|  |  |
| --- | --- |
| **[Names]** | **[Date]** |

**Graph Analysis Project:   
Study of the formation of one network**

*In this project, you will investigate two mechanisms driving the formation of a real network.*

*1/ The first mechanism should be related to node or edge attributes. You can study either assortativity or dissortativity on a node attribute (i.e., similar or dissimilar individuals tend to be connected), sociability or popularity on a node attribute (i.e., some individuals connect to more people or receive more connections), or the influence of another network variable (individuals connected through another network tend to be connected).*

*2/ The second mechanism should be related to an endogenous process: reciprocity, transitivity, or rich-get-richer effects…*

*Please leave the instructions in italic. Attach your code as a RmD file.*

1. **Problem Statement**

*Describe the network you are studying. What mechanisms are you interested in and why is it interesting or relevant to study them in this network? (max 200 words)*

1. **Expectations**

*Formulate your expectations about the causes and consequences of these mechanisms: Why do you expect these mechanisms to occur? How should these mechanisms affect the graph structure? (max 300 words)*

1. **Research design**

*Explain how to define a QAP test to find evidence for the first mechanism you are studying. Clearly define all your statistical variables with equations. (max 300 words)*

*Explain how to define a CUG test to find evidence for the second mechanism you are studying. Clearly specify the reference model you choose and define your test statistic with equations. (max 300 words)*

1. **Data collection**

*Describe your data. Which network did you select, how did you collect and store the data? (max 200 words)*

1. **Exploration and Analysis**

*Define centralities for the nodes in your graph, using the measure of your choice. Explain your choice, describe the distribution of these measures, and provide a graph visualization of these centralities. (max 200 words)*

*Identify communities in your graph, using the community detection algorithm of your choice. Explain your choice, describe the communities you found and provide a graph visualization of the communities. (max 300 words)*

*Perform the CUG test and the QAP test previously described. Report all test statistics and empirical p-values. (max 300 words)*

1. **Interpretation and conclusions**

*Summarize what you learned from your data exploration and statistical tests. Are they all in line with your expectations regarding the mechanisms you are studying? (max 300 words)*

*Discuss the limitations of this study and identify possible ways to improve or enrich it**. (max 200 words)*