Online Social Networks and Media

Strong and Weak Ties

Chapter 3, from D. Easley and J. Kleinberg book

Issues

- How simple processes at the level of individual nodes and links can have complex effects at the whole population

- How information flows within the network

- How different nodes play structurally distinct roles

The Strength of Weak Ties Hypothesis

Mark Granovetter, in the late 1960s

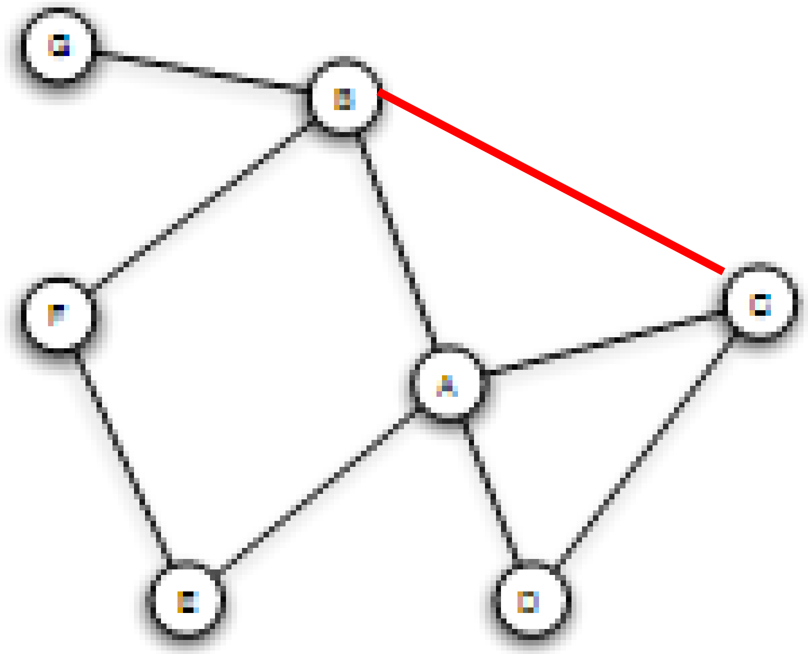
Many people learned information leading to their current job ***through personal contacts***, often described as ***acquaintances*** rather than closed friends

Two aspects

- Structural
- Local (interpersonal)

Triadic Closure

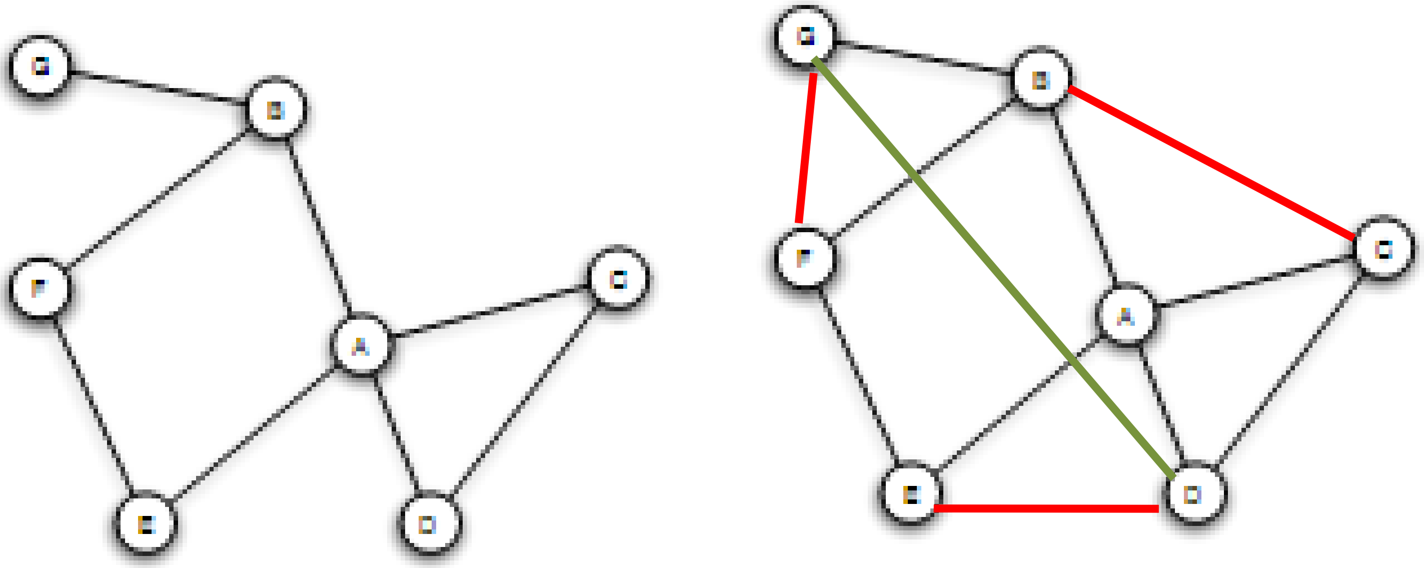
If two people in a social network have a friend in common, then there is an increased likelihood that they will become friends themselves at some point in the future



Triangle

Triadic Closure

Snapshots over time:



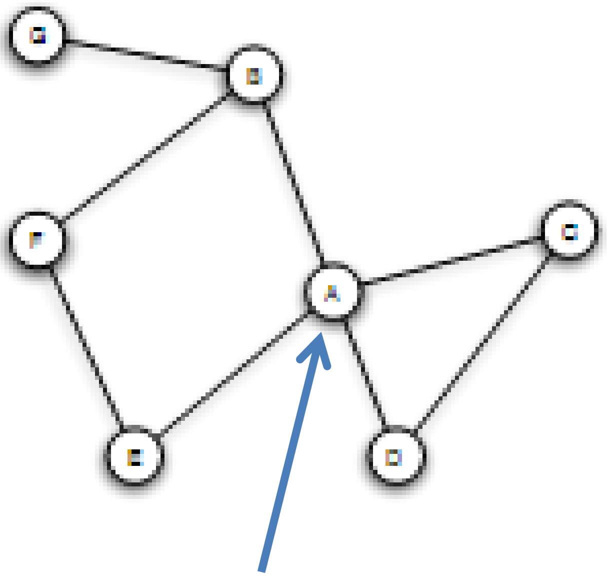
Clustering Coefficient

(Local) clustering coefficient for a node is the probability that two randomly selected friends of a node are friends with each other

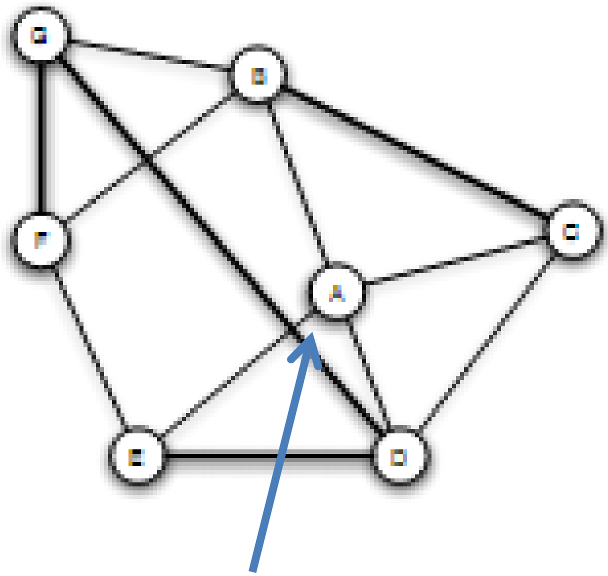
$$C_i = \frac{2\,|\{e_{jk}\}|}{k_i(k_i - 1)} \qquad e_{jk} \in E, u_i, u_j \in N_i, k \text{ size of } N_i, N_i \text{ neigborhood of } u_i$$

Fraction of the friends of a node that are friends with each other (i.e., connected)

Clustering Coefficient



1/6

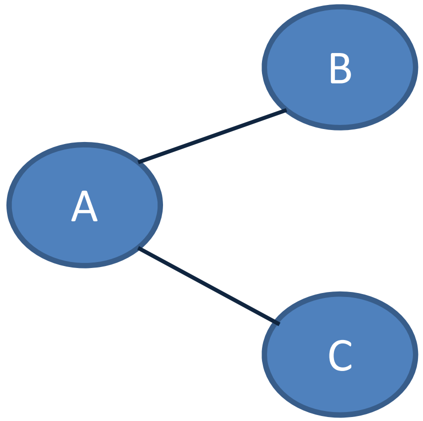


1/2

Ranges from 0 to 1

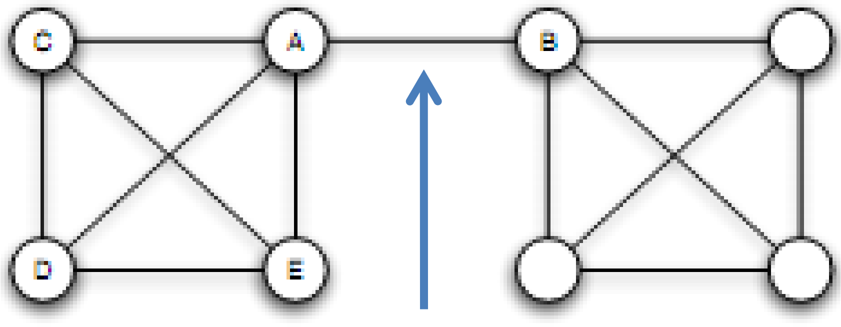
Triadic Closure

If A knows B and C, B and C are likely to become friends, but WHY?



1. Opportunity
2. Trust
3. Incentive of A (latent stress for A, if B and C are not friends, dating back to social psychology)

Bridges and Local Bridges



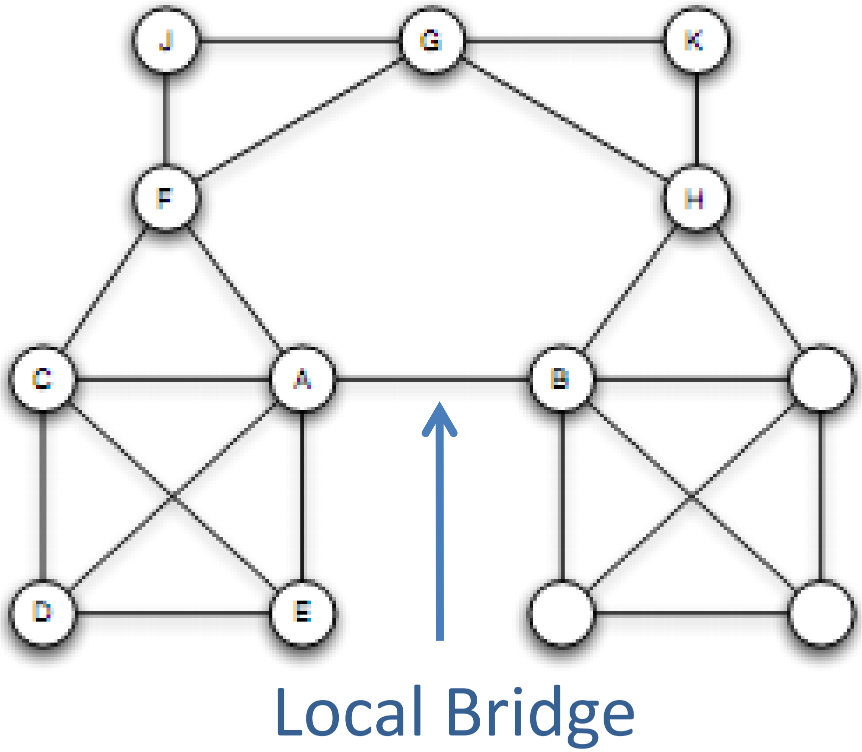
Bridge
(aka cut-edge)

An edge between A and B is a *bridge* if deleting that edge would cause A and B to lie in two different components

AB the only "route" between A and B

extremely rare in social networks

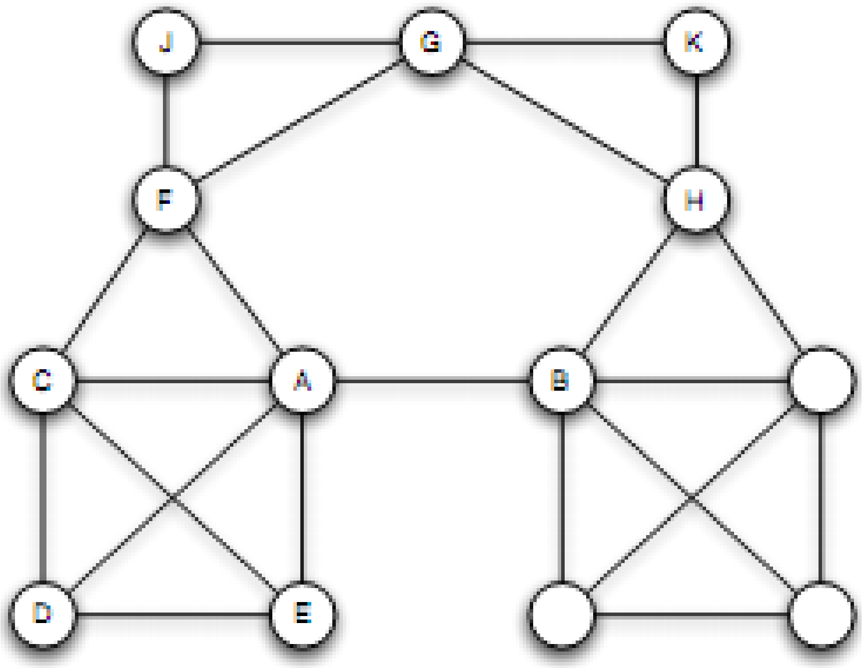
Bridges and Local Bridges



An edge between A and B is a local bridge if deleting that edge would increase the distance between A and B to a value strictly more than 2

Span of a local bridge: distance of the its endpoints if the edge is deleted

Bridges and Local Bridges



An edge is a local bridge, if an only if, it does not form a side of any triangle in the graph

Back to job seeking:

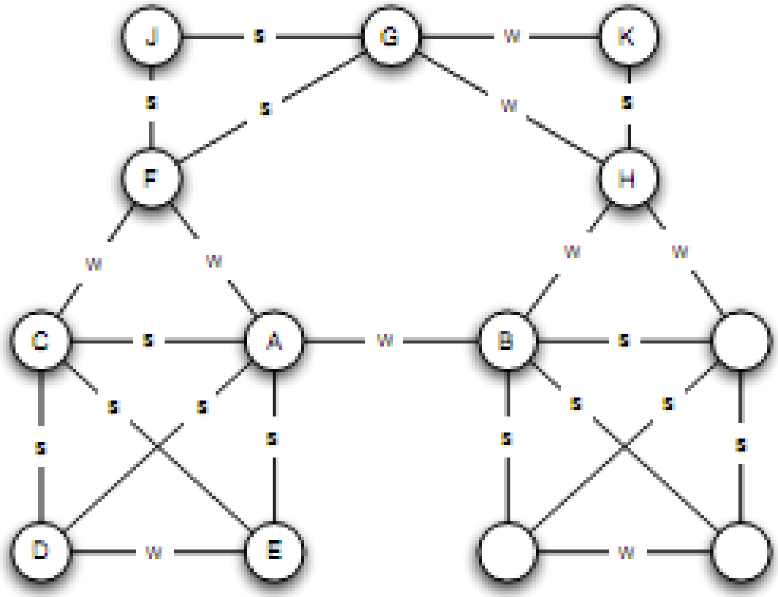
If you are going to get truly new information, it may come from a friend connected by a local bridge

But why distant acquaintances?

The Strong Triadic Closure Property

- Levels of strength of a link
- Strong and weak ties
- Vary across different times and situations

Annotated graph

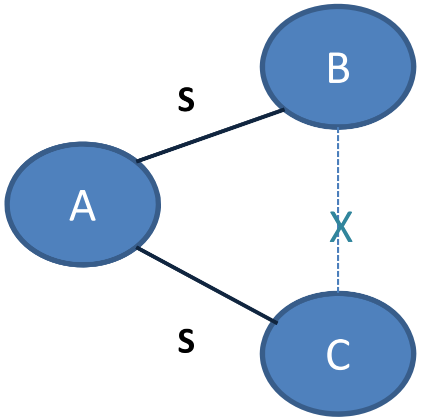


The Strong Triadic Closure Property

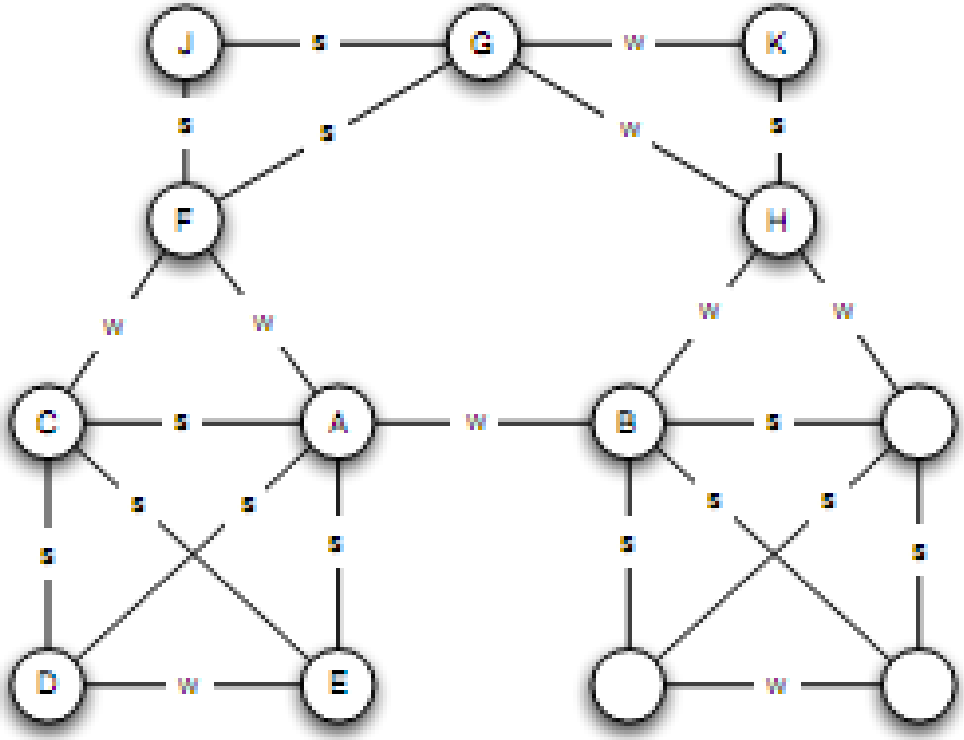
If a node A has edges to nodes B and C, then the B-C edge is <u>especially likely</u> to form if both A-B and A-C are <u>strong ties</u>

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 (strong or weak tie) between B and C.

A node A satisfies the Strong Triadic Property if it does not violate it



The Strong Triadic Closure Property



Local Bridges and Weak Ties

✓Local distinction: weak and strong ties
✓Global structural distinction: local bridges or not

Claim:
If a node A in a network satisfies the Strong Triadic Closure and is involved in at least two strong ties, then any local bridge it is involved in must be a weak tie

Proof: by contradiction

Relation to job seeking?

The role of simplifying assumptions:

- *Useful when they lead to statements robust in practice, making sense as qualitative conclusions that hold in approximate forms even when the assumptions are relaxed*

- *Stated precisely, so possible to test them in real-world data*

- *A framework to explain surprising facts*

Tie Strength and Network Structure in Large-Scale Data

How to test these prediction on large social networks?

Tie Strength and Network Structure in Large-Scale Data

Communication network: "who-talks-to-whom"
Strength of the tie: time spent talking during an observation period

Cell-phone study [Omnela et. al., 2007]

"who-talks-to-whom network", covering 20% of the national population

- Nodes: cell phone users
- Edge: if they make phone calls to each other in both directions over 18-week observation periods

<u>Is it a "social network"?</u>
Cells generally used for personal communication + no central directory, thus cell-phone numbers exchanged among people who already know each other
Broad structural features of large social networks (giant component, 84% of nodes)

Generalizing Weak Ties and Local Bridges

- ✓ Either weak or strong
- ✓ Local bridge or not

Tie Strength

From weak and strong -> Numerical quantity (= number of min spent on the phone)

Quantify “local bridges”, how?

Generalizing Weak Ties and Local Bridges

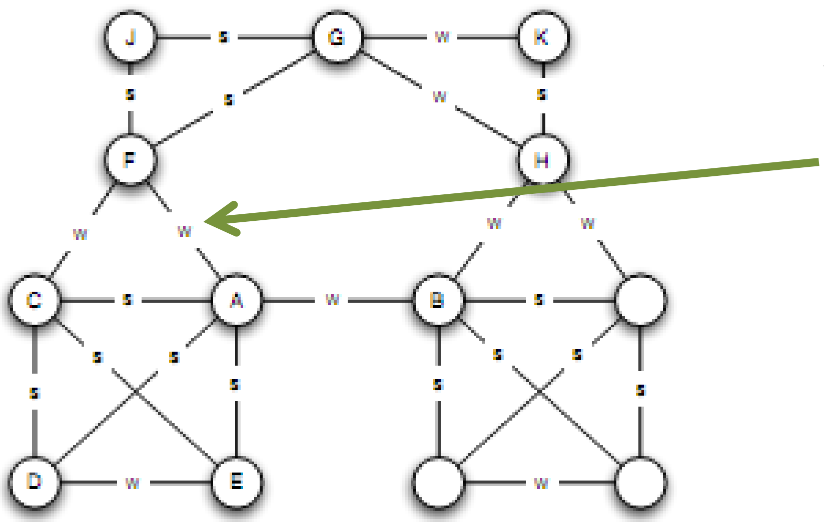
Bridges

"almost" local bridges

Neighborhood overlap of an edge e_{ij}

$$\frac{|N_i \cap N_j|}{|N_i \cup N_j|}$$

(*) In the denominator we do not count A or B themselves



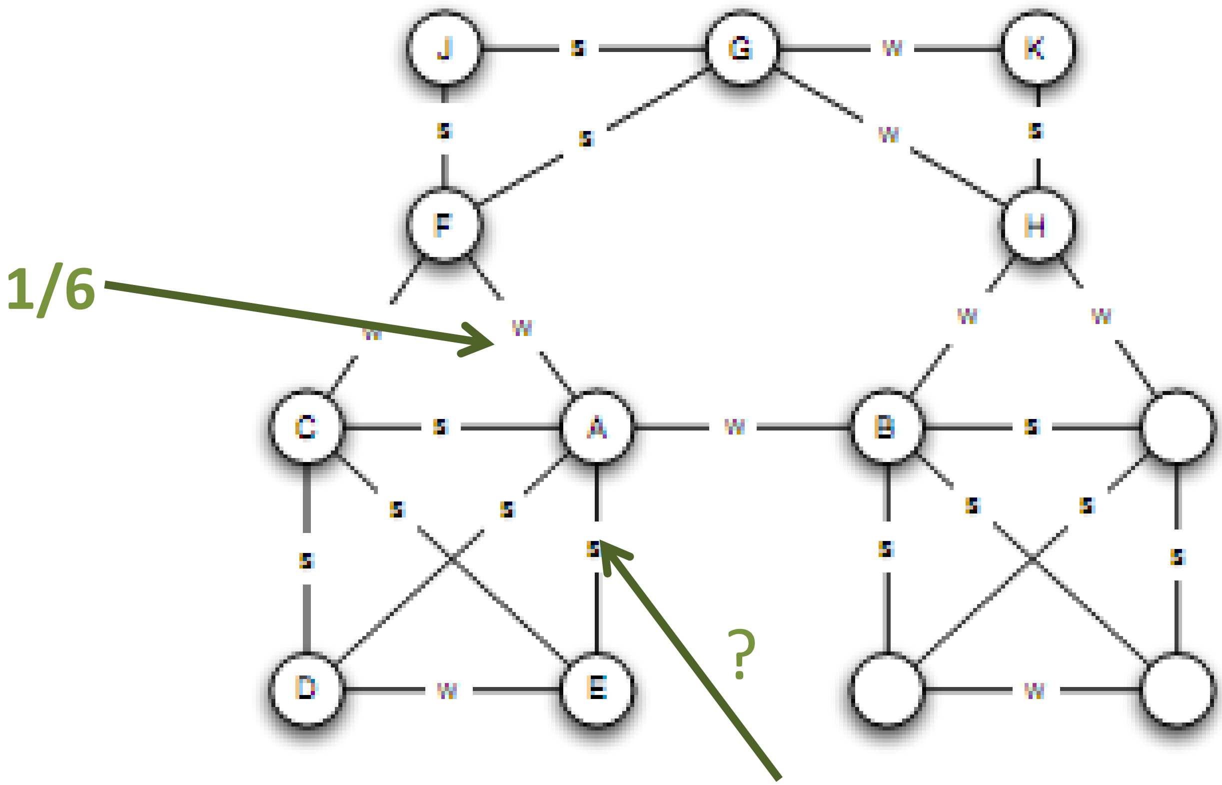
A: B, E, D, C
F: C, J, G

1/6

When is this value 0?

Generalizing Weak Ties and Local Bridges

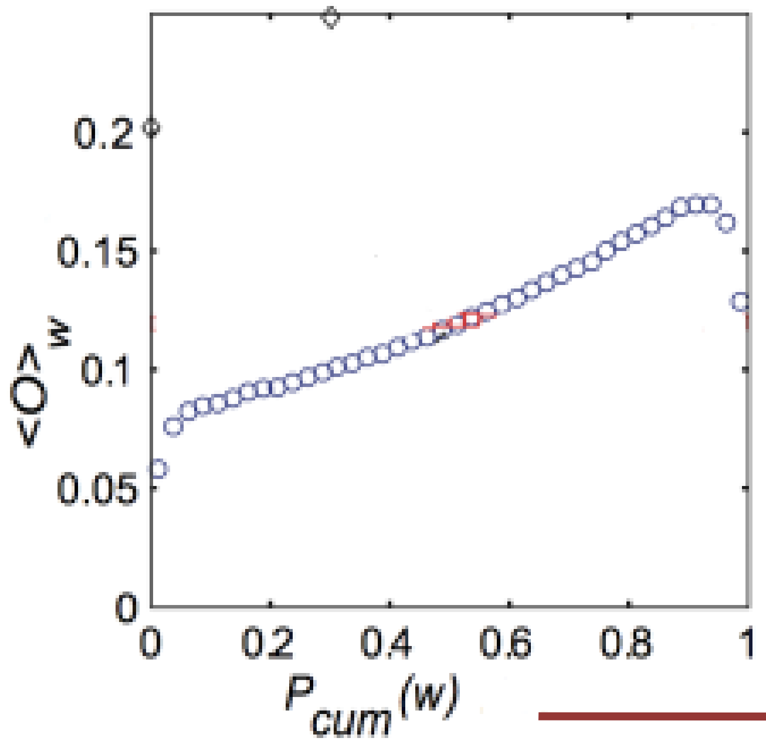
= 0 : edge is a local bridge
Small value: “almost” local bridges



Generalizing Weak Ties and Local Bridges:
Empirical Results

How the neighborhood overlap of an edge depends on its strength

(Hypothesis: the strength of weak ties predicts that neighborhood overlap should grow as tie strength grows)



(*) Some deviation at the right-hand edge of the plot

sort the edges -> for each edge at which percentile

Strength of connection (function of the percentile in the sorted order)

Local level -?-> global level: weak ties serve to link different tightly-knit communities that each contain a large number of stronger ties – How would you test this?

Generalizing Weak Ties and Local Bridges: Empirical Results

Hypothesis: weak ties serve to link different tightly-knit communities that each contain a large number of stronger ties

Delete edges from the network one at a time

- Starting with the strongest ties and working downwards in order of tie strength
    - giant component shrank steadily

-Starting with the weakest ties and upwards in order of tie strength
    - giant component shrank more rapidly, broke apart abruptly as a critical number of weak ties were removed

Social Media and Passive Engagement

People maintain large explicit lists of friends

Test:
How online activity is distributed across links of different strengths

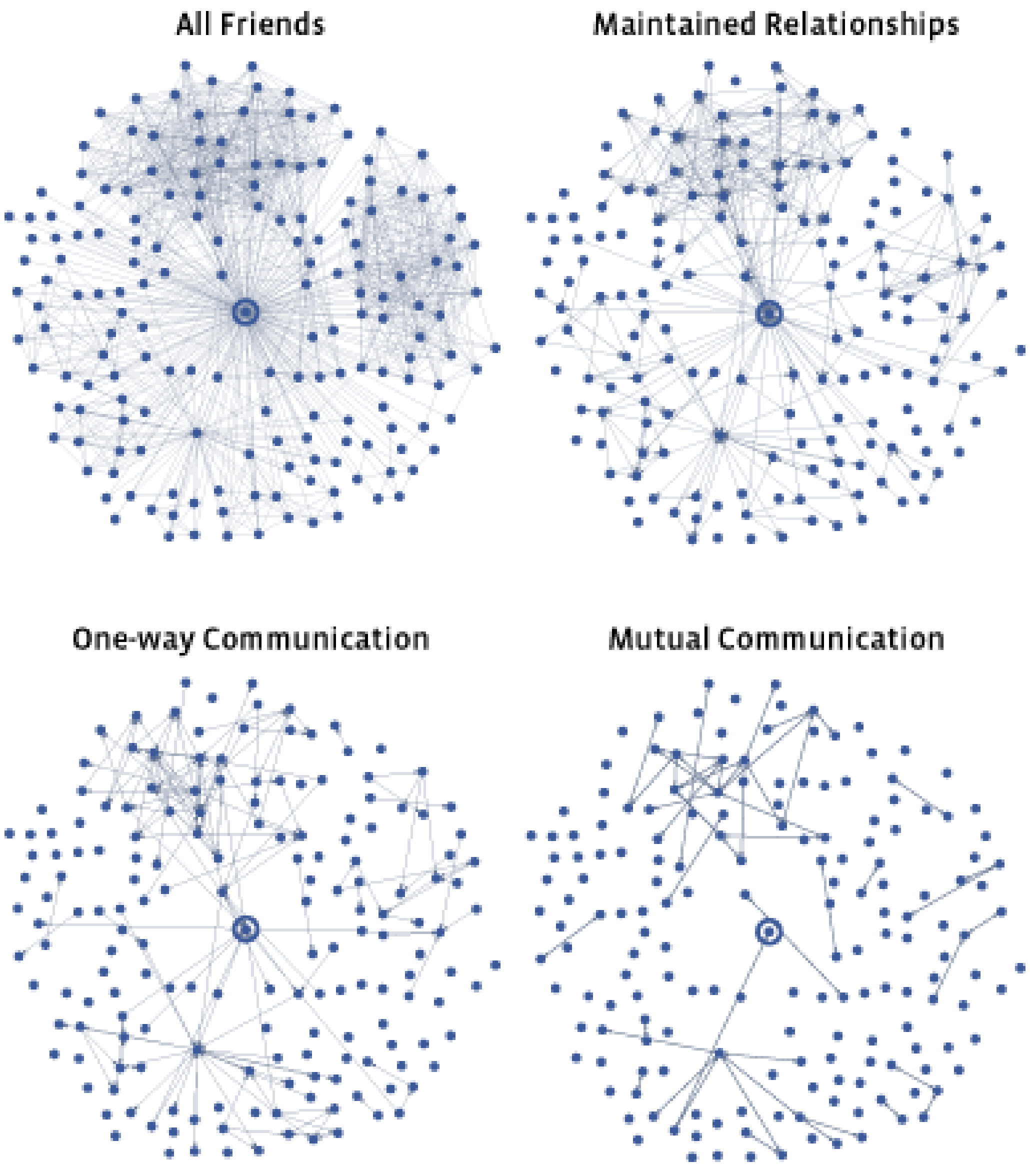
Tie Strength on Facebook

Cameron Marlow, et al, 2009

At what extent each link was used for social interactions

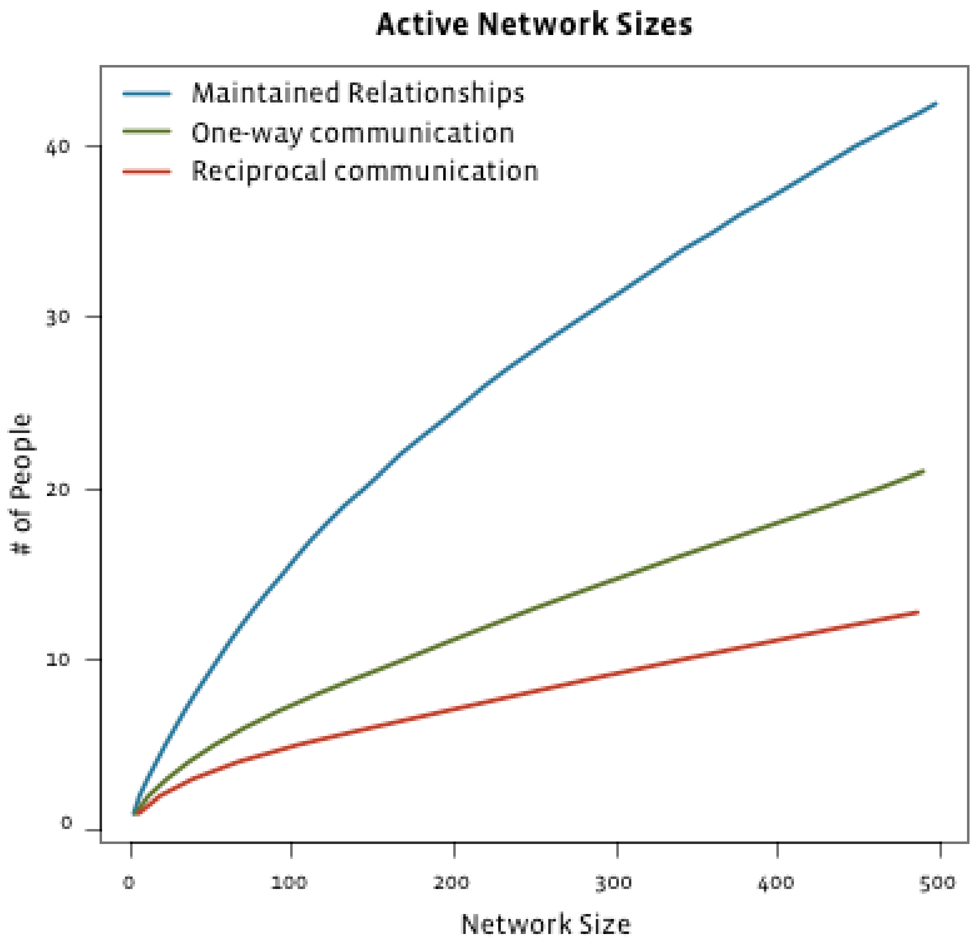
1. Reciprocal (mutual) communication: both send and received messages to friends at the other end of the link
2. One-way communication: the user send one or more message to the friend at the other end of the link
3. Maintained relationship: the user followed information about the friend at the other end of the link (click on content via News feed or visit the friend profile more than once)

Tie Strength on Facebook



Two distinct regions

Tie Strength on Facebook



Total number of friends

Even for users with very large number of friends

- actually communicate : 10-20
- number of friends follow even passively <50

Passive engagement (keep up with friends by reading about them even in the absence of communication)

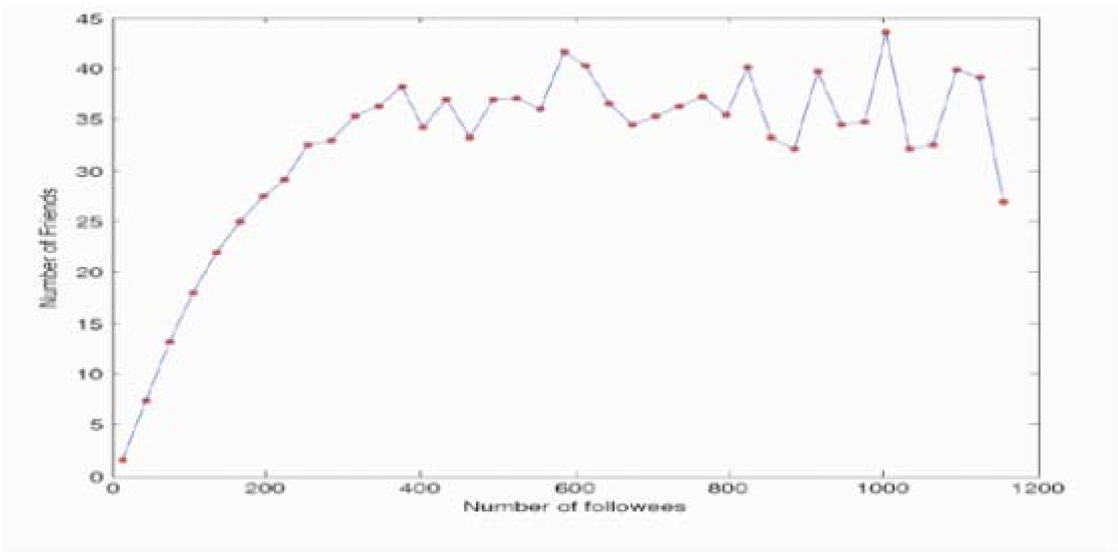
Passive as a network middle ground

Tie Strength on Twitter

Huberman, Romero and Wu, 2009

Two kinds of links

- Follow
- Strong ties (friends): users to whom the user has *directed at least two messages* over the course if the observation period



Social Media and Passive Engagement

- Strong ties require continuous investment of time and effort to maintain (as opposed to weak ties)

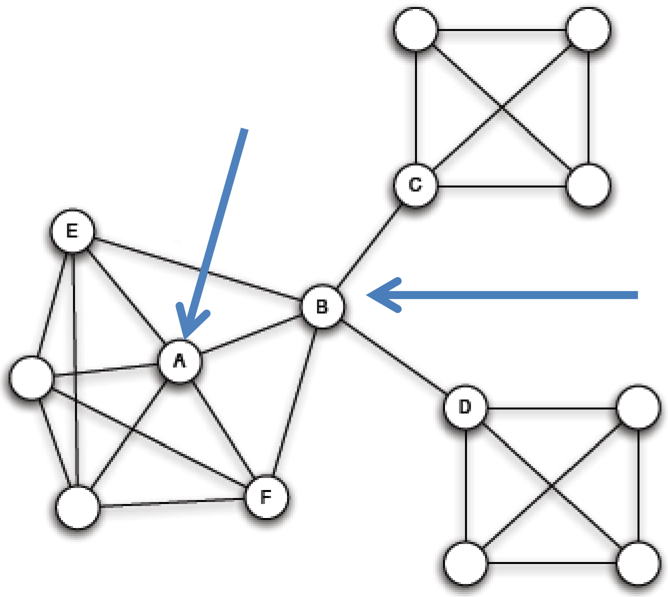
- Network of strong ties still remain sparse

- How different links are used to convey information

Closure, Structural Holes and Social Capital

Different roles that **nodes** play in this structure

Access to edges that span different groups is not equally distributed across all nodes

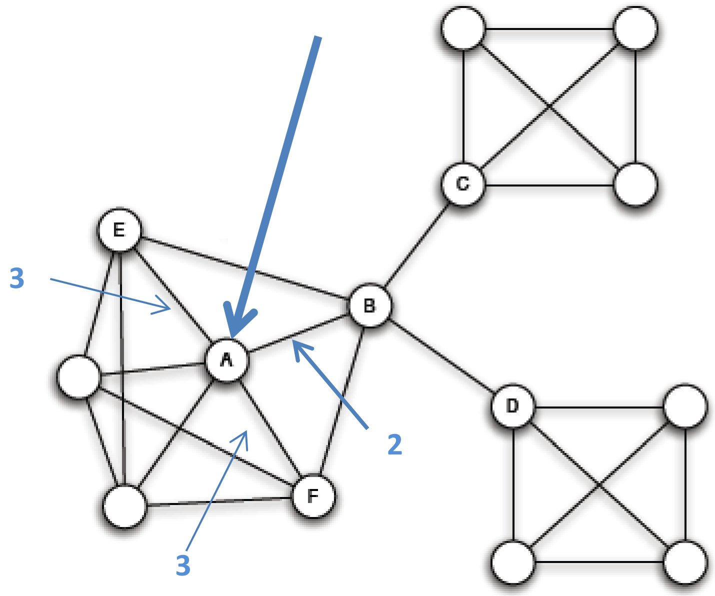


Embeddedness

Large clustering coefficient

- Embeddedness of an edge: number of common neighbors of its endpoints (neighborhood overlap, local bridge if 0)

A all its edges have significant embeddedness



(sociology) if two individuals are connected by an embedded edge => trust

- "Put the interactions between two people on display"

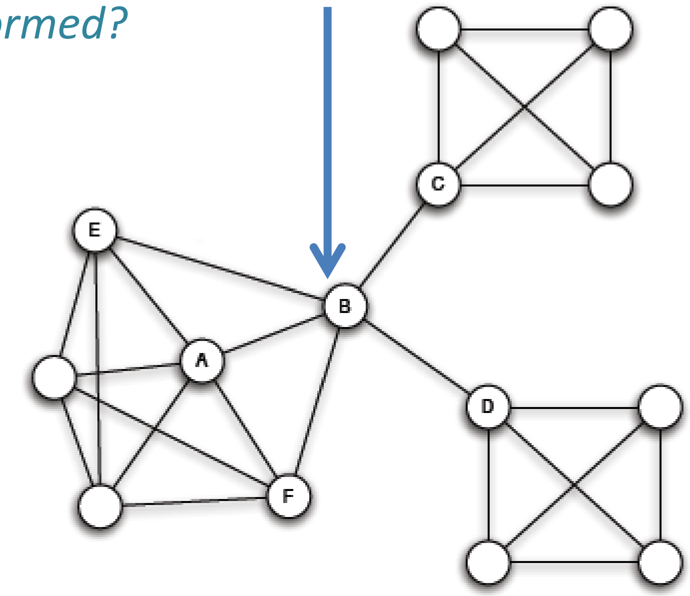
Structural Holes

(sociology) B-C, B-D much riskier, also, possible contradictory constraints
Success in a large cooperation correlated to access to local bridges

B “spans a structural hole”

- B has access to information originating in multiple, non interacting parts of the network
- An amplifier for creativity
- Source of power as a social “gate-keeping”

Will a triangle be formed?

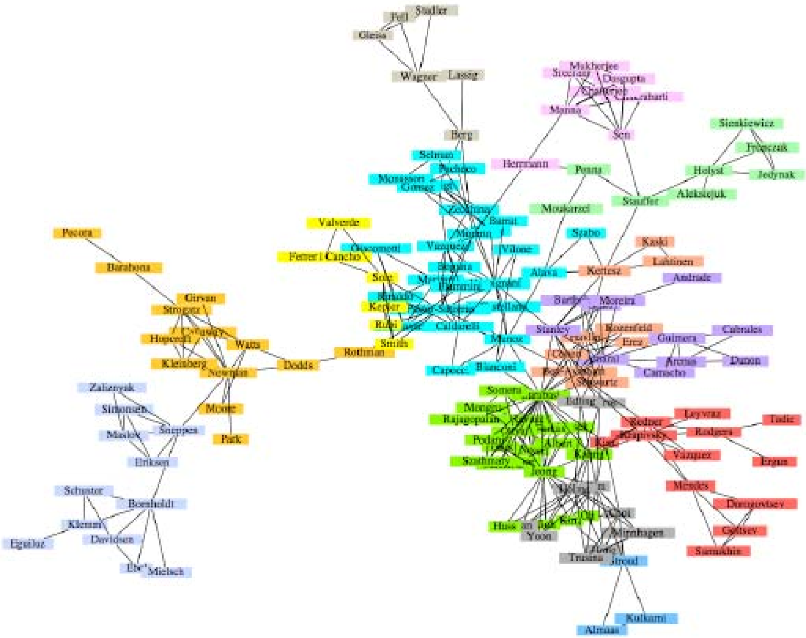


Closure and Bridging as Forms of Social Capital

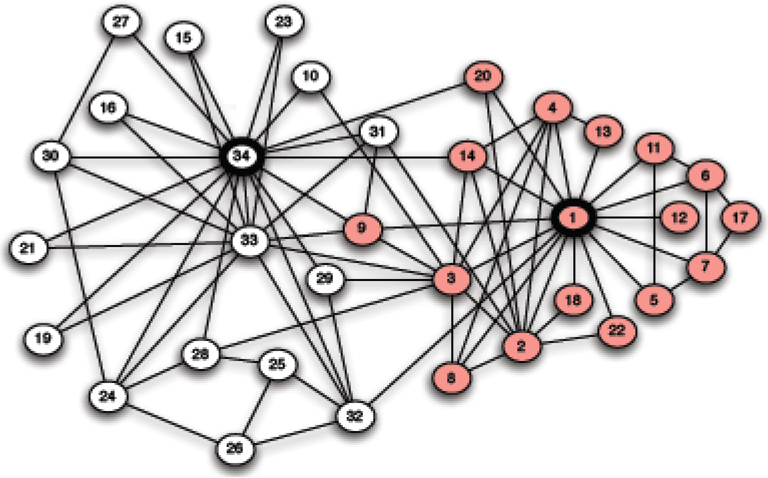
Social capital: benefits from membership in social networks and other social structures

Betweenness and Graph Partitioning

Graph partitioning
Given a network dataset, how to identity densely connected groups of nodes within it



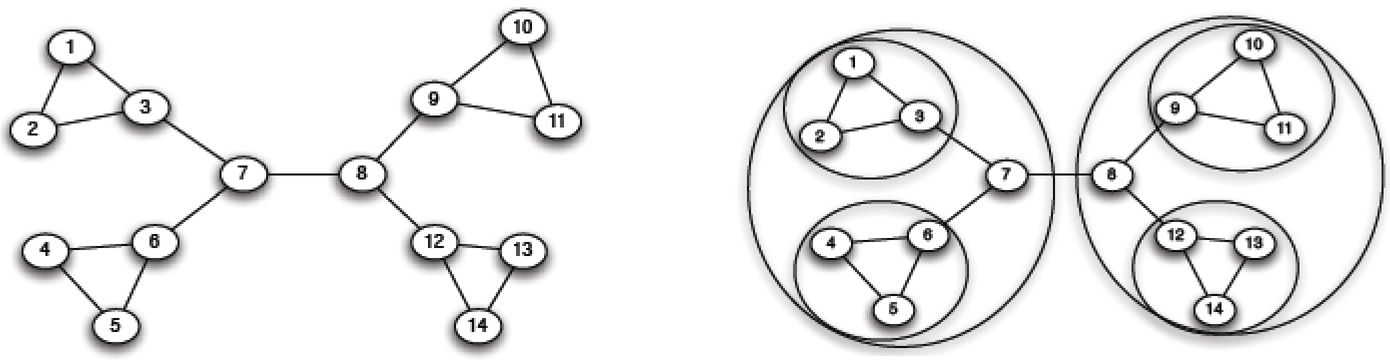
Co-authorship network of physicists and applied mathematicians



Karate club

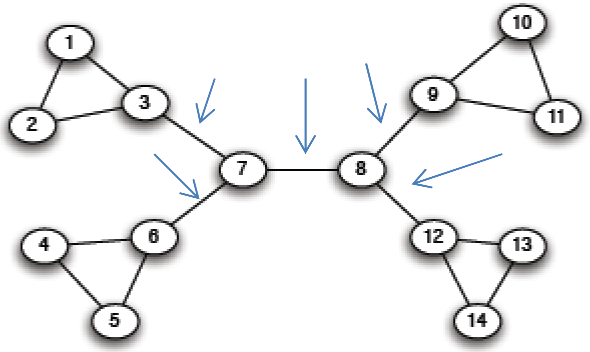
Betweenness and Graph Partitioning

- Divisive methods: try to identify and remove the "spanning links" between densely-connected regions
- Agglomerative methods: Find nodes that are likely to belong to the same region and merge them together (bottom-up)



Girvan and Newman

- Divisive method

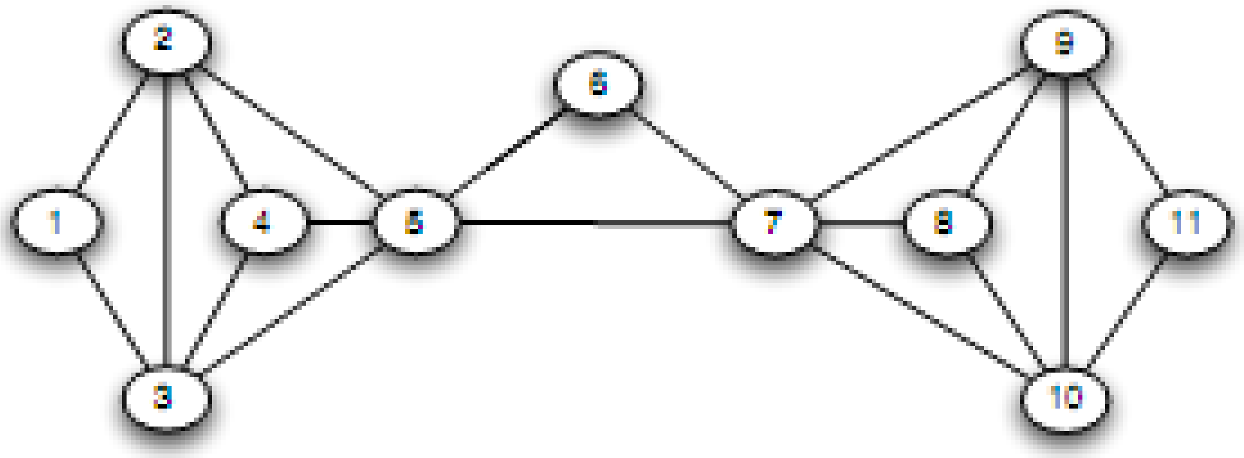


Finding bridges and local bridges?

Which one to choose?

Girvan and Newman

There is no local bridge



Girvan and Newman: Betweenness

vertices that have a high probability to occur on a randomly chosen shortest path between two randomly chosen nodes have a high betweenness.

Traffic (unit of flow)

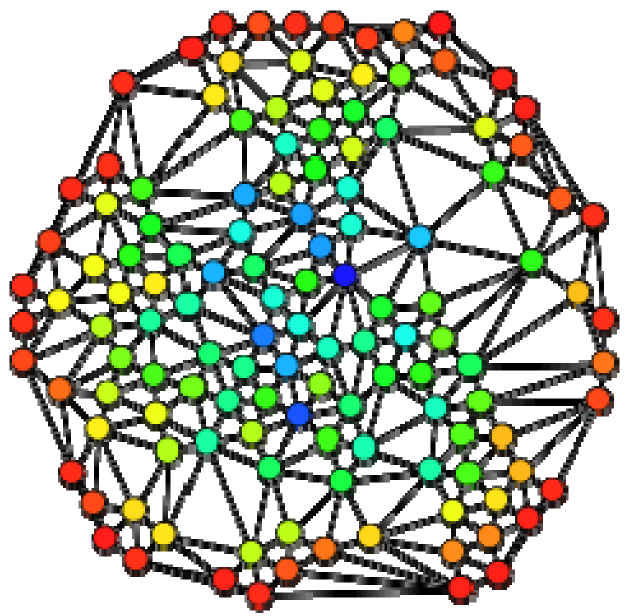
Betweenness of an edge (a, b): number of pairs of nodes x and y such that the edge (a, b) lies on the shortest path between x and y.

Since there can be several shortest paths between x and y, edge (a, b) is credited with the fraction of those shortest paths that include the edge (a, b).

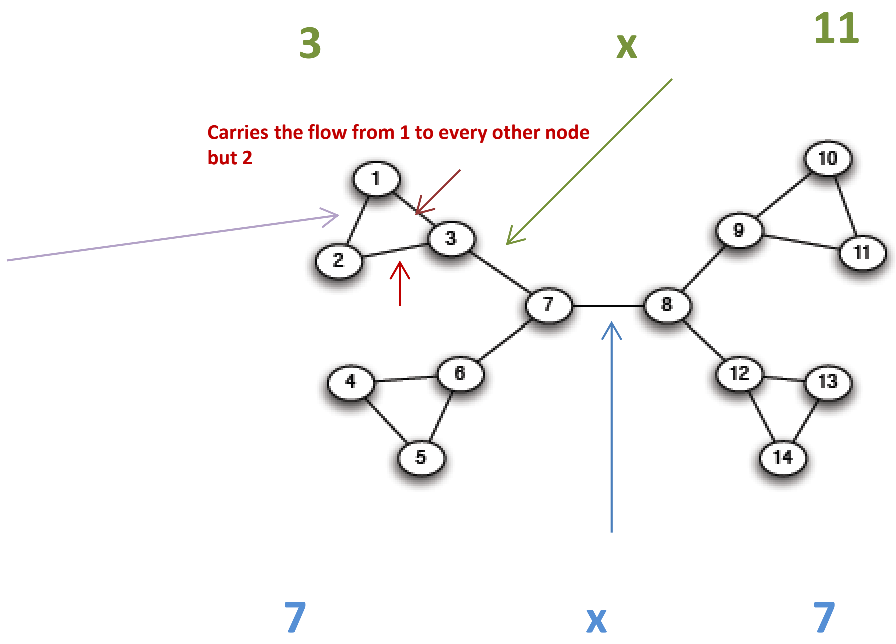
Similar notion exists for nodes

Blue (max)

Red (0)



Girvan and Newman: Betweenness

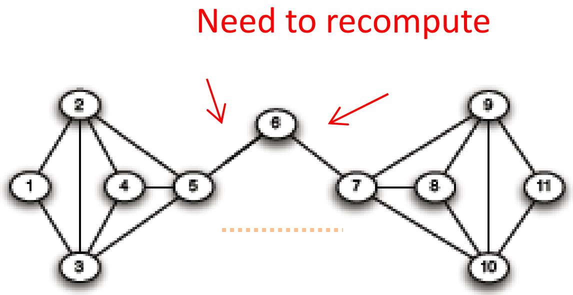


Girvan and Newman

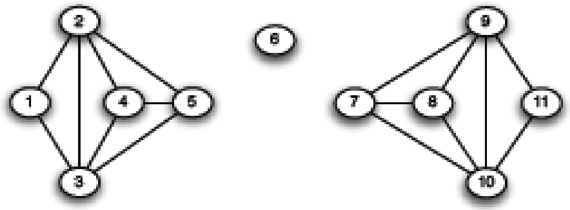
1. The betweenness of all existing edges in the network is calculated first.
2. The edge with the highest betweenness is removed.
   If this separates the graph -> partition.
3. The betweenness of all edges affected by the removal is recalculated.

Steps 2 and 3 are repeated until no edges remain.

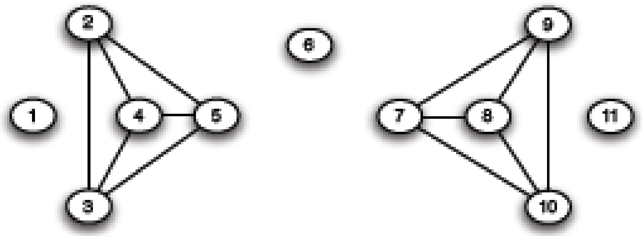
Girvan and Newman



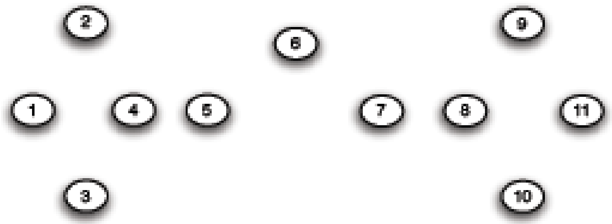
(a) *Step 1*



(b) *Step 2*

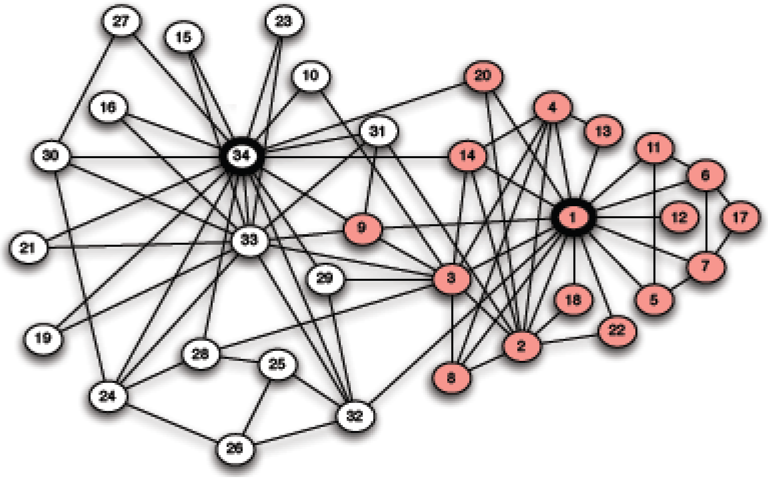


(c) *Step 3*



(d) *Step 4*

Girvan and Newman



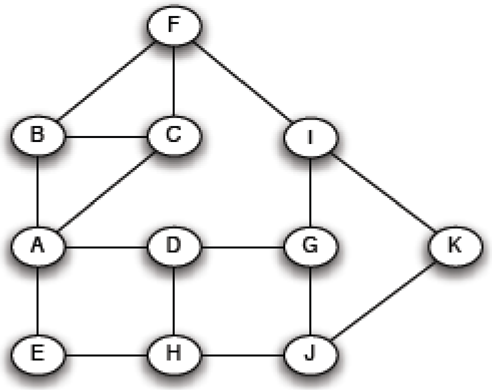
34 president
1 instructor
Correct but node 9 (attached it to 34)

Minimum cut approach – the same outcome

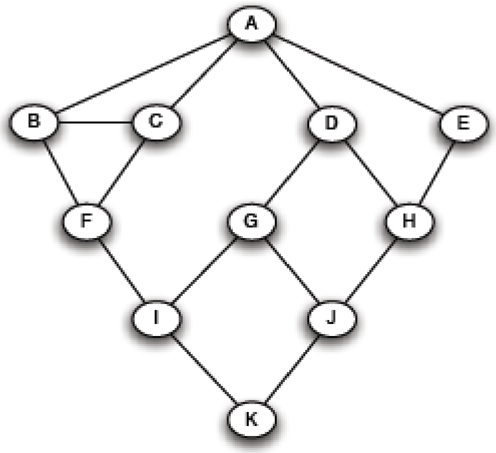
Computing Betweenness

1. Perform a BFS starting from A
2. Determine the shortest path from A to each other node
3. Based on these numbers, determine the amount of flow from A to all other nodes that uses each edge

Computing Betweenness



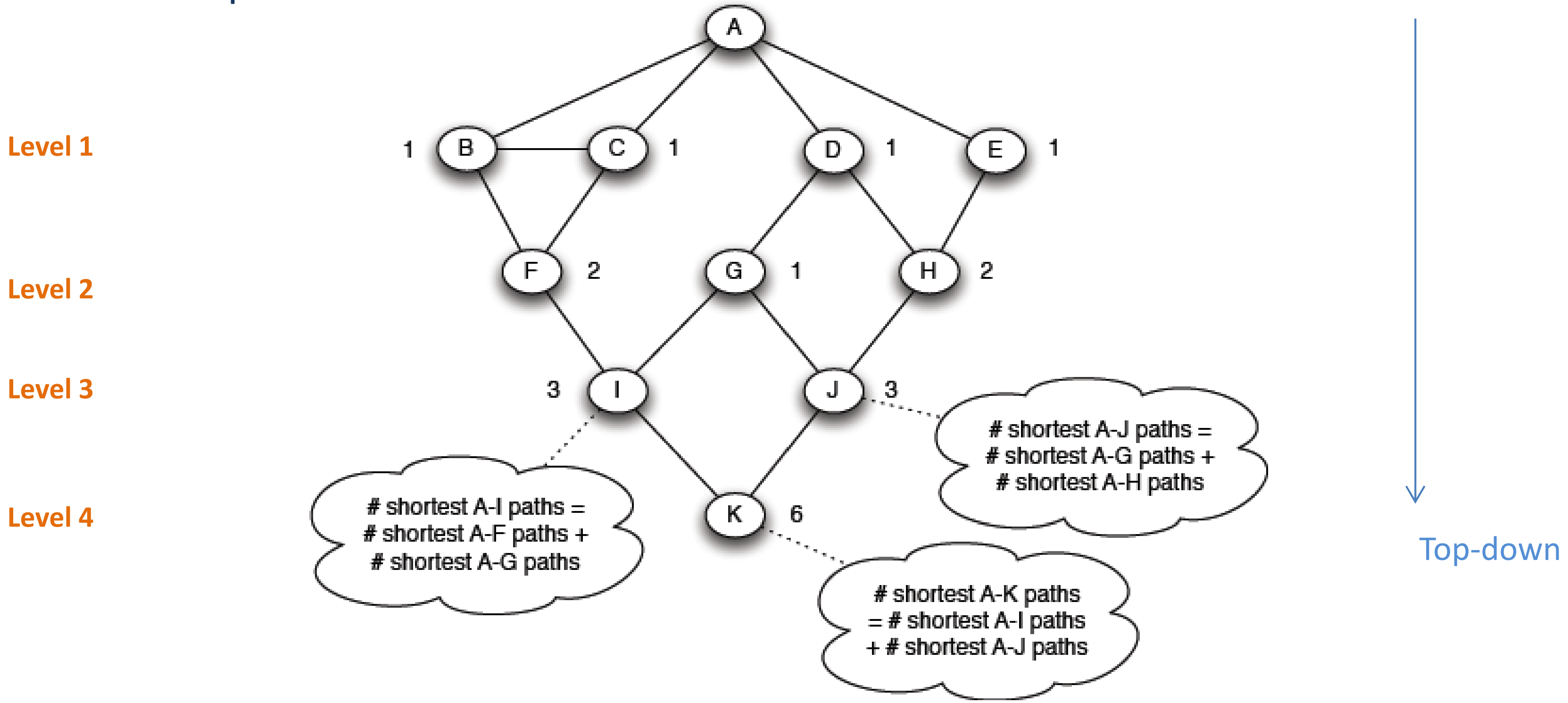
Initial network



BFS on A

Computing Betweenness

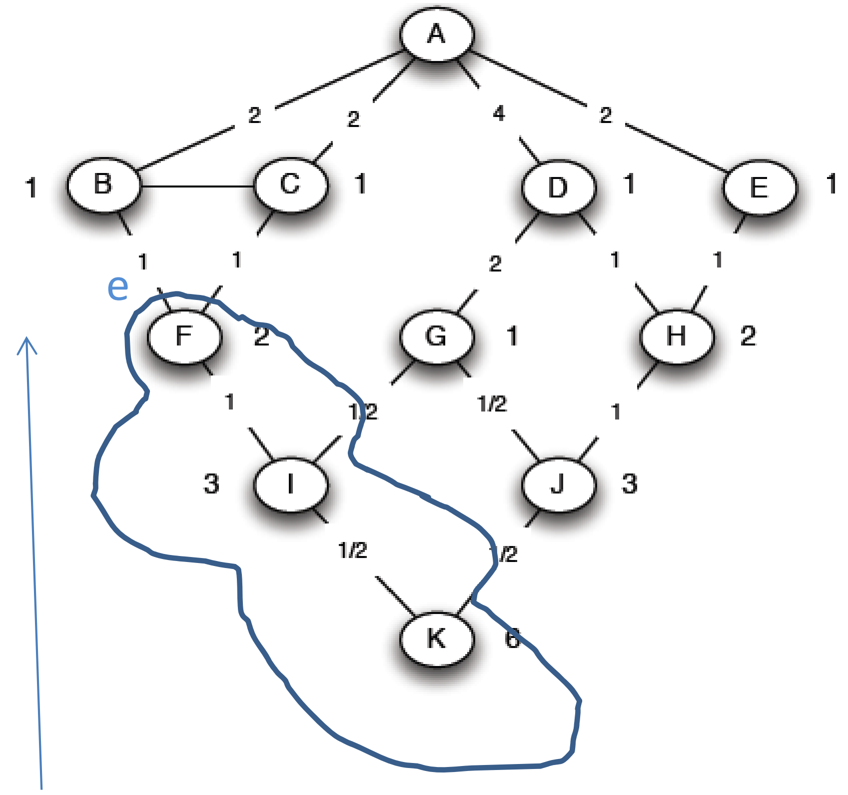
Count how many shortest paths
from A to a specific node



Computing Betweenness

Compute the flow
calculate for each edge e the sum over all nodes
Y of the fraction of shortest paths from the root X to Y
that go through e.

for each edge e: the sum over all nodes Y of the fraction of shortest paths from the root A to Y that go through e.
Each edge participates in the shortest-paths from the root to nodes (at levels) below it



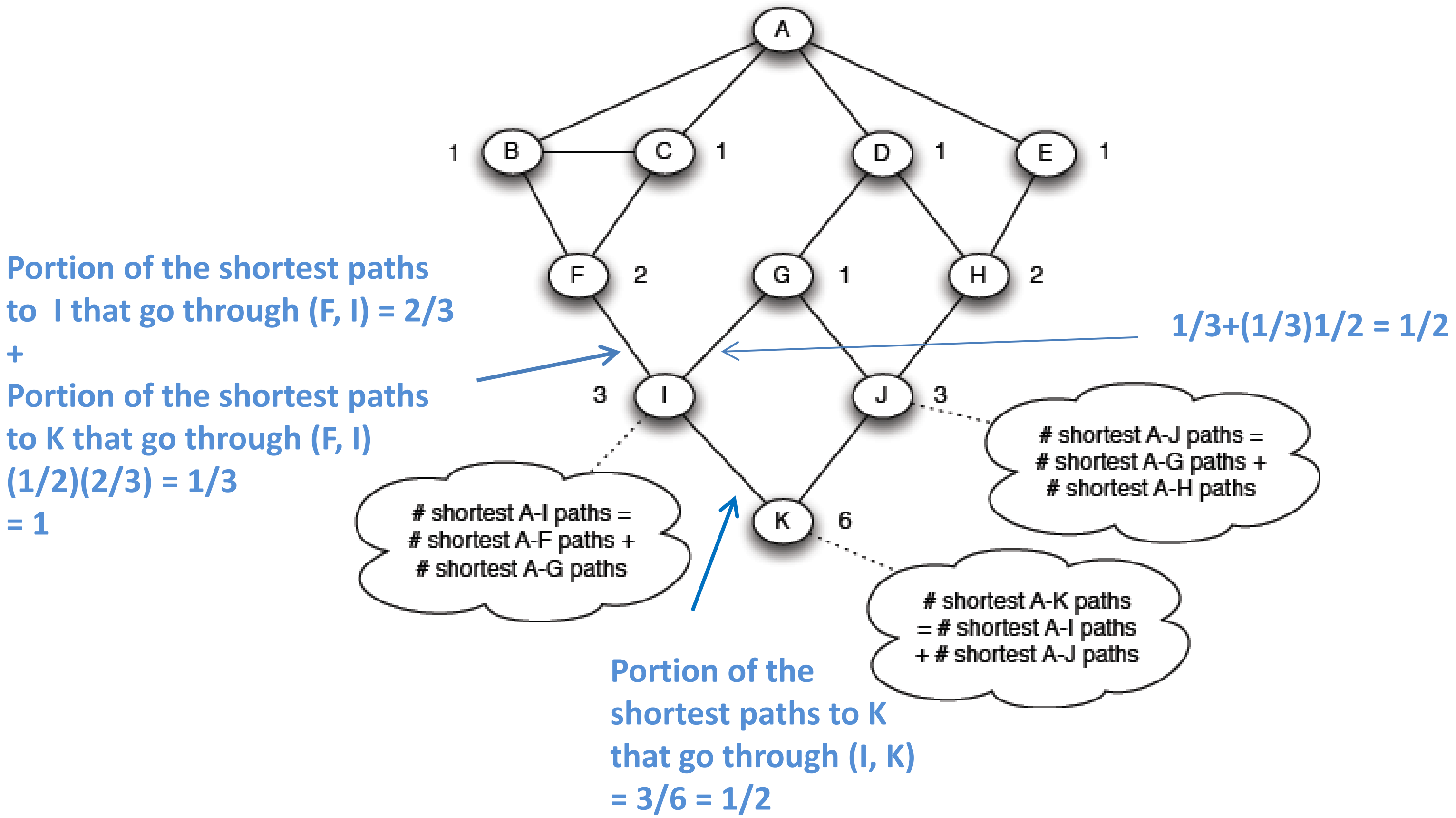
Each node other than the root is given a credit of 1, representing the shortest path to that node.

This credit may be divided among nodes and edges above, since there could be several different shortest paths to the node.

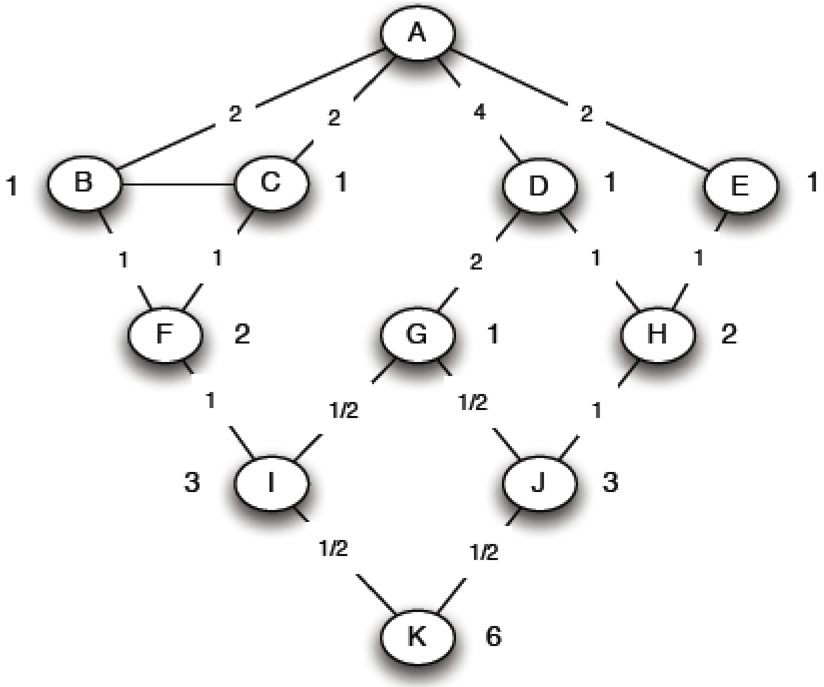
Computing Betweenness

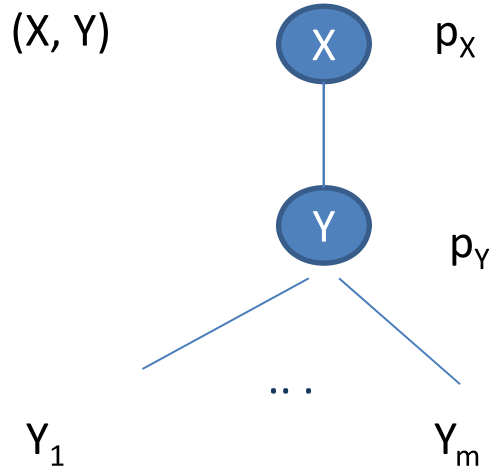
Count the flow through each link

$$credit(e) = \sum_{X,Y} \frac{|\, shortest - path(X,Y) \text{through } e \,|}{|\, shortest\_path(X,Y)\} \,|}$$



Computing Betweenness



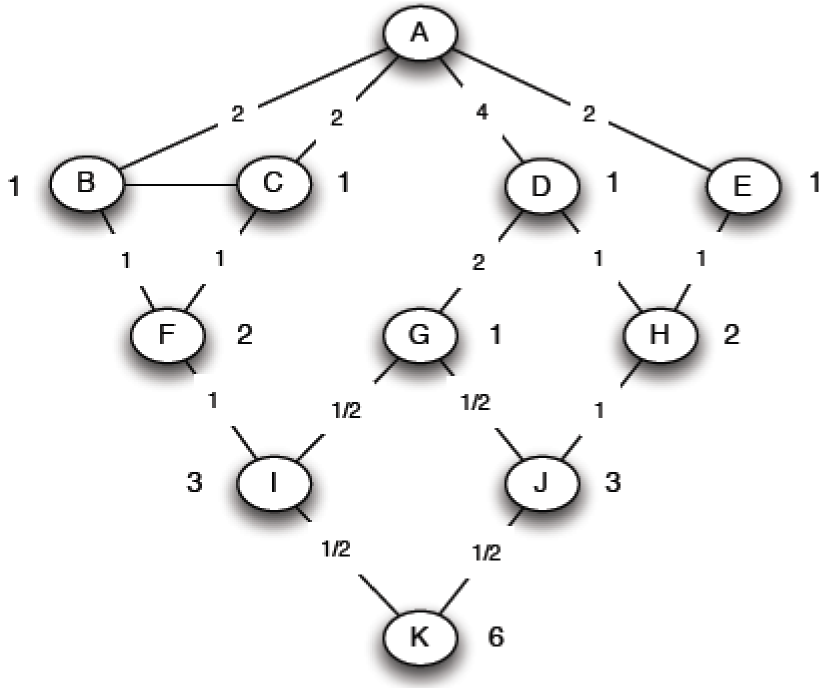


$$flow(X,Y) = p_X / p_Y \sum_{Y_i child of Y} (p_X / p_Y)\, flow(Y, Y_i)$$

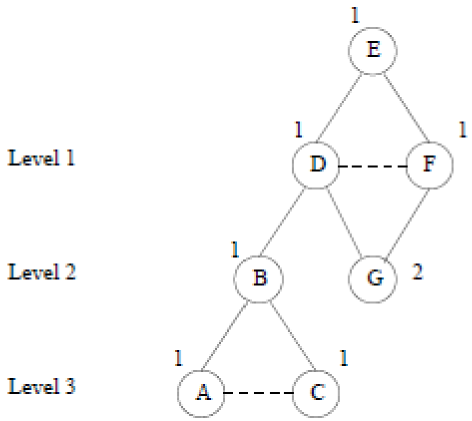
Computing Betweenness

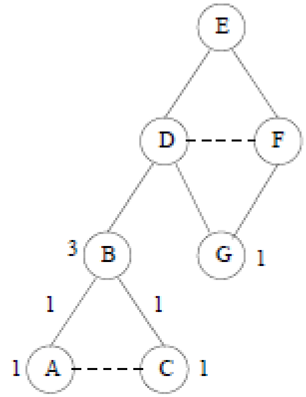
Final for A

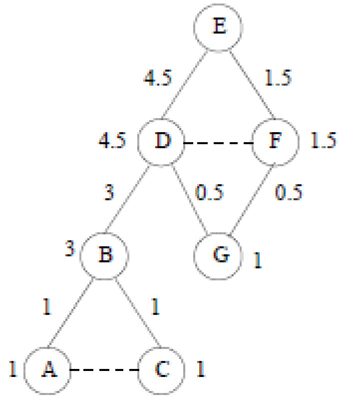
Sum over all BFSs



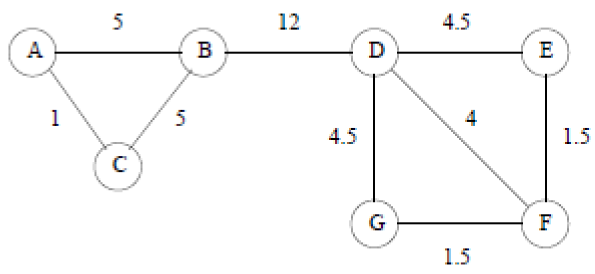
Computing Betweenness

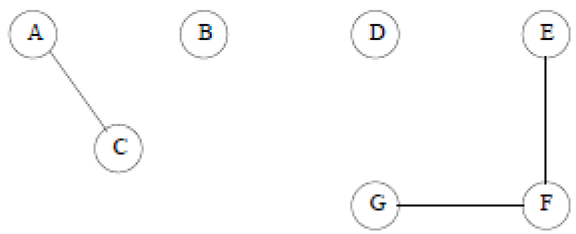






Computing Betweenness



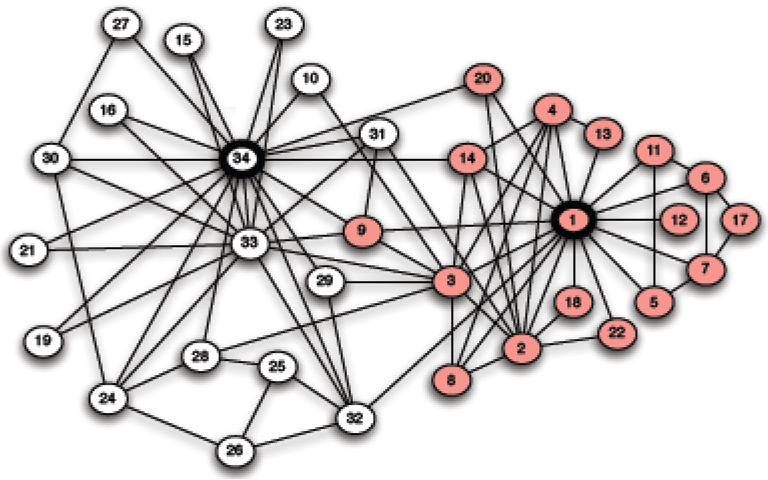


Computing Betweenness

Issues

- Test for connectivity?
- Re-compute all paths, or only those affected
- Parallel computation
- “Sampling

Girvan and Newman



34 president
1 instructor
Correct but node 9 (attached it to 34) – why? 3 weeks away from getting a black belt

Minimum cut approach – the same outcome