

(i) The belief at node 1 is:

$$b_1(x_1) = \frac{1}{Z} \phi_1(x_1) m_{2,1}(x_1) = \frac{1}{Z} \phi_1(x_1) \sum_{x_2} \phi_2(x_2) \psi_{2,1}(x_2, x_1) m_{3,2}(x_2) m_{4,2}(x_2) =$$

$$\frac{1}{Z} \phi_1(x_1) \sum_{x_2, x_3, x_4} \phi_2(x_2) \phi_3(x_4) \phi_4(x_4) \psi_{2,1}(x_2, x_1) \psi_{3,2}(x_3, x_2) \psi_{4,2}(x_4, x_2)$$

where,

$$\phi_i(x_i) = \phi_i(x_i, y_i)$$

and Z denotes the normalizing constant that ensures elements in $b_1(x_1)$ sum to 1.

(ii)

$$P(x_1|\mathbf{y}) = \sum_{x_2, x_3, x_4} P(\mathbf{x}|\mathbf{y}) = \frac{1}{Z} \sum_{x_2, x_3, x_4} \prod_{i,k \text{ adjacent}} \psi_{i,k}(x_i, x_k) \prod_{j=1}^4 \phi_j(x_j) =$$

$$\frac{1}{Z} \sum_{x_2, x_3, x_4} \psi_{2,1}(x_2, x_1) \psi_{3,2}(x_3, x_2) \psi_{4,2}(x_4, x_2) \phi_1(x_1) \phi_2(x_2) \phi_3(x_3) \phi_4(x_4) =$$

$$b_1(x_1)$$

Question 2.1, Homework 1, CS224W

An example for a simple graph for which when minimizing the following loss, $\mathcal{L}_{simple} = \sum_{(h,\ell,t) \in S} d(\mathbf{h} + \ell, t)$ is insufficient is a star graph. For example the center node's embedding will be \mathbf{h}_{center} , the i^{th} satellite node embedding will be $\mathbf{h}_{sat}^{(i)}$ and the relation's embedding is ℓ . Therefore, in order to minimize the loss (to 0) $\mathbf{h}_{sat}^{(1)} = \mathbf{h}_{sat}^{(2)} = \dots = \mathbf{h}_{sat}^{(k)}$ for k satellites. Thus, we can't distinguish between the satellites nodes.

When $d(\mathbf{h}' + \ell, \mathbf{t}') > d(\mathbf{h} + \ell, \mathbf{t})$ for all the entities.

Set $d(\mathbf{h}' + \ell, \mathbf{t}') - d(\mathbf{h} + \ell, \mathbf{t}) > \gamma$

Question 2.4, Homework 1, CS224W

When two entities has the same relation (such as siblings in a family tree) than $\mathbf{h} = \mathbf{t}$ and $\ell = 0$ which is meaningless. Another problem is many-to-one (or one-to-many) as in Question 2.1.

- (i) Given these 2 graphs, since the 2^{nd} hop neighbors are different, a GNN with 3 message passing layers can distinguish them.
- (ii) As we have learn, every layer in a GNN can capture information of the local K -hop neighborhood for a given route. Since a cyclic subgraph fo length 10 can be model by minimum two 5-hop routes, this task can perform perfectly with at least 5 GNN layers.

- (i) The transition matrix of a 1-step uniform random walk, $h_i^{(l+1)} = \frac{1}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} h_j^{(l)}$, is

$$P = D^{-1}A$$

since the transition from node u to node v is $P_{uv} = \Pr(\text{going from } u \text{ to } v \mid \text{we are at } u) = \frac{1}{d_u}$ if $(u, v) \in E$ (and 0 otherwise).

- (ii) In the case $h_i^{(l+1)} = \frac{1}{2}h_i^{(l)} + \frac{1}{2}\frac{1}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} h_j^{(l)}$ the corresponding transition matrix of a uniform random walk is

$$P = \frac{1}{2}I + \frac{1}{2}D^{-1}A$$

- (i) The standard random walk, $h_i^{(l+1)} = \frac{1}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} h_j^{(l)}$, is a Markov chain with a transition matrix P as in Question 3.2. Then, starting at embedding for node i is $h_i^{(0)}$, the stationary distribution after infinite number of steps in the limit:

$$\lim_{l \rightarrow \infty} (P^l) h_i^{(0)}$$

Thus, under certain conditions (irreducible and aperiodic transition matrix) the limiting distribution is identical to one that satisfies the following:

$$\left(h_i^{(l)}\right)^* = P^T \left(h_i^{(l)}\right)^*$$

and in other words, the node embedding will converge as $l \rightarrow \infty$.

- (i) The embedding of node i when the *update rule* of BFS is applied is:

$$h_i^{(t+1)} = \begin{cases} 1, & h_i^{(t)} = 1 \\ 1, & \exists j; (j, i) \in E \wedge h_j^{(t)} = 1 \\ 0, & \text{otherwise} \end{cases}$$

where $h_i^{(1)} = 1$ only for the source node (and 0 otherwise).

- (ii) In order to learn the task perfectly, the *aggregation function* is:

$$h_{\mathcal{N}_v}^{(l)} = \max(\{h_u^{(l-1)}, \forall u \in \mathcal{N}_v\})$$

and the *message function* is:

$$h_v^{(l)} = h_v^{(l-1)} \vee h_{\mathcal{N}_v}^{(l)}$$

Information sheet

CS224W: Machine Learning with Graphs

Assignment Submission Fill in and include this information sheet with each of your assignments. This page should be the last page of your submission. Assignments are due at 11:59pm and are always due on a Thursday. All students (SCPD and non-SCPD) must submit their homework via GradeScope (<http://www.gradescope.com>). Students can typeset or scan their homework. Make sure that you answer each (sub-)question on a separate page. That is, one answer per page regardless of the answer length. Students also need to upload their code on Gradescope. Put all the code for a single question into a single file and upload it.

Late Homework Policy Each student will have a total of *two* late periods. *Homework are due on Thursdays at 11:59pm PT and one late period expires on the following Monday at 11:59pm PT.* Only one late period may be used for an assignment. Any homework received after 11:59pm PT on the Monday following the homework due date will receive no credit. Once these late periods are exhausted, any assignments turned in late will receive no credit.

Honor Code We strongly encourage students to form study groups. Students may discuss and work on homework problems in groups. However, each student must write down their solutions independently, i.e., each student must understand the solution well enough in order to reconstruct it by him/herself. Students should clearly mention the names of all the other students who were part of their discussion group. Using code or solutions obtained from the web (GitHub/Google/previous year's solutions etc.) is considered an honor code violation. We check all the submissions for plagiarism. We take the honor code very seriously and expect students to do the same.

Your name: _____

Email: _____ **SUID:** _____

Discussion Group: _____

I acknowledge and accept the Honor Code.

(Signed) _____