Instituto Superior Técnico | Universidade de Lisboa
Complex Networks 2016/17

Solution of the 1st Exam | Wednesday, 18 January 2017 (1.5h)

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

1. Consider a Erdős–Rényi (ER) network with N=4x10^3 nodes, connected to each other with probability p = 10^{-3}. Estimate the number of links and the average degree of the network.

R: $\langle E \rangle = p\frac{(N-1)N}{2} \approx 10^{-3}\frac{16\times10^6}{2} = 8\times10^3, \langle k \rangle = p(N-1) \sim 4$

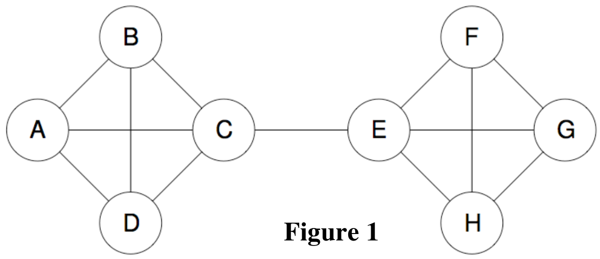


Figure 1

2. Briefly discuss the difference in aims of *betweenness centrality* and *closeness centrality* of a node of an (unweighted) undirected graph. Discuss also how they can be computed and the underlying computational complexity.

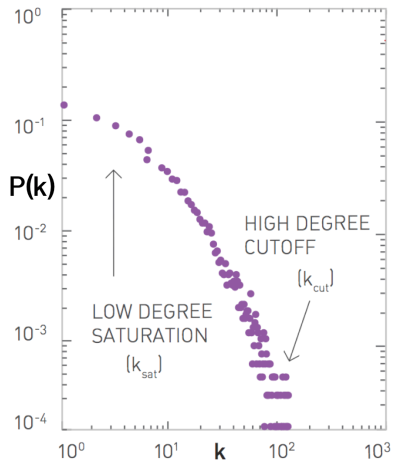
R: (one possible answer) Closeness centrality of a node A is the average length of the shortest path between A and all other nodes in a graph of V nodes and E edges. It can be computed through a BFS starting at node A, having a time complexity of O(V+E). Betweenness centrality quantifies the number or fraction of times a node acts as a bridge along the shortest path between two other nodes. It involves computing all shortest paths among all pairs of nodes, which can be also be computed through a BFS starting from every node, taking O(V(V+E)).

3. Consider the graph G in Fig. 1. Indicate the partition of G which maximizes modularity. Justify.

R: Modularity measures the quality of a given partition, by computing the difference between the network's real wiring diagram and the expected number of links within partitions if the network is randomly wired. In Fig. 1 it is clear that the partitions [ABCD][EFGH] maximizes modularity.

4. Briefly explain the origin of the term scale-free when applied to complex networks with power-law degree distributions.

R: (one possible answer) For a large network ($N \to \infty$) with a power-law degree distribution ($P(k) \sim k^{-\gamma}$) with γ < 3 the first moment of the distribution is finite but the second moment is infinite. The divergence of **<**k^2**>** *(and of σ_k) for large N indicates that the fluctuations around the average can be arbitrary large. This means that when we randomly choose a node, we do not know what to expect: The selected node's degree could be tiny or arbitrarily large. Hence networks with γ<3 do not have a meaningful internal scale, but are "scale-free".*



5. In real systems we rarely find perfect power-laws (see Fig. 2). Often one observes a flattened degree distribution $P(k)$ for low degrees (i.e., $k<k_{sat}$), and high-degree cutoffs (rapid drop in $P(k)$ for $k>k_{cut}$). Suggest a cause for the emergence of each feature.

R: High degree cutoffs may emerge from various factors: Finite size effects, ageing of nodes or node/link removal, maximum degree of nodes (e.g., airports), etc. Low degree saturation may emerge from death of low degree nodes, preferential attachment with initial attractiveness, etc.

6. Propose an algorithm capable of generating a network portraying a power-law degree distribution with a high prevalence of triangular motifs (high clustering coefficient).

R: Consider a growing network model where at each time-step a new vertex is added. Each new vertex attaches to both ends of a randomly chosen edge. As such, this rule favors the creation of triangular relations between individuals, thereby greatly enhancing the cluster coefficient. While choosing a random edge, one is implicitly promoting the preferential choice of highly connected nodes, just like the Barabási-Albert model, leading to the same degree distribution. This model has been proposed by Dorogotsev et al. in 2001.

7. Consider a large-scale simulation of an epidemic outbreak in a networked population (e.g., consider an SIS model). Discuss how the representation of the network alters the running time of your simulation.

R: At each time step, for each vertex X, one needs to access the states of each neighbor of X. If the network is represented by adjacency matrix, it is necessary to loop over each row of the matrix to find the states of each neighboring vertex, an operation which scales with the total number of vertices. Differently, an adjacency list allows an immediate access to the list of the neighbors of a given vertex. Thus, while the update of a node scales with V in the first case, in the second case it scales with the degree of the node.

8. Consider the SIS model and that pathogens spread on a network. Discuss two vaccination strategies likely to perform better than random vaccination.

R: In heterogeneous networks a virus can be eradicated by increasing the epidemic threshold through hub immunization. The more hubs are immunized, the smaller is the effective maximum degree, and the larger is the epidemic threshold, increasing the chance that the disease dies out. While effective, this technique involves a complete knowledge of the network. Alternatively, one could start by vaccinating a set A of randomly chosen nodes. Then, select random partners of the nodes in A and vaccinate them. It can be shown that by picking partners of random chosen nodes it is highly probable to reach highly connected nodes, as if we had complete information about all node's degrees.

9. You have probably come across *Golden Balls*, a British TV daytime game show. This game involves a segment called "Split or Steal" in which two participants make the decision to "split" or "steal" to determine how the final jackpot is divided. If both contestants choose to split, the jackpot is split equally between them. If one contestant chooses to split and the other chooses to steal, the stealer gets all the money and the splitter gets nothing. If both contestants choose to steal, they both get nothing. One may also assume that the individual success is also influenced by an additional non-monetary factor representing the disutility derived from losing, as a result of being cheated by the other player (when a player is the only one "splitting"). a) Propose one payoff matrix for this 2-player, 2-strategy game, and indicate the associated *Nash equilibrium* or *equilibria*. Justify. *b)* Let us now consider the evolution of a population of individuals that interact following the payoff matrix indicated in (a). Identify the evolutionary stable strategy or strategies of this dilemma.

R: To understand the dilemma at stake one can draw the following payoff matrix, which is in all aspects equivalent to a Prisoner's dilemma.

	Split	*Still*
Split	*50 (half of the pot)*	*-1 (disutility for being cheated)*
Steal	*100 (entire pot)*	*0 (nothing)*

a) (Steal, Steal) is the only Nash Equlibrium. Irrespectively of what you opponent does, the best option is always to Steal. If both players opt for Stealing, none of the two will have any reason to switch to split, even if this equilibrium is not ideal for any of the two players.

b) The same type of argument can be used to understand why Steal is the only evolutionary stable strategy (ESS). An ESS is a strategy which, 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 Split will be always disadvantageous with respect to a resident population of Steal strategists, then Steal will be an ESS. Split is not an ESS since a rare Steal strategist will always be advantageous.