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

**Graph Analysis Project:   
Study of the formation of several networks**

*In this project, you will investigate one mechanism driving the formation of several networks. You have two options:*

*Option 1/ The mechanism can 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).*

*Option 2/ The mechanism can 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 networks you are studying. What mechanism are you interested in and why is it interesting or relevant to study it in these networks? (max 200 words)*

1. **Expectations**

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

1. **Research design**

*Option 1/ IF your mechanism is related to attributes, 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.*

Option 2/ *IF your mechanism is endogenous, 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 networks did you select, how did you collect and store the data? (max 250 words)*

1. **Exploration and Analysis**

*Define centralities for the nodes in your graphs, using the measure of your choice. Explain your choice, describe the distributions of these measures in the different graphs, and provide graph visualizations (for each graph) of these centralities. (max 250 words)*

*Identify communities in your graphs, using the community detection algorithm of your choice. Explain your choice, describe the communities you found and provide graph visualizations (for each graph) of the communities. (max 350 words)*

*Perform the CUG test OR the QAP test previously described. Report all test statistics and empirical p-values. (max 200 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 mechanism you are studying? (max 300 words)*

*Discuss the similarities and differences found for your different graphs. Can you explain them? (max 300 words)*

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