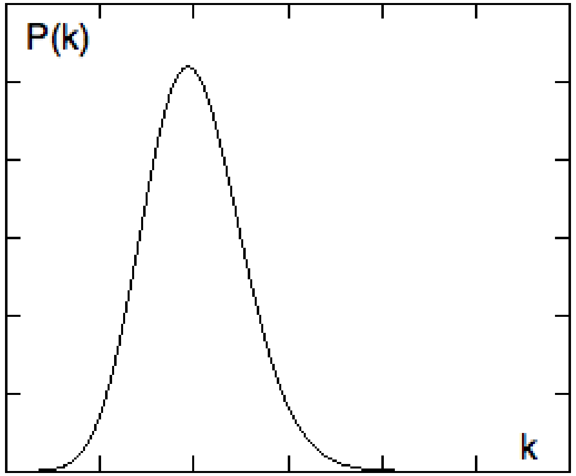
e.g. power grid (Barabasi-Albert)



e.g. Flickr (Leskovec)

Degree Distribution

- Here P(k) is a Poisson distribution.



average degree is meaningful

Degree-Degree correlation

- It is important to know if the nodes with degree k are connected to nodes with degree k'. How?
  - 1) Method proposed by Pastor Satoras et al. and plot the mean degree of the neighbors as a function of the degree

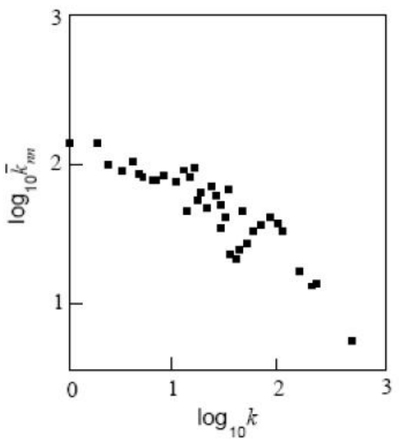


FIG. 3.13. Correlations of the degrees of nearest-neighbour vertices (autonomous systems) in the Internet at the interdomain level (after Pastor-Satorras, Vázquez, and Vespignani 2001). The empirical dependence of the average degree of the nearest neighbours of a vertex on the degree of this vertex is shown in a log–log scale. This empirical dependence was fitted by a power law with exponent approximately 0.5.

Degree-Degree correlation

- It is important to know if the nodes with degree k are connected to nodes with degree k'. How?
  - 2) method proposed by Newman and compute the correlation coefficient

$$r = \frac{\frac{1}{E}\sum_{j>i} k_i k_j a_{ij} - \left[\frac{1}{E}\sum_{j>i}\frac{1}{2}\left(k_i + k_j\right)a_{ij}\right]^2}{\frac{1}{E}\sum_{j>i}\frac{1}{2}\left(k_i^2 + k_j^2\right)a_{ij} - \left[\frac{1}{E}\sum_{j>i}\frac{1}{2}\left(k_i + k_j\right)a_{ij}\right]^2}$$

E is the total number of edges

a_{ij} is the entry (i,j) of the adjacency matrix

k_i is the degree of node i

Degree-Degree correlation

- $r > 0$: the network is called assortative
  - Node with large degree intent to connect to those with large degrees and nodes with low degrees intend to connect to those with low degrees (rich with rich and poor with poor)
- $r < 0$: the network is called disassortative
  - Node with large degree intent to connect to those with low degrees and nodes with low degrees intend to connect to those with high degrees (rich with poor)
- $r = 0$: no correlations
  - There is no specific intention in the connection between the nodes in the sense of their degrees

Any Question?