Complex Networks

Teachers: D. Garlaschelli, F. den Hollander, A. Plaat.
Written re-examination: Thursday, 26 March 2015, 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 each and every sheet. Provide full explanations with each of the answers. Simply repeating formulas from the lecture notes without explaining their background is not sufficient.

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: 40% for homework assignments and 60% for this re-exam.

Success!

1. [**5 points**] Give 4 examples of real-world networks, each with a brief description.
2. [**7 points**]
   - 2a. What is the definition of the degree distribution of a graph G?
   - 2b. Given a sequence $\mathcal{G} = (G_n)_{n\in\mathbb{N}}$ of random graphs, when is $\mathcal{G}$ called small world?
   - 2c. Given a sequence $\mathcal{G} = (G_n)_{n\in\mathbb{N}}$ of random graphs, when is $\mathcal{G}$ called scale free with exponent $\tau \in (1,\infty)$?
3. [**8 points**]
   - 3a. Give the definition of the preferential attachment model $\mathrm{PA}(1,\delta)$ with $n \in \mathbb{N}$ vertices and parameter $\delta \in (-1,\infty)$. Except for the self-loop created at the first iteration step, is the resulting random graph always simple or not?
   - 3b. Give the formula for $\lim_{n\to\infty} f_{n,\delta} = f_\delta$ when $f_{n,\delta}$ is the degree distribution of $\mathrm{PA}(1,\delta)$.
   - 3c. Compute the exponent τ of f_0 with the help of Stirling's formula: $n! \sim n^n e^{-n}\sqrt{2\pi n}$, $n \to \infty$.

4. **[16 points]** Given a real-world graph with adjacency matrix

$$\mathbf{A}_1 = \begin{pmatrix} 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{pmatrix},$$

consider an ensemble of networks built on the Chung-Lu version of the configuration model of the graph.

4a. Write the probability matrix $\mathbf{P} = (p_{ij})$, where p_{ij} is the probability of connection of nodes i and j. Note: put $p_{ii} = 0$ for all i.

4b. In this particular case, is the Chung-Lu model giving an acceptable probability for all i and j?

4c. Write the probability $P(\mathbf{A})$ of a generic graph with adjacency matrix $\mathbf{A}$, with $\{p_{ij}\}$ as parameters.

4d. Use the previous result to calculate the probability of occurrence of the two graphs with adjacency matrices

$$\mathbf{A}_2 = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{pmatrix}, \qquad \mathbf{A}_3 = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix}.$$

4e. The graphs described by $\mathbf{A}_2$ and $\mathbf{A}_3$ have the same degree sequence. Are the probabilities $P(\mathbf{A}_2)$ and $P(\mathbf{A}_3)$ the same? Why?

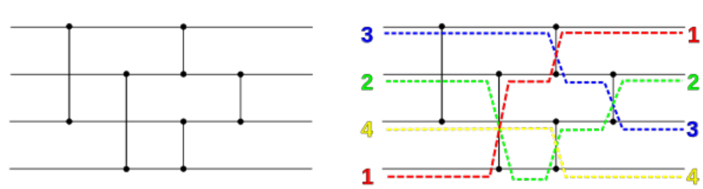
4f. Draw the most likely graph generated by the model.

5. **[3 points]** Dijkstra computes the single-source shortest path length, Girvan-Newman is a community detection algorithm, while Pagerank computes a centrality measure.

5a. Describe how shortest path, community detection and centrality measure are related.

5b. Can one be used as part of the other? Which ones? How?

6. **[9 points]** You are given the following sorting network:



6a. You are given the following sequence of numbers: 4, 2, 6, 5. Sort them by using the network. List for each step the comparisons and the intermediate sequence.

6b. You are given the following seqence of numbers: 4, 2, 6, 5, 7. You try to sort them by using the network. What happens? What can you do to address this issue?

7. [**8 points**]

7a. Describe the construction of the invasion percolation cluster C_{IPC} on $\mathbb{Z}^d$.

7b. What percentage of $\mathbb{Z}^d$ is covered by C_{IPC}?

7c. How is it possible that C_{IPC} "looks like" the cluster of ordinary percolation at criticality, even though it has no parameter?

8. [**7 points**] Consider the contact process with parameter $\lambda \in (0, \infty)$ on the graph G consisting of 2 vertices connected by 1 undirected edge. Let X_t denote the number of infected sites at time t. Then $X = (X_t)_{t \geq 0}$ is a continuous-time Markov process with state space $\{0, 1, 2\}$.

8a. Write down the transition rates of X.

8b. Compute the expectation $\mathbb{E}(\tau_0 | X_0 = 2)$ when $\tau_0 = \inf\{t \geq 0\colon\ X_t = 0\}$ is the time until the infection dies out. *Hint:* The expected time X stays at $i \in \{0, 1, 2\}$ equals 1 divided by the sum of the transition rates out of i.

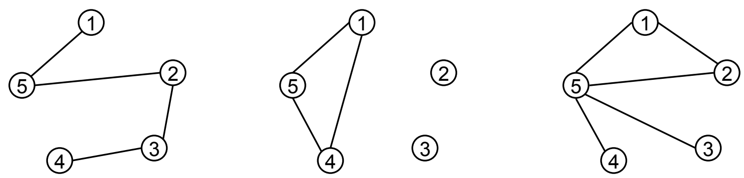
9. [**16 points**] A real social network with N nodes (actors) and L links (friendship relations) is modelled as an (undirected) Erdős-Rényi random graph with probability p^*, where the star indicates the value that maximises the likelihood of the model, given the data. Based on this model, the expected average nearest-neighbour degree is estimated as $\langle k^{nn} \rangle \approx 90$, and the expected average clustering coefficient as $\langle C \rangle \approx 0.6$.

9a. Determine the total number N of actors, and the probability p^* that any two actors are friends of each other.

9b. Determine the expected number $\langle k \rangle$ of friends of each actors, and the total expected number $\langle L \rangle$ of friendships.

9c. Assume that p^* remains equal to the value calculated above even when more and more actors are considered, i.e., for arbitrarily large N. What can be concluded about the connected components in the network model as $N \to \infty$?

9d. Consider the following three possible configurations for the subgraph involving the first 5 actors only:



Subgraph 1 *Subgraph 2* *Subgraph 3*

Calculate the probability of occurrence for each of the three possible subgraph configurations.

9e. Now draw the most probable configuration for the subgraph involving the first 5 actors only, and calculate its probability.

9f. Write down a general expression for the probability of occurrence of a subgraph with n nodes and l links.

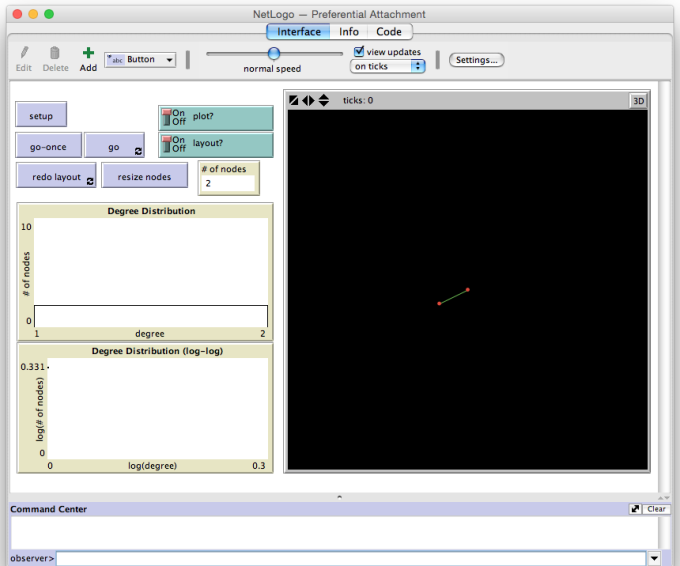
10. **[11 points]** Create a System Dynamics diagram of inner city crime. Situation: the general public complains about inner city crime levels. Shop lifting, pickpockets and drug use create a sense of insecurity among the shoppers, the shopkeepers and the general public. The city council wants to act. You are asked to analyse the situation with the help of a system dynamics approach. Use the following entities:

    - local tax level
    - number of police
    - popularity of the council
    - city budget
    - attractiveness of the city
    - crime level
    - crime complaints

    Hint: The lecture notes contain a list of guidelines. Points are awarded for correctness, clarity, and loops.

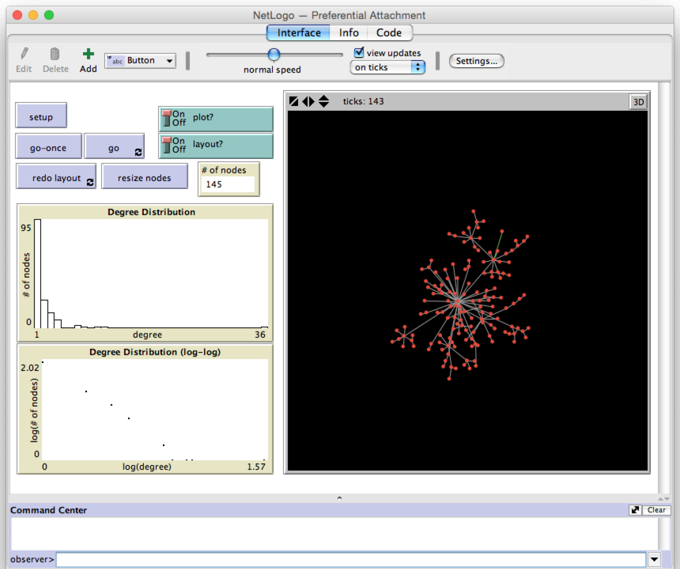
11. **[10 points]**

    11a. The Netlogo example application Preferential Attachment has the following initial state (after hitting Setup):



Explain what we see: What are the line and dots, and what are the preferential attachment rules that govern the dynamic process?

11b. After hitting the Go-once button a few times some structure is starting to appear, and after some time the system stabilises in a dynamic situation, in the sense that new lines and dots appear, but the general structure of the picture stays the same. The following screenshot was taken:



How does the macro structure follow from the micro rules. Why is the end result not a regular random graph?