Complex Networks Exam (20 points)
December 19, 2019

All documents allowed. Read all directions carefully and write well-argued answers.
Try to be concise but precise.
Write your master specialty (e.g., Physics, Computer science, etc.) on your answer sheet.
You can write your answers in English or French.
Don't forget questions on the other side of the page

Read carefully the provided article: *A community-aware approach for identifying node anomalies in complex networks*, and answer the following questions.

1. (2 points) In section 1, page 2, the authors introduce nodes with a star-like ego-network as typical anomalies. Give examples of *heuristics* that you could use to detect such a star-like property. You can provide the name of a known heuristic (e.g., Adamic-Adar, Degree centrality, etc.), and/or propose your own formula. Comment your answer.

2. (3 points) In section 4, three methods for anomaly detection are introduced (Oddball, Embed, CADA). Each method corresponds to a different definition of what an *anomaly* means. Write a short description of the type of anomaly Oddball and CADA aim to find, and compare what they agree/disagree about. Give examples taken from real networks (e.g., social networks) of nodes that would be anomalous according to each of these definitions. (Note: If you have followed the class about embeddings, you can do the same for Embed method, for 1 bonus point).

3. (4 points) In section 4.2, the authors provide a definition of an anomaly score based on the community structure. The authors define g_i^j and g_i^* using natural language.

   - Draw a small example graph (10 nodes max) with communities, examples of vectors associated to some nodes, and corresponding scores.
   - Describe, formally and/or in natural language, the values that could be taken by this anomaly score.
   - Propose a more formal definition of g_i^j and g_i^* using the graph or matrix formalism.

4. (4 points) In section 5.2 and 5.3, the authors summarize the LFR benchmark, its parameters, and how they are set for the experiments.

   - For classic network models, we have discussed some key properties, in particular: average shortest path length, clustering coefficient, degree distribution. Comment on the impact of LFR parameters on those properties for generated networks.
   - Are there other relevant random network models that the authors could have used? If yes, what would be the difference?
   - Explain in details the effect of parameters γ_1,γ_2 and $k_{\max}$, and what are the consequences of setting them as in the paper.

5. (2 points) From the experimental section, it is clear that the anomaly detection problem has similarities to the node classification problem, with two classes being *anomalous/normal* nodes. Would node classification methods introduced in the class be adapted to solve the anomaly detection problem? Describe briefly a supervised method for anomaly detection, and explain the strenghts and weaknesses of such an approach compared with the method introduced in this paper.

6. (3 points) Give your opinion about this paper ($\approx 1/2$ page), in particular:
   - The strengths and weaknesses of the method itself
   - The quality and reliability of the experimental section
7. (2 points) Some networks are not organized in communities, but instead as *spatial networks*. Could you propose a variant of the method proposed in this article, designed to discover anomalous nodes in spatial networks?