Complex Networks

Teachers: L. Avena, M. Emmerich, D. Garlaschelli.
Written examination (retake): 31 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**] Briefly describe the different approaches that Mathematics, Physics and Computer Science adopt in the study of Complex Networks.

2. Consider the Erdős-Rényi random graph model $ER_n(p)$ with n vertices and connection probability $p \in (0, 1)$.

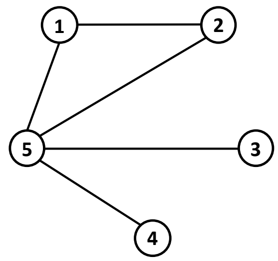
   a. [**6 points**] Fix $k < n$. Compute the expectation of the total number of (unoriented) *k-cycles* in $ER_n(p)$.
   Recall that a *k-cycle* is a collection of k distinct vertices $\{v_1, v_2 \ldots, v_k\}$ such that v_i is connected to v_{i+1} for $i < k$ and v_k is connected to v_1.

   b. [**4 points**] Fix $a \in (0, 2)$. How should p scale with n to guarantee that the average total number of edges in Erdős-Rényi is of order n^a?

   c. [**6 points**] We have seen that when p scales like $1/n$ (i.e. sparse regime) Erdős-Rényi undergoes the so-called Percolation Transition. Describe what sort of *Branching Process* is needed to prove such a Percolation Transition and describe shortly the main idea of the underlying proof.

   d. [**4 points**] We have seen that the Percolation Transition is also true for the Configuration Model with i.i.d. degrees sampled from a distribution with finite second moment. Explain in which sense this Percolation Transition for the Configuration Model differs from the one for Erdős-Rényi in the sparse regime.

3. Consider the graph $\mathbf{G}^*$ in the figure below.



a. **[2 points]** Compute the empirical link density (or connectance) $c(\mathbf{G}^*)$, the empirical average nearest neighbour degree $k_i^{nn}(\mathbf{G}^*)$ of each vertex i and the empirical local clustering coefficient $c_i(\mathbf{G}^*)$ of each vertex i.

b. **[2 points]** Consider the Erdős-Rényi model with $n = n(\mathbf{G}^*)$ vertices and connection probability p. Write the probability $\mathbb{P}(\mathbf{G}|p)$ to generate any graph $\mathbf{G}$ in the model.

c. **[2 points]** Write the log-likelihood $\lambda(p) \equiv \ln \mathbb{P}(\mathbf{G}^*|p)$ for the model, given the data, explicitly as a function of $L_u(\mathbf{G}^*)$, where L_u is the total number of undirected edges and $\mathbb{P}(\mathbf{G}^*|p)$ is the probability to generate the specific graph $\mathbf{G}^*$.

d. **[2 points]** Find the value p^* of the connection probability that maximises $\lambda(p)$.

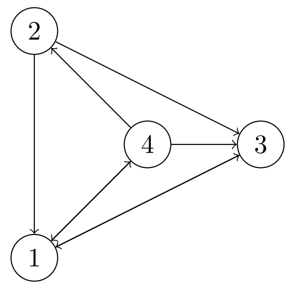
e. **[2 points]** For the value p^* found above, compute the expected value $\langle k_i^{nn} \rangle$ of the average nearest neighbour degree of each vertex i under the model. Compare each $\langle k_i^{nn} \rangle$ with the empirical value $k_i^{nn}(\mathbf{G}^*)$ found above. Comment your result.

f. **[2 points]** For the value p^* found above, compute the expected value $\langle c_i \rangle$ of the local clustering coefficient of each vertex i under the model. Compare each $\langle c_i \rangle$ with the empirical value $c_i(\mathbf{G}^*)$ found above. Comment your result.

g. **[2 points]** What do the graphs drawn from the model described above look like, compared with the original graph $\mathbf{G}^*$?

h. **[2 points]** What would the graphs drawn from the corresponding *microcanonical* ensemble (with a *hard* constraint on L_u) look like?

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 edge list 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 counting the number of isolated nodes, i.e., nodes with no connection to other nodes, if the graph is stored as an *adjacency matrix* data structure?

c. [**2 points**] Given a non-sparse undirected graph with n nodes. What is the time complexity in terms of n in the big O notation of counting the number of isolated nodes if the graph is stored as an *adjacency list* data structure?

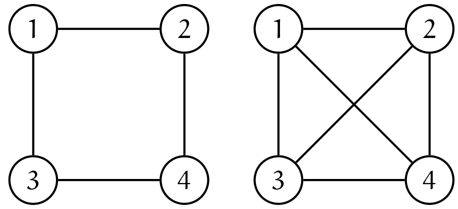
5. Consider a directed graph $G = (V, E)$ with a set of n nodes (V) and a set of m edges (E). A subgraph $G' = (V', E')$of this graph is a graph for which it holds that $V' = V$ and $E' \subseteq E$.

   a. [**2 points**] How many different subgraphs with k edges can be formed?

   b. [**5 points**] Describe an efficient procedure of producing a (uniformly) random subgraph of G with $k \leq m$ edges (k is provided by the user). 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.

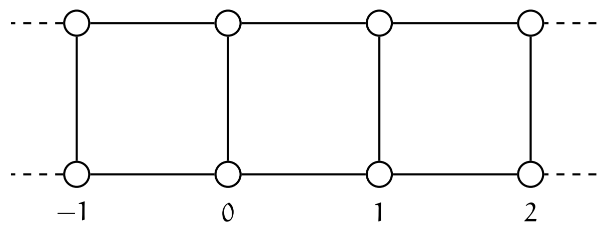
6. a. [**3 points**] Define the so-called Invasion Percolation model on an arbitrary connected (possibly) finite graph exactly in the same way in which it is defined on the integer lattice.

   b. [**2 points**] Consider Invasion Percolation on the two graphs on 4 vertices in the picture below (a 4-cycle graph on the left, and a complete graph of size 4 on the right, respectively).



   Will the iterative algorithm to build the Invasion Percolation Cluster invade the whole vertex set in the same number of steps in both graphs? Motivate the answer.

c. **[7 points]** Consider now Invasion Percolation model on the *infinite ladder graph* in the figure below.



Describe the resulting Invasion Percolation Cluster. Justify your answer with a rigorous mathematical argument.

7. Consider a simple undirected graph $\mathbf{G}_1$ with $n = 6$ vertices. The graph contains the edge $(3, 6)$. Add edges to the graph in such a way that the degree sequence $\vec{k} = (5, 4, 3, 2, 2, 2)$ is realised.

a. **[4 points]** Calculate the average nearest neighbour degree k_i^{nn} for each vertex and make a scatter plot where each point represents a vertex and has coordinates (k_i, k_i^{nn}). From the trend of the plot, would you conclude that the graph $\mathbf{G}_1$ is assortative or disassortative?

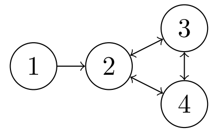
b. **[4 points]** Calculate the local clustering coefficient c_i for each vertex and make a scatter plot where each point represents a vertex and has coordinates (k_i, c_i). Is the trend of the plot increasing or decreasing?

c. **[4 points]** For each vertex, calculate the expected value $\langle c_i \rangle$ of c_i under the Erdős-Rényi model having an expected number of edges equal to the realised number of edges of $\mathbf{G}_1$. For each vertex, compare the realised value c_i in $\mathbf{G}_1$ with the expected value $\langle c_i \rangle$: which vertices have a higher or lower clustering coefficient in $\mathbf{G}_1$ with respect to the model?

d. **[2 points]** Draw another graph $\mathbf{G}_2$ with the same degree sequence as $\mathbf{G}_1$ and where the edge $(3, 6)$ is not necessarily present. Can you find a move of the local rewiring algorithm that transforms $\mathbf{G}_1$ into $\mathbf{G}_2$?

e. **[2 points]** Which of the quantities calculated above for $\mathbf{G}_1$ (i.e. k_i^{nn} and c_i for all i) are different in $\mathbf{G}_2$?

8. Consider a network with four nodes:.



Assume SI model simulated by a Continuous time Markov chain and that node 1 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. **[2 points]** What is the size of the state space of this SI process? What is the size (columns, rows) of the generator matrix of the CTMC process?

b. **[2 points]** What is the expected time (from t_0, in terms of λ) until two nodes in the network are infected?

c. **[2 points]** What is the expected time (from t_0, in terms of λ) until three nodes in the network are infected?

d. **[2 points]** What is the probability that the third node that gets infected will be node 4?

e. **[3 points]** Write up and solve the specific equation system that determines the *page rank centrality* for the four-node graph above for a damping factor $\alpha = 0.5$ in terms of x_1, ..., x_4.

f. **[3 points]** Solve the equation system and thereby determine the page rank centrality of every node.

g. **[3 points]** Which nodes in the network share the same page rank centrality? Which node has the highest centrality? Which node has has the lowest centrality?