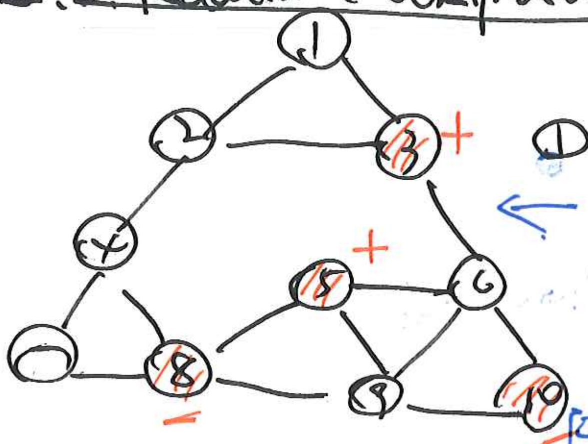
HW2. Q1.

1.1 Relational classification

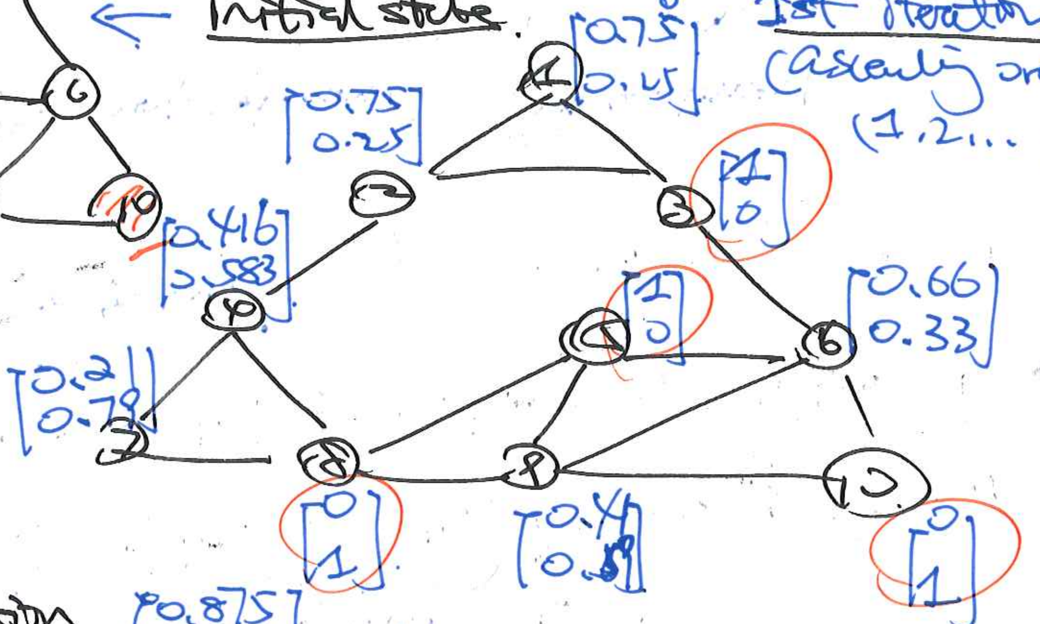
i) ∘ uniformed edge weight $W=1$.

∘ $P(Y_i=c) = \frac{1}{|N_i|}\sum_{(i,j)\in E} W(i,j)\, P(Y_j=c)$

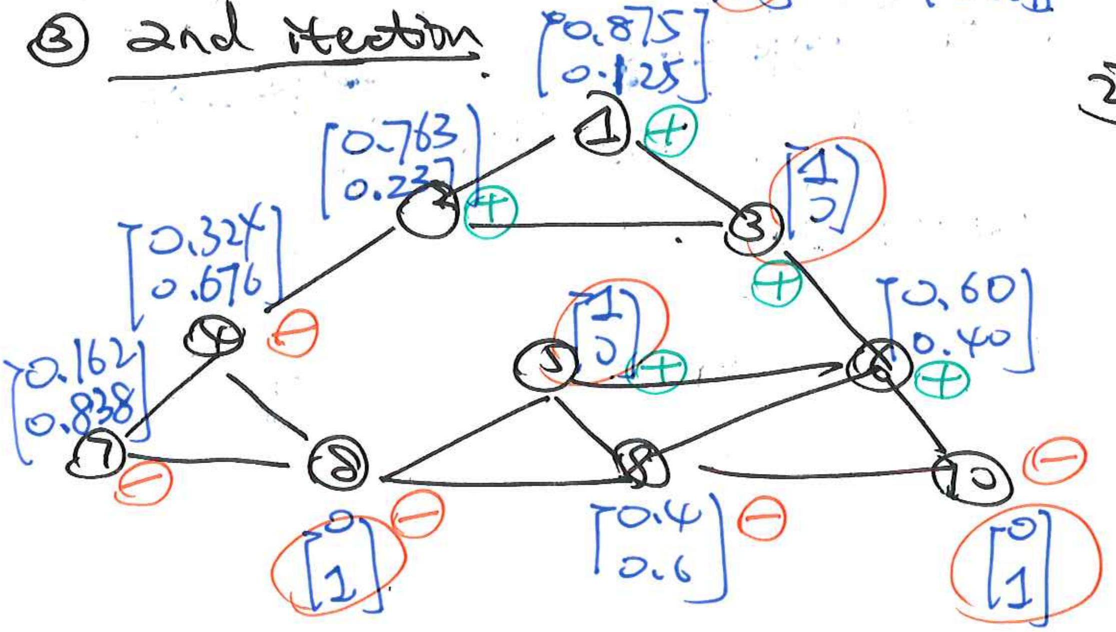
① Initial state ←



② 1st iteration (ascending order) (1, 2, ...)



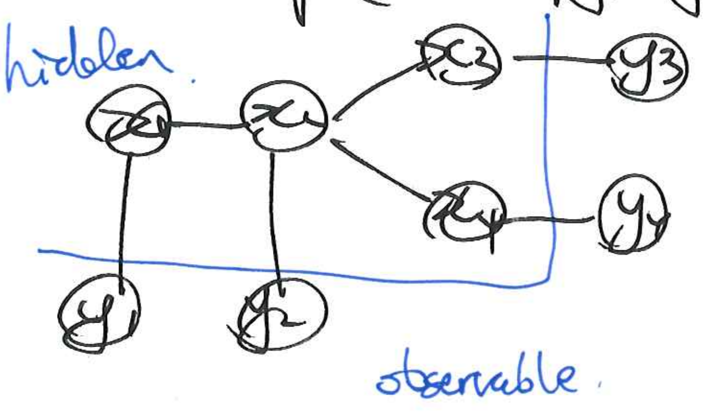
③ 2nd iteration



ii) use 0.5 as threshold. we classify the nodes ↵

1.2. Belief Propagation (BP)

∘ CRF: $p(x_1 \dots x_n | y_1 \dots y_m) = \left(\prod_{c \in C} \phi_c(x_c, y_c)\right) \frac{1}{Z}$



∘ given the observation of $(y_1 \dots y_4)$
we want to analysis the property of node hidden.

i). belief of node 1

$$b_1(x_1) = \frac{1}{Z}\phi_1(x_1)\, m_{21}(x_1)$$

$$= \frac{1}{Z}\underbrace{\phi_1(x_1,y_1)}_{\text{message from } y_1}\underbrace{m_{21}(x_1)}_{\text{message from } x_2}$$

ii) $p(x_1|y_2,y_3,y_4,y_1)$

$$= p(x_1|y_1)\, \overbrace{p(x_1|x_2)\, p(x_2|y_2\, y_3, y_4)}^{\text{edge 1-2}}$$

$$= p(x_1|y_1)\, p(x_1|x_2)\, p(x_2|y_2)\, \underbrace{p(x_2|x_3)\, p(x_3|y_3)}_{\text{edge 2-3}}\, \underbrace{p(x_2|x_4)\, p(x_4|y_4)}_{\text{edge 2-4}}$$

$$= p(x_1|y_1)\, p(x_1|x_2)\, m_{32}\cdot m_{42}\cdot \phi(x_2)\, \frac{1}{Z}$$

$$= p(x_1|y_1)\, m_{21}(x_1)\, \frac{1}{Z}$$

$$= \frac{1}{Z}\phi_1(x_1,y_1)\, m_{21}(x_1) = b_1(x_1) \qquad \square$$

iii). 1 \ \ \

[HW2, Q2]: Node embedding with TransE

- multi-relational graph. $G = (E, S, L)$
  - $S = \{(h, l, t)\}$ edges.

head. relationship (called here) tail.

f: • E: entities (nodes) $\xrightarrow{\text{into}} \mathbb{R}^k$ (embedding)

- $f(h) + f(l) = f(t)$.

2.1. Warmup: Why the comparative loss?

$$\mathcal{L}_{simple} = \sum_{(h,l,t) \in S} \cancel{d(h+l, t)}\ d(f(h) + f(l), f(t))$$

→ will lead to a useless embedding

- for an example. take $f(l) = 0$, $f(h) = f(t) =$ const. then we will have $\mathcal{L}_s = 0$.

2.2. The Purpose of the Margin. γ:

$$\mathcal{L}_{no\ margin} = \sum_{(h,l,t) \in S} \sum_{(h',l',t') \in S'} \left[d(f(h) + f(l), f(t)) - d(f(h') + f(l'), f(t')) \right]_+$$

→ lead to useless embedding.

- $f(l) = 0$, $f(h) = f(t) =$ const

  → $\mathcal{L} = 0$

- ADD Margin γ. will force the embedding to diverge.

2.3. Why are Entity Embeddings Normalized?

- TransE normalises every entity embedding to have unit length.

  → as we are calculating distance

- if not normalized, we will simply have embeddings all converges to zero.

2.4. Where TransE fails.

for graph without perfect embedding

i.e. no embedding perfectly satisfies $u + l = v$

for all $(u, l, v) \in S$, and $u + l \neq v$ for $(u, l, v) \notin S$.

HW2. Q3

3. GNN Expressiveness

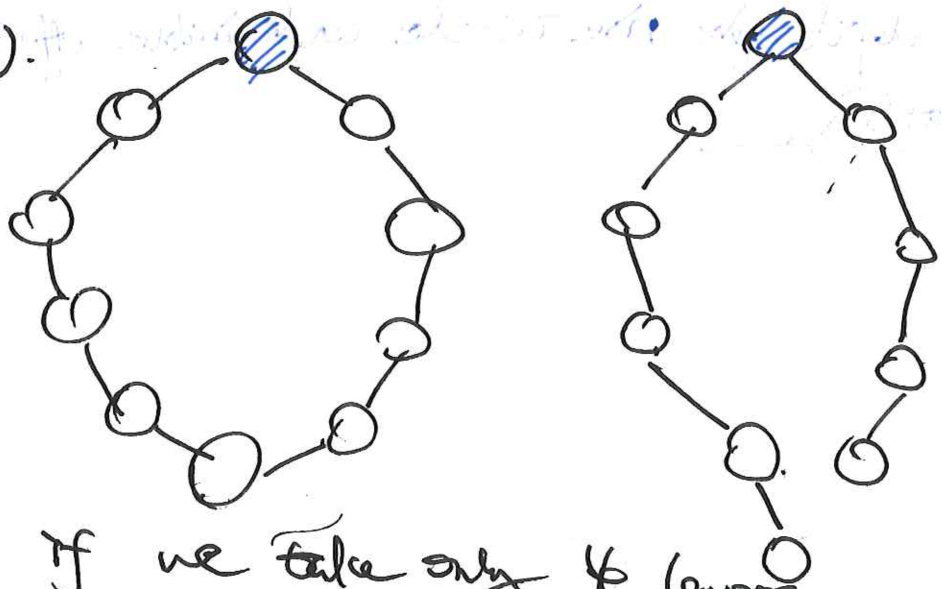
$$h^{k+1} = \sigma(D^{-1/2} A D^{-1/2} h^k W^k)$$

3.1 effect of depth on expressiveness

- $d=1$.

i). need 3 layers to distinguish.

ii).



increasing depth could give more expressive power.

- If we take only 4 layers, we can not distinguish the upper two blue nodes.
- no. GNN with fewer than 5 layers can perfectly perform this task.

2.

3.2. Relation to Random walk

i). message passing & random walk.

- take $\sigma(h)=h$, and average aggregator:

$$h_i^{(k+1)} = \frac{1}{|N_i|} \sum_{j \in N_i} h_j^{(k)}$$

so equivalent to take step 1 random-walk

$$H^{k+1} = D^{-1} A H^k$$

2). $$h_i^{(l+1)} = \frac{1}{2} h_i^{(l)} + \frac{1}{2} \frac{1}{|N_i|} \sum_{j \in N_i} h_j^{(l)}$$

$$H^{l+1} = \frac{1}{2}(H^l + D^{-1}AH^l)$$

$$= \frac{1}{2}(I + D^{-1}A) H^l.$$

(o note take $D^{-1} = \mathrm{diag}(P_{ii}^{-1})$ $D^{-1}A = D^{-1/2}AD^{-1/2}$)

3.3. Over-Smoothing Effect.

2) very large depth give rise to the undesirable effect of oversmoothing.

eg. take $h_i^{(l+1)} = \frac{1}{|N_i|} \sum_{j \in N_i} h_j^{(l)}$

. assume the graph is connected & no bipartite components.

o take $l \to \infty$.

$$H^{l+1} = D^{-1}A\, H^l.$$

$$= (D^{-1}A)^l H^0. = D^{-l} A^l H^0$$

o if $l \to \infty$.

each node takes its neighbor to average.

• it will spread to the whole graph.

→ it will similar to some sort of average of the whole graph. → smooth effect

Hint: prove using Markov chain

3.4. Learning BFS with GNN

breadth-first search.

→ fcn that determines whether a node is reachable from source at step t. given its previous reachability and its neighbor's previous reachability

~ water flow