Social and Technological Network Analysis

Lecture 3: Centrality Measures and Community Detection

Dr. Cecilia Mascolo
(some material from Lada Adamic's
course lectures)

UNIVERSITY OF CAMBRIDGE

In This Lecture

- We will introduce the concept of centrality and the various measures which have been associated to this concept.
- We will use centrality (betweenness) to isolate communities.
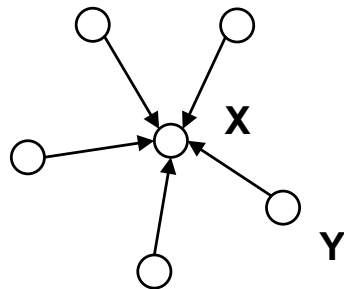
UNIVERSITY OF
CAMBRIDGE

Centrality

- Finding out which is the most central node is important:
  - It could help disseminating information in the network faster
  - It could help stopping epidemics
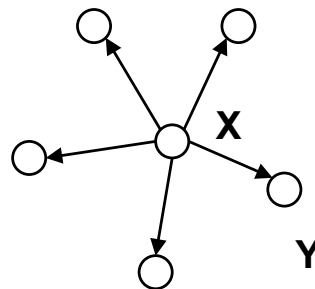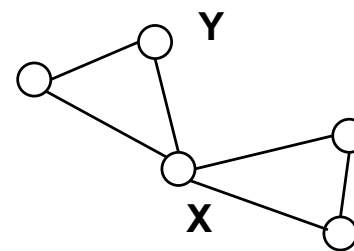 - It could help protecting the network from breaking

Centrality: visually
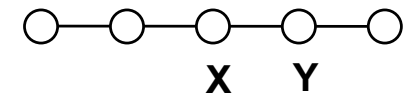
- Centrality can have various meanings:
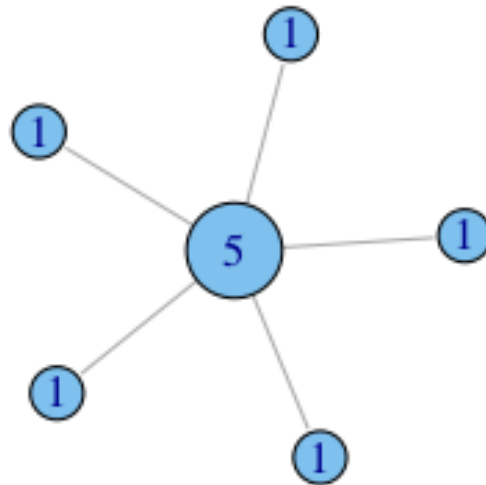

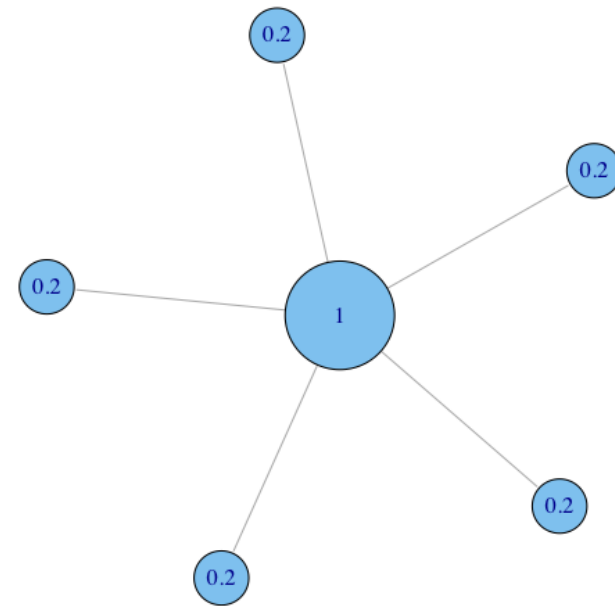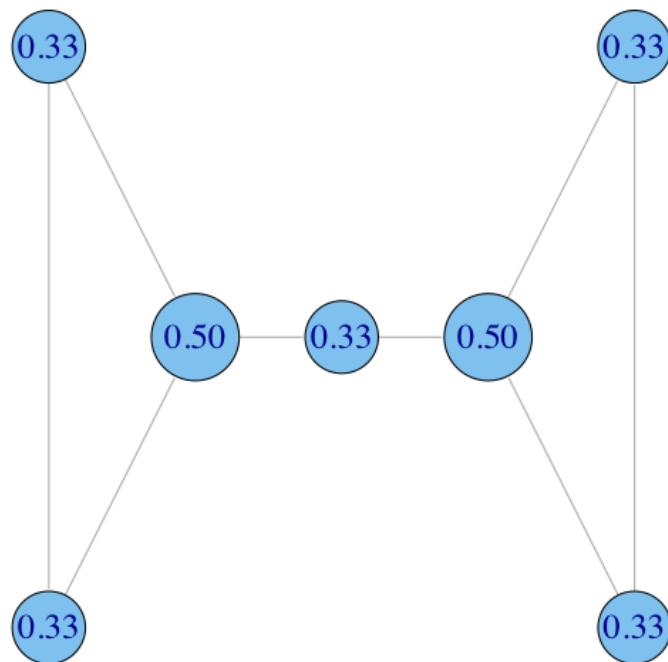| indegree | outdegree | betweenness | closeness |

Degree Centrality

When is the number of connections the best centrality measure?
- o people who will do favors for you
- o people you can talk to / have a beer with

UNIVERSITY OF
CAMBRIDGE

Normalization

- Divide for the max number of nodes (N-1)

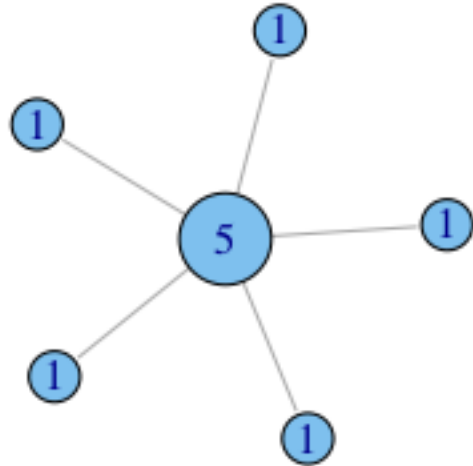UNIVERSITY OF
CAMBRIDGE

Freeman's Network Centrality

- How do we calculate the value of centrality of the network
    - To check how much variation there is among the nodes (heterogeneity?)

Max value of Degree Centrality in the Network

Max value of the above: when network is a star: 1 node has C=N-1 and all others (N-1) have 1.

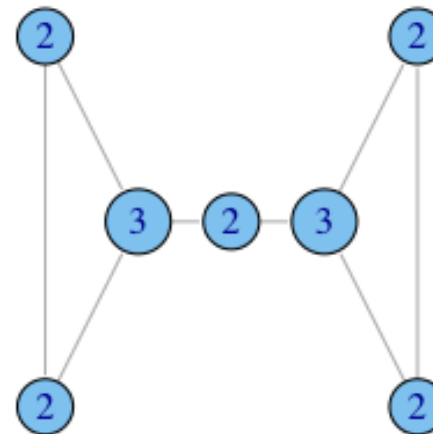$$C_D = \frac{\sum_{i=1}^{s}\left[C_D(n^*) - C_D(i)\right]}{[(N-1)(N-2)]}$$

UNIVERSITY OF
CAMBRIDGE

Freeman's Network Centrality
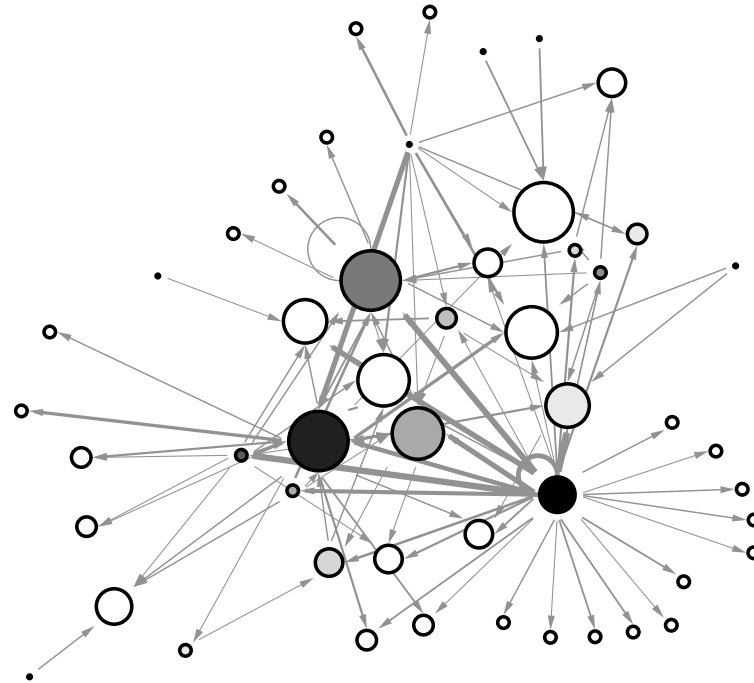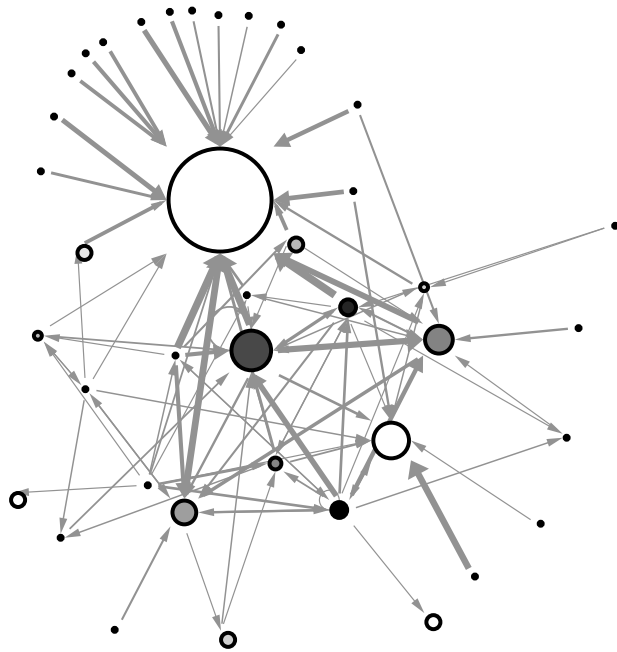

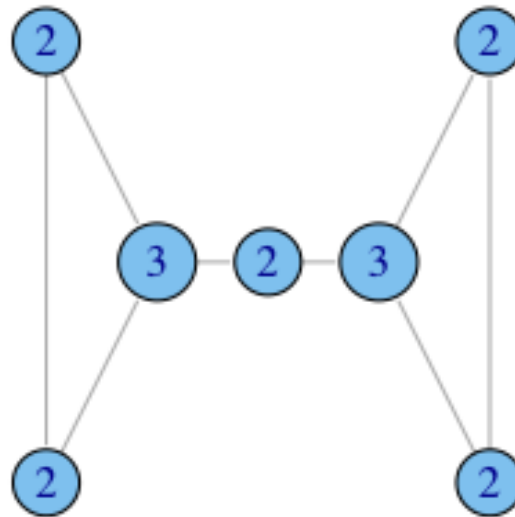$C_D = 1.0$

$C_D = 0.167$

$C_D = 0.167$

UNIVERSITY OF
CAMBRIDGE

Examples: Financial Networks

When is Degree Centrality not so good?

When is Degree Centrality not so good (2)?
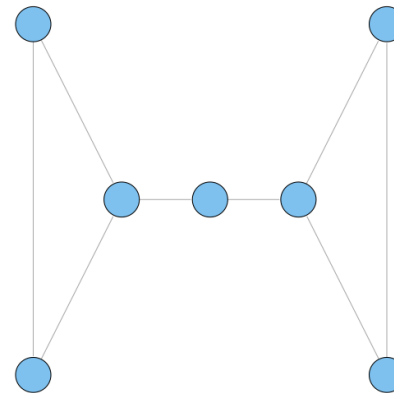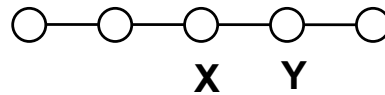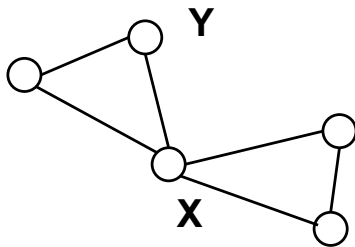
- Ability to broker between groups
- Likelihood that information originating anywhere in the network reaches you...

UNIVERSITY OF CAMBRIDGE

Betweenness Centrality

- intuition: how many pairs of individuals would have to go through you in order to reach one another in the minimum number of hops?
- who has higher betweenness, X or Y?

UNIVERSITY OF
CAMBRIDGE

Betweenness (Formally)

$$C_B(i) = \sum_{j<k} g_{jk}(i) / g_{jk}$$

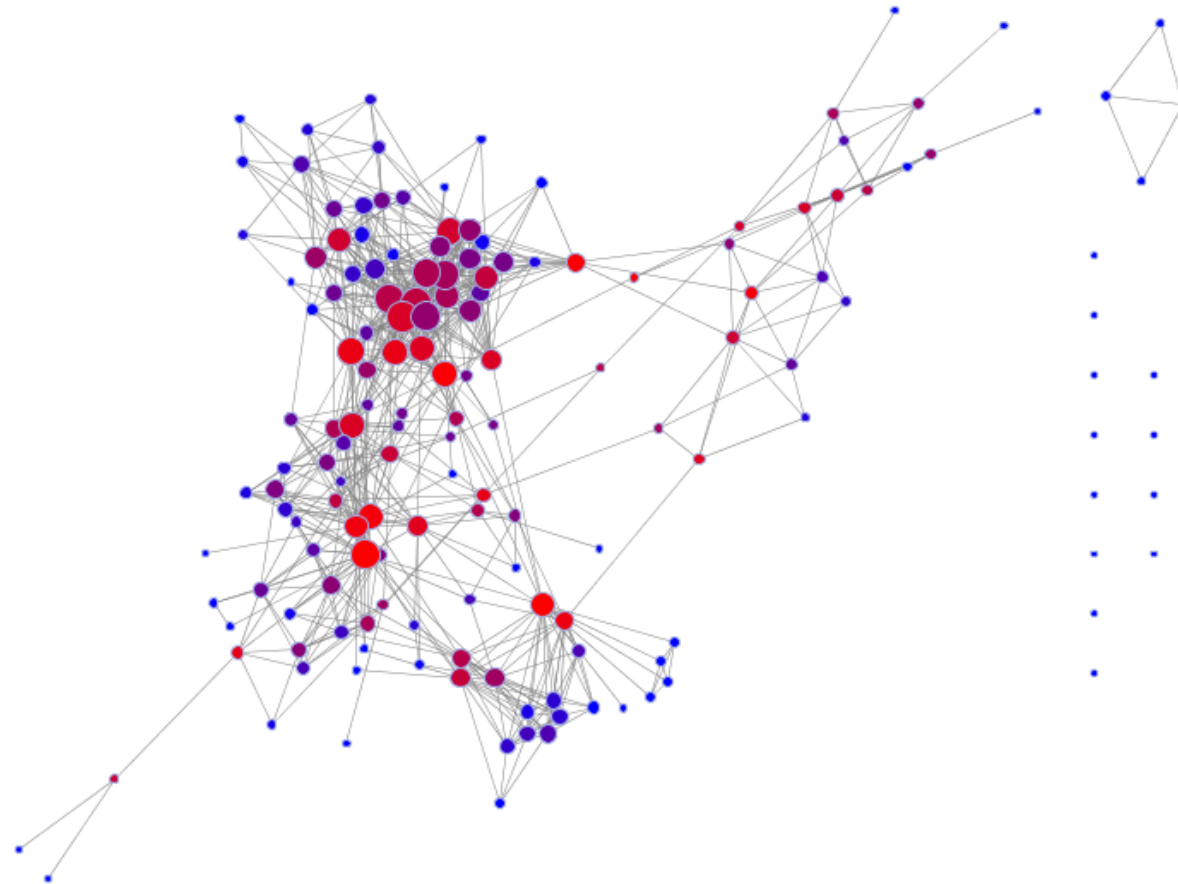Where $g_{jk}(i)$ = the number of shortest paths connecting *jk* passing through i

g_{jk} = total number of shortest paths

Usually normalized by:

$$C'_B(i) = C_B(i) \boxed{/[(n-1)(n-2)/2]}$$

number of pairs of vertices excluding the vertex itself

Facebook Example (Adamic)

- A lies between no two other vertices
- B lies between A and 3 other vertices: C, D, and E
- C lies between 4 pairs of vertices (A,D),(A,E), (B,D),(B,E)

- note that there are no alternate paths for these pairs to take, so C gets full credit

0 — 3 — 4 — 3 — 0

Closeness Centrality

- What if it is not so important to have many direct friends?
- Or be "between" others

- But one still wants to be in the "middle" of things, not too far from the center

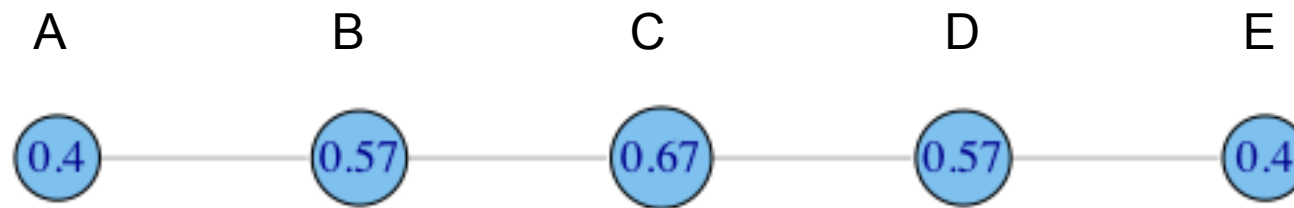UNIVERSITY OF CAMBRIDGE

Closeness Centrality (Formally)

- Closeness is based on the length of the average shortest path between a vertex and all vertices in the graph

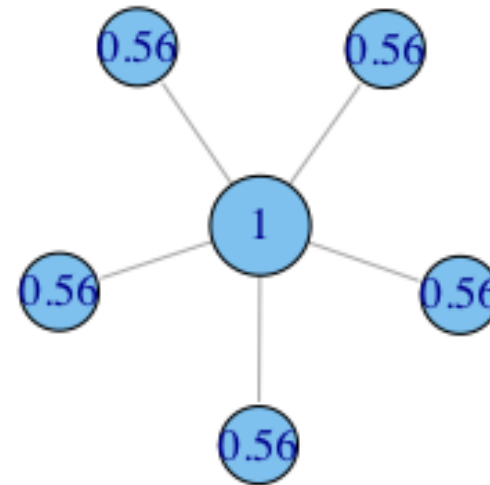$$C_c(i) = \left[\sum_{j=1}^{N} d(i,j) \right]^{-1}$$

$$C_C'(i) = (C_C(i)) / (N-1)$$

Closeness: Example

A   B   C   D   E

0.4 — 0.57 — 0.67 — 0.57 — 0.4
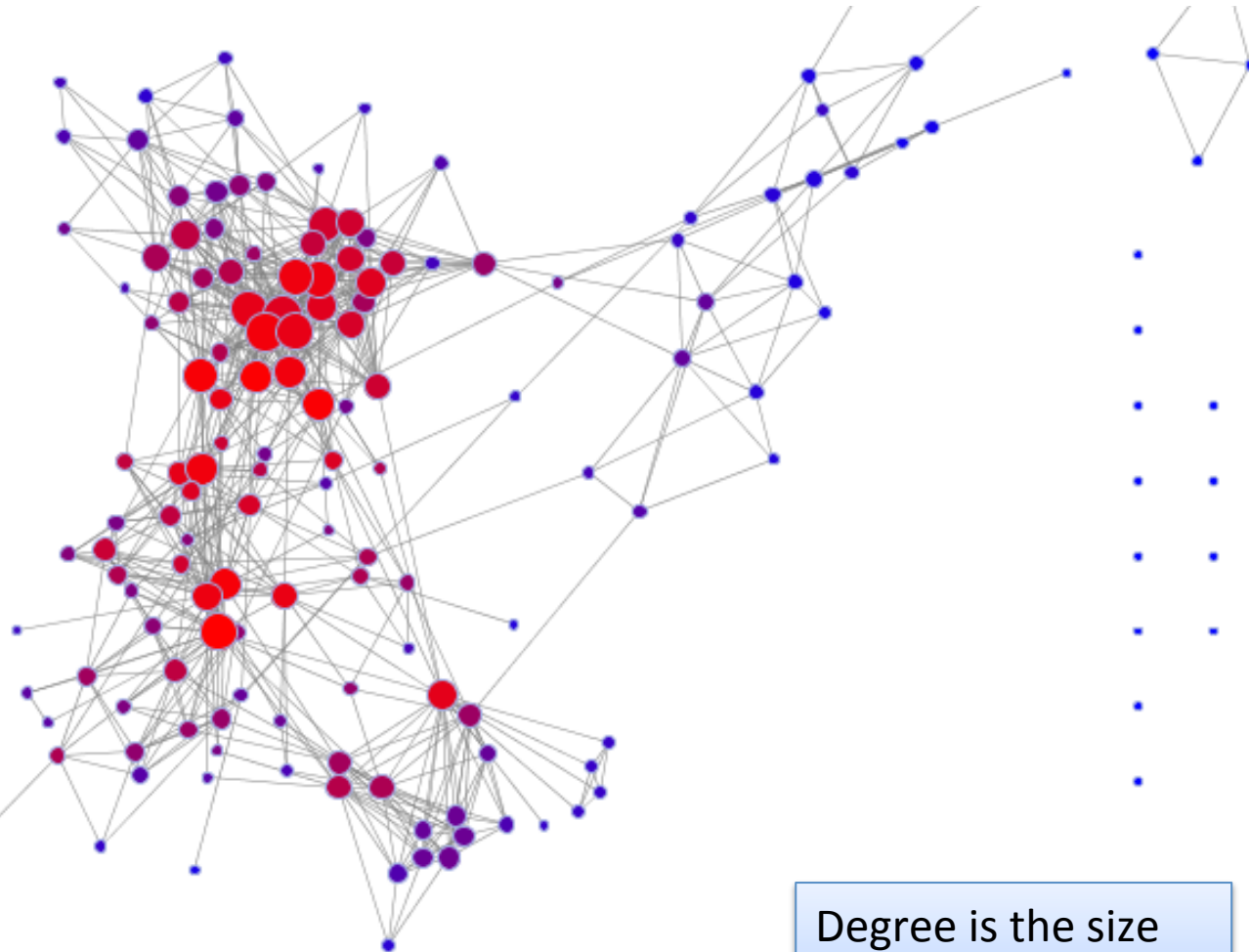
$$C'_c(A) = \left[\frac{\sum_{j=1}^{N} d(A,j)}{N-1} \right]^{-1} = \left[\frac{1+2+3+4}{4} \right]^{-1} = \left[\frac{10}{4} \right]^{-1} = 0.4$$

UNIVERSITY OF
CAMBRIDGE
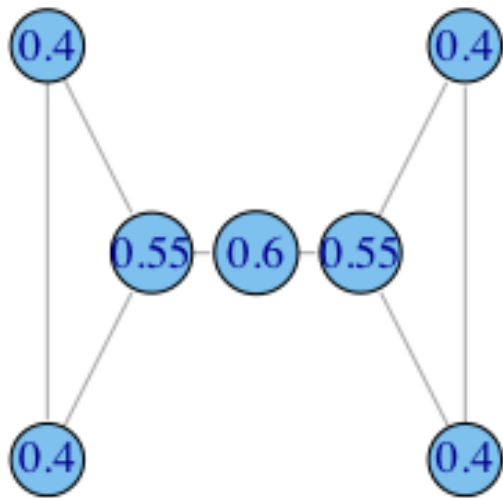
Examples

Example: Facebook (Adamic)

Degree is the size
Color is closeness

UNIVERSITY OF
CAMBRIDGE

Other measures

- The influence of a node is another measure of centrality: information centrality.
- Betweenness counts the number of shortest paths, but one could count the number of paths (Katz Centrality).

UNIVERSITY OF
CAMBRIDGE

Communities

- Weak ties (Lecture 2) seemed to bridge groups of tightly coupled nodes (communities)
- How do we find these communities?

UNIVERSITY OF
CAMBRIDGE

network communities

utomatically find

nodes?

UNIVERSITY OF
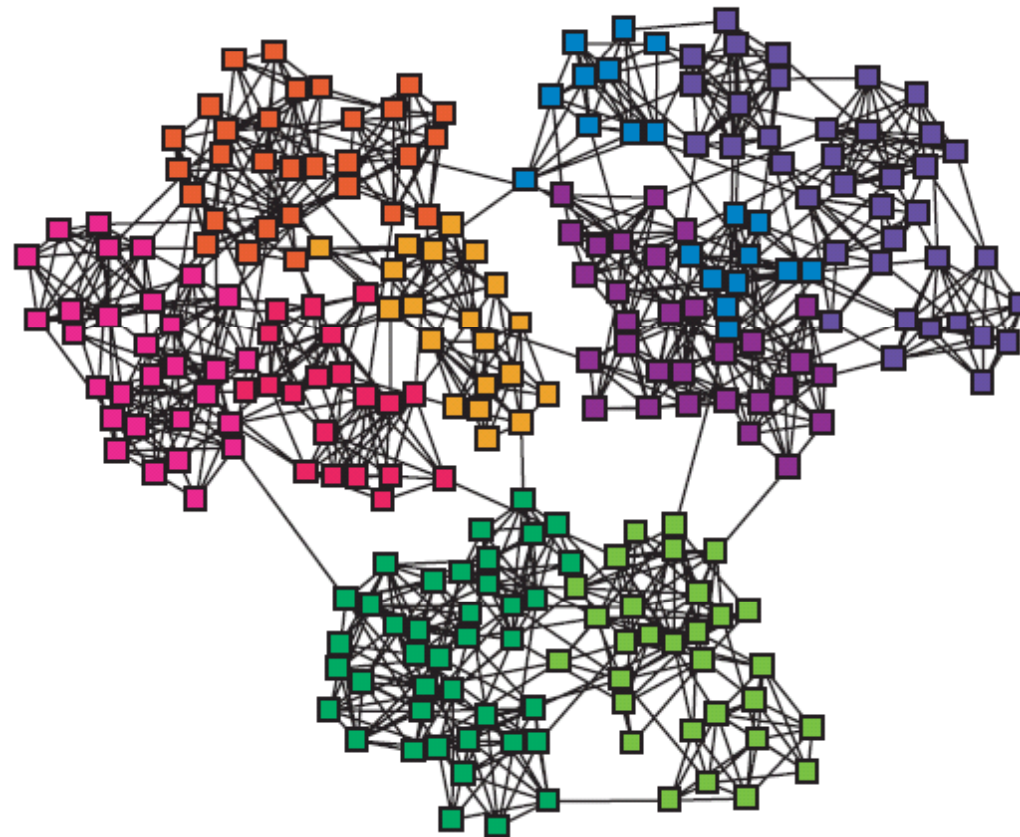
Why do we want to find partitions/communities?

- 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
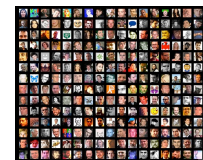over 3 years. Edges: interaction outside the club
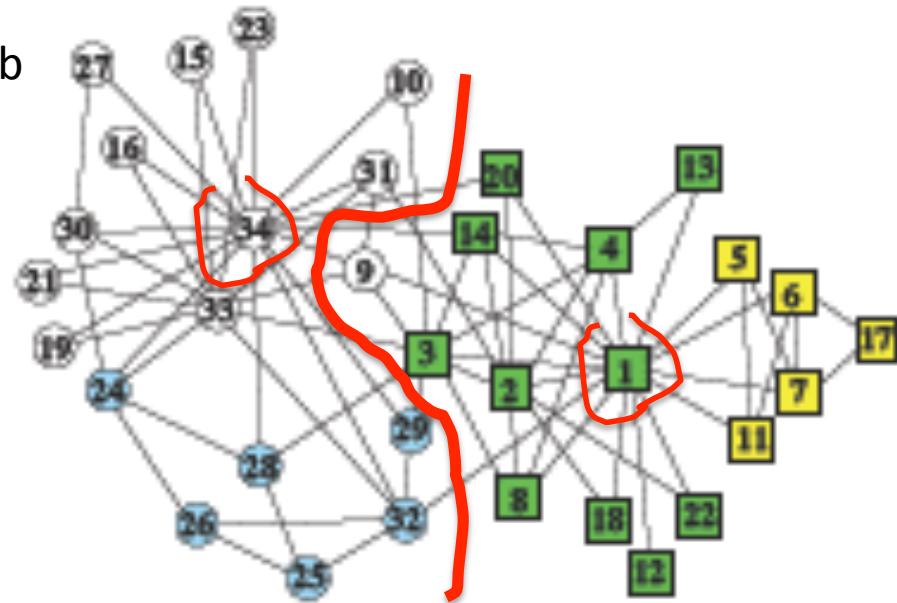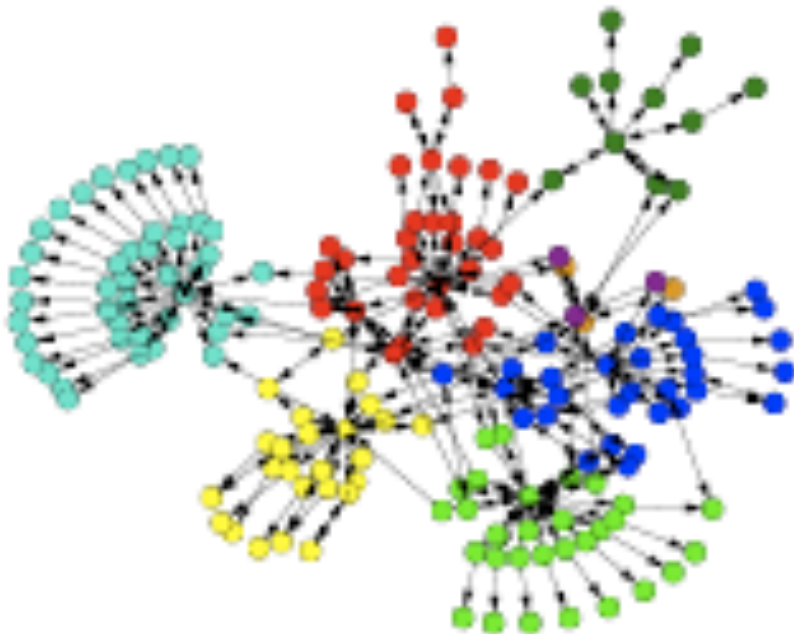
WWW: pages and hyperlinks
Identification of clusters can improve
pageranking

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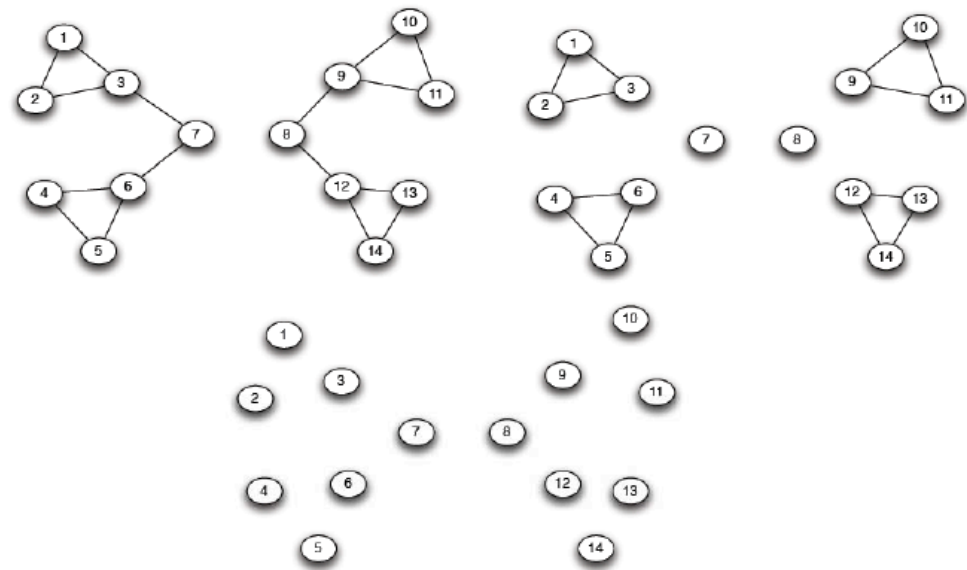
- Edge Betweenness: the number of shortest paths between pairs of nodes that run along the edge.

7 * 7
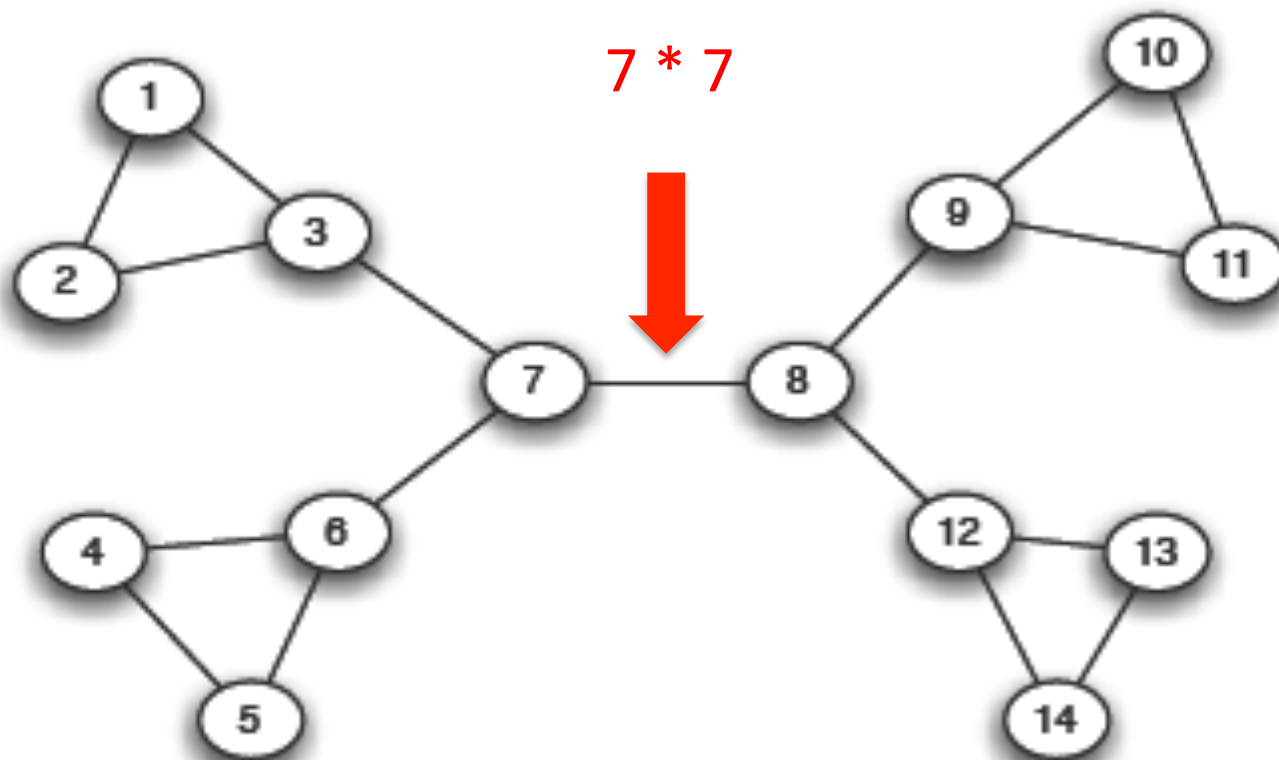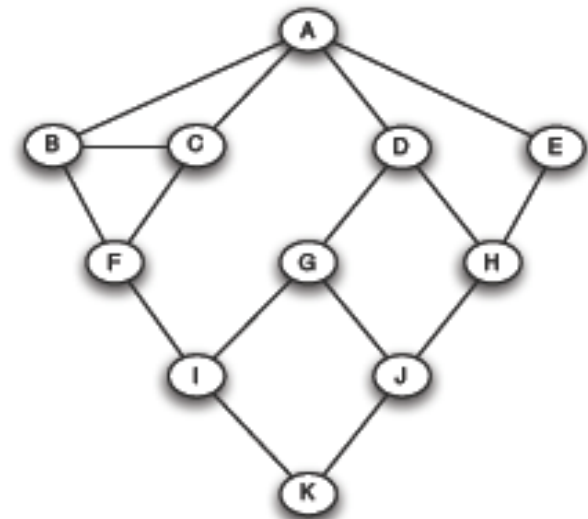
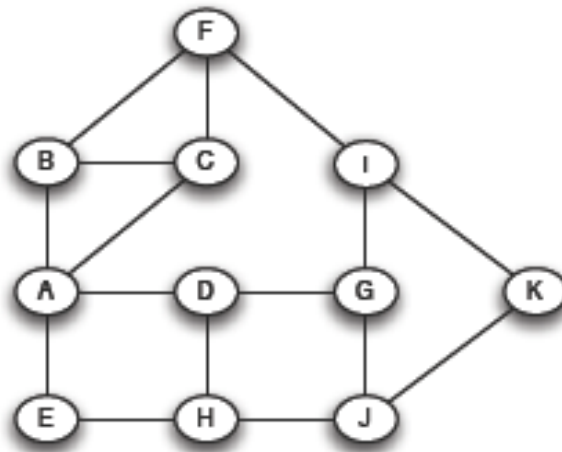# Algorithm of Girvan-Newmann (PNAS 2002)

- Calculate the betweenness of all edges
- Cut the edge with highest betweenness
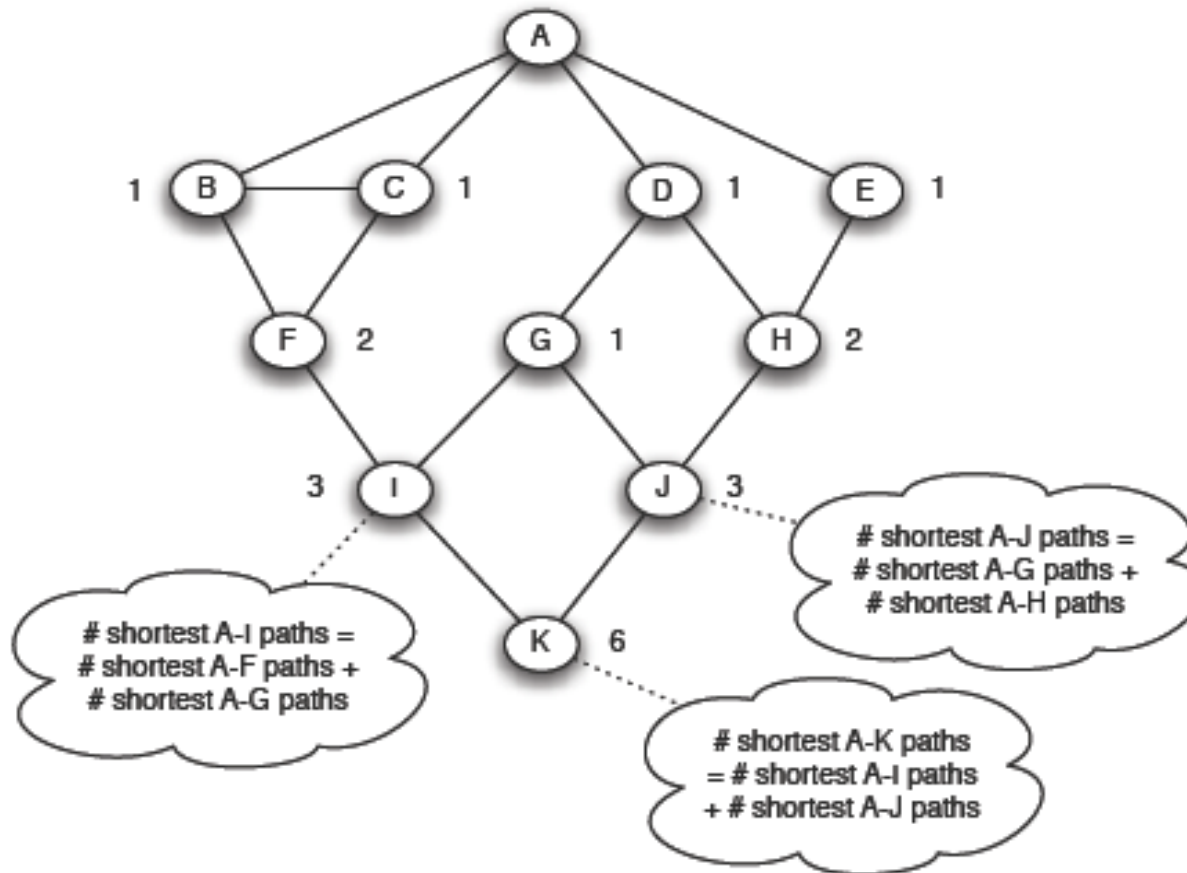- Recalculate edge betweenness
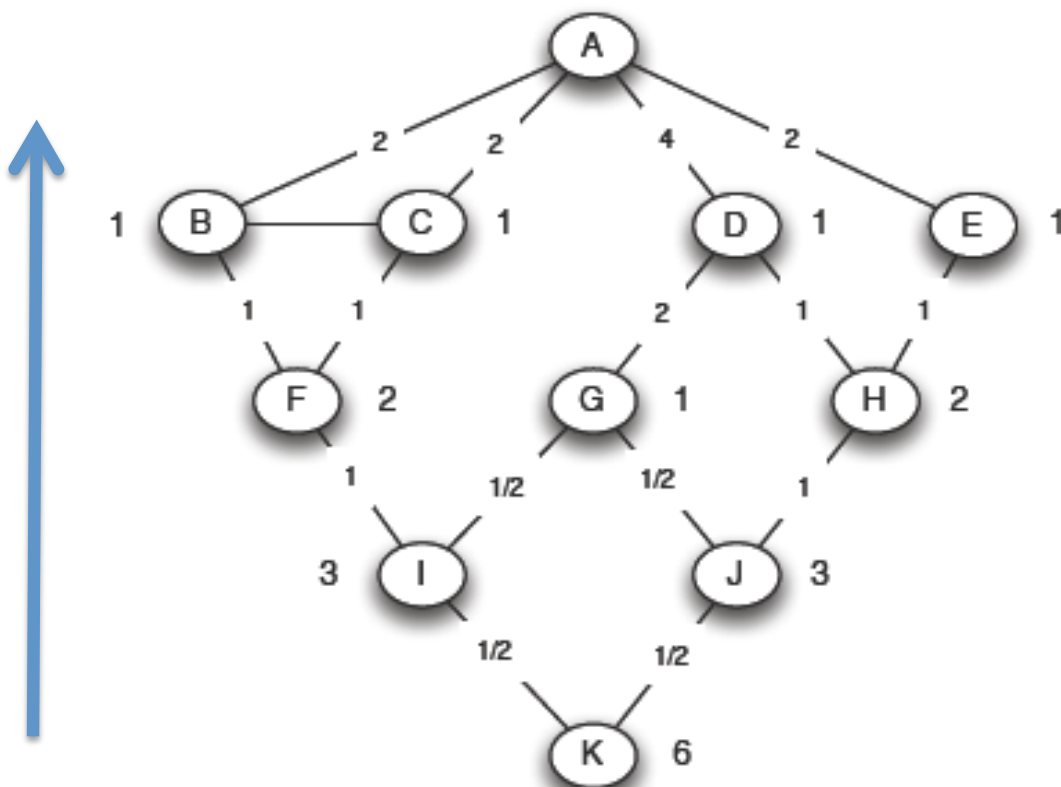
How is the betweenness computed?

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

Calculating number of shortest paths

Calculating flows
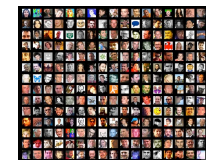
When we get to a node X in the breadth-first search structure, working up from the bottom, we add up all the flow arriving from edges directly below X, plus 1 for the flow destined for X itself. We then divide this up over the edges leading upward from X, in proportion to the number of shortest paths coming through each.

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

UNIVERSITY OF
CAMBRIDGE

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

UNIVERSITY OF
CAMBRIDGE

Summary

- We have described various measures of centrality.
- We have shown how betweenness can be used to isolate communities.

References

- Kleinberg's book: Chapter 3.
- **Ack: L. Adamic's slides**

- M. Newmann. **Networks**. Oxford University Press. April 2010.
- **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.

UNIVERSITY OF
CAMBRIDGE