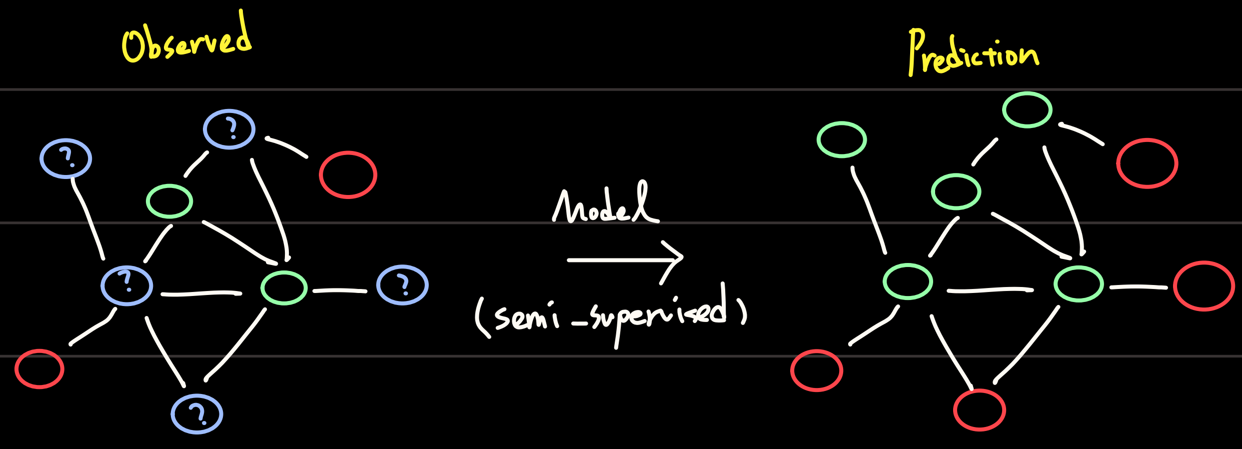
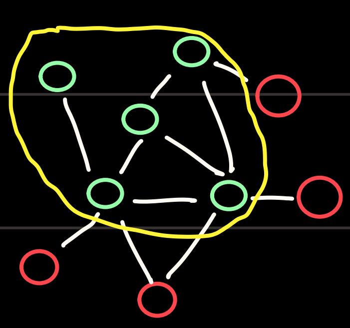
Message Passing and Node Classification

Question: How do we know labels of all other nodes



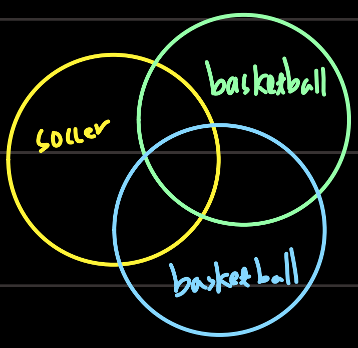
We can use correlation exist in networks (collective classification)

Correlation: nearby nodes have same class



Phenomenon:

Homophly: The tendency of individuals to associate and bond with similar others



Influence: Social connections can influence the individual characteristics of a person

Property

Classification label of a node v in network may depend on:

① Features of v

② Labels of the nodes in v's neighborhood

③ Features of the nodes in v's neighborhood

KNN

Semi-Supervised Learning:

Given $G = G(V, E)$. Few Node Labels → Plz predict class for unlabeled nodes in G

Assump G exist Homophly property (Markov property)

e.g. Binary Classification:

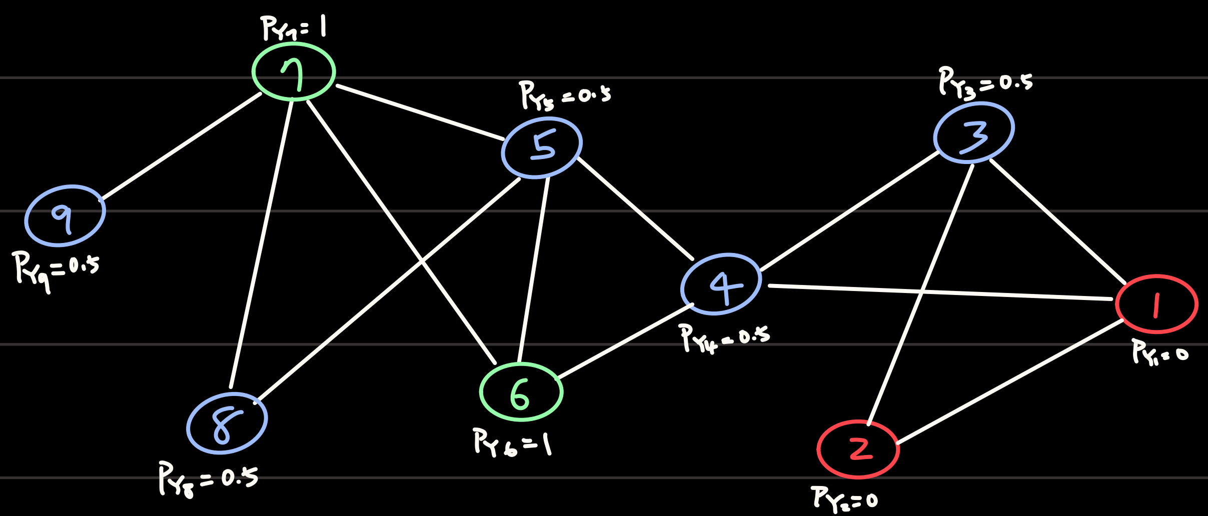
$G = G(V, E)$, $|V| = m$, Adjacency matrix $A \in \{0,1\}^{n \times n}$ with labels $Y = \{0,1\}^n$, where $m > n$

（根本是K-Nearest Neighbors）

Def Relational Classifiers (v) → $P(Y_v) = ?$

For some class c, $P(Y_v=c)$ = weighted average of $P(Y_u=c)$, $u \in N(v)$

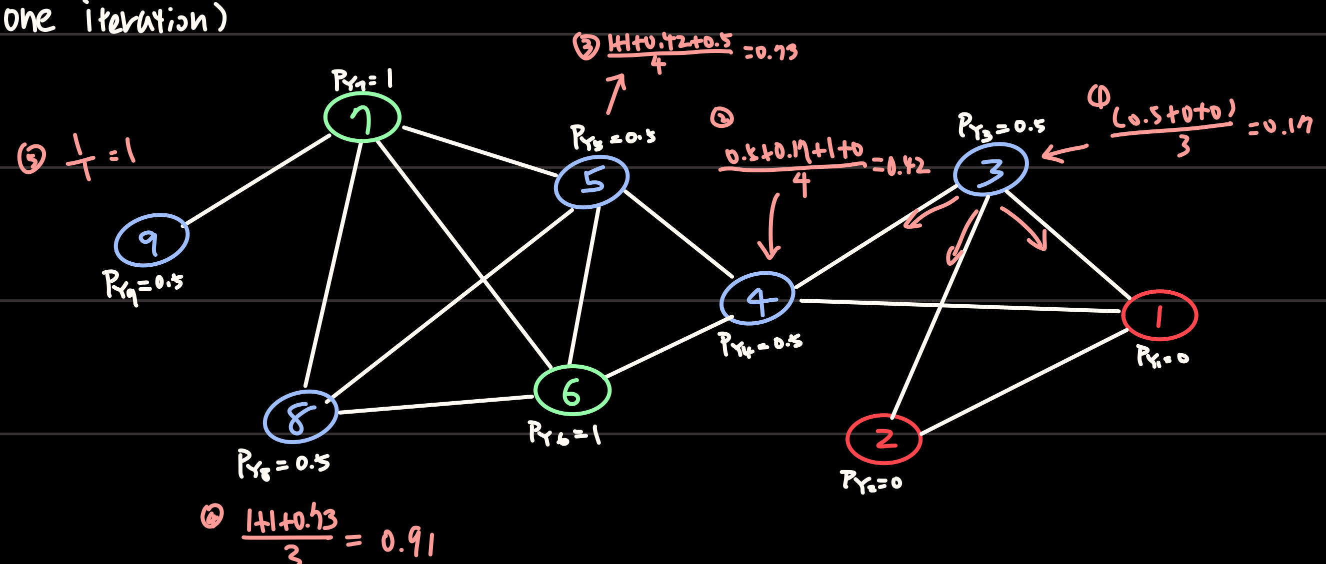
Init: labeled: $Y_v = 1$ (v_1) unlabeled: $Y_v = 0.5$ (v)



Update:

$$P(Y_v=c) = \frac{1}{\sum_{(v,u)\in E} A_{v,u}} \sum_{(v,u)\in E} A_{v,u} P(Y_u=c)$$, where $A_{v,u}$ is edge weight of (v,u)

(one iteration)



⟶ After Convergence : $\begin{cases} 1 & \text{if } P(Y_v) > 0.5 \\ 0 & \text{o.w} \end{cases}$

Challenges: ① Convergence not guaranteed

② Model cannot use node feature information

Def Iterative Classification

Given $G=G(V,E)$, some of nodes with labels, each $v \in V$ have feature f_v

Train 2 Classifier : ① $\phi_1(f_v)$: predict node label based on node feature vector f_v

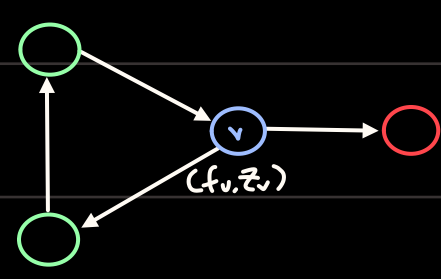
② $\phi_2(f_v, z_v)$: predict label based on feature vector f_v and Summary z_v of labels of v's neighbors

the Summary z_v:

① Histogram of the number (or fraction) of each label in $N(v)$

② Most common label in $N(v)$

③ Number of different labels in $N(v)$



Algorithm : $G=G(V,E)$, some $v \in V$ have labels

Step 1 : classify based on node attributes alone

train 2 clf e.g. SVM · RFs

① $\phi_1(f_v)$ predict Y_v based on f_v

② $\phi_2(f_v, z_v)$ to predict Y_v based on f_v and z_v

Step 2 Iterate till convergence

Test unknow: $\phi_1(f_v) = Y_v$, compute $z_v \rightarrow \phi_2 = Y_v$

Iterate $v \in V$ {

Update z_v based on Y_u $\forall u \in N(v)$

Update Y_v based on the new $z_v(\phi_2)$ } until ① class label not changed (Convergence not guaranteed) ② reach max iteration

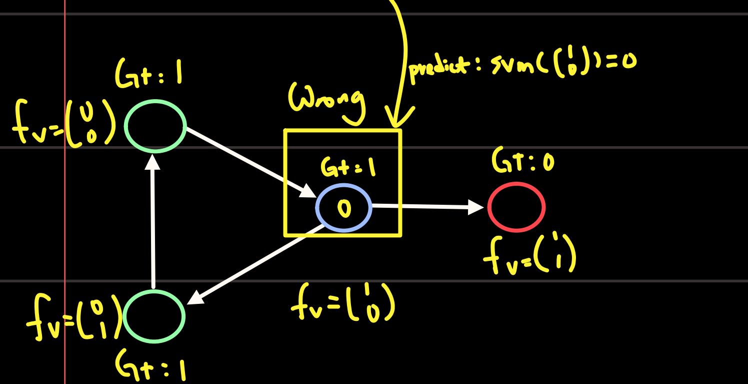
Example : Web Page Classification

Given
Input : $G=G(V,E)$, V: Web Pages E: Hyper-link between Web pages (directed)
Node features : Webpage description (for simplicity, only consider 2 binary features)

→ Plz predict the topic of the webpage

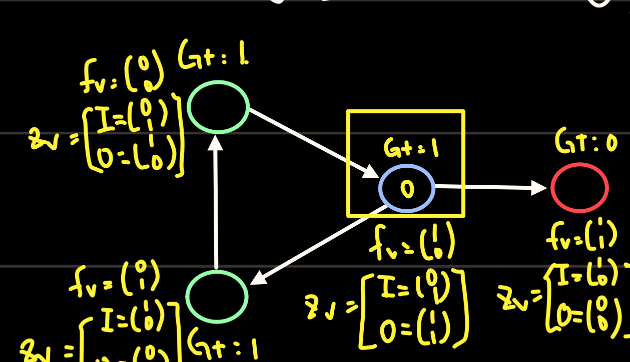
① $\phi_1(f_v)$ predict Y_v based on f_v

train: $\phi_1 = svm([\binom{0}{0}, \binom{0}{1}, \binom{1}{1}]) \leftrightarrow [1, 1, 0]$



② $\phi_2(f_v, z_v)$ to predict Y_v based on f_v and z_v

$z_v \rightarrow$ I: incoming neighbor feature, O: outgoing neighbor feature



Rules: $I_0=1$ if one of incoming labeled 0
$I_1=1$, " "
$O_0=1$, " outgoing
$O_1=1$, " "

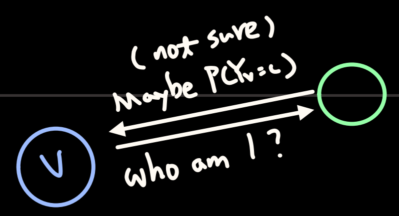
train: $\phi_2 = svm(\{(f_u, I, O) \mid \forall u \in N(v)\})$ $\forall v \in V$

Test : $\overset{\text{update}}{z_v} = [I, O]$ Iterate $\phi_2(\{(f_u, I, O) \mid \forall u \in N(v)\}) = Y_v$

(Dynamic Programming)

Def Collective Classification : Belief Propagation

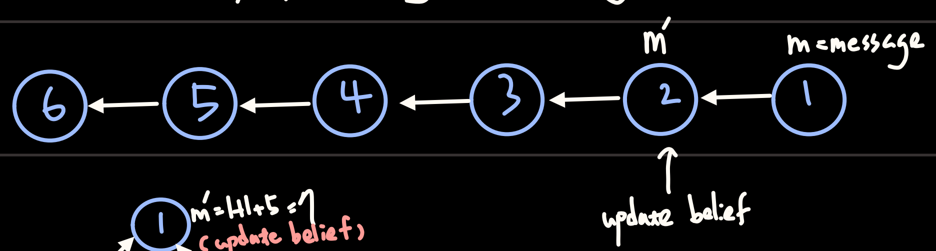
Idea :



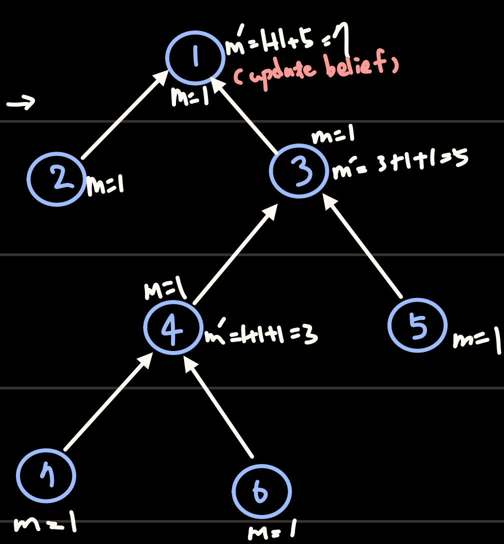
Message Passing Concept:

Condition: Each node can only pass message to its neighbors (Conclude recived from in-link neighbors and pass to out-link)

Ordered nodes :



e.g. Sum Tree →



Notation :

Label-label potential matrix ψ : $\psi(Y_i, Y_j) = P(\text{node } j \text{ is being in class } Y_j \mid j\text{'s neighbor } i \text{ in class } Y_i)$

Prior belief ϕ : $\phi(Y_i) = P(\text{node } i \text{ being in class } Y_i)$

$m_{i\to j}(Y_j)$: i's message (estimation) of j being in class Y_j

L : the set of all labels

Algorithm :

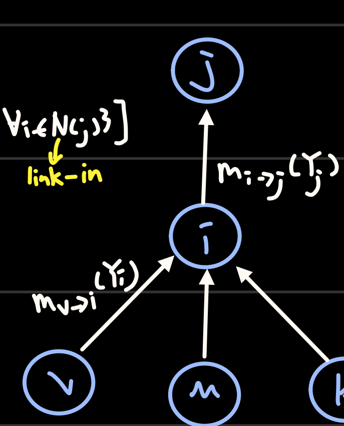
Given $G = G(V, E)$ directed

Init : all message = 1 (Prior)

Iter each node $\in V$ $\left\{ m_{i\to j}(Y_j) = \sum_{Y_i\in L} \underbrace{\psi(Y_i,Y_j)\cdot\phi(Y_i)}_{P(Y_j|Y_i)\cdot P(Y_i) = P(Y_i,Y_j)} \prod_{k\in N(i)/j} m_{k\to i}(Y_i) \quad \forall Y_j \in L \right\} = E[P(Y_j) \mid Y_i \mid \forall i \in N(j)\}]$

link-in

only link-in neighbors



Convergence $\longrightarrow$ belief of node i being in class Y_i : $b_i(Y_i) = \phi_i(Y_i) \prod_{j\in N(i)} m_{j\to i}(Y_j)$

Very Similar to Bellman Equation (But focus on j's link-out neighbors)

$\forall s_i \in N(s_j)$ (link-out), State S_j take an action to the state S_i with policy π, and define the value function $V_\pi(S_j)$

$$V_\pi(S_j) = E_\pi\left[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \mid S_t = S_j\right] = \sum_a \pi(a|s_j) \sum_{s_i\in N(s_j)} \sum_r p(s_i, r|s_j, a)[r + \gamma v_\pi(s_i)]$$

(link-out)