Lecture 5

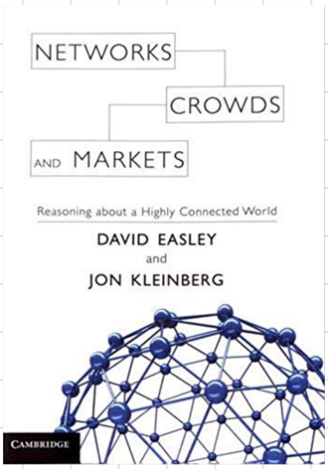
Network Science

Strong and Weak Ties:

Partitions and Betweenness

Today's Topics

- Closure, Structural Holes, and Social Capital
- Betweenness Measures and Graph Partitioning



Chapter 3

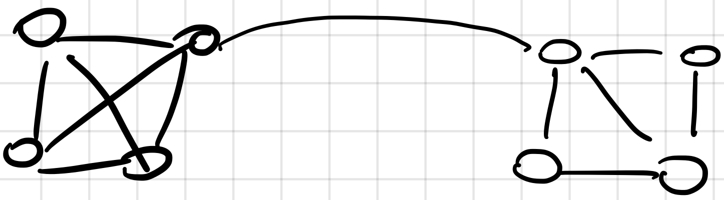
"Strong and Weak ties"

Heteogeneity

tightly connected groups

vs

week ties (local bridges)



- different roles
- measures

focus on edges that span across different regions

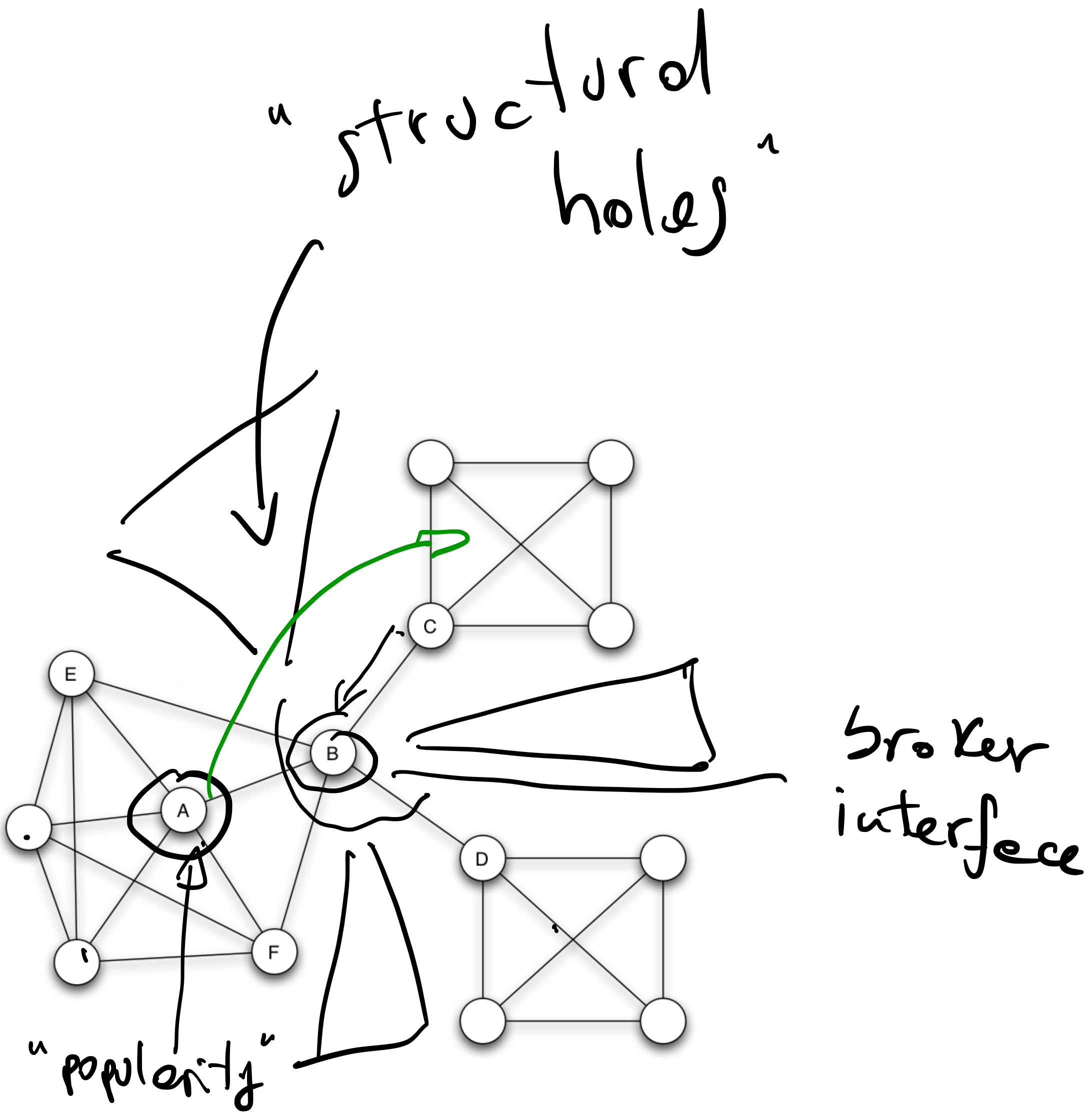


Figure 3.11: The contrast between densely-knit groups and boundary-spanning links is reflected in the different positions of nodes A and B in the underlying social network.

Ronald Burt, 2000

"organizations"

Embeddedness

$$NO_{AB} = \frac{\#\ \text{common neigh. of } A \& B}{\#\ \text{neigh. of } A \& B}$$

Embeddedness = numerator of NO

"similar" the clustering coefficient

three measures that correlate "a lot"

When the interaction is "on Display"

embeddedness $\Leftrightarrow$ misbehavior

$\Downarrow$

Social consequences

weak ties have a kind of privilege :

not the same consequences on misbehavior

Structural Holes

Burt

structural holes

advantage in organisations

- "influence"
- informational advantage
- "creativity"

latent power

Bad things:

- : manipulation
- "gate keeper"

Social Capital

"the ability of actors to secure benefits by virtue of membership in social netw. or other social structures"

(A. Portes)

individuals vs organizations

Social Capital and triadic Closure

Burt : "social capital is a tension between closure and brokerage"

bounding vs bridging

Take home message

- social structures are facilitators of actions by individual and groups

- networks are at the hearth of such discussions

Betweenness Measures and Graph Partitioning

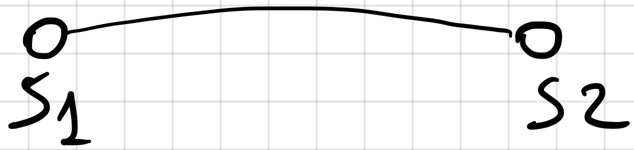
Networks are made of tightly-knit regions connected by means of sparser interconnetions

We do not have introduced so far a formal definition of such regions although we have some useful measures and definitions:

- clustering coefficient
- local bridges
- neighborhood overlap
- triadic closure ...

the problem of finding denser regions in a network is called: Graph Partitioning or also Community Detection

Example 1: co-authorship network



s_1 and s_2 are network scientists that co-authored at least one scientific paper

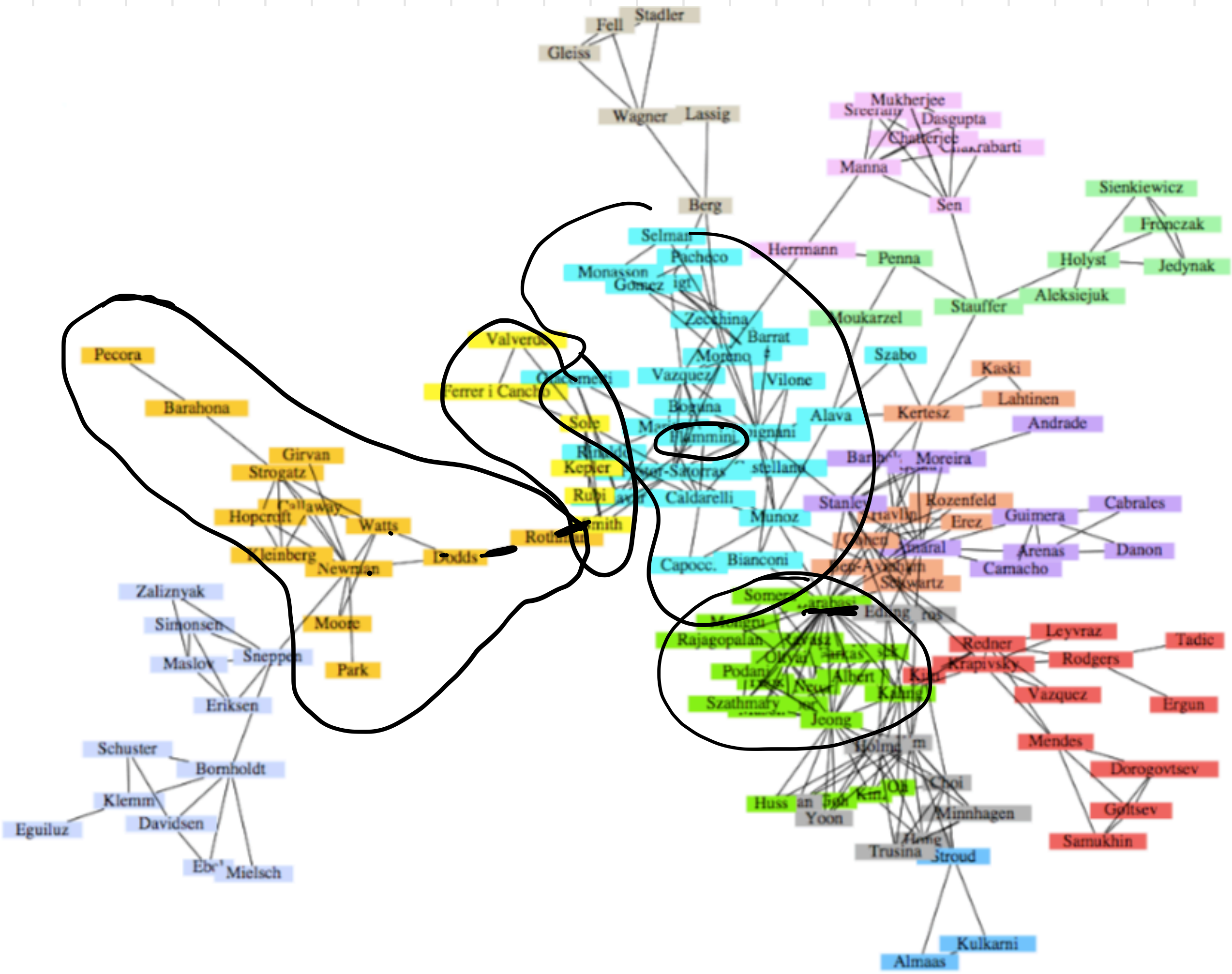


Figure 3.12: A co-authorship network of physicists and applied mathematicians working on networks [322]. Within this professional community, more tightly-knit subgroups are evident

Example 2: Zachary's Karate Club

can I use structure itself to predict the "fault line"

president

minimum cut

instructor

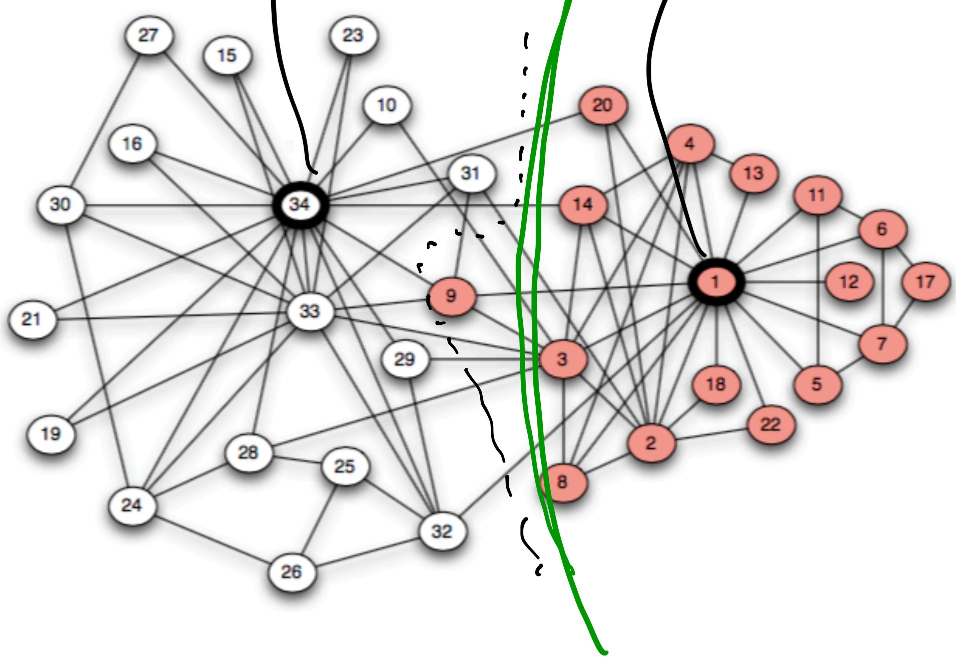


Figure 3.13: A karate club studied by Wayne Zachary [421] — a dispute during the course of the study caused it to split into two clubs. Could the boundaries of the two clubs be predicted from the network structure?

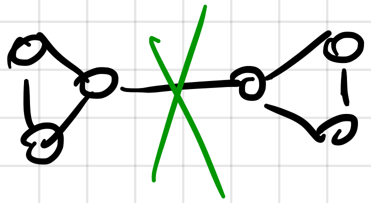
General Approaches to Graph Partitioning

divisive

top-down

"spanning links"

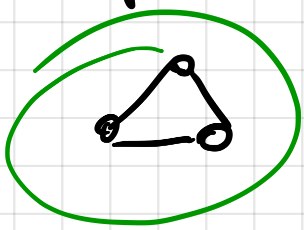
goal: remove



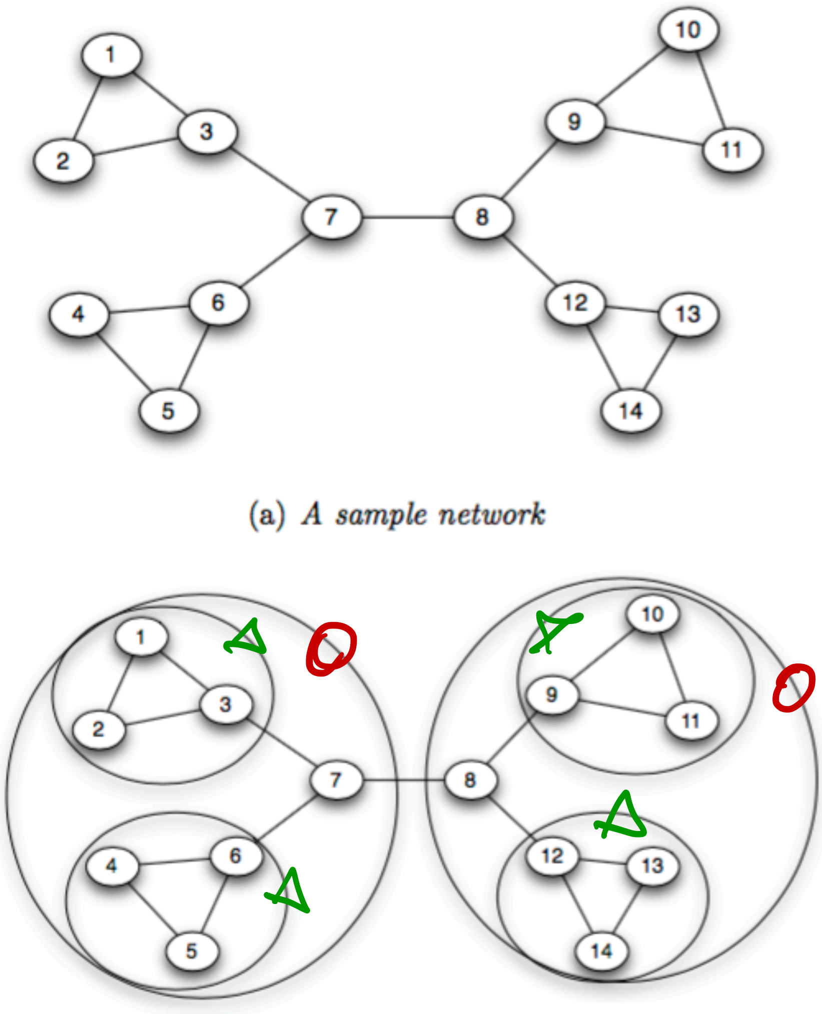
agglomerative

bottom-up

nodes merged with other nodes to form groups



Nested structure



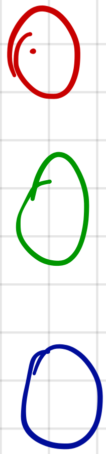
(a) *A sample network*

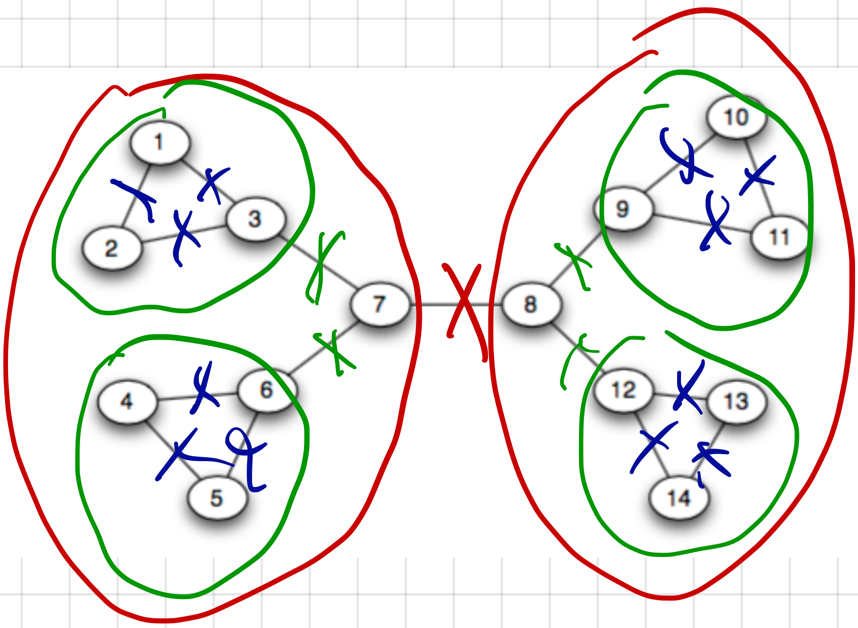
(b) *Tightly-knit regions and their nested structure*

Figure 3.14: In many networks, there are tightly-knit regions that are intuitively apparent, and they can even display a *nested* structure, with smaller regions nesting inside larger ones.

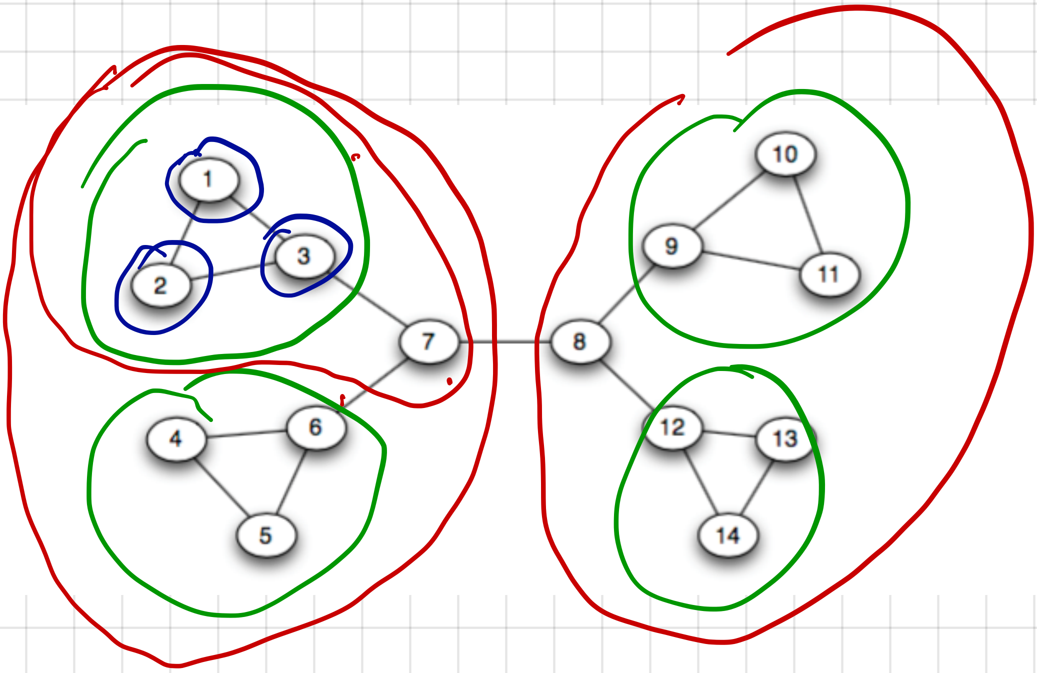
Girven - Newman

Divisive method





Agglomerative Method



Which edge to remove first in divisive methods?

removing local bridges ...

if disconnected: distances will be longer ($\to \infty$)

how many shortest paths will be affected if edge (x, y) is removed?

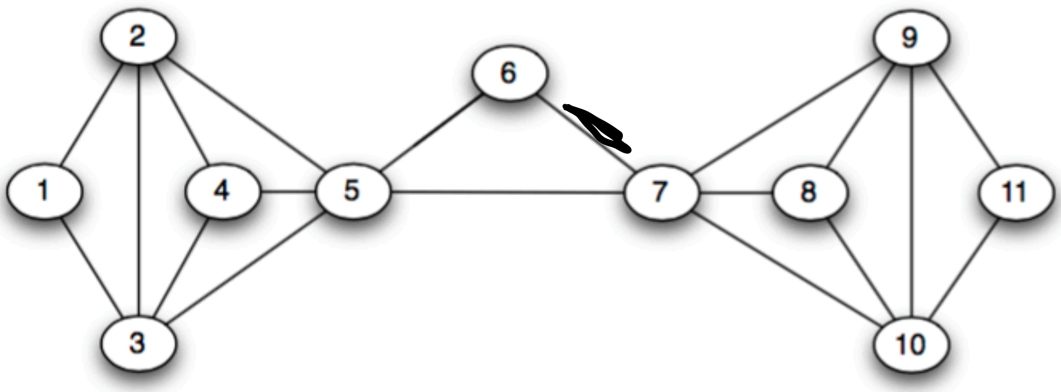


Figure 3.15: A network can display tightly-knit regions even when there are no bridges or local bridges along which to separate it.

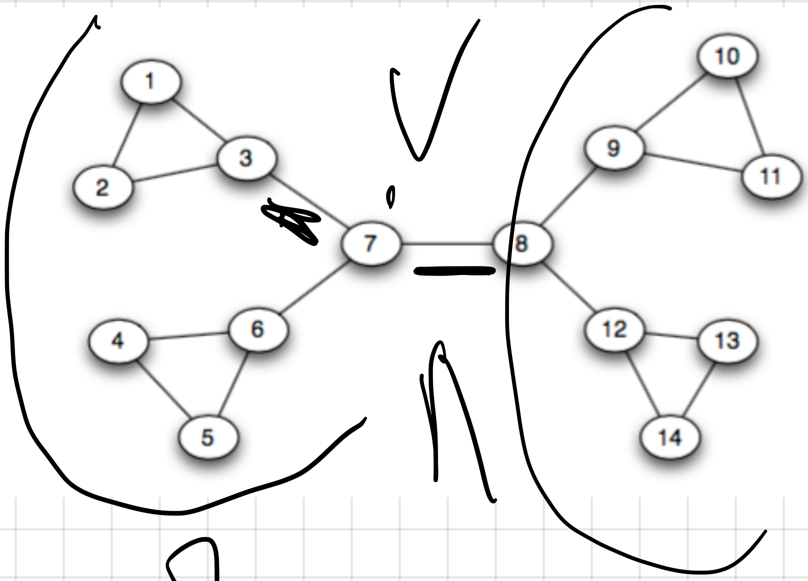
Betweenness of edge (x, y)

"number of shortest paths that cross through (x, y)"

$B(7,8) = 49$

$B(3,7) = 33$

. . . .



7 7

$$B(i) = \sum_{j<k} \frac{d_{jk}(i)}{d_{jk}}$$

normalized betweenness of a node

Some observations on Betweenness

local bridges

NO inversely correlated

" structural holes " ?

" traffic " or " flow "

the Girvan - Newman Method

1. a. find edge(s) with the highest betweenness
   b. remove those edges
   c. → components

2. Re-calculate betweenness for all the edges

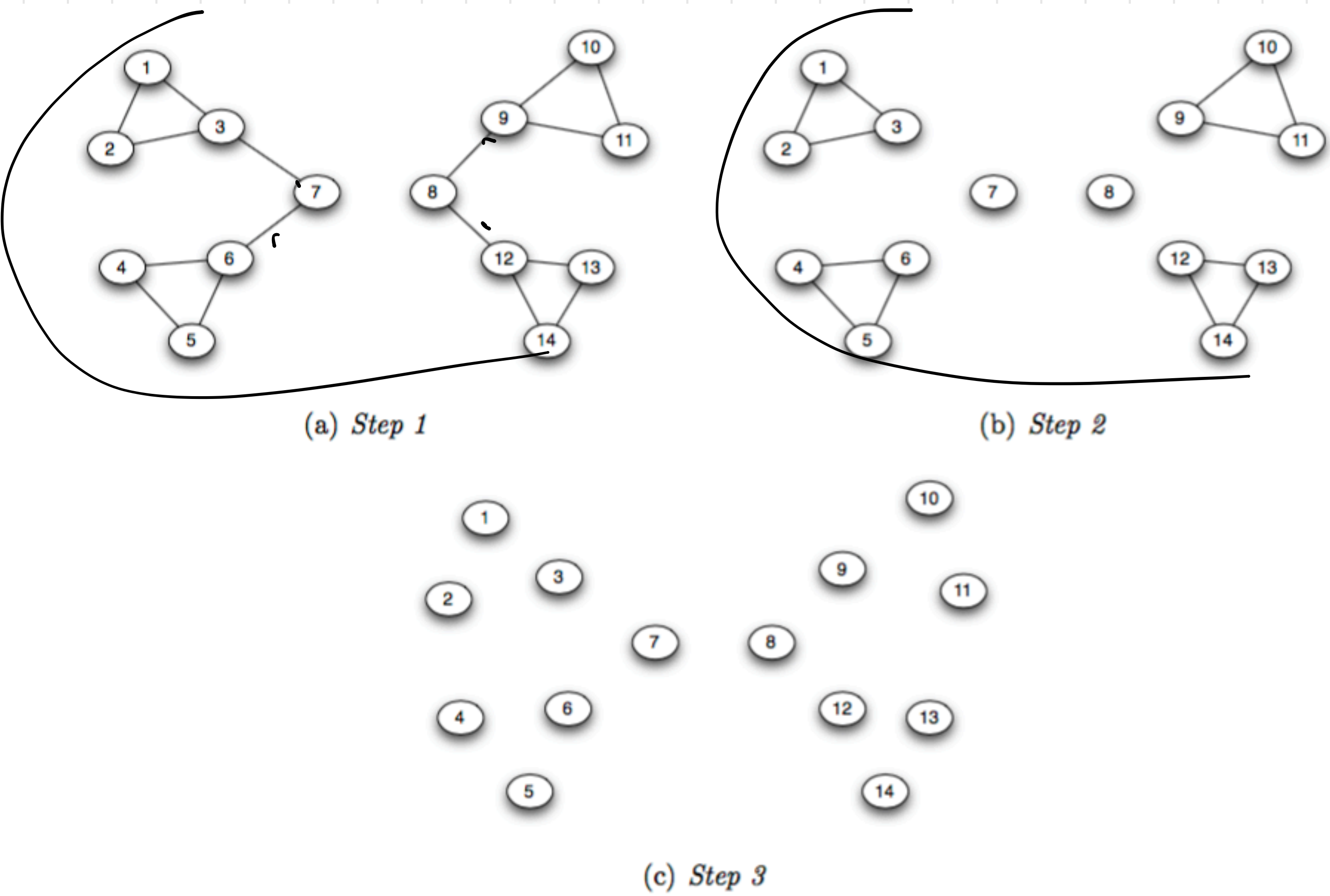


Figure 3.16: The steps of the Girvan-Newman method on the network from Figure 3.14(a).

Another G-N example

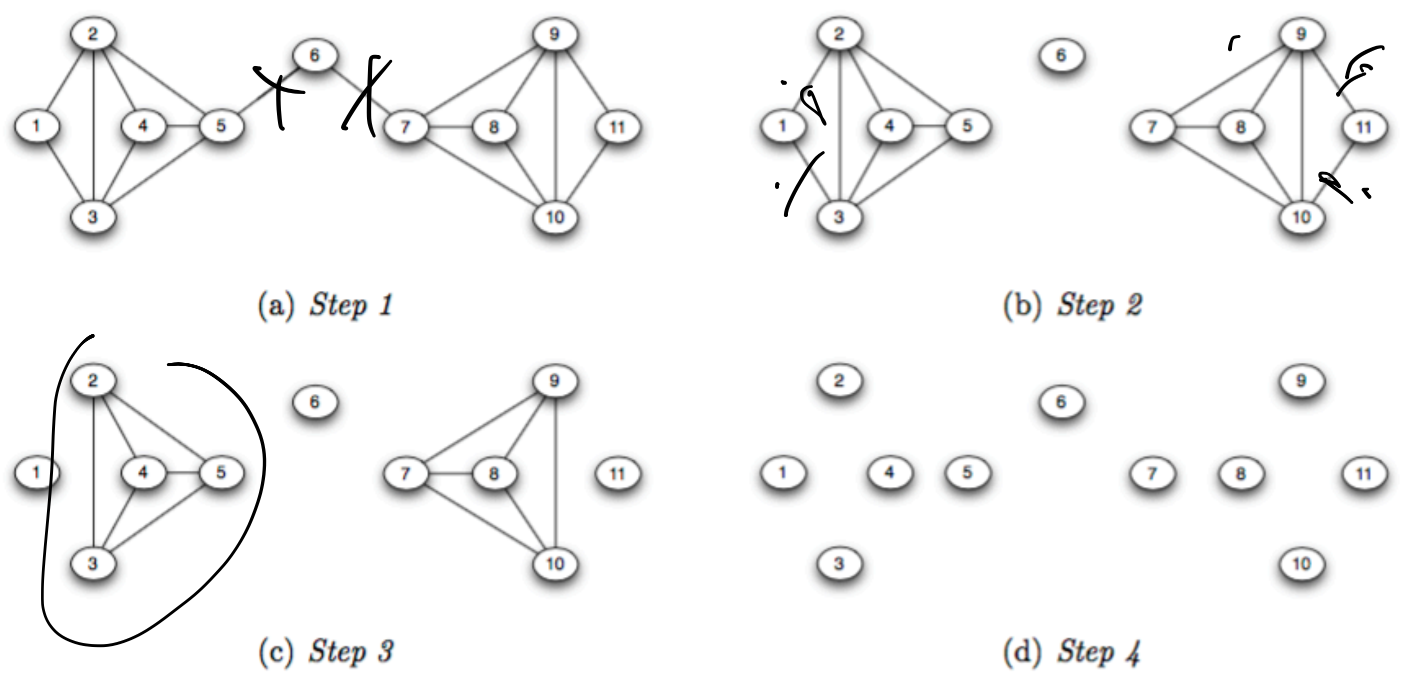
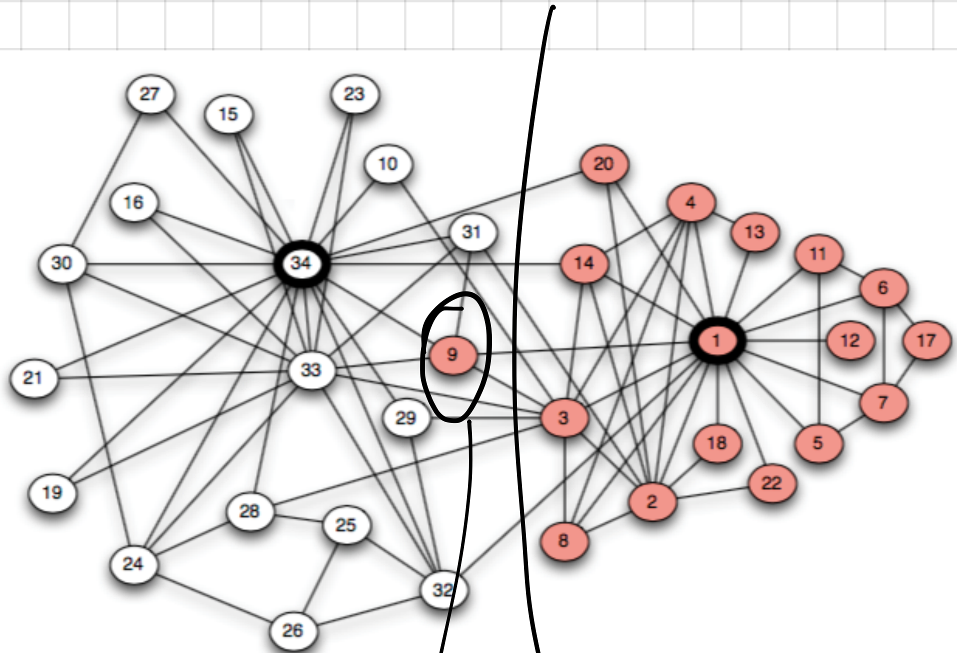


Figure 3.17: The steps of the Girvan-Newman method on the network from Figure 3.15.

We do not need to find local bridges anymore

⇒ betweenness is an excellent approximation that is calculated through a global property of the network

Applying G-N method to Karate Club



"minimum cut"

?

Computing Betweenness Values

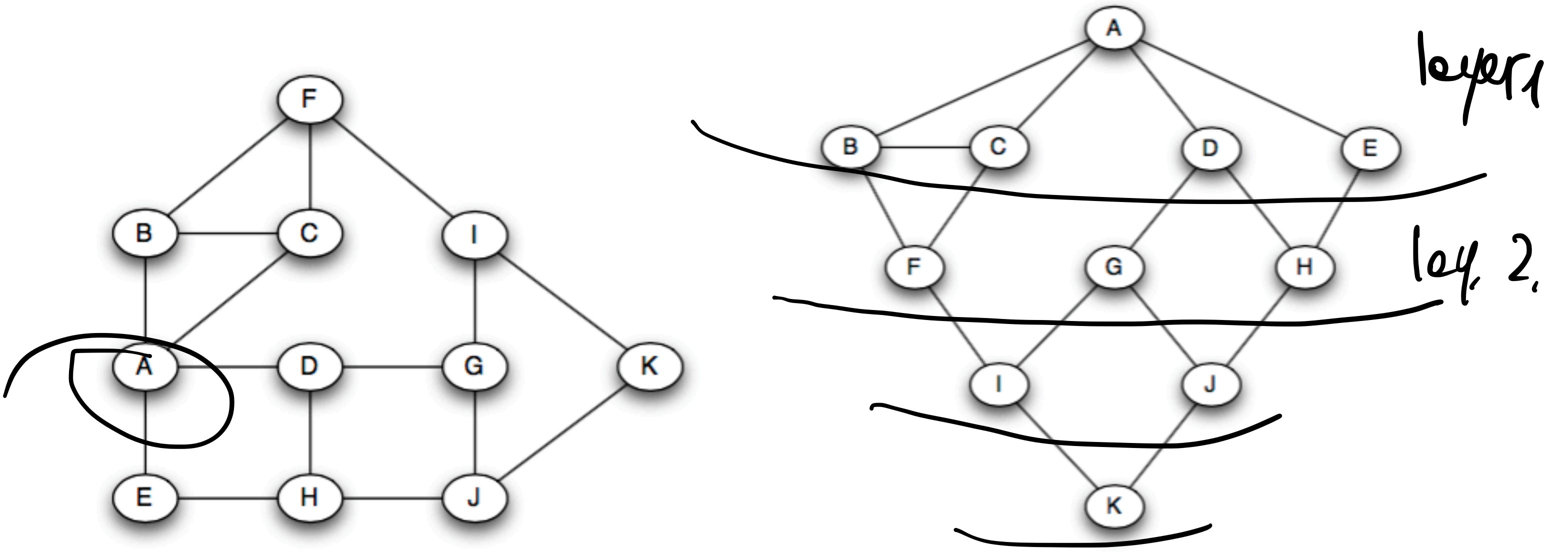
BFS

$\forall$ node A

1. perform BFS
2. determine the number of shortest paths from A
3. calculate the amount of "flow" from A

First step: BFS from node A

top down



(a) *A sample network*

(b) *Breadth-first search starting at node A*

Figure 3.18: The first step in the efficient method for computing betweenness values is to perform a breadth-first search of the network. Here the results of breadth-first from node A are shown; over the course of the method, breadth-first search is performed from each node in turn.

Second step: Counting Shortest Paths

bottom - up

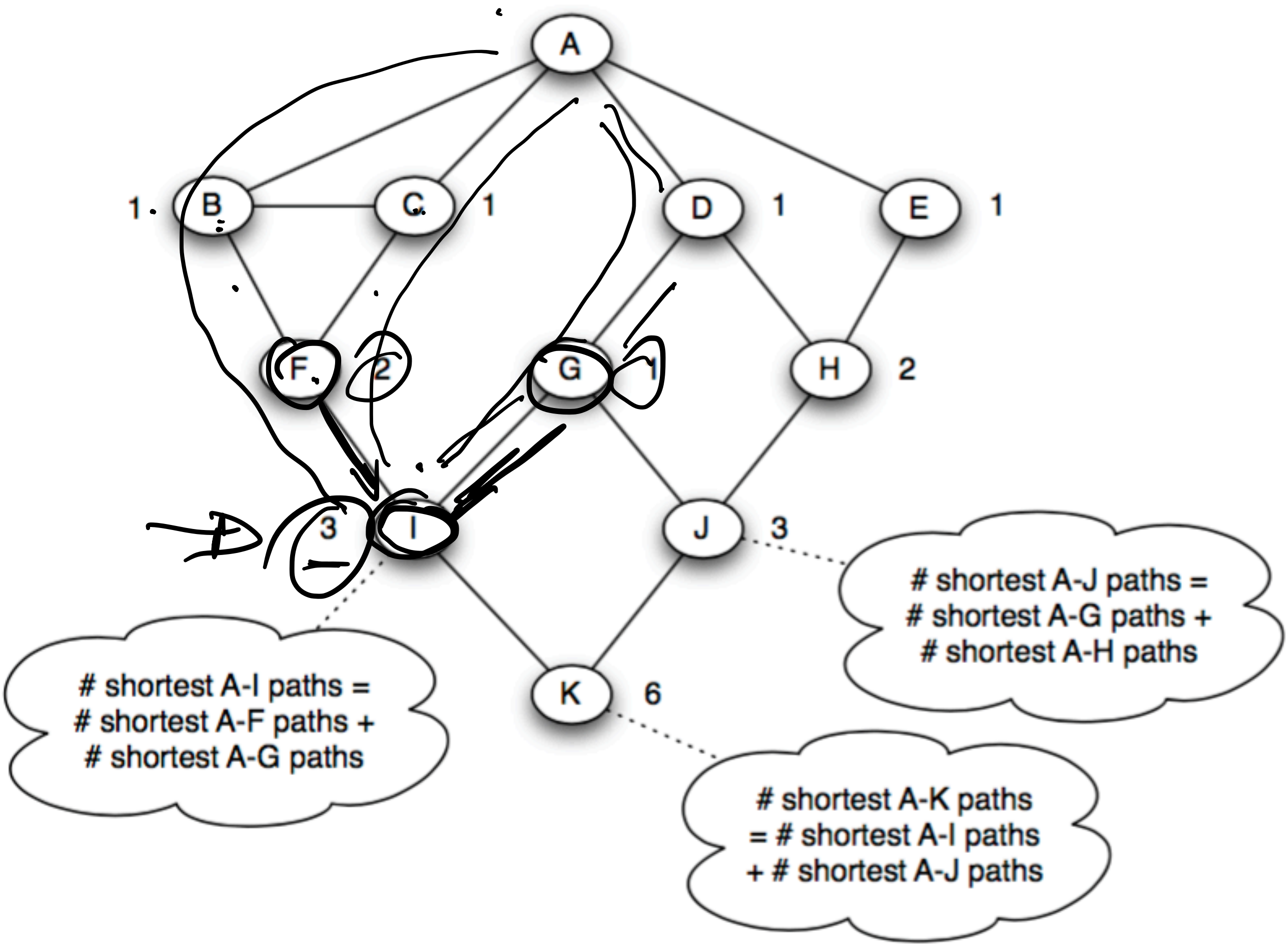


Figure 3.19: The second step in computing betweenness values is to count the number of shortest paths from a starting node A to all other nodes in the network. This can be done by adding up counts of shortest paths, moving downward through the breadth-first search structure.

Final Step: Computing the "flow"

bottom-up

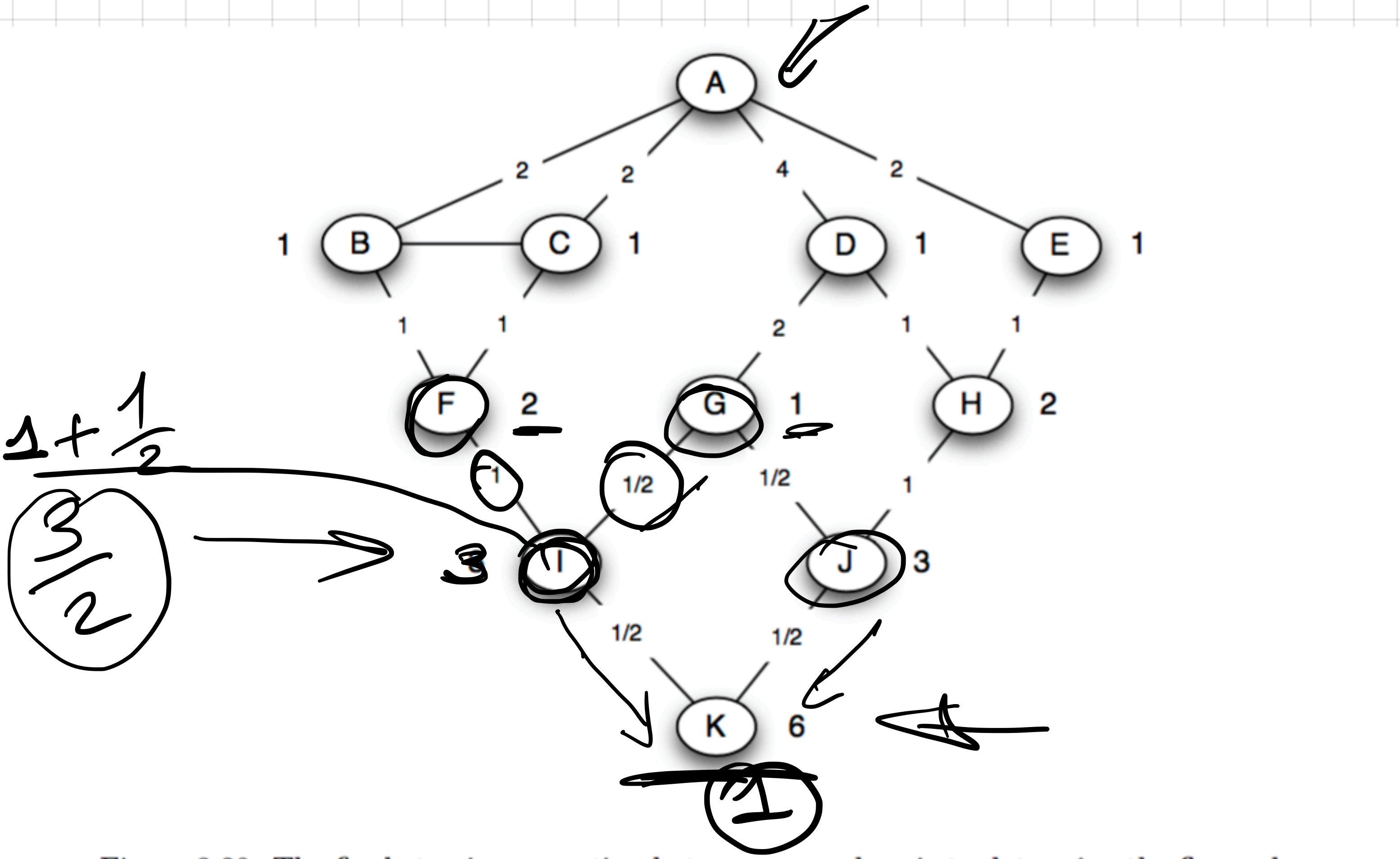


Figure 3.20: The final step in computing betweenness values is to determine the flow values from a starting node A to all other nodes in the network. This is done by working up from the lowest layers of the breadth-first search, dividing up the flow above a node in proportion to the number of shortest paths coming into it on each edge.

Output: Betweenness values for each edge

Repeat $\forall$ nodes in G

$$\frac{\text{Sum up all the "flow" values}}{2}$$

$$\Downarrow$$

<u>betweenness values</u>

Take Home Message

Girvan - Newman: $O(N^3)$

dense graph

$O(N^2)$

sparse graphs

"Leuven"
(modularity algorithm in Gephi)

$O(N)$

sparse graph

open research question
how to better define groups in large scale networks