Social and Information Network Analysis
Strong and Weak Ties

Cornelia Caragea

Department of Computer Science and Engineering
University of North Texas

Acknowledgement: Jure Leskovec

June 13, 2016

Announcements

- Class notes:
  - http://www.cse.unt.edu/~ccaragea/kdsin16.html
- Lab sessions: Wednesdays, June 15 and 22, 4:30pm - 6:00pm.
- Homework assignment available at:
  - http://www.cse.unt.edu/~ccaragea/kdsin16/assignments.html
- Project topic presentations: Monday June 20, 4:30pm - 6:00pm.
- Exam: Tuesday June 21, 4:30pm - 6:00pm.

To Recap

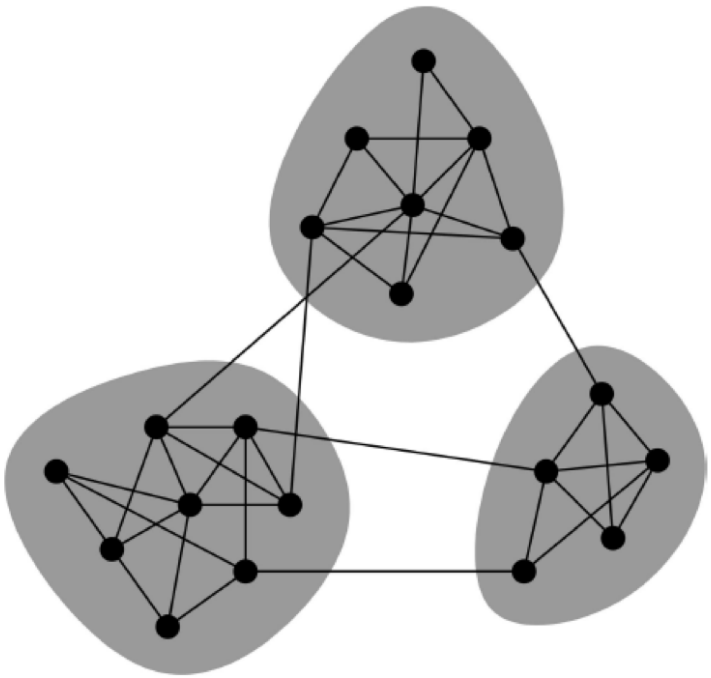
- What is the structure of a network?
- How to characterize a network structure?
  - Connectivity
  - Path length
  - Small world phenomenon
  - Node degree and degree distribution
  - Clustering coefficient

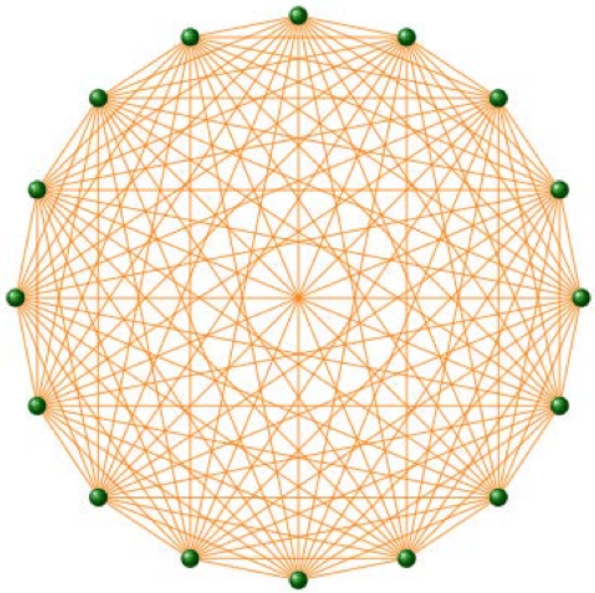
For Today's Class

- We look at some fundamental social network issues:
  - How information flows through a social network
  - How different nodes can play structurally distinct roles in this process
  - How these structural considerations shape the evolution of the network itself over time

Networks and Communities

- How do we often think of networks? “Looking" which way?





Networks: Flow of Information

- Networks play powerful roles in "bridging the local and the global"
  - Specifically, in offering explanations for how simple processes at the level of individual nodes and links can have complex effects on a population as a whole.

Motivating Question

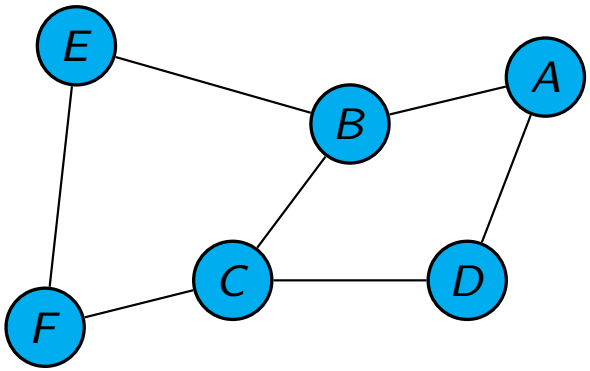
- How people find out about new jobs?
  - Mark Granovetter, as part of his PhD in 1960s, interviewed people who had recently changed employers to learn how they discovered their new jobs
  - People find the information through personal contacts
- But: Contacts were often acquaintances rather than close friends
  - This is surprising: One would expect your close friends to help you out more than casual acquaintances
- Why is it that your more distant acquaintances are more helpful?

Granovetter's Answer

- Two perspectives on friendships:
  - Structural: Friendships span different parts of the network
  - Interpersonal: Friendship between two people is either strong or weak
- Understanding the answer offers a way of thinking about the architecture of social networks more generally
- But first, let us develop some general principles about social networks and their evolution

Triadic Closure

- Static structures - a snapshot of the nodes and edges at a particular moment in time
- Dynamic structures - How does a network evolve over time?
  - How nodes arrive and depart
  - How edges form and vanish



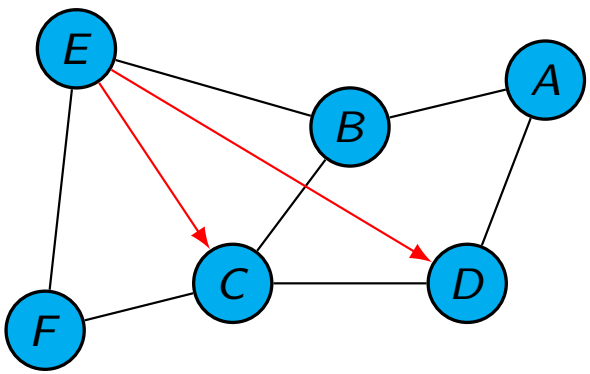
Which edge is more likely $E - C$ or $E - D$?

Triadic closure:

If two people in a network have a friend in common, there is an increased likelihood they will become friends themselves at some point in the future.

Triadic Closure

- Static structures - a snapshot of the nodes and edges at a particular moment in time
- Dynamic structures - How does a network evolve over time?
  - How nodes arrive and depart
  - How edges form and vanish



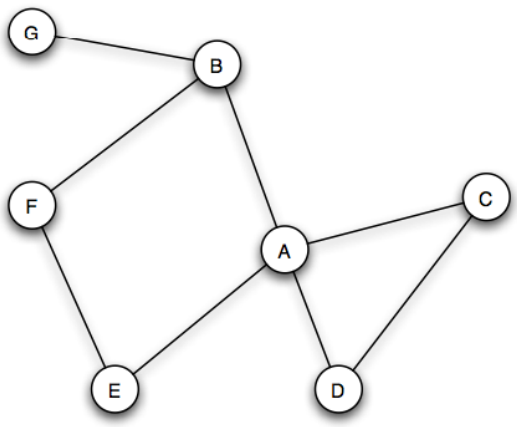
Which edge is more likely $E - C$ or $E - D$?

Triadic closure:

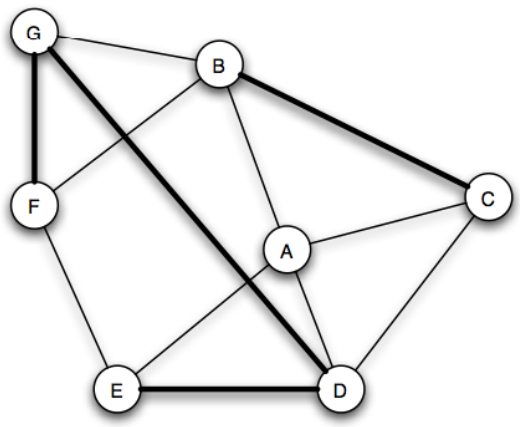
If two people in a network have a friend in common, there is an increased likelihood they will become friends themselves at some point in the future.

Triadic Closure

- If we observe snapshots of a social network at two distinct points in time, then in the later snapshot, we generally find a significant number of new edges that have formed through this triangle-closing operation, between two people who had a common neighbor in the earlier snapshot.



(a) *Before new edges form.*



(b) *After new edges form.*

Triadic Closure and Clustering Coefficient

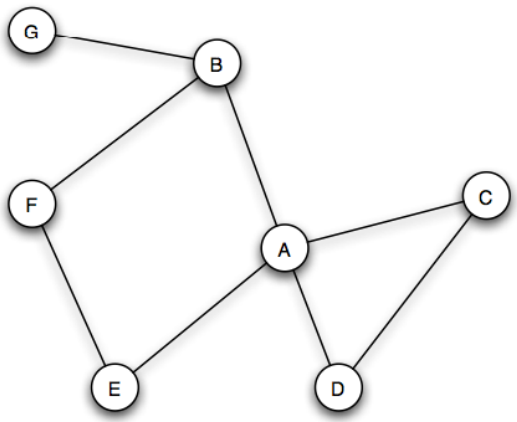
Triadic closure - high connection with the clustering coefficient!

- Remember that the clustering coefficient of a node A is defined as the fraction of pairs of A's friends that are connected to each other by edges.

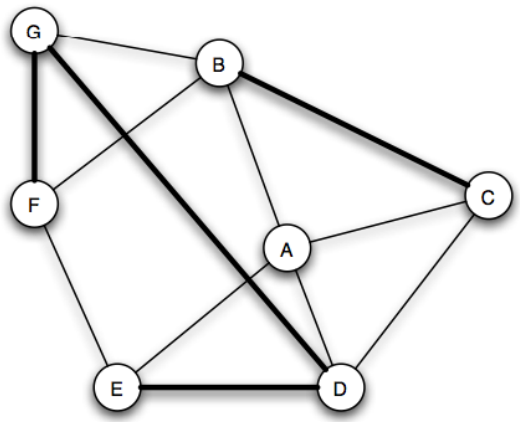
Triadic Closure and Clustering Coefficient

Triadic closure - high connection with the clustering coefficient!

- Remember that the clustering coefficient of a node A is defined as the fraction of pairs of A's friends that are connected to each other by edges.
- What is the clustering coefficient of node A in (a) and in (b)?



(a) *Before new edges form.*

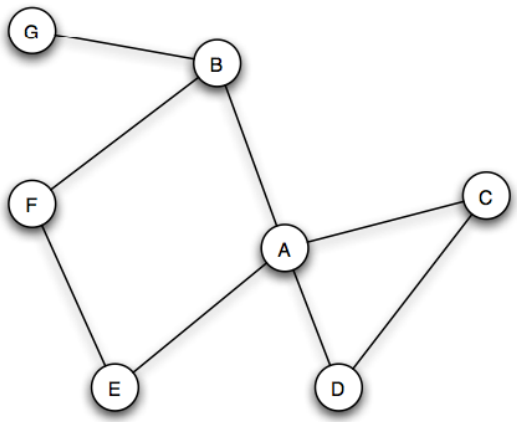


(b) *After new edges form.*

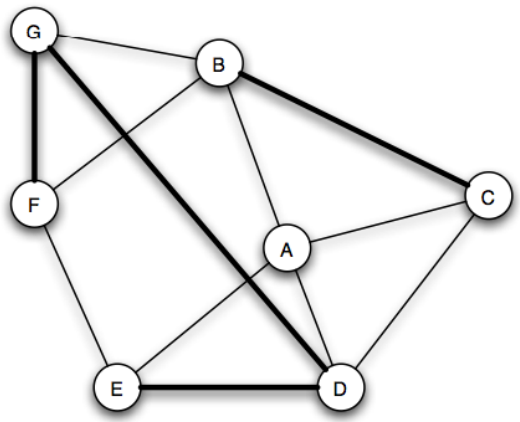
Triadic Closure and Clustering Coefficient

Triadic closure - high connection with the clustering coefficient!

- Remember that the clustering coefficient of a node A is defined as the fraction of pairs of A's friends that are connected to each other by edges.
- What is the clustering coefficient of node A in (a) and in (b)?



(a) *Before new edges form.*



(b) *After new edges form.*

- 1/6 in (a) and 1/2 in (b).

Triadic Closure and Clustering Coefficient

Triadic closure - high connection with the clustering coefficient!

- The more strongly triadic closure is operating in the neighborhood of the node, the higher the clustering coefficient will be.

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) - opportunity
  - B and C trust each other (since they have a friend in common) - trust
  - A has incentive to bring B and C together (sometimes it is hard to maintain two disjoint relationships) - incentive
- Empirical study on public-health data by Bearman and Moody:
  - Teenage girls with low clustering coefficient are more likely to contemplate suicide.

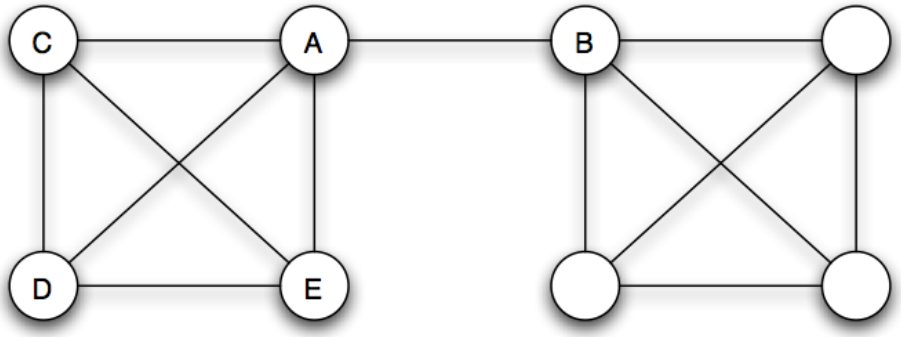
Back to Granovetter's Question

How does triadic closure relate to Granovetter's question?

- Assume that information about good jobs is relatively scarce
  - Hearing about a promising job opportunity from someone suggests that they have access to a source of useful information that you don't.
- What connects you to these people?
  - Bridges and local bridges!

Bridges

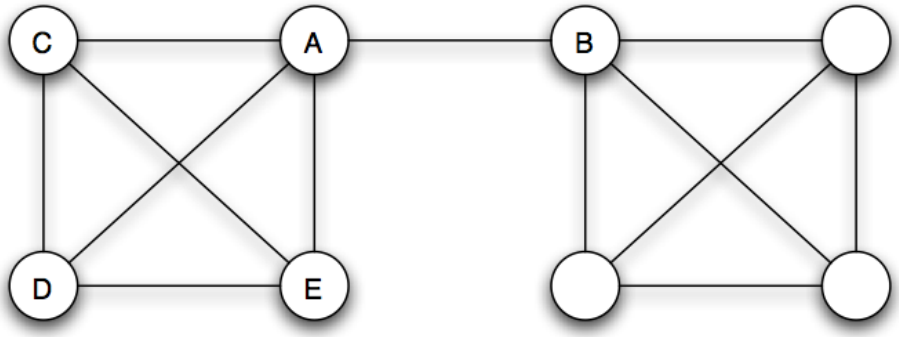
Bridge edge: If removed, it disconnects the graph.



- A has 4 friends C, D, E and B
- Structural differences
  - A, C, D, E form a tightly-knit group of friends who all know each other (similar opinions and sources of information)
  - The link to B seems to reach into a different part of the network and may offer access to information that A would not necessarily hear about.

Bridges

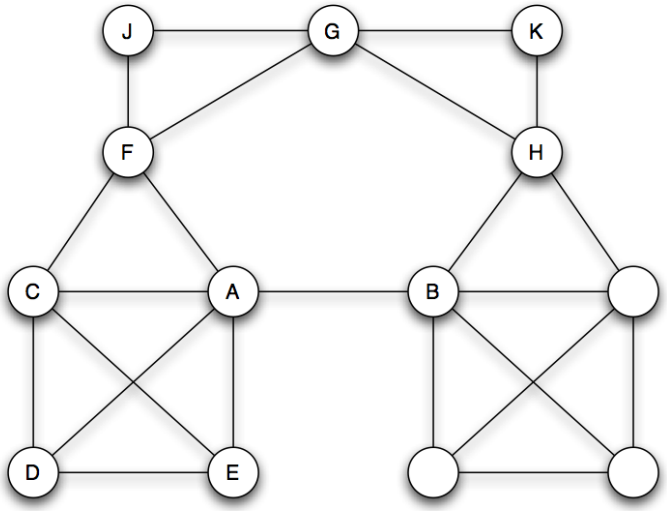
Bridge edge: If removed, it disconnects the graph.



- A has 4 friends C, D, E and B
- Structural differences
  - A, C, D, E form a tightly-knit group of friends who all know each other (similar opinions and sources of information)
  - The link to B seems to reach into a different part of the network and may offer access to information that A would not necessarily hear about.
- Bridges are extremely rare in real social network data!

Local Bridges

- The local network embedded into a larger social network



  - A-B edge isn't the only path that connects them
  - Local bridge edge:
    - If removed, it increases the distance between A and B to a value strictly more than two (A and B have no friends in common).
  - Span of a local bridge is the distance its endpoints would be from each other if the edge were deleted.

Local Bridges and Triadic Closure

Connection between Local Bridges and Triadic Closure

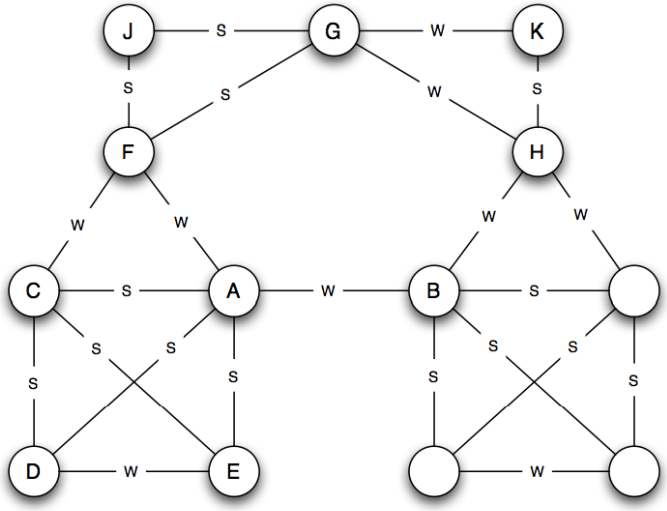
- Two opposite concepts:
  - An edge is a local bridge precisely when it does not form a side of any triangle in the graph.

Granovetter's Explanation

- The truly new information might come unusually often (though certainly not always) from a friend connected by a local bridge
- The closely-knit groups that you belong to, though they are filled with people eager to help, are also filled with people who know roughly the same things that you do
- How do we quantify the level of strength in the links of a social network?
  - Stronger links represent closer friendship and greater frequency of interaction
    - Strong ties (the stronger links, corresponding to friends)
    - Weak ties (the weaker links, corresponding to acquaintances)

The Strong Triadic Closure Property

- If the nodes annotate the network with strong and weak ties, we can get:

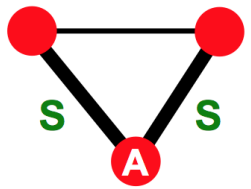


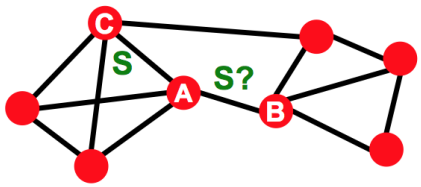
- What can we say about the triadic closure?
  - The Strong Triadic Closure Property: *We say that a node A violates the Strong Triadic Closure Property if it has strong ties to two other nodes B and C, and there is no edge at all (either a strong or weak tie) between B and C. We say that a node A satisfies the STCP if it does not violate it.*

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
  - Let A-C another strong tie
  - Then B-C must exist because of Strong Triadic Closure
  - But then A-B is not a local bridge! Contradiction.





The Strength of Weak Ties

- The argument made completes the connection between the local property of tie strength and the global property of serving as a local bridge
  - It provides a way to think of how interpersonal properties of social-network links are related to the network's structure
  - It provides a concrete framework for thinking about information flow
    - Weak social ties connect us to new sources of information and new opportunities
  - Their dual role as weak connections but also valuable conduits to hard-to-reach parts of the network is the surprising strength of weak ties

Tie Strength and Network Structure in Large-Scale Data

- The connection between the tie strength and structural properties of an underlying social network make intriguing theoretical predictions about the organization of social networks in real life
- For many years the Granovetter's theory was not tested due to the difficulty in finding data that reliably captured the strengths of edges in large-scale, realistic settings
- But, today we have large "who-talks-to-whom" graphs:
  - Email, Messenger, Cell phones, Facebook

Tie Strength and Network Structure in Large-Scale Data

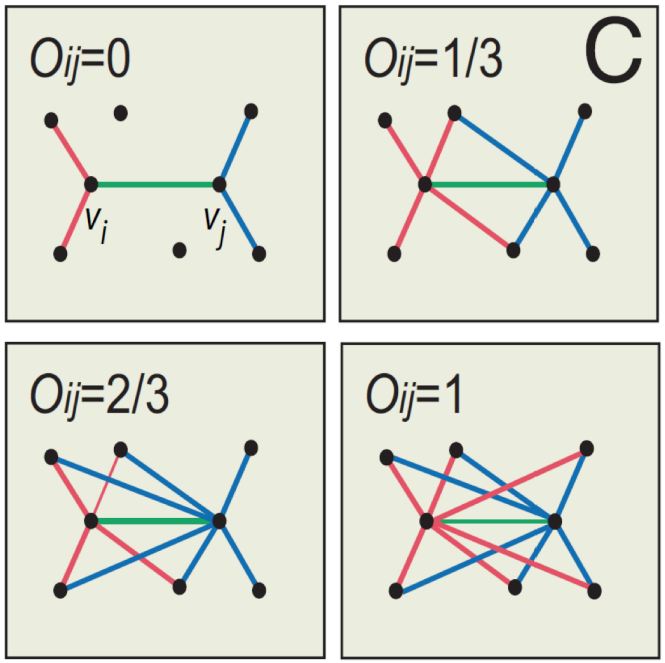
- Onnela et al. 2007:
  - Cell-phone network of 20% of a country's population
  - The nodes correspond to cell-phone users, there is an edge joining two nodes if they made phone calls to each other in both directions over an 18-week observation period.
  - Edge strength: # phone calls
  - The networks exhibits a giant component, a single connected component containing most (in this case 84%) of the individuals in the network
- The strong assumption made so far: an edge is either a strong tie or a weak tie, and it is either a local bridge or it is not
- Generally, a very small fraction of the edges in the cell-phone data are local bridges
- To examine real data at a large scale, it is important to soften this assumption and define "almost" local bridges

Neighborhood Overlap

- Edge overlap:

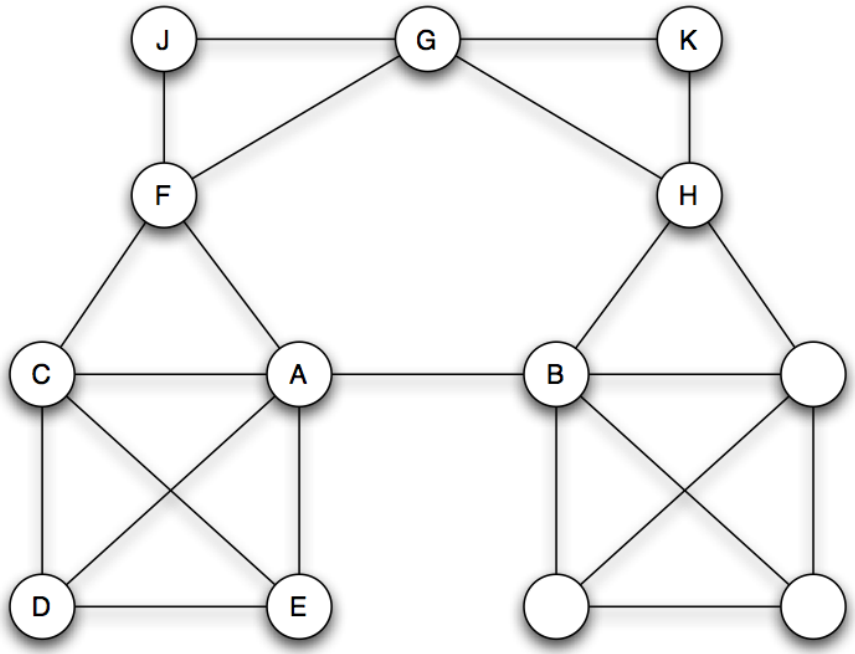
$$O_{ij} = \frac{N(i) \cap N(j)}{N(i) \cup N(j)}$$

  - $N(i)$ is a set of neighbors of node i.



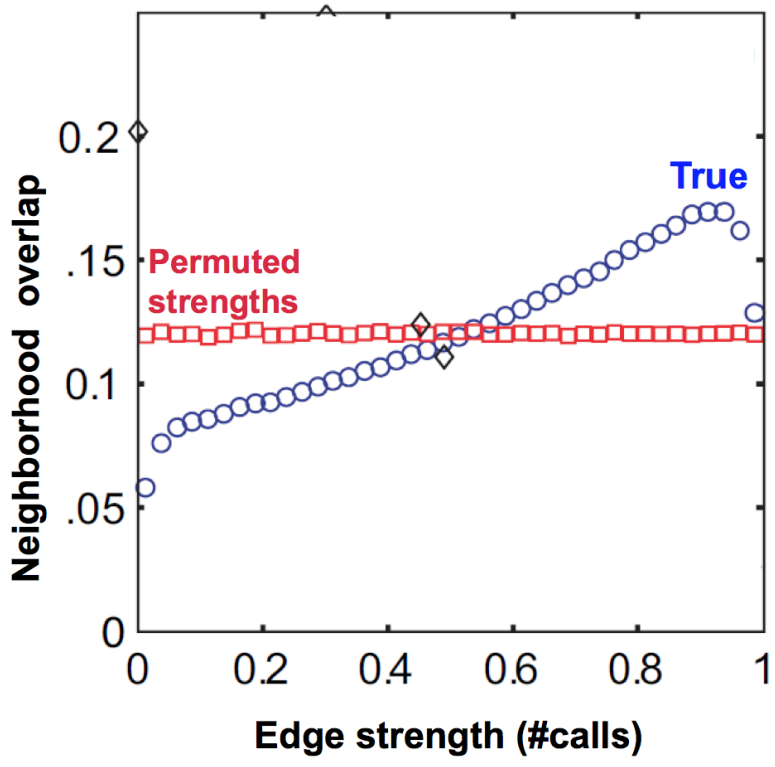
Several Questions

- Is the local bridge contained in the neighborhood overlap?
- What is the neighborhood overlap of A-F?
- Which edge is closer to being a local bridge A-F or A-E?



Empirical Results on the Cell-Phone Data

- Onnela et al's question #1: How does the neighborhood overlap of an edge depend on its strength?

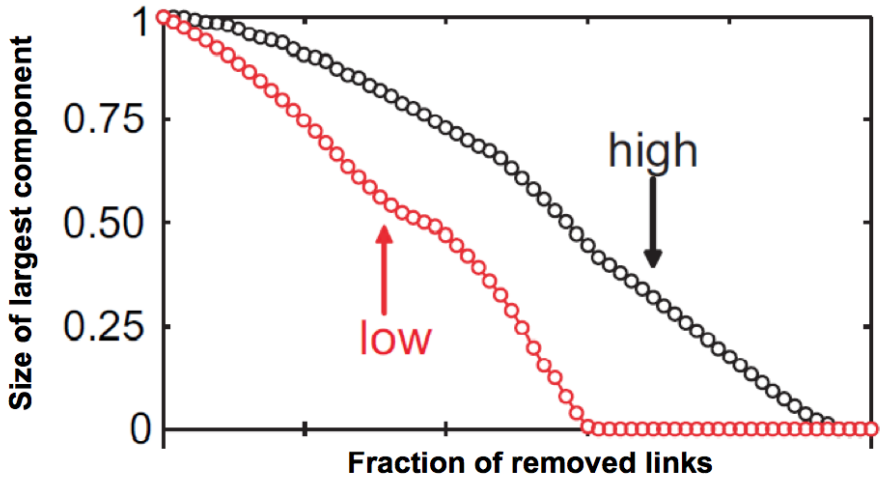


- Observation: Highly used links have high overlap!
- Legend: True: The data; Permuted strengths: Keep the network structure but randomly reassign edge strengths.

Empirical Results on the Cell-Phone Data

- Onnela et al's question #2: Do weak ties serve to link together different tightly-knit communities that each contain a large number of stronger ties?
- Deleted edges from the network one at a time:
  - Starting with the highest overlap (strongest ties) and working downward in order of overlap (tie strength).
  - Starting from the lowest overlap (weakest ties) and working upward in order of overlap (tie strength).

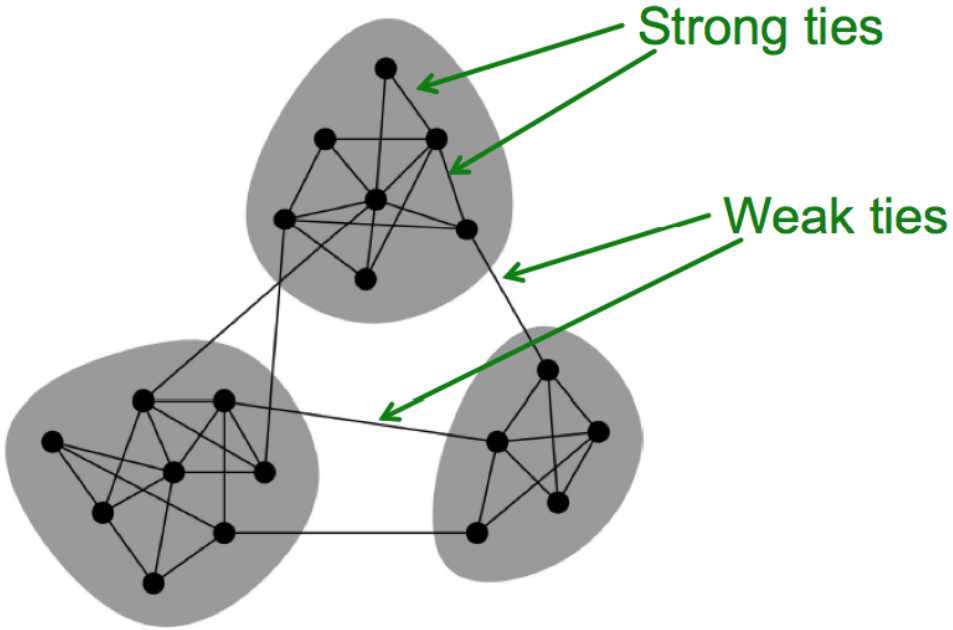
Link Removal by Overlap



- Removing links based on overlap
  - Low to high
  - High to low
- Low disconnects the network sooner.
- Link removal based on overlap broke apart abruptly once a critical number of weak ties had been removed
- Similar results can be obtained for link removal by strength (#calls).

Conceptual Picture of Networks

- Granovetter's theory leads to the following conceptual picture of networks



Tie Strength on Social Media

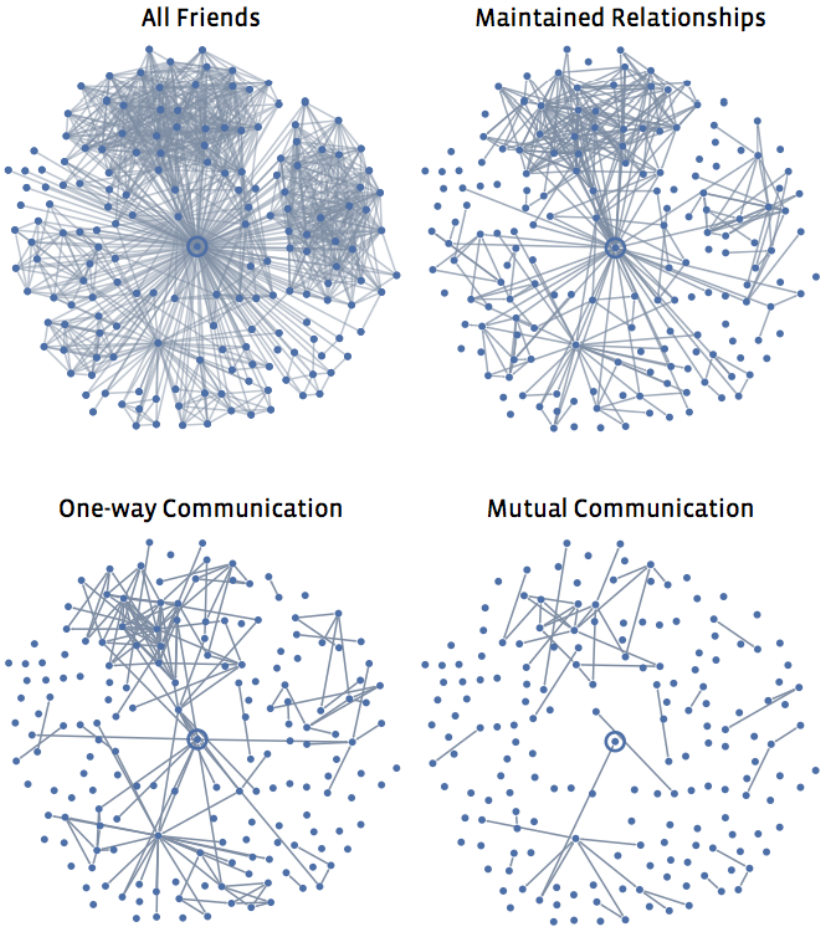
- People maintain large explicit lists of friends in their profiles on these sites
- What effect does this have on social network structure more broadly?
  - Weak ties provide the more crucial connective structure for holding together disparate communities, and for keeping the global structure of the giant component intact
- Tie strength can provide a language for answering how on-line social activity is distributed across links of different strengths
  - Among hundreds of friendship links maintained by people on social-networking sites, how many of these correspond to strong ties that involve frequent contact, and how many correspond to weak ties that are activated relatively rarely

Tie Strength on Facebook

- Research on Facebook:
  - To what extent each link was actually used for social interaction, beyond simply being reported in the profile?
- Three types of links:
  - reciprocal (mutual) communication
  - one-way communication
  - maintained relationship, i.e., user followed information about the friend at the other end of the link, whether or not actual communication took place
    - "following information" = either clicking on content via Facebook's News Feed service (providing information about the friend) or visiting the friend's profile more than once

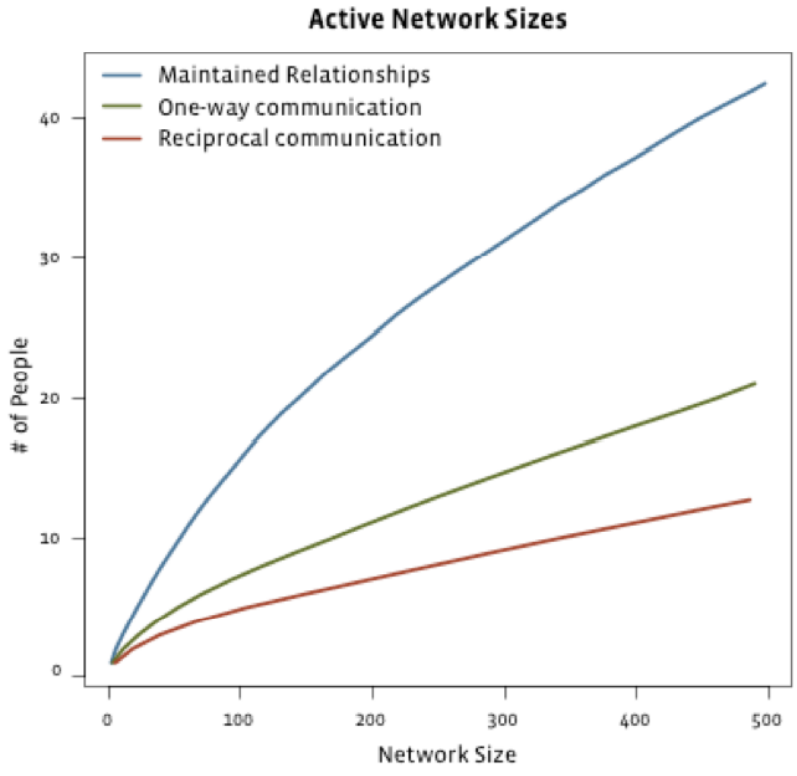
Tie Strength on Facebook

- Four views of a Facebook user's network neighborhood



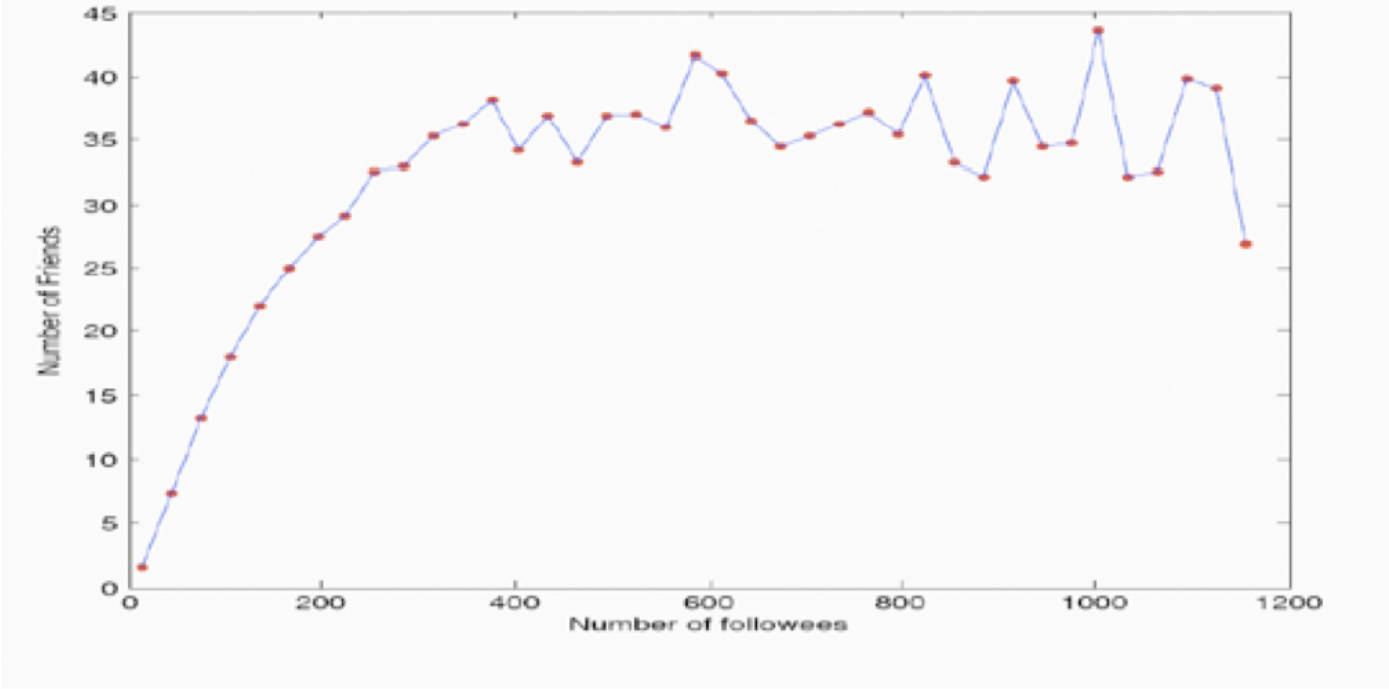
- The links becomes sparser moving from maintained relationships to reciprocal communication.

Tie Strength on Facebook - Aggregate View



- Even users with high neighborhood sizes actively communicate with $\approx$ 10 to 20 friends
- High *passive engagement*, one keeps up with friends by reading news about them even in the absence of communication
  - "... news, e.g., baby or engagement, can propagate very quickly through this highly connected network"

Tie Strength on Twitter



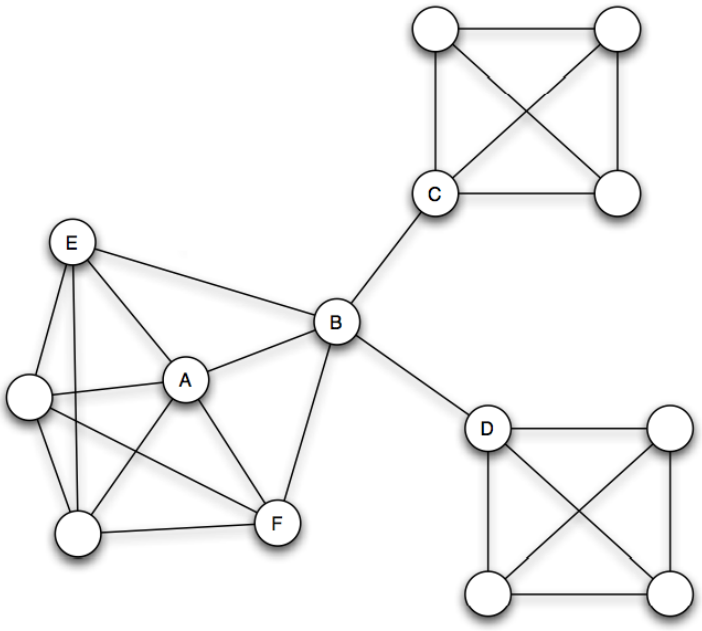
- The total number of a user's strong ties (defined by multiple directed messages) as a function of the number of followees he or she has on Twitter.

Tie Strength on Social Media

- Another useful way to think about the contrast between the ease of forming links and the relative scarcity of strong ties in environments like Facebook and Twitter is that:
  - Each strong tie requires the continuous investment of time and effort to maintain
  - Even people who devote a lot of their energy to building strong ties will eventually reach a limit - imposed by the hours in a day - on the number of ties that they can maintain
  - Weak ties do not need to be maintained continuously - easier for someone to accumulate them in large numbers

The Roles of Nodes in the Network Structure

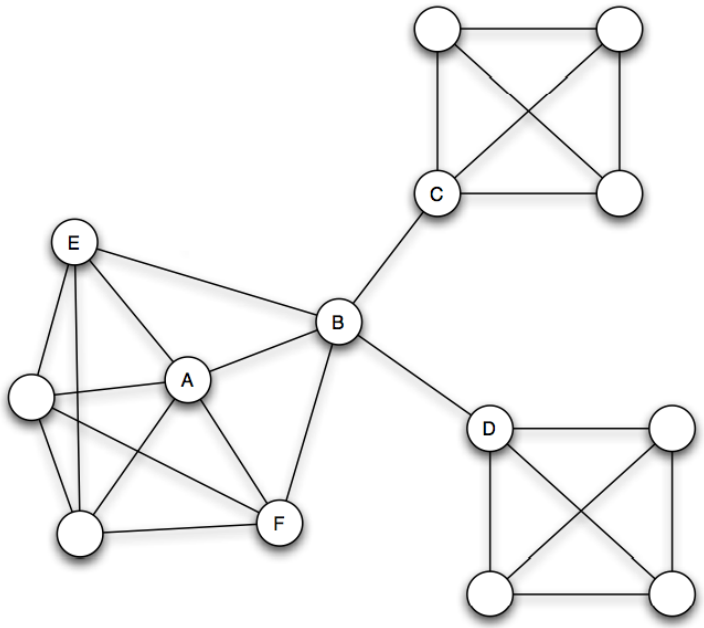
- We have now a general view of social networks in terms of tightly-knit groups and the weak ties that link them
  - We focused on the roles that different edges play in a network structure - with a few edges spanning different groups while most are surrounded by dense patterns of connections
- What about the roles that different nodes play in this structure?



- Some nodes are positioned at the interface between multiple groups, with access to boundary-spanning edges (B)
- Others are positioned in the middle of a single group (A)

The Embeddedness of an Edge

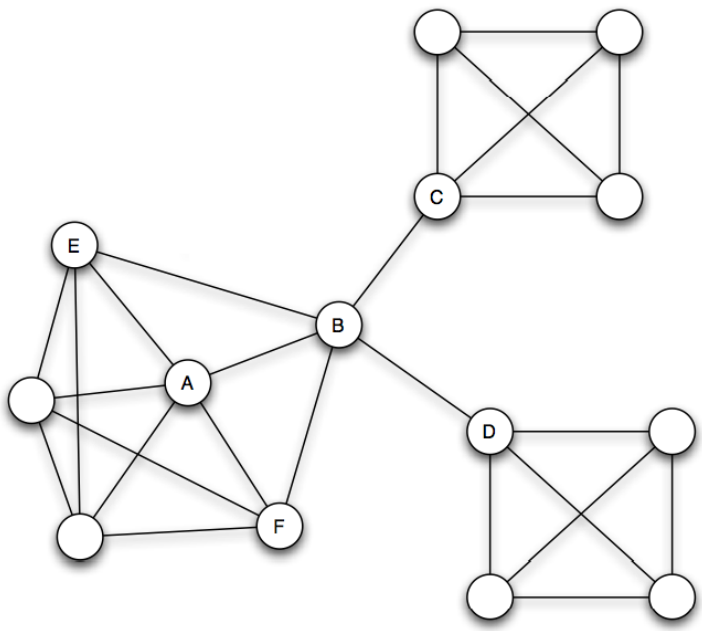
- The structure of a node in a network relates to the notion of *embeddedness of an edge*, defined as the number of common neighbors of the two endpoints.
- What is the embeddedness of A-B?



- Observations:
  - The embeddedness of an edge is equal to the numerator in the ratio that defines the neighborhood overlap
  - Local bridges are precisely the edges that have an embeddedness of zero

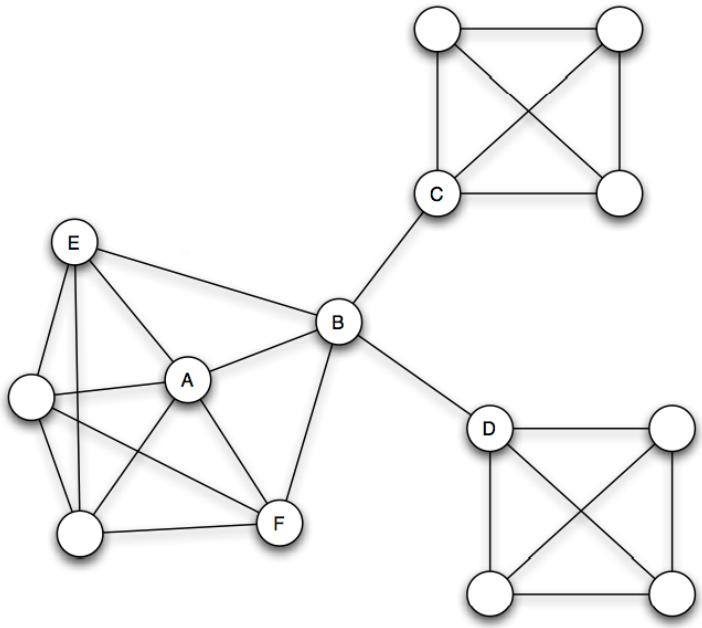
The Closure of a Node

- What can we say about node A?



The Closure of a Node

- What can we say about node A?



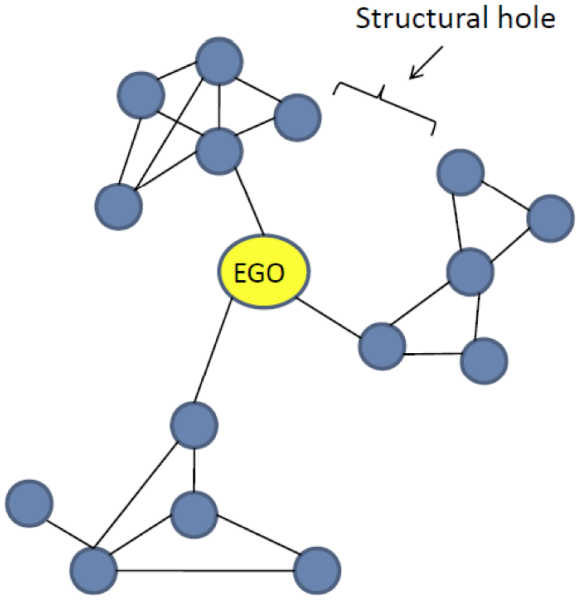
- Node A has a high clustering coefficient
- All of A's edges have a significant embeddedness

Remarks from Research in Sociology

- If two individuals are connected by an embedded edge
  - Easier for them to trust one another and have confidence in the integrity of transactions that take place between them
  - The presence of mutual friends puts the interactions between two people “on display" in a social sense
    - In case of misbehavior, there is the potential for social sanctions and reputational consequences from their mutual friends
- Granovetter writes:
  “My mortification at cheating a friend of long standing may be substantial even when undiscovered. It may increase when a friend becomes aware of it. But it may become even more unbearable when our mutual friends uncover the deceit and tell one another."

Structural Holes

- What can we say about node B?
  - Any advantages for sitting at ends of multiple local bridges?
  - Informally, we say that node B *spans* a structural hole in an organization
- Structural holes: the "empty space" in the network between two sets of nodes that do not otherwise interact closely



Many structural holes

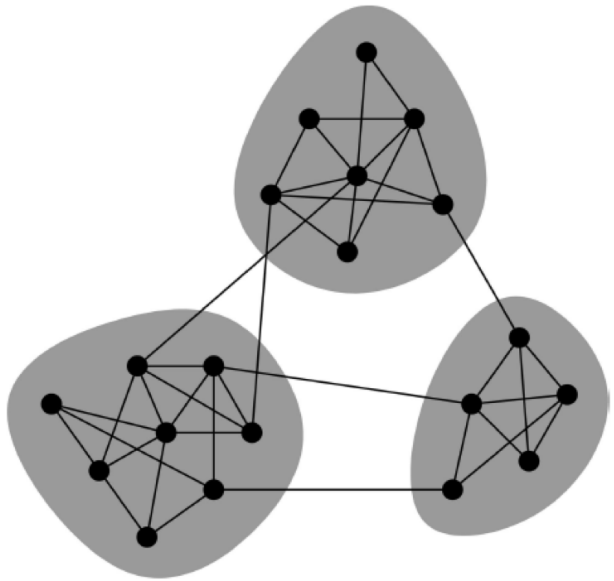
Remarks from Research in Sociology

Advantages of node B relative to node A:

- B has early access to information originating in multiple, non-interacting parts of the network
  - Anybody has a limited amount of energy they can invest in maintaining contacts across an organization
  - B is investing energy efficiently by reaching out to different groups rather than basing all B's contacts in the same group
- Standing at one end of a local bridge can be an amplifier for creativity
  - Innovations often arise from the unexpected synthesis of multiple ideas from distinct and unrelated bodies of expertise
  - B's position gives access to combined information from these groups, as well as the opportunity for novel ideas
- An opportunity for a kind of social "gate-keeping": regulates what information can flow in the network
  - A source of power

Network Communities

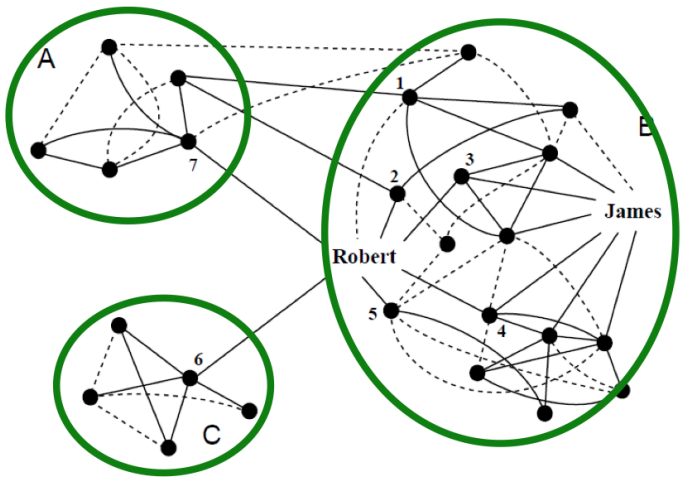
- Granovetter's theory (and common sense) suggest that networks are composed of tightly connected sets of nodes and weak ties that link them together

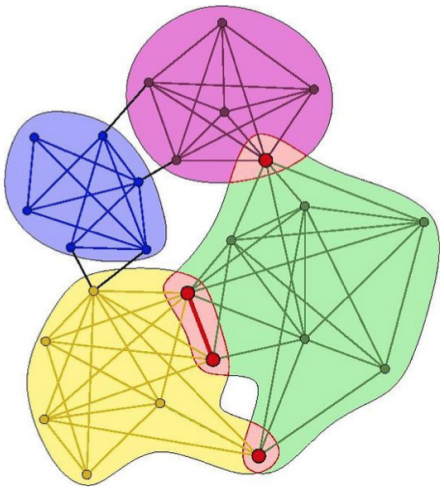


  - We define network communities as:
    - Sets of nodes with lots of connections inside and few to outside (the rest of the network)

Finding Network Communities

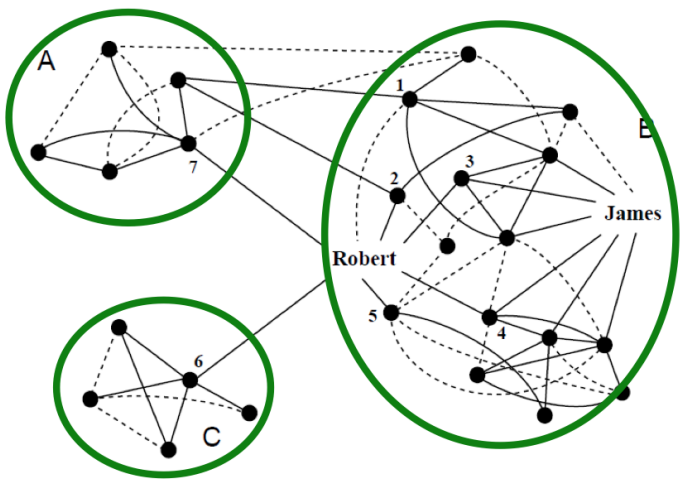
- How to automatically find densely connected groups of nodes?
- Ideally such automatically detected clusters would then correspond to real groups

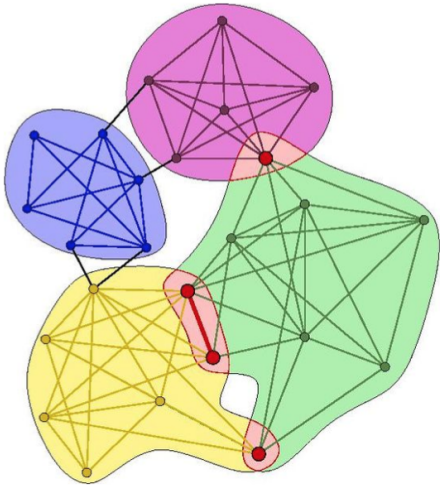




Finding Network Communities

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

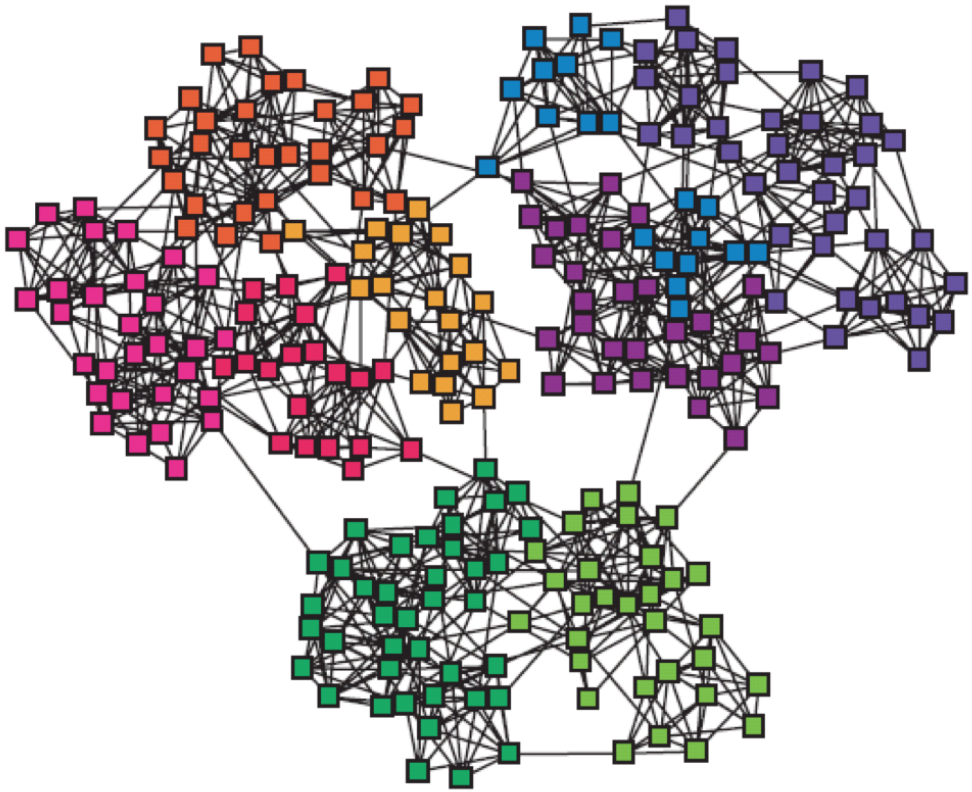




- Graph partitioning or community detection algorithms

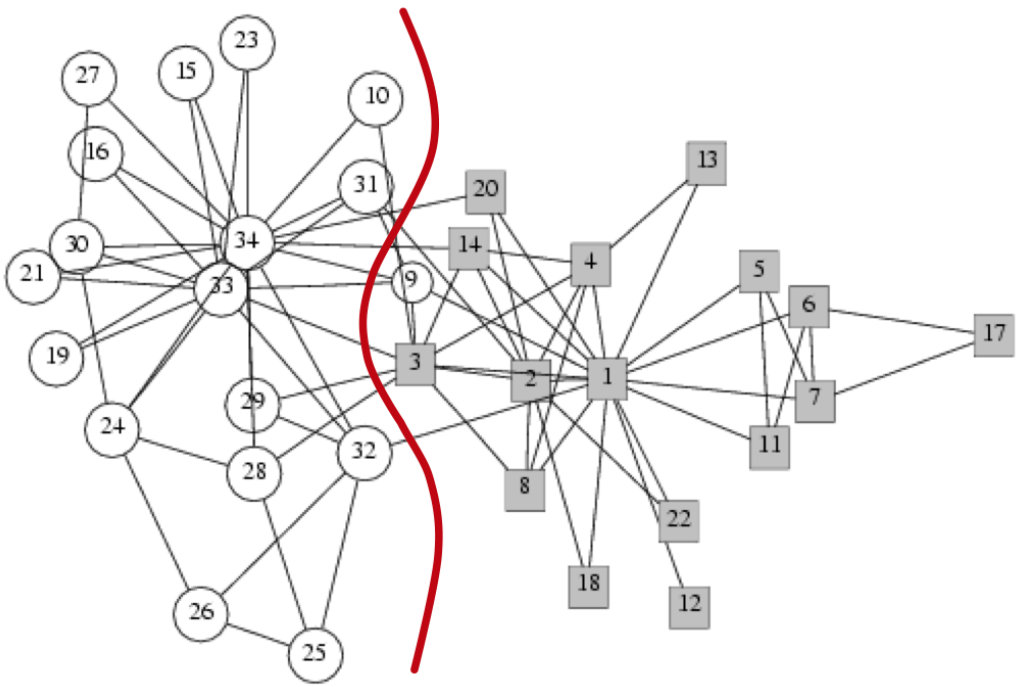
Motivation for Finding Network Communities

- Flow of information



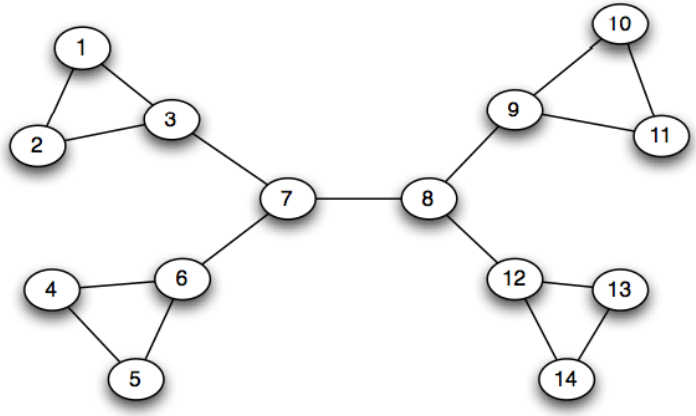
Motivation for Finding Network Communities

- Conflict identification
  - 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

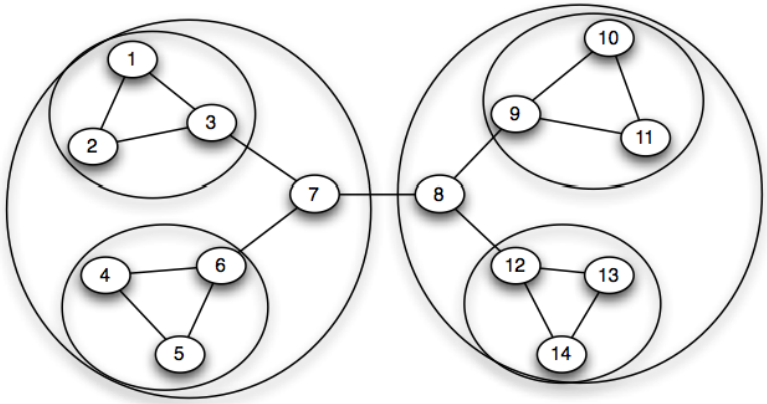


General Approaches to Graph Partitioning

- Divisive methods: identifying and removing the “spanning links" between densely-connected regions
  - Once links are removed, the network begins to fall apart into large pieces
- Agglomerative methods:
  - Finding nodes that are likely to belong to the same region and merging them together



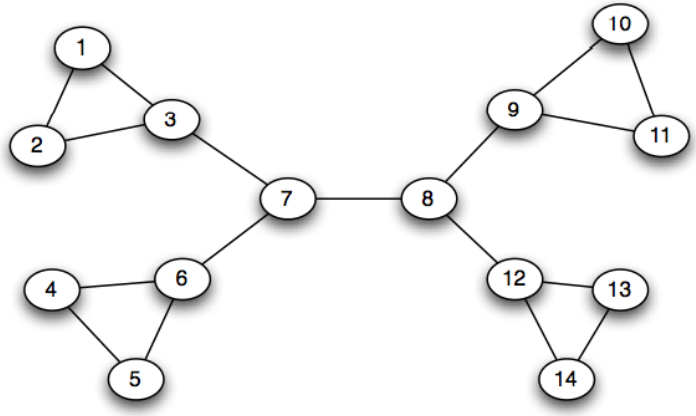
(a) *A sample network*



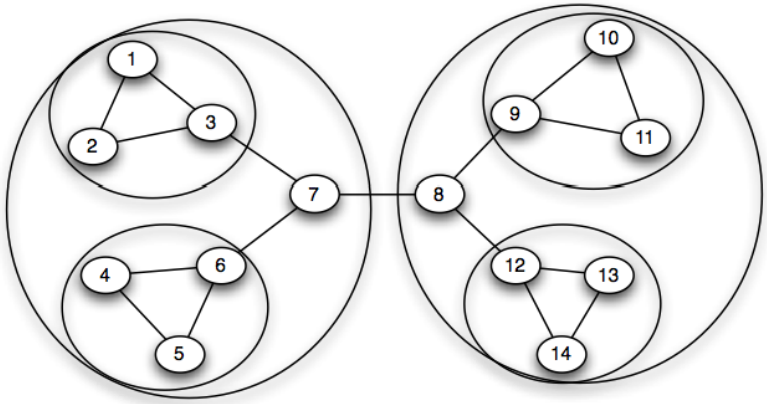
(b) *Tightly-knit regions and their nested structure*

General Approaches to Graph Partitioning

- Divisive methods: identifying and removing the "spanning links" between densely-connected regions
  - Once links are removed, the network begins to fall apart into large pieces
- Agglomerative methods:
  - Finding nodes that are likely to belong to the same region and merging them together



(a) *A sample network*



(b) *Tightly-knit regions and their nested structure*

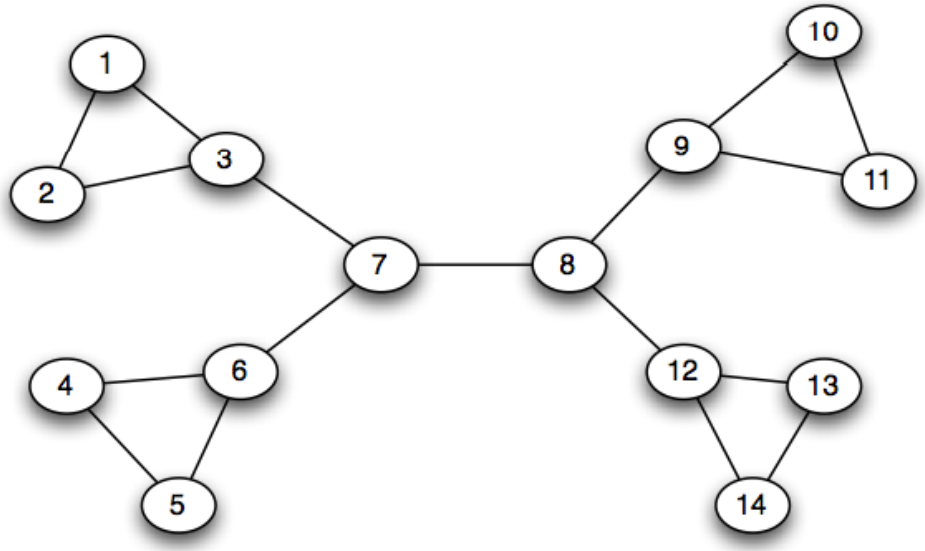
- Our focus: a divisive method.

A Divisive Method: Strength of Weak Ties

- Idea: Bridges and local bridges that connect weakly interacting parts of a network should be removed first
  - Local bridges form part of the shortest path between pairs of nodes in different parts of the network - without a particular local bridge, paths between many pairs of nodes may have to be “re-routed" a longer way
  - We define an abstract notion of “traffic" on the network, and look for the edges that carry the most of this traffic
- Betweenness of an edge: the total amount of flow it carries, counting flow between all pairs of nodes using this edge (if multiple shortest paths from A to B, the flow between A and B divides evenly among all paths)
  - More precisely, edge betweenness is the number of shortest paths passing over the edge

Betweenness of an Edge

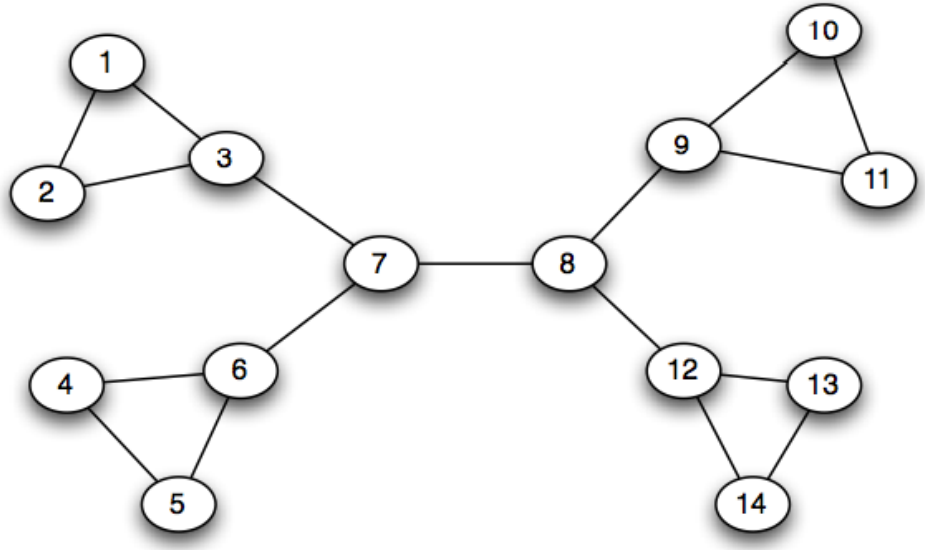
- What is the betweenness of edge 7-8?
- How about 3-7, 1-3, 1-2?



(a) *A sample network*

Betweenness of an Edge

- What is the betweenness of edge 7-8?
- How about 3-7, 1-3, 1-2?



(a) *A sample network*

- Betweenness 7-8: 49; 3-7: 33; 1-3: 12; 1-2:1

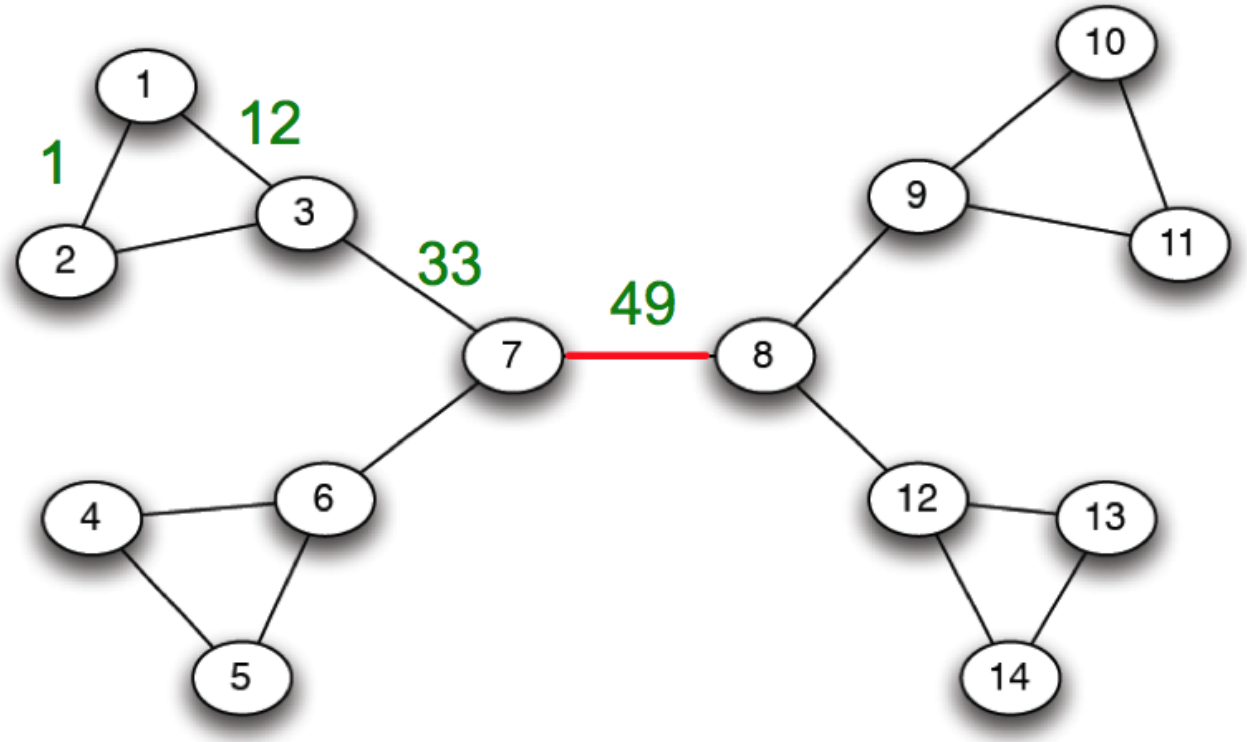
The Girvan-Newman Method

Successively Deleting Edges of High Betweenness

- Divisive hierarchical clustering based on the notion of edge betweenness (i.e., the number of shortest paths passing through the edge)
- Girvan-Newman Algorithm:
  - Repeat until no edges are left:
    - Calculate betweenness of edges
    - Remove edges with highest betweenness
  - Connected components are communities
  - Gives a hierarchical decomposition of the network

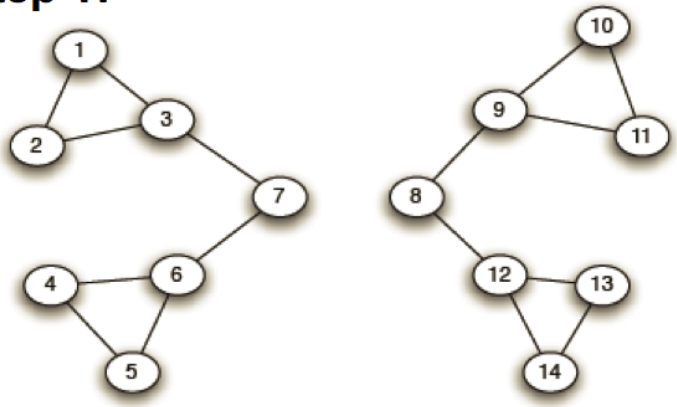
Girvan-Newman: Example

- Needs to re-compute betweenness at every step!

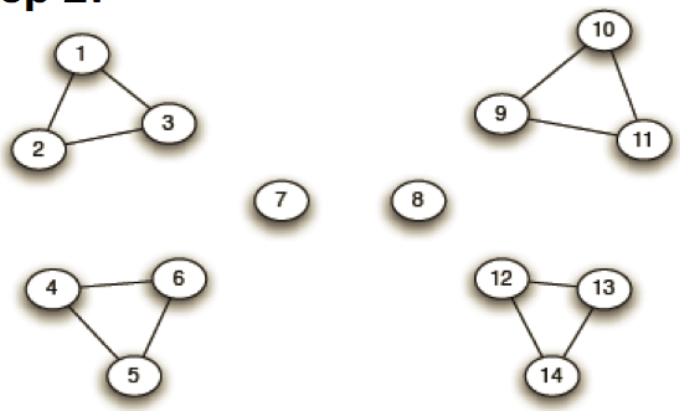


Girvan-Newman: Example

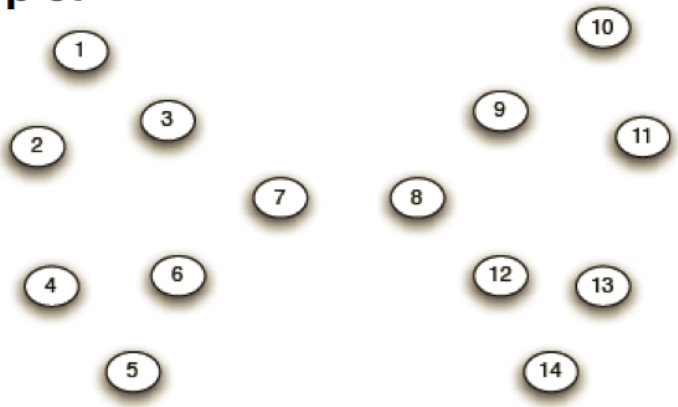
Step 1:



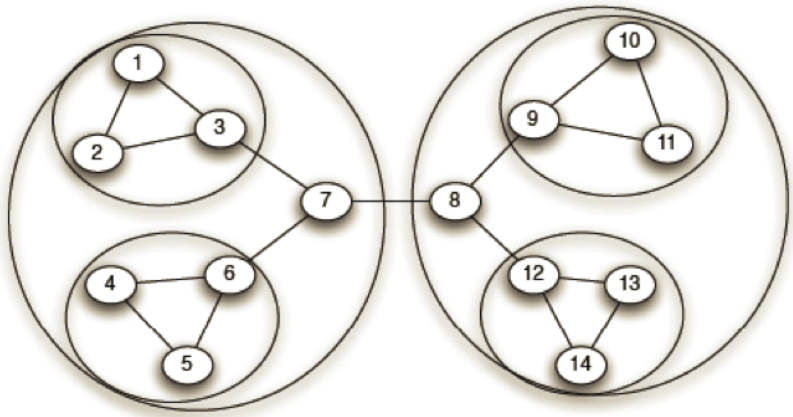
Step 2:



Step 3:

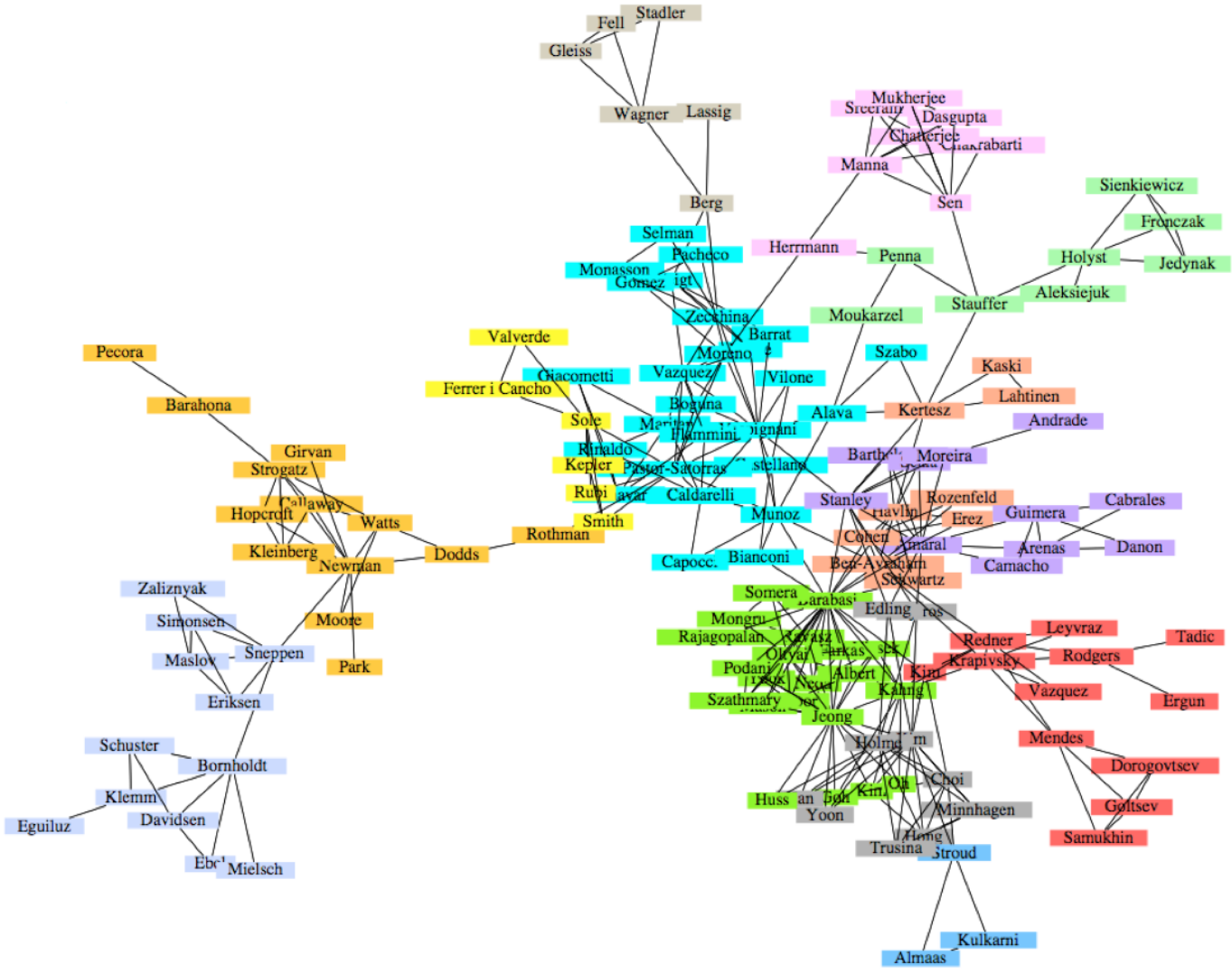


Hierarchical network decomposition:



Girvan-Newman: Results

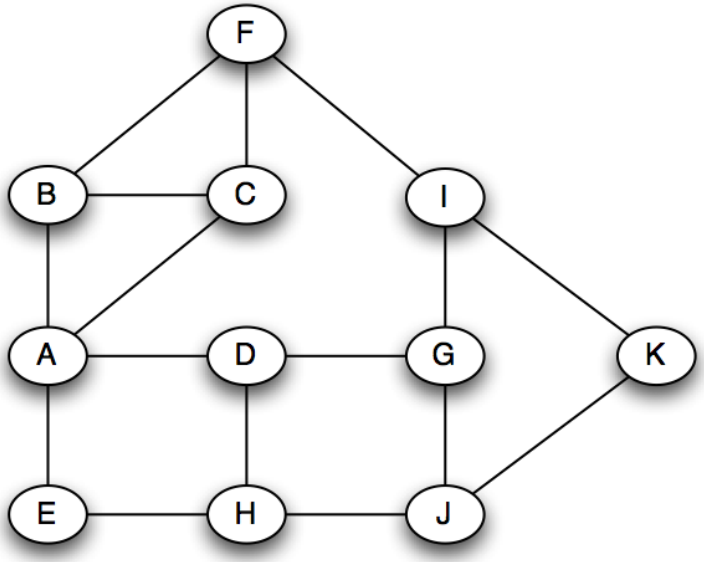
Successfully applied to a co-authorship network

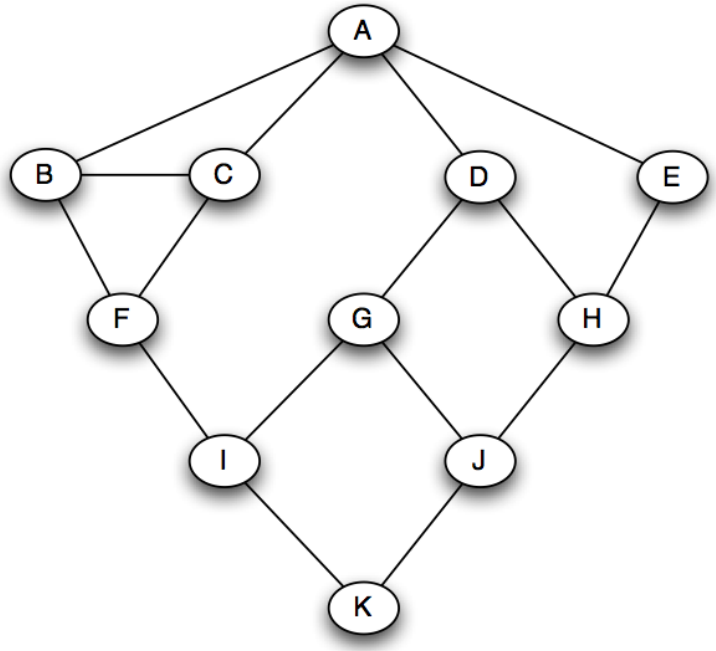


How to Compute Betweenness?

What is an efficient way to compute betweenness without the overhead of listing out all paths in the graph?

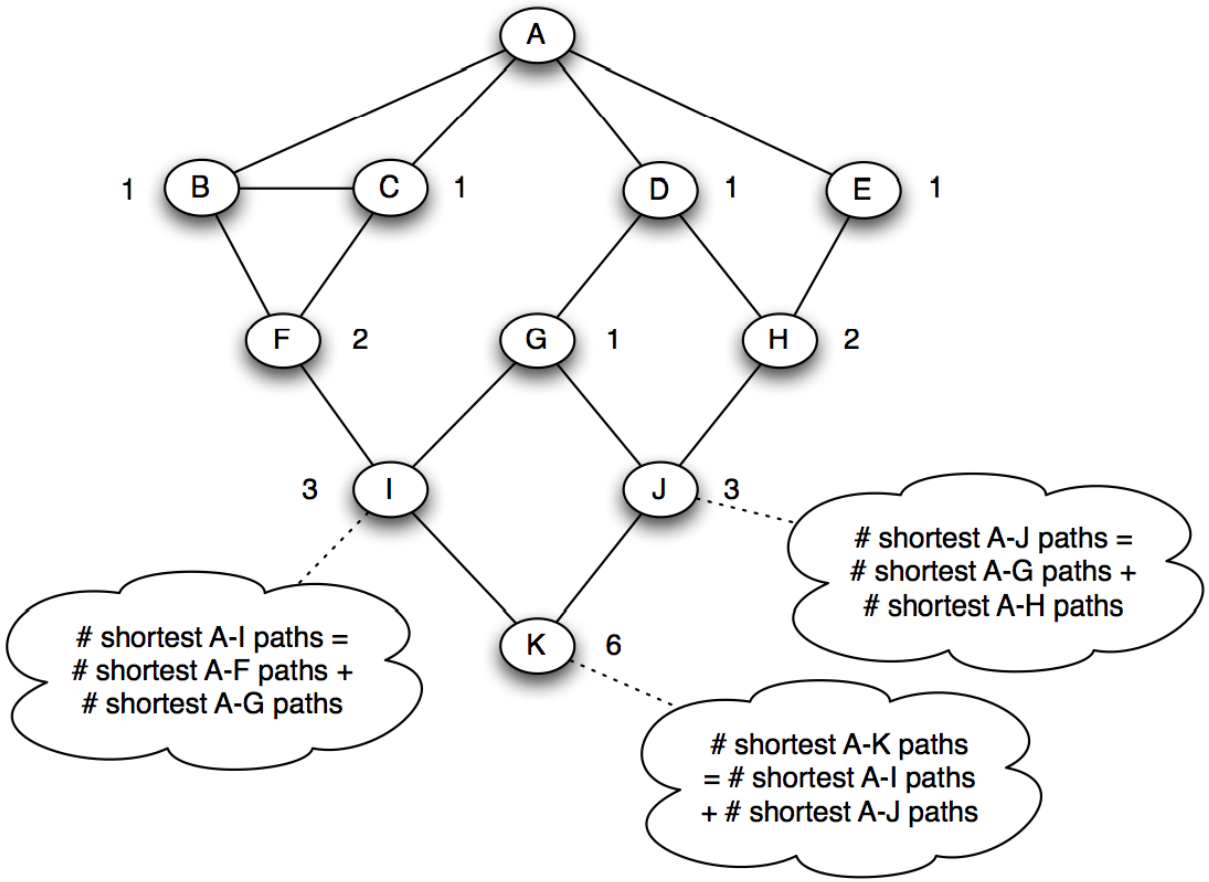
- Breath first search (Consider one node at a time)
- **Step 1**:
  - Perform breath first search starting from A:





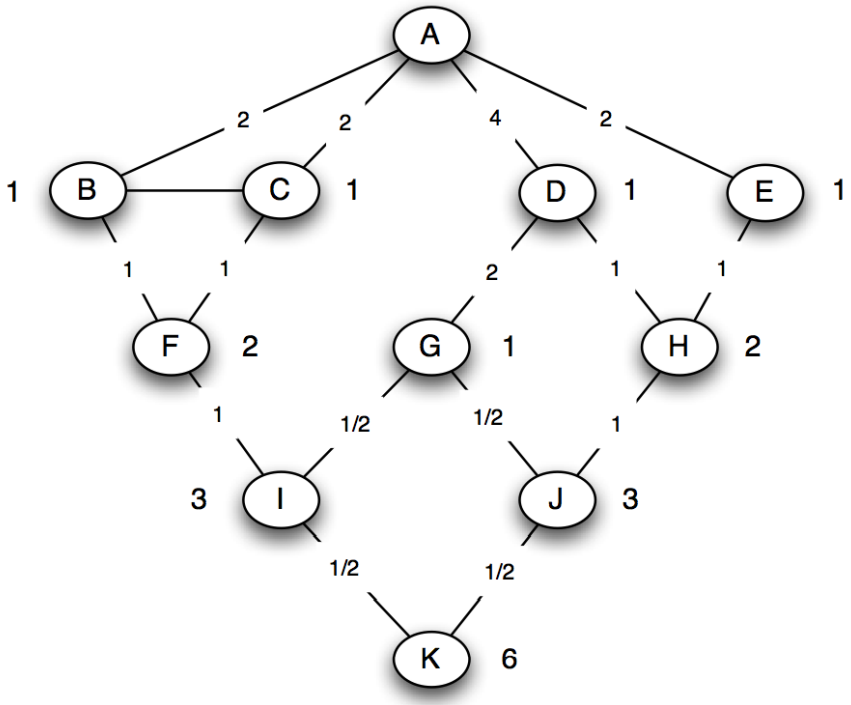
How to Compute Betweenness?

- **Step 2**:
  - Count the number of shortest paths from A to all other nodes of the network:

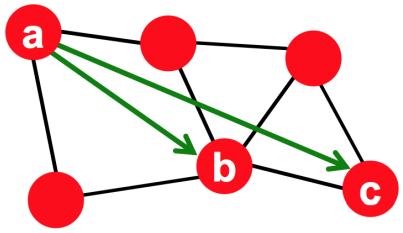


How to Compute Betweenness?

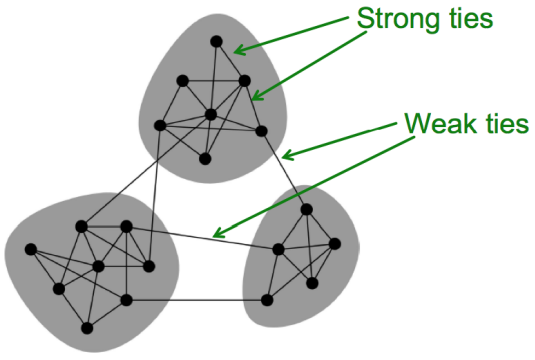
- **Step 3**:
  - Compute betweenness by working up the tree: if there are multiple paths, count them fractionally
    - Add up all the flow arriving from edges directly below X, plus 1 for the flow destined for X itself
    - Divide it up over the edges upward from X, *in proportion to the number of shortest paths coming through each*

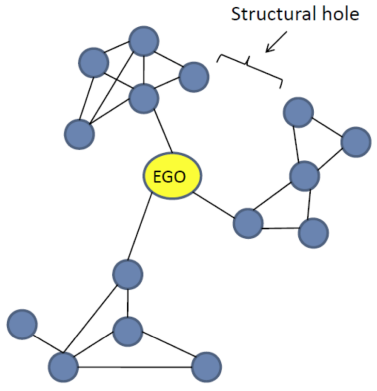


We Learned...



Which edge is more likely a-b or a-c?





- Triadic Closure
- The Strength of Weak Ties
- Tie Strength in Social Media
- Closure and Structural Holes
- Community Detection