Social and Technological Network Analysis

Lecture 2: Weak Ties and Community Detection

Dr. Cecilia Mascolo

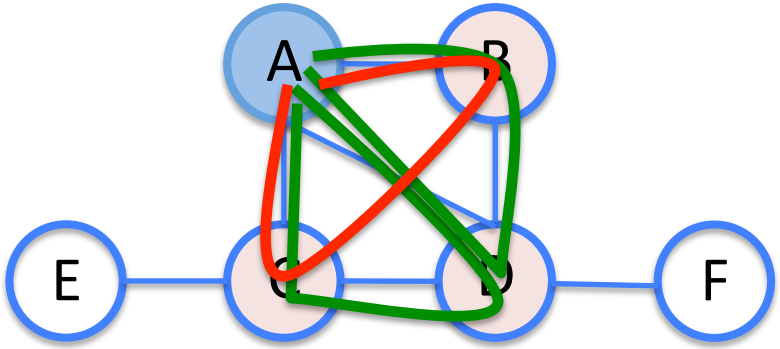
UNIVERSITY OF CAMBRIDGE

In This Lecture

- We will introduce the concept of weak ties and illustrate their importance
- From weak ties we will discuss some basic community detection methods

Again on Clustering Coefficient

- We have introduced the clustering coefficient. This indicates:
  - The number of triangles including node A.
  - How connected the friends of A are.
- **Triadic closure**: if C and B are connected to A there is an increased likelihood that they will be connected in future.



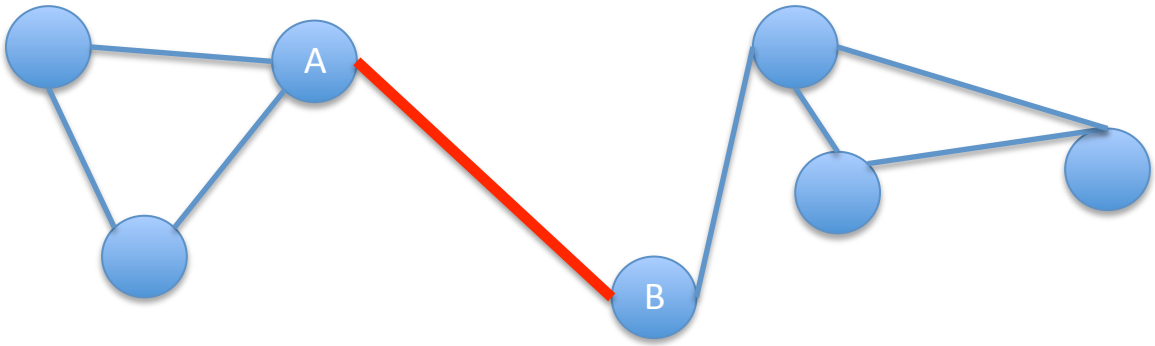
[Granovetter'74]

- Granovetter interviewed people about how they discovered their jobs
  - Most people did so through personal contacts
  - Often the personal contacts described as acquaintances and not close friends

- Basic intuition on this is: close friends are part of triad closures and would know what you know and would know others who would know what you know
- We will explain this more formally…

Bridges

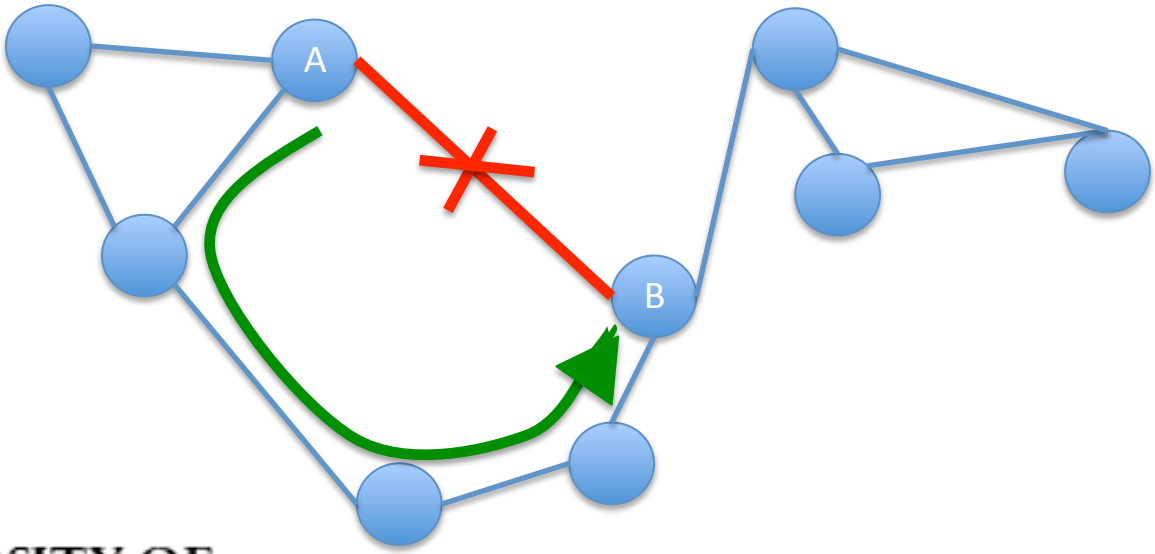
- Edge between A and B is a **bridge** if, when deleted, it would make A and B lie in 2 different components



UNIVERSITY OF CAMBRIDGE

Local Bridges

- An edge is a local bridge if its endpoints have no friends in common
  - If deleting the edge would increase the distance of the endpoints to a value more than 2.



UNIVERSITY OF CAMBRIDGE

Strong Triadic Closure Property (STPC)

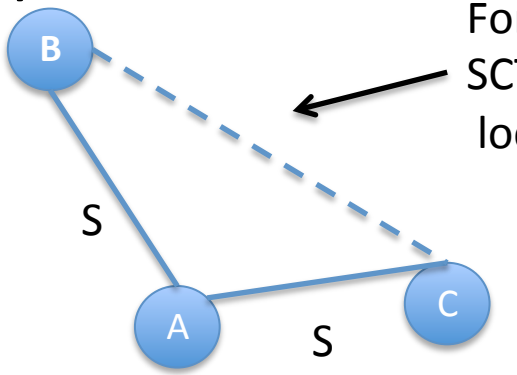
- Links between nodes have different "value": **strong and weak ties**
  - E.g: Friendship vs acquaintances

- **Strong Triadic Closure Property (Granovetter)**: If a node has two strong links (to B and C) then a link (strong or weak) must exist between B and C.

UNIVERSITY OF CAMBRIDGE

Local Bridges and Weak Ties

- If node A satisfies the SCTP 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)



For AC and AB to be a strong link
SCTP says BC must exist but
local bridge definition says it must not

- **Local bridges must be weak ties**

UNIVERSITY OF CAMBRIDGE

Real Data Validation

- Granovetter's theory remained not validated for years for large social networks due to the lack of data.
- [Onnela et al '07] tested it over a large cell-phone network (4 millions users):
  - Edge between two users if they called each other within the 18 months period.
  - Data exhibits a giant component (84%).
  - *Edge weight: time spent in conversation.*

UNIVERSITY OF CAMBRIDGE

Onnela et al. 2007

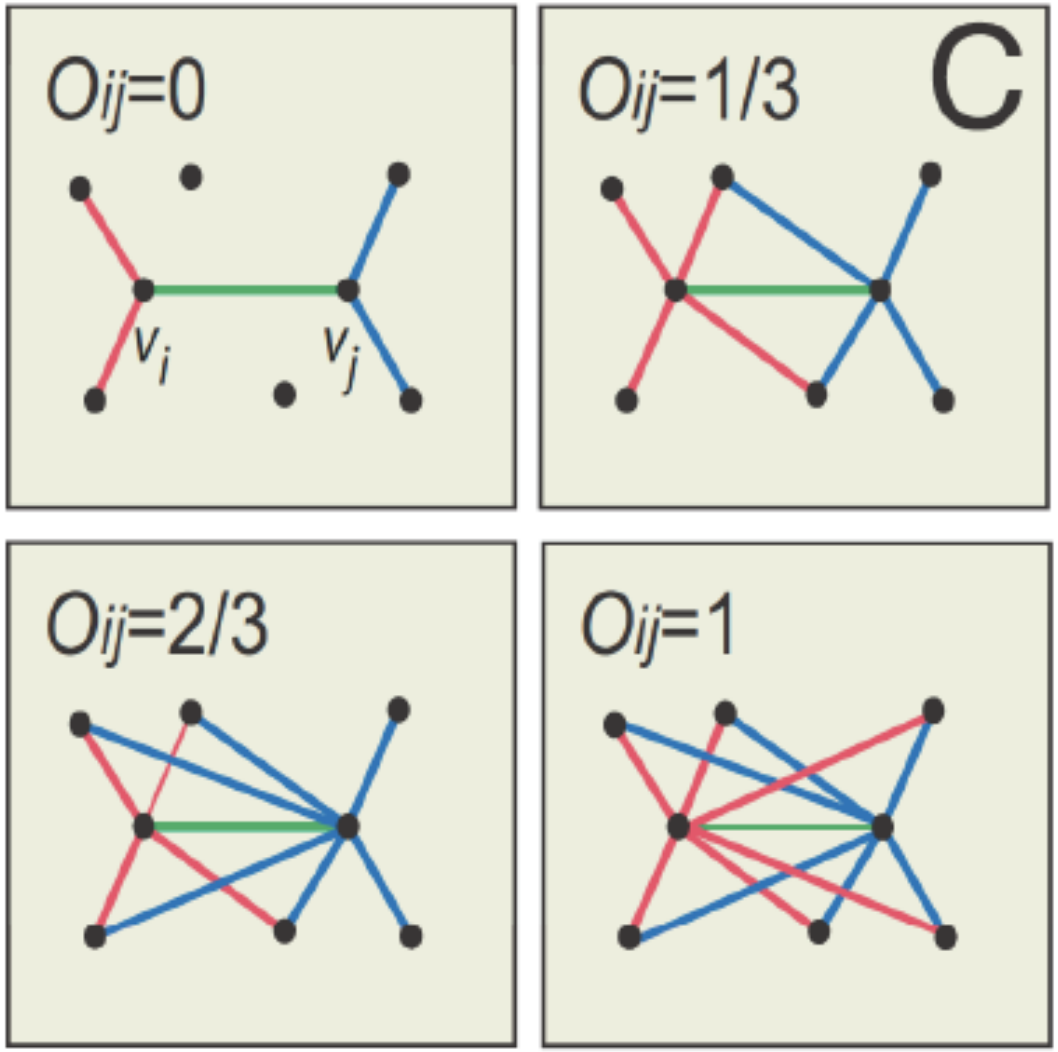
- Extending the definition of local bridge
- Given:

- **Neighbourhood overlap:**

$$\frac{\text{Number of nodes who are neighbours of both A \& B}}{\text{Number of nodes who are neighbours of at least A or B}}$$

- When the numerator is 0 the quantity is 0.
  - Numerator is 0 when AB is a local bridge
- The definition finds "almost local bridges" (~0)

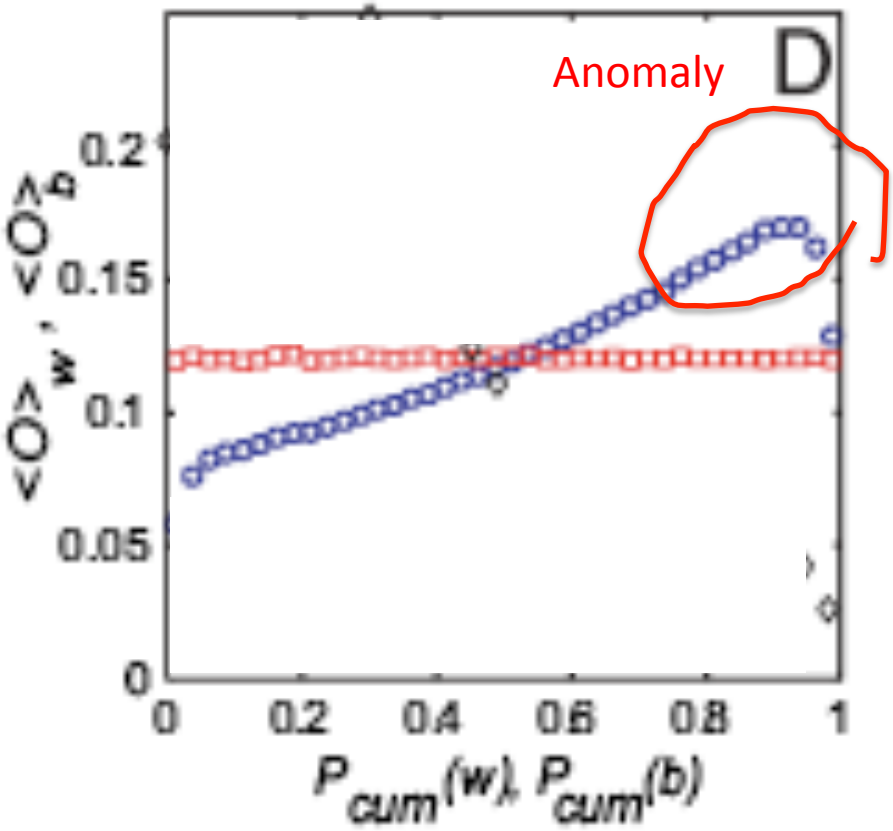
Neighbourhood overlap



UNIVERSITY OF CAMBRIDGE

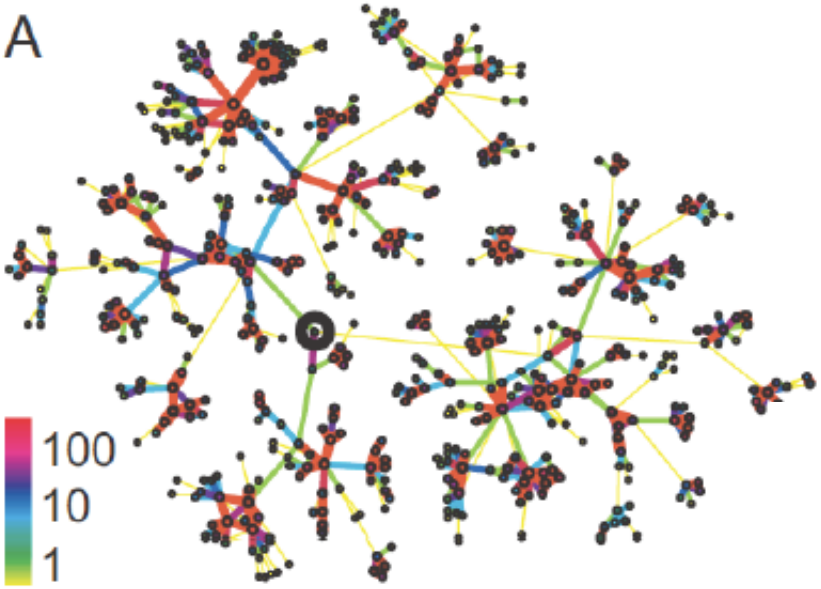
Relationship of Overlap with Tie Strength

- Red: random shuffled weights over links.
- Blue: real ones. Correlation with tie strength.



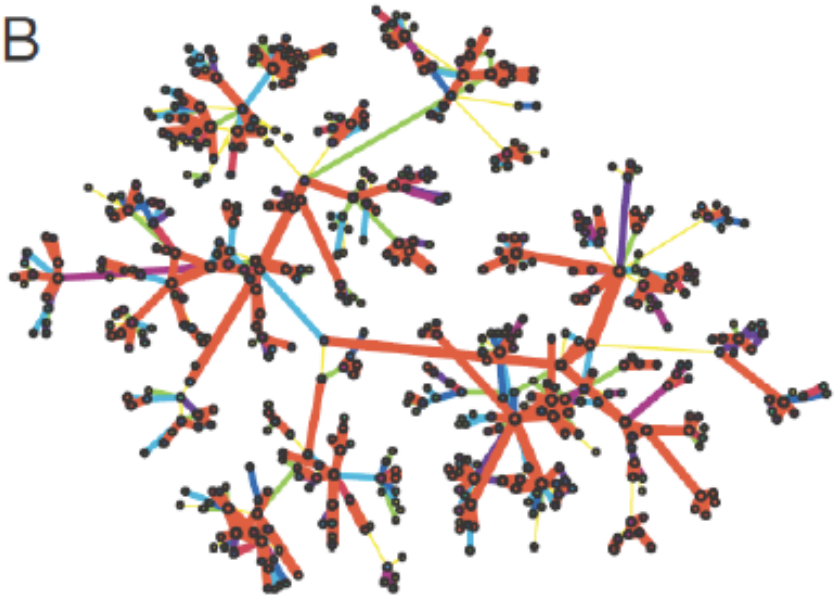
UNIVERSITY OF CAMBRIDGE

Real tie weights in a portion of the graph (around a random node)



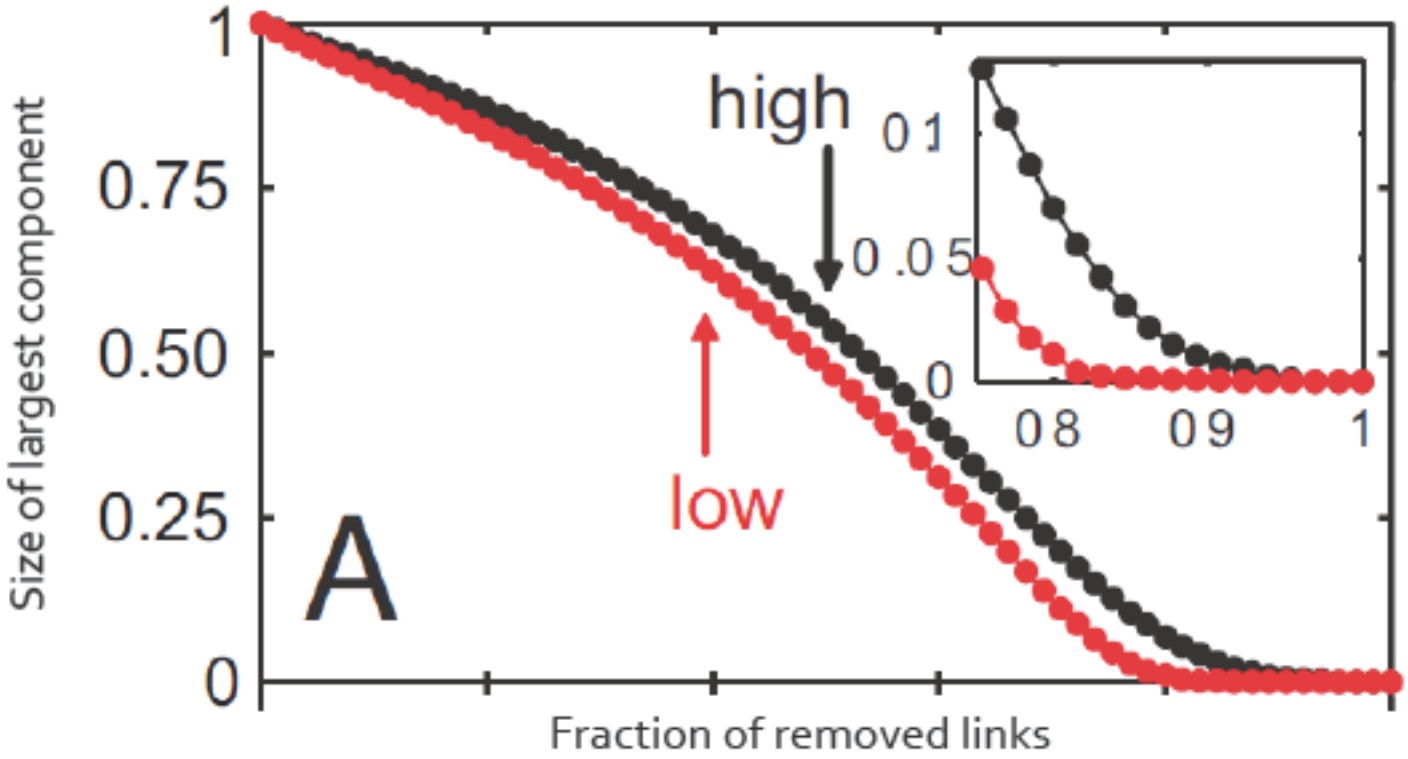
A= Real

B= Randomly shuffled



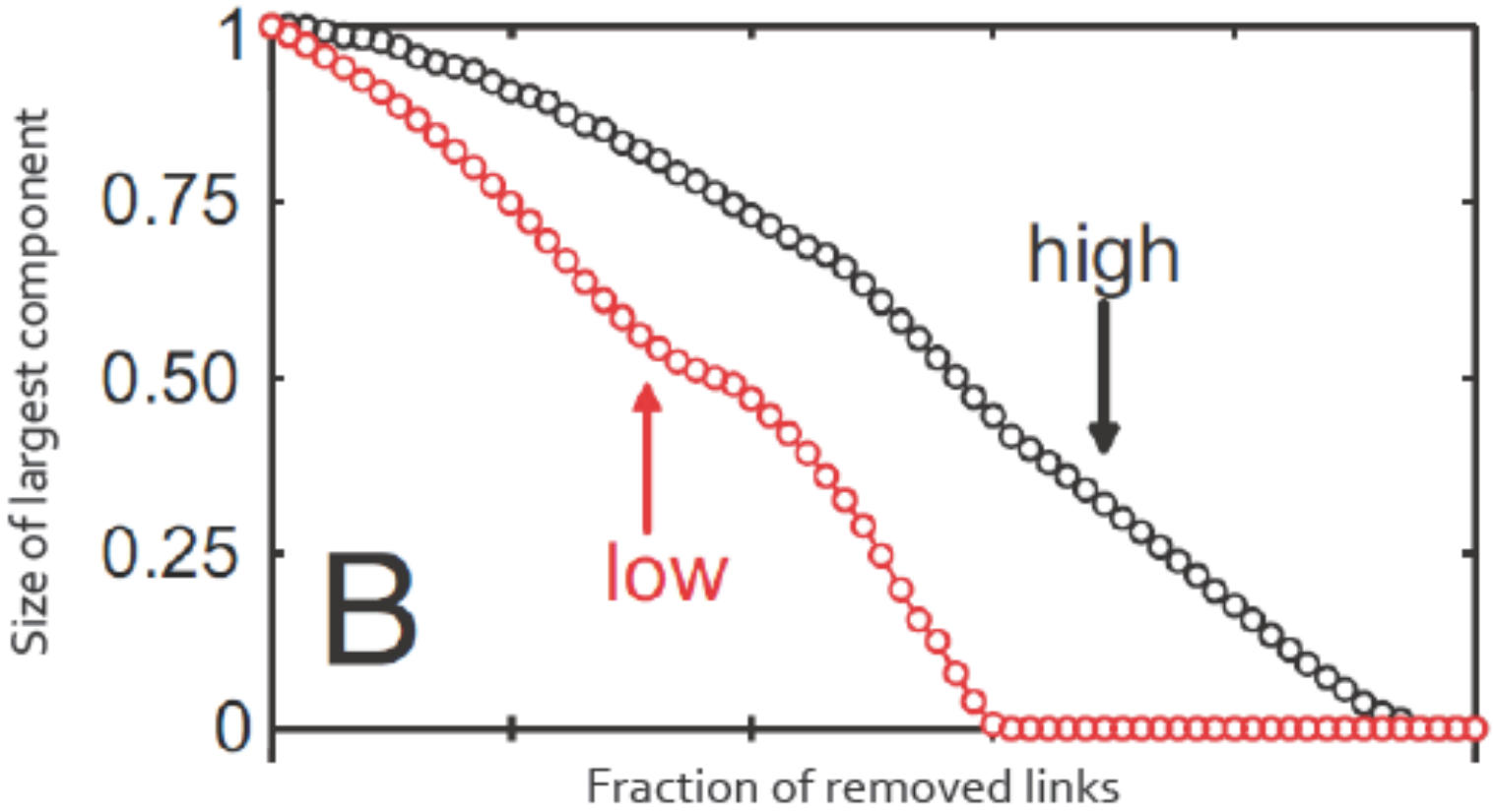
UNIVERSITY OF CAMBRIDGE

Effect of edge removal



UNIVERSITY OF CAMBRIDGE

Overlap based link removal



UNIVERSITY OF CAMBRIDGE

Weak ties matter!

- We have just seen that weak ties matter and if they are removed, they lead to a breakdown in the network.

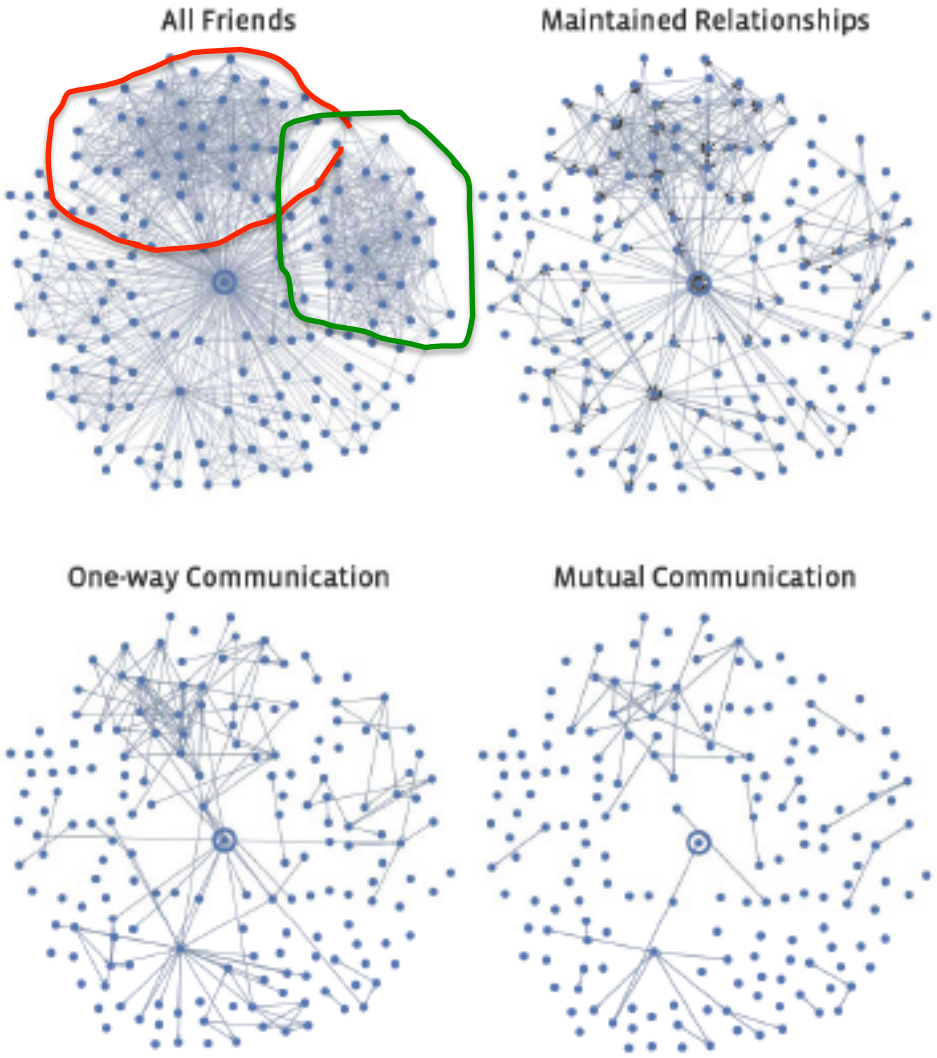
- If strong ties are removed they lead to a smooth degrading of the network

Facebook Example

- Facebook data analysis of one month of data
- Four networks:
  - Declared friendship
  - Reciprocal communication (messages)
  - One way communication
  - Maintained relationship: clicking on content on news feed from other friend or visiting profile more than once.

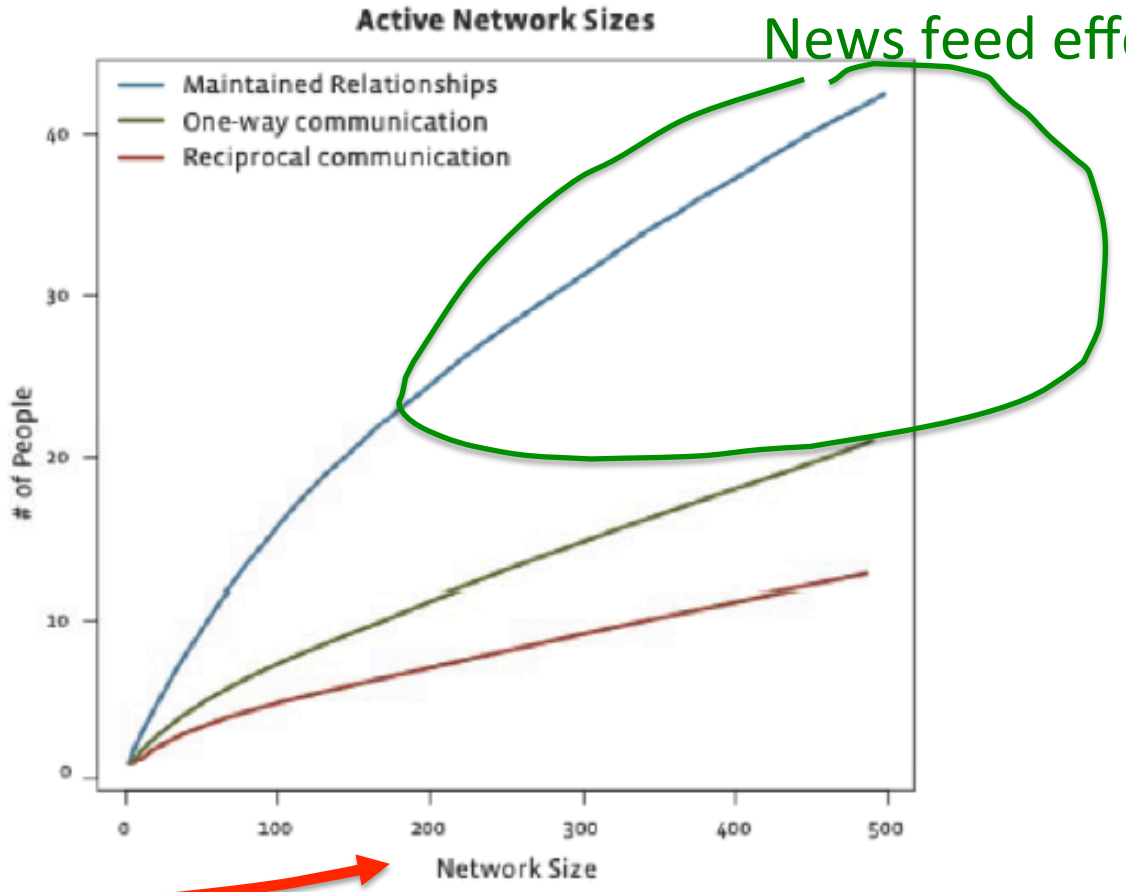
UNIVERSITY OF CAMBRIDGE

What does it look like? (one random user)



UNIVERSITY OF CAMBRIDGE

Active Network Size: number of links



News feed effect

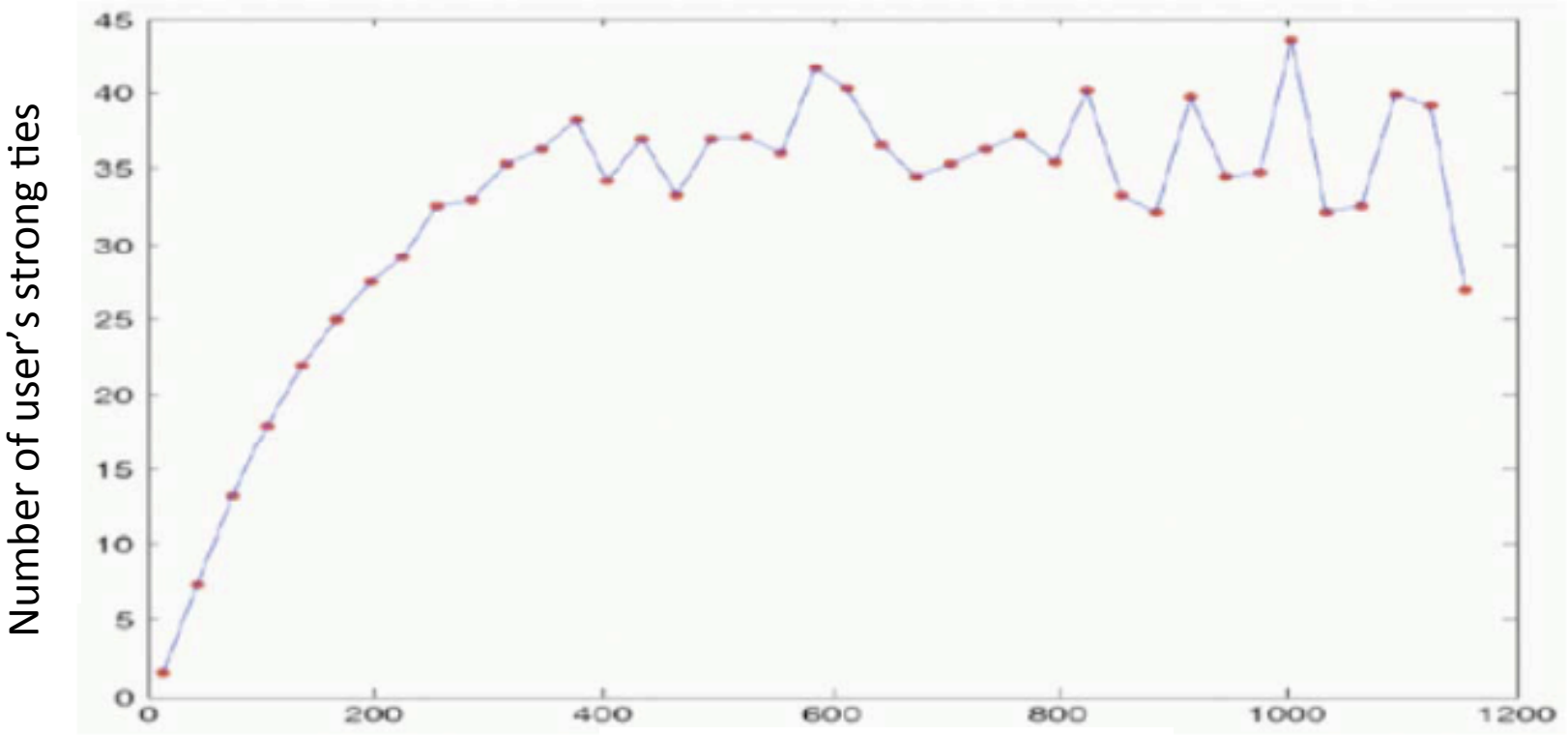
Declared friends

UNIVERSITY OF CAMBRIDGE

Twitter Analysis

- Huberman at al. have analyzed strong and weak ties in Twitter.
- The "followers" graph in Twitter is directed
  - Someone can follow someone else who does not follow him
- Messages of 140 chars can be posted
- Messages can be addressed to specific users (although they stay readable to all)
- **Weak ties**: users followed
- **Strong ties**: users to whom the user sent at least 2 messages in the observation period

Twitter



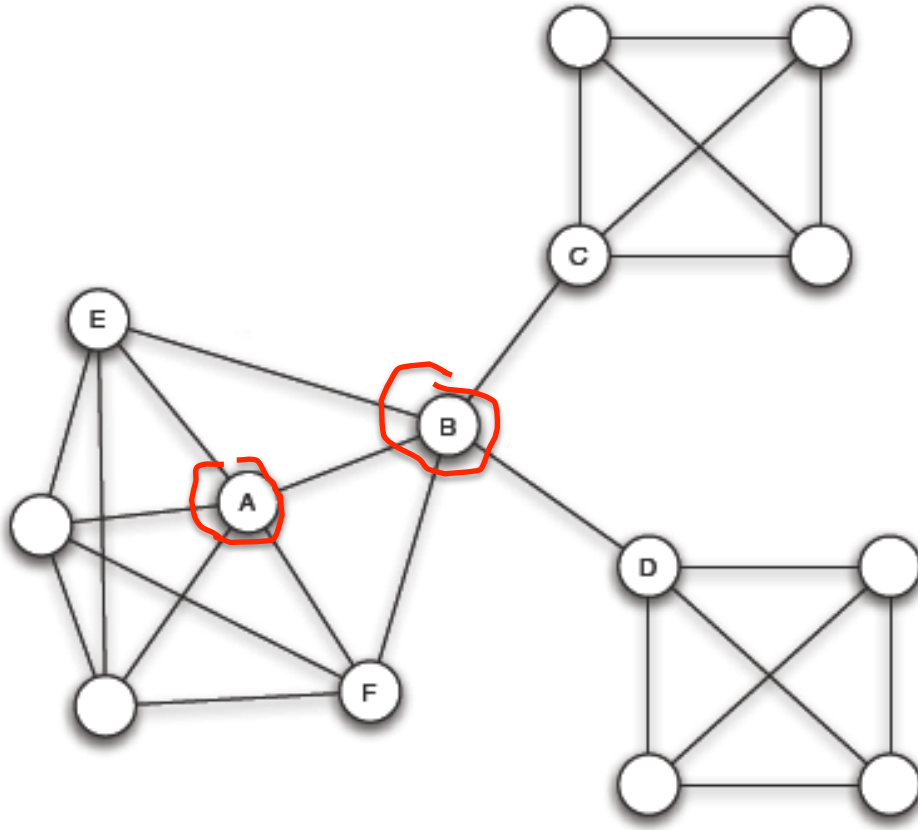
UNIVERSITY OF CAMBRIDGE

Number of strong ties stays below ~50

Embeddedness

- Emdeddedness of an edge: number of common neighbours of the 2 end points.
- A-B value is 2
- A has high clust. coeff.
- B *spans a structural hole*
- Local bridges have

Embeddedness of 0



UNIVERSITY OF CAMBRIDGE

Weak ties and Communities

- Weak ties seem to bridge groups of tightly coupled nodes (communities)
- How do we find these communities?

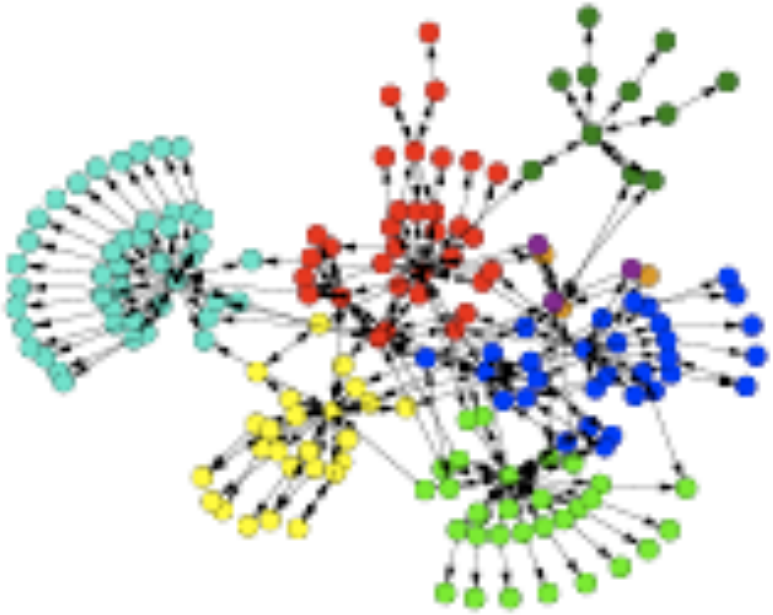
Why do we want to find partitions/communities?

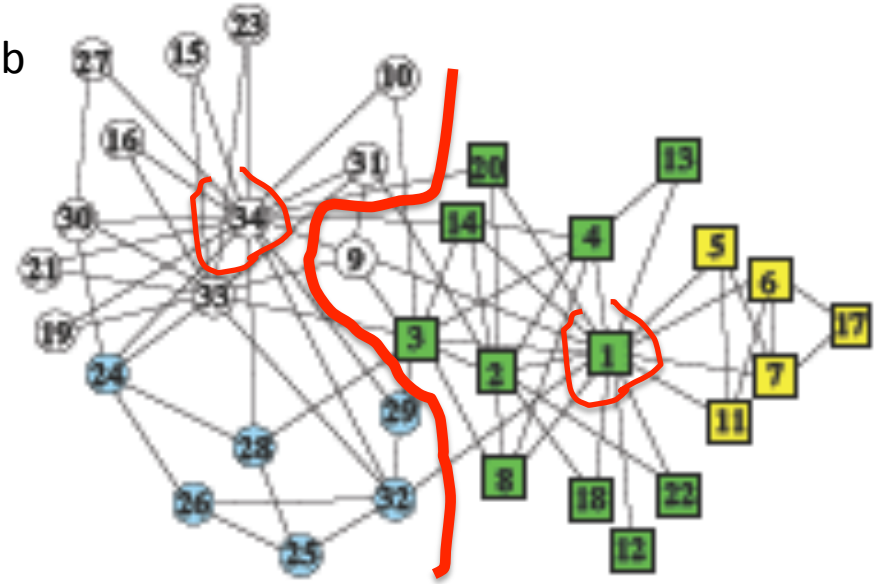
- Clustering web clients with similar interest or geographically near can improve performance
- Customers with similar interests could be clustered to help recommendation systems
- Clusters in large graphs can be used to create data structures to efficient storage of graph data to handle queries or path searches
- Detect artificial improvements of PageRank
- Study the relationship/mediation among nodes
  - Hierarchical organization study

UNIVERSITY OF CAMBRIDGE

Example

Zachary's Karate club: 34 members of a club over 3 years. Edges: interaction outside the club





WWW: pages and hyperlinks
Identification of clusters can improve pageranking

UNIVERSITY OF CAMBRIDGE

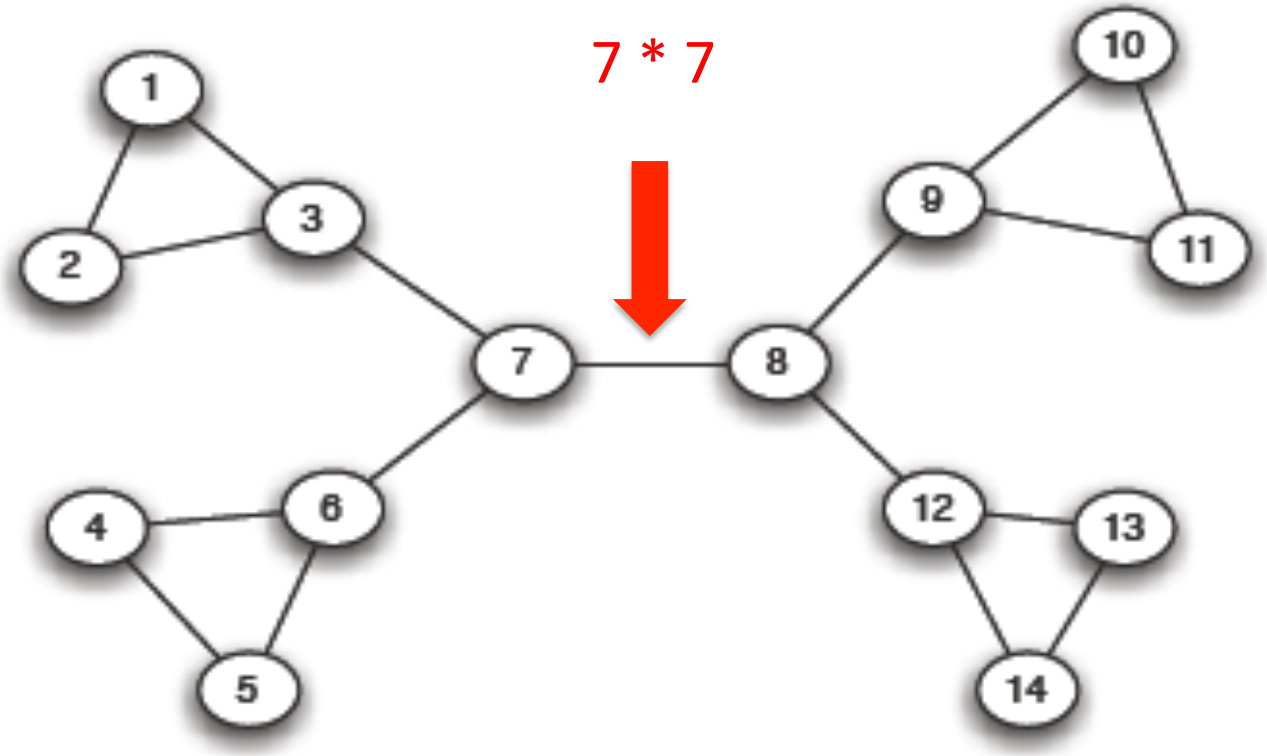
Remove weak ties

- Local bridges connect weakly interacting parts of the network
- What if we have many bridges: which do we remove first? Or there might be no bridges.

- Note: **Without those bridges paths between nodes would be longer**

Edge Betweenness

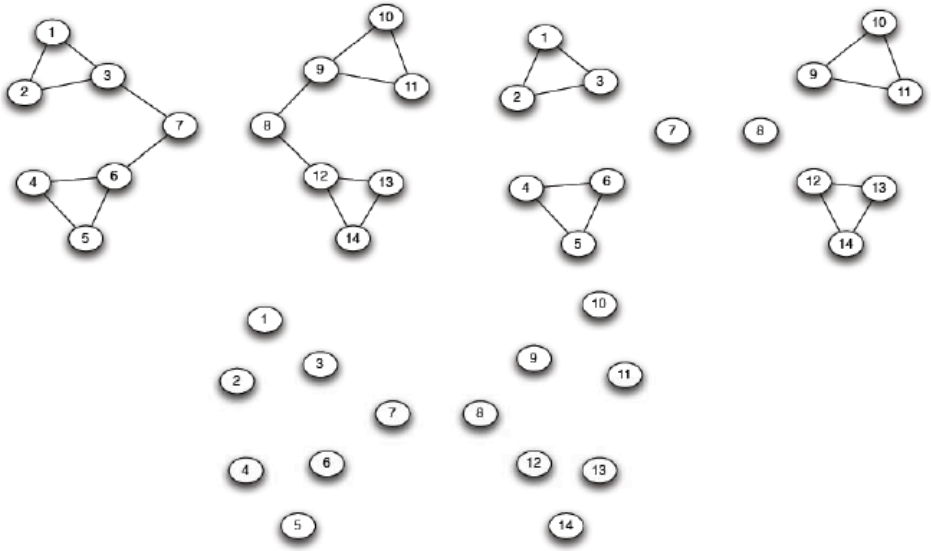
- Edge Betweenness: the number of shortest paths between pairs of nodes that run along the edge.



UNIVERSITY OF CAMBRIDGE

Algorithm of Girvan-Newmann (PNAS 2002)

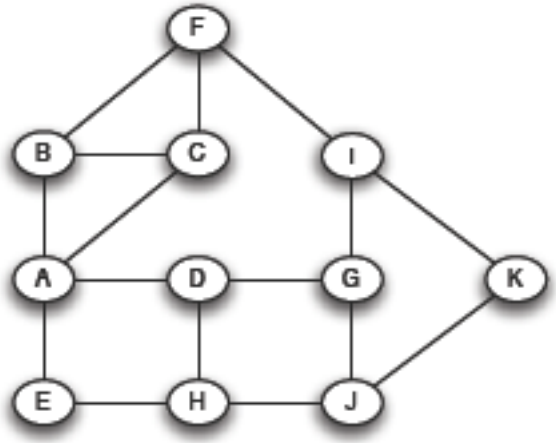
- Calculate the betweenness of all edges
- Cut the edge with highest betweenness
- Recalculate edge betweenness

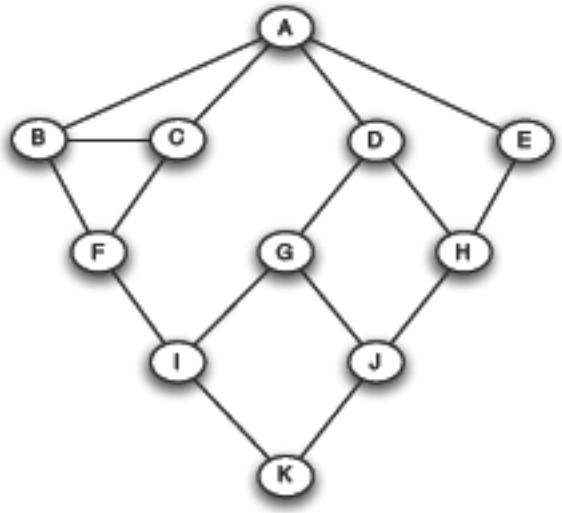


UNIVERSITY OF CAMBRIDGE

How is the betweenness computed?

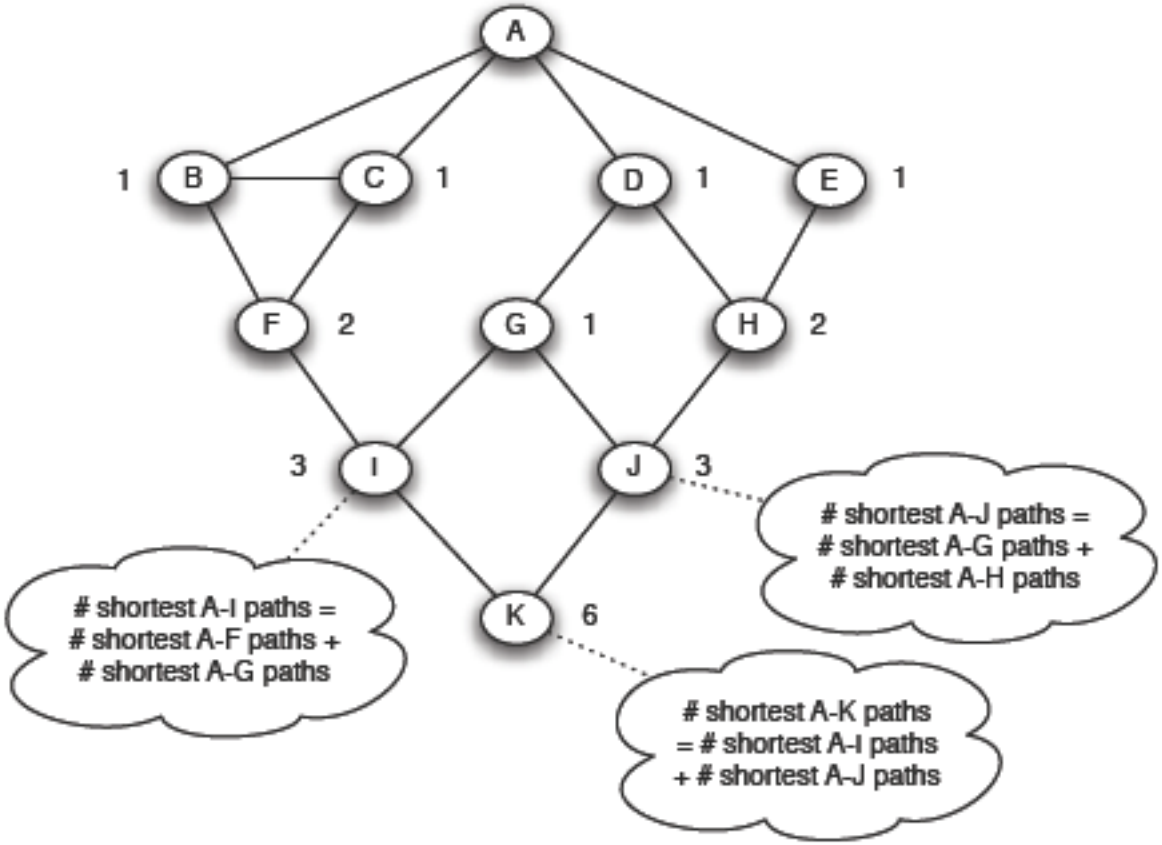
- Calculate the shortest paths from node A
  - BFS search from A.
  - Determine number of shortest paths from A to each node.





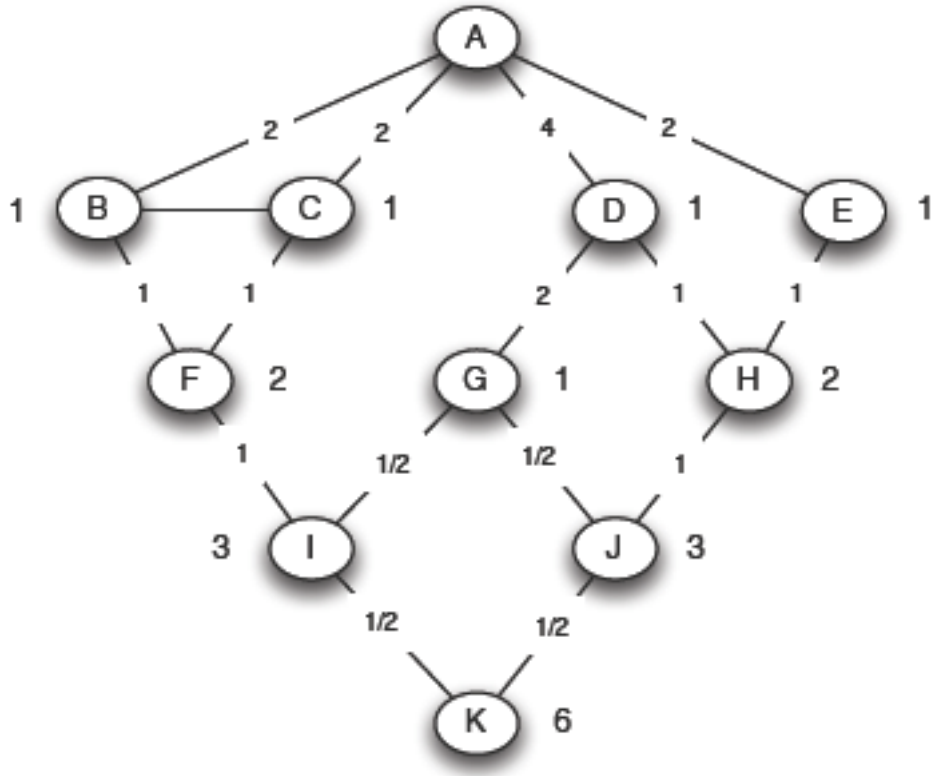
UNIVERSITY OF CAMBRIDGE

Calculating number of shortest paths



UNIVERSITY OF CAMBRIDGE

Calculating flows



UNIVERSITY OF CAMBRIDGE

Calculating Edge Betweenness

- Build one of these graphs for each node in the graph
- Sum the values on the edges on each graph to obtain the edge betweenness

Community Detection

- How do we know when to stop?

- When X communities have been detected?
- When the level of cohesion inside a community has reached Y?

- There is no prescriptive way for every case
- There are also many other ways of detecting communities

References

- Kleinberg's book: Chapter 3.

- **Structure and tie strengths in mobile communication networks.** J. P. Onnela, J. Saramaki, J. Hyvonen, G. Szabo, D. Lazer, K. Kaski, J. Kertesz, A. L. Barabasi. *Proceedings of the National Academy of Sciences*, Vol. 104, No. 18. (13 Oct 2006), pp. 7332-7336.
- **Maintained relationships on facebook.** Cameron Marlow, Lee Byron, Tom Lento, and Itamar Rosenn. 2009. On-line at http://overstated.net/2009/03/09/maintained- relationships-on-facebook.
- **Social networks that matter: Twitter under the microscope.** Bernardo A. Huberman, Daniel M. Romero, and Fang Wu. First Monday, 14(1), January 2009.
- **Community structure in social and biological networks** Michelle Girvan and Mark E. J. Newman. Proc. Natl. Acad. Sci. USA, 99(12): 7821–7826, June 2002.