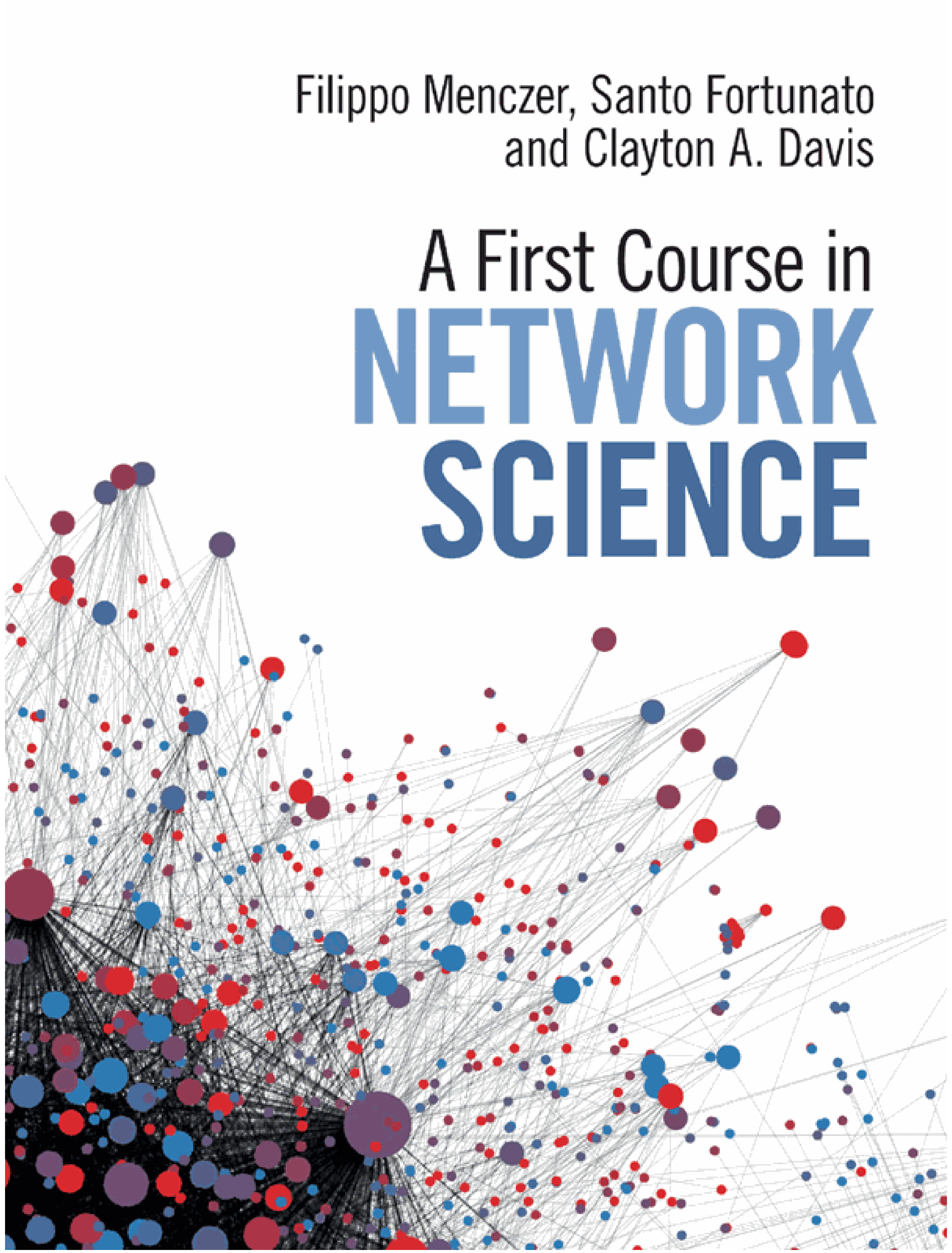
netherlands eScience center

First course in Network Science

Section 6.2

The content of this presentation is based on the book: A First Course in NETWORK SCIENCE. Filippo Menczer, Santo Fortunato, Clayton A. Davis, ISBN: 9781108471138, Cambridge University Press.

Sarah Alidoost
Analytics SIG, 07/06/2021


Filippo Menczer, Santo Fortunato
and Clayton A. Davis
A First Course in
NETWORK
SCIENCE

6.2. Related problems:

- **6.2.1 Network partitioning**
  - The minimum cut problem
  - Kernighan-Lin algorithm
  - Limitations

- **6.2.2 Data clustering**
  - Similarity measures
  - Hierarchical clustering
    - Agglomerative method
    - Dendrogram
    - Limitations

6.2.1 Network partitioning

Identifying well-separated subnetworks.

6.2.1 Network partitioning

Identifying well-separated subnetworks.

Dividing a network into clusters of given sizes, such that the total number of links connecting nodes in different clusters is minimized.

6.2.1 Network partitioning

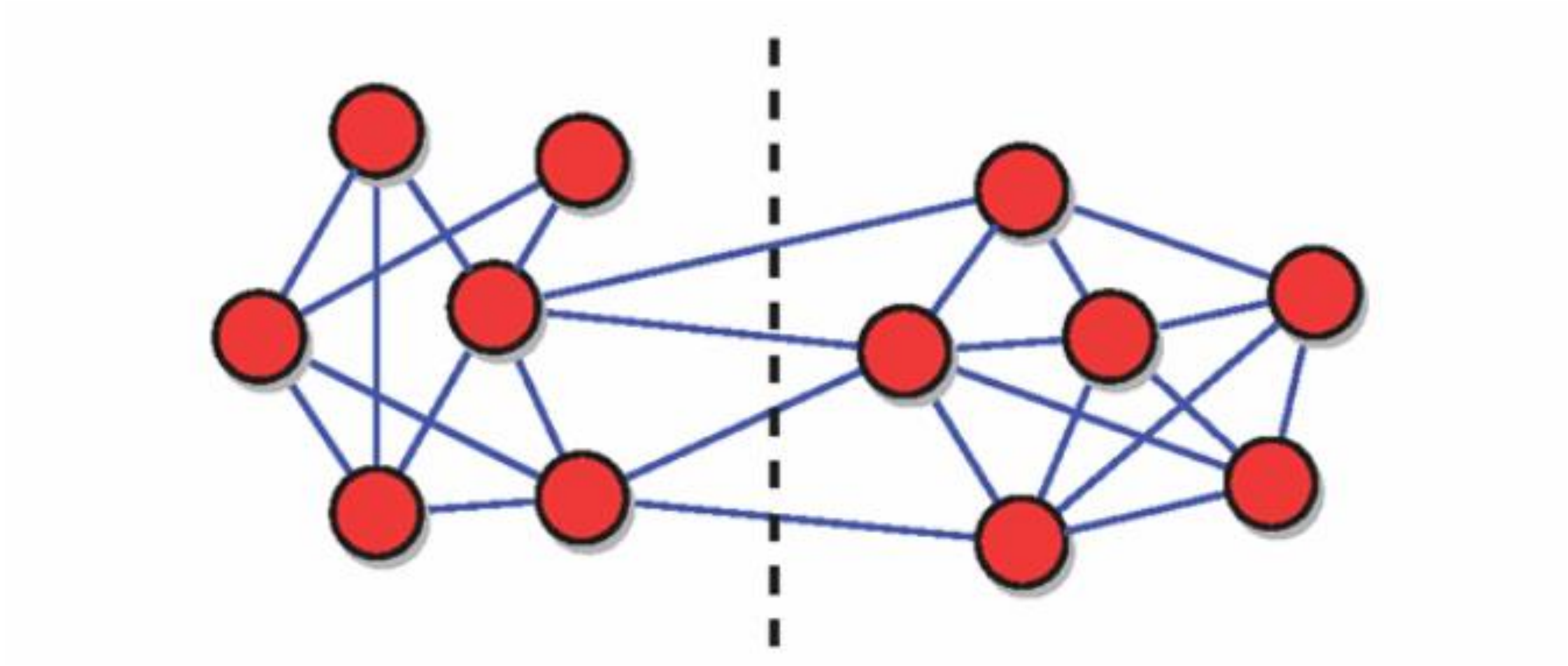
Identifying well-separated subnetworks.

Dividing a network into clusters of given sizes, such that the total number of links connecting nodes in different clusters is minimized.

Partitioning techniques in combination with other procedures are used to detect communities.

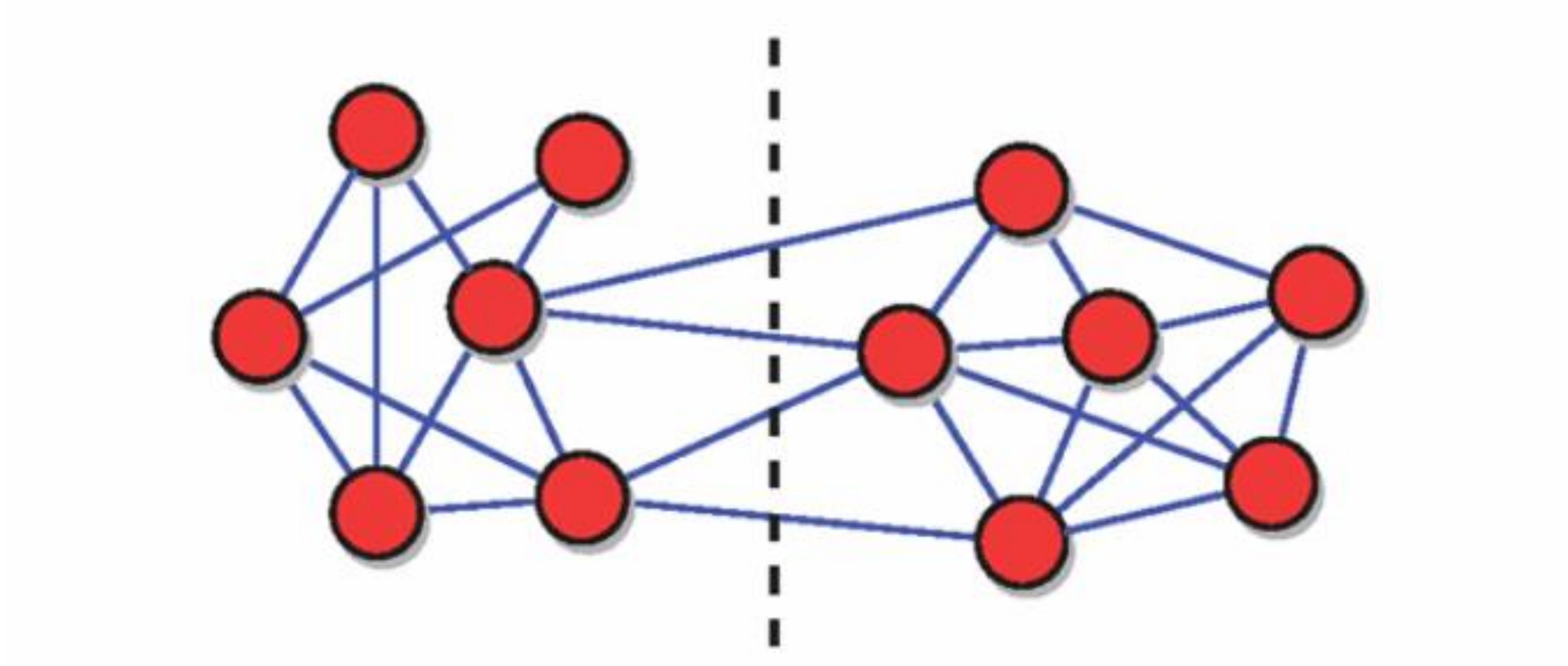
Applications of network partitioning:

- parallel computing,
- partial differential equations,
- sparse linear systems of equations,
- image processing,
- fluid dynamics,
- road networks,
- mobile communication networks,
- air traffic control.



Graph bisection.

Figure 6.8 in the book: A First Course in NETWORK SCIENCE. Filippo Menczer, Santo Fortunato, Clayton A. Davis, ISBN: 9781108471138, Cambridge University Press.

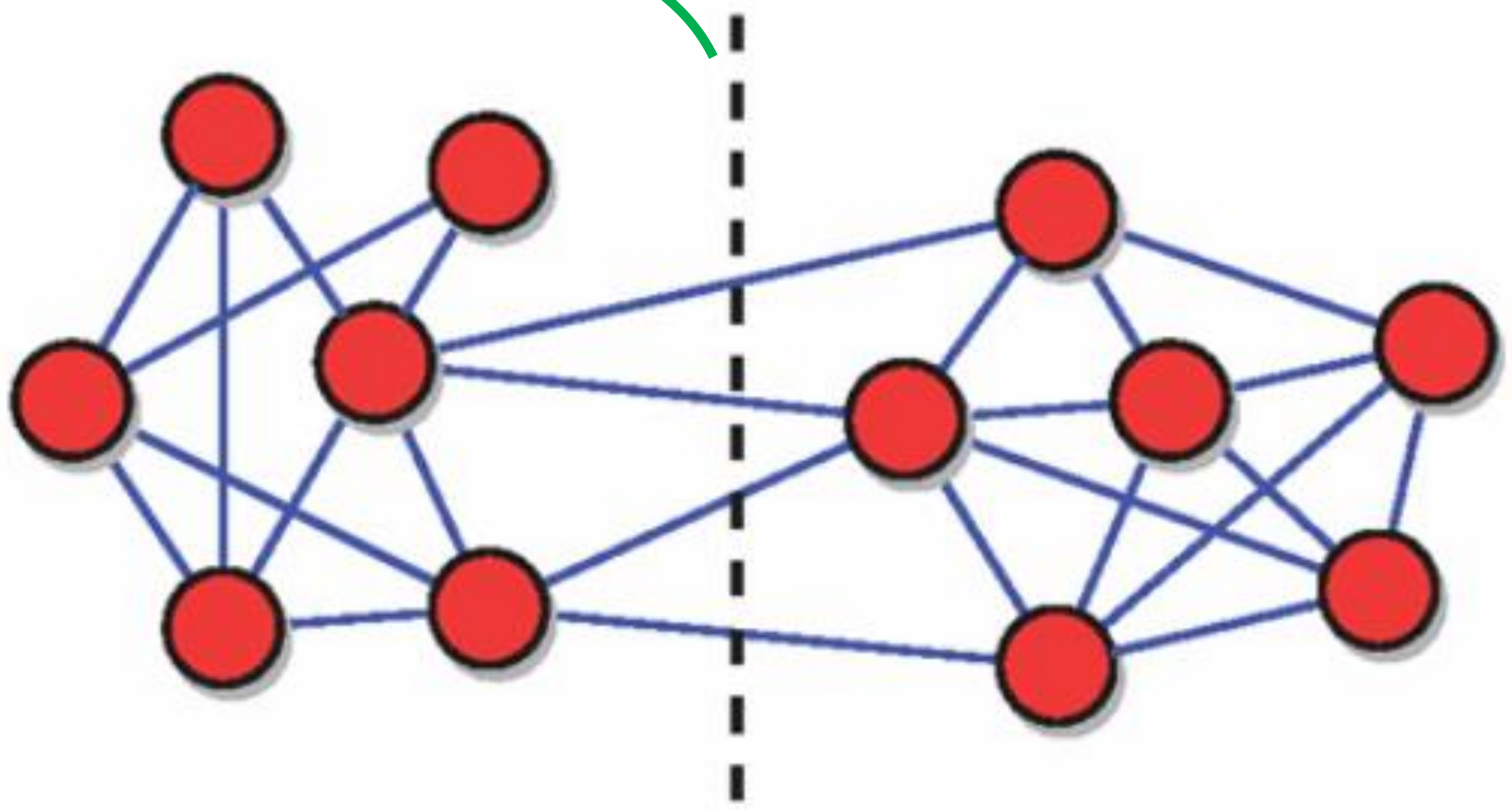


Graph bisection.

Figure 6.8 in the book: A First Course in NETWORK SCIENCE. Filippo Menczer,
Santo Fortunato, Clayton A. Davis, ISBN: 9781108471138, Cambridge University Press.

The minimum cut problem: two clusters of equal size are desired.

The set of links joining the subnetworks: ***cut.***



Graph bisection.

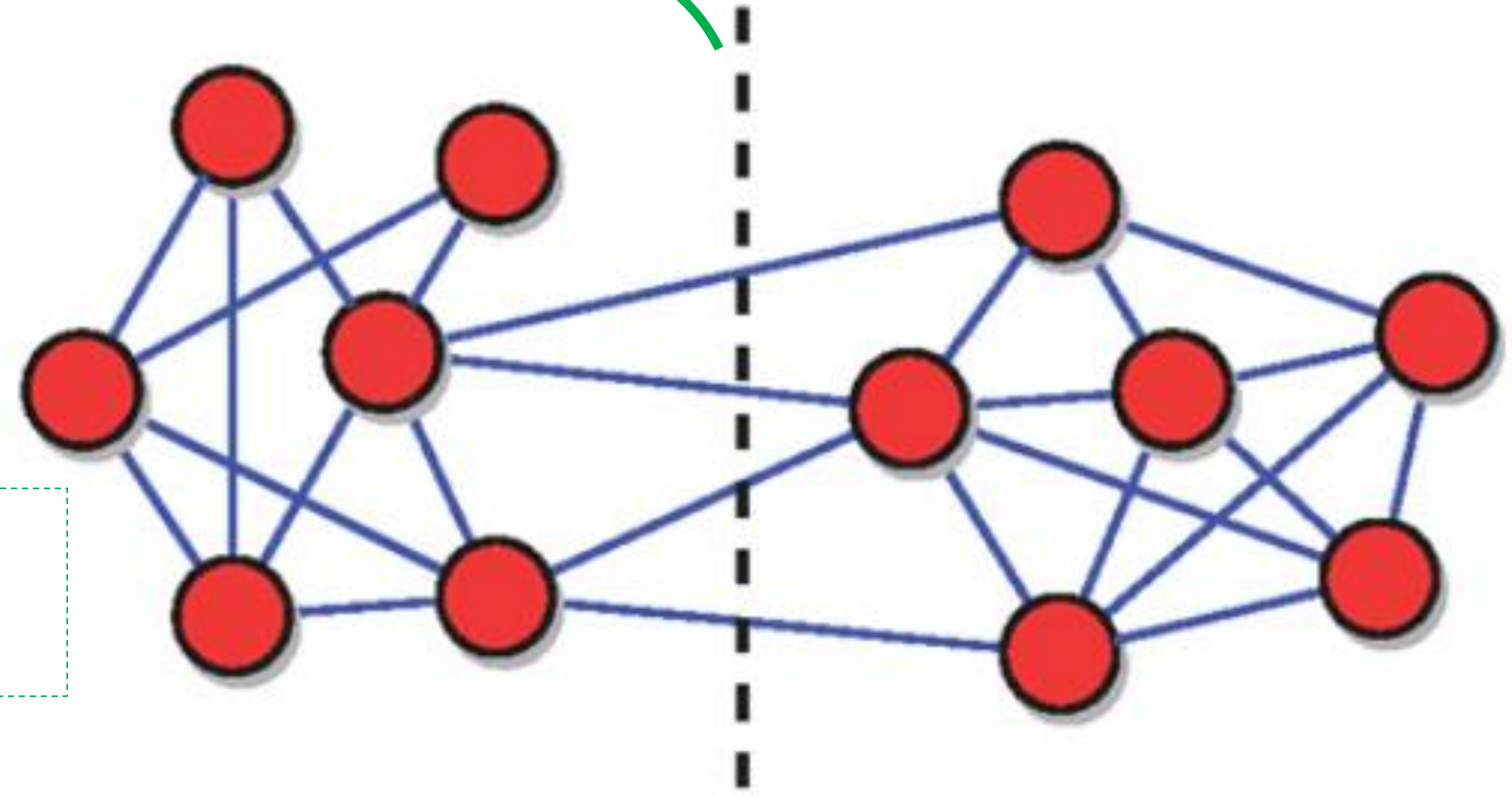
Figure 6.8 in the book: A First Course in NETWORK SCIENCE. Filippo Menczer, Santo Fortunato, Clayton A. Davis, ISBN: 9781108471138, Cambridge University Press.

The minimum cut problem: two clusters of equal size are desired.

The set of links joining

the subnetworks: ***cut.***

Their number: ***cut size.***



Graph bisection.

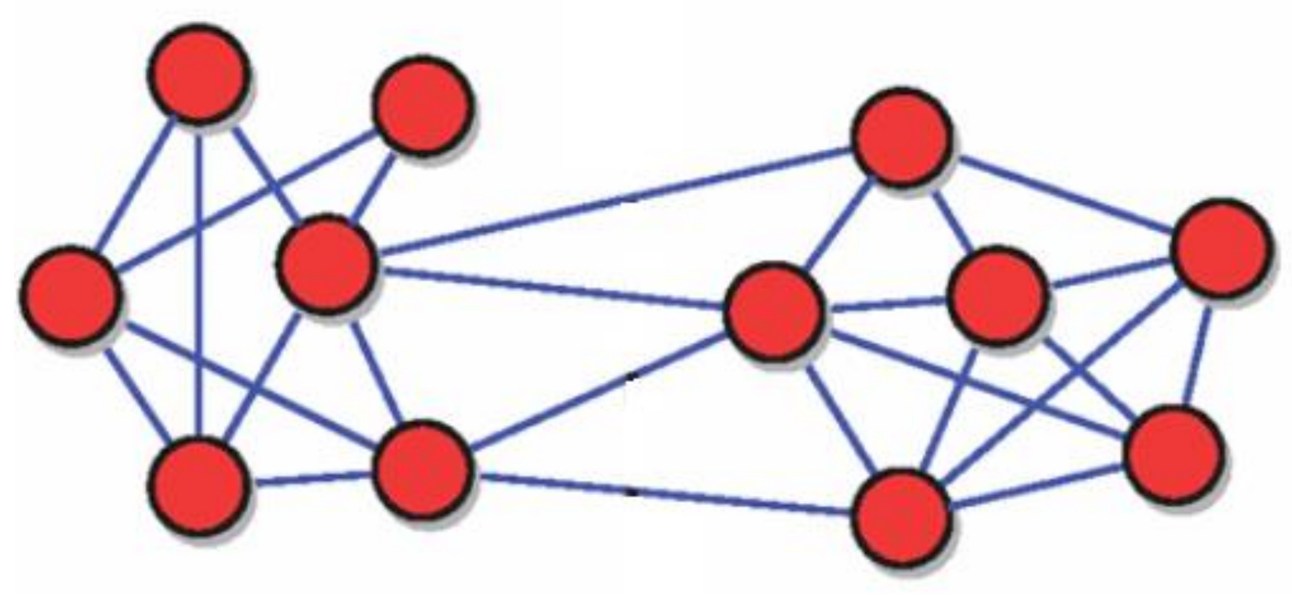
Figure 6.8 in the book: A First Course in NETWORK SCIENCE. Filippo Menczer,
Santo Fortunato, Clayton A. Davis, ISBN: 9781108471138, Cambridge University Press.

The minimum cut problem: two clusters of equal size are desired.

Question:

Why is it necessary to specify the number of clusters beforehand?

We could let the partitioning procedure find the optimal number.



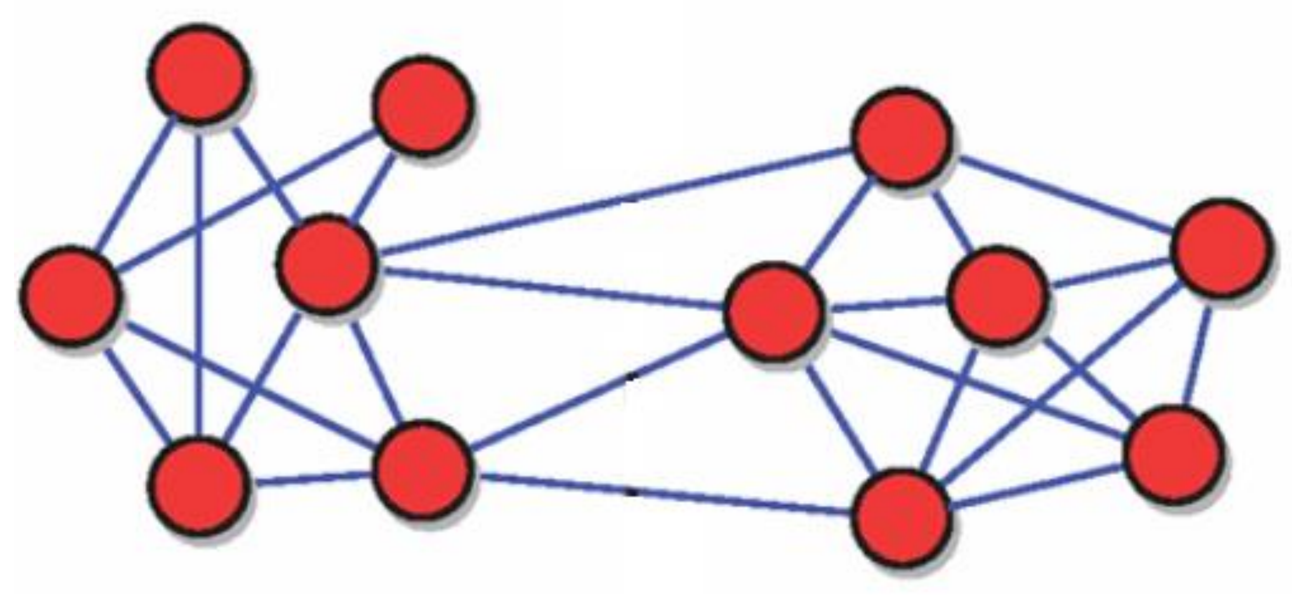
Question:

Why is it necessary to specify the number of clusters beforehand?

We could let the partitioning procedure find the optimal number.

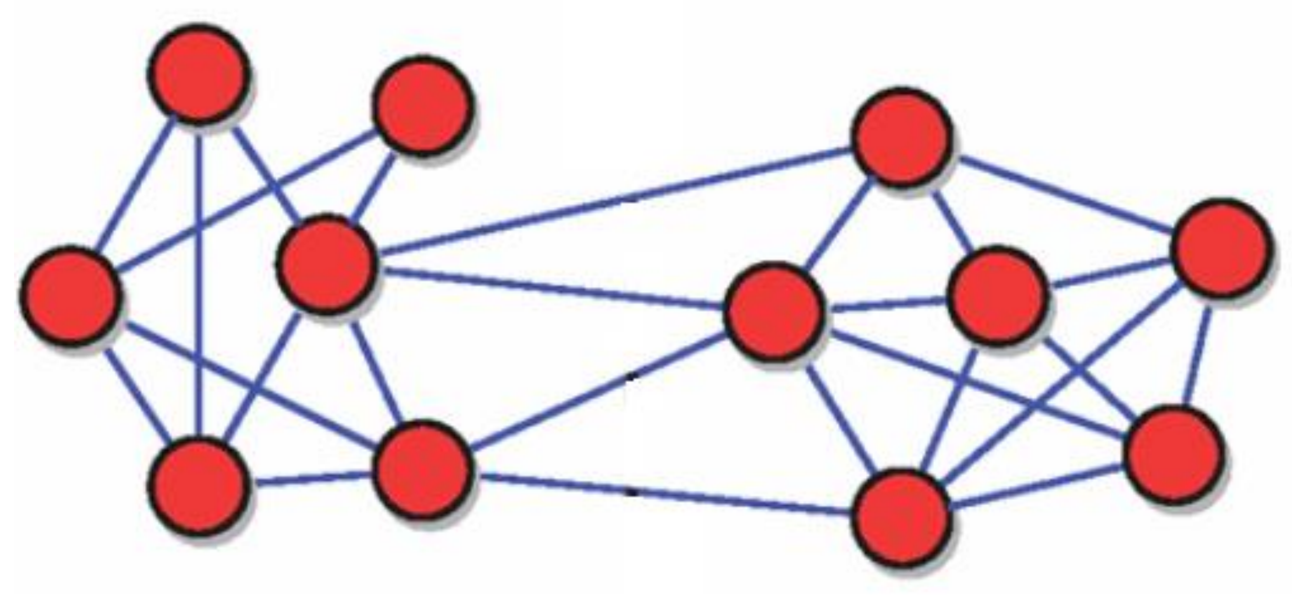
Answer:

It gives a trivial solution: the entire network has cut size zero.



Question:

Why should we specify the sizes of the clusters?

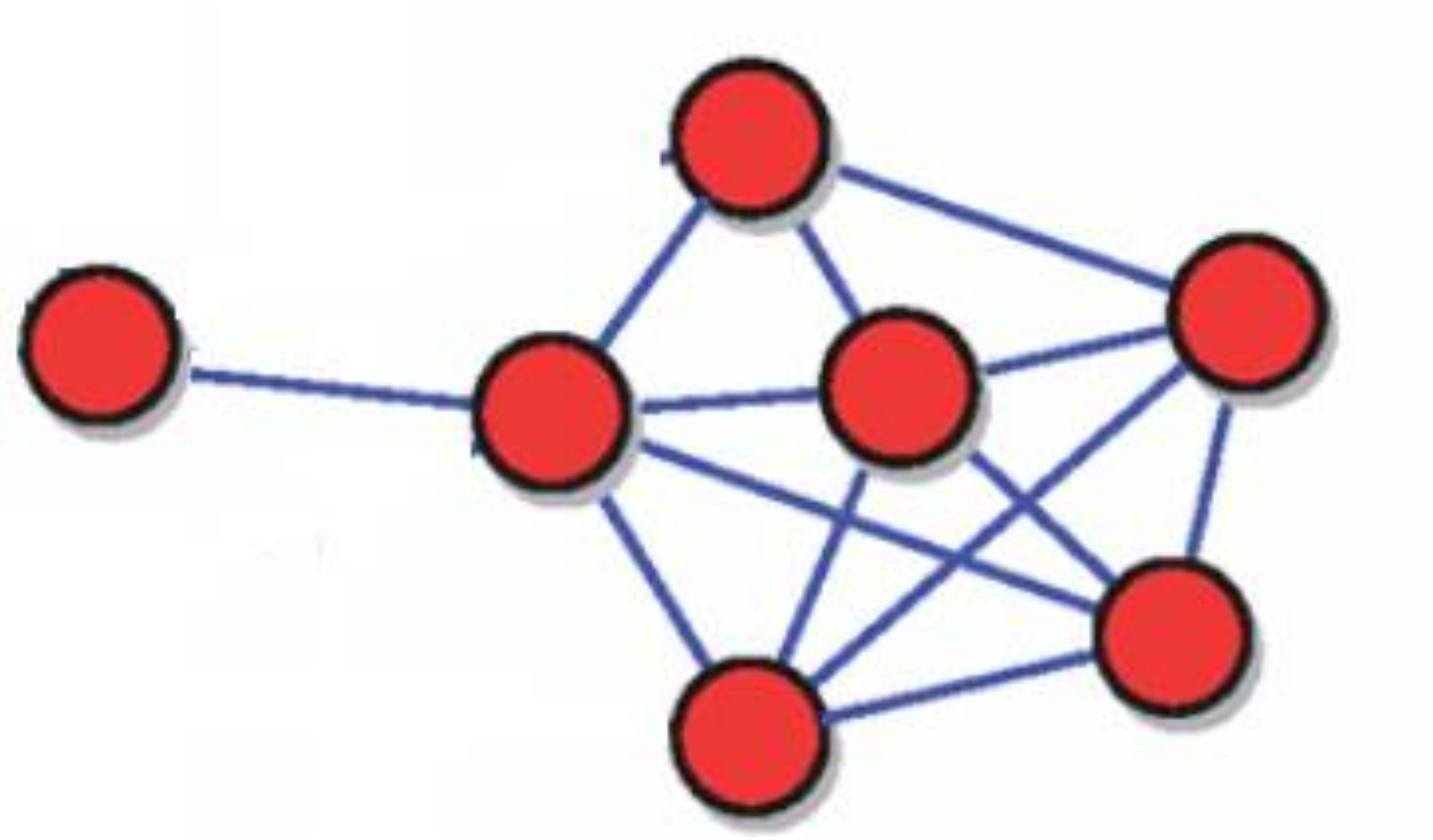


Question:

Why should we specify the sizes of the clusters?

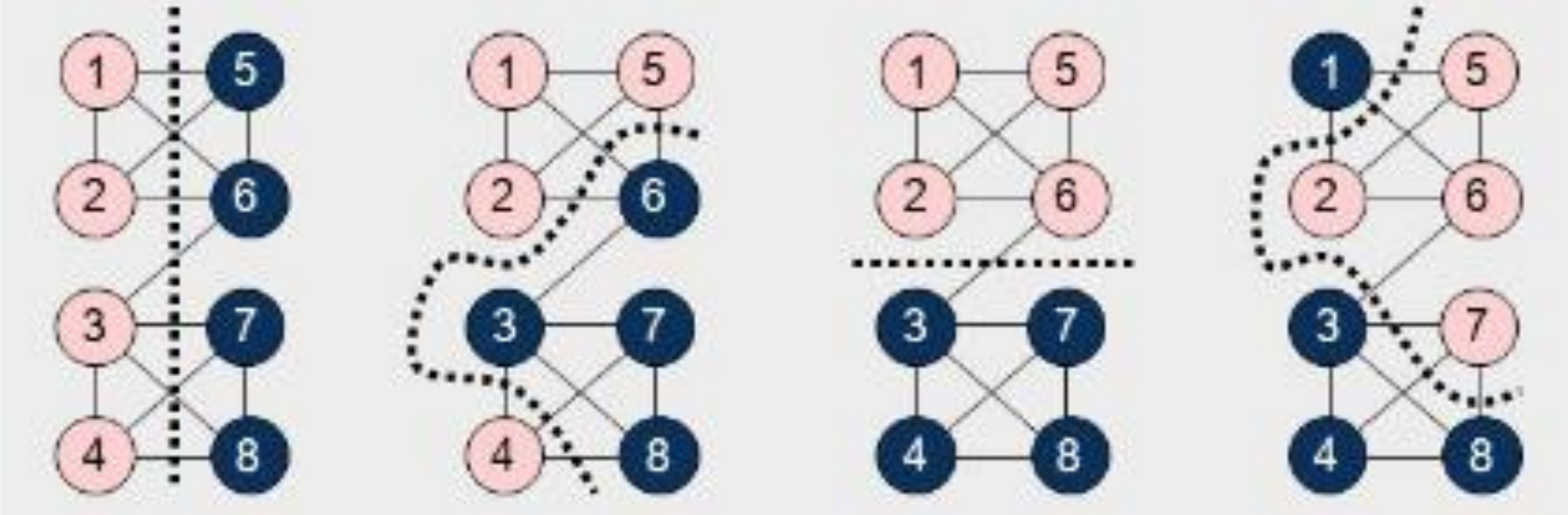
Answer:

If the network has one leaf (a node with degree one), there is a single link separating clusters.



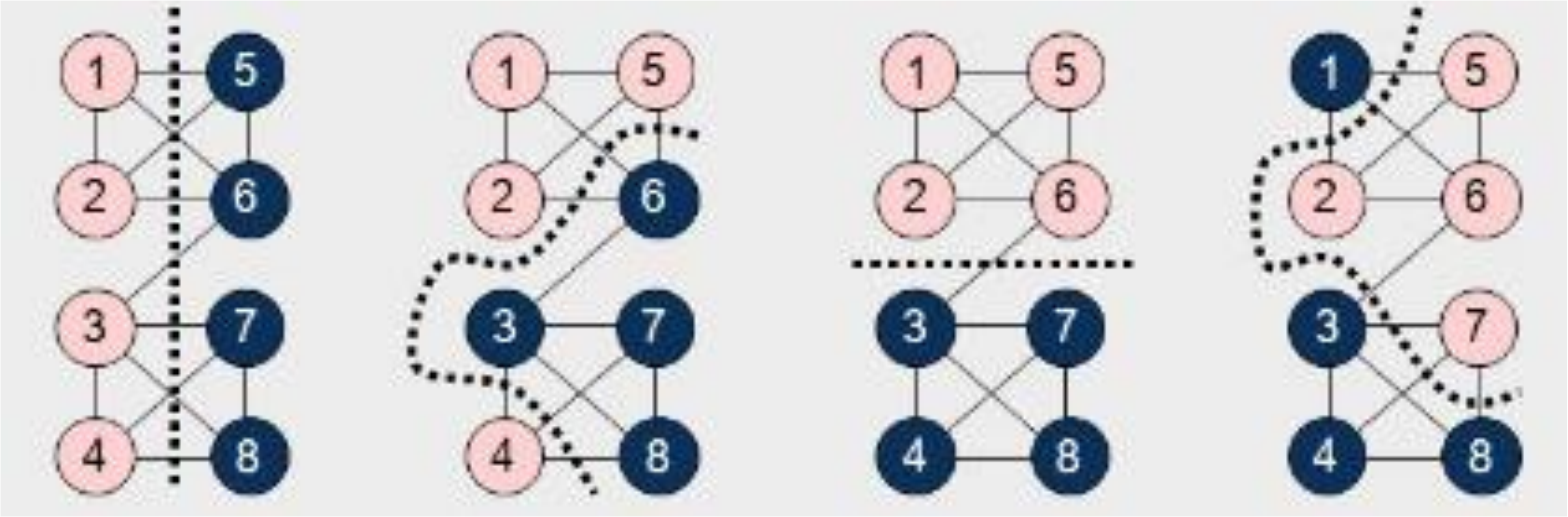
Kernighan-Lin algorithm:

We swap pairs of nodes between the clusters, while the size of the clusters does not change, and the greatest decrease of the cut size is obtained.



Chapter 2 – Netlist and System Partitioning
VLSI Physical Design: From Graph Partitioning to Timing Closure
Authors: Andrew B. Kahng, Jens Lienig, Igor L. Markov, Jin Hu

Kernighan-Lin algorithm:



Cut size= 9 6 1 7

Kernighan-Lin algorithm limitations:

- initial bisection,
- local minimum.

Kernighan-Lin algorithm limitations:

- initial bisection,
- local minimum.

As a post-processing technique to improve partitions delivered by other methods.

Kernighan-Lin algorithm limitations:

- initial bisection,
- local minimum.

As a post-processing technique to improve partitions delivered by other methods.

```
# minimum cut bisection: returns a pair of sets of nodes
partition = nx.community.kernighan_lin_bisection(G)
```

6.2.2 Data clustering

- papers or websites dealing with the same or elated topics,
- people working in the same area or department,
- proteins having similar cellular functions.

6.2.2 Data clustering

- papers or websites dealing with the same or elated topics,
- people working in the same area or department,
- proteins having similar cellular functions.

Community detection is a special version of data clustering.

Two main clustering algorithms:

- Hierarchical clustering
- Partitional clustering

Two main clustering algorithms:

- Hierarchical clustering
- Partitional clustering

Hierarchical clustering is much more frequently adapted in community detection. The main ingredient is a similarity measure between nodes.

Similarity measures:

Similarity measures can be presented as distance or derived from the structure of the network.

Similarity measures:

Similarity measures can be presented as distance or derived from the structure of the network.

An example is structural equivalence which expresses the similarity between the neighborhoods of a pair of nodes.

Similarity measures:

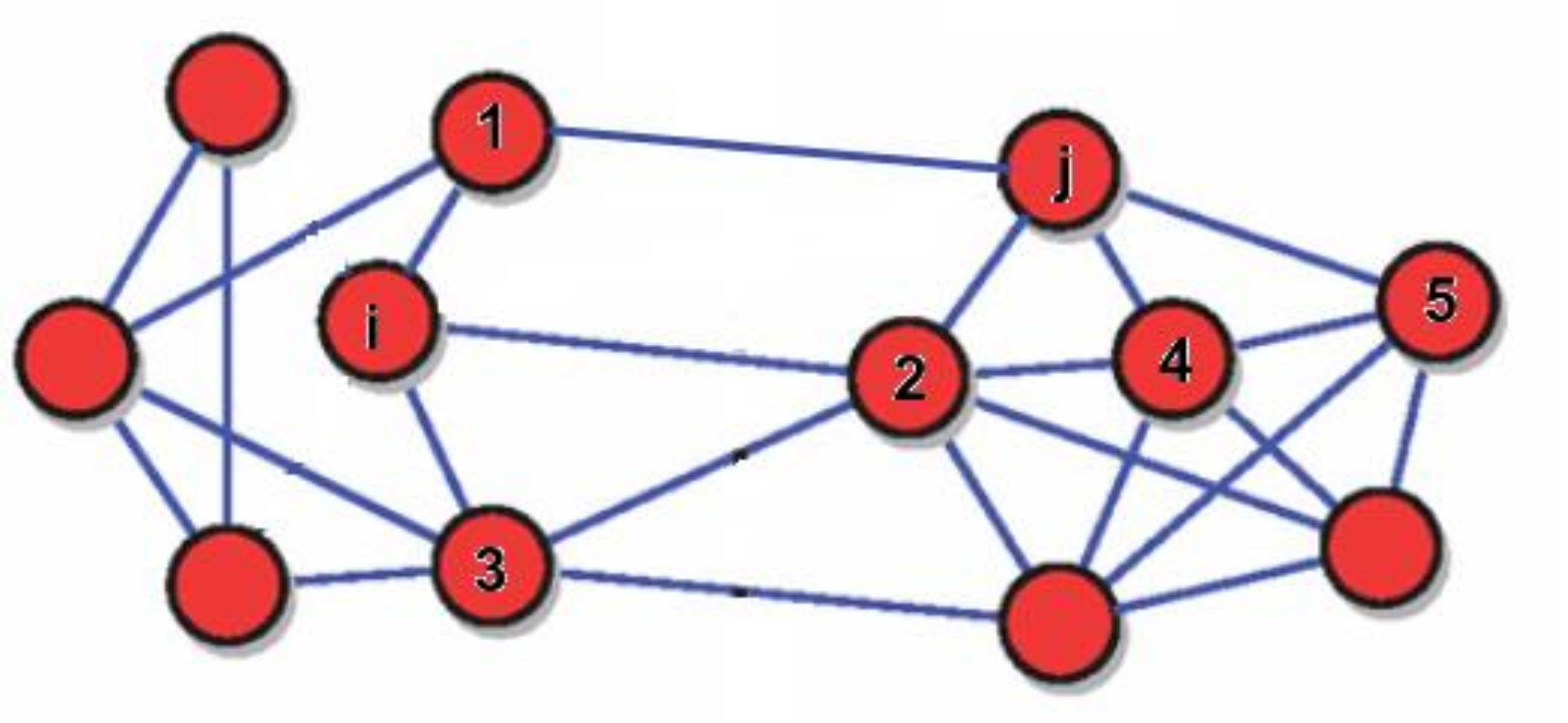
$$S_{ij}^{SE} = \frac{\text{number of neighbors shared by } i \text{ and } j}{\text{total number of nodes neighboring only } i\text{, only } j\text{, or both}}\,.$$

Similarity measures:

$$S_{ij}^{SE} = \frac{\text{number of neighbors shared by } i \text{ and } j}{\text{total number of nodes neighboring only } i\text{, only } j\text{, or both}}.$$

Question:

What is the similarity between i and j?

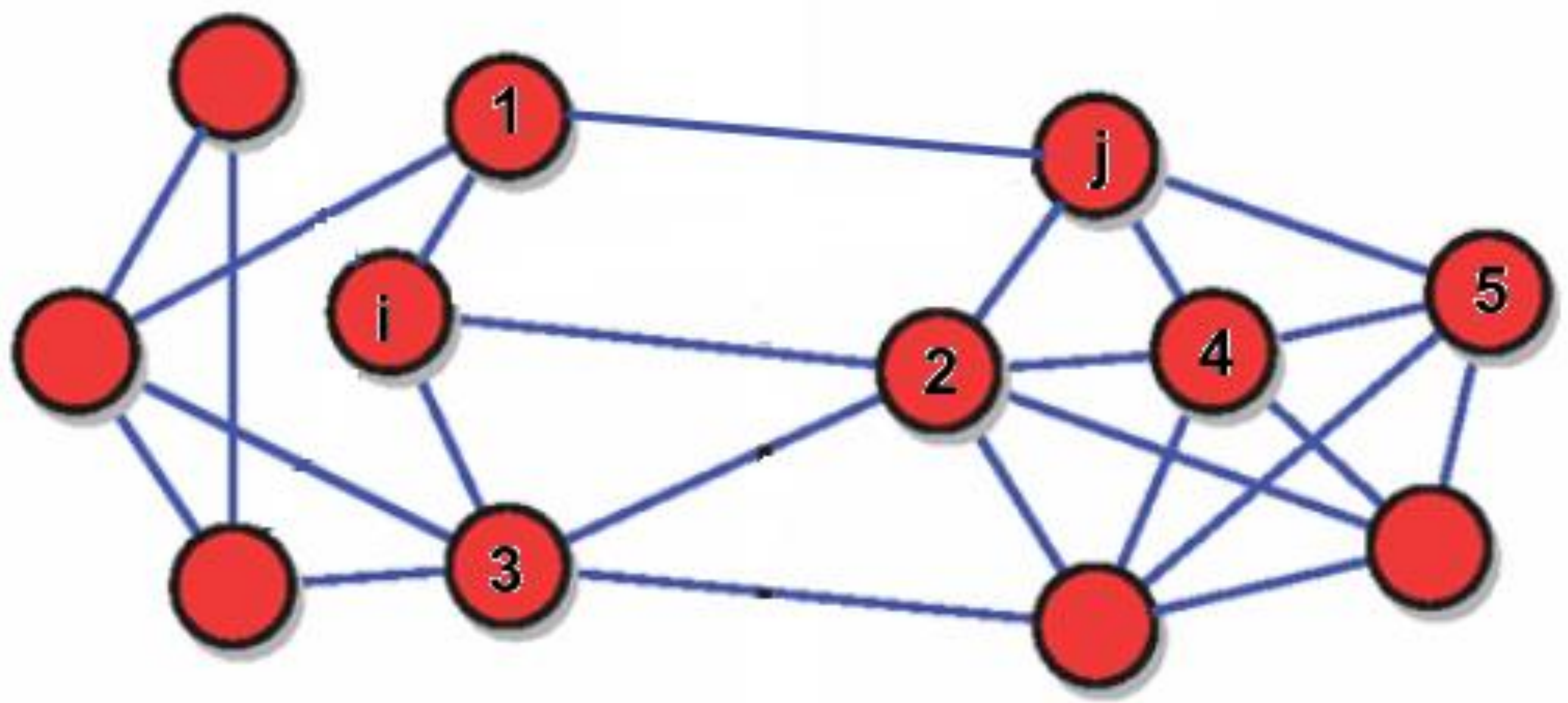


Similarity measures:

$$S_{ij}^{SE} = \frac{\text{number of neighbors shared by } i \text{ and } j}{\text{total number of nodes neighboring only } i\text{, only } j\text{, or both}}.$$

Question:

What is the similarity between i and j?



Answer:

The neighbors of i are (1, 2, 3), and

the neighbors of j are (1, 2, 4, 5),

then Sij= 2 / 5.

Similarity measures:

Similarity between two groups (G1 and G2) of nodes can be measured using:

- single linkage
- complete linkage
- average linkage

Similarity measures:

Similarity between two groups (G1 and G2) of nodes can be measured using:

- single linkage: maximum of Sij
- complete linkage: minimum of Sij
- average linkage: average of Sij

where i in G1 and j in G2.

Hierarchical clustering techniques:

- Agglomerative: partitions are generated by iteratively merging groups of nodes.

- Divisive: partitions are generated by iteratively splitting groups of nodes.

Hierarchical clustering techniques:

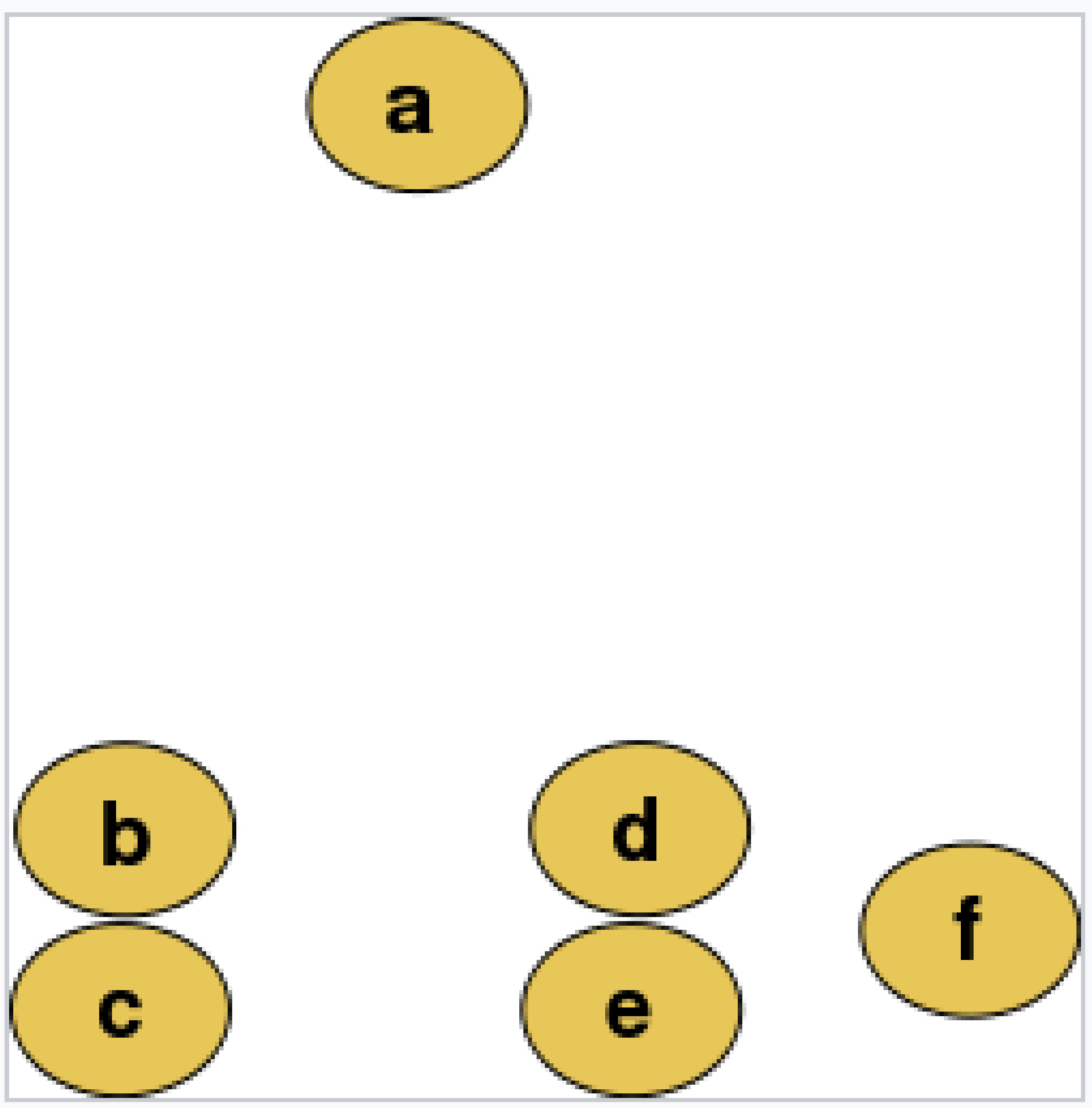
- Agglomerative: partitions are generated by iteratively merging groups of nodes.

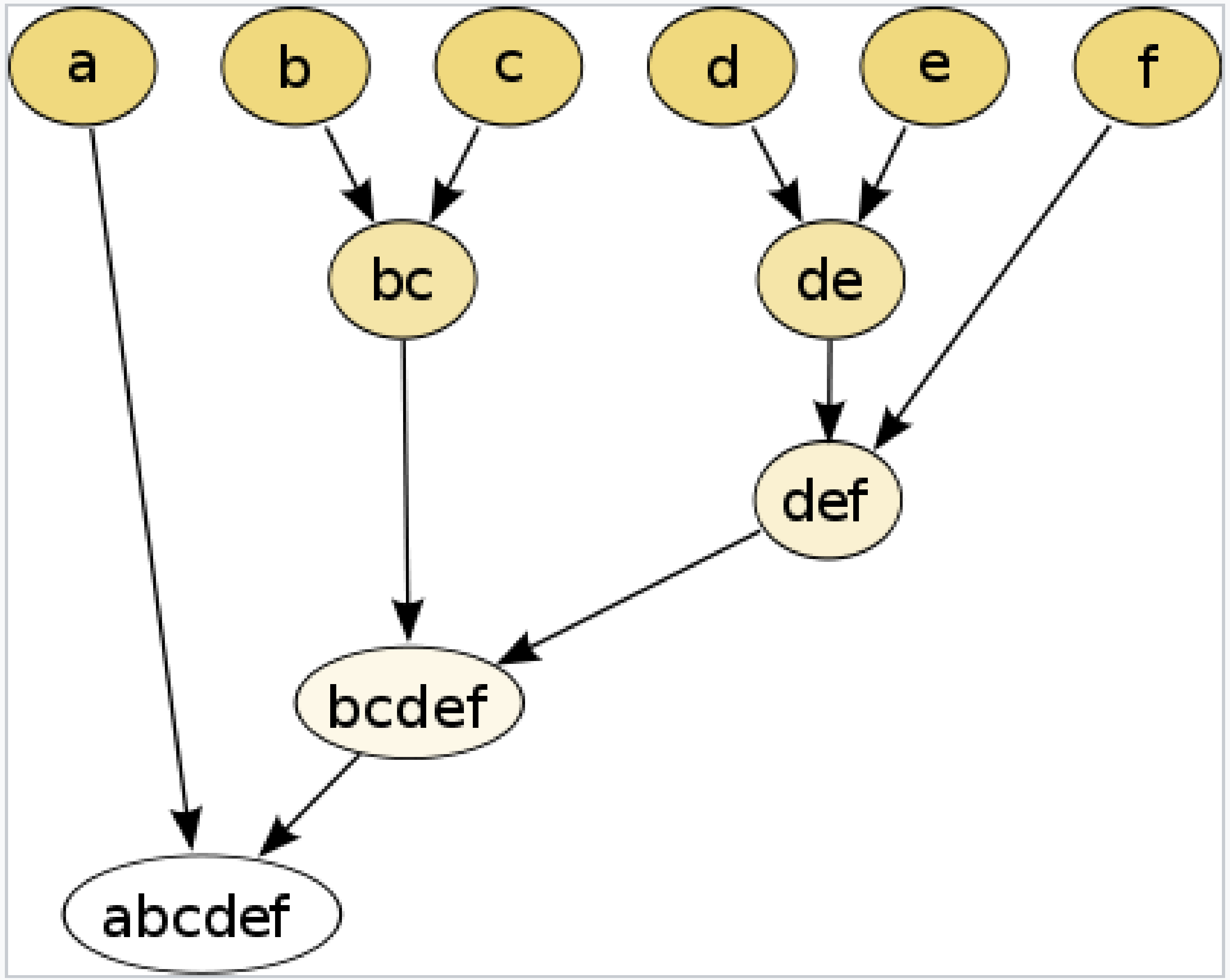
- Divisive: partitions are generated by iteratively splitting groups of nodes.

Here, we focus on agglomerative procedures

which are popular in the literature.

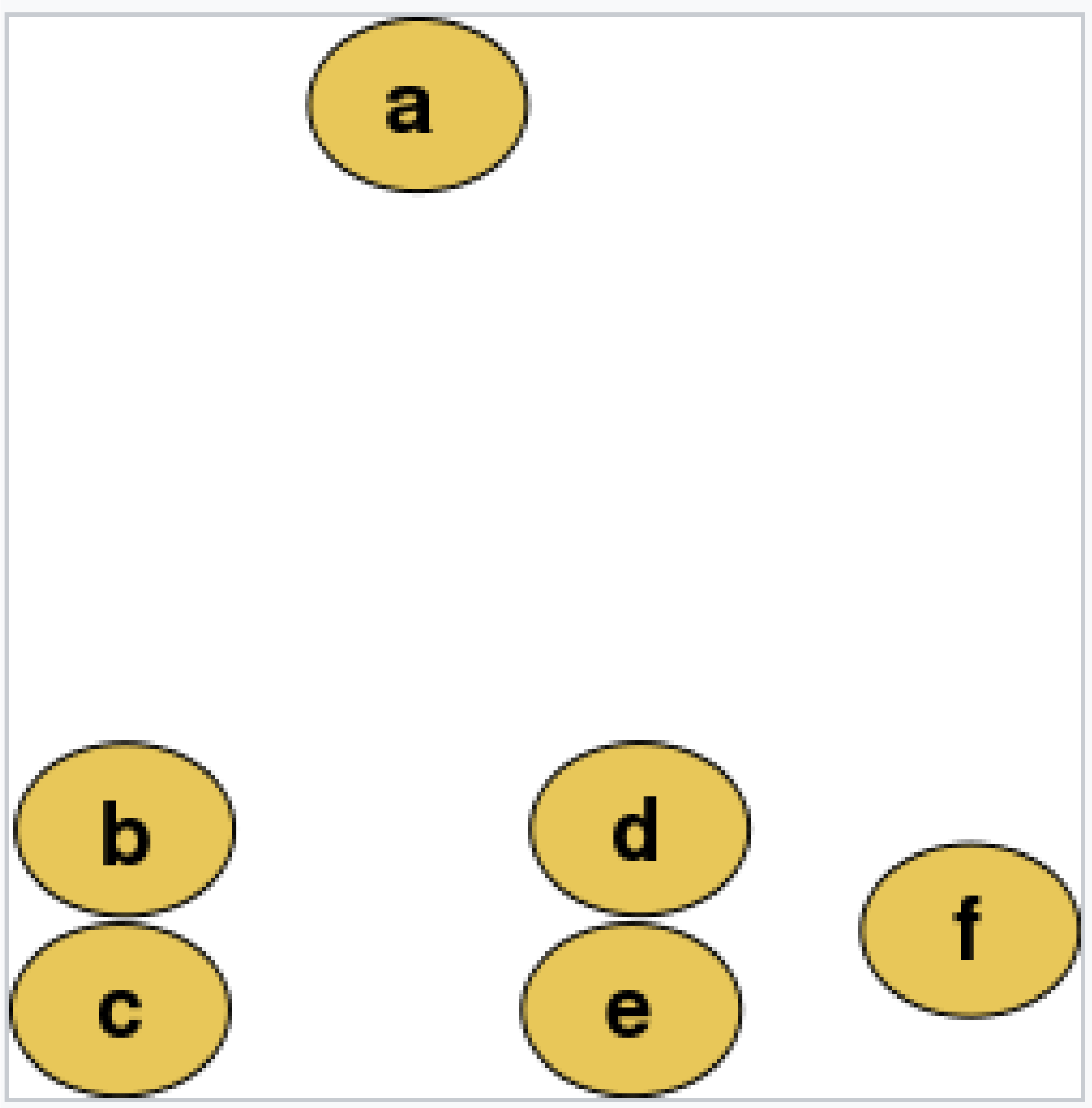
Hierarchical clustering techniques:

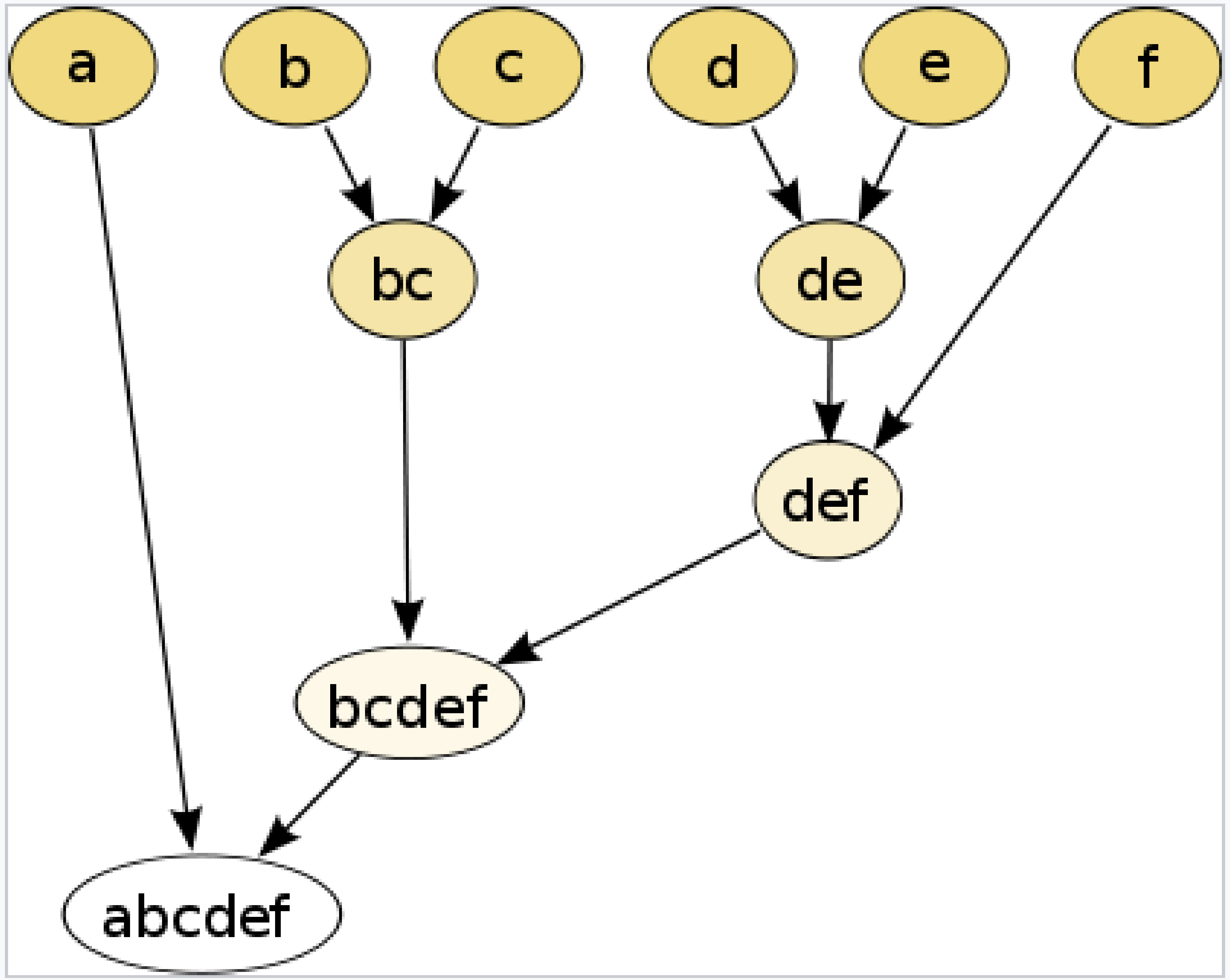




Figures from Wikipedia.

Hierarchical clustering techniques:





Figures from Wikipedia.

A dendrogram of hierarchical tree.

Example: a dendrogram for partitions of a small network.

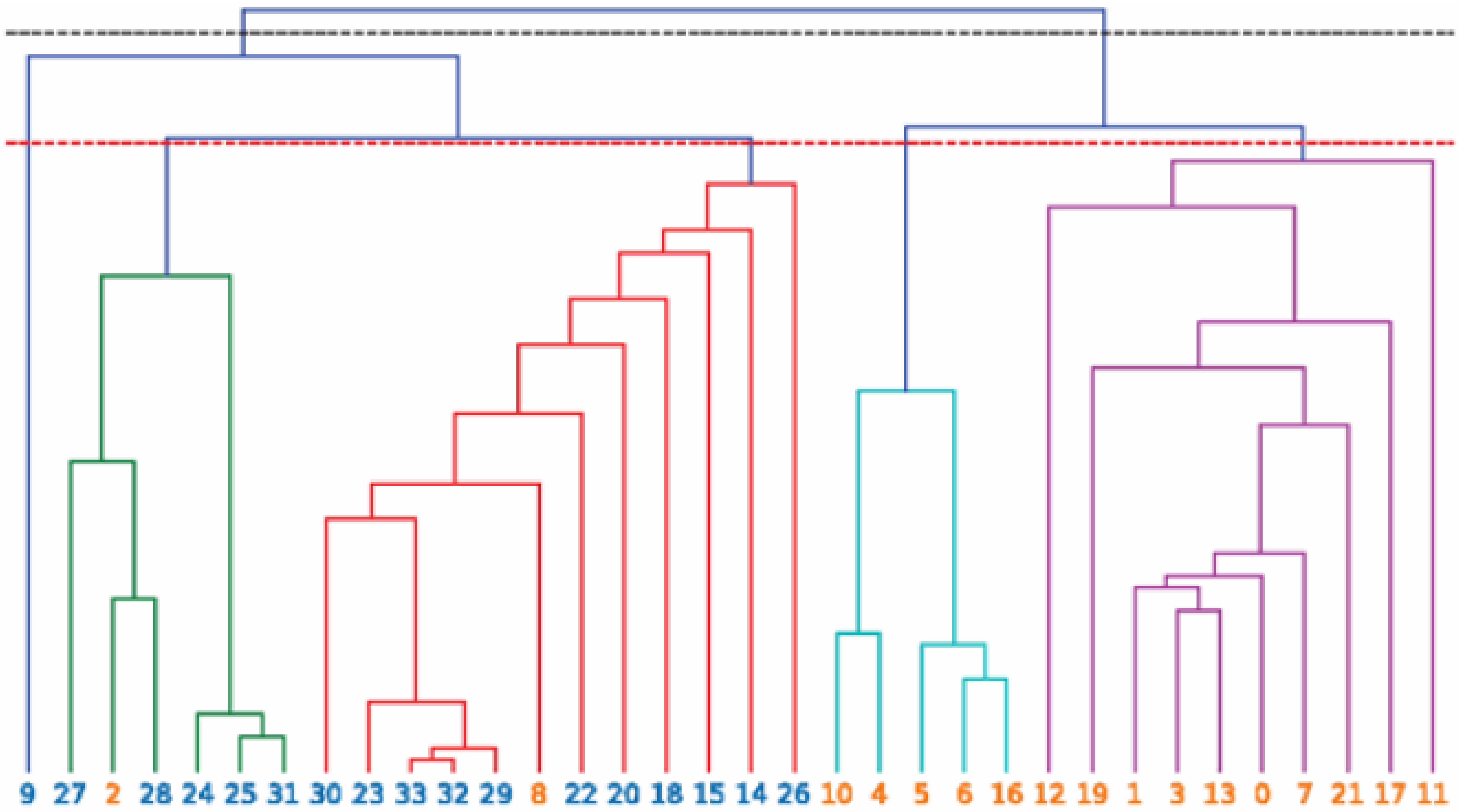
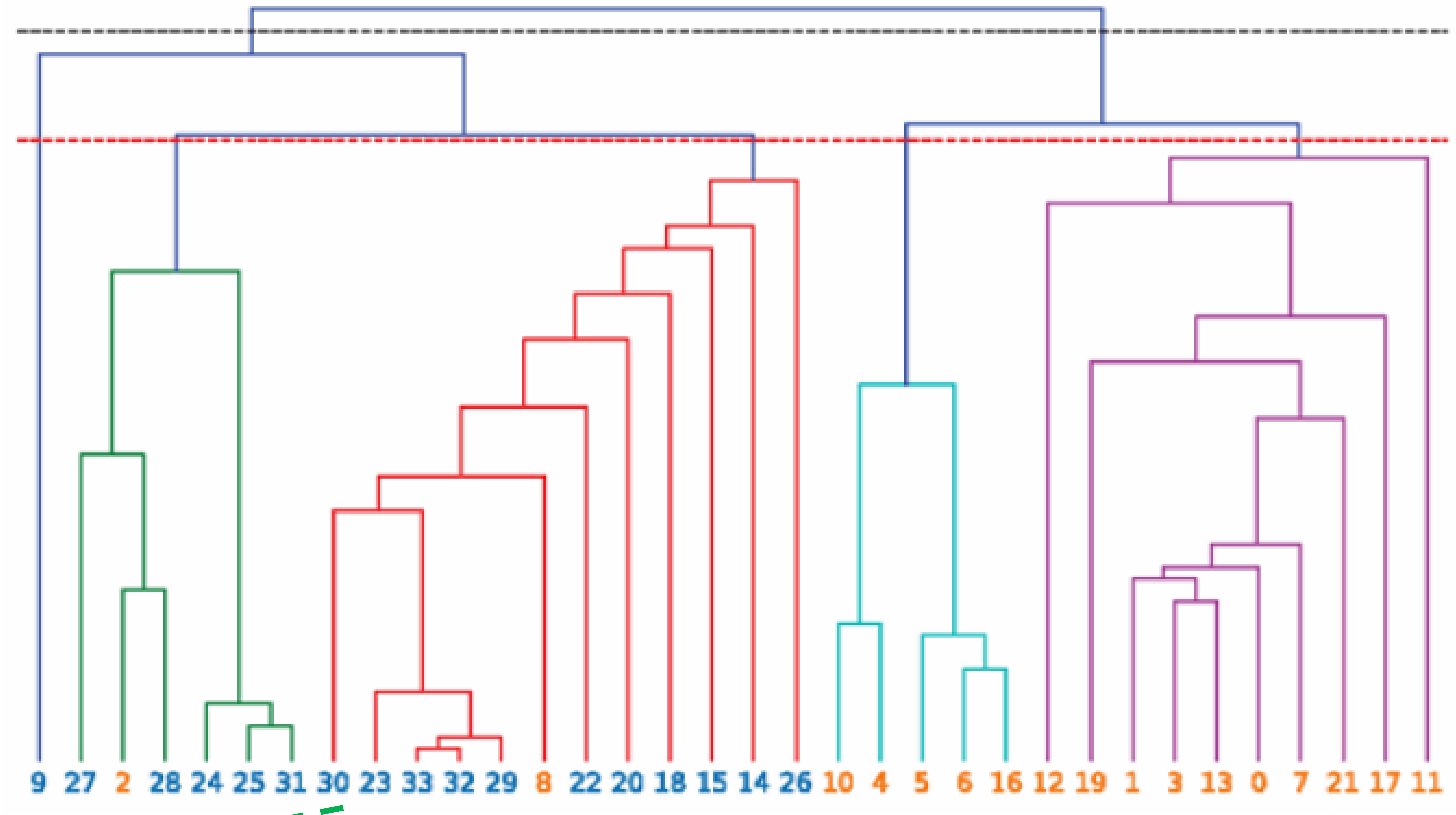


Figure 6.9 in the book: A First Course in NETWORK SCIENCE. Filippo Menczer, Santo Fortunato, Clayton A. Davis, ISBN: 9781108471138, Cambridge University Press.

Example: a dendrogram for partitions of a small network.



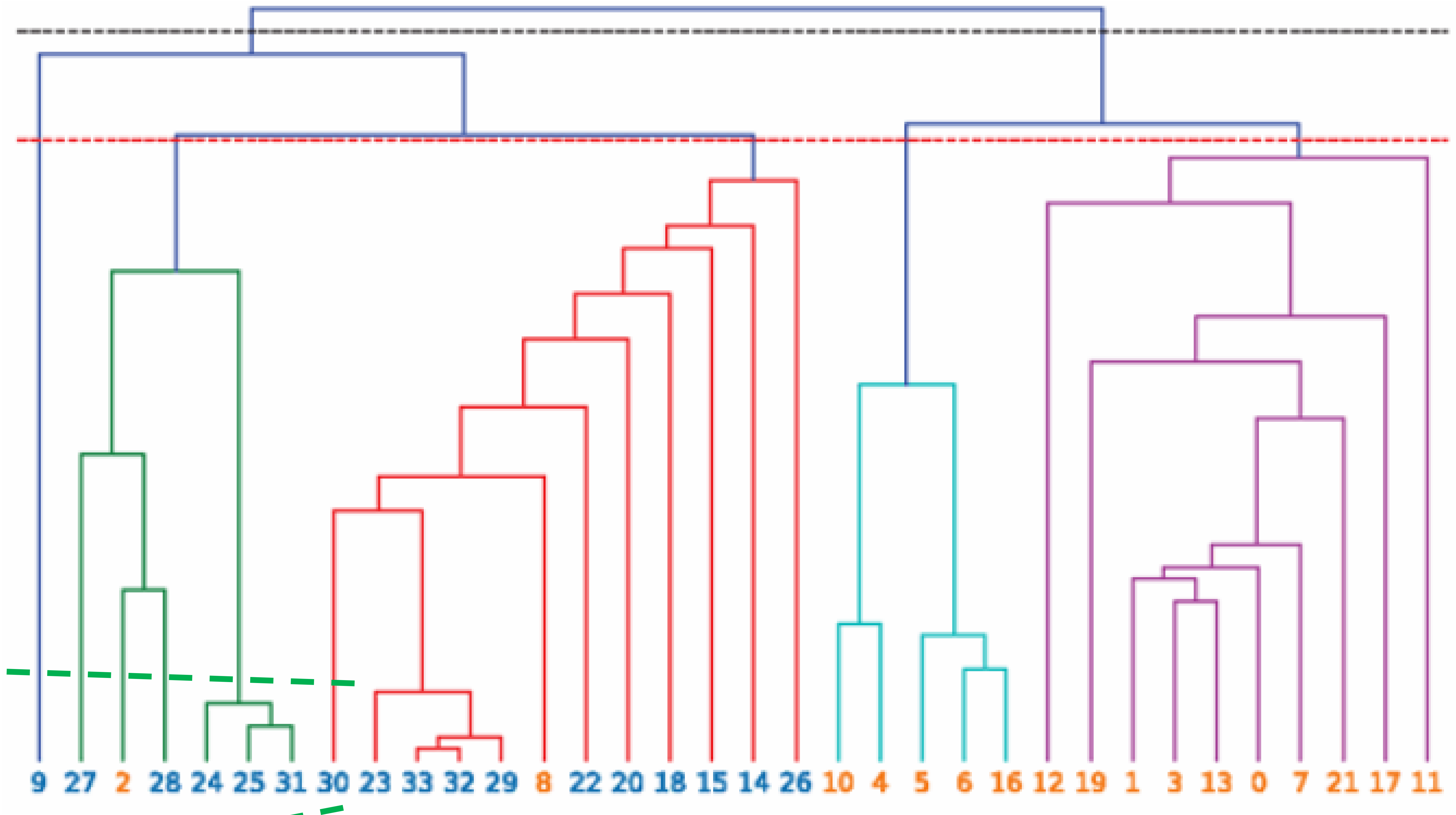
The leaves of the tree:

individual nodes.

Example: a dendrogram for partitions of a small network.

Each merge is illustrated by a horizontal line.

The leaves of the tree: individual nodes.

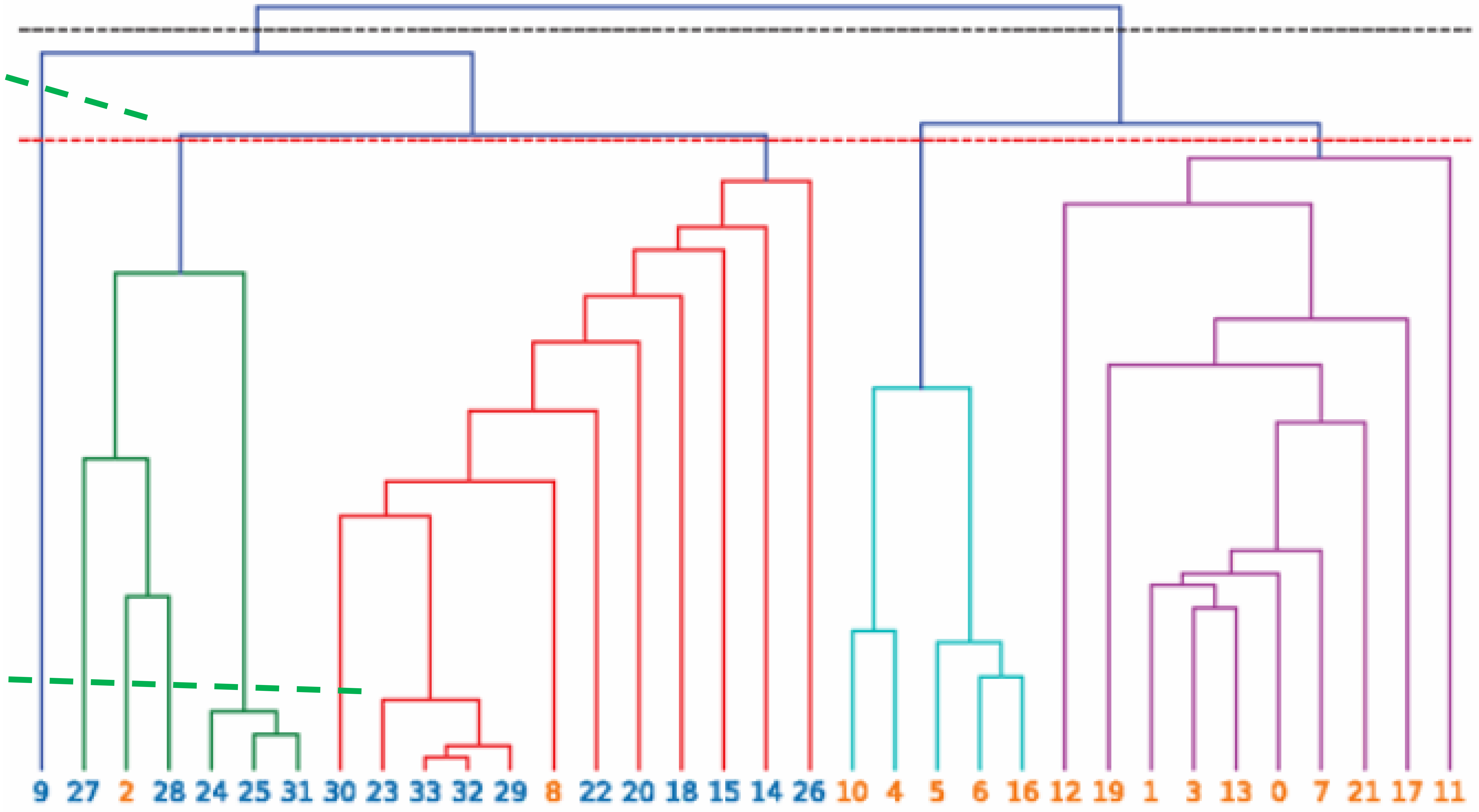


Example: a dendrogram for partitions of a small network.

To single out one of the partitions, we cut the dendrogram with a horizontal line.

Each merge is illustrated by a horizontal line.

The leaves of the tree: individual nodes.



Hierarchical clustering limitations:

- It delivers as many partitions as there are nodes without providing a criteria.

Hierarchical clustering limitations:

- It delivers as many partitions as there are nodes without providing a criteria.

- The results depend on the similarity measure.

Hierarchical clustering limitations:

- It delivers as many partitions as there are nodes without providing a criteria.

- The results depend on the similarity measure.

- The algorithms are slow.

Summary:

- Network partitioning searches for well-separated subnetworks.
- Hierarchical clustering groups nodes based on their similarity.
- Hierarchical clustering limitation is the lack of a criterion to select meaningful partitions.