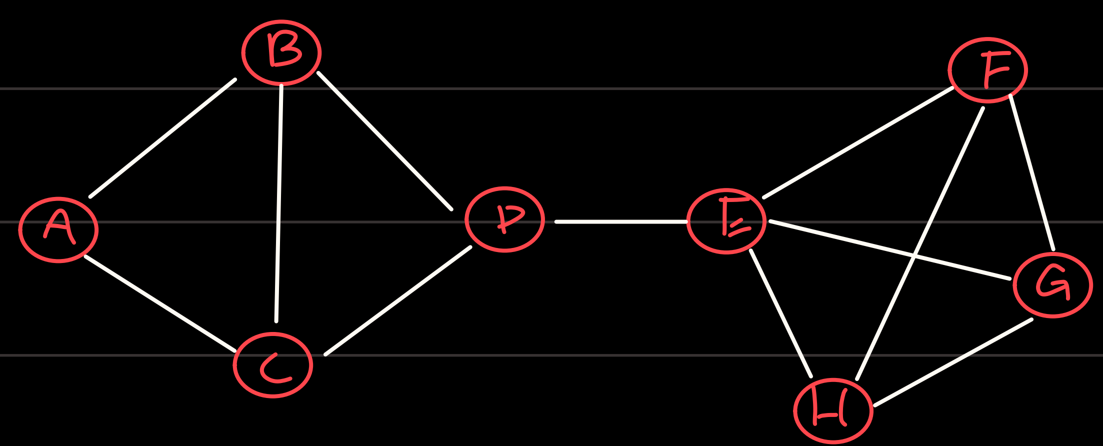
Graph Level Feature



How do we describe entire Graph Feature

Def Kernels (e.g. kernel SVM · kernel PCA)

G : Graph

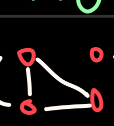
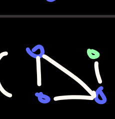
Kernel : $K(G, G') \in \mathbb{R}$ measure Similarity of 2 graphs

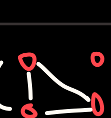
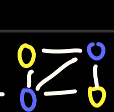
Kernel Matrix : $K = (K(G, G'))_{G, G'}$, pos-semi define : ① pos eigenvalue ② symmetric

$\exists\, \phi$ (feature Rep) s.t $K(G, G') = \phi(G)^T \phi(G')$

Key Design : Bag of words $\longrightarrow$ Bag of node degrees

o : 1 Degree o : 2 Degree o : 3 Degree

$\phi($  $) = \phi($  $) = (1, 3, 0)$

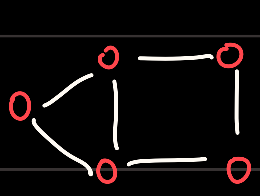
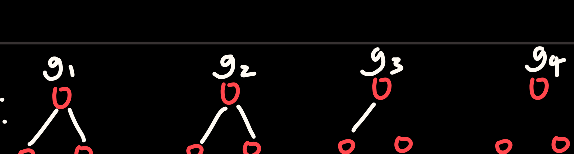
$\phi($  $) = \phi($  $) = (0, 2, 2)$

Def Graphlet Feature (Not Node-Level Graphlet Feature)

$g_k = (g_1, \ldots, g_{n_k})$: a list of graphlets of size k

Given a graph G, graphlet list g_k, define the graphlet count $f_G \in \mathbb{R}^{n_k}$

$(f_G)_i = \mathrm{Count}(g_i \subseteq G)\ \forall i = 1, \ldots, n_k$

G :  Graphlet $k=3$: 

$f_G = (1, 3, 6, 0)$

Graphlet Kernel: $K(G,G') = f_G^T f_{G'}$

$\xrightarrow[\text{skew}]{\text{prevent}}$ normalize: $h_G = \frac{f_G}{\sum_i f_G^{(i)}}$, $K(G,G') = h_G^T h_{G'}$

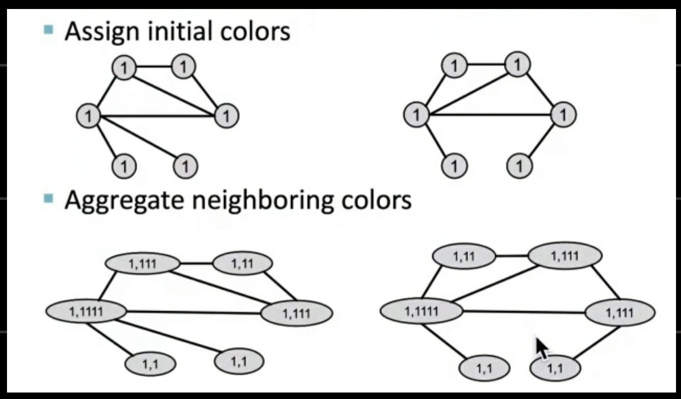
limitation: Computation Expensive (NP-Hard, $O(nd^{k-1})$; graph's node degrees bdd by d)

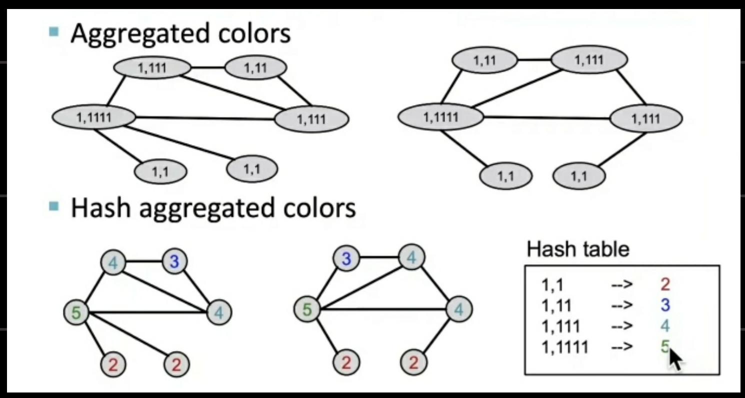
Def Weisfeiler-Lehman Kernel

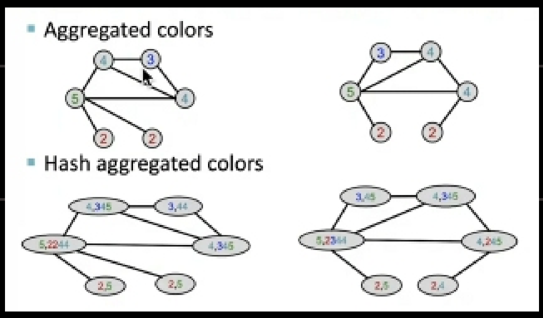
Color Refinement Algorithm: use neighborhood structure to iteratively enrich node vocabulary

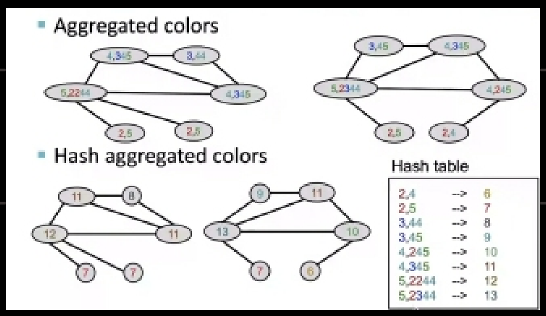
> Given a graph G with a set of nodes V
>
> Init color $c^{(0)}(v)$ to each node v in V
>
> Iterate: $c^{(k+1)}(v) = \text{HASH}(\{c^{(k)}(v), \{c^{(k)}(u)\}_{u \in N(v)}\})$
>
> node v's color
>
> v's neighbors' colors
>
> Map: v → color
>
> $\xrightarrow{\text{K steps}}$ $c^{(K)}(v)$ summarizes the structure of K-hop neighborhood

Init









the count of the color appeared in all iter step