CS544 2021 End Semester Quiz

There are 2 sections in this quiz. Section A contains both multiple choice as well as fill in the blanks. For the multiple choice questions, only one answer is correct. Section B is fill in the blanks questions. Strictly follow the instructions on the answer format that would be given with each question. YOU CAN SUBMIT ONLY ONCE, SO BE CAREFUL WHILE SUBMITTING.

Points:

20/20

Section A

Multiple choice questions have only one correct answer. For fill in the blanks follow the instruction in the question.

1.In random walk based shallow embeddings, if z_u and z_v denote the embeddings of nodes u and v respectively, then, transpose(z_u)z_v may represent which of these possibilities most strongly

(2/2 Points)

○ Probability that u and v are in the same connected component

○ **Probability that u and v co-occur on a same random walk over the graph**

○ Probability that u and v are directly connected by an edge

○ None of these

2.In negative sampling

(2/2 Points)

○ the probability $P(v|z_u)$ is calculated by normalizing over most probable neighbor of u

○ the probability $P(v|z_u)$ is calculated by normalizing over all neighbors of u

○ the probability $P(v|z_u)$ is calculated by normalizing over k nodes co-occuring in the random walks from u

○ **None of these**

3.Which of the following is true for neighbor sampling that is used to improve GNN scalability?

(2/2 Points)

○ Increasing the number of samples per node lead to better node aggregation but unstable training

○ Neighbor sampling reduces the size of the computational graph from exponential size to polynomial with respect to the number of layers K of the GNN

○ **Random sampling of the neighbors may be fast but selects unimportant nodes**

○ Random walk with restarts generates better samples for Poisson networks (homogeneous node degrees) as compared to scale free networks

4.Suppose a random network with 1000 nodes and 2500 edges follows a G(n,m) model. The average degree of the nodes in the network is
(2/2 Points)

5

5.In a node2vec algorithm, the values of p and q are 2 and 3, respectively. Suppose a random walker starting from u, has reached to node w through v. The distance of v and w from u is 1 and 2 respectively. There are 2 more neighbors of w which are farther away from u. The probability of jumping to v in the next hop is _____ (round to 2 decimal places)
(2/2 Points)

0.50

Section B
Questions are fill in the blank types. Write the answers in the format as instructed in each question

6.Consider a large network (n --> \infty) with an average node degree of 3, formed based on the G(n,p) model. The variance in the node degree distribution would be _____ (Round to 1 decimal, e.g., 1.3, 2.0)
(3/3 Points)

3.0

7.Consider a network of 6 nodes labeled 1 to 6. The adjacency list is given as (1: 2, 6) (2: 1, 3, 6) (3: 2, 4, 5) (4: 3, 5) (5: 3, 4) (6: 1, 2). A community detection algorithm identified 2 communities : (1, 2, 6) and (3, 4, 5). The modularity Q of the network is _____ (Round to 2 decimal places, e.g. 0.19, 0.20)
(4/4 Points)

0.36

8.For the network in Q7, we need to predict whether a node is illicit. Nodes 3, 4 and 6 are already labeled, with 3 and 4 being licit and 6 being an illicit node. We use a probabilistic relational classifier to classify the nodes. The final probability of node 2 being illicit is _____ (Round to 2 decimal places, e.g. 0.19, 0.20)
(3/3 Points)

0.60