# 3USU0 Networks Project Description

The goal of this project is to take the concepts you learned during the lectures and apply them to a real-world network. To this end, each group will be given a data set corresponding to an undirected real-world network. For every lecture there are several small assignments you as a group will have to do that involve this network. By the end of the course each group should hand in a small report that includes the worked out version of each assignment[1].

We will use Python Jupyter Notebooks[2] for working with the network data sets. These Notebooks provide a very simple way to run Python code and see the results. Basic tutorials can be found on the Canvas page of the course. A very basic installation guide is also available there and also included at the end of this document.

## Lecture N1: Networks are everywhere

**Assignment 1** *Display your network using pyvis[3]. Describe its structure. What interesting features can you see? How are these related to the data that your network represents?*

*Note that you might have to play a bit with the visualization settings to get a picture that is helpful.*

**Assignment 2** *Compute the following characteristics of your network:*

1. *Average degree*

2. *Average clustering*

3. *Average distance*

4. *Diameter*

*Explain the outcomes, based on the type of data that this network represents.*

## Lecture N2: Structure of real-world networks

**Assignment 3** *Compute the degree distribution of your network and plot it. In addition, compute the second and third moment of your degree distribution. Do you think you can call your network scale-free? Explain your answer.*

**Assignment 4** *Select 100 nodes at random, compute their distance and plot the histogram of the results. Combining this plot with the average distance you computed before, would you say that your network is a small-world? Explain your answer.*

---

[1] If you use your own Python code then please include the corresponding Jupyter Notebooks in the report.

[2] For more information see `https://www.python.org/downloads/` and `https://jupyter.org/`

[3] See Canvas for a tutorial on this.

## Lecture N3: Mathematical models for networks

**Assignment 5** *Would any of the models we have covered so far be a good model for the structure of your network? Explain your answer. Which characteristics of your network are captured by the model and which are not.*

**Assignment 6** *Use the configuration model to create a random network with the same degree sequence as your network. Compute the four characteristics:*

1. *Average degree*

2. *Average clustering*

3. *Average distance*

4. *Diameter*

*and compare them to those of the original network. What differences do you see? What does this tell you about the structure of your network?*

## Lecture N4: Spreading processes on networks

**Assignment 7** *Compute the epidemic threshold of your network for the SIS and SIR infection models. What does this information tell you about the behavior of infection spreading on your network?*

**Assignment 8** *Select 5 different nodes at random in your network. For each of these nodes, run 5 complex contagions with different values of $\alpha$. For each case, document whether there is a complete cascade or else, how large the number of infected nodes is.[4] Visualize the outcome. What does this tell you about the structure of your network.*

**Assignment 9** *Give an example of a relevant process on your network that could be modeled as an infection processes or a complex contagion. Explain your answer.*

## Lecture N5: Centrality and Assortativity

**Assignment 10** *Find the four nodes in your network with the highest:*

- *Closeness centrality score*

- *Harmonic centrality score*

- *Betweenness centrality score*

- *PageRank centrality score*

---

[4]Could you find a node so that there is a complete cascade? If so, how did you find it?

*For each of these nodes, explain what its highest centrality score tells you about the importance of the node in your network.*

**Assignment 11** *Visualize your network and find each of the four nodes from the previous assignment.[5] Can you explain why their centrality score is high, based on their position in the network and other properties?*

**Assignment 12** *Based on the data that your network represents. Which of the four centrality scores would be the best measure for importance of nodes in the network? Explain your answer.*

## Basic installation guide

The following is a very basic guide for setting up the Python environment and Jupyter notebooks on a computer (with either Windows or Linux) using Anaconda.

1. Download the anaconda installer at `https://www.anaconda.com/products/individual`

2. Run installer

3. Install pyvis
   On Windows:

   (a) Start Anaconda Navigator
   (b) Run Anaconda CMD.exe
   (c) Run: pip install pyvis

   On Linux: Run: pip install pyvis

   To start Jupyter notebooks simply start the Anaconda Navigator and Launch Jupyter Notebooks.

---

[5]Hint: you can increase the node size of these nodes so that you can easily spot them in the visualization.