Network analysis practice exercises

1. Create the network on Fig. 1 in Python using the list of edges (i.e. create a list with the edges and use add_edges_from function). Calculate the most important descriptive measures of the network.

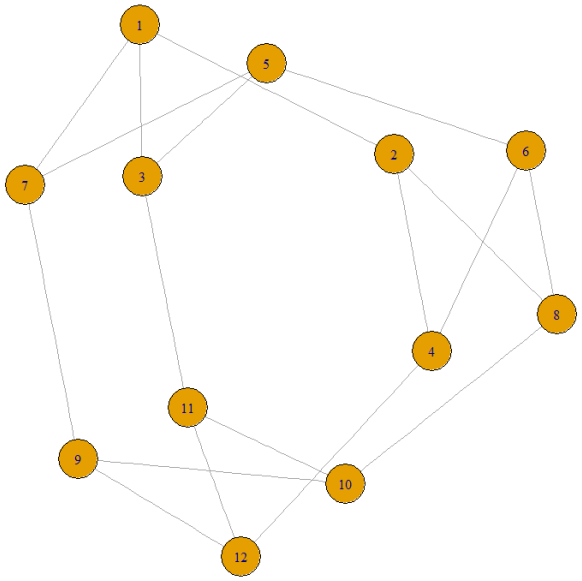

Figure 1: Network for Exercise 1.

2. In this exercise, you will have to work with the data in the file star star_wars.csv. The file contains the relationships between characters in the Star Wars movie. Your task is to identify the most central characters based on the provided network structure. (Note: the file contains two columns with the start and end node of links). Calculate the discussed centrality measures for the network (degree centrality, betweenness, closeness, PageRank). Discuss the results, first by comparing the 5 most central characters based on every method and examining the rankings (calculate the common characters in the list of top 5 characters for each pair of methods), secondly by calculating the correlations between the measures. Which methods are the most similar to each other in terms of the most central characters and in terms of correlation?

3. In the exercise session, you will have to use network data representing the friendship relationships of a group of registered users of a social network site. The adjacency matrix depicting a relationship among 300 users can be found in the file network_data.csv.

   Perform community detection as introduced in the lecture, including determining the optimal number of communities based on modularity.

   The 300 users spend some of their time and money on a specific game available on the social network platform. Additionally to the network, there are several variables recording demographic, general social network related and specific game related information on the users. The dataset (sngame.csv) contains the following variables:

(a) *user.id* : the identification code of the social network user

(b) *gender*: the gender of the user

(c) *age*: the age of the user

(d) *edu*: the highest education of the user

(e) *salary*: monthly salary of the user

(f) *sn.conn*: connections that the user has on the social network site

(g) *sn.min*: number of minutes spent on the social network site

(h) *game.min*: number of minutes spent on the social network site playing the game

(i) *game.purchase*: amount of money spent on purchasing extra features, upgrades, power-ups etc. in the game

Analyze the community structure obtained from the community detection method in terms of these attributes. Are the sub-communities different on some of the attributes? (Use groupby and some aggregation functions as in the lecture videos).