Homework 3 10/22/2021

Xinming Zhang

The first paper by Stanley Milgram and Jeffrey Travers is about the experimental study of the Small World Problem. From what I understand from the paper, it is a real experiment, and the "data" is provided. They are saying that "within this group (296 individuals) the mean number of intermediaries between starters and targets is 5.2". I did not trust the result at the first, because 5.2 is a small number compared to the group number 296. Also, the experiment was held in 1967 which is a long time ago. Those factors made me think the result will not make that much sense. However, the rest of the paper convinced me by performing the data and the graph.

The way of holding the experiment is in old style, and it makes sense because it was in 1967. In my opinion, the result will be that accurate causing by the huge error. Mistakes could occur during deciding whether you know the target person or not. Even they made those choices right, but the path might be the shortest path between starters and targets. If there is a better person (can result a shorter path) they could choose, but they did not. Then, the length of the path will be increased, and we do not want that. This could make the result inaccurate. It is still an extremely well-designed experiment back then, because if we are trying to solve the average upper bound, then the effect of the error will be decreased. Assume we try solving the same problem (Small World Problem) right now, it will be different. The technology is in a much higher level than before. We know a lot more people as well because of the internet. That is the first difference. We can also use computer to solve the problem, and we just need to collect the information from the individuals we want to test on. The number of people in the group can be increased to a huge number, and the computer should solve 10,000-induvial problem quick. The time of experiment could be shortened because of the internet and computer. Also, we are able to hold the experiment remotely, so we do not need contact the volunteers in person. (Especially under covid time, we want to avoid that as much as possible)

Suppose I am doing the same experiment, I would hire 10,000 volunteers on a platform online like Facebook or Instagram. The volunteers could be anyone from anywhere, because there are millions of different users on the platforms. I can start the experiment simply by posting a survey online to hire the volunteers. Then, after the number of volunteers gets up to 10,000 or the limit number we set, I would close the survey. When we have enough volunteer, we don't have to collect much information from them. Next step is to build our own program. Undirected graph will be my choice, because "know" will be in both directions as already explained in Jeffery's paper. Then we will need an algorithm to find the shortest path between starters and targets.

At first, we will need to set up the node carrying the useful information (include the follower if possible). After that, we could insert them in the graph G, and connect the people's node if they followed each other. Then we should have a huge graph with many different components. For the problem to be solved faster, we can select the target as the person who know the most people in the volunteers (the node with most edges).

If we are trying to find the shortest path between starters and target. We will apply depth-first-search (DFS) on the starters, and look for the shortest path. The runtime will be O(n), where n is the number of the nodes in the graph. The result will be the shortest possible in the graph we have, but there might be a case there is no possible path between them. That means that starter could not connected to the target, and that is not true in real life. If we want to have a better result, we can spend more time on modifying the algorithm.

We can also build a method to find the shortest outside of the graph we have. The method could go through the follower of the starter to find a path to target, if there is no possible path on the current graph. Then the runtime will be increased, but eventually we will have a path for every starter. Finally, we should have the "shortest" path between every starter and the target. The final step will be calculating the mean length of those paths, and that will be our final result.

We are holding the experiment on the platform, so there will be a path between everyone who has an account. This might not be true in our real life. In some areas, people might be disconnected to everyone else around the world. Then there will be a graph component disconnected to other connected components. We have internet now days, so the probability of this case happening will be extremely small, but it still might happen.

The way of experiment above might be the best at our time, and there are more smarter ways to achieve the same goal, but this is my best try. If the algorithm is smart enough, I believe there is one day the experiment can be applied to the whole population around world through the contact list on our phone. On the other hand, the experiment form Milgram gives me the idea the experiment, and that might happen on lots of people who read the paper. The result of the experiment now days might be more accurate than the one from paper, and that is caused by the improvement of the technology. I am still impressed by the Milgram's paper, because the experiment was done late 1960s. The experiment is not simple even now days, so I really appreciate the work they have done.

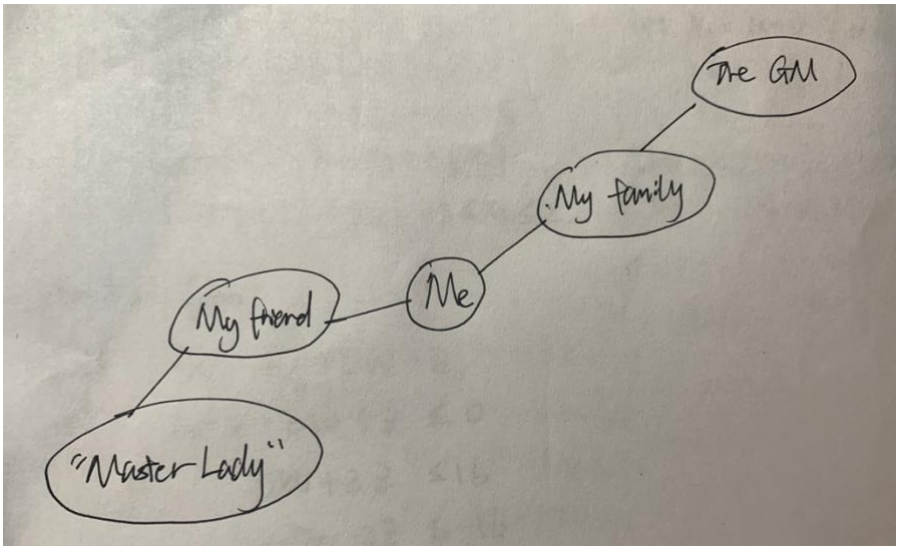
Frigyes Karinthy was talking about the same problem in his short story *Chain Links*. Compare it to Milgram's paper, Karinthy was not from an academic aspect to analysis the problem. Milgram held the real experiment to get the data to get and prove his result. Then Karinthy did not even analysis anything to get the solution of the problem. The solution form him for the "Small World Problem" is 5, and that is close to 5.2 which is the result from Milgram. That surprised me, because I did not believe the story at all. I am a "science" type of person, I will not consider anything is a fact until I see some proves. I still do not get it, because the solution seems reasonable, but there is no proves we can rely on. What's more, the story was from 1929, it is even a lot earlier than Milgram's paper. That is the most amazing point to me, it is from more than ninety years ago, and the solution he got is close to the one from Milgram's real experiment. I still do not have any ideas why he got that, and I am not able to analysis his story from my point of view.

Other than those paper, I also tried to explore the acquaintance graph from my perspective. After asking my friends and families, there are two outstanding examples affect the graph. One of my families work s in a fishing company, and I am still student, so I do not know any of his colleagues. I will not be directly connected to any of his colleagues. The general

manager (GM) in his company will be one of the examples I will introduce. The GM is 72-year-old gentlemen, and I am 22-year-old male student. Our age difference is huge, and our income will be extremely different as well. He is the general manager of the company, but I am still student with a part-time job. However, we both live in Washington state, so we still have some similarities. He is one of the examples, and the other one is my friend's friend.

My friend's friend is a lady who is in a master program in Shanghai, China. We do not know each other, so we are not connected on the graph. At first, we are in different genders. I am still an undergraduate student, but she is in graduated program. Even though we all come from China, I am here, and she is in Shanghai. She has a good full-time job, so our incomes are different. She is a bit older than me, but I am not sure exactly how much older. Above those two people will be two examples.

The GM and "Master Lady" will affect the acquaintance graph from my perspective. If we want to find the shortest path between me and other elder people on the graph, the GM might be one of the intermediaries on the path. Another possible case will be looking for the shortest path between me and a GM in some company, the GM is most likely "on" the path. "Master Lady" will most likely be "on" the shortest path between me and master students in China or other employees in Shanghai. On the other hand, the GM and "Master Lady" can be connected through me on the graph. They are different as well, they are in different countries, and they do not know each other. They don't have direct friends in common, but still, they will be connected through me in four edges. The graph is shown below:



The two huge graph components connected to the GM and "Master Lady" will be connected to me on graph because of them, that would give some more shortest path between some people.

For the last two question from the assignment, I briefly answered them on first page. The occupation of the target would affect the connection on target, so if the target has an occupation which has less connections to others. Then the result from the experiment will be increase, that means it needs more edges to connect starters and the target on average.