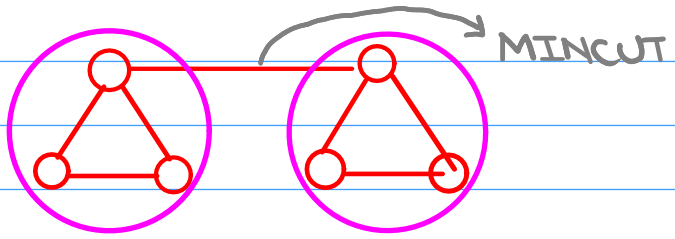
* Graph Partitioning
* Clustering



→ Natural Communities

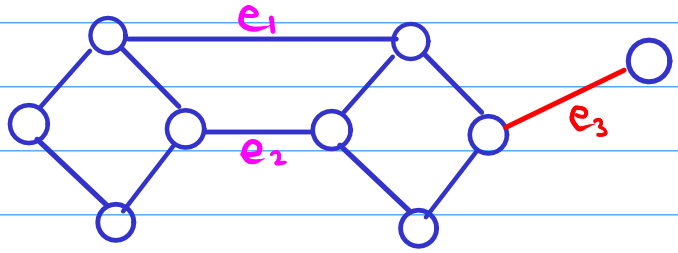
* Girvan - Newman Algorithm (GN algo)

The algorithm's steps for community detection are summarized below

1. The betweenness of all existing edges in the network is calculated first.
2. The edge(s) with the highest betweenness are removed.
3. The betweenness of all edges affected by the removal is recalculated.
4. Steps 2 and 3 are repeated until no edges remain.

* edge betweeness :
→ edges in max no. of paths

→ Flaw / Limitation :



* GN requires removal of e_1, e_2 whereas job could be done by removing e_3.

* Internal Connectivity

① Internal Density $\frac{m_s}{n_s(n_s-1)/2}$

$G : \langle V, E \rangle$ S : Set of nodes
m_s : edges in set , $n_s = |S|$, $m = |E|$
c_s : edges (outgoing) from S

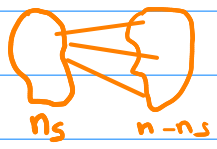
② Avg Degree $2m_s/n_s$ (Σdegrees = 2 x edges)

③ Fraction over median degree $\frac{|\{u : u \in S, |\{(u,v) : v \in S\}| > d_m\}|}{n_s}$

* External Connectivity

① Expansion c_s/n_s

② Cut-Ratio $c_s/n_s(n-n_s)$



* Ideal Community :
Small Expansion / Cut-Ratio & High Internal Density

✱ Internal + External Connectivity

① Conductance $\frac{C_S}{2m_S + C_S}$ ← Sum of degrees of S

✱ Ideally low conductance

② Normalized-Cut $\frac{C_S}{2m_S + C_S} + \frac{C_S}{2(m - m_S) + C_S}$ ← Sum of other degrees

✱ Network Model

Modularity

$$Q = \frac{1}{4m} \sum_{i,j} \left(A_{ij} - \frac{d_i d_j}{2m}\right)(s_i s_j + 1)$$

$\frac{d_i d_j}{2m}$ → Expected number of edges between Nodes i, j

$m = |E|$, A : Adjacency matrix

d_i → degree of i^{th} node

$s_i = 1$ (Community 1)

$= -1$ (Community 2)

✱ Maximizing modularity beneficial