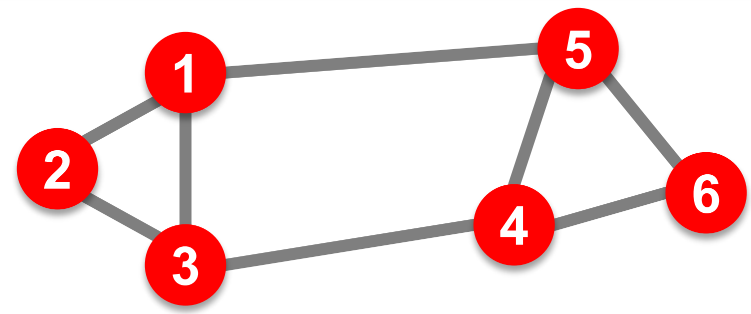
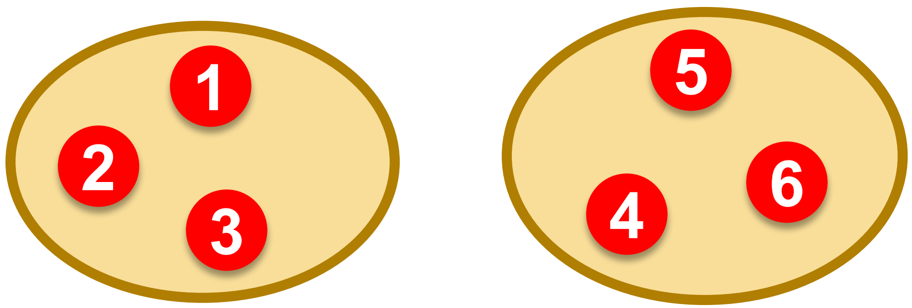
What makes a good cluster?

- **Undirected graph $G(V, E)$:**



- **Partitioning task:**
  - Divide vertices into 2 disjoint groups $A, B = V\setminus A$

A

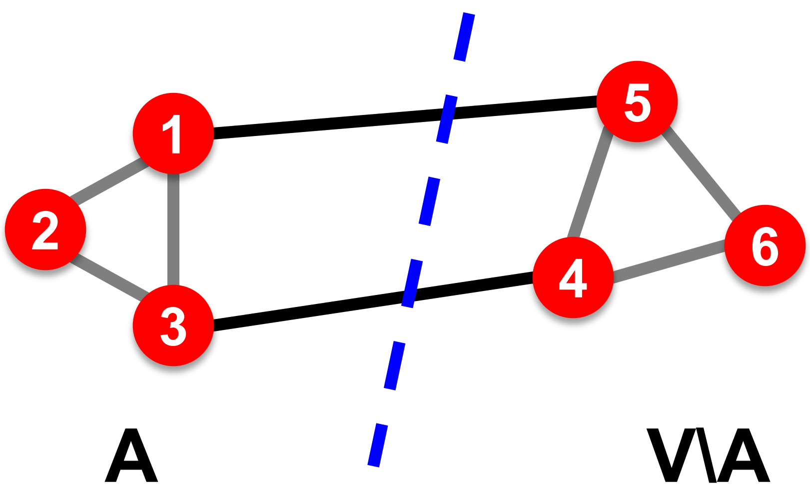


$B=V\setminus A$

- **Question:**
  - **How can we define a "good" cluster in G?**

What makes a good cluster?

- What makes a good cluster?
  - Maximize the number of within-cluster connections
  - Minimize the number of between-cluster connections

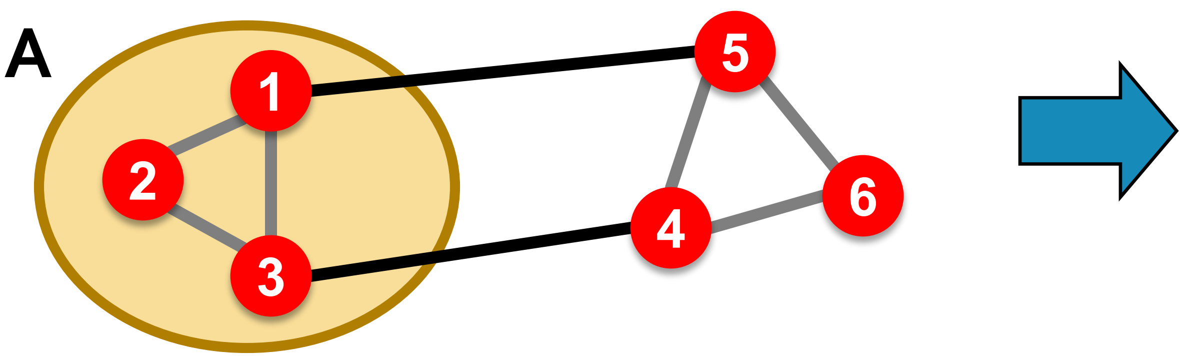


Graph Cuts

- **Express cluster quality as a function of the "edge cut" of the cluster**
- **Cut:** Set of edges with only one node in the cluster:

$$cut(A) = \sum_{i \in A, j \notin A} w_{ij}$$

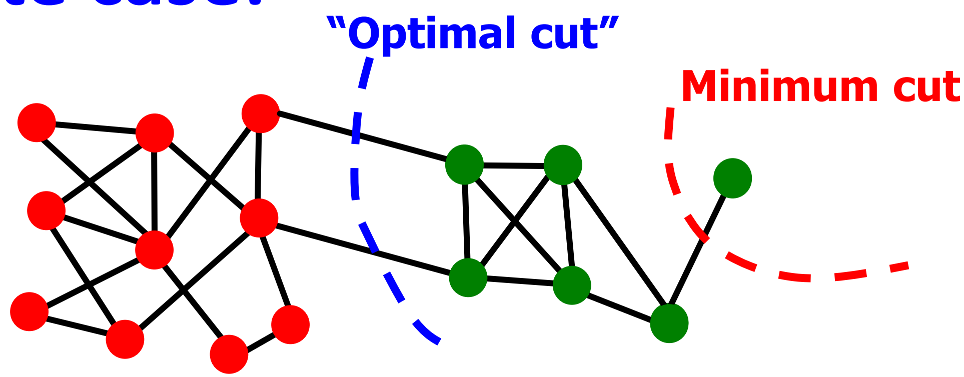
Note: This works for weighed and unweighted (set all $\mathbf{w_{ij}=1}$) graphs



$$cut(A) = 2$$

Cut Score

- **Partition quality: Cut score**
  - Quality of a cluster is the weight of connections pointing outside the cluster
- **Degenerate case:**



- **Problem:**
  - **Only considers external cluster connections**
  - **Does not consider internal cluster connectivity**

Graph Partitioning Criteria

- **Criterion: Conductance:**
  Connectivity of the group to the rest of the network relative to the density of the group

$$\phi(A) = \frac{|\{(i,j) \in E; i \in A, j \notin A\}|}{\min(vol(A),\ 2m - vol(A))}$$

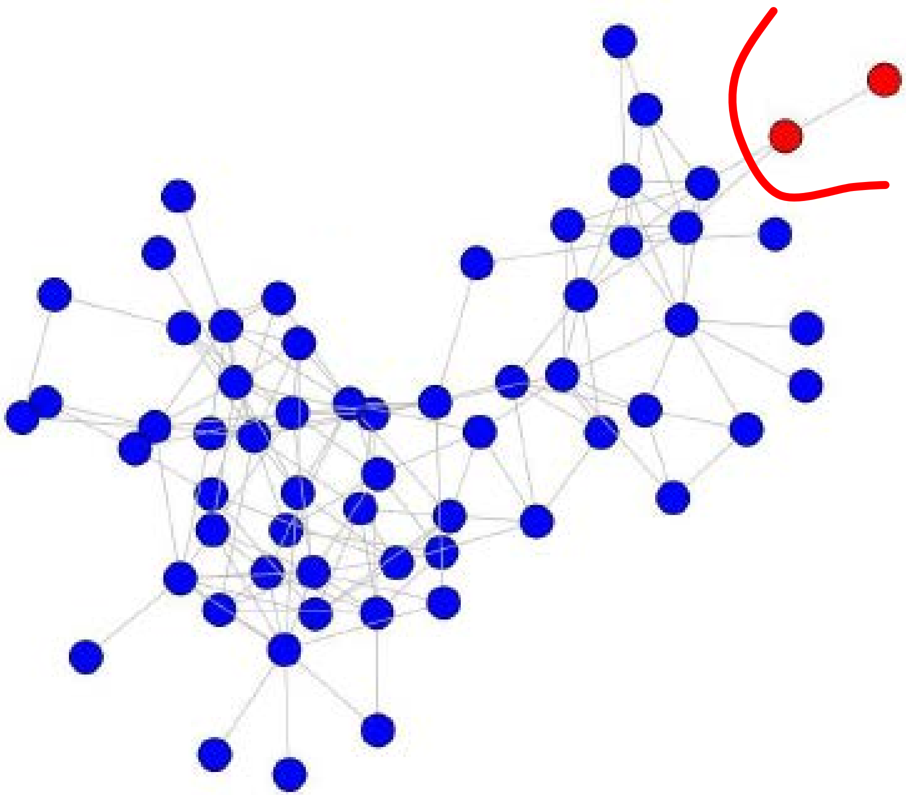
$\boldsymbol{vol(A)}$: total weight of the edges with at least one endpoint in $\boldsymbol{A}$: $\mathbf{vol}(\boldsymbol{A}) = \sum_{i \in A} \boldsymbol{d_i}$

m number of edges of the graph

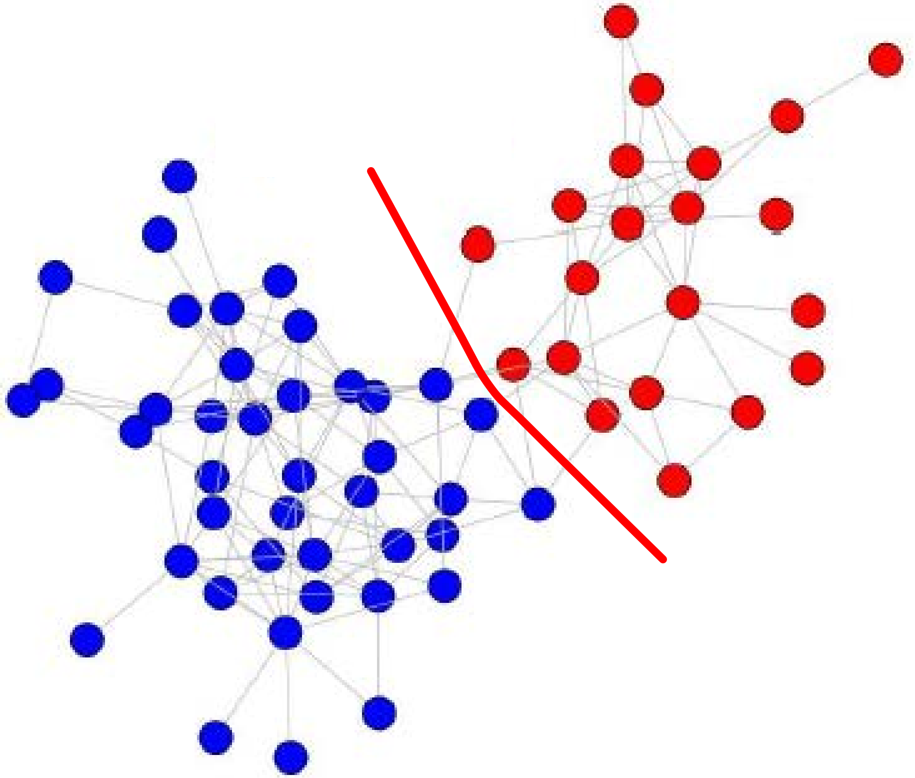
d_i degree of node ***i***

- **Why use this criterion?**
  - Produces more balanced partitions

Example: Conductance Score



$$\phi = 2/4 = 0.5$$



$$\phi = 6/92 = 0.065$$