Peer-to-Peer and Social Networks

Centrality measures

Centrality measures

Centrality is related to the potential importance of a node. Some nodes have greater “influence” over others compared to the rest, or are more easily accessible to other, or act as a go-between in most node-to-node communications. These are represented by various centrality measures. Some important ones are

- Degree centrality
- Closeness centrality
- Betweenness centrality

Degree centrality

The more neighbors a given node has, the greater is its influence.

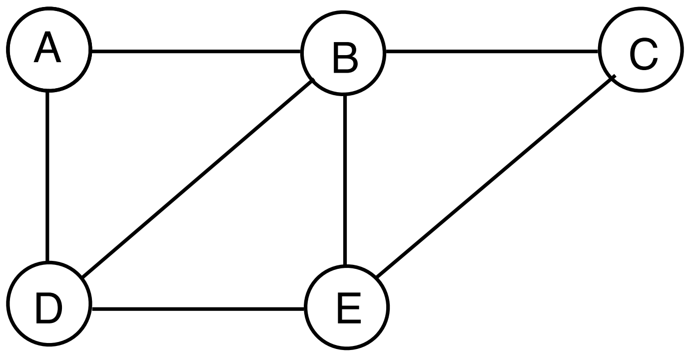
In human society, a person with a large number of friends is believed to be in a favorable position compared to persons with fewer.

This leads to the idea of **degree centrality**, which refers to the degree of a given node in the graph representing a social network.

Closeness centrality

Nodes that are able to reach other nodes via shorter paths, or who are "more reachable" by other nodes via shorter paths, are in more favored positions. This structural advantage can be translated into power, and it leads to the notion of **closeness centrality**.

Closeness centrality



(a)

	A	B	C	D	E		normalized
A	0	1	2	1	2	6	4/6
B	1	0	1	1	1	**4**	1
C	2	1	0	2	1	6	4/6
D	1	1	2	0	1	5	4/5
E	2	1	1	1	0	5	4/5

(b)

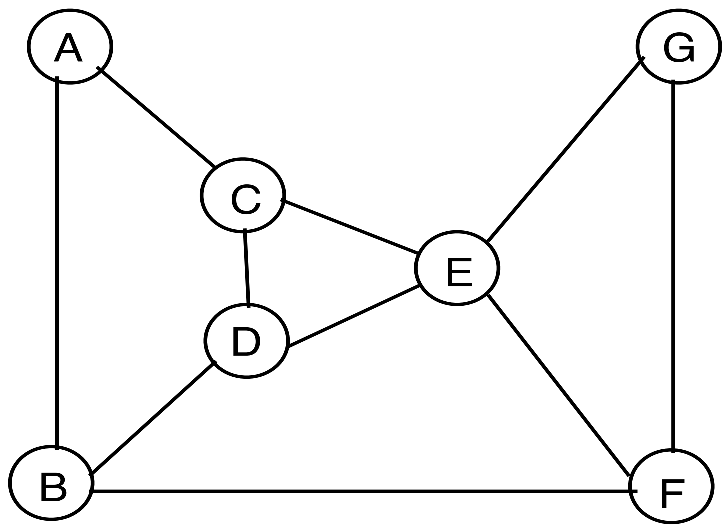
Let V denote the set of nodes, and $d(i,j)$ represent the distance between i and j .

A yardstick of the closeness of a node i from the other nodes is $\sum_{j \in V} d(i,j)$

This number is normalized in various ways: one is $\frac{1}{(n-1)} \sum_{j \in V} d(i,j)$. However, a problem occurs when the graph is partitioned, so an alternative is to consider the inverse of it

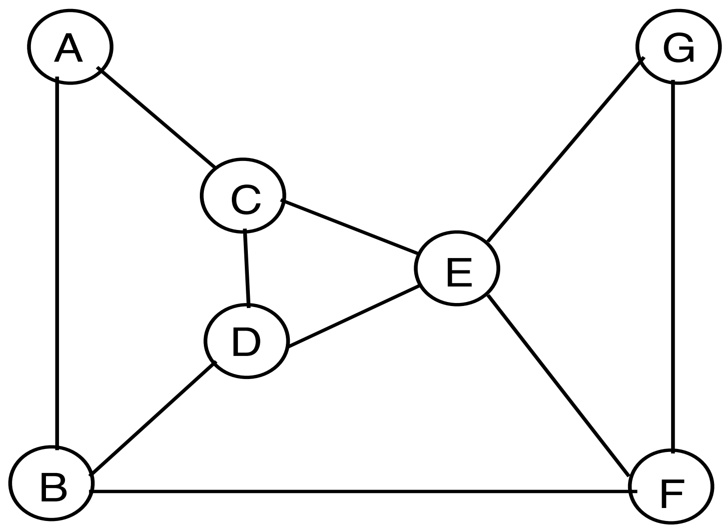
i.e. $$\frac{n-1}{\sum_{j \in V} d(i,j)}$$ as shown above.

Betweenness centrality



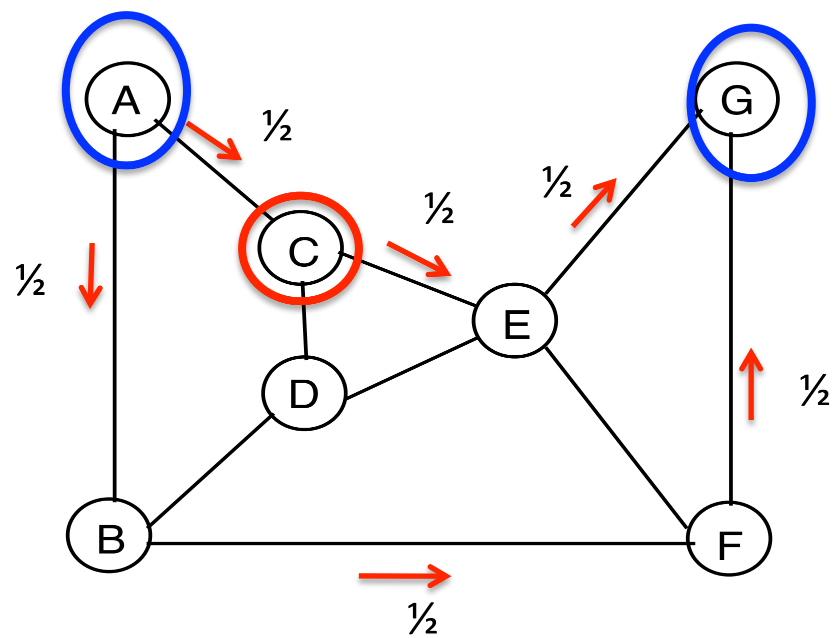
For each pair of nodes in a social network, consider one of the shortest paths –- all nodes in this path are intermediaries. The node that belongs to the shortest paths between the **maximum number** of such communications, is a special node – it is a potential deal maker and is in a special position since most other nodes have to channel their communications through it. Such a node has a high **betweenness centrality.**

Computing betweenness centrality



From each source node **u** to each destination node **v**, push 1 unit of flow via shortest paths. If there are multiple shortest paths, then the flow will be evenly split. The amount of flow handled by a node or an edge is a measure of its betweenness centrality.

Computing betweenness centrality

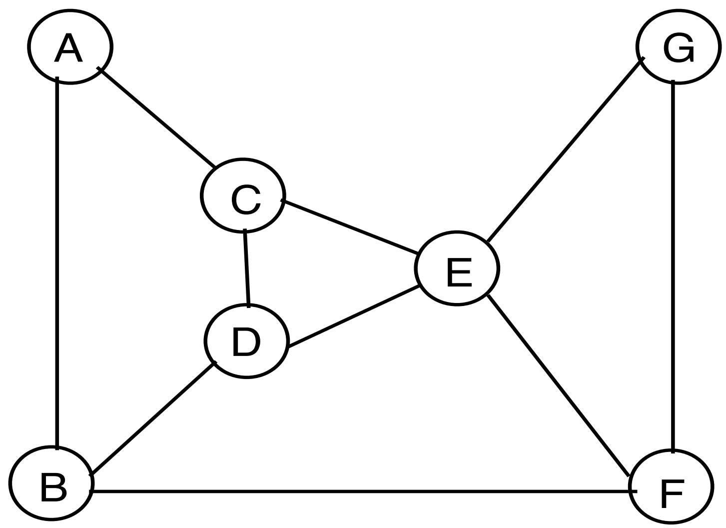


Betweenness centrality of **C** with respect to (A, G)

If each node sends 1 unit of flow to every other node (excluding C),

then C will handle only 2. (1 + ½ + ½) = 4 units of flow. Why?

Computing betweenness centrality

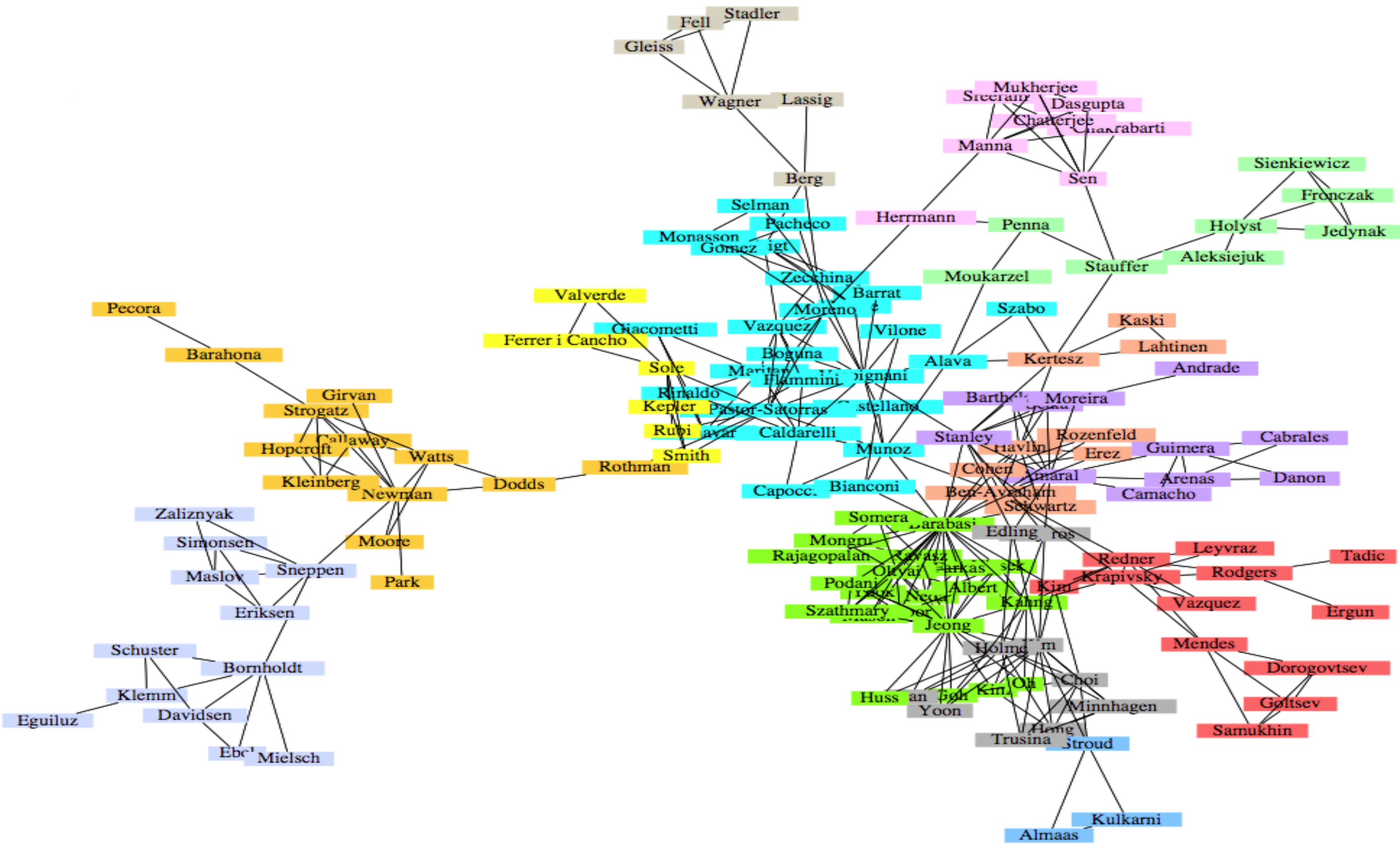


What is the betweenness centrality of node E?

Random walk centrality measures

It a different measure that takes into account the distance from the source to the distance node via a **random walk. Examples?** Follow the class lecture.

Community detection

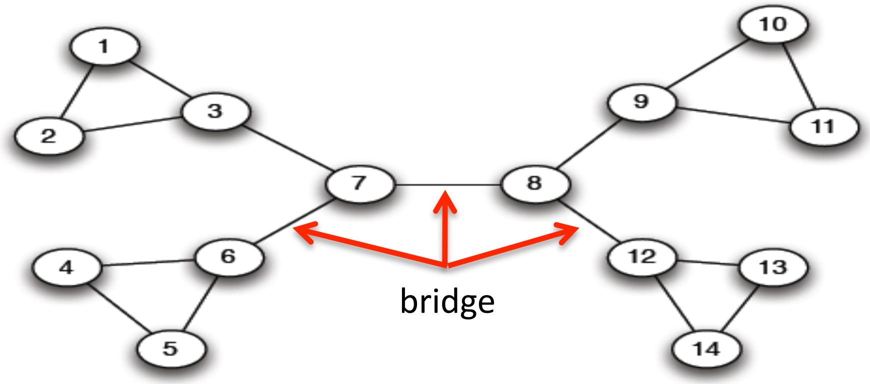


A co-authorship network of physicists and mathematicians
(Courtesy: Easley & Kleinberg)

What is a community?

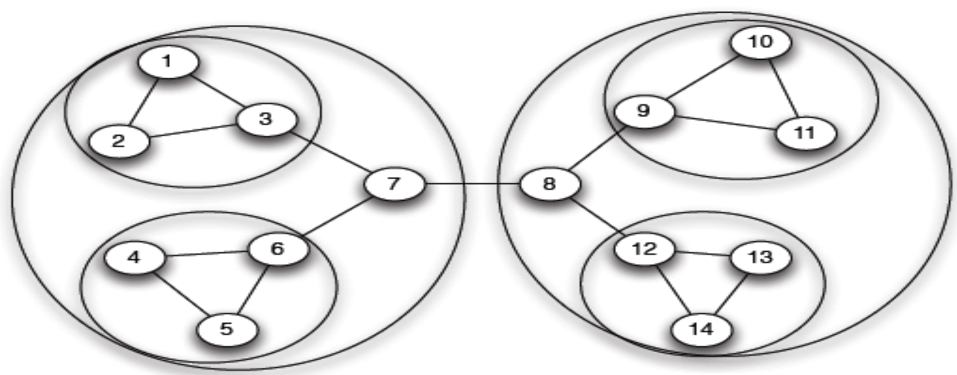
Informally a community is a “tightly-knit region” of the network. How to identify such a region, and how to separate them from a different tightly-knit region? It depends on how we quantify the notion of tightly-knit. Visual check of the graph is not enough, and can sometimes be misleading. Also, such regions can be nested.

What is a community?



(a) *A sample network*

Removal of a bridge separates the graph into disjoint components

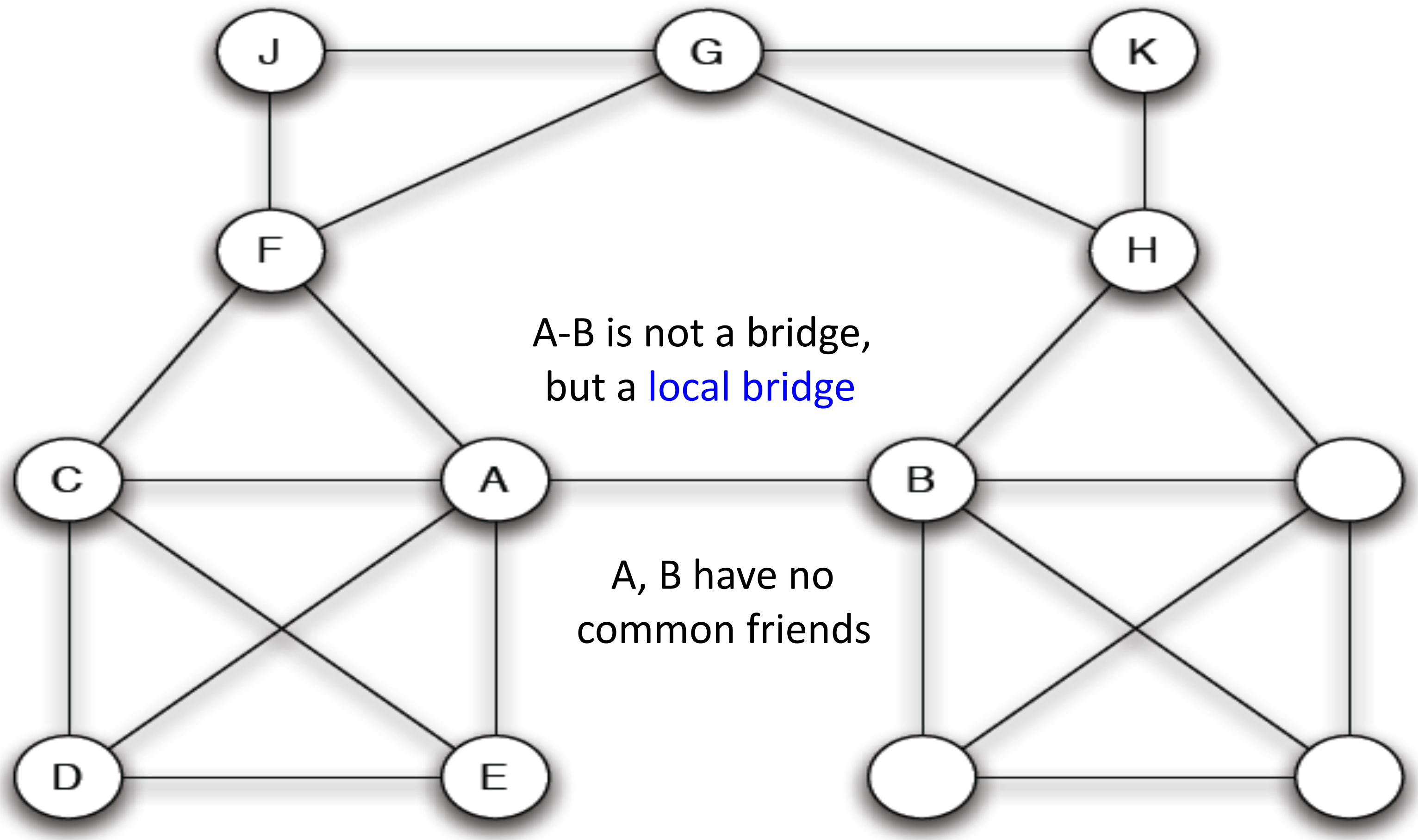


(b) *Tightly-knit regions and their nested structure*

An example of a nested structure of the communities

(Courtesy: Easley & Kleinberg)

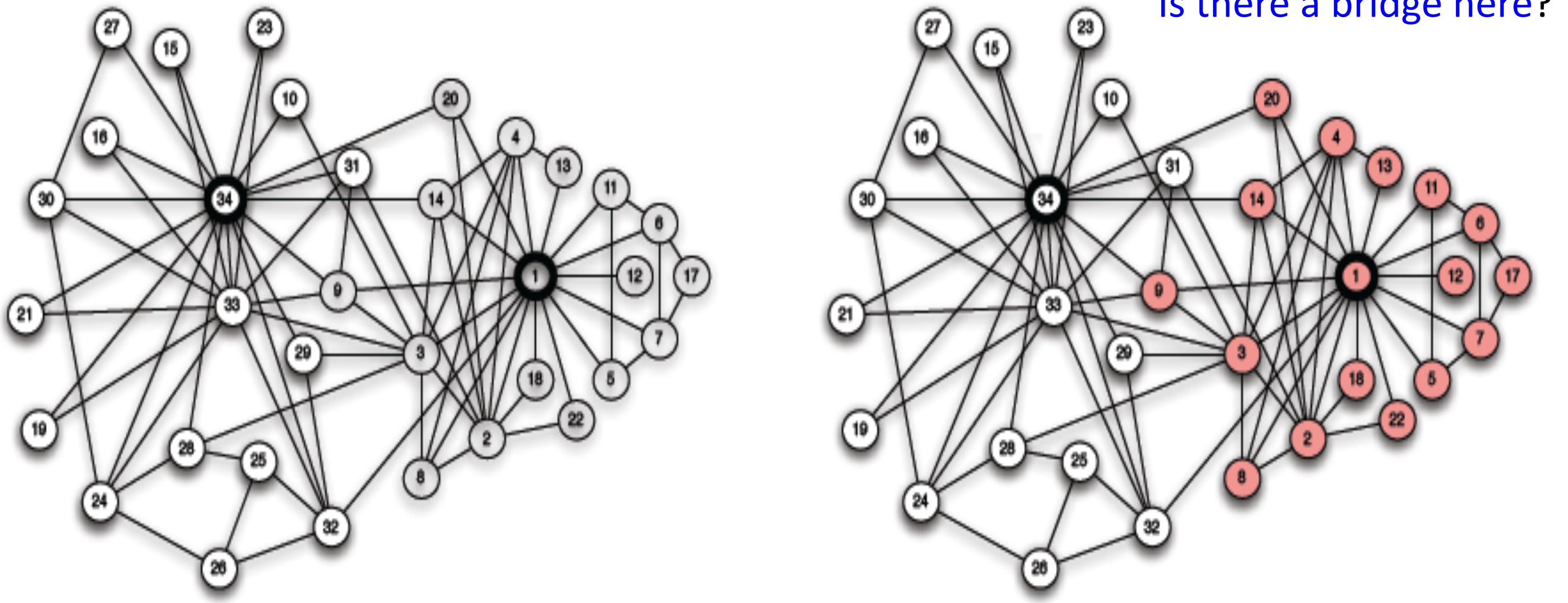
What is a community?



(Courtesy: Easley & Kleinberg)

What is a community?

Is there a bridge here?



(a) *Karate club network* (b) *After a split into two clubs*

A dispute caused the club to split into two clubs. How can you identify them?

(Courtesy: Easley & Kleinberg)

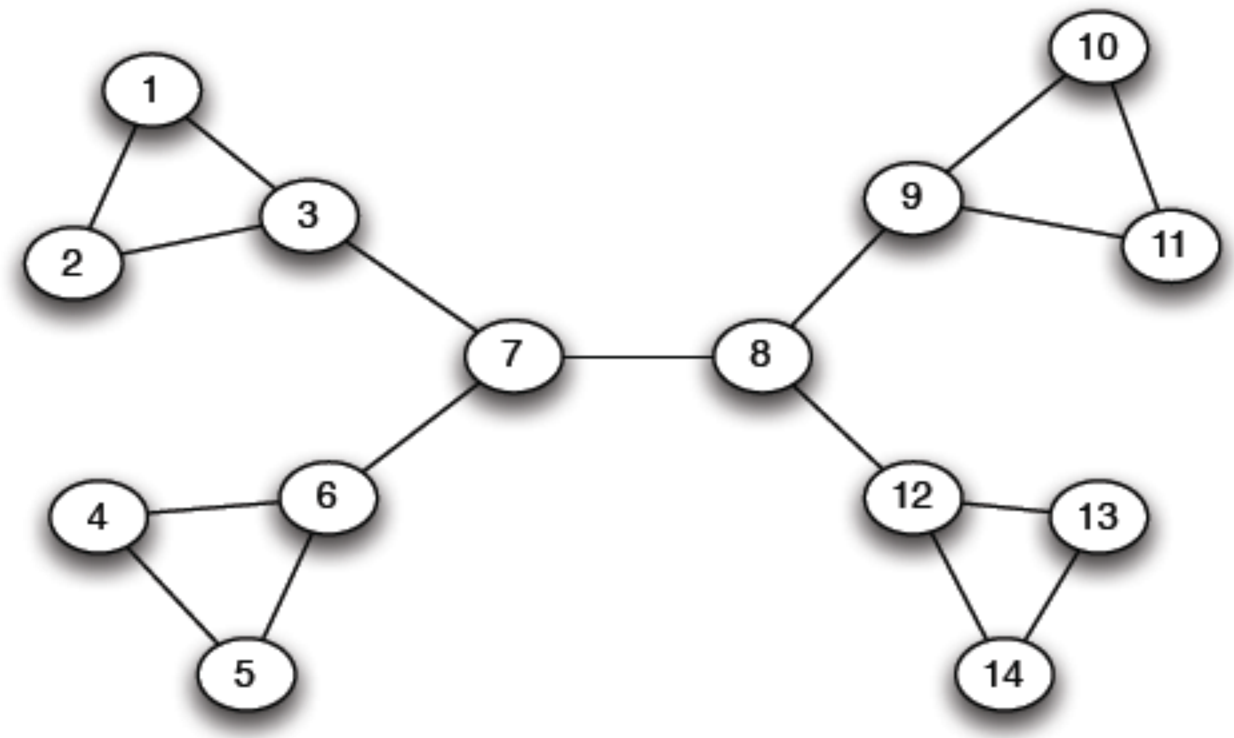
Community detection

Girvan-Newman Method

- Remove the edges of **highest betweenness** first.
- Repeat the same step with the remainder graph.
- Continue this until the graph breaks down into individual nodes.

As the graph breaks down into pieces, the tightly knit community structure is exposed.

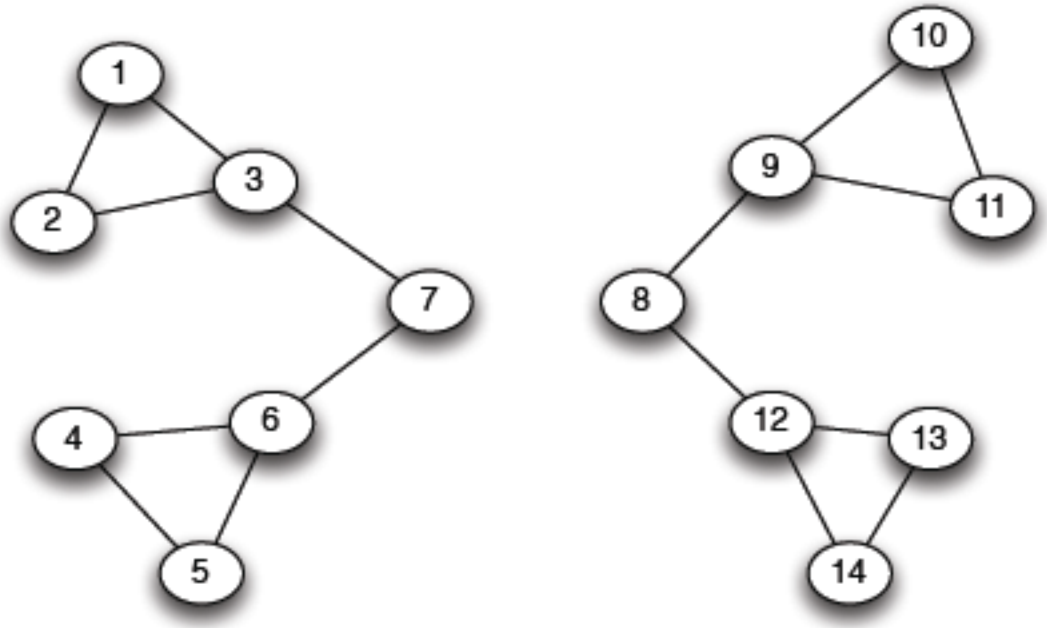
Girvan Newman method: An example



Betweenness(7-8)= 7x7 = 49 Betweenness(1-3) = 1X12=12

Betweenness(3-7)=betweenness(6-7)=betweenness(8-9) = betweenness(8-12)= 3X11=33

Girvan Newman method: An example

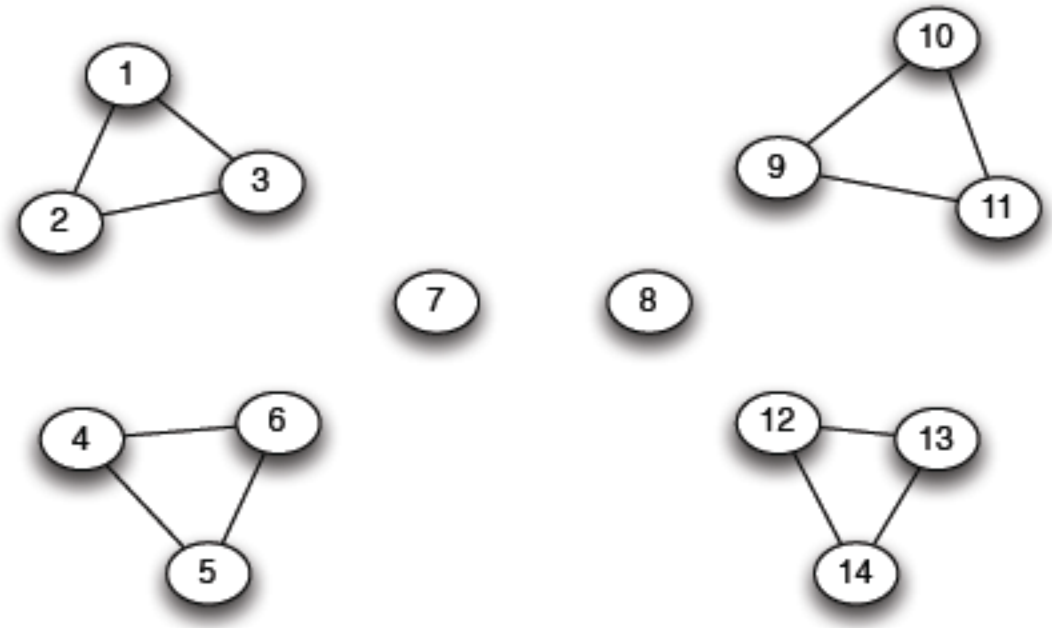


(a) *Step 1*

Betweenness(1-3) = 1X5=5

Betweenness(3-7)=betweenness(6-7)=betweenness(8-9) = betweenness(8-12)= 3X4=12

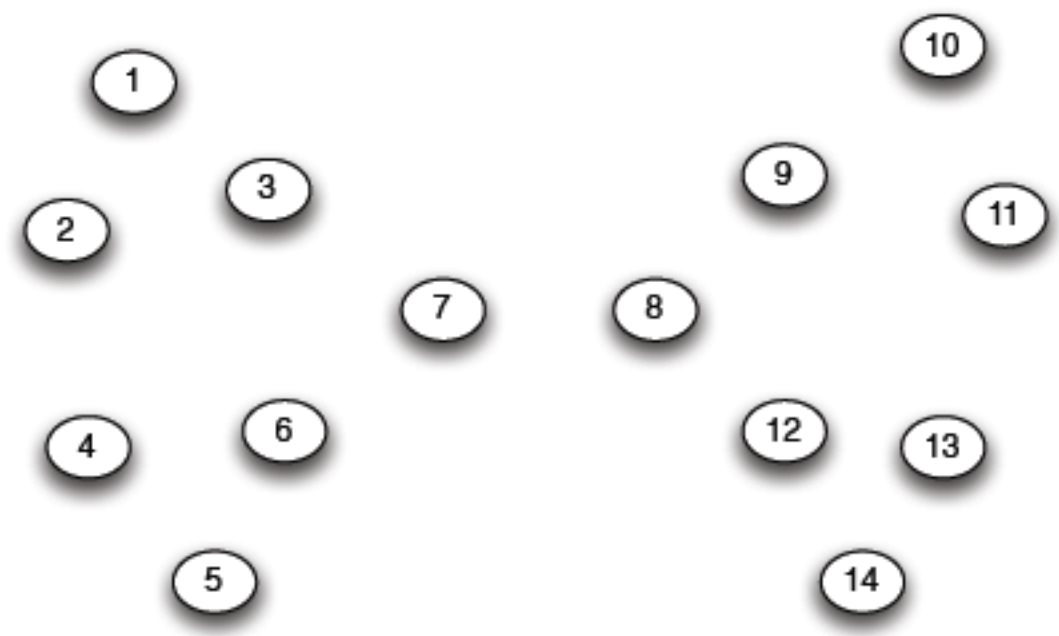
Girvan Newman method: An example



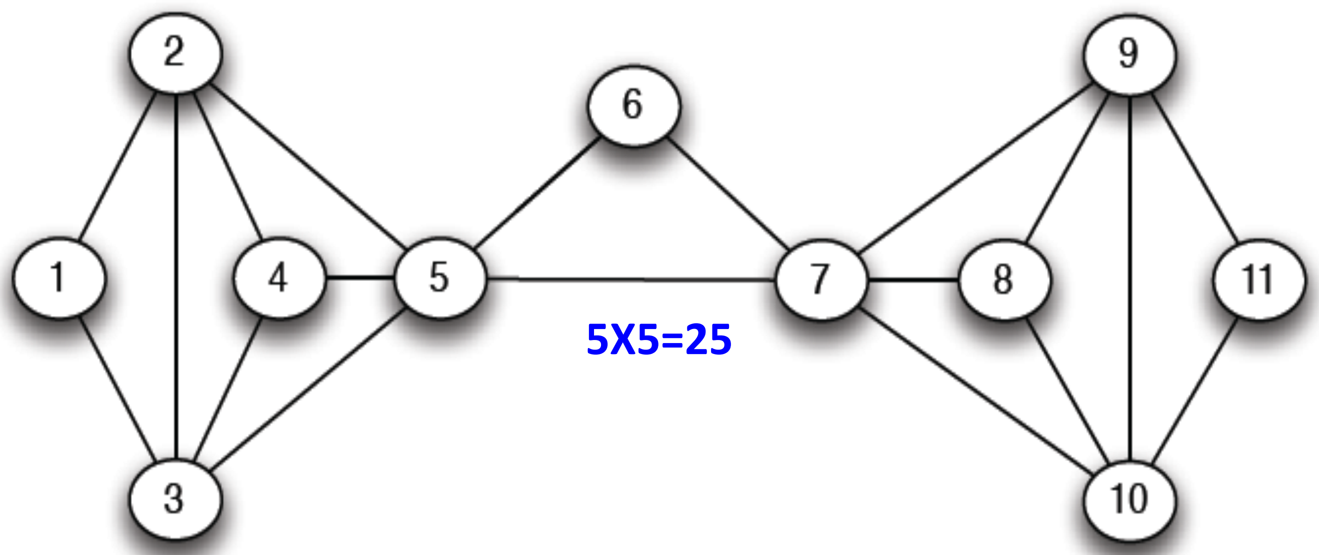
(b) *Step 2*

Betweenness of every edge = 1

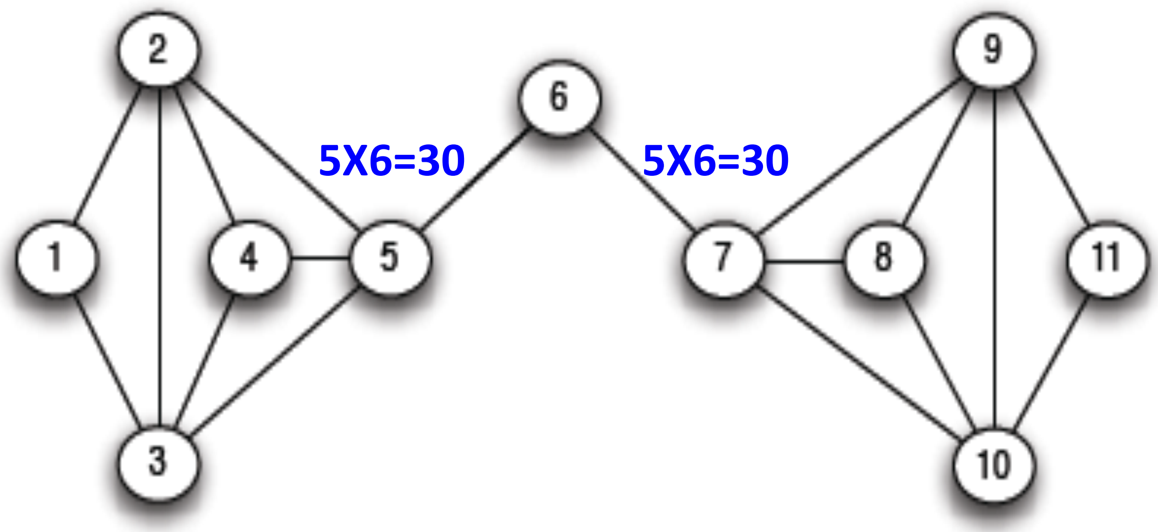
Girvan Newman method: An example



Another example

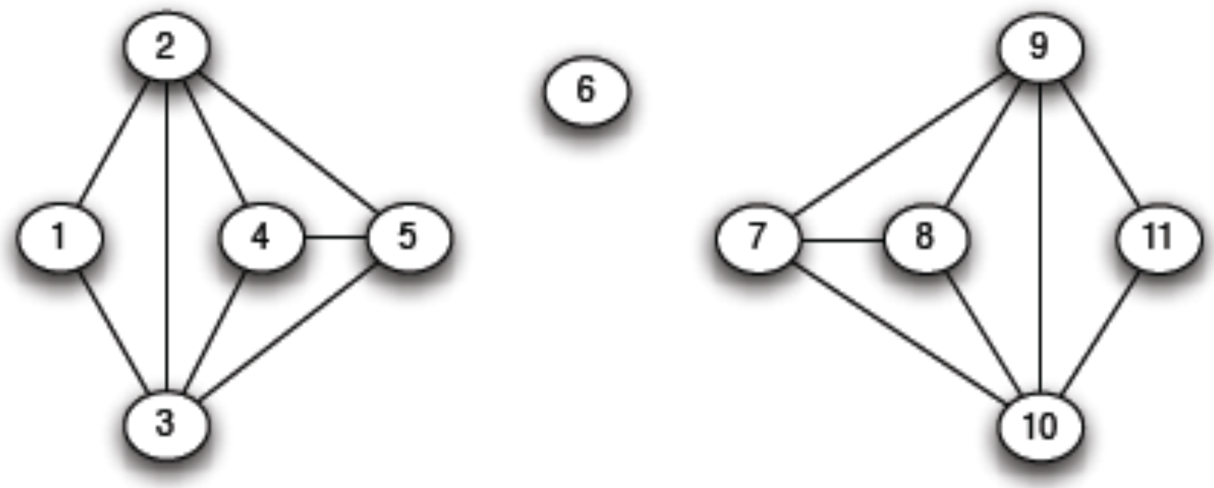


Another example



(a) *Step 1*

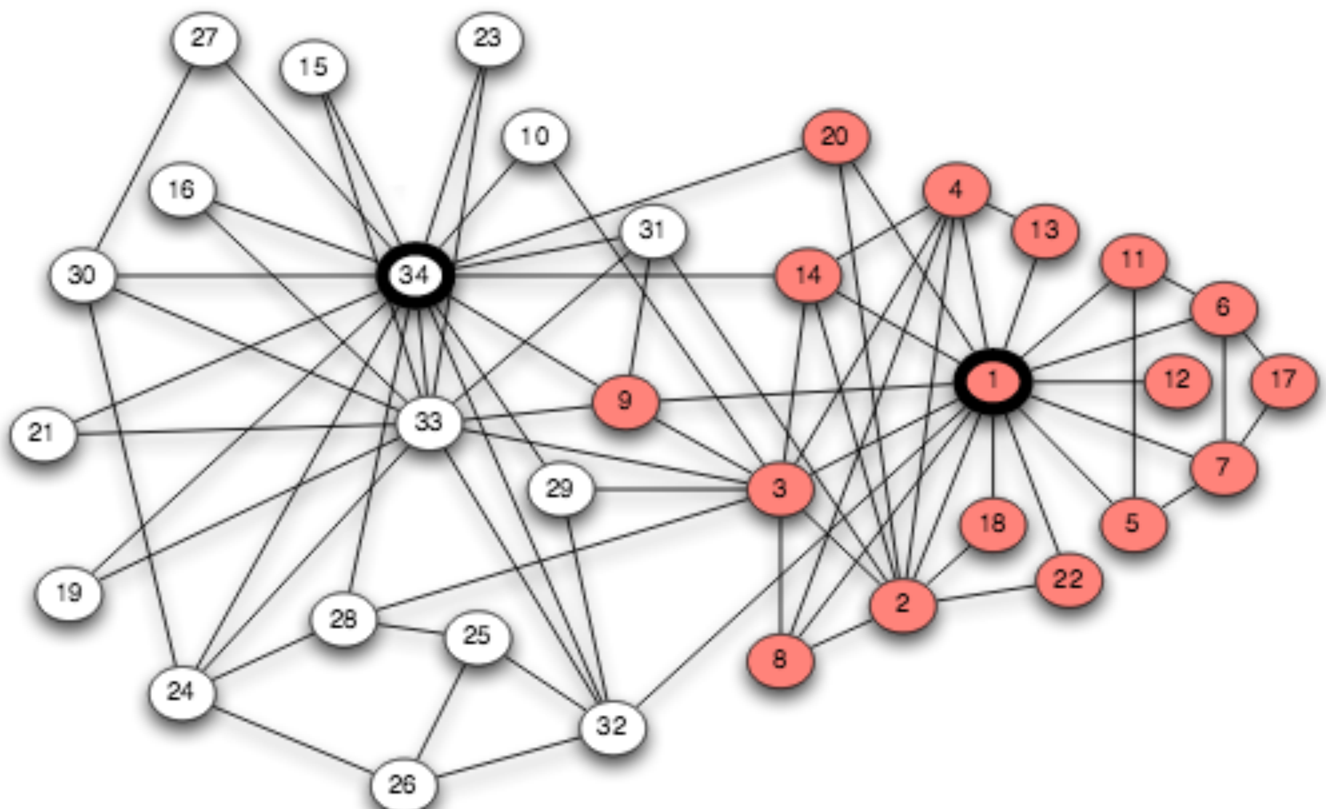
Another example



(b) *Step 2*

Revisiting Karate Club Network

A set of minimum total strength whose removal will disconnect the rival leaders. Zachary followed this to identify the split.



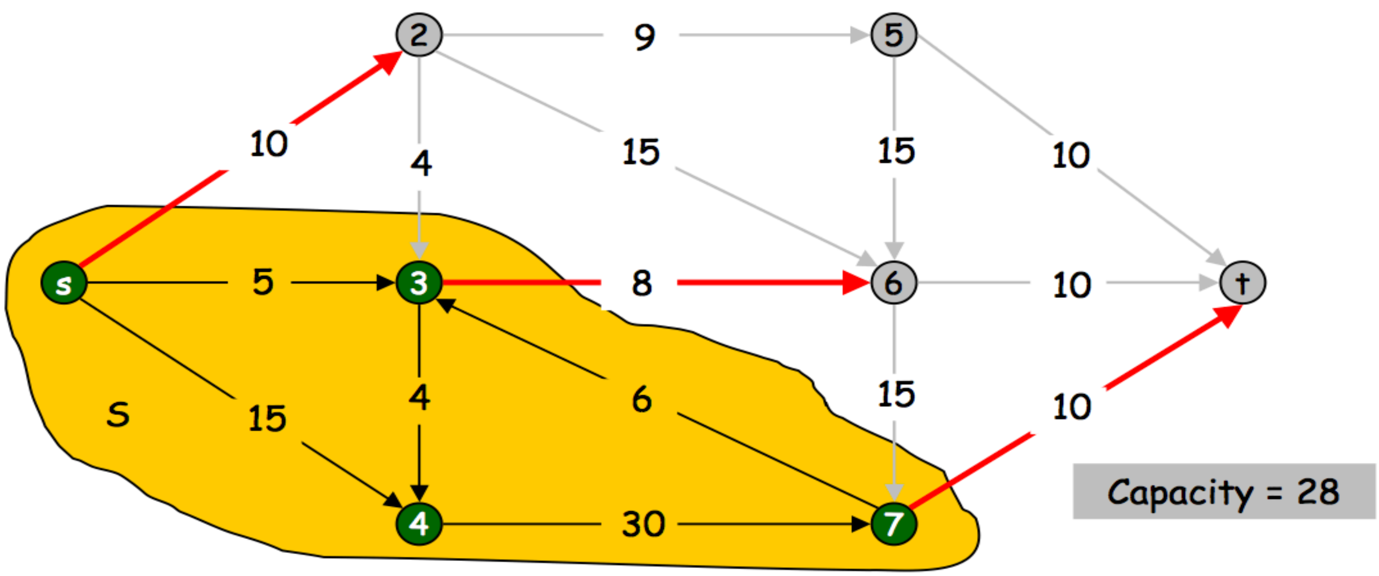
Min Cut Problem

Minimum Cut Problem

A cut is a node partition (S, T) such that s is in S and t is in T.

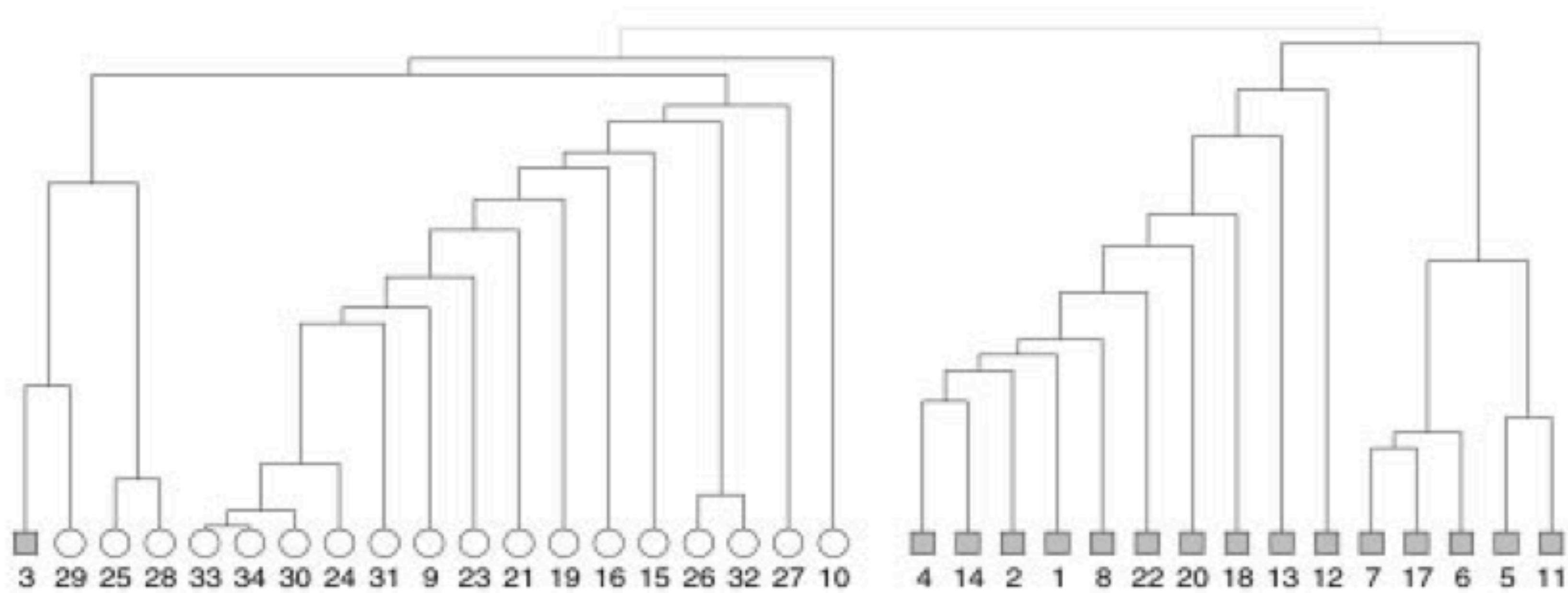
- capacity(S, T) = sum of weights of edges leaving S.

Min cut problem. Find an s-t cut of minimum capacity.



Revisiting Karate Club Network

Zachary initially assigned weight to each edge and then partitioned the graph by computing the min-cut. Successive application of this produces a dendogram



A Dendogram

Revisiting Karate Club Network

Girvan-Newman's community detection algorithm produced almost the exact solution, except for one node (node 9)

However, it is not computationally efficient.