The homework is due on **Thursday, April 14, 14:00.** You can hand it in before the class starts in person or drop it in the mailbox of Evimaria (at room A335) before the due date. Collaboration is allowed, but you have to write the solutions by yourself and mention the names of your collaborators. You will receive 20 points for each problem that you solve.

Question 1 Consider the following evolving model for generating an undirected graph. Initially there are only three nodes connected into a triangle. At every time step, an edge of the current network is selected uniformly at random, and a new node is added to the network that links to both the endpoints of the edge.
Prove that p_k, the fraction of nodes with degree k, follows a power law with exponent 3. Specifically, show that

$$p_k = \frac{12}{k(k+1)(k+2)}$$

Provide an intuitive explanation as to why this model is the same as the preferential attachment model.

Question 2 Consider the following small world model. There are n nodes arranged on a **directed** ring. Every node is connected to the one on its left with a directed link of unit length. We add a new central point to the network, and connect it with every other vertex on the circle with probability p, using an undirected link of length $1/2$. Essentially, this means that we select some vertices with probability p and then connect them with an undirected link of length 1.

1. Compute the probability $P(l,k)$ that the shortest path from one vertex to another is l when the distance between the vertices along the ring is k. (Hint: $\sum_{l=1}^{k} P(l,k) = 1$)
2. Find the distribution $P(l)$ of the shortest path between pairs of vertices.
3. [**Bonus question (5 points)**] Compute the average shortest path between the vertices of the graph

Question 3 Consider the two clustering coefficients $C^{(1)}$ and $C^{(2)}$ defined in class. Construct a graph where $C^{(1)} = \Theta(1)$ and $C^{(2)} = \Theta(1/n)$. Give a short explanation for the intuition behind your construction.

Question 4 Let X be a random variable that holds the income of an individual. Initially X takes a value chosen uniformly in the interval $[m, 2m)$. The income is updated iteratively according to the following process. If X_t is the income at time t, then at time $t+1$ the income X_{t+1} is computed as follows:

- If $X_t \geq 2m$, then $X_{t+1} = 2X_t$ with probability $1/3$, and $X_{t+1} = \frac{1}{2}X_t$ with probability $2/3$.
- If $X_t < 2m$, then $X_{t+1} = 2X_t$ with probability $1/3$, and $X_{t+1} = X_t$ with probability $2/3$.

Define an income to be of class c, if it is in the interval $[2^{c-1}m, 2^c m)$. According to the previous rules, at each time step, the income moves from class c to class $c+1$ with probability $1/3$, or it moves to class $c-1$ with probability $2/3$, except for class 1, where the income stays at the same class. When $t \to \infty$, the probability that the income belongs to a certain class converges. We now study the distribution of X when $t = \infty$

1. Prove that the probability $p(c)$ that the income belongs to class c is $1/2^c$.
2. Prove that
$$Pr[X \geq 2^{c-1}m] = \frac{1}{2^{c-1}}$$
Conclude that X follows a power law with exponent 2.
3. Prove that X follows a power law with exponent 2, using the combination of exponentials method.

This exercise shows that when we have a multiplicative process that is bounded from below, then instead of obtaining a log-normal distribution, we obtain a power law.

Question 5 In this question you are asked to implement the algorithms for measuring power-laws, and other properties of networks. You should submit a printout of your code, as well as e-mail the code to Evimaria (trezi@cs.helsinki.fi). There should be clear instructions on how to run the experiments. If you have pre-processed the data, you should either submit the code for preprocessing, or submit also the pre-processed dataset.

1. Generate a dataset of 1M values following a power-law distribution with exponent 2.5. Then compute experimentally the exponent of the distribution, using the following 4 methods:
   - Fitting a line to the frequency distribution.
   - Fitting a line to the frequency distribution with logarithmic binning.
   - Fitting a line to the cumulative distribution.
   - Using the maximum likelihood estimate.
2. Download one of the AS Internet graphs, and the co-authorship graph from the home page of the course, and perform the following measurements.
   - Demonstrate that the degree distributions follow a power law, and compute the exponent of the distribution with the four different methods.
   - Compute the clustering coefficients $C^{(1)}$ and $C^{(2)}$.
   - Compute the degree correlation, using the method of Newman.
3. **[Bonus points (5 point)]** Perform the same measurements for some of the other networks on the course home page. If you work on directed graphs, for degree related measurements you can focus on the in, or out degree, or do separate measurements for each, or view the graph as undirected. For the clustering coefficient you should view the graph as undirected.