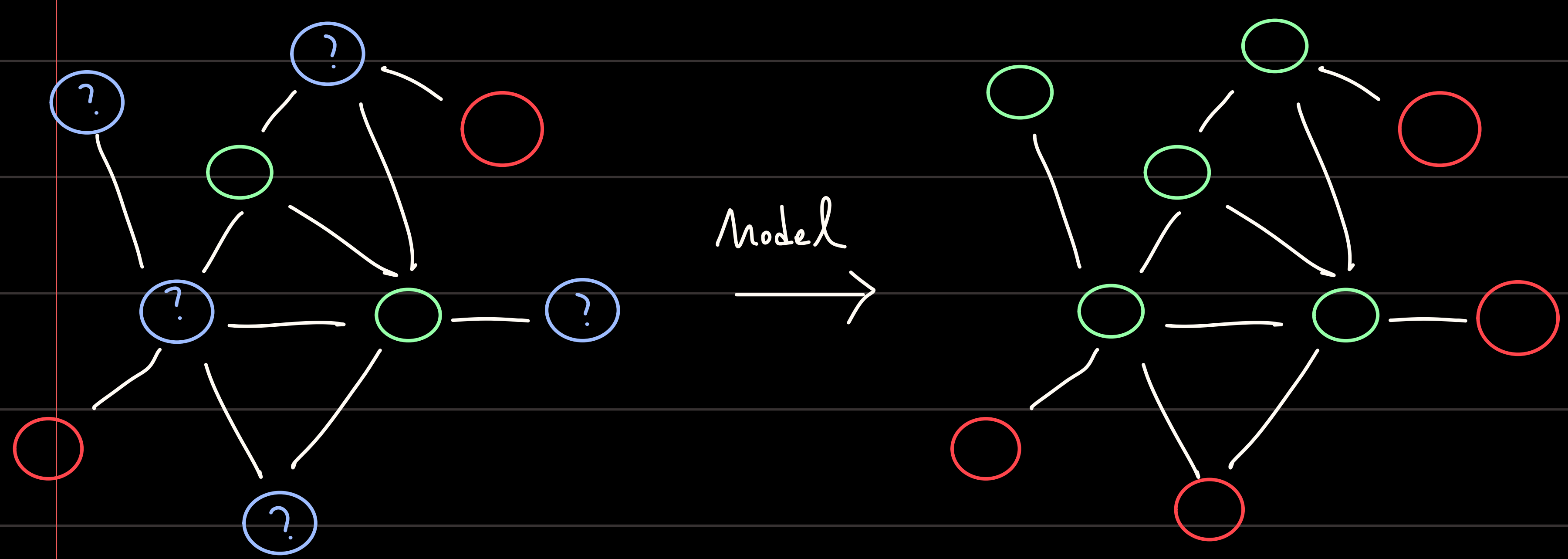
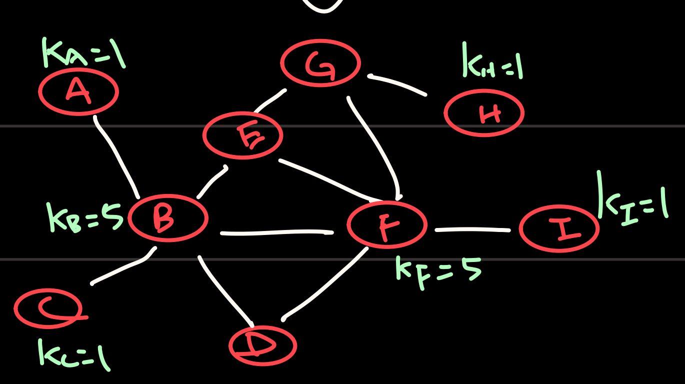
Node-Level Features

Node Classifaction → Condisider 2 labels : {◯, ◯}



Charactirize Node Features : Node Degree · Node Cenerality
Clusturing Coefficient · Graphlets

① Node Degree



Weak : $k_A = k_C = k_H = k_I$
$k_B = k_F$
⇒ model can't distinguish

② Node Centrality → How Important the Node is

(Def) Eigenvector Centrality

Node v is Important if surrounding by important nodes at $N(v)$

$C_v = \frac{1}{\lambda}\sum_{u\in N(v)} C_u$, $\lambda > 0$ iff $\lambda c = Ac$, where A: Adjacency Matrix
C: Centrality Vector

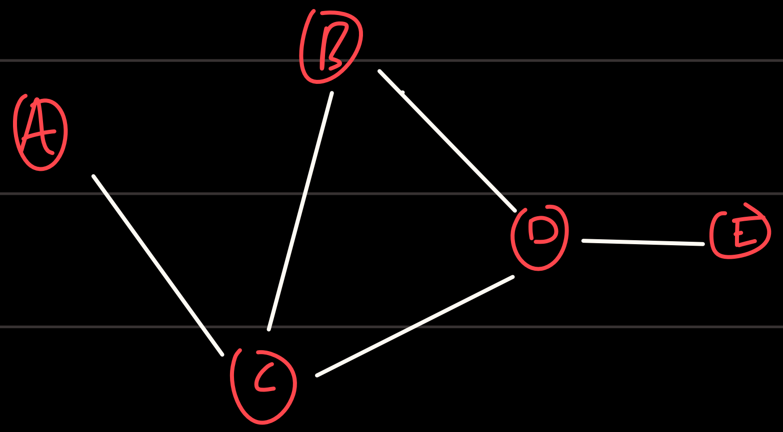
By Perron-Frobenius Thm $\rightarrow \lambda_{max} > 0$ and unique

$\rightarrow C_{max}$ is used for Centrality

Def Betweeness Centrality

Node v is important if lies on many shortest path between other nodes

$$C_v = \sum_{s \neq v \neq t} \frac{\text{Count: shortest paths between } s \text{ and } t \text{ that contains } v}{\text{Count: shortest paths between } s \text{ and } t}$$



$C_A = C_B = C_E = 0$

$C_C = 3 \rightarrow \overset{A\to B}{ACB},\ \overset{A\to D}{ACD},\ \overset{A\to E}{ACDE}$

$C_D = 3 \rightarrow \overset{B\to E}{BDE},\ \overset{C\to E}{CDE},\ \overset{A\to E}{ACDE}$

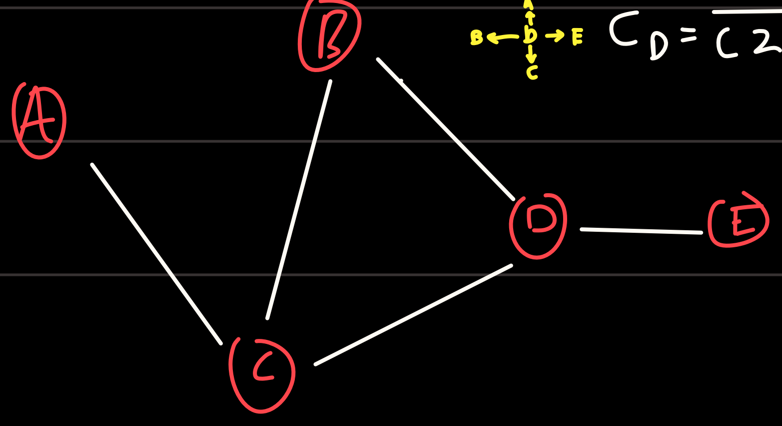
Def Closeness Centrality

Node v is important if it has small shortest path length to all other nodes

$$C_v = \frac{1}{\sum_{u \neq v} \text{shortest path length between } u \text{ and } v}$$

$C_A = \frac{1}{(2+1+2+3)} = \frac{1}{8} \rightarrow (ACB,\ AC,\ ACD,\ ACDE)$

$C_D = \frac{1}{(2+1+1+1)} = \frac{1}{5} \rightarrow (DCA,\ DB,\ DC,\ DE)$

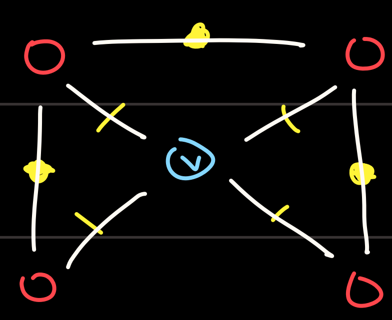


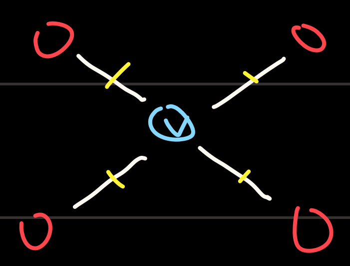
Importance : $C_D > C_A$

Def Clustering Coefficient

Measures how connected v's neighboring nodes

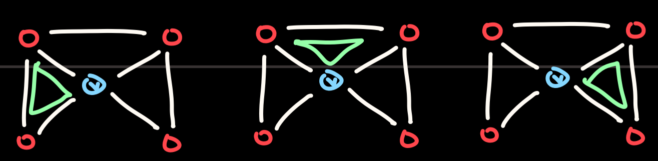
$$e_v = \frac{\text{Count: edges among neighboring nodes}}{\binom{k_v}{2}} \in [0, 1]$$





$$e_v = \frac{1}{\binom{4}{2}} \cdot 3 = \frac{1}{2}$$

$$e_v = \frac{1}{\binom{4}{2}} \cdot 0 = 0$$

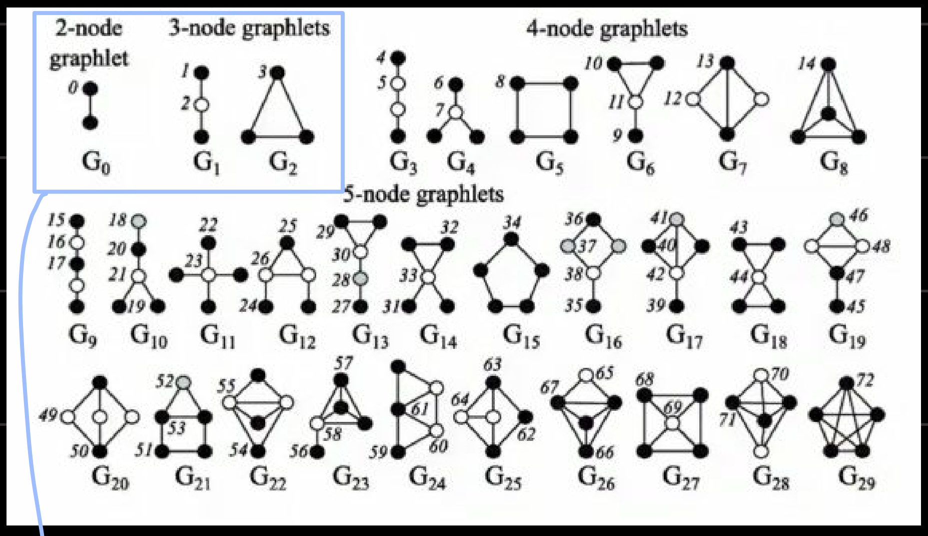


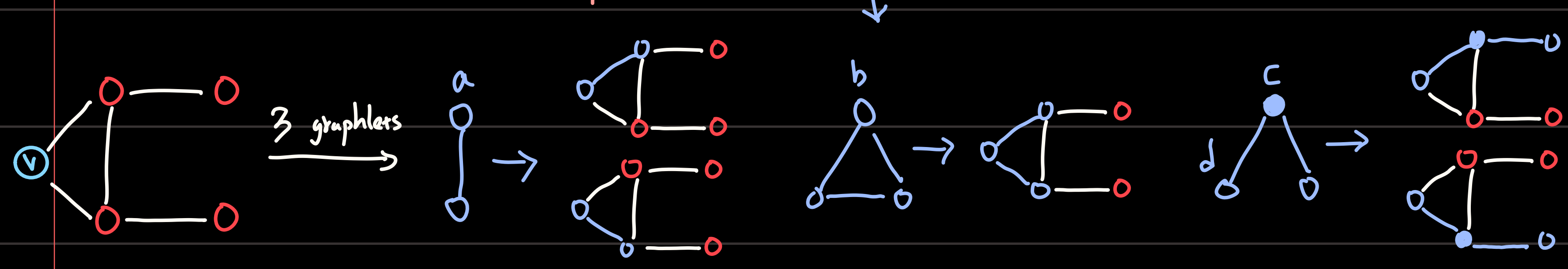
Social Network → friend of friend also my friend

Def Graphlets → Rooted Connected None-Isomorphic Subgraphs

Graphlet Degree Vector (GDV): Graphlet-base feature for node
(A count vector of graphlets rooted at a given node)

- Degree counts that a Node touches Edges
- Clustering counts that a Node touches Triangles
- GDV counts that a Node touches Graphlets





GDV of Node v on 2 to 3 nodes: $(a, b, c, d) \longrightarrow (2, 1, 0, 2)$

If on 2 to 5 → $GDV \in V^{1\times 73}$

GDV provide a measure of a Node's Local Network Topology

Comparing GDV of 2 Nodes provides a more detailed measure of local topological similarity than Node Degrees or Clustering Coefficient