Assignment 2

Please make sure that you always use notations consistent with lecture notes. Different notations will not be accepted.

The deadline for assignment 2 is: Fri 5, August 9:00 pm (Sydney Time)

Question 1 (7 marks)

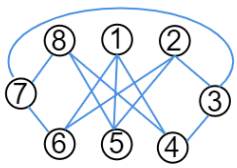


Figure 1

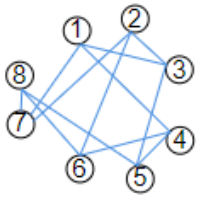


Figure 2

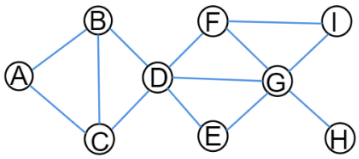


Figure 3

1) Are the graphs in Figure 1 and Figure 2 isomorphic? If so, demonstrate an isomorphism between the sets of vertices. If these two graphs are not isomorphic, prove it by finding a structure (node and/or edge) in one graph which is not present in the other. (2 marks)

2) Compute the betweenness centrality and closeness centrality of nodes 'A', 'D' and 'G' in Figure 3 step by step. (2 marks)

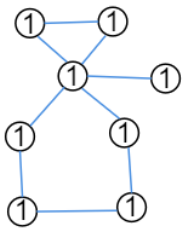


Figure 4

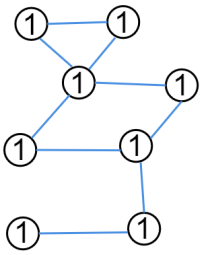


Figure 5

3) Please implement the color refinement algorithm to compute the Weisfeiler-Lehman kernel value for the graph in Figure 4 and graph in Figure 5 with 2 steps. You can only use different integer values to present different colors. Please show each intermediate graph, the color count vectors of the two graphs and resulted kernel value. (3 marks)

Question 2 (5 marks)

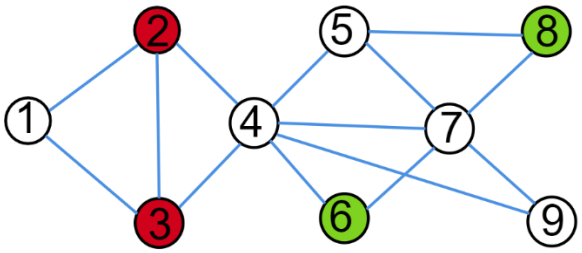


Figure 6

1) Consider the graph in Figure 6, red nodes have label 1, green nodes have label 0 and other nodes are unlabeled. Please implement the probabilistic relational classifier to update the labels of all the nodes. Can it reach the convergence? If so, how many iterations of computation does it need to confirm (the convergence)? Please show the intermediate graph of each iteration. If not, please explain why by also showing the intermediate graph of each iteration. (3 marks)

2) Consider a second order biased walk with p = 4 and q = 3 on the undirected graph with 9 nodes in Figure 6. Assume a walker just traversed edge (3, 4) and now resides at node 4. Please compute the unnormalized transition probabilities of the nodes in the table based on the distance from node 3 (fill in the blanks in the table below). Please show intermediate steps and justify your answers. (2 marks)

Target t	Probability	Dist(3, t)
3		
2		
5		
6		
7		
9		

Question 3 (4 marks)

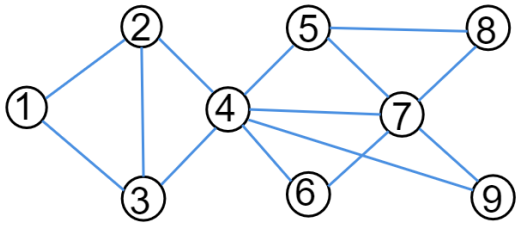


Figure 7

Consider the simple neighborhood aggregation GCN model on the above undirected graph with 8 nodes. **Please compute the output of the first graph convolutional layer(*i.e.*,H^1) step by step base on the following formula:**

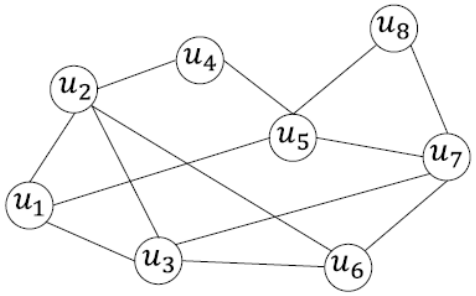
$$h_v^{(l+1)} = \sigma(W^l \sum_{u \in N_v} \frac{h_u^{(l)}}{|N(v)|} + B^l h_v^{(l)})$$

where $h_v^{(l)}$ indicates the d_l-dimensional embedding of node v in layer l, and $H^{(l)} = \left[h_{v1}^{(l)}, h_{v2}^{(l)}, \ldots, h_{v9}^{(l)}\right]$. $W^l \in \mathbb{R}^{d_{l+1} * d_l}$ denotes the weight matrix for the neighbors of v in layer l, d_l denotes the dimension of the node embedding in layer l, $N(v)$ denotes the neighbor set of the node v, $B^l \in \mathbb{R}^{d_{l+1} * d_l}$ denotes the self-looping weight matrix in layer l, σ denotes the ReLU non-linear function. The initial embeddings for all nodes are stacked in H^0. W^0 and B^0 are the weight matrices of layer 0.

$$H_0 = \begin{vmatrix} 0.30 & -0.60 & 0.10 & 0.20 \\ 0.60 & 0.30 & 0.40 & -0.10 \\ 0.20 & 0.70 & -0.40 & 0.50 \\ -0.40 & 0.60 & 0.10 & 0.80 \\ 0.40 & 0.90 & -0.20 & 0.10 \\ 0.30 & -0.30 & 0.90 & 0.70 \\ -0.20 & 0.20 & 0.70 & 0.40 \\ -0.60 & 0.50 & 0.10 & 0.70 \\ 0.20 & 0.60 & 0.70 & -0.10 \end{vmatrix}$$

$$W_0 = \begin{vmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{vmatrix} \qquad B_0 = \begin{vmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{vmatrix}$$

Question 4 (4 marks)



$$H = \begin{vmatrix} 0.73 & 0.59 & 0.46 & 0.14 \\ 0.73 & 0.12 & 0.31 & 0.68 \\ 0.64 & 0.19 & 0.40 & 0.62 \\ 0.22 & 0.29 & 0.55 & 0.72 \\ 0.04 & 0.51 & 0.80 & 0.73 \\ 0.48 & 0.34 & 0.46 & 0.50 \\ 0.50 & 0.48 & 0.71 & 0.11 \\ 0.31 & 0.72 & 0.32 & 0.51 \end{vmatrix}$$

Figure 8

Consider the graph in Figure 8. The node embeddings for all nodes are stacked in H. We use the COS-Similarity which is defined as follows to measure the similarity between two vectors $u, v \in \mathbb{R}^k$.

$$COS(u,v) = \frac{\sum_{i=1}^{k} u_i * v_i}{\sqrt{\sum_{i=1}^{k} (u_i)^2} * \sqrt{\sum_{i=1}^{k} (v_i)^2}}$$

1) Calculate the similarity of node pairs (u_4, u_5) and (u_2, u_7). (2 marks)
2) If the similarity between nodes u and v larger than a threshold t, we consider there is a link between u and v. Otherwise, there is no link between the nodes. If we set the threshold t=0.6, does the link prediction result is correct or not for node pairs (u_4, u_5) and (u_2, u_7). What if we set the threshold t=0.8? Justify your answer. (2 marks)

Assignment Submission

- Students must submit an electronic copy of their answers to the above questions to the course website in Moodle.
- Only **.doc** or **.pdf** file is accepted. The file name should be ass1_**studentID**.doc or ass1_**studentID**.pdf (e.g., **ass1_z5100000.doc** or **ass1_z5100000.pdf**).

Note:

1. For any problems in submissions, please email to comp9312unsw@gmail.com
2. All submissions will be checked for plagiarism.
3. We do not accept e-mail submissions.

Warning: Before submission, please keep a copy in your university account or other reliable cloud servers (such as dropbox or google drive). If you are not sure how, please have a look at taggi. Usually, the submission should be successful. In case it fails, we do **not** accept backups from your own computers as the modification time can be edited.

The university regards plagiarism as a form of academic misconduct and has very strict rules regarding plagiarism. For UNSW policies, penalties, and information to help avoid plagiarism, please see: https://student.unsw.edu.au/plagiarism as well as the guidelines in the online ELISE tutorials for all new UNSW students: https://subjectguides.library.unsw.edu.au/elise

Late Submission Penalty

- 5% of your final mark will be deducted for each additional day (24hr) after the specified submission time and date.
- Submissions that are more than five days late will not be marked.