Instituto Superior Técnico | Universidade de Lisboa

Network Science 2018/19

1st Exam | Wednesday, 16 January 2019 (2h), 18h30

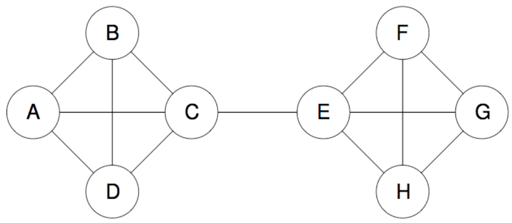
The exam can be answered in Portuguese or in English in a standard IST sheet. All exercises are equally valued.

1. Given the centrality measures that you know, which one would you use to identify the users in a social network that are candidates to control the flow of information? Justify.

R: Betweeness centrality (BC). The BC of a vertex is the number of shortest paths that pass through the vertex. It represents the degree of which nodes stand between each other. For example, in a telecommunications network, a node with higher BC would have more control over the network, because more information will pass through that node. One may apply this measure to a wide range of problems in network theory where information spreading is important, including problems related to social networks, biology, transport and scientific cooperation.

2. Consider the graph *G* pictured on your right. Indicate the partition of *G* that 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 (the null model). In Fig. 1 it is clear that the partitions [ABCD][EFGH] maximize modularity.



3. The Barabási–Albert (BA) model is likely the most famous algorithm for generating scale-free networks. Describe the BA algorithm and indicate the values you would adopt to have a network with $N=10^4$ nodes, a power-law degree distribution, and an average degree of $<k>=12$.

R: The Barabási–Albert (BA) algorithm combines growth and preferential attachment for generating scale-free networks. 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 $m<m_0$ existing nodes with a probability that is proportional to the number of links that the existing nodes: the probability p_i that a node i with a degree k_i receives a link from the new node is given by $p_i = k_i / \sum k_j = k_i/(2E)$. The average degree is given by 2m, such that, for having a SF network with <k>=12 one needs to adopt m=6. The algorithm stops when the desired number of vertices is achieved.

4. Discuss how the representation of the graph (e.g., adjacency lists and matrices) alters the complexity of BFS and DFS algorithms used to compute the measures of centrality recurrent in network science.

R: In both cases, the runtime depends on how long it takes to iterate across the edges of a given node. With an adjacency list, the runtime is directly proportional to the number of edges. Since each node is visited once, the cost is the number of nodes plus the number of edges, which is O(N+E). With an adjacency matrix, the time required to find all outgoing edges is O(N) because all N columns in the row for a node must be inspected. Summing up across all n nodes, this works out to $O(N^2)$.

5. Immunization strategies specify how vaccines, treatments or drugs are distributed in the population. Often cost considerations, the difficulty of reaching all individuals at risk, and real or perceived side effects of the treatment prohibit full coverage. Briefly justify why random vaccination is a bad idea in scale-free contact networks, and suggest a better alternative.

R: Taking the SIS model as a suitable example, one can compute the expected fraction g_c of immune individuals needed to escape an endemic state, and how it depends on the degree distribution of the network, on the transmition rate β, and on the recovery rate δ: $g_c = 1 - \frac{\delta \langle k \rangle}{\beta \langle k^2 \rangle}$. Since for scale-free with γ < 3 the second moment $<k^2>$ diverges for large networks (N→∞), we would need to immunize virtually all nodes to stop the epidemic (g_c=1). This prediction is consistent with the finding that for many diseases (e.g., measles, digital viruses, etc.) we must immunize 80%-100% of the population to eradicate the pathogen. As an alternative, immunizing the hubs drastically reduces the second moment of the degree distribution of the network on which the disease spreads, making the hubs invisible to the pathogen, leading to an effective halt of the epidemics with a limited fraction of immune individuals. The problem with a hub-based immunization strategy, however, is that for most epidemic processes we lack a detailed map of the contact network. Indeed, for instance, we do not know the number of sexual partners each individual has in a population, nor can we accurately identify the super-spreaders during an influenza outbreak. As an alternative, it can be shown that a strategy based on immunization of acquaintances partially solves this problem. The idea is to *vaccinate the acquaintances of a randomly selected individual, indirectly targeting the hubs without having to know precisely which individuals are hubs.*

6. Let $G=(V, E)$ be an undirected and acyclic graph with more than 4 vertices. Then:

   a. All vertices have local clustering coefficient equal to 1.
   b. There is at least one vertex with local clustering coefficient 1.
   c. The clustering coefficient of G is 0.
   d. The clustering coefficient of G is 1.

R: c

7. Provide an intuition for the observed resilience of scale-free networks against random failures and justify the famous *Achilles' Heel* of these topologies.

R: One can analyze the robustness of a graph as an inverse percolation problem. It can be shown that the critical fraction threshold follows $f_C = 1 - \frac{1}{\langle k^2 \rangle / \langle k \rangle - 1}$. *For scale-free networks with* $\gamma < 3$ *the second moment diverges in the* $N \to \infty$ *limit. If we insert this limit in the previous expression, we find that* $f_c = 1$ *for large networks. This means that to fragment a scale-free network we must remove all of its nodes. In other words, the random removal of a finite fraction of its nodes does not break apart a large scale-free network. The Achilles' Heel of SF networks relies on its low resilience against targeted attacks. Let us remove first the highest degree node, followed by the next highest degree, and so on. In this case, it is sufficient to remove only a few hubs to break a scale-free network into disconnected components, as the* 2^{nd} *moment of the degree distribution no longer diverges for large N.*

8. Direct and indirect reciprocity have been proposed as mechanisms for the evolution of cooperation. *i*) Suggest one appropriate individual strategy (or action rule) for direct reciprocity and one for indirect reciprocity. *ii*) Explain what is a social norm in the context of indirect reciprocity.

R: i) Direct reciprocity: one possibility is the famous Tit-for-tat (first cooperate, and then subsequently replicate an opponent's previous action). Indirect reciprocity: one possible choice could be a Discriminator (cooperate with those with a good reputation, and defect with those with a bad reputation). ii) Indirect reciprocity is based on reputations assigned by a third party who observes each interaction. To do so, this observer has to rely on some sort of rule that defines which conditions lead to a good or a bad reputation. These rules are known as social norms.

9. Define graph Laplacian and detail one of its uses in network analysis.

R: The Laplacian of a graph is a matrix representation of a graph. Given a graph G with V vertices, its Laplacian matrix L is a V by V matrix given by L = D-A where D is the degree matrix (a diagonal matrix which contains information about the degree of each vertex) and A is the adjacency matrix of the graph (containing 1s or 0s and its diagonal elements are all 0s). The Laplacian can be used to calculate the number of spanning trees for a given graph (together with Kirchhoff's theorem), and for graph (spectral) partitioning, among other applications.

10. Please indicate whether each of the following statements is TRUE or FALSE.
    Note: For each wrong answer we discount a correct one.

    a) The Erdős–Rényi model is a graph generation model that illustrates the emergence of small-world properties, including high average path lengths and a low clustering coefficient.
    b) Let us consider a network that evolves in time through growth and a linear preferential attachment. Adding an initial attractiveness $A>0$ to each node generates a saturation for low degrees, i.e., we observe a decrease in the number of low degree nodes when compared with the same model with $A=0$.
    c) Computing the eigenvector centrality for all vertices can be solved in polynomial time.
    d) The Basic Reproductive Number, R_0, provides the number of individuals an infectious individual infects if all its contacts are susceptible.
    e) Complex contagion is the phenomenon in social networks in which multiple sources of exposure to (e.g.) an innovation are required before an individual adopts a change of behavior.

R: a: False; b: True; c: True; d: True; e: True