Assignment 5

Steven Raaijmakers Irene Vega Ramón

November 2019

Q A 1. 1. The average degree value is 2 which is a logical value since all girls are asked for two dining tables creating two out-connections for every girl.
2. Two persons have the highest in-degree being 6.
3. It represents how many persons have listed this person as a dining-table companion.
4. Irene, Ruth, Laura, Ella and Alice are the most isolated since they all have zero in-degree.
5. Eva and Marion have the highest in-degree and are the most popular among the other girls and therefore have the most influence.

2. 1. Yes it does.
2. Changing the color of nodes according to their in-degree.

3. 1. There are 11 different strongly connected components.
2. Since the graph has a few isolated notes (nodes which cannot be reached because they have no incoming connections), the graph as a whole will be weakly connected.
3. No, the colors do not give information about the isolation of nodes. However, if we take a closer look into a color that will only occur on one node, we can say it is isolated. If we look into the data laboratory we can see that this only happens 4 times, with the same girls.
4. Yes, we could use a higher opacity for nodes which have a higher in-degree and lower opacity for for nodes with lower in-degree creating shades of the color.

4. 1. Marion has the highest betweenness-centrality (BC) (79.833).
2. The BC measures how many times a node lies on the shortest path between any other node pair in the network while the in-degree measures the amount of incoming links.
3. The people with the highest BC, like Marion and Anna. They lie on the shortest path between many node pairs and therefore have a lot of influence on the communication between many pairs.

4. People with the lowest BC. The BC of the isolated nodes is 0 in this scenario. In other words, when their node + connections would be removed from the graph, no shortest path between any other two nodes would be affected.
5. To identify the most influential nodes we would use the BC in combination with closeness centrality (CC). The BC and CC values of each node have to be ranked an normalized (NR) in order to determine a new score via $0.8 \cdot NR_{BC} + 0.2 \cdot NR_{CC}$. In this way we not only take into account on how many shortest path a node is but also how far its distance is to other nodes.
   For isolated nodes it would be best to measure the in-degree.
   See Figure 1 for the graph visualization of this network.

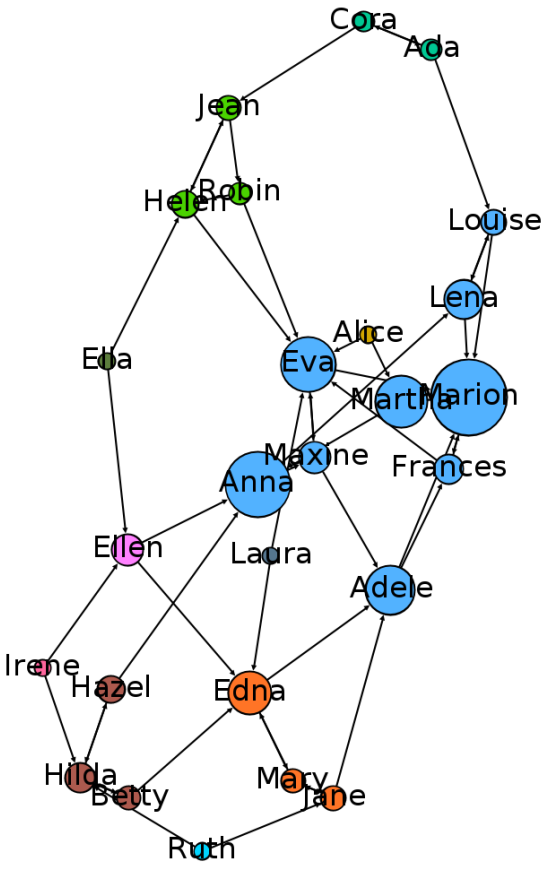

Figure 1: Graph representation of network with different communities denoted by different colors.

Q B 1. 1. There are 388 nodes in the network with 3598 connections.
2. The average degree is 9.273 and Stanley has the most connections (93)
3. Kelly
4. We would rank them according to their betweenness-centrality and use the top 50. This metric defines the importance of a node upon how many times it occurs in the shortest path between all pairs of nodes in a graph [1]. Using nodes with a high BC we target people that are crucial to the communication since they connect many nodes with each other.
5. We could give weights to the connections. If people exchange message more often the weight would become higher. We could incorporate the weight into calculating a new value (also using BC) and use the top 50 of those values instead of just the BC.

2. 1. There are 26 communities (using the modularity algorithm with default inputs)
2. The largest communities holds 92 nodes.
3. If two persons from two unconnected communities respond to each other, we can create a connection to them with a certain weight.
4. See Figure 2 for the graph visualization of this network.

[1] https://towardsdatascience.com/graph-analytics-introduction-and-concepts-of-centrality-8f5543b55de3

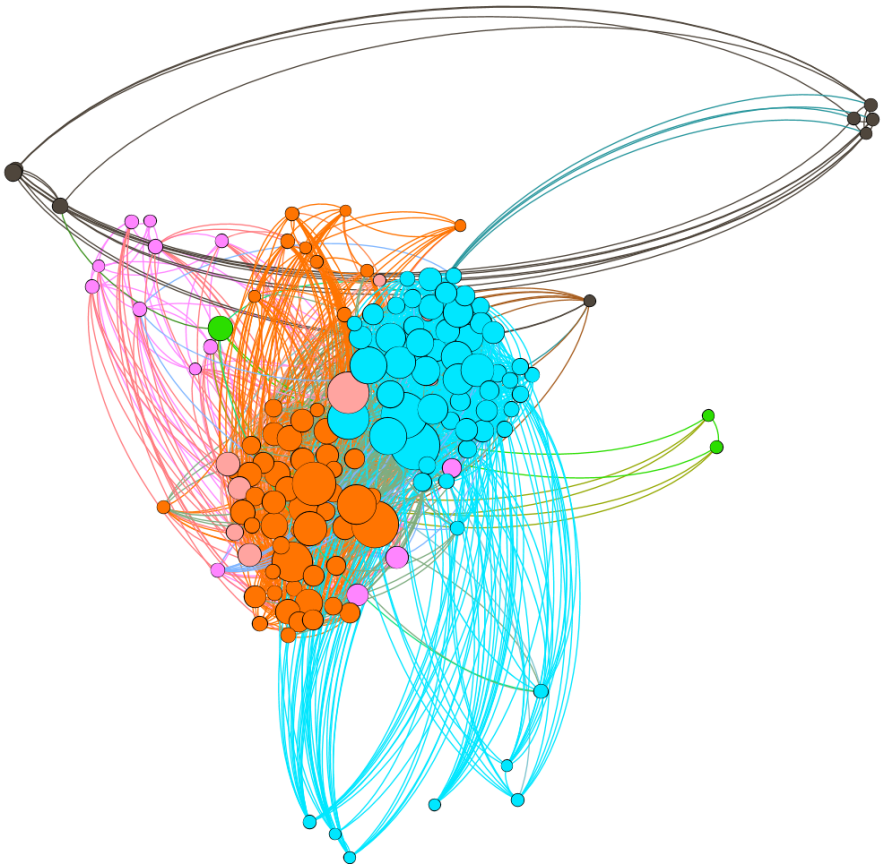

Figure 2: Graph representation of network with different communities denoted by different colors.