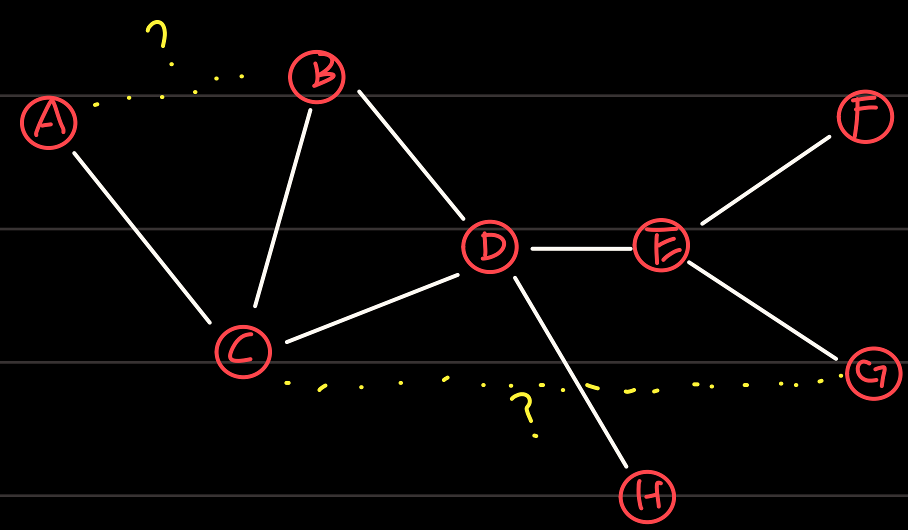
Link - Level Features



Key : Design features for pairs of nodes

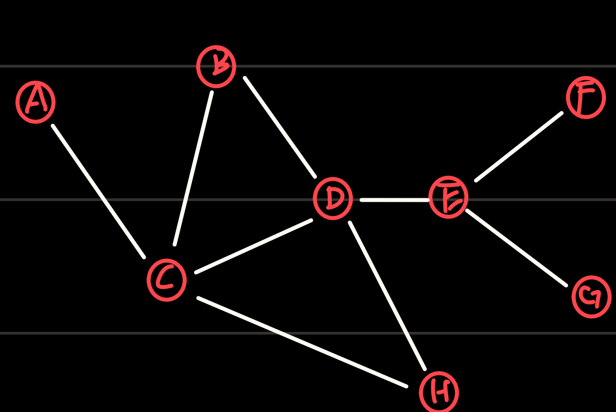
Formulations for link predition :

① Random missing links : Remove links randomly and try to predict them

② links over time :

$G[t_0, t_0']$ $G[t_1, t_1']$

① Distance - Based Feature

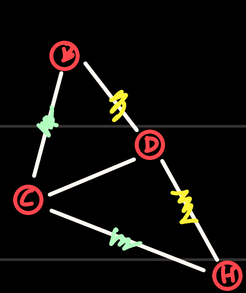


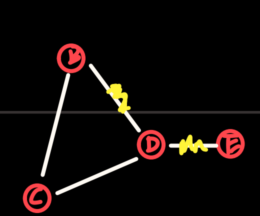
shortest distance from Node1 to Node2

$S_{BH} = S_{BE} = S_{AB} = 2$

$S_{BG} = S_{BF} = 3$

Weak : not capture the degree of neighborhood overlap

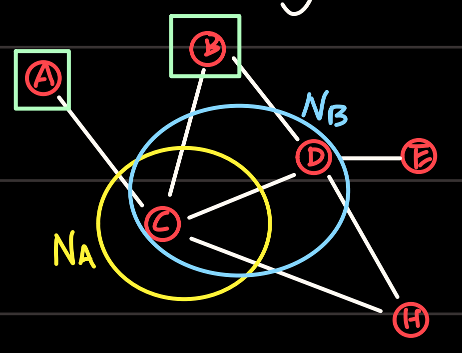




S_{BH} : 2 paths S_{BE} : 1 paths

② Neighborhood Overlaps Feature

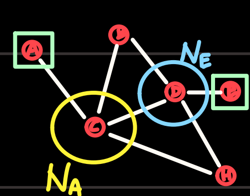
Def Local Neighborhood Feature



Capture num of neighboring nodes shared between nodes $\boxed{V_1}$ and $\boxed{V_2}$

- Common Neighbors: $|N(A) \cap N(B)| = |\{C\}| = 1$
- Jaccard's coefficient: $\dfrac{|N(A) \cap N(B)|}{|N(A) \cup N(B)|} = \dfrac{|\{C\}|}{|\{C,D\}|} = \dfrac{1}{2}$
- Adamic-Adar Index: $\displaystyle\sum_{u \in N(A) \cap N(B)} \frac{1}{\log(k_u)} = \frac{1}{\log(k_c)} = \frac{1}{\log 4}$

limitation $\longrightarrow$



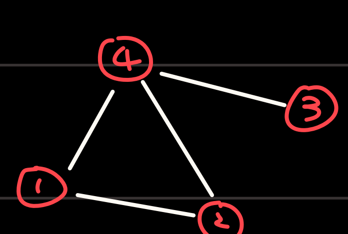
$N_A \cap N_E = \varnothing$

$|N_A \cap N_E| = 0$

Def Global Neighborhood Feature

Katz Index: Count the num of the path of all length between a given pair of nodes

(Use Power Of Adjacency Matrix to compute num of paths between 2 nodes)



$P^{(1)}_{12} = A_{12}$

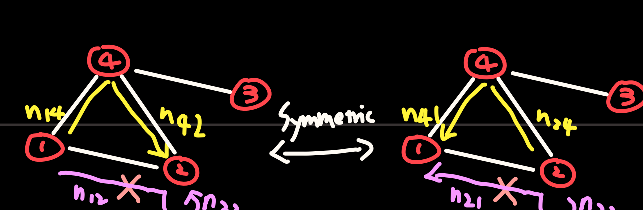
$$A = \begin{pmatrix} 0 & \boxed{1} & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}, \quad A_{uv} = 1 \text{ if } u \in N(v)$$

Let $P^{(K)}_{uv}$ denote num of path length is K between node u and v

Property : $P^{(K)} = A^K$

$P^{(2)}_{12} = A^2_{12}$

$$A^2 = A \cdot A^T = \begin{array}{cccc} n_{11} & n_{12} & n_{13} & n_{14} \end{array} \begin{pmatrix} 0 & \boxed{1} & 0 & \boxed{1} \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & n_{12}\,1 & 0 & 1 \\ 1 & n_{22}\,\boxed{0} & 0 & 1 \\ 0 & n_{32}\,0 & 0 & 1 \\ 1 & n_{42}\,\boxed{1} & 1 & 0 \end{pmatrix} = \begin{pmatrix} 2 & \boxed{1} & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 3 \end{pmatrix}$$



Katz Index between node v_1 and v_2 :

$$S_{v_1 v_2} = \sum_{l=1}^{\infty} \beta^l A^l_{v_1 v_2}, \quad \beta = \text{discount factor} \in (0, 1)$$

$\downarrow$ closed form

$$S = (I - \beta A)^{-1} - I$$

Proof $\displaystyle\sum_{l=1}^{\infty} \beta^l A^l = (I - \beta A)^{-1} - I$

$S = \beta A + \beta^2 A^2 + \beta^3 A^3 \ldots$

$(I - \beta A)(I + S)$

$= (I - \beta A)(I + \beta A + \beta^2 A^2 + \beta^3 A^3 + \ldots)$

$= (I + \beta A + \beta^2 A^2 + \ldots) - (\beta A + \beta^2 A^2 + \ldots)$

$= I$

$\rightarrow I + S = (I - \beta A)^{-1}$

$\rightarrow S = (I - \beta A)^{-1} - I$