Homework 2

CSIT6000K- Social Networks and Social Computing: A Data Science Perspective

Name: ……

Student ID: ……

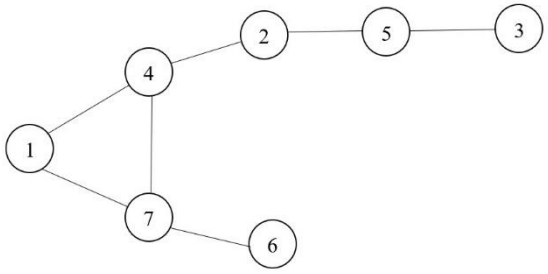
E-mail Address: ……

Instructions:

- In this HW, we have 9 theoretical questions.
- Both hand-written (scanned) and typed answers are accepted.
- You can discuss your ideas with your friends; however, the final answers should all be your own work. ZERO-Tolerance on Plagiarism: All involved parties will get zero mark.
- Put all of your files neatly into a zip file and name it "hw2_studentID_studentName"
- For each hour delay, you will lose 2% of your homework's grade. We will not accept submissions received 24 hours after the deadline.

Questions:

1- Consider the graph below to answer the following questions. Explain your answers. **[6 points]**

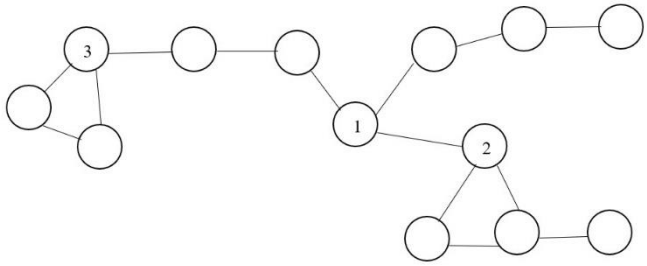


1-1- Which of the node(s) have the largest Degree centrality?

1-2- What is the Closeness centrality for Node 2?

1-3- Which of the node(s) have largest Betweenness centrality?

2- Consider the graph below to answer the following questions. Explain your answers. **[4 points]**

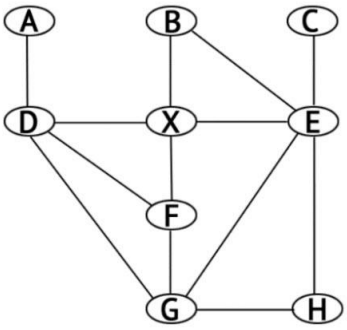


2-1- What is the normalized degree centrality of node 1?

2-2- What is the closeness centrality of node 1?

3- Consider the figure shown below. Suppose this graph represents a network. You are consulting for a large social-networking company, and your job is to help design the part of the system that recommends new friendships to their members (i.e., to the users of the system). The idea is to look at the position of a member X in the overall social network, and try to automatically recommend, based on the pattern of links, the name of one other user to whom X is not currently connected, but to whom X might want to connect. If Y is chosen by the system as the recommendation for X, then X receives a screen prompt asking if X would like to add Y as a friend. The recommendation is viewed as successful if X accepts the suggestion and adds Y.

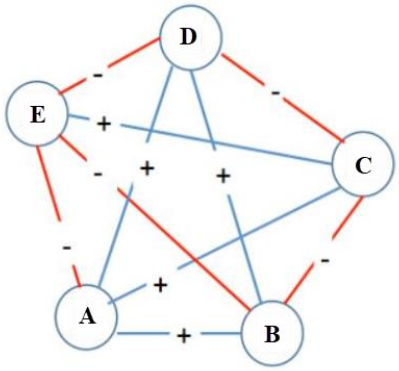
You have been attending meetings in which the team in charge of the recommendation system shows examples of real-life situations, and asks you, based on your knowledge of networks, what the system should recommend. They seek your feedback to fine-tune the behavior of the system. Which node do you think is the best choice to recommend to X? Explain your answer. **[5 points]**



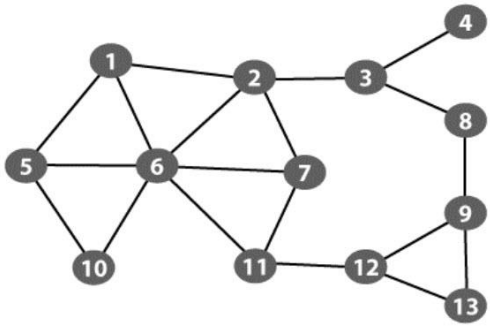
4- Suppose you are talking to a group of people who are managing a website where people download music. They believe that the popularity of songs on this site closely follows a power law with exponent 2: the number of songs that get k downloads is equal to c/k^2 for some constant c. From their log files, they find that there are 10,000 songs that have each received exactly 10 downloads.

If you believe their claim that the downloads follow a power law with exponent 2, what do you expect to see as the number of songs that have each received 100 downloads? **[10 points]**

5- Consider the network below, which depicts a five-node social network. Each edge is labeled as either a friend (positive) or enemy (negative) relation. Which of the triangles are not balanced? Explain. **[10 points]**



6- Suppose we have a behavior that is spreading through the social network shown in the figure below. It begins with nodes 5 and 6; after this, each node will switch to the new behavior if at least half of its neighbors have already adopted the new behavior. **[10 points]**



6-1- Consider which nodes will be affected when a new behavior begins spreading from nodes 5 and 6. In addition to 5 and 6, which other nodes will eventually switch to the new behavior?

6-2- Suppose you view the spread of this behavior as a bad thing (maybe, for example, it's the spread of a false and dangerous rumor) and you'd like to limit its reach by removing a link from the network. You want to select a single edge in this figure that will help limit the reach of the behavior as follows: If you delete this edge and then the behavior spreads from nodes 5 and 6 according to the same rule as before, on this graph with the edge deleted, only four nodes will be affected in total (including the original nodes 5 and 6). Which edge is it?

(**Note**: there are more than one correct answers to this question, indicate all of them.)

7- Suppose people are sequentially making decisions about whether or not to adopt a product. Assume that the product is with probability 1/2 good, and with probability 1/2 bad. Each person gets a signal independently about the quality of the product. For a good product, the signal is H with probability 4/5th, and L with probability 1/5. For a bad product, the signal is L with probability 4/5th, and H with probability 1/5. Each person gets a reward of +1 for accepting a good product, and a cost of -1 for accepting a bad product. Each person in turn, observes the decision of the previous person, but not their signal. **[20 points]**

7-1- What is the probability that the product is good, if a person receives an H signal? Explain your answer.

7-2- What is the probability that the product is good, if a person receives an L signal? Explain your answer.

7-3- What is the probability two people in a row receive an H signal, assuming the product is actually bad? Explain your answer.

7-4- Now suppose that the cost of accepting a bad technology is actually -3, while the value for a good technology remains 1. Suppose the first person receives a H signal. What would be the expected value of the technology for the person? Explain your answer.

8- Consider a regular lattice (similar to what we saw in Watts-Strogatz model), prove that the number of connections between neighbors is $\frac{3}{8}\ c\ (c-2)$, where c is the average degree. Explain your proof thoroughly. **[10 points]**

9- Given the definition of following models:

- **Static geographic model:** Given N nodes randomly dispersed in the space, each node has a fixed position, and each node is connected to its *m* nearest neighbors, *m* is the same for all the nodes.
- **Random growth model:** Similar with preferential attachment growth model (Barabasi-Albert model), except that for each new added node, the probability p_i that the new node

is connected to node *i* is the same for all the existed nodes $p_i = \frac{1}{n}$, n is the total number of existed nodes when the new node is added.

In this question we will compare these two models with **Erdos-Renyi model** and **preferential attachment model**. **[25 points]**

9-1- For an Erdos-Renyi model with N nodes and link probability p, what is the expected degree $\bar{k}$?

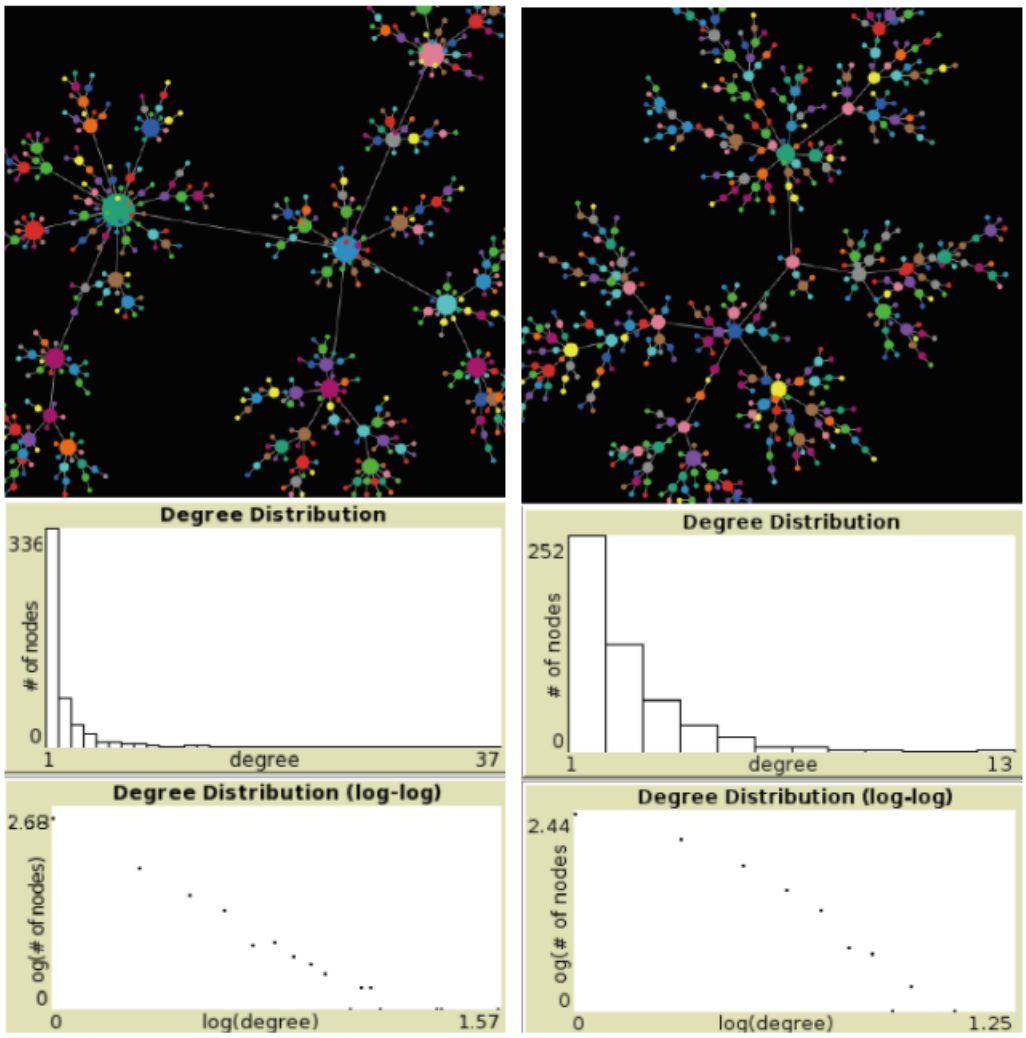
9-2- Consider a static geographic model with N nodes and $m \approx \bar{k}$, then compare it with the previous Erdos-Renyi model, whether the following statements are true or false, and give your explanations?

a) the static geographic model has stronger locality

b) the static geographic model has shorter average shortest path

9-3- Compare random growth model with Erdos-Renyi model, explain why links are unevenly distributed in random growth model? Which model has more nodes with degree 1?

9-4- Compare random growth model with preferential attachment model, for the following two pictures, each has 500 nodes, which model fits best to each picture and explain why?



(a) (b)

(a) Model 1; (b) Model 2.

Good Luck