Complex Networks

Teachers: D. Garlaschelli, F. den Hollander, A. Plaat.
Written re-examination: Wednesday, 27 January 2016, 10:00–13:00.

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

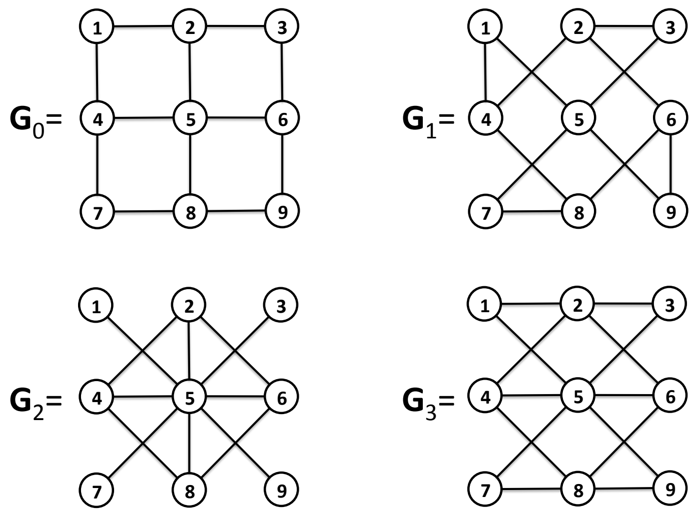
Answer each question on a separate sheet. Put your name, student number and the number of the question you are answering on each and every sheet. Provide full explanations with each of the answers.

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 homework assignments and 70% for this exam.

Success!

1. [**5 points**] Summarise what is known about the collaboration graph for mathematicians.
2. [**8 points**]
   Consider the Erdős-Rényi random graph $\mathrm{ER}_n(p)$ with n vertices and parameter $p \in (0,1)$.
   - 2a. Compute the average number of open paths of length 2 between two given vertices.
   - 2b. Compute the average number of open triangles containing two given vertices.
   - 2c. Explain the relation between the answers to questions 2a and 2b.
3. [**8 points**]
   - 3a. Draw the possible outcomes of $\mathrm{PA}_3(1,0)$, the preferential attachment random graph after three iterations, with 1 edge per incoming vertex and bias parameter 0.
   - 3b. What are the probabilities of each outcome?
   - 3c. Are all outcomes simple?
4. [**16 points**]
   - 4a. Consider the graph $\mathbf{G}_0$ shown in the figure below. Which of the three graphs $\mathbf{G}_1$, $\mathbf{G}_2$, $\mathbf{G}_3$ can be obtained from $\mathbf{G}_0$ via iterations of the Local Rewiring Algorithm (LRA)? Denote this graph by $\mathbf{G}'_0$, and specify the pairs of edges to which the LRA should be applied to go

transform $\mathbf{G}_0$ into $\mathbf{G}_0'$.



4b. Assume that the graph $\mathbf{G}_0$ is modeled using the canonical version of the Configuration Model (CM), with connection probabilities $p(x_i, x_j) = x_i x_j/(1 + x_i x_j)$ for $i \neq j$ and $p_{ii} = 0$. Also assume that, after applying the Maximum Likelihood Principle, the hidden variables turn out to be $\{x_i^*\}$ where $x_3^* = a$, $x_4^* = b$, $x_5^* = c$ (with $a, b, c > 0$). Find the correct ordering of a, b, c from the largest to the smallest and write, as a function of a, b, c alone, the probabilities $p(x_1^*, x_8^*)$, $p(x_7^*, x_9^*)$, $p(x_2^*, x_6^*)$, $p(x_5^*, x_7^*)$, $p(x_9^*, x_5^*)$.

4c. Now assume that the graph $\mathbf{G}_0'$ is modeled using the canonical version of the CM, and let $\{x_i'\}$ denote the hidden variables obtained using the ML principle in this case. What relation exists between $\{x_i'\}$ and $\{x_i^*\}$ found previously in 4b? Write, as a function of a, b, c alone, the probabilities $p(x_8', x_2')$, $p(x_1', x_5')$, $p(x_9', x_1')$, $p(x_5', x_8')$, $p(x_6', x_9')$. Order these probabilities from the largest to the smaller.

4d. Identify the pairs of vertices having maximum probability of being connected under the CM in $\mathbf{G}_0$ and $\mathbf{G}_0'$. In which of the two graphs are the edges connecting these pairs of vertices realized?

4e. Calculate the nearest neighbour degree of node 5 in $\mathbf{G}_0$ and in $\mathbf{G}_0'$. Which of the two graphs has the largest value of this quantity?

4f. Consider the shortest paths between all pairs of vertices. Find the longest of such paths in $\mathbf{G}_0$ and in $\mathbf{G}_0'$. Which of the two graphs has the "longest shortest path"?

5. **[8 points]**

Of an undirected graph the following edge list is given:

v_1	v_2
1	2
2	3
2	4
2	5
2	6
4	6

5a. Draw the graph.

5b. Provide the adjacency matrix of this graph in algorithmic Pseudo-code.

5c. What is the variance of the number of edges of the vertices? Provide an algorithm to compute this number.

5d. What is the space complexity of this algorithm? Why?

5e. Is the mean of the number of edges of 100 random graph different from 10 graphs? What about the variance? Why?

6. [**12 points**]

6a. Assume you are given an algorithm for the Configuration Model that is based on an adjacency matrix. Provide an algorithm that detects self-loops and multi-edges. Describe any assumptions, imperfections or limitations of your program.

6b. What is the computational time complexity of the program?

7. [**10 points**]

7a. Consider invasion percolation on the triangle. Draw all the possible outcomes of the invasion percolation process $I(n)$, $n = 0, 1, 2$.

7b. Consider invasion percolation on $\mathbb{Z}$. What is C_{IPC}, the invasion percolation cluster? Motivate your answer.

7c. Consider invasion percolation on the binary tree. Do you expect that all sites will eventually be invaded? Motivate your answer.

8. [**5 points**]

8a. What does it mean that the contact process on $\mathbb{Z}^d$ with parameter $\lambda \in (0, \infty)$ is attractive.

8b. What important consequence does attractiveness have for the limiting behaviour?

8c. Give a heuristic explanation why the contact process on $\Lambda_N = \mathbb{Z}^d \cap [0, N)^d$, $N \in \mathbb{N}$, with periodic boundary conditions, exhibits a dichotomy in its average extinction time behaviour in the limit as $N \to \infty$.

9. **[16 points]**

   9a. Draw a simple undirected graph consistent with the degree sequence $\vec{k} = (2, 2, 4, 2, 2)$. Let $\mathbf{G}$ denote your graph.

   9b. Now consider the Erdős-Rényi (ER) random graph with probability p as a model for $\mathbf{G}$. Find the value p^* that maximises the likelihood to generate $\mathbf{G}$.

   9c. Calculate the expected value of the degrees of all vertices under the ER model. For each vertex, determine whether the realized degree in $\mathbf{G}$ is larger or smaller than the corresponding expected degree.

   9d. Calculate the standard deviation of the degrees of all vertices under the ER model. Identify the vertices whose realized degree in $\mathbf{G}$ differs from the expected degree by more than one standard deviation.

   9e. Calculate the probability of occurrence of $\mathbf{G}$ under the ER model.

   9f. Rewire some edges of $\mathbf{G}$ in such a way that the degree sequence is preserved. Let $\mathbf{G}'$ denote the new graph. Recalculate p^* for $\mathbf{G}'$ and calculate the probability of occurrence of $\mathbf{G}'$ under the ER model.

10. **[6 points]**

   10a. Provide an algorithm to compute the Empirical Average Local Clustering Coefficient. Assume as data structure an adjacency *matrix*.

   10b. What is the computational time complexity of the program? Why?

11. **[6 points]**

   11a. Provide an algorithm to compute the Empirical Average Local Clustering Coefficient. Assume as data structure an adjacency *list*.

   11b. What is the computational time complexity of the program? Why?

Wikipedia's entry on Pseudo-code says the following:

It is an informal high-level description of the operating principle of a computer program or other algorithm.

It uses the structural conventions of a programming language, but is intended for human reading rather than machine reading. [...] The purpose of using Pseudocode is that it is easier for people to understand than conventional programming language code, and that it is an efficient and environment-independent description of the key principles of an algorithm. It is commonly used in textbooks and scientific publications that are documenting various algorithms.

Whenever this exam asks to "provide an algorithm," you are requested to do so in a programming language that is close to Pyhton, Java, C, or C++. Here "close" means that we are interested in the algorithmic essence. No points will be deducted for simple syntactic omissions such as missing semi-colons. The algorithmic meaning, however, should be clear beyond doubt. Do provide declarations of essential variables.