Tie strength, social capital, betweenness and homophily

Rik Sarkar

Social and technological networks
Edinburgh 2019

Networks

- Position of a node in a network determines its role/importance
- Structure of a network determines its properties

Today

- Notion of strong ties (close friends) and weak ties (remote acquaintances)
  - How they influence the network and spread of information
- Friendships and their evolution
- "Central" locations
- Several small, but related concepts

- [Reference for most: Kleinberg-Easley, Chapter 3,4]
  - Also see end of chapter exercises

Strong and weak ties

- Survey of job seekers show people often find jobs through social contacts
- More important: people more often find jobs through *acquaintances (weak ties)* than close friends (strong ties)

- Strength of weak ties. Mark S. Granovetter, American journal of Sociology, 1973

Strong and weak ties

- Explanation:
  - A close friend is likely in the same community and has the same information sources
  - Person in a different community is more likely to have “new” information, that you do not already know
- *Weak* ties are more critical: they can act as *bridges* across communities

- Other observation: Job information does not travel far – long paths are not involved

Weak ties in social action

- Psychology: People do not often act on global information (radio, tv) etc
- People are more likely to act when confirmed by friends (creates trust)
- Therefore, people are more likely trust a leader when confirmed by direct familiarity or common friends acting as intermediaries
- A society without *bridges* is fragmented
  - The leader does not reach a large number of people that trust him

Weak ties in social action

- Example (from Granovetter): A small town needs to coordinate action on a social issues
  - If everyone works at different places in nearby industries
    - Then people only know their families. There are no work-acquaintances, etc.
    - Organizing a protest is hard
  - If everyone works at the same large industry
    - Likely there are work-acquaintances (weak ties)
    - Social action works better
- See also:
  - Ted talk: Online social change: Easy to organize, hard to win (can you model and explain this?)

Homophily

- We are similar to our friends
  - Not always explained by things intrinsic to the network like simple triadic closure
- External contexts like Culture, hobbies, interests influence networks
- Suppose the network has 2 types of nodes (eg. Male, female), fractions p and q
  - Expected fraction of cross-gender edges: 2pq
- A test for homophily:
  - Fraction of cross gender edges < 2pq

Homophily: The obesity epidemic

- Christakis and fowler (See ted talk: hidden influence of social networks)
- Is it that:
  - People are selecting similar people?
  - Other correlated hommophilic factors (existing food/cultural habits…) affecting data?
  - Are obese friends influencing the habits causing more people to be obese?
- Authors argue that tracking data over a period of time shows significant evidence of the influence hypothesis
  - It is an epidemic

Clustering in social networks

People with mutual friends are often friends

If A and C have a common friend B

- Edges AB and BC exist

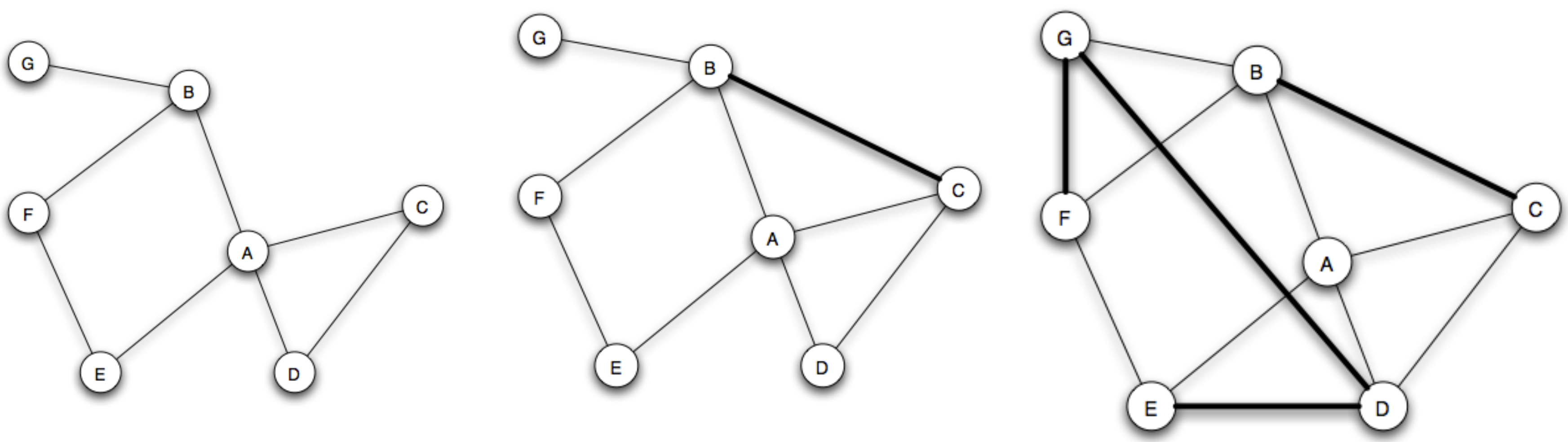
Then ABC is said to form a *Triad*

- Closed triad : Edge AC also exists
- Open triad: Edge AC does not exist

Exercise: Prove that any connected graph has at least n triads (considering both open and closed).

Triadic closure: Friends of Friends

- If two people have a friend in common, they are more likely to become friends
  - *Triadic closure*
- If B & C both know A
  - They are likely to meet, may be for extended time
  - Likely to trust each-other



Clustering coefficient (cc)

Measures how tight the friend neighborhoods are:
   frequency of closed triads

cc(A) fractions of pairs of A's neighbors that are
   friends

Average cc : average of cc of all nodes

Global cc : ratio:

$$\frac{\#\text{ closed triads}}{\#\text{ all triads}}$$

Avg CC In real networks

Facebook (old data) ~ 0.6

https://snap.stanford.edu/data/egonets-Facebook.html

Google web graph ~0.5

https://snap.stanford.edu/data/web-Google.html

In general, cc of ~ 0.2 or 0.3 is considered 'high'

that the network has significant clustering/ community structure

CC of a graph model

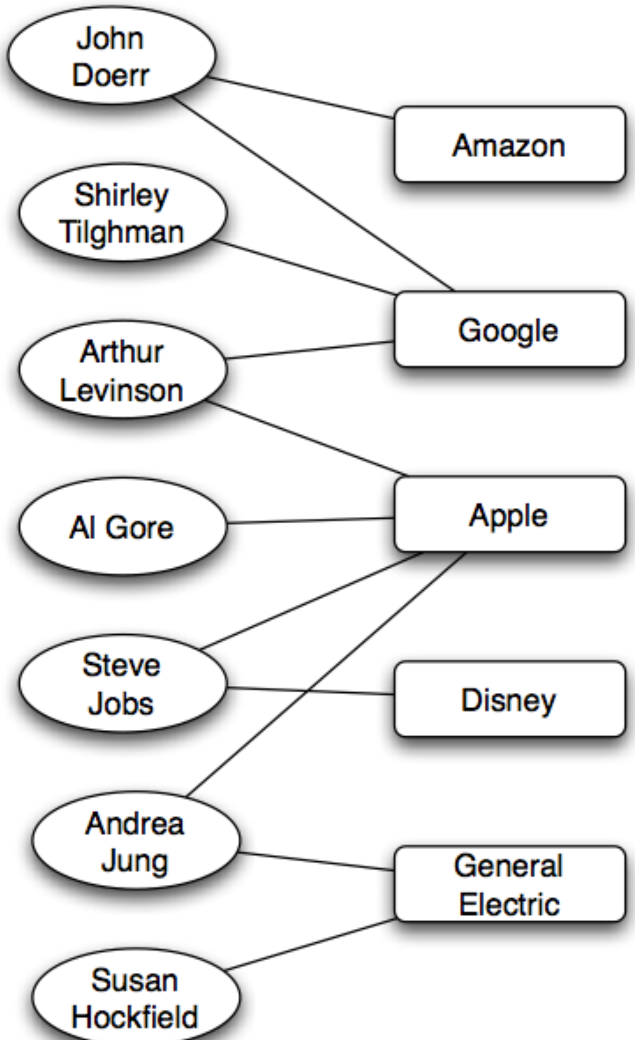
If we are given a model of graphs

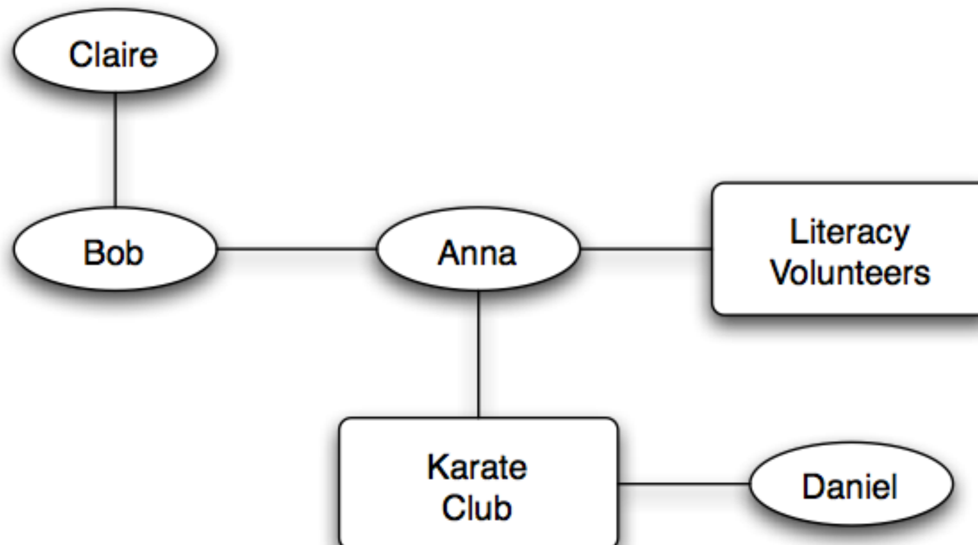
- Clustering is considered significant if
- CC is bounded from below by a constant
  - E.g. cc(G) > 0.1
  - Note that cc(G) > 1/n does not help, since this can be very small

Example problems:

- What can you say about CC of Trees?
- Complete graphs?
- Grids?
- Grids with diagonals added?

Social foci: affiliation networks





Triadic closure in affiliation networks

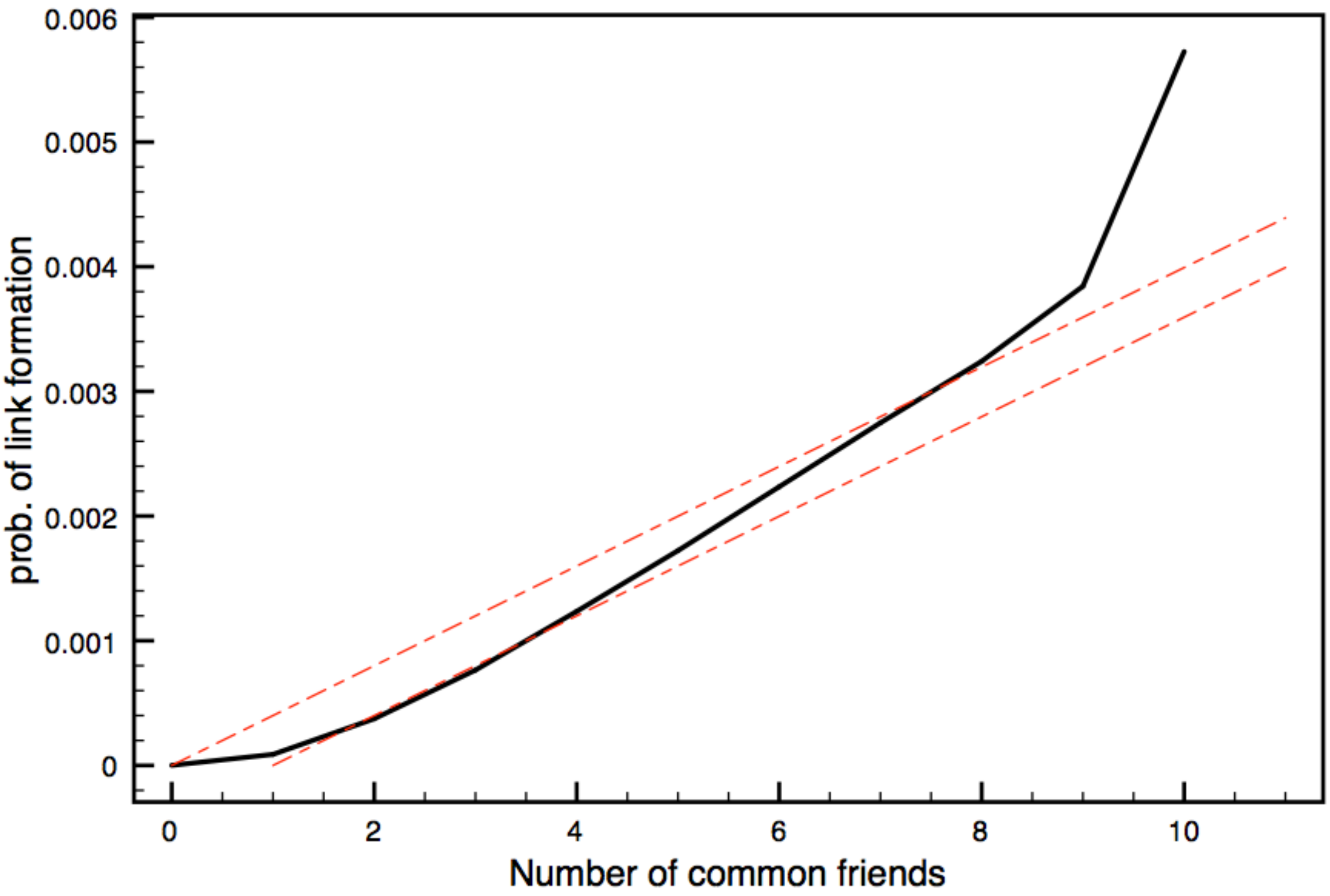
(i) Bob introduces Anna to Claire.

(ii) Karate introduces Anna to Daniel.

(iii) Anna introduces Bob to Karate.

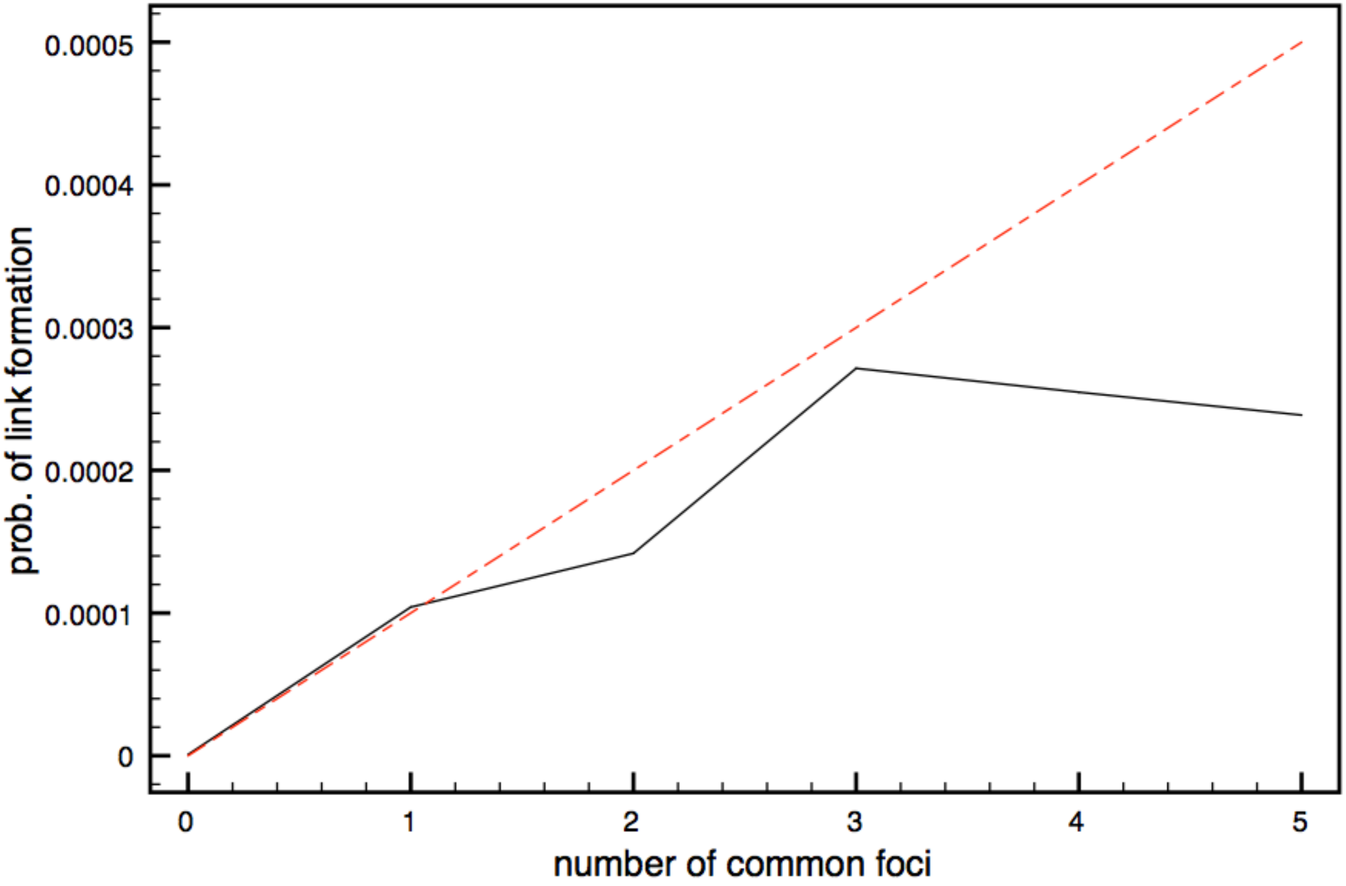
Triadic Closures

- From student email dataset

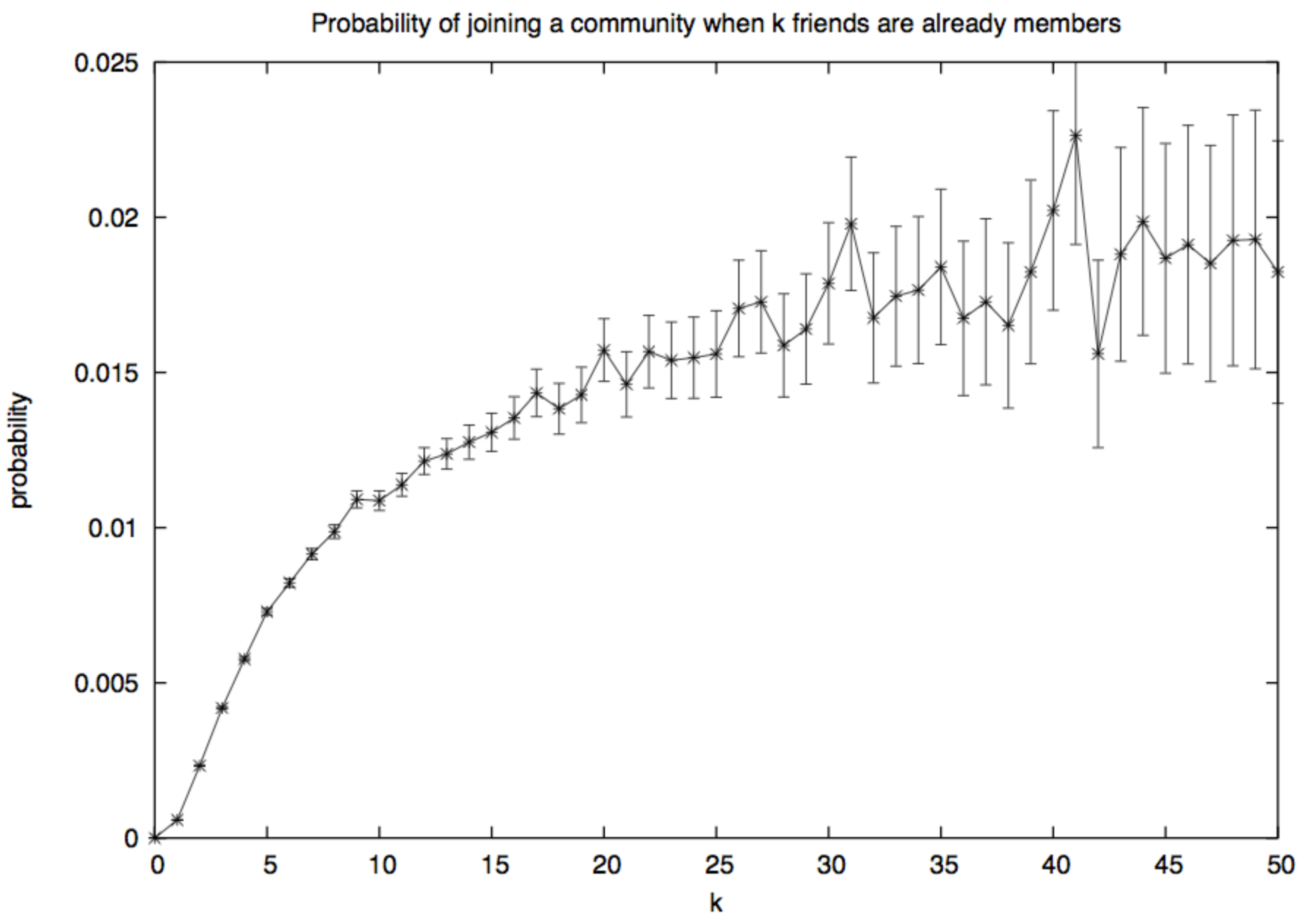


Focal closure

- Classes as foci

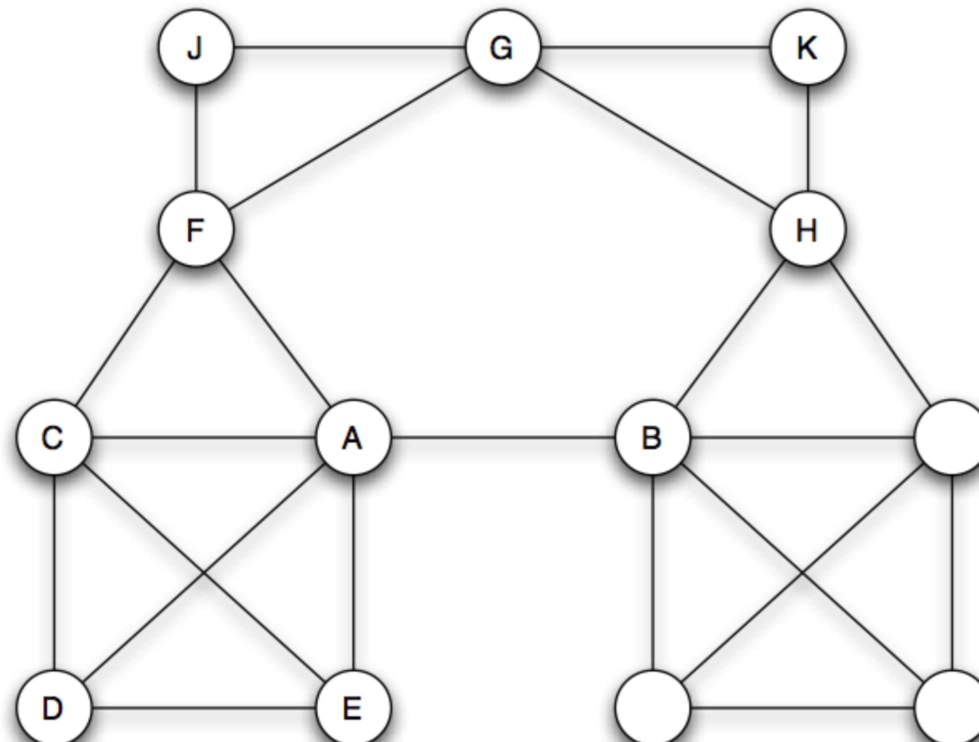


Membership closure



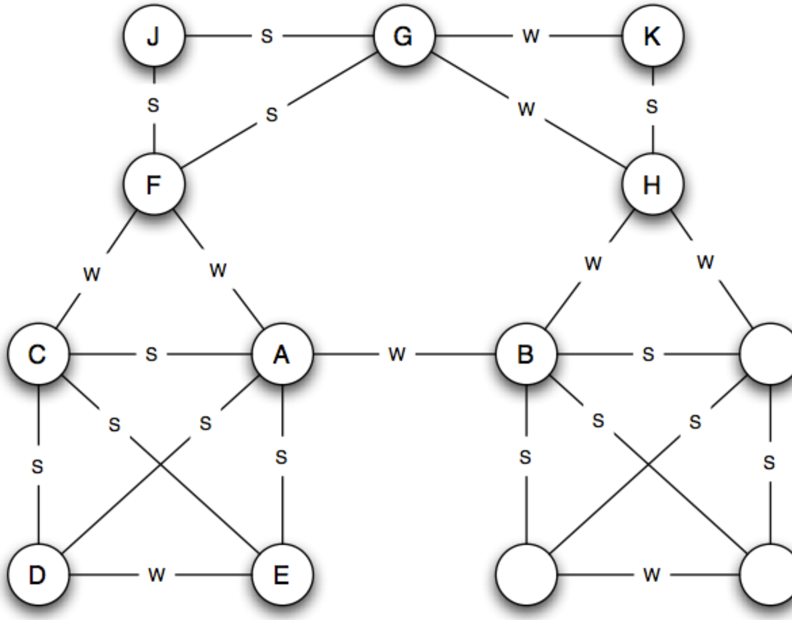
Bridges

- Bridge: Removing a bridge will disconnect network
  - Rare in real networks
- Local bridge (A, B): If A, B have no friends in common
  - Deleting (A, B) will increase distance to $d > 2$
  - d Is called the *span* of the bridge (A, B)



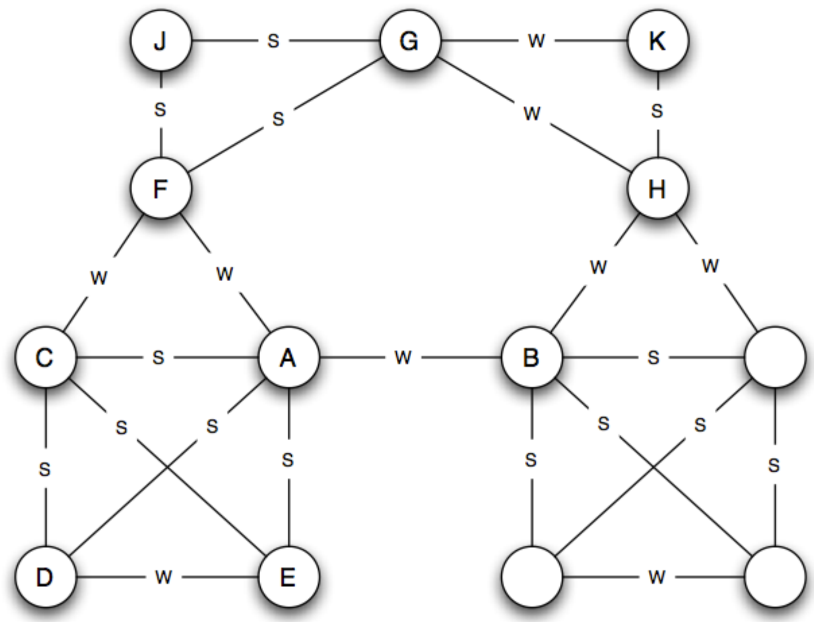
Strong triadic closure

- Suppose we know some ties to be strong, some to be weak
  - For some definition of strong/ weak
- Strong triadic closure: If ab and bc are strong, then edge ac exists (may be weak, but it is there)



Strong triadic closure

- Theorem: if a network satisfies strong triadic closure and node A has ≥ 2 strong ties then any bridge involving A must be a weak tie.
- Proof: Easy!



- In real world, triadic closure is reasonably important
  - Many examples
  - People want their friends to be friends (otherwise it is hard to have groups)
  - Absence of triadic closure implies poor relation between friends, stress etc

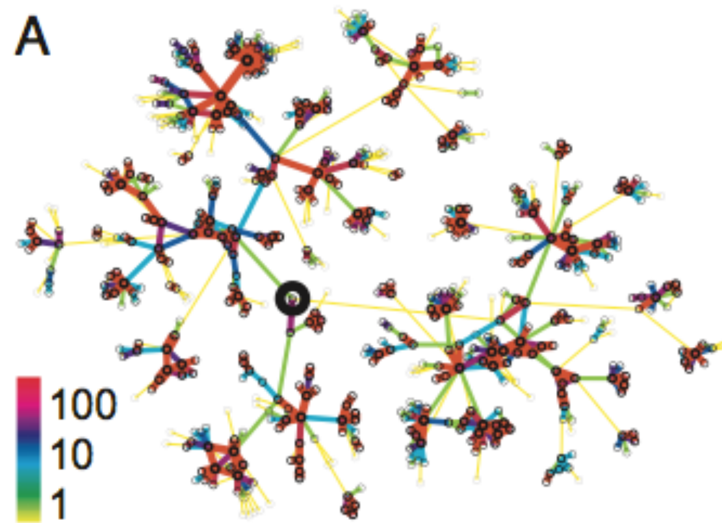
An experiment: Cell phone social net

- Network from phone conversations
- 18 weeks of all mobile calls for ~20% of US population, 90% had a mobile phone
- link: at least 1 reciprocating call.
- tie strength : aggregated duration of calls

- Onella et al. Structure and tie strengths in mobile communication networks. PNAS 2007

Observations

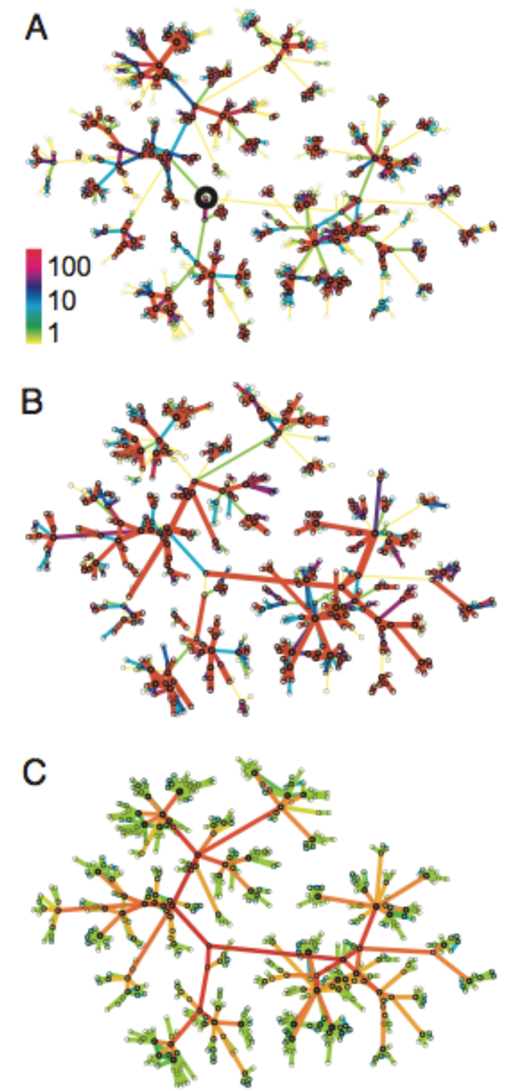
- Most people talk to few others, few talk to many people
  - Power law-like distribution
  - “Hubs” are relatively rare
- Strong ties are within clusters



- Onella et al. Structure and tie strengths in mobile communication networks. PNAS 2007

Possible network structures

- Efficiency: Inter-cluster ties are strong
  - Eg. Highways, Internet routers, water distribution, etc, to allow large flows (C)
- Dyadic: tie strength depends on individual relationship only
  - Simulated as random(B)
- Strength of weak ties (A)
  - Opposite of c
  - Argument: Social Information does not have a conservation requirement like transport or water



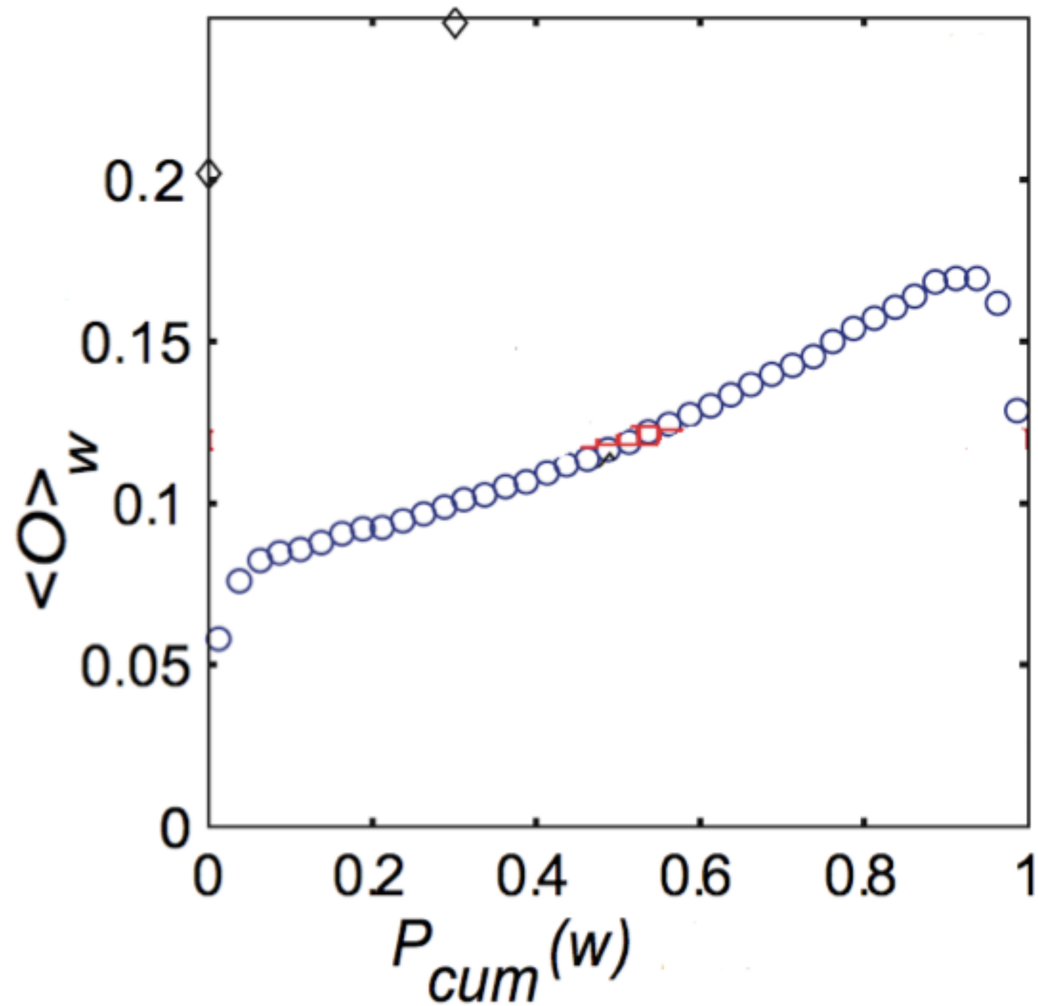
Other observations

- When strong ties are removed, network degrades slowly, but remains largely connected
- When the weak ties are removed, the network quickly and suddenly (phase transition) falls apart. i.e disconnects into chunks
- Experiment: Spread a rumor in this network. Anyone having the rumor is likely to transmit probabilistically: ie. More likely in a longer conversation
  - Observation: In majority of cases, people learn of it through ties of *intermediate strength*.

Neighborhood based estimate tie strength

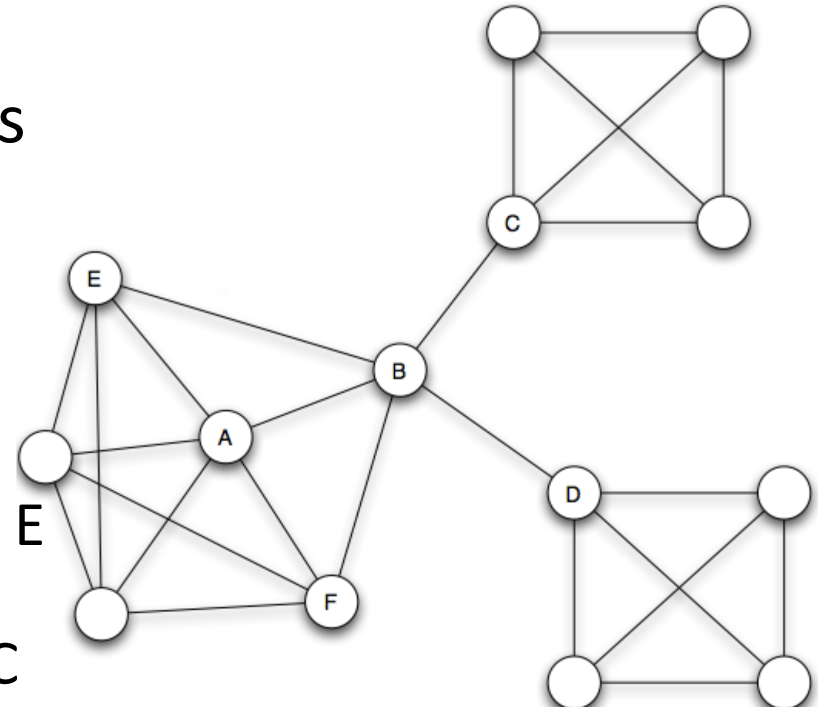
- When we do not have a real observation for tie strength
- $N_r(p)$: neighborhood of r hops centered at p. Sometimes written as $B_r(p)$
  - $N(p) = N_1(p)$
- Neighborhood overlap of ab: $$\frac{|N(a) \cap N(b)|}{|N(a) \cup N(b)|}$$
  - A more continuous notion of strength
  - And derived from the network
  - Potential experiment : compare with other definitions of strengths
- Zero (or small, depending on definition of N) when ab is a local bridge

Neighborhood overlap Vs phone call duration

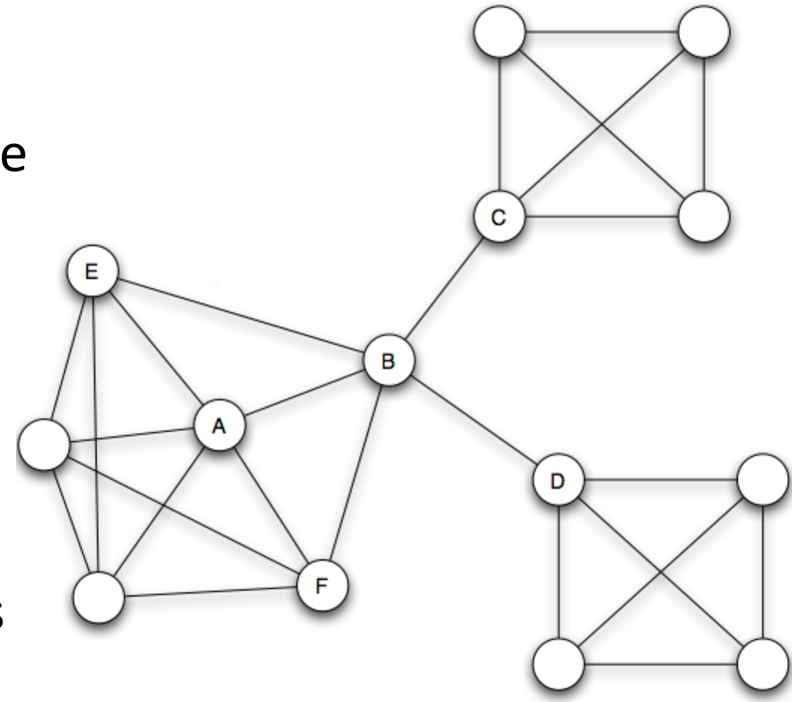


Embeddedness of an edge

- The number of common friends
- Higher embeddedness implies more people monitoring the relation
  - B does not want to cheat A since E will no longer trust B
  - But B can sacrifice relation with C without losing any direct friend
- What is the embeddedness of a bridge?



Structural holes



- B is part of a bridge that spans a gap/hole in the network (called structural holes)

- B has early access to information from other parts of network
- Interesting ideas occur as synthesis of multiple topics
- B has control over what the group learns from c and d
- B has reason to not allow triangles to form
- On the other hand, B's relations are not so protected by embeddedness
- How people actually behave in such situations is not well understood
  - Tension between closure and *brokerage*

Social capital

- The ability to secure benefits by virtue of membership (and position) in social networks or other social structures

- Sometimes used as a property of a group

Betweenness centrality

- Bridges are “central” to the network
  - They lie on shortest paths
- Betweenness of edge (e) (or vertex (v)):
  - We send 1 unit of traffic between every pair of nodes in the network, and measure what fraction passes through e, assuming the flow is split equally among all shortest paths.

Other Centrality measures

- Degree centrality – nodes with high degree
- Pagerank
- Eigen vector centrality (similar to pagerank, but undirected graphs)

-

- Closeness centrality
  - Average distance to all other nodes

$$\ell_x = \frac{1}{n}\sum_y d(x,y)$$

    - Decreases with centrality
    - Inverse is an increasing measure of centrality

$$C_x = \frac{1}{\ell_x} = \frac{n}{\sum d(x,y)}$$

k-core of a graph G

- A maximal connected subgraph where each vertex has a degree at least k
  - *Inside that subgraph.*
- Obtained by repeatedly deleting vertices of degree less than k