Complex Networks

Teachers: M. Emmerich, D. Garlaschelli, F. den Hollander.
Written examination: 11 January 2019, 14:00-17:00.

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. a. **[3 points]** List the four main classes of complex networks.
   b. **[4 points]** Give one example in each class and provide a brief description of each example.

2. Consider the graph $G(V, E)$ with vertex set $V = \{1, 2, 3, 4\}$ and edge set $E = \{12, 23, 34, 41, 13, 24\}$.
   a. **[3 points]** Compute the typical distance.
   b. **[4 points]** Compute the number of wedges and the number of triangles.
   c. **[2 points]** Compute the global clustering coefficient.

3. Consider the Erdős-Rényi random graph with n vertices and with retention probability λ/n, where $\lambda \in (0, \infty)$. Start an exploration of the graph from a given vertex $*$.
   a. **[4 points]** What branching process stochastically dominates the number of vertices explored at the k-th generation of the exploration process for $0 \leq k \leq n$?
   b. **[5 points]** For which values of k is this branching process a good approximation of the actual number of vertices? Why?
   c. **[4 points]** Explain what the 'depletion effect' in the exploration process is.

X 4. Consider a real-world binary undirected graph $\mathbf{G}^*$ with n vertices and degree sequence $\{k_i^*\}_{i=1}^n \equiv \{k_i(\mathbf{G}^*)\}_{i=1}^n$, where $k_i^* = k_0$ for all i.
   a. **[2 points]** Describe the procedure how to construct the maximum-

entropy ensemble of binary undirected graphs with n vertices and expected degree sequence $\{\langle k_i \rangle\}_{i=1}^n = \{k_i^*\}_{i=1}^n$.

b. **[2 points]** Write the resulting probability $P(\mathbf{G})$ of any graph $\mathbf{G}$ in such ensemble, as a function of only the degree sequence $\{k_i(\mathbf{G})\}_{i=1}^n$ of $\mathbf{G}$ (note that $\mathbf{G}$ is *any* graph, not necessarily $\mathbf{G}^*$).

c. **[3 points]** Write the equation(s) required to fix the value of the parameters of $P(\mathbf{G})$ and solve these equations explicitly, writing the parameters as a function of k_0.

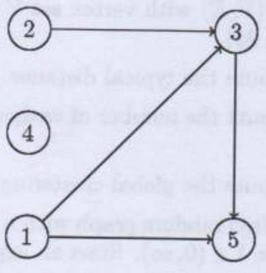
d. **[3 points]** Discuss how this ensemble relates to the Configuration Model and to the Erdős-Rényi random graph model.

e. **[3 points]** Now consider two graphs $\mathbf{G}_A$ and $\mathbf{G}_B$ belonging to this ensemble. Calculate the ratio $P(\mathbf{G}_A)/P(\mathbf{G}_B)$ of the probabilities of generating the two graphs, only as a function of the numbers $L(\mathbf{G}_A)$ and $L(\mathbf{G}_B)$ of undirected links of $\mathbf{G}_A$ and $\mathbf{G}_B$ respectively.

f. **[3 points]** Discuss how one may choose $\mathbf{G}_A$ and $\mathbf{G}_B$ in order to maximize $P(\mathbf{G}_A)/P(\mathbf{G}_B)$, based on the value of k_0.

5. Complex networks can be represented as adjacency matrix, adjacency list, and edge list.

   a. **[4 points]** Describe the graph in the picture as an adjacency matrix and as an igraph graph formula ($\triangleq$ adjacency list). Use the syntax of igraph (R or Python).



   b. **[5 points]** A sparse graph is a graph where the largest degree is bounded by a constant, say k. What is the time complexity of computing the degree of all nodes of a sparse graph when the graph is stored (1) as an edge list, (2) as an adjacency matrix? Provide upper and lower time complexity bounds in the Big O, respectively, Big Ω notation.

6. Hypergraphs are networks in which one edge can consist of more than two nodes. Such edges are called *hyperedges.* Let us consider now triplet-graphs, where hyperedges consists of triplets, that consist of exactly three different nodes. Assume the nodes in the edges are ordered but there is no repetition of a node in one edge.

   a. **[3 points]** What is the maximum number of triplet-graphs with m

hyperedges?

b. **[6 points]** Discuss an efficient way to (uniformly) randomly generate a random triplet-graph with m random hyperedges for a given set of n nodes (denoted by the indexes 1, …, n). What is the time complexity and space complexity of your algorithm in terms of m and n.

7. a. **[4 points]** Describe the algorithm that generates the invasion percolation cluster on $\mathbb{Z}^d$, $d \geq 2$.

   b. **[4 points]** Give three characteristic properties of the invasion percolation cluster.

   c. **[6 points]** For each of the three properties, provide a heuristic explanation.

8. Given a generic binary undirected graph, let k_i denote the degree of vertex i and k_i^{nn} the arithmetic average of the degrees of the vertices connected to vertex i. Imagine that you produce a scatter plot where each vertex i is represented as a point with coordinates (k_i, k_i^{nn}) in the plane.

   a. **[4 points]** Describe what the scatter plot looks like in typical realizations of the Erdős-Rényi random graph model with n nodes and connection probability p. Explain your answer.

   b. **[4 points]** Describe what the scatter plot looks like in typical realizations of the canonical configuration model, as a function of the input degree sequence. Explain your answer.

   c. **[5 points]** Describe what the scatter plot looks like in different real-world networks, and what can be concluded from it about the assortativity of these networks.

9. Consider a star-graph with n nodes $V = \{v_1, \ldots, v_n\}$, a central node v_1 and $n-1$ edges $E = \{(v_1, v_2), (v_1, v_3), \ldots, (v_1, v_n)\}$. Consider the SI model of epidemiology. At time $t = 0$ the central node v_1 gets infected, while all the other nodes are in the susceptible state.

   a. **[4 points]** Describe the **generator matrix** of the process. Note that, due to symmetry, the number of infected nodes can be used to represent the state of the graph.

   b. **[3 points]** What is the average time of a Continuous Time Markov Chain process on the graph until it infects every node in the star network. The initial state at time t_0 is the state where the first node just got infected. Assume $n > 1$ and a contagiousness (infection rate) λ.

   c. **[3 points]** What is the closeness centrality of the central node (v_1) in a star graph? What is the closeness centrality of a peripheral node (e.g., v_2) in terms of n?