Questions – 50pt

**Question 1** (15pt) – Consider a set of items \( U = \{a_1, \ldots, a_n\} \), each item \( a_i \) has a weight \( w_i \). We have a knapsack of capacity \( B \). We say that the activities in a set \( X \subseteq U \) are independent if their collective weight is less than or equal to \( B \), i.e.
\[
\sum_{a_i \in X} w_i \leq B.
\]
We define the set \( S \subseteq 2^U \) as the set of subsets of \( U \) whose items are independent, i.e. \( S = \{ X \text{ s.t. } X \subseteq U \text{ and } \sum_{a_i \in X} w_i \leq B \} \).

Prove that \((U, G)\) is an independence system, but it is not a matroid. (5pt)

Prove that if \( w_i = 1 \) for each item \( i = 1, \ldots, n \), then \((U, G)\) is a matroid. (5pt)

What is the implication of the above proof? (5pt)

**Question 2** (15pt) – What is the probability that an Erdos-Rényi graph \( G(n, p) \) is a star, i.e. one node has degree \( n-1 \) and all other nodes have degree 1? Note that, the number of edges is a necessary but not a sufficient condition for a graph to be a star. Additionally, there can be several possible stars with \( n \) nodes.

**Question 3** (10pt) – Define a giant component of a graph and discuss the phase transition in the formation of such a component in an Erdos-Rényi graph \( G(n, p) \).

**Question 4** (10pt) – Describe the Spectral Modularity Maximization approach to the community detection problem.
**Mini-project – 50pt**

We define an extension of the Erdos-Rényi (ER) model in which we have two probabilities $p_1$ and $p_2$. We generate a graph $G(n, p_1, p_2)$ as follows.

- Assume that vertices are numbered $1,...,n$.
- Given two vertices $i$ and $j$, the probability of having an edge between $i$ and $j$ is $p_1$, if $(i + j) \mod 2 = 0$, and $p_2$ otherwise.

Develop a program to generate graphs according to the modified ER model. The program should take as an input the number of nodes $n$ and the probabilities $p_1$ and $p_2$.

**A)** Generate ER a networks of 50 nodes and visualize with “Force Atlas 2” layout – 10pt

- Network 1: use $p_1 = 0.8$ and $p_2 = 0.1$
- Network 2: use $p_1 = 0.4$ and $p_2 = 0.5$
- Provide a snap-shot of the resulting networks
- Comment the resulting networks

**B)** Study the performance of the simple modularity maximization (for network bisection) heuristic applied to the modified ER model. The study should be performed in terms of three experiments. The results of each experiment should be shown as a separate graph. The heuristic should be compared to a random approach that assigns each vertex to one of the two classes randomly. The results in each graph should be discussed in detail.

**B.1)** Fix $n = 50$, $p_1 = 0.5$, and vary $p_2$ from 0 to 1. Measure the modularity achieved by the heuristic in comparison with the random approach. (10pt)

The graph has on the x-axis the value of $p_2$ and on the y-axis the value of the modularity.

**B.2)** Consider three different settings of $(p_1; p_2)$ that is, $(0;1)$, $(0.1;0.9)$, $(0.4;0.5)$. Show the modularity achieved by the heuristic in comparison with the random approach by increasing the network size $n$ from 0 to 500. (10pt)

The graph has on the x-axis the network size $n$, on the y-axis the value of the modularity. Each setting should be a different line on the graph.

**B.3)** Fix $n = 50$ and increase $p_1$ and $p_2$ from 0 to 1. Measure the modularity achieved by the heuristic in comparison with the random approach.

Generate a 3d graph, which has on the x-axis the value of $p_1$, on the y-axis the value of $p_2$, and on the z-axis the value of the modularity. (20pt)