**MAA507 – Seminar 2 - A**

1. Elif Cemre Durgut

Group 4

1. Our problem was to identify researchers/scientists who are good at working with other people. We have a network that includes the research papers published by scientists.

We turned this network into an undirected graph in which the nodes represent the scientists, and the edges represent the collaboration between scientists. Initially, we thought that we can weigh the edges as the number of papers published. However, after the discussion during the presentation, we decided to make modifications to weights. Some problems we realized:

1. Older the scientist, more the paper.

Solution: We can only take the research papers published in the last 10 years into account.

1. Not all papers are successful.

Solution: We can give extra weight to the papers that have been published in important journals / awarded by well-known committees.

1. International group works are usually harder.

Solution: We can give extra weights based on the distance between the scientists.

The centrality measure is the combination of two methods. The first one is the clustering coefficient. Because our graph consists of clusters based on scientific domains since people from the same domain are more likely to work together.

The second centrality measure is degree centrality. Because the success of a scientist depends on the edges that are connected themselves. Other edges should not affect that scientist’s success. There

Using this approach, we miss people who don’t have any published papers yet.

**Example:**

Scientists: {Rivest, Shamir, Adleman, Elif}

Papers: {RSA Cryptography(by R, S and A in 1977), 2 papers (by E, A in 2020), 1 paper (by R, S in 1970), 1 paper (by R, S in 2012)}

Extra info: RSA Cryptography paper has 3 awards. Elif and Adleman live in different countries.

Nodes are the scientists represented by their initials. For the edges:

-We exclude the papers that are published before 1975 to compensate for the age difference. (That limit should be higher in reality.)

-RSA paper is awarded 3 times 🡪 extra weight: 3α

-Elif and Adleman live in different countries 🡪 extra weight: β

(α: award coefficient β: distance coefficient (These coefficients can be decided based on further research))

3α + 2

3α + 1

3α + 1

β + 1

1. Questions of this week were more discussion-based rather than mathematical questions, unlike the previous seminar. Therefore, we struggled a little bit more compared to the last seminar. We equally contributed to the discussion and presentation.
2. In my opinion, the biggest problem is to measure the collaboration between the scientists. Counting the published papers is not enough and it has its problems such as measuring the quality of the paper, deciding on the coefficients of extra weights(different award types, distances). Another issue might be “Should there be an extra weight for people who work in different disciplines due to lack of common scientific terminology?”.

Also, some people work together but have not published any paper yet.

1. Our graph has features of both. It can be considered as a small world because the clustering coefficient is large but computing the shortest path in our case does not make sense. On the other hand, it can be considered scale-free because most of the scientists work with a small network and they do not have many publications. But some scientists have lots of papers with others as well.
2. I think, it is reasonable for small networks. However, it can get quite slow and complicated for large networks. Another idea might be to first cluster the network using an ML algorithm based on some general domains such as mathematics, economics, social sciences and then apply the centrality degree measurement.