INSTITUTO SUPERIOR TÉCNICO | UNIVERSIDADE DE LISBOA
COMPLEX NETWORKS 2017/18

Solution of the 1st Exam | Wednesday, 17 January 2018 (2h)

This exam can be answered in Portuguese or in English in a standard IST exam sheet.
All exercises are equally valued.

1. a) Suggest a network measure able to assess the number of people you have to communicate with (on average) to reach a complete stranger in a social network, or, for instance, the average number of clicks needed to lead you from one website to another.

b) Suggest an algorithm for your network measure (considering directed, unweighted graphs) and indicate its complexity (indicate the network representation you assumed).

R: a) Average path length (APL). APL is the average distance (= length of shortest path) between all pairs of nodes in the network. b) It may be calculated through a BFS starting from every node. Assuming a linked list, it takes O(V(V+E)), where V is the number of vertices and E the number of edges.

2. Network science offers several ways of measuring the relevance of a node for different purposes and contexts. Briefly describe the differences in goals of the following three centrality measures: *Closeness Centrality* (CC), *Page Rank* (PR), and *Betweenness Centrality* (BC).

R: Closeness centrality tries to rank nodes in terms of their distance to all other nodes: it is calculated as the sum of the length of the shortest paths between the node and all other nodes in the graph. Betweenness centrality quantifies the number of times a node acts as a bridge along the shortest path between two other nodes. Let's think on a telecommunications network. A node with higher betweenness would have more importance and control because more information will pass through that node. Page Rank (the same applies to Eigenvector and Katz centralities) tries to rank the importance of nodes based on the importance of their peers. It assigns scores to all nodes such that connections to high-scoring nodes contribute more than equal connections to low-scoring nodes.

3. One of the main goals of computational and mathematical epidemiology is to model the establishment and spread of pathogens. A predominant method of doing so is to abstract the population into compartments, which represent their health status with respect to the pathogen in the system. In **Fig. 1**, we show a typical time-evolution of one of these classic models.

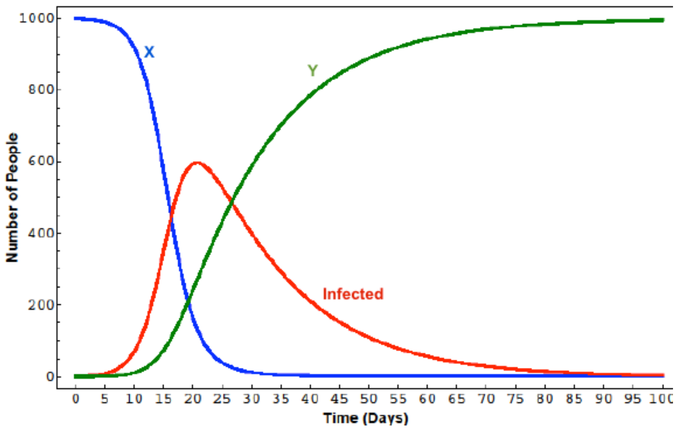

Figure 1 - Typical time-evolution of a famous compartment model in epidemiology. Can you guess it? (see question 3).

a) Indicate one possible model underlying the dynamics described in **Fig. 1**, and the compartments associated with *types X* and *Y*.

b) The basic reproduction number (often denoted by R_0) is a key measure used to determine whether or not an infectious disease can spread through a population. By looking at **Fig. 1**, what can you say about R_0? Justify.

R: a) SIR model: Susceptible(X) - Infected - Recovered (Y). b) R_0 of an infection can be thought of as the number of cases one case generates on average over the course of its infectious period, in an otherwise uninfected population. Clearly, we have an outbreak of the disease. Thus, $R_0>1$.

4. Let us consider the small-world model proposed by Watts and Strogatz in 1998. Describe the dependence of the *i)* average path length and *ii)* clustering coefficient on the rewiring probability *p*, and use it to discuss the main contribution of the model.

R: The WS model tries to understand how networks that are strongly locally structured (high clustering coeff.) are, at the same time, small worlds (low APL). Watts & Strogatz invented a very simple model which interpolates between regular and random graphs. Starting from a regular (ring) graph, each edge is randomly rewired with a probability p. The average path length falls very rapidly with increasing p, quickly approaching its limiting value. Differently, in the intermediate region the clustering coefficient remains quite close to its value for the regular lattice, and only falls at relatively high p. This model illustrates a very simple principle: it takes a lot of randomness to ruin clustering, but a very small amount to overcome locality.

5. Finding the degree distribution is often part of analyzing the properties of a network.

   a) What is a degree distribution of an undirected graph?

   b) Plotting heterogeneous degree distributions is often a difficult task. Suggest two usual procedures to properly plot and extract the main properties of power-law degree distributions.

   R: a) Degree distribution d(k) is the probability distribution of the degrees over the whole network. The process starts by computing the number of number N_k of nodes with degree k. From N_k we calculate $d(k)=N_k/N$. b) Use log-log plots; use cumulative degree distributions; avoid linear binning (e.g., use logarithmic binning).

6. A scale-free network may be defined as a network whose degree distribution follows a power law. Describe in detail the steps of an algorithm capable of creating a network with a degree distribution of the form $d(k) \sim k^{-3}$ and an average degree $\left\langle k \right\rangle = 4$.

   R: Barábasi-Albert model. This algorithm combines growth and preferential attachment. The network begins with an initial connected network of m_0 nodes. New nodes are added to the network one at a time. Each new node is connected to $m \le m_0$ existing nodes with a probability that is proportional to the number of links that the existing nodes already have, such that the probability p_i that a new node connects with a node i of degree k_i is given by $p_i = k_i \Big/ \sum_j k_j$, where the sum is made over all existing nodes j. The degree distribution resulting from the BA model is a power-law $d(k) \sim k^{-\gamma}$, with an exact exponent of $\gamma = 3$, and an average degree $\left\langle k \right\rangle = 2m$. Thus one should have $m = 2$.

7. a) Explain what *degree assortativity* of a network is (also known as *degree correlations*)?

   b) Explain one possible way of measuring the *degree assortativity* of a network.

   R: Degree assortativity is the preference of network's nodes to attach to others that have a similar degree. It can be measured through degree correlation function $k_{nn}(k)$ which is the average degree of the neighbors of all degree-k nodes. For neutral networks k_{nn} is independent of the degree. For assortative (dissassortaive) networks, $k_{nn}(k)$ grows (decreases) with k, following a power-law of the form $k_{nn}(k) \sim k^u$. Thus, if u>0 we have an assortative net; if u<0 we get a disassortative net.

8. The *Prisoner's dilemma* is one of the most famous dilemmas of cooperation. In its simplest form, it considers symmetric pairwise interactions involving costs and benefits. Cooperation (*C*) corresponds to offering the other player a benefit *b* at a personal cost *c*, with $b > c$. Defection (*D*) means offering nothing.

   a) Propose a payoff matrix associated with the problem and justify why mutual Defection is the only Nash Equilibrium.

   b) Let us now consider that there is an expected punishment $K > c$ of those opting for defection against a cooperator. Consider a well-mixed population of individuals that interact following this game of cooperation with punishment. Indicate the payoff matrix associated with this setting and find the evolutionary stable strategy/strategies of this dilemma. As an example, you may consider $b = 1.5$, $c = 1$ and $K = 2$.

   R: To understand the dilemma at stake let us write down the payoff matrix:

	C	*D*
C	*b-c*	*-c*
D	*b-K*	*0*

a) *For K=0 (see (b)), (D, D) is the only Nash Equlibrium. Irrespectively of what you opponent does, the best option is always to Defect (b>b-c and 0>-c). If both players opt for D, none of the two will have any reason to switch to C, even both would be better with mutual cooperation.*

b) *An Evolutionary Stable strategy ESS is a strategy that, if adopted by a population in a given environment, cannot be invaded by any alternative strategy that is initially rare. Since a rare mutant choosing C will be always disadvantageous with respect to a resident population of Ds, then D will be an ESS. However, for K>c, the same reasoning applies to resident populations of Cs. Thus, both C and D are evolutionary stable strategies in the presence of punishment. In other words, punishment transforms a prisoner's dilemma into a stag-hunt game.*

   This answer is fine. You may also wish to show it formally assuming the replicator equation

$$G(x)=\dot{x}=\frac{dx}{dt}=x(1-x)\left(f_C(x)-f_D(x)\right),$$

   where x is the the fraction of Cs in the population. The average fitness of Cs and Ds is given by $f_C(x)=x(b-c)+(1-x)(-c)$ and $f_D(x)=x(b-K)$. *There's one internal fixed point for* $f_C(x)-f_D(x)=0 \Rightarrow x^*=c/K$. *Since,* $f_C(x)-f_D(x)$ *is a monotonous increasing function,* x^* *must be unstable. Please find on your right an illustration of the gradient G(x) for the parameters given.*