Strength of Weak Ties and Community Structure in Networks

CS224W: Social and Information Network Analysis

Jure Leskovec, Stanford University

http://cs224w.stanford.edu

Networks: Flow of information

- How information flows through the network?
- How different nodes can play structurally distinct process in roles in this process?
- How different links (short range vs. long range) play different roles in diffusion?

Strength of weak ties

[Granovetter '73]

- How people find out about new jobs?
  - Mark Granovetter, part of his PhD in 1960s
  - People find the information through personal contacts
- But: Contacts were often acquaintances rather than close friends
  - This is surprising:
    - One would expect your friends to help you out more than casual acquaintances when you are between the jobs
- Why is it that distance acquaintances are most helpful?

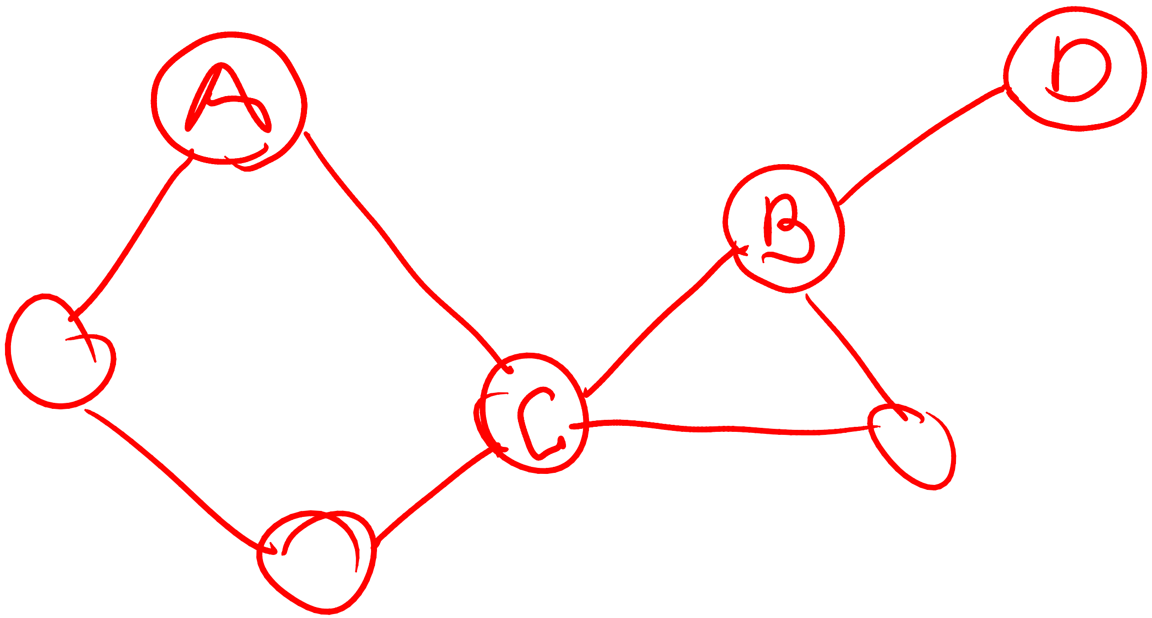
[Granovetter '73]

Granovetter's answer

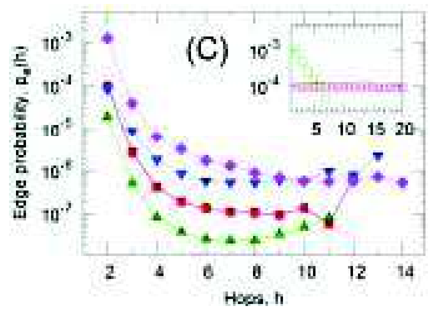
- Two perspectives on friendships:
  - Structural:
    - Friendships span different portions of the network
  - Interpersonal:
    - Friendship between two people is either strong or weak

Triadic closure

- Which edge is more likely A-B or A-D?



- Triadic closure: If two people in a network have a friend in common there is an increased likelihood they will become friends themselves



Triadic closure

- Triadic closure == High clustering coefficient

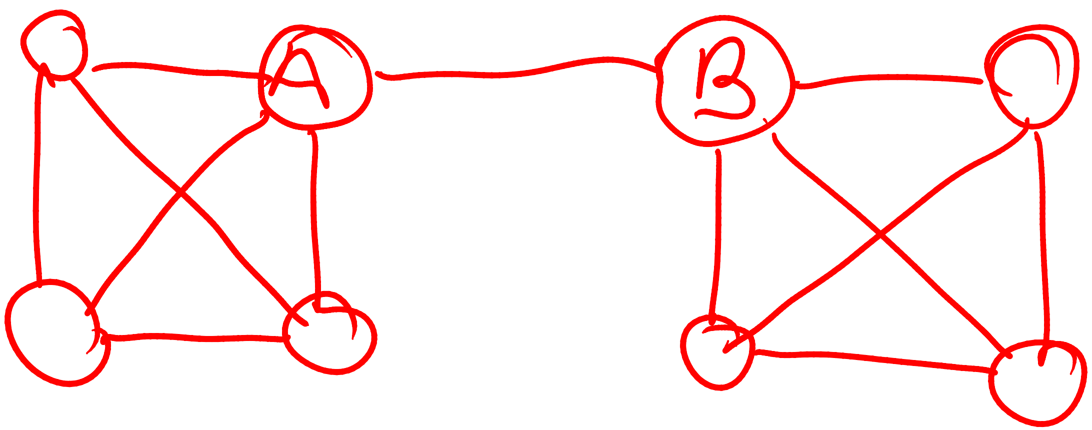
Reasons for triadic closure:

- If B and C have a friend A in common, then:
  - B is more likely to meet C
    - (since they both spend time with A)
  - B and C trust each other
    - (since they have a friend in common)
  - A has incentive to bring B and C together
    - (as it is hard for A to maintain two disjoint relationships)

- Empirical study by Bearman and Moody:
  - Teenage girls with low clustering coefficient are more likely to contemplate suicide

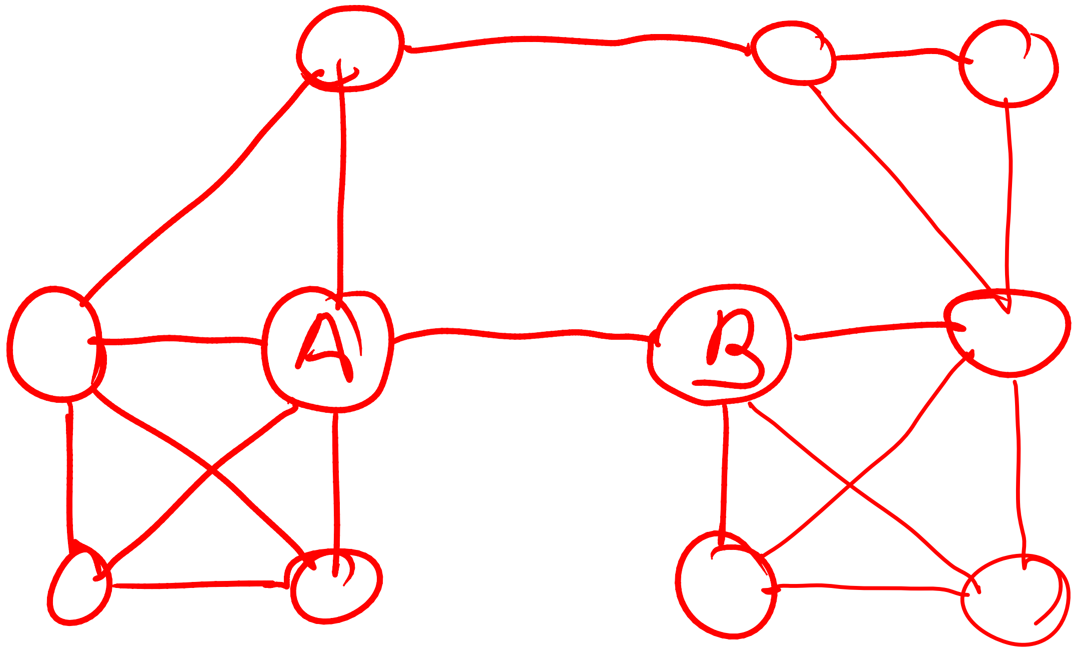
Bridges and Local Bridges

- Edge (A,B) is a bridge if deleting it would make A and B in be in two separate connected components.



Bridges and Local Bridges

- Edge (A,B) is a local bridge A and B have no friends in common
- Span of a local bridge is the distance of the edge endpoints if the edge is deleted



(local bridges with long span are like real bridges)

Strong Triadic Closure

- Links in networks have strength:
  - Friendship
  - Communication
- We characterize links as either Strong (friends) or Weak (acquaintances)
- Def: **Strong Triadic Closure Property:**
  If A has strong links to B and C, then there must be a link (B,C) (that can be strong or weak)

Local Bridges and Weak ties

- Claim: If node A satisfies Strong Triadic Closure and is involved in at least two strong ties, then any local bridge adjacent to A must be a weak tie.
- Proof by contradiction:
  - A satisfies Strong Triadic Closure
  - Let A-B be local bridge and a strong tie
  - Then B-C must exist because of Strong Triadic Closure
  - But then (A,B) is not a bridge

Summary of what we just did

- Defined Local Bridges:
  - Edges not in triangles
- Set two types of edges:
  - Strong and Weak Ties
- Defined Strong Triadic Closure:
  - Two strong ties imply a third edge
- → Local bridges are weak ties

Tie strength in real data

- For many years the Granovetter's theory was not tested
- But, today we have large who-talks-to-whom graphs:
  - Email, Messenger, Cell phones, Facebook
- Onnela et al. 2007:
  - Cell-phone network of 20% of country's population

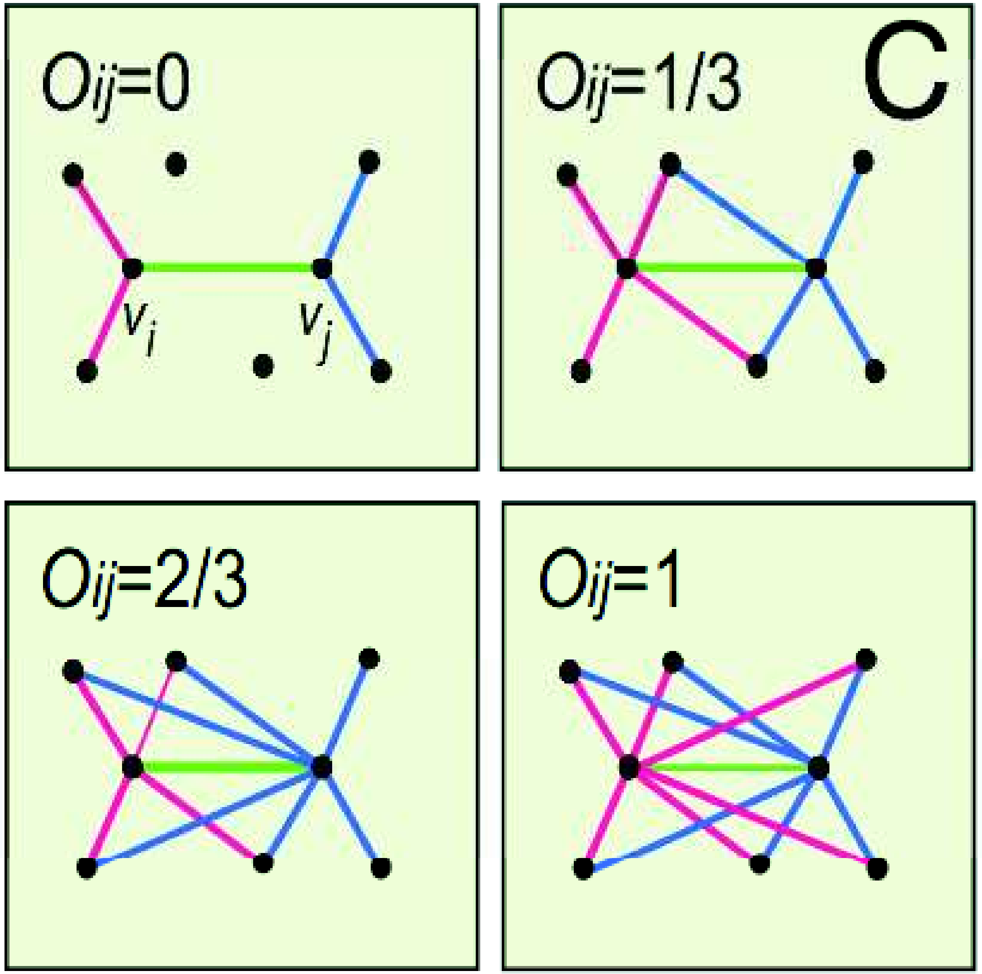
Neighborhood Overlap

- Overlap:

  $$O_{ij} = \frac{n(i) \cap n(j)}{n(i) \cup n(j)}$$

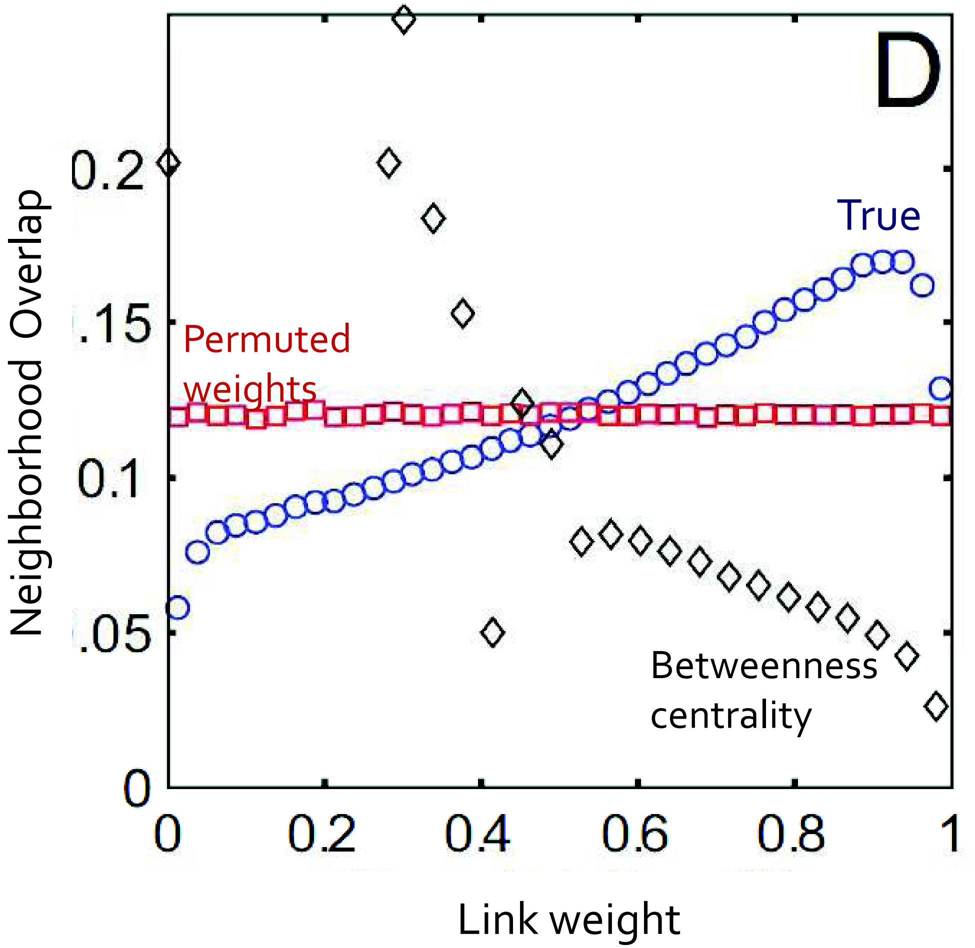
  - n(i) … set of neighbors of A

- Overlap $= 0$ when an edge is a local bridge



[Onnela et al. '07]

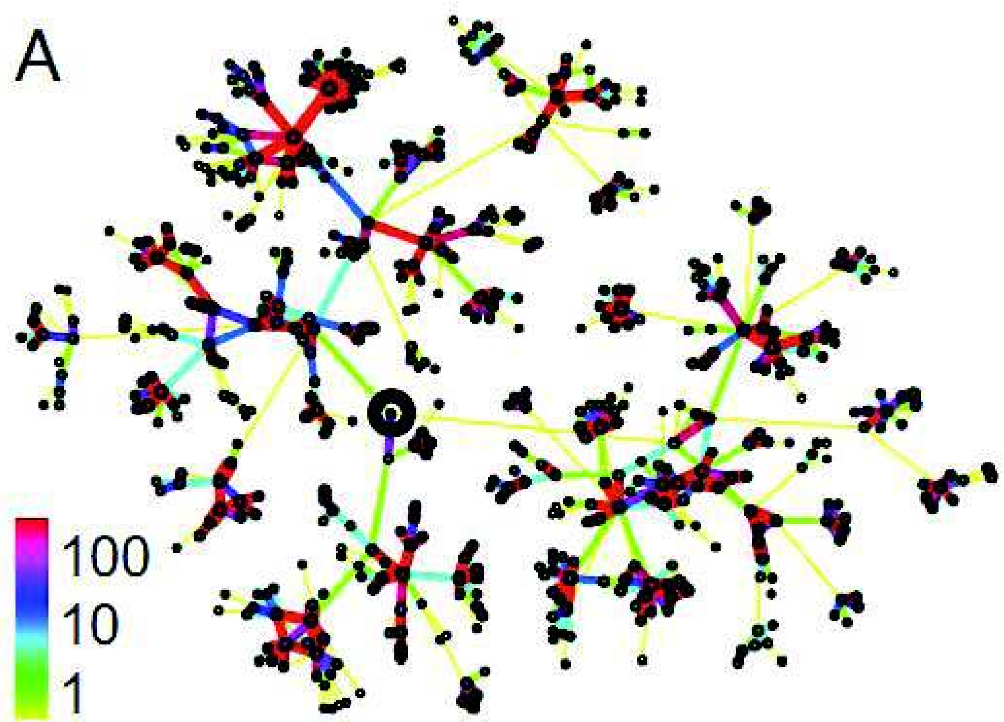
Mobile phones: Overlap vs. Weight



- Permuted weights: Keep the structure but randomly reassign edge weights
- Betweenness centrality: Number of shortest paths going through an edge

Real network tie strengths

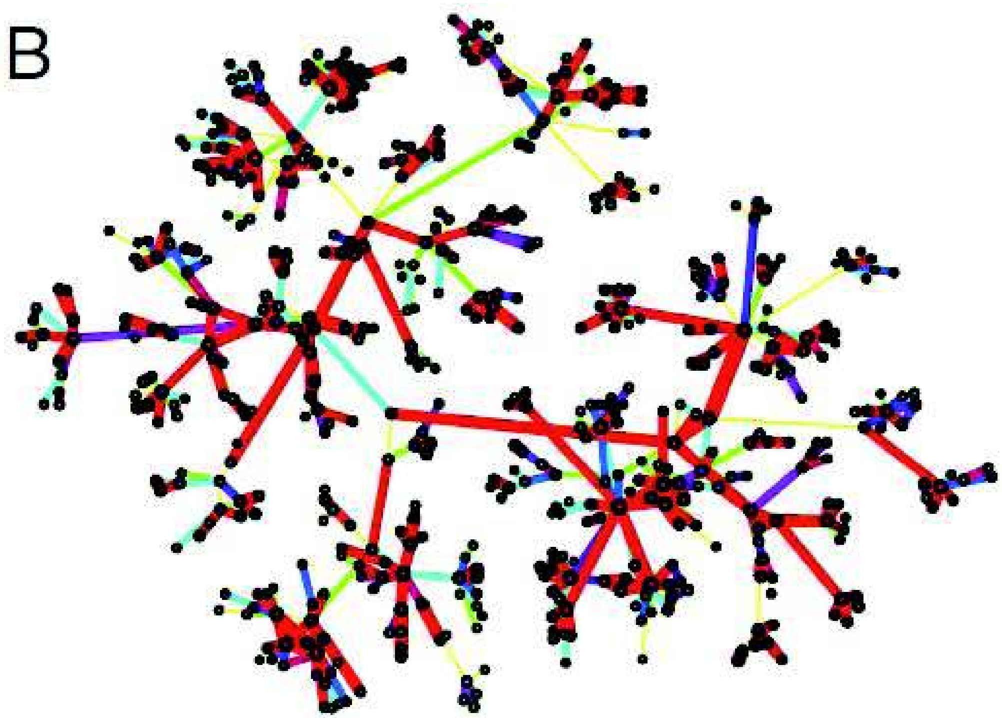
[Onnela et al. '07]



- Real edge strengths in mobile call graph

Permuted tie strenghts

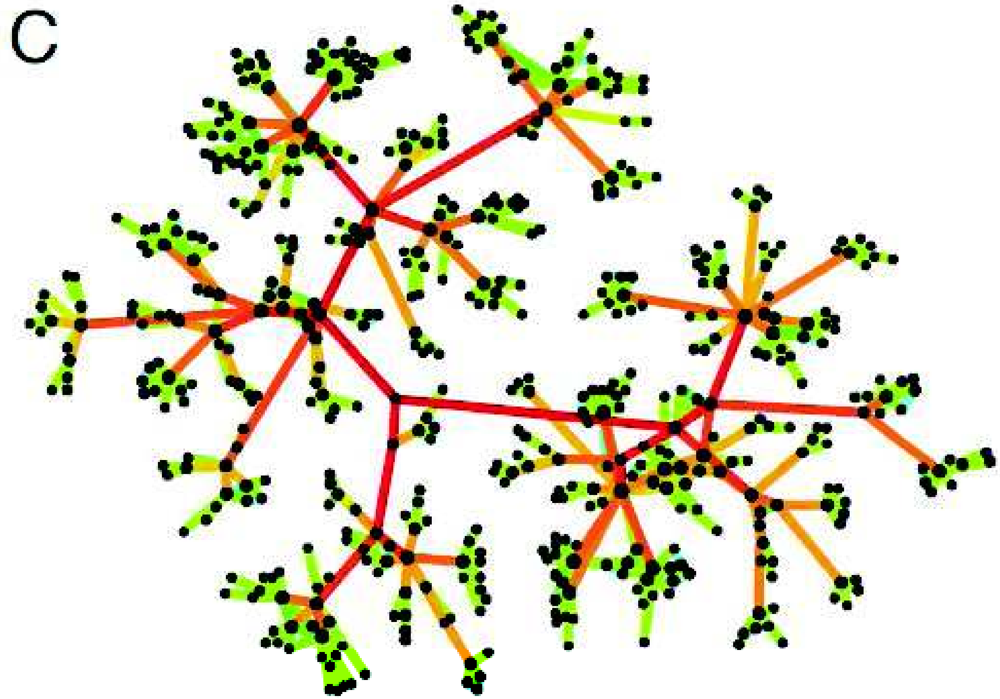
[Onnela et al. '07]



- Same network, same set of edge strengths
- But now strengths are randomly shuffled over the edges

[Onnela et al. '07]

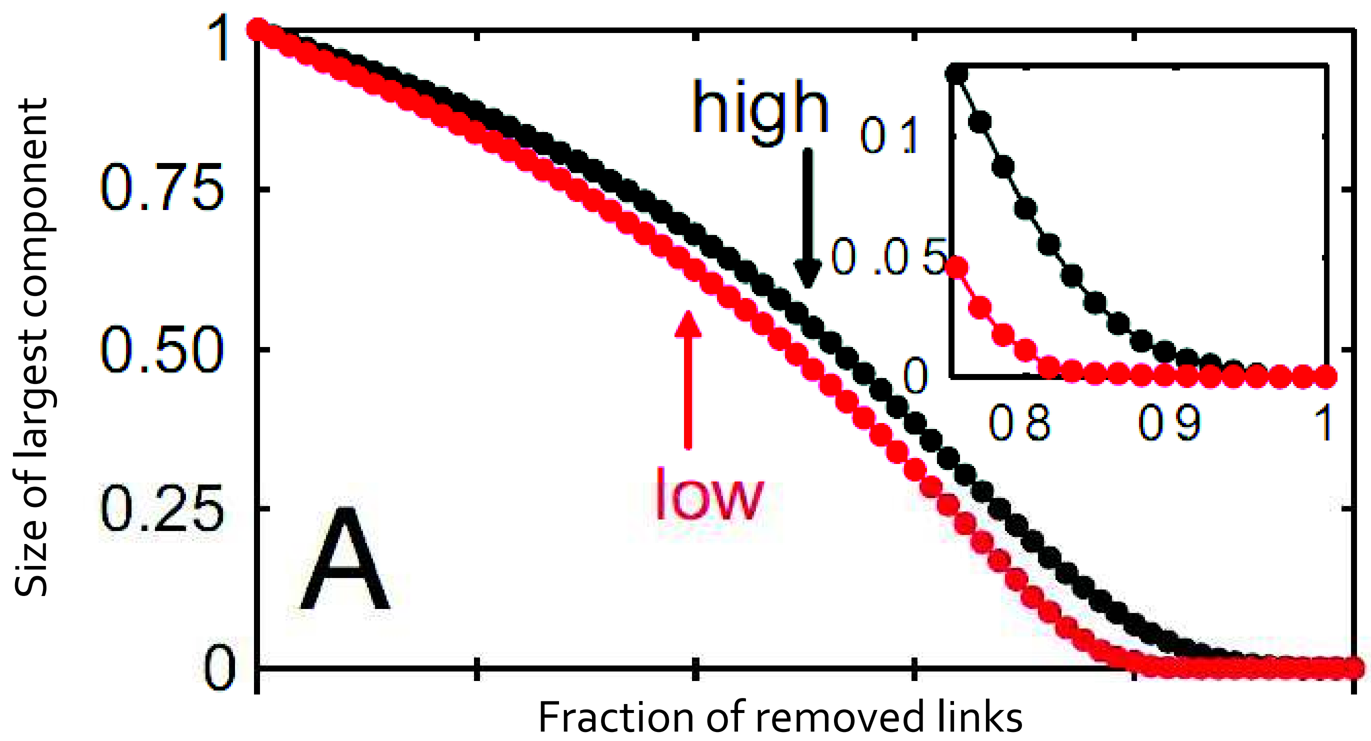
Edge betweenness centrality



- Edges labeled based on betweenness centrality (number of shortest paths going through an edge)

Link removal: Weight

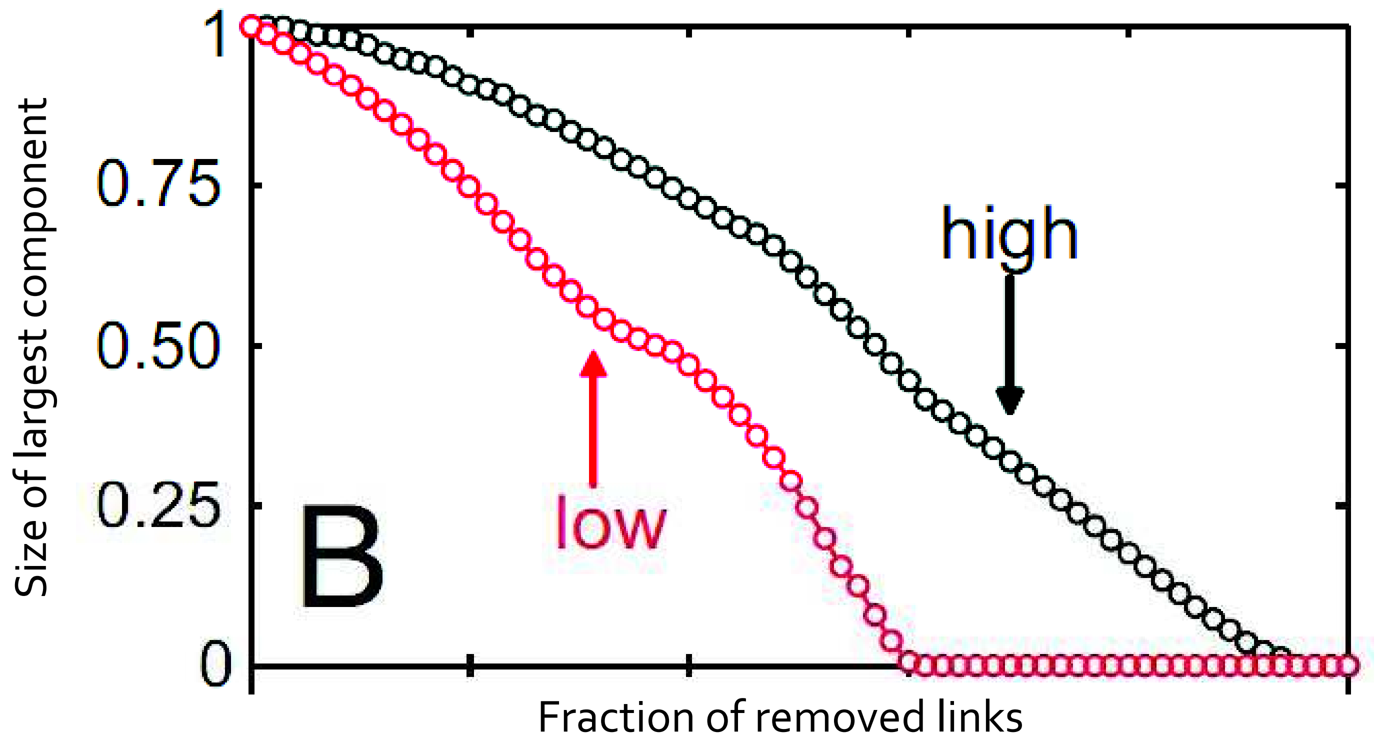
[Onnela et al. '07]



- Removing links based on **strength** (# conversations)
  - Low to high
  - High to low

[Onnela et al. '07]

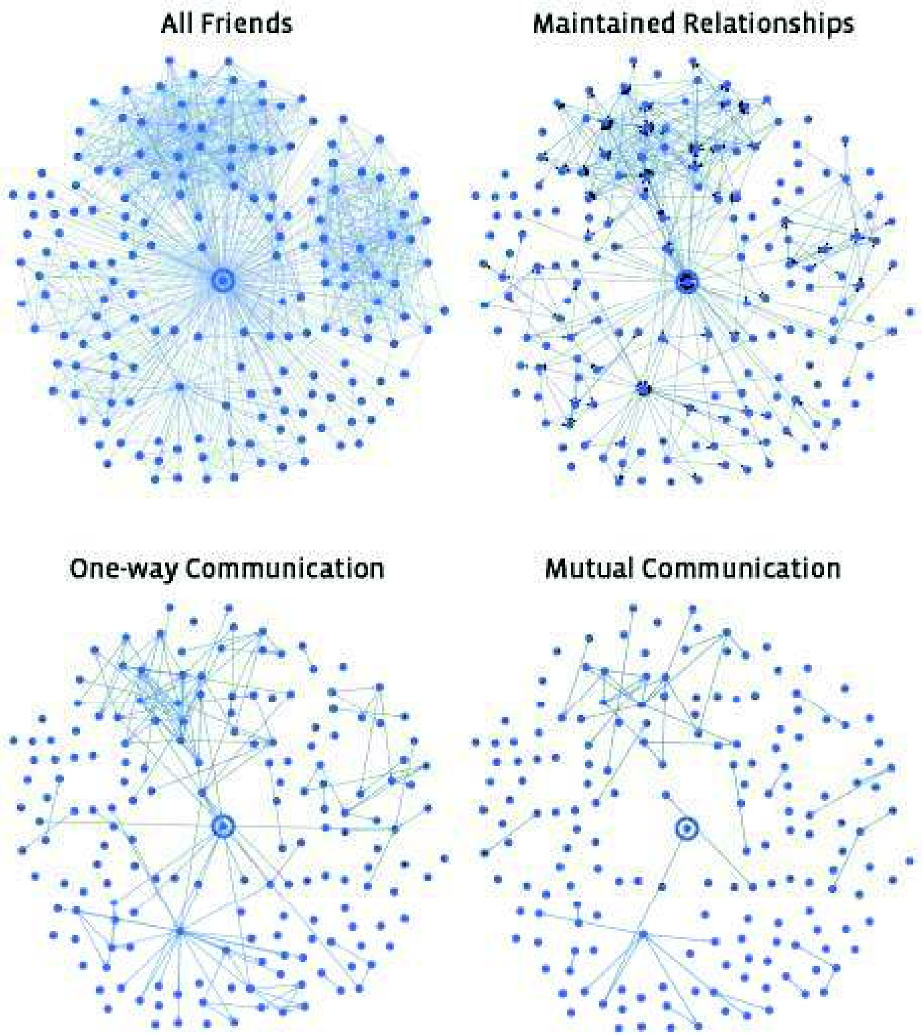
Link removal: Overlap



- Removing links based on **overlap**
  - Low to high
  - High to low

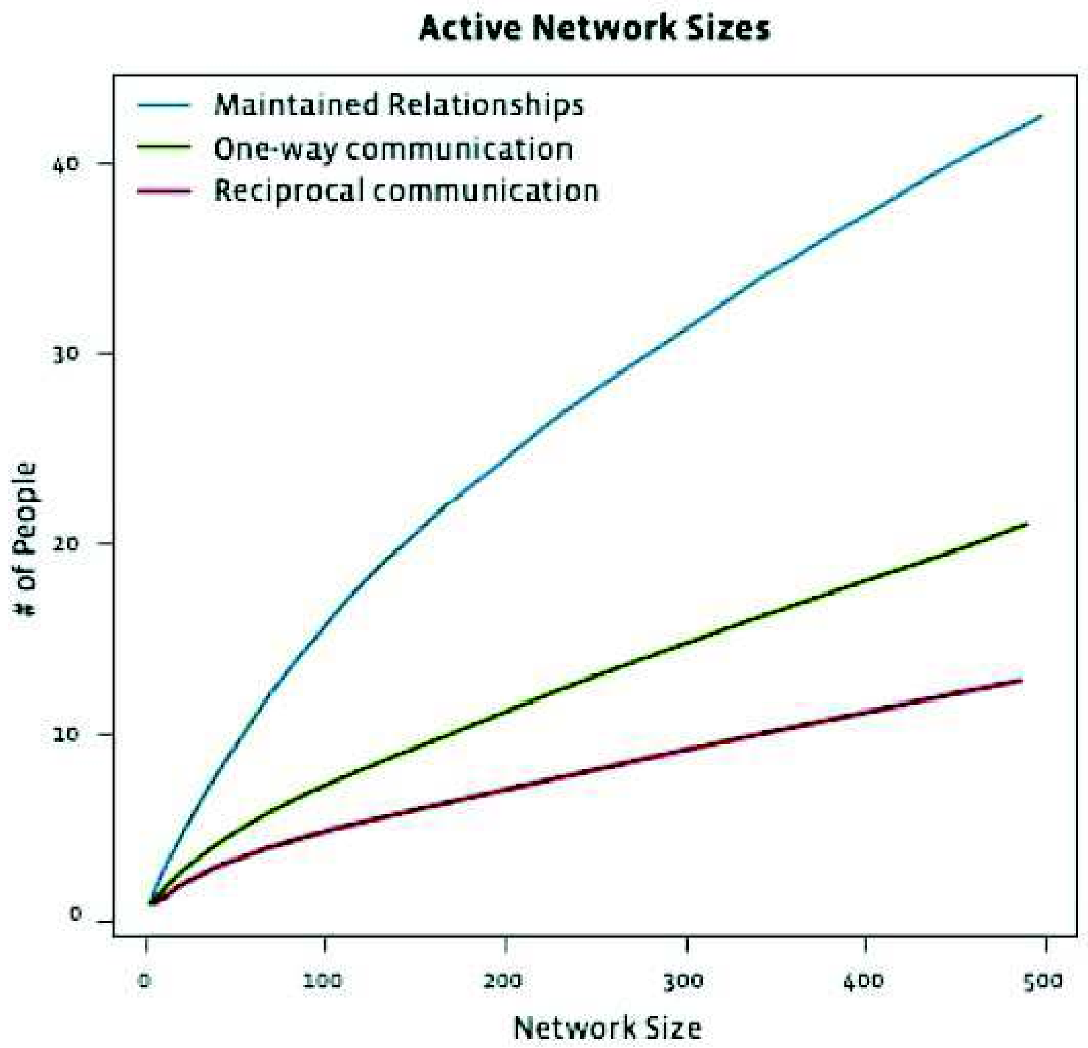
Another example: Facebook

[Marlow et al. '09]



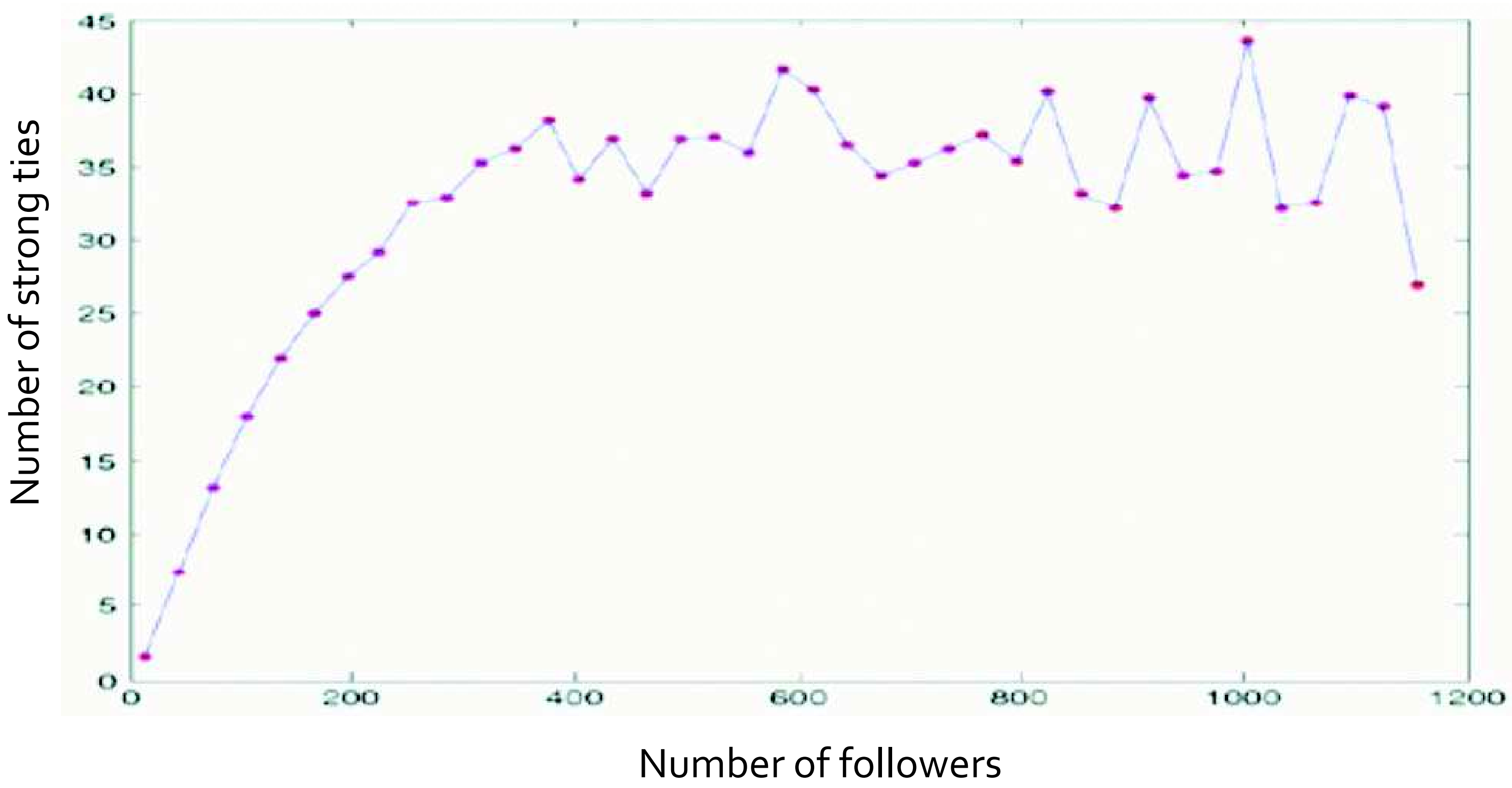
Facebook: Number of ties

[Marlow et al. '09]



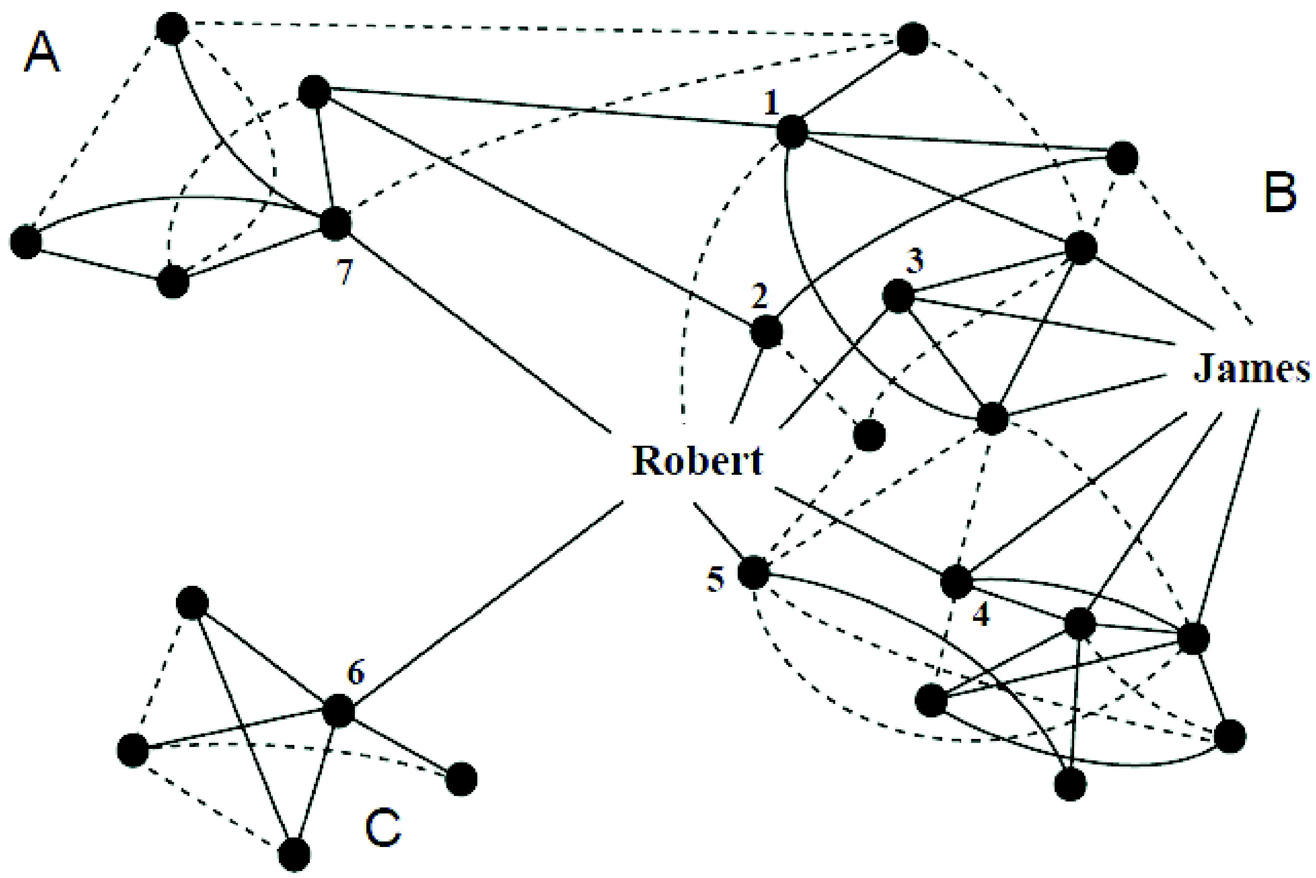
Twitter: Strong ties vs. Followers

[Huberman et al. '09]

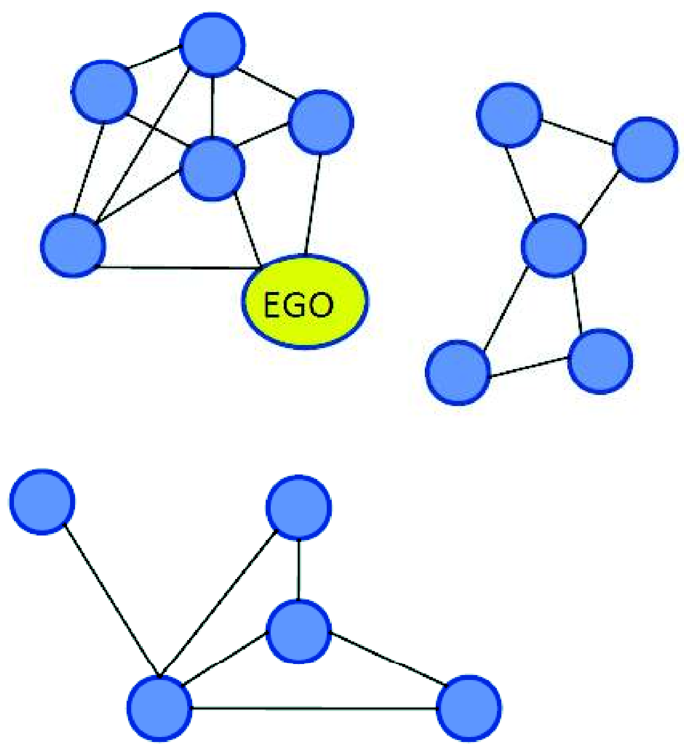


Structural Holes

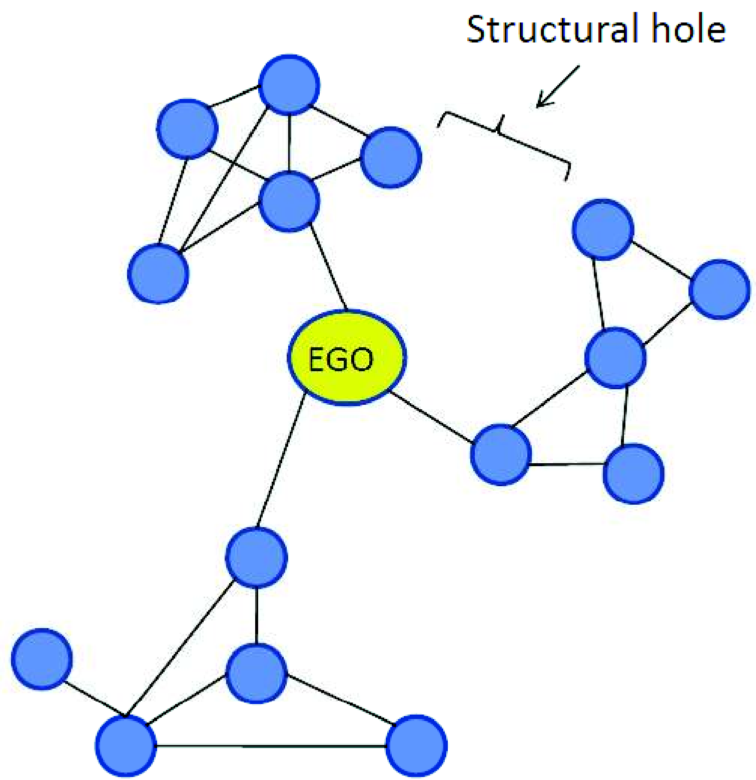
[Ron Burt]



Structural Holes



Few structural holes

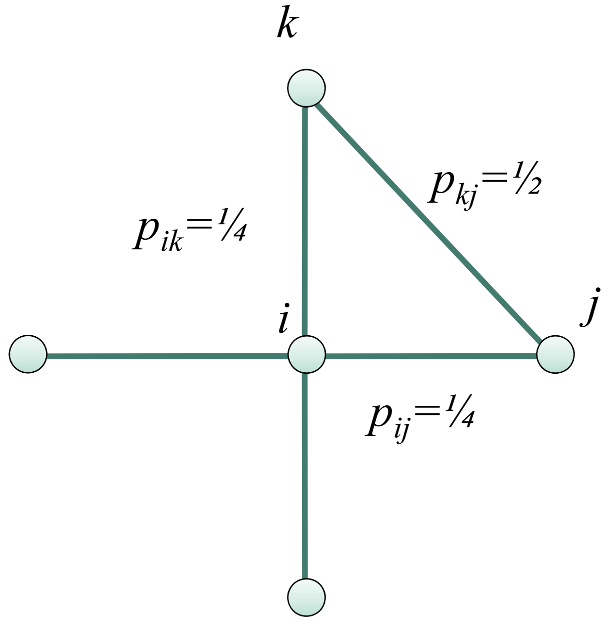


Many structural holes

Structural Holes provide Ego with access to novel information, power, freedom

Structural Holes: Network constraint

- The 'network constraint' measure [Burt]:



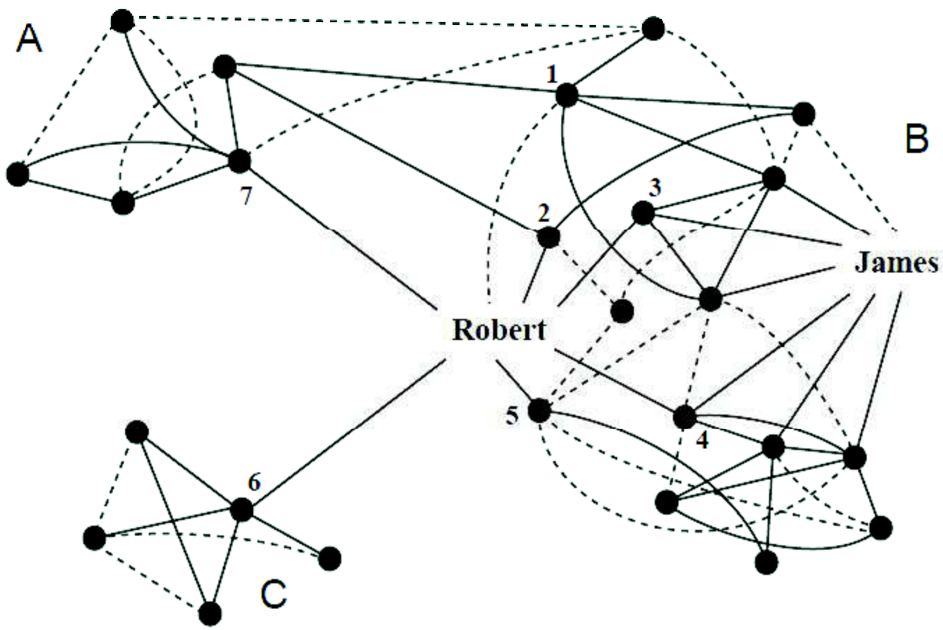
- To what extent are person's contacts redundant
  - Low: disconnected contacts
  - High: contacts that are close or strongly tied

$$c_i = \sum_j \left[p_{ij} + \sum_k \left(p_{ik} p_{kj} \right) \right]^2 :$$

$$c_i = \left[\frac{1}{4} + \left(\frac{1}{4}\frac{1}{2}\right)\right]^2 + \left[\frac{1}{4} + \left(\frac{1}{4}\frac{1}{2}\right)\right]^2 + \left[\frac{1}{4}\right]^2 + \left[\frac{1}{4}\right]^2 = \frac{13}{32}$$

p_{ij} … proportion of i's energy invested in relationship with j

Example: Robert vs. James

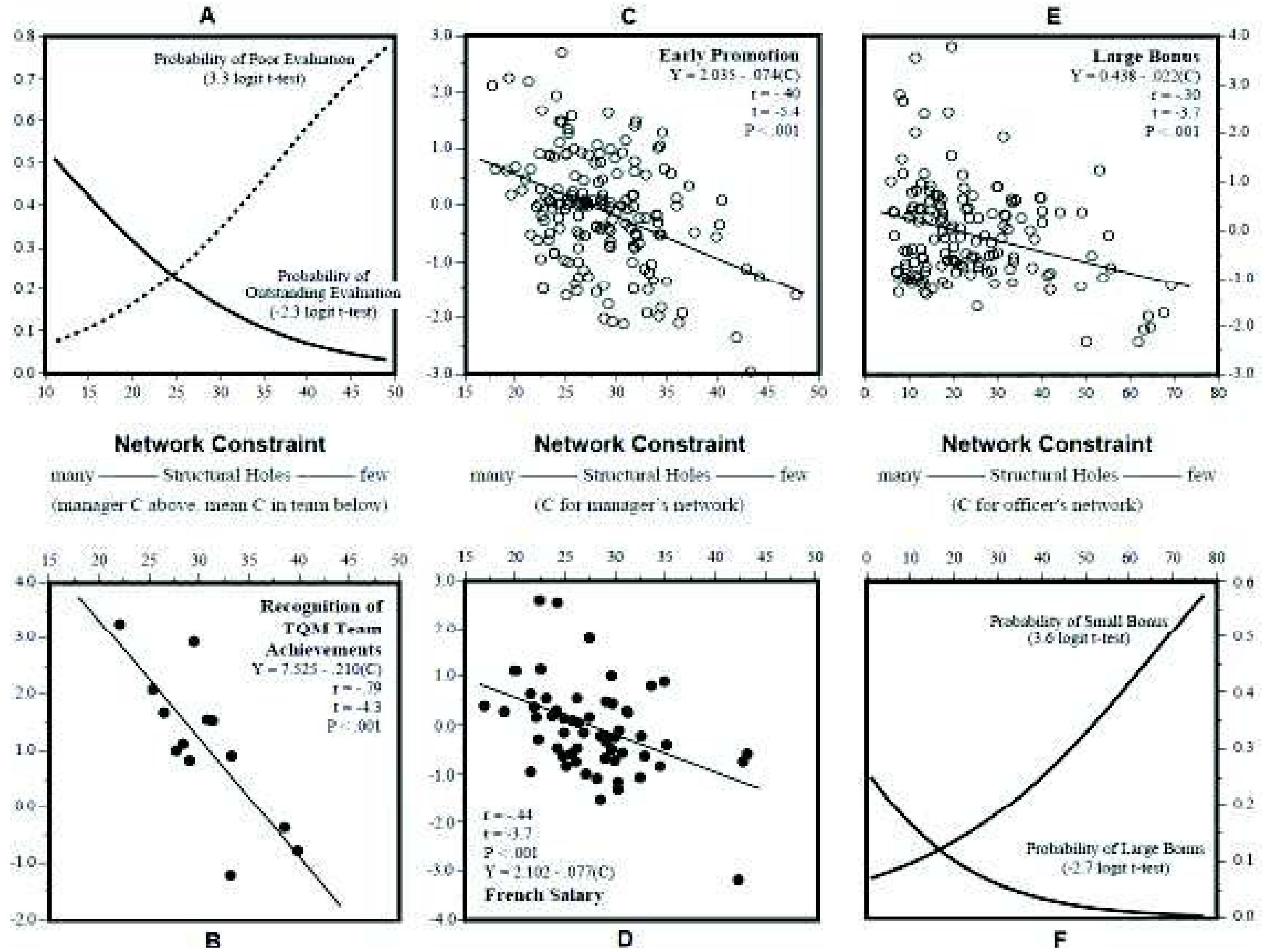


- Constraint: To what extent are person's contacts redundant
  - Low: disconnected contacts
  - High: contacts that are close or strongly tied

- Network constraint:
  - James: c_j=0.309
  - Robert: c_r=0.148

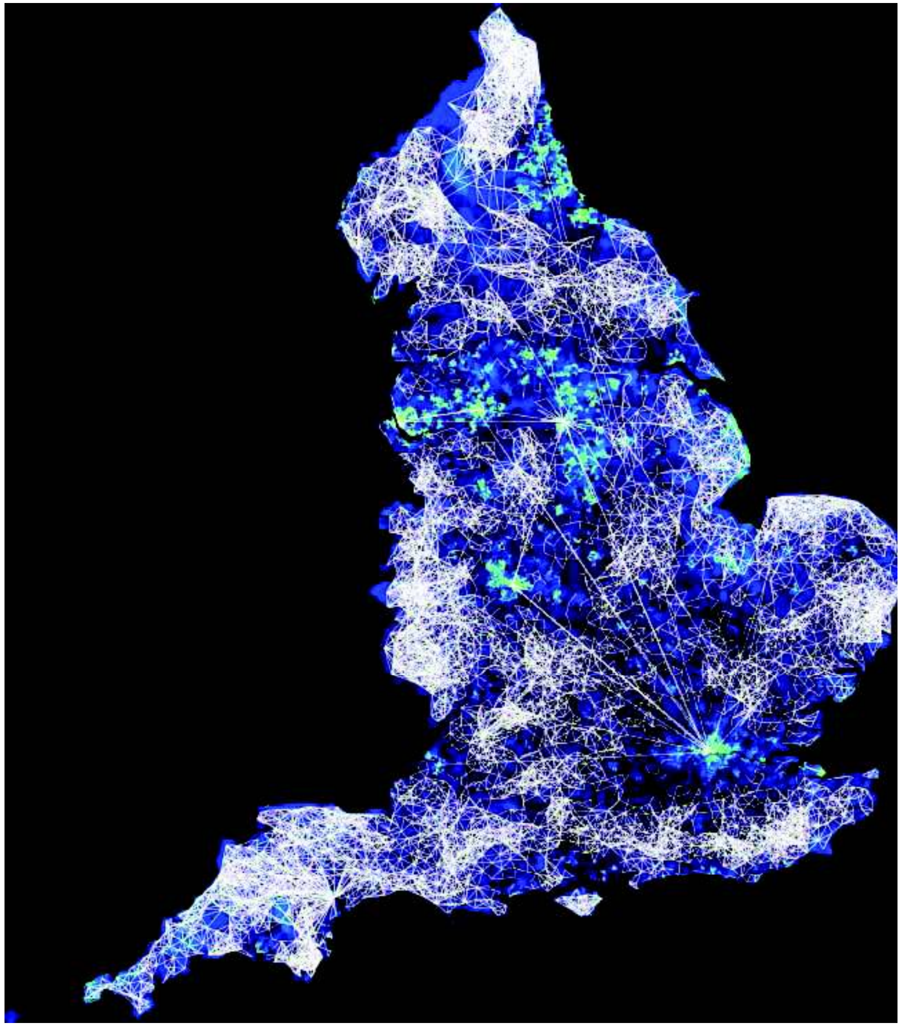
[Ron Burt]

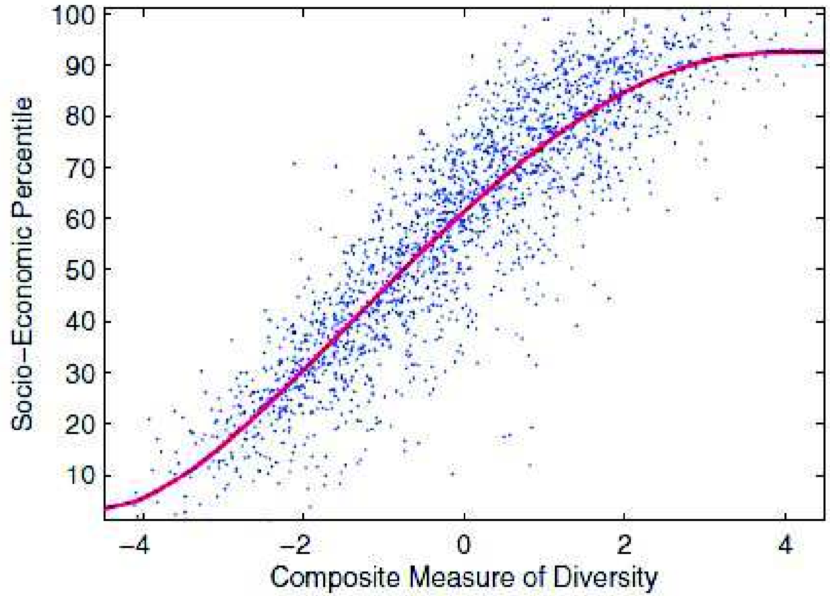
Spanning the holes matters



Diversity & Development

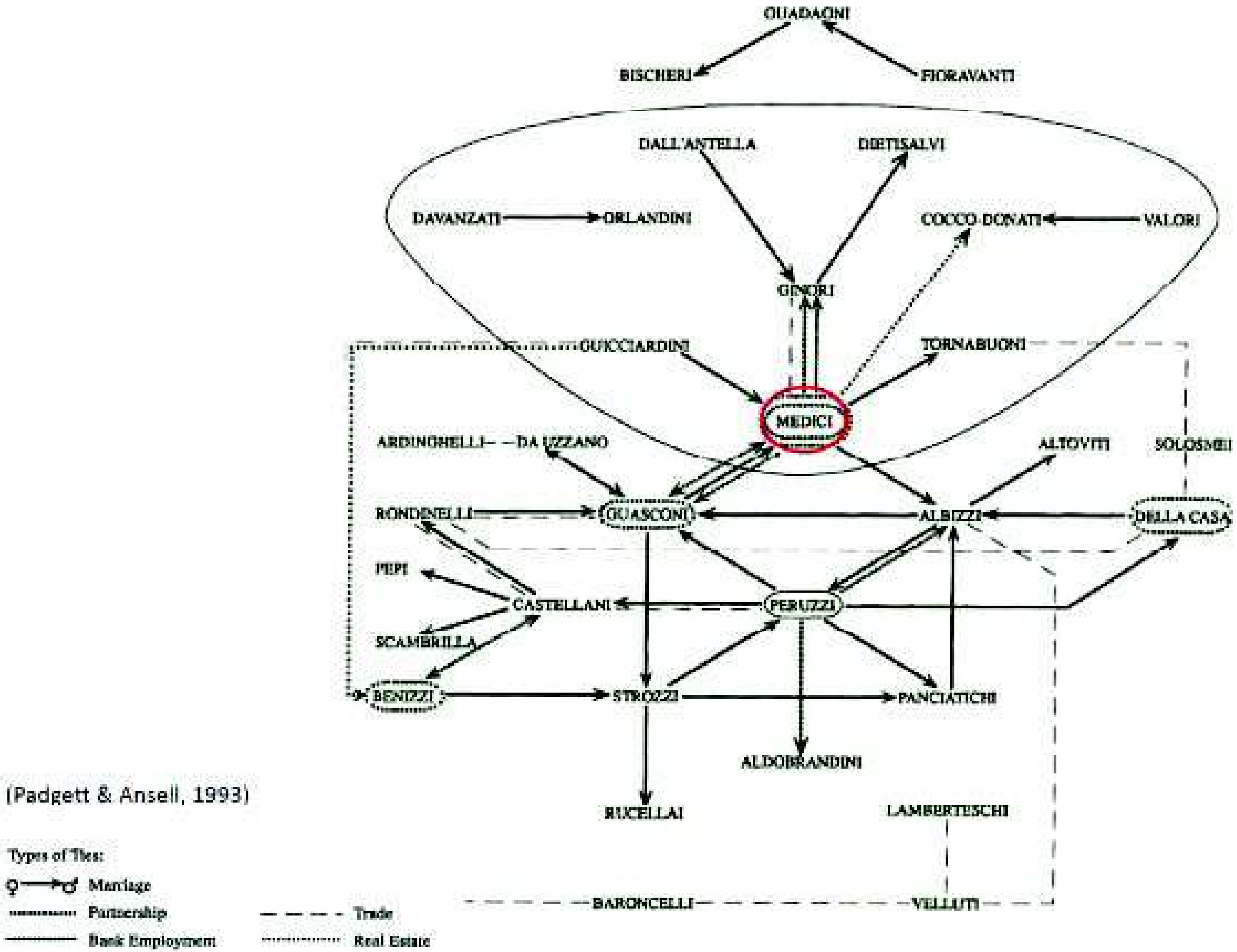
[Eagle-Macy, 2010]





- Measure of diversity:
  - Structural holes + entropy of edge strengths (call volume)

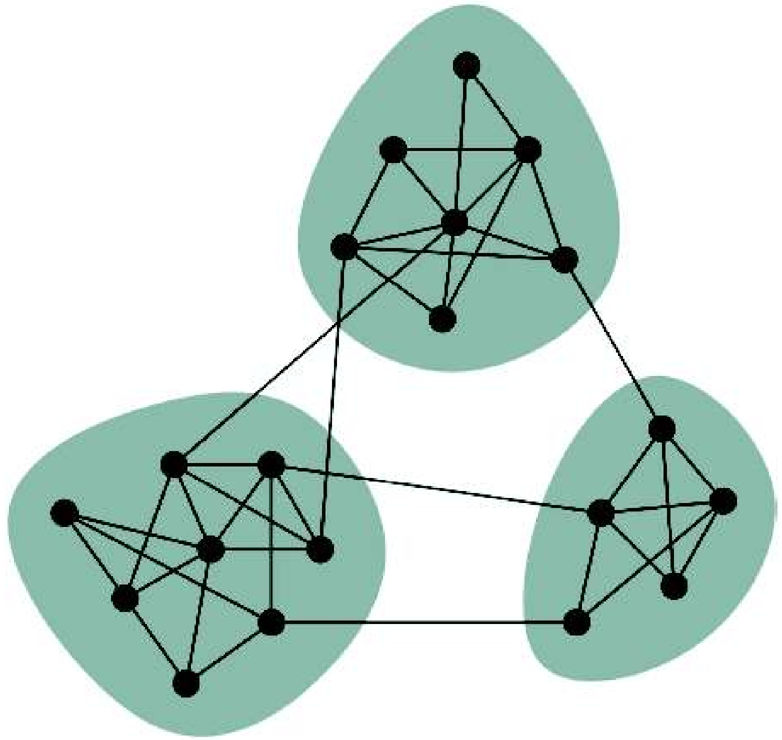
Florentine families: Power



(Padgett & Ansell, 1993)

Network communities

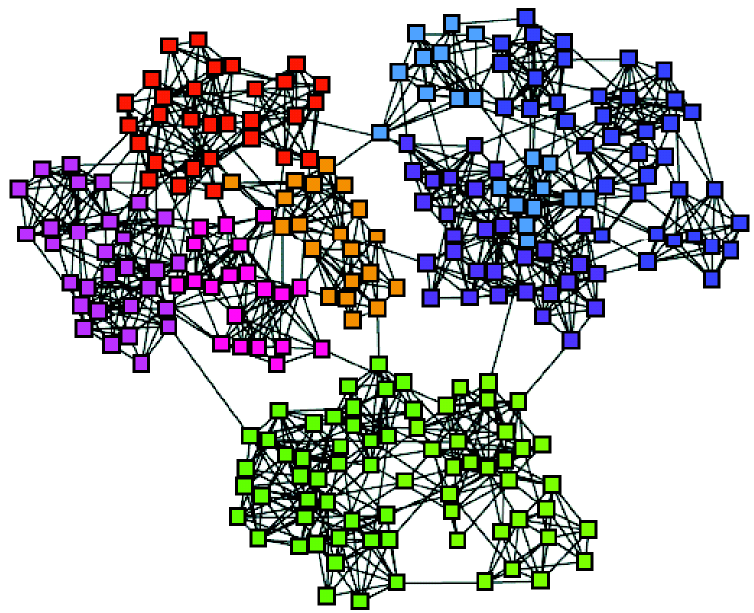
- Networks of **tightly connected groups**
- **Network communities:**
  - Sets of nodes with **lots** of connections **inside** and **few** to **outside** (the rest of the network)



Communities, clusters, groups, modules

Finding network communities

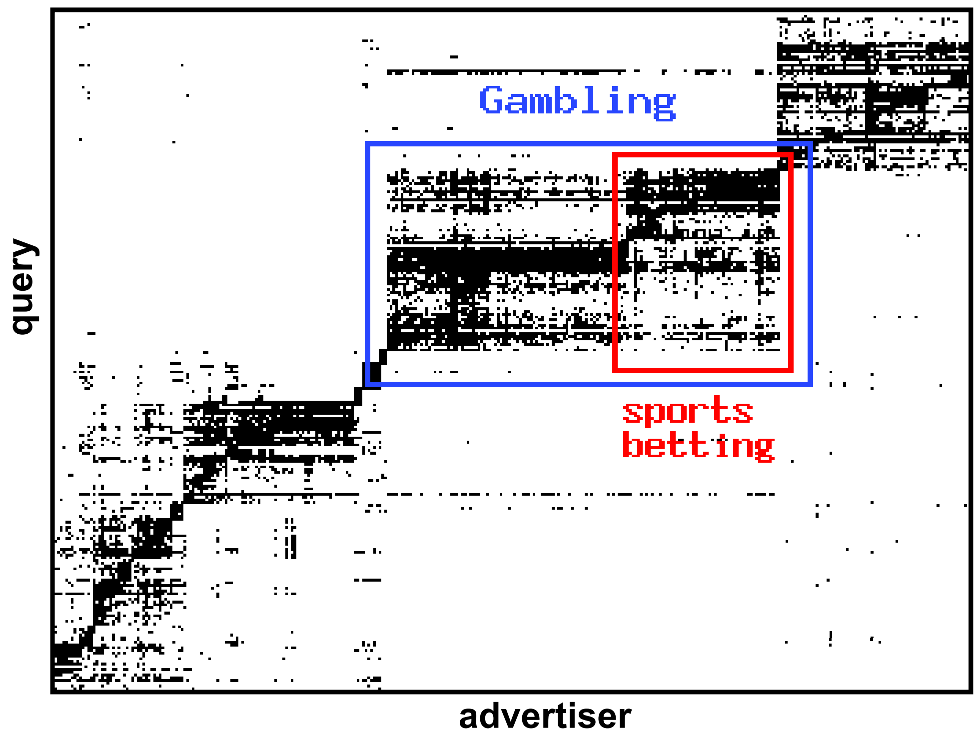
- How to automatically find such densely connected groups of nodes?
- Ideally such automatically detected clusters would then correspond to real groups
- For example:



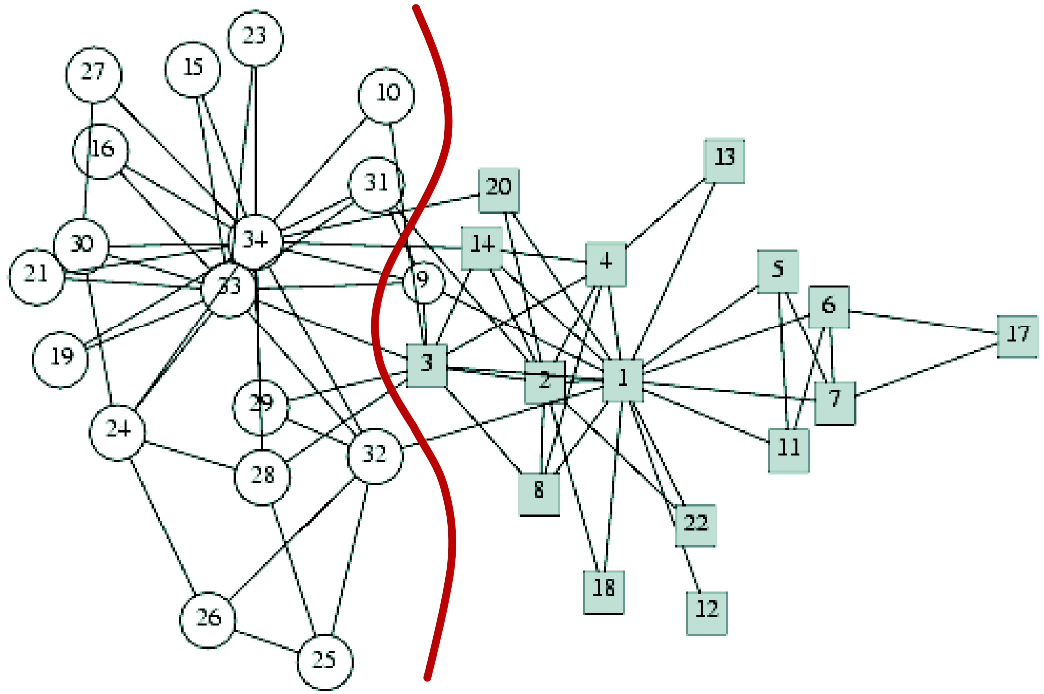
Communities, clusters, groups, modules

Micro-markets in sponsored search

Find micro-markets by partitioning the “query x advertiser” graph:



Social Network Data

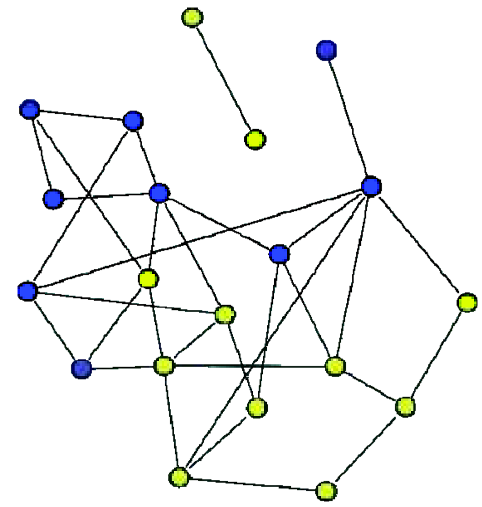


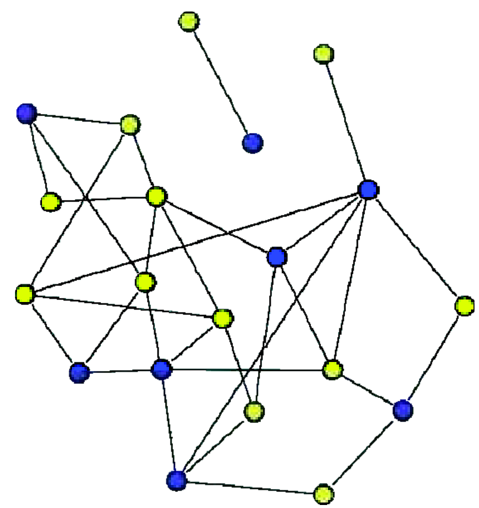
- Zachary's Karate club network:
  - Observe social ties and rivalries in a university karate club
  - During his observation, conflicts led the group to split
  - Split could be explained by a minimum cut in the network
- **Why would we expect such clusters to arise?**

[Backstrom et al. KDD '06]

Group formation in networks

- In a social network nodes explicitly declare group membership:
  - Facebook groups, Publication venue
- Can think of groups as node colors
- Gives insights into social dynamics:
  - Recruits friends? Memberships spread along edges
  - Doesn't recruit? Spread randomly
- What factors influence a person's decision to join a group?

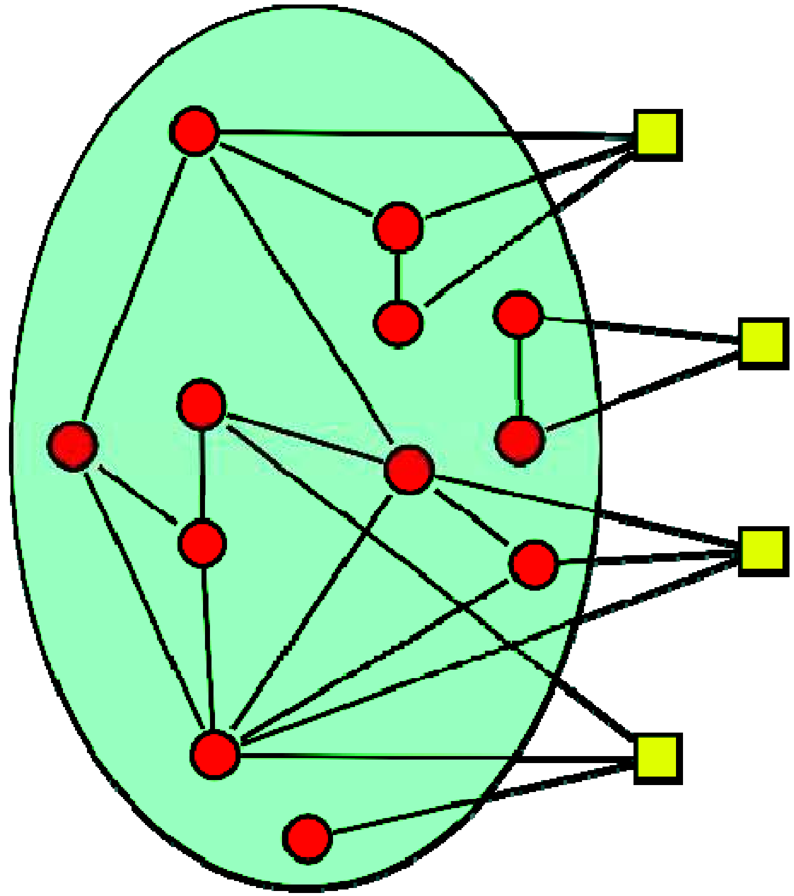




[Backstrom et al. KDD '06]

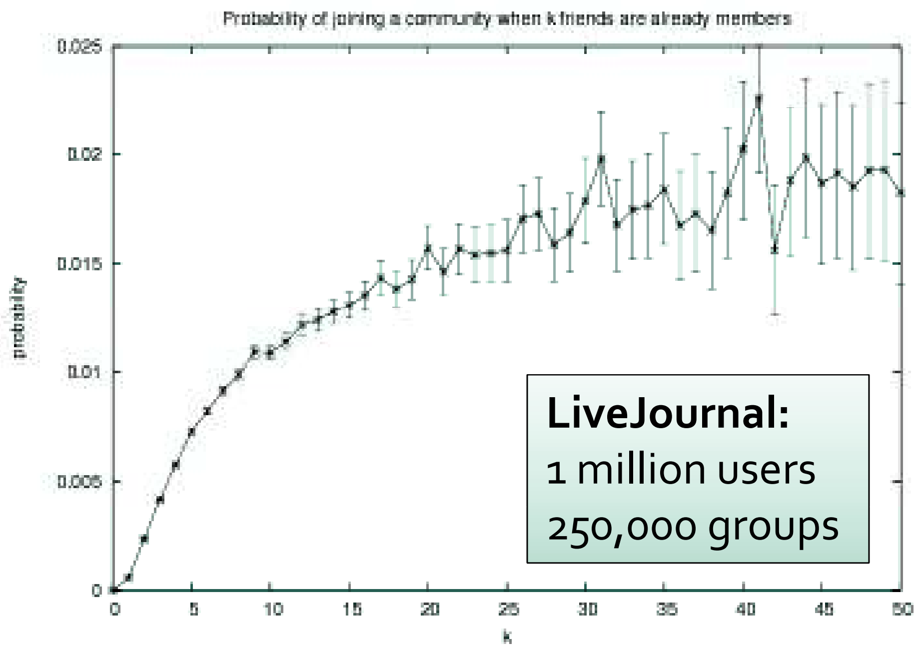
Group growth as diffusion

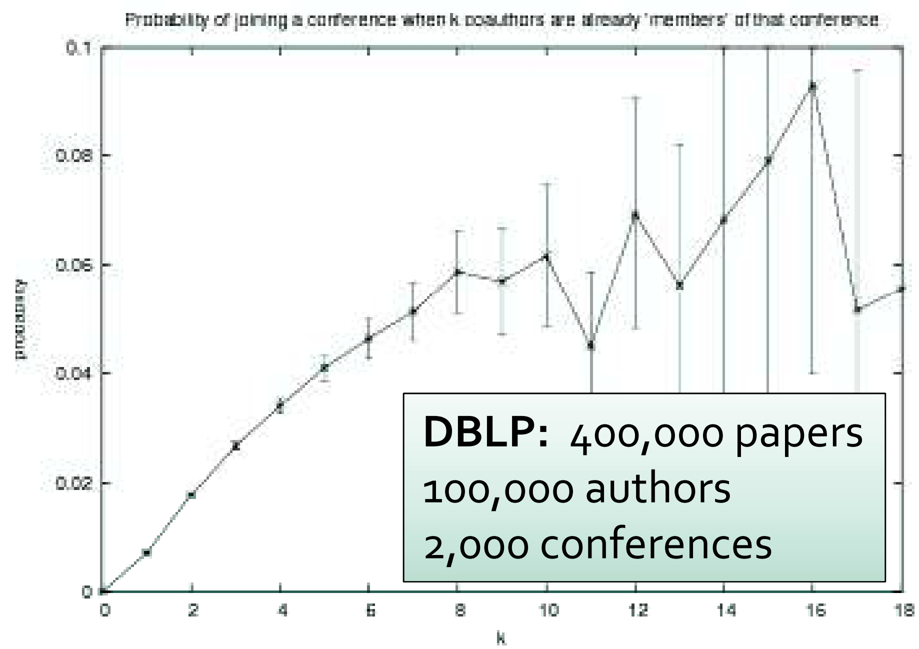
- Analogous to diffusion
  Group memberships spread over the network:
  - Red circles represent existing group members
  - Yellow squares may join
- Question:
  - How does prob. of joining a group depend on the number of friends already in the group?



[Backstrom et al. KDD ‘06]

P(join) vs. # friends in the group





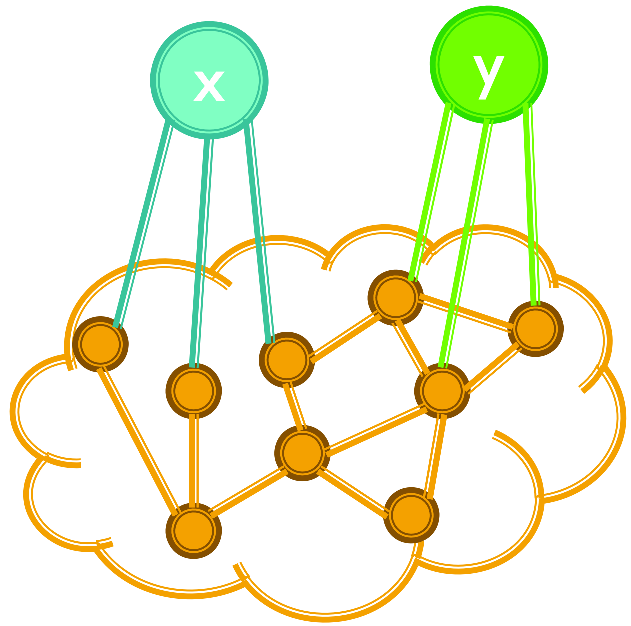
- Diminishing returns:
  - Probability of joining increases with the number of friends in the group
  - But increases get smaller and smaller

[Backstrom et al. KDD '06]

Groups: More subtle features

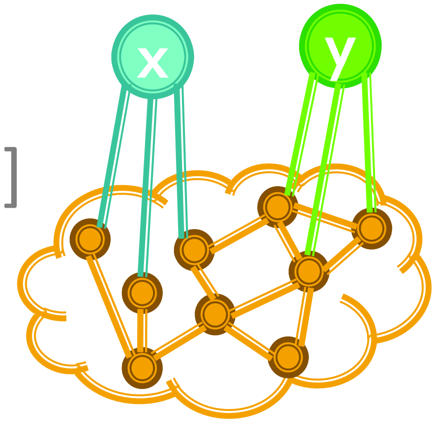
- Connectedness of friends:
  - *x* and *y* have three friends in the group
  - *x*'s friends are independent
  - *y*'s friends are all connected

Who is more likely to join?



[Backstrom et al. KDD '06]

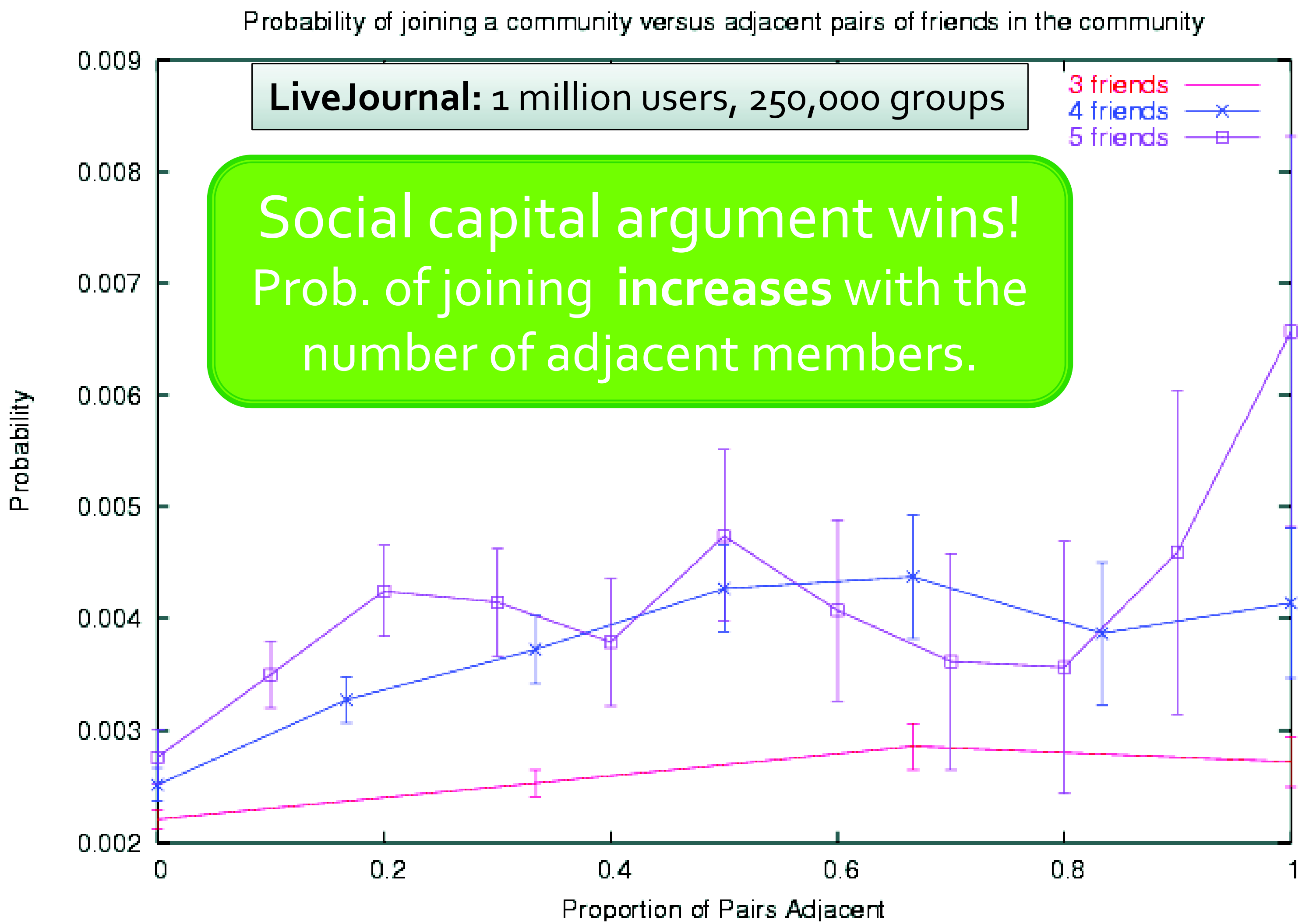
Connectedness of Friends



- Competing sociological theories:
  - Information argument [Granovetter '73]
  - Social capital argument [Coleman '88]
- Information argument:
  - Unconnected friends give independent support
- Social capital argument:
  - Safety/trust advantage in having friends who know each other

[Backstrom et al. KDD '06]

Connectedness of friends



LiveJournal: 1 million users, 250,000 groups

Social capital argument wins!

Prob. of joining **increases** with the number of adjacent members.

So, this means that

- A person is more likely to join a group if
  - she has more friends who are already in the group
  - friends have more connections between themselves
- So, groups form clusters of tightly connected nodes