Complex Networks

Teachers: L. Avena, M. Emmerich, D. Garlaschelli.
Written examination: 5 January 2023.

Open book exam: the lecture notes may be consulted, but no other material.

Answer each question on a separate sheet. Put your name, student number and the number of the question you are answering on every sheet. Provide full explanations with each of the answers!

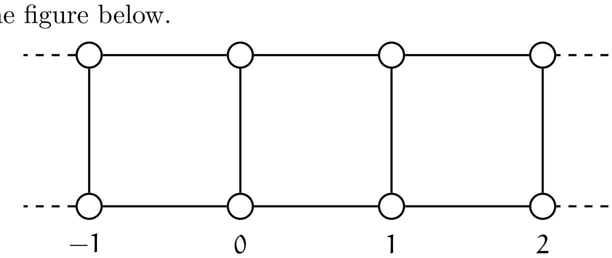
Each question is weighted by a number of points, as indicated. The total number of points is 100. The final grade will be calculated as a weighted average: 30% for the homework assignments and 70% for the exam.

Success!

1. [**4 points**] What is a random graph ensemble and why is it useful in the study of real-world complex networks?

2.
   a. [**2 points**] Give the definition of Ordinary Percolation with parameter p on a general connected simple graph.

   b. [**3 points**] Consider Ordinary Percolation on the infinite ladder graph L in the figure below.



   What is the critical percolation value $p_c(L)$? Justify your answer with a clear mathematical argument.

   c. [**5 points**] Consider Ordinary Percolation on the two dimensional integer lattice $\mathbb{Z}^2$ with *supercritical* parameter $p > 1/2 = p_c(\mathbb{Z}^2)$. What is the probability that all the edges on the horizontal axis traversing the origin belong to the unique infinite cluster? In other words, you are required to compute the following probability:

$$\mathbb{P}(\text{ every } e \in A \text{ belongs to the unique infinite cluster }),$$

   where $A := \{(x, y) \in E(\mathbb{Z}^2) : x = (z, 0) \text{ and } y = (z + 1, 0), z \in \mathbb{Z}\}$

and $E(\mathbb{Z}^2)$ denotes the edge set of $\mathbb{Z}^2$.

d. **[6 points]** Fix $n \geq 2$ and $k < n$. Consider Ordinary Percolation of parameter p on a complete graph K_n with n nodes. Compute the average total number of distinct k-*star* subgraphs contained in the resulting p-clusters.
We recall that a k-*star* graph is a graph built on $k+1$ vertices where one of them, called *center of the star*, has one connection with all the other k nodes.

3. Consider a simple graph $\mathbf{G}^*$ with $n^* = 6$ vertices, $L^* = 9$ undirected edges and degree sequence equal to $\vec{k}^* = (5, 4, 3, 2, 2, 2)$.

a. **[3 points]** How many simple (undirected) graphs (including $\mathbf{G}^*$ itself) exist with n^* vertices and degree sequence equal to $\vec{k}^*$?

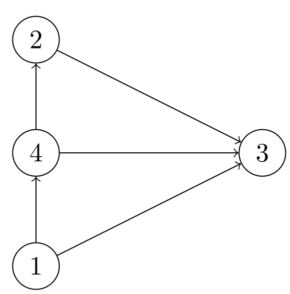
b. **[3 points]** How many simple (undirected) graphs (including $\mathbf{G}^*$ itself) exist with n^* vertices and L^* edges?

c. **[3 points]** Calculate the probability $\mathbb{P}(\mathbf{G}^*|p^*)$ of generating the graph $\mathbf{G}^*$ under the Erdős-Rényi random graph model with n^* vertices and connection probability p^*, where p^* is the value of the connection probability p that maximises the log-likelihood $\ln \mathbb{P}(\mathbf{G}^*|p)$ for the model, given the data.

d. **[3 points]** Calculate the probability of generating *any* of the simple (undirected) graphs (including $\mathbf{G}^*$ itself) with n^* vertices and L^* edges, under the Erdős-Rényi random graph model with connection probability p^*, where p^* is the same as above.

e. **[4 points]** Given the degree sequence $\vec{k}^*$, calculate the numerical value of the connection probability p_{ij} for all pairs of nodes under the Chung-Lu model and arrange the resulting 'probabilities' into a 6×6 matrix with entries $\{p_{ij}\}$. Comment your result.

4. a. **[4 points]** Write a small program in the syntax of igraph (R or Python) to construct this graph and plot it. Use the adjacency matrix representation.



b. **[2 points]** Given a sparse, undirected graph with n nodes. What is the time complexity in terms of n in the big O notation of finding the node with the highest degree when the graph is stored as an *adjacency matrix* data structure?

c. [**2 points**] Given a sparse, undirected graph with n nodes. What is the time complexity in terms of n in the big O notation of finding the node with the highest degree when the graph is stored as an *adjacency list* data structure?

5. Consider bipartite-directed graphs (BD graphs) with $2n$ nodes. In contrast to a general directed graph, there is the restriction that nodes with odd numbers as an index (v_1, v_3, ...) can only be directly linked with nodes with even numbers as an index (v_2, v_4, ...), and vice versa. Direct links between two even or two odd nodes are not allowed.

   a. [**2 points**] How many different BD graphs can be formed?

   b. [**5 points**] Describe an efficient procedure of producing a (uniformly) random graph from the above bipartite graph model with $2n$ nodes and m edges. Describe your algorithm as a pseudo-code and analyze its computational time complexity. Strive for an efficient way to generate a random graph, for instance, by considering the Fisher-Yates Shuffle.

6. a. [**3 points**] Consider a generic random graph ensemble $\mathcal{G} = (G_n)_{n\in\mathbb{N}}$. Explain what it means mathematically that $\mathcal{G}$ is *scale free with parameter* $\tau \in (1, \infty)$.

   b. [**5 points**] Let $CM(\vec{k})$ be a configuration model with degree sequence (possibly random) $\vec{k}$, and let $PA(m, \delta)$ denote the preferential attachment model, for some given $m \in \mathbb{N}$ and $\delta \in [-m, \infty)$, Give sufficient assumptions both for the configuration model $CM(\vec{k})$ as well as for the preferential attachment $PA(m, \delta)$ which guarantee that these ensembles are scale free with some parameter $\tau \in (1, \infty)$.

   c. [**2 points**] What is the probability that vertex 1 in $PA(m, \delta)$ has at least one self-loop?

   d. [**6 points**] Consider the constant degree sequence $\vec{k} = (k_i)_{i\in\mathbb{N}}$ with $k_i = d$ for all $i \in \mathbb{N}$ and some fixed $d \geq 2$. What is the probability that vertex 1 in $CM(\vec{k})$ has at least one self-loop?

7. Consider two simple undirected graphs $\mathbf{G}_1$ and $\mathbf{G}_2$ with $n = 5$ vertices and adjacency matrices given by

$$\mathbf{G}_1 = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix}, \qquad \mathbf{G}_2 = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{pmatrix}.$$

a. **[2 points]** If $\mathbf{G}_1$ and $\mathbf{G}_2$ have the same degree sequence, then find a move of the Local Rewiring Algorithm that changes $\mathbf{G}_1$ into $\mathbf{G}_2$. If they do not, then describe which changes are needed to turn $\mathbf{G}_1$ into $\mathbf{G}_2$.

b. **[2 points]** For both $\mathbf{G}_1$ and $\mathbf{G}_2$, calculate the average nearest neighbour degree k_i^{nn} for all nodes ($i = 1, \ldots 5$) and its node average $\overline{k^{nn}} \equiv \frac{1}{n}\sum_{i=1}^{n} k_i^{nn}$.

c. **[2 points]** For both $\mathbf{G}_1$ and $\mathbf{G}_2$, calculate the local clustering coefficient c_i for all nodes (recall that, for nodes with degree 0 or 1, the local clustering coefficient is conventionally set to zero) and its node average $\overline{c} \equiv \frac{1}{n}\sum_{i=1}^{n} c_i$.

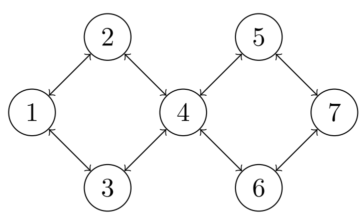
d. **[3 points]** For both $\mathbf{G}_1$ and $\mathbf{G}_2$, calculate the expected average nearest neighbour degree $\langle k_i^{nn} \rangle$ for all nodes ($i = 1, \ldots 5$) under the Erdős-Rényi (ER) random graph with the same number of nodes and expected number of links as the given network. For which of the two graphs is the realized value of $\overline{k^{nn}}$ closer to the corresponding expected value $\langle \overline{k^{nn}} \rangle$ under the ER model?

e. **[3 points]** For both $\mathbf{G}_1$ and $\mathbf{G}_2$, calculate the expected local clustering coefficient $\langle c_i \rangle$ for all nodes ($i = 1, \ldots 5$) under the Erdős-Rényi random graph with the same number of nodes and expected number of links as the given network. For which of the two graphs is the realized value of $\overline{c}$ closer to the corresponding expected value $\langle \overline{c} \rangle$ under the ER model?

f. **[2 points]** Which of the two graphs has the highest probability of occurrence under the ER model?

g. **[2 points]** Which of the two graphs has the highest probability of occurrence under the Configuration Model?

8. Consider a 2-ring graph with 7 nodes.

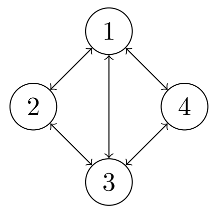


Assume SI model simulated by a Continuous time Markov chain and that node 4 got just infected at time t_0 and all other nodes are in a susceptible state. Assume λ denotes the contagiousness of the virus (for a single contact or edge).

a. **[4 points]** What is the expected time (from t_0, in terms of λ) until the second node gets infected. What is the probability that node 2 will be the second node that gets infected?

b. **[4 points]** Assume, again, node 4 got first infected at time t_0. What is the expected time from t_0 until the third node gets infected? What is the probability that this will be node 1?

c. **[5 points]** Next, consider a network with 4 nodes:



Write up the specific equation system that determines the *page rank centrality* for the 2-ring graph for a damping factor $\alpha = 0.5$ in terms of x_1, ..., x_4. Which nodes in the network share the same page rank centrality? (hint: it is not required to solve the equation system to answer this question)