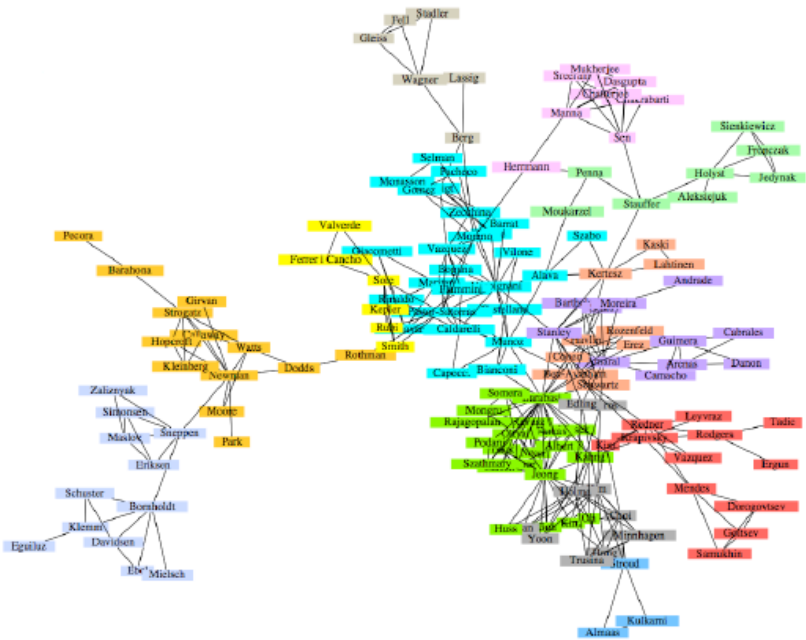
Strong and Weak Ties

Ana Paula Couto
Computer Science Department
Universidade Federal de Minas Gerais

Based on http://www.cs.uoi.gr/~tsap/teaching/2013-cs-l14/references.html

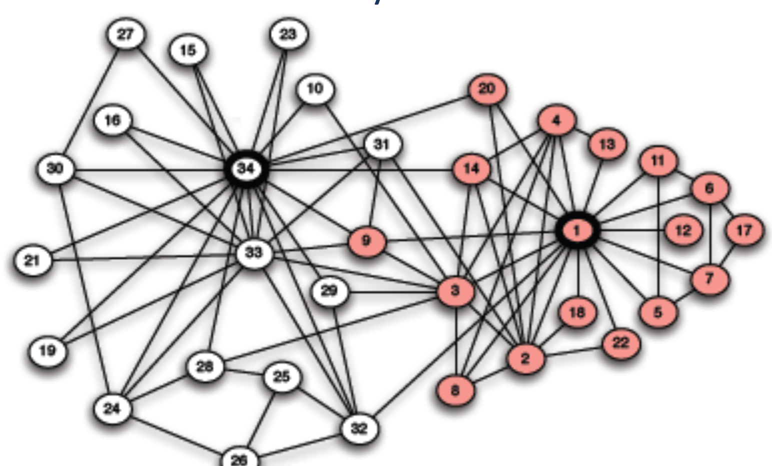
Betweenness and Graph Partitioning

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



Co-authorship network of physicists and applied mathematicians

Still heavily interconnected!

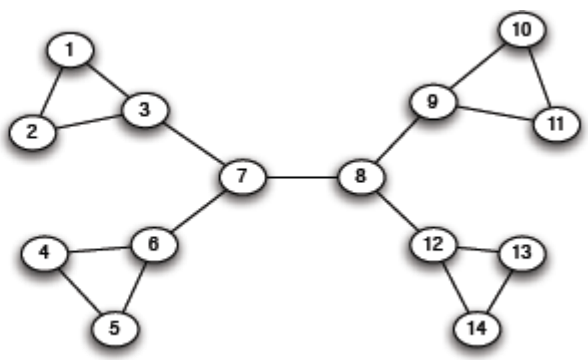


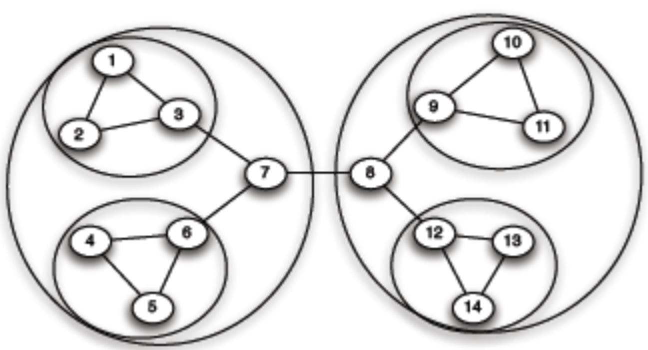
Karate club

Dispute between the president (34) and the Instructor (1) led the club to split into two rival clubs.
Community-finding algorithms are often tested based on their ability to infer these two communities.

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)

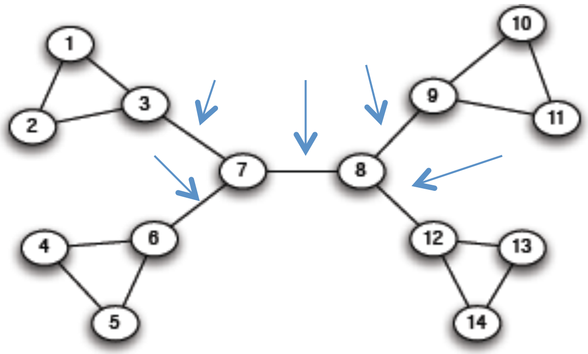




Tightly-knit regions and nested structure

Girvan and Newman

- Divisive method

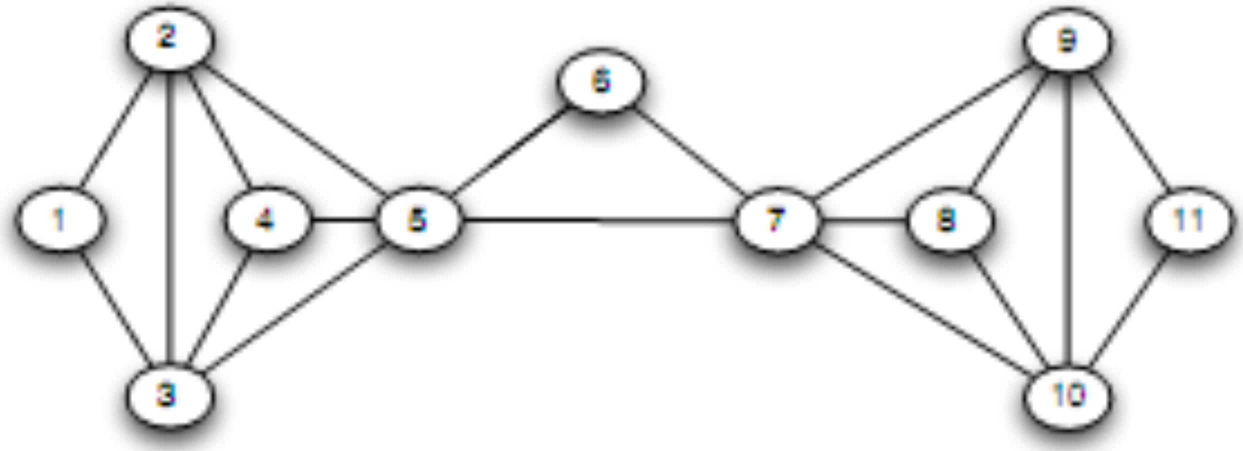


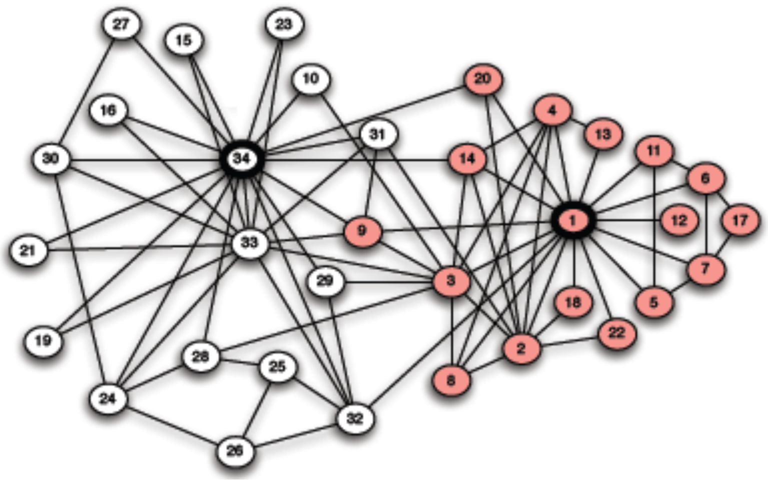
Finding bridges and local bridges?

Which one to choose?

Girvan and Newman

There is no local bridge

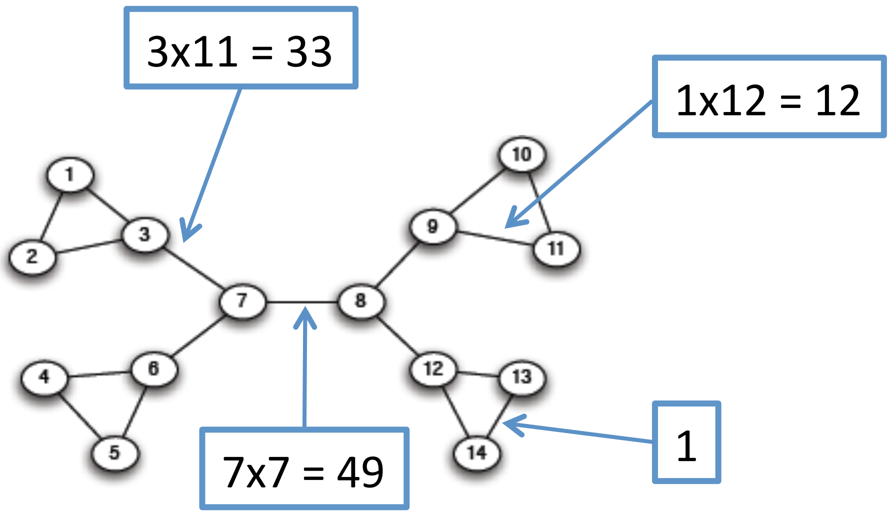




Edge Betweenness

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 such shortest paths edge (a, b) is credited with the fraction of those shortest paths that include (a, b).

$$\mathrm{bt}(a,b) = \sum_{x,y} \frac{\#shortest\_paths(x,y)through(a,b)}{\#shortest\_paths(x,y)}$$



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

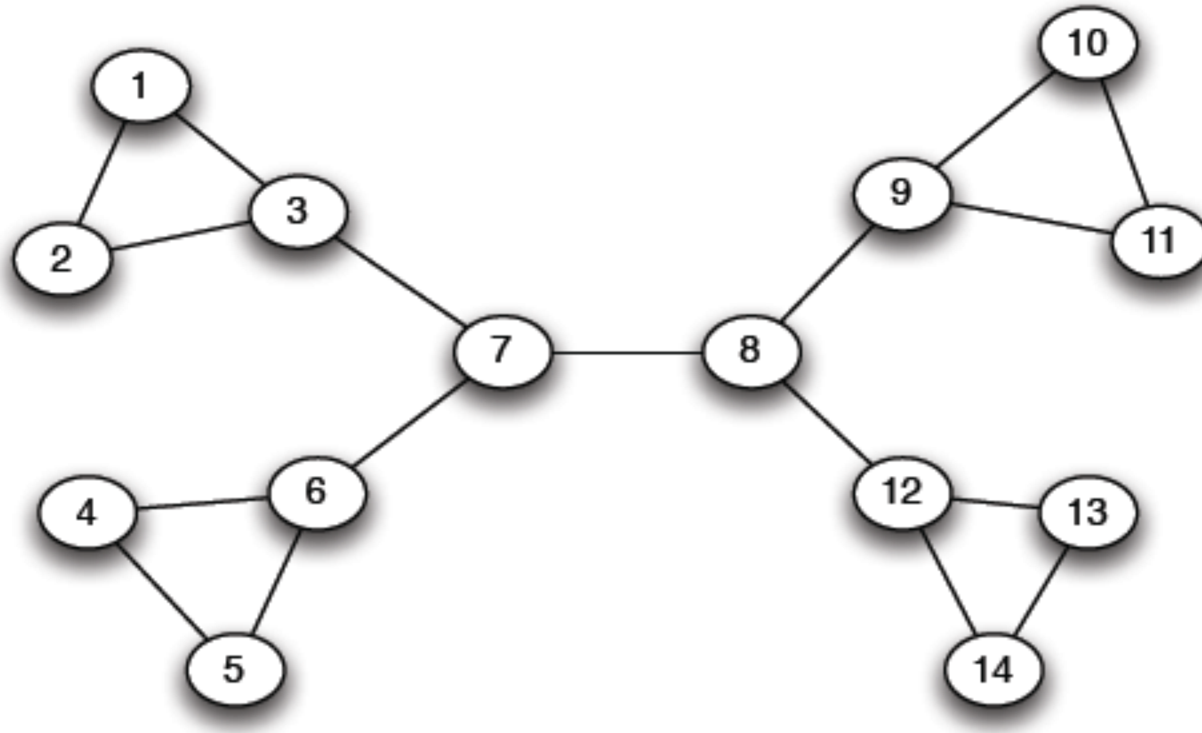
Traffic (unit of flow)

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.

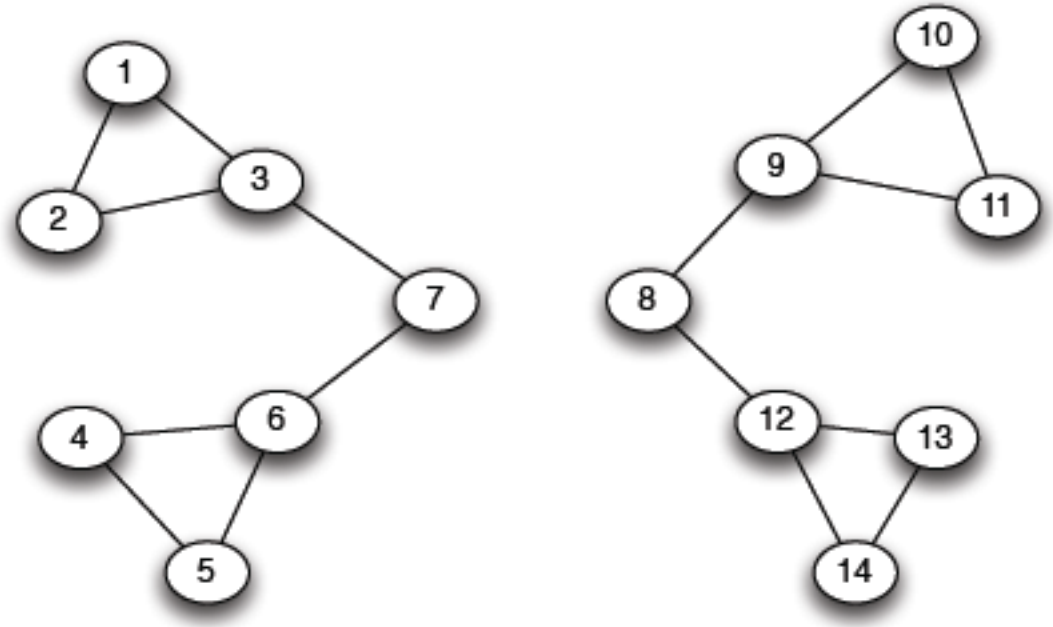
Girvan Newman method: An example



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

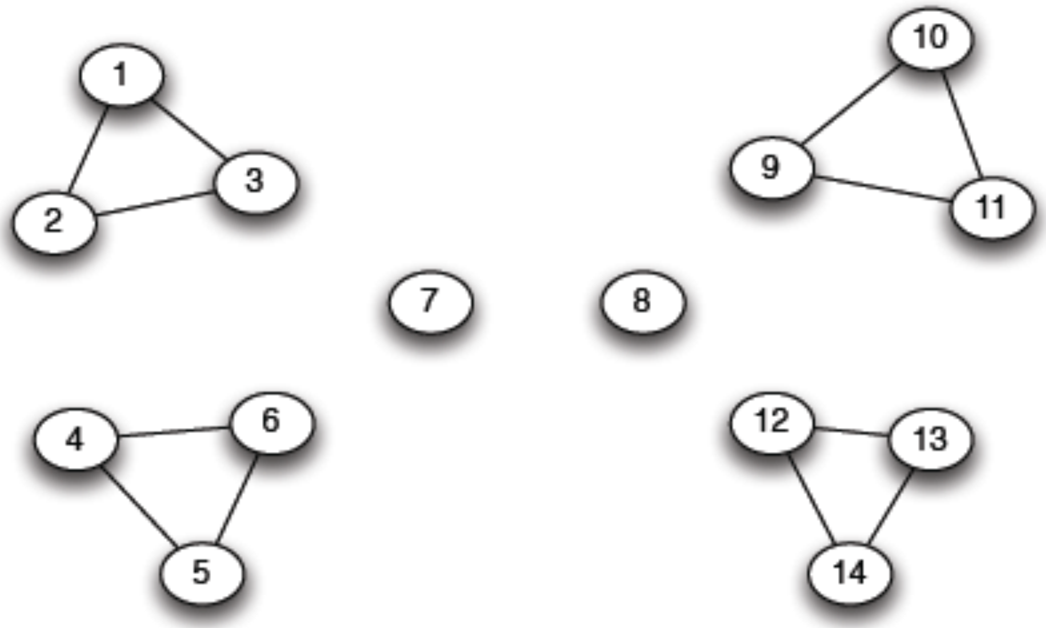


(a) *Step 1*

Betweenness(1, 3) = 1X5=5

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

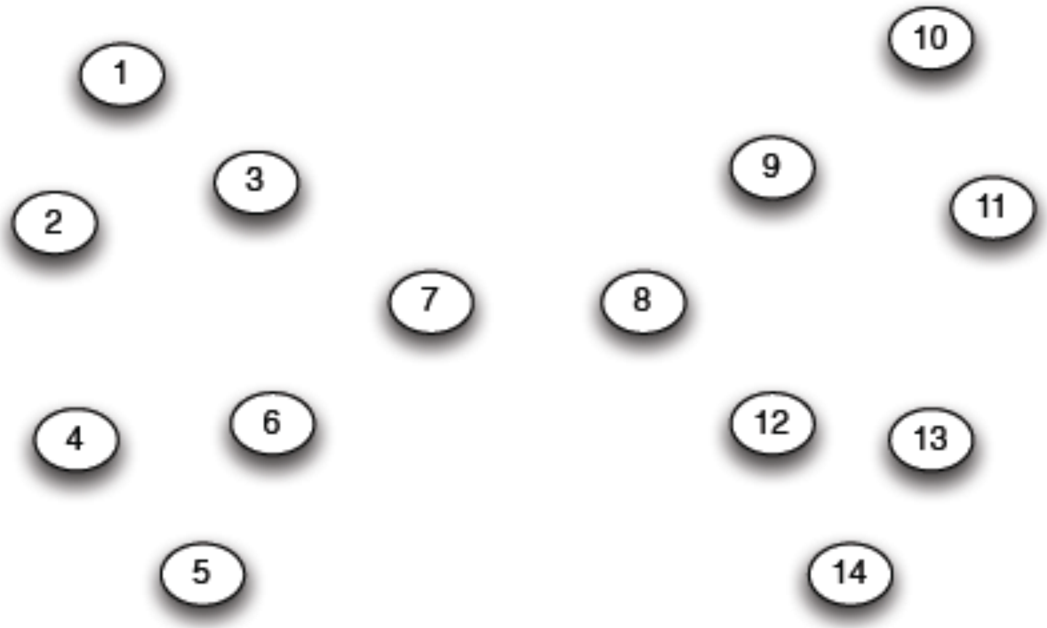
Girvan Newman method: An example



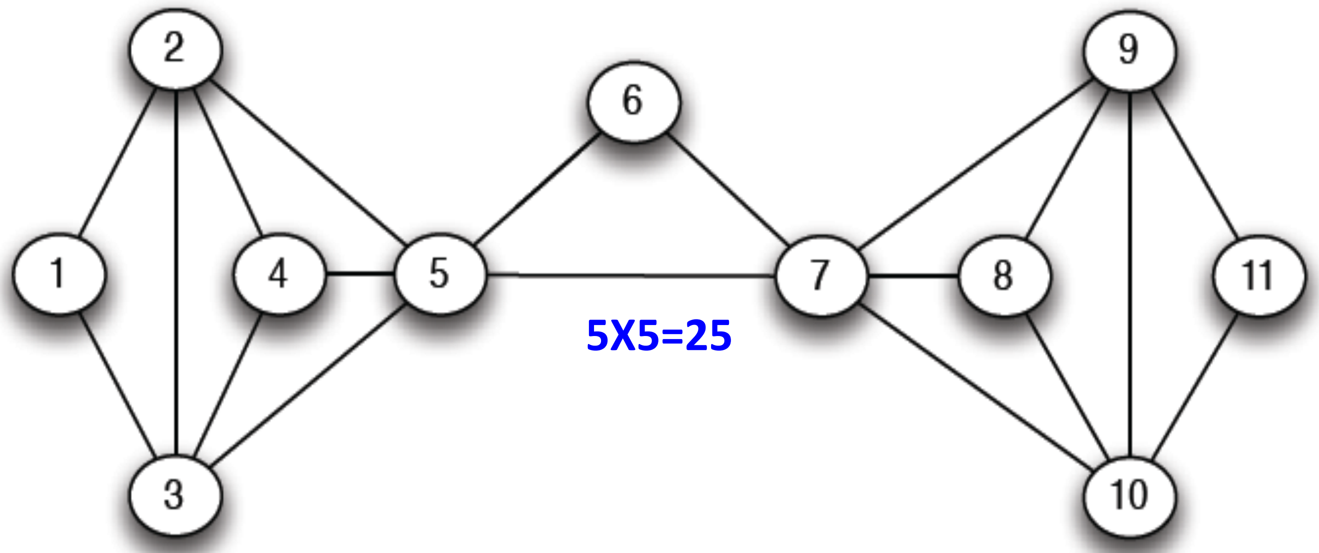
(b) *Step 2*

Betweenness of every edge = 1

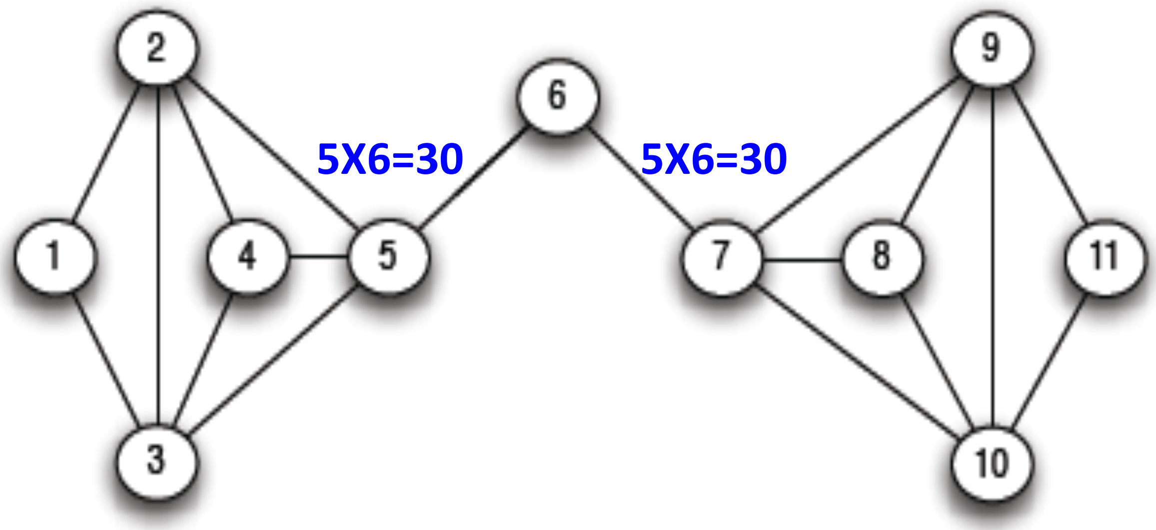
Girvan Newman method: An example



Another example

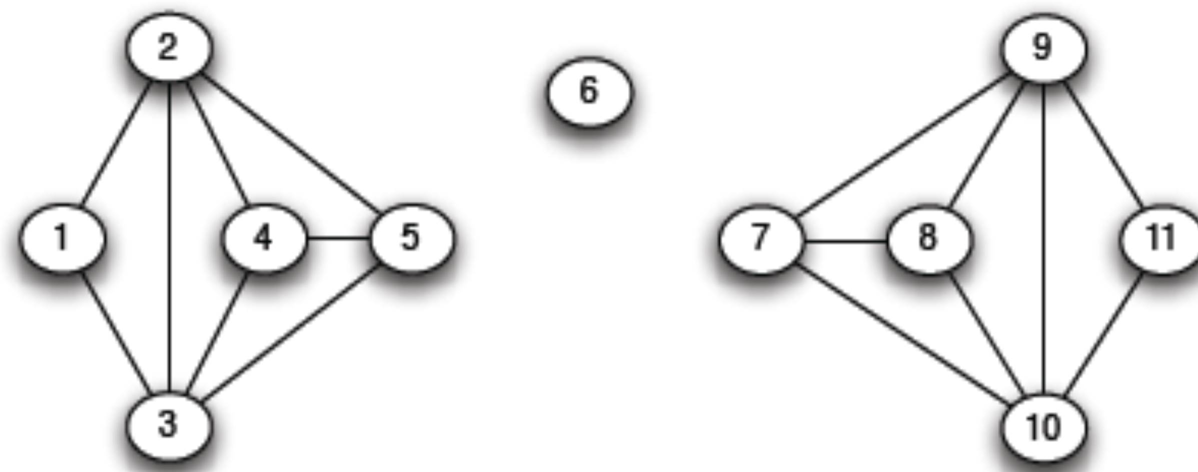


Another example



(a) *Step 1*

Another example



(b) *Step 2*

Girvan and Newman

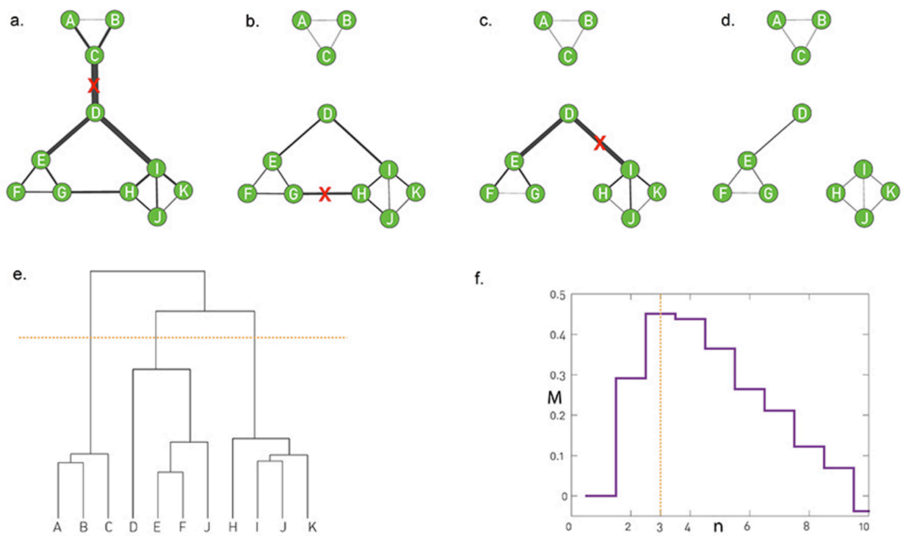


Image 9.12

The Girvan-Newman Algorithm

Divisive algorithms require a centrality measure that is high for nodes that belong to different communities and is low for node pairs in the same community. Two frequently used measures can achieve this:

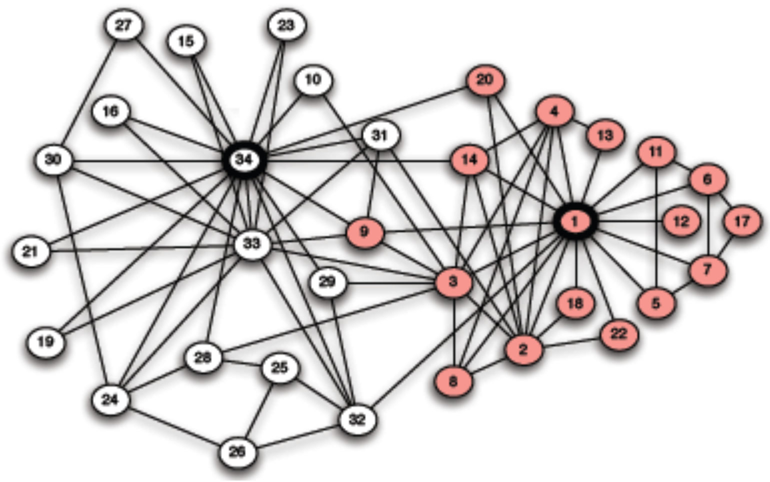
(a) The divisive hierarchical algorithm of Girvan and Newman uses link betweenness (Image 9.11a) as centrality. In the figure the link weights, assigned proportionally to x_{ij}, indicate that links connecting different communities have the highest x_{ij}. Indeed, each shortest path between these communities must run through them.

(b)-(d) The sequence of images illustrates how the algorithm removes one-by-one the three highest x_{ij} links, leaving three isolated communities behind. Note that betweenness needs to be recalculated after each link removal.

(e) The dendrogram generated by the Girvan-Newman algorithm. The cut at level 3, shown as an orange dotted line, reproduces the three communities present in the network.

(f) The modularity function, *M*, introduced in SECTION 9.4, helps us select the optimal cut. Its maxima agrees with our expectation that the best cut is at level 3, as shown in (e).

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