NAME	Uxxxxxx	GRADE

Introduction to Networks Science (2023-2024)

FINAL EXAM

WRITE YOUR ANSWERS BRIEFLY and CLEARLY IN THE BLANK SPACES. Please underline key words in your answers. Please if you include intermediate calculations, circle the final result. If for some reason (e.g., if after you have written the solution you realize that there is some mistake), you can attach an extra sheet to your exam. In this case, indicate clearly that the solution can be found in the extra sheet.

Problem 1 *0.5 point*

In the Erdős–Rényi (ER) model, the degree follows a Binomial distribution with parameters p for the "probability of success" and $N-1$ for the "number of trials."

What is the value at which this distribution has the maximum probability? Explain briefly why.

Problem 2 *0.5 point*

One interpretation of Milgram's experiment from 1967 is that, given that the average path length of delivered letters was 6, this is the average path length in the underlying social graph.

Why do we claim this is an underestimate of average path length?

Problem 3 *1 point*

Consider the Linear Threshold propagation model in a graph of N nodes, in which the threshold θ_i for every node $i = 1, 2, \ldots, N$ is strictly larger than zero and strictly smaller than $\min_{(i,j)\in E} w(i,j)$ in which $w(i,j)$ is the influence weight from node i to node j.

If we run this propagation model in an ER graph starting from a random node, what should be the average degree on the ER graph to ensure that, with high probability, all or almost all of the nodes will be infected in the end?

Your answer here, justifying your answer:

Problem 4 *1 point*

TikTok has approximately 1.5×10^9 users, and the distribution of the number of followers on it follows a power law with exponent $\gamma = 2.1$. Remember that in a scale-free network $p_k = k^{-\gamma}/\zeta(\gamma)$, and $\zeta(2.1) \approx 1.56$.

We observe that there are 3 accounts that have about 100 million followers each.

Are those 3 accounts with 100 million followers each (a) too many, (b) too few, or (c) approximately what we would expect? Justify.

Problem 5 *1 point*

A malicious actor wants to attack a communications network, which we assume is a scale-free network, by destroying some of its nodes. The goal of the attacker is to increase average path length.

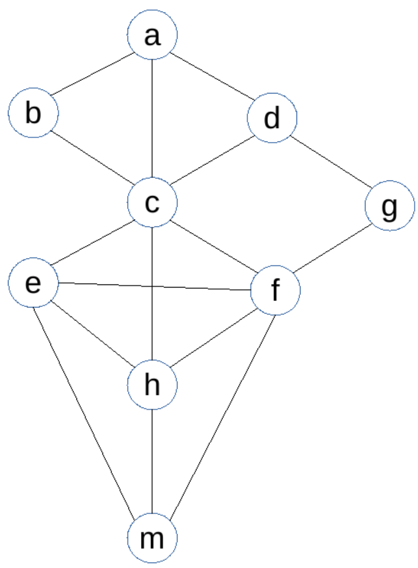
There are two possible attack modalities: one that picks two nodes at random and destroys them, and one that picks a link at random and destroys the two nodes attached to it.

Which attack is more likely to be effective, or both are equally effective? Justify your answer by using relevant concepts in a precise manner:

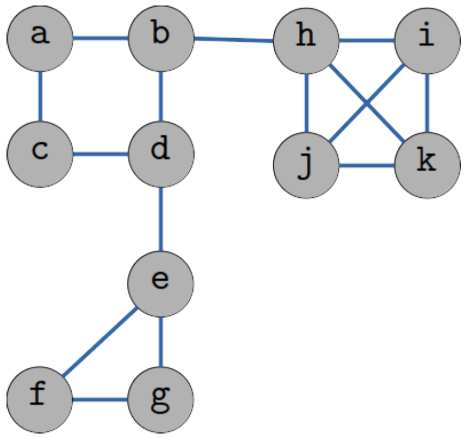
Problem 6 *1 point*

Perform a k-core decomposition of the graph shown on the right, indicating clearly which nodes belong to the 1-core, 2-core, 3-core, ...

Nodes in each of the cores of the graph:



Problem 7 *2 points*



We want to divide the graph on the right into communities, but we are not sure about whether there should be:

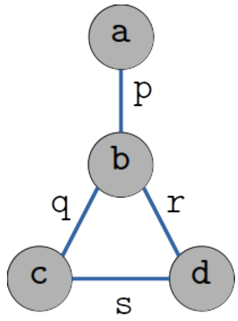
- two communities: $\{a, b, c, d, e, f, g\}, \{h, i, j, k\}$, or
- three communities: $\{a, b, c, d\}, \{e, f, g\}, \{h, i, j, k\}$.

We want to choose between these two alternatives using the modularity criterion. The formula for modularity is $Q = \frac{1}{L} \sum_C \left(L_C - \frac{k_C^2}{4L} \right)$ where C are the communities, L is the total number of links, L_C is the number of links internal to community C, and k_C is the summation of the degree of the nodes in C.

Which alternative is best? Justify your answer by computing modularity. In your answer, mark clearly the modularity of the first and second option, and state clearly whether a smaller or larger modularity is preferable.

Problem 8 *1 point*

Consider a propagation under the Independent Cascade model, which starts by infecting node "a". Indicate the probability that at the end of this process, nodes "a" and "b" are infected, but nodes "c" and "d" are not infected. In the figure, the values next to each edge p, q, r, s correspond to propagation probabilities along their respective edges.



Probability that nodes "a" and "b" are infected, but nodes "c" and "d" are not infected? Justify your answer.

Problem 9 *2 points*

Prove that a constant vector (i.e., a vector in which all coordinates are equal) is an eigenvector of the Laplacian L of a graph.

How? Write the definition of Laplacian (0 points), multiply L by a constant vector and conclude (2 points).

Define each of the variables you use.

Do not give a concrete example of a specific graph – write a general proof.

Your proof here.