Noga Starobinski

For this Final Project, I analyzed a dataset containing information on the top 1000 Instagram influencers –who are ranked by the number of quality and engaged followers. Given the dataset, I aimed to discover a relationship between Instagram influencer accounts and location. To determine whether the Instagram influencers in the same Audience Country tend to have similar follower counts, I created a graph that explores the idea of Six Degrees of Separation by finding the average distance between pairs of vertices in the graph.

I used the BFS algorithm to find the shortest path between the different users' audience country and follower counts. The average distance is calculated by iterating over each account in the country and calculating the average BFS distance between that account and all other accounts in the country. The average number of followers is calculated by dividing the total number of followers in the country by the number of Instagram accounts in that country. The bfs_distance function works in the following way: it takes a reference to the graph HashMap and a start username as input and returns the average distance between the start node and all other nodes in the graph. The function initializes a visited set and a queue with the start node, and then iteratively dequeues the first node from the queue, enqueues its unvisited neighbors, and updates the visited set and the total distance. The distance is incremented when all nodes at the current level have been dequeued. If the start node is the only node in the graph, the distance is zero. Implementing BFS allowed me to efficiently analyze the large dataset and estimate the sampling distribution of the mean distance between different combinations of user accounts.

When running the code, I obtained a range of values for the average distance, revealing that perhaps there is not such a significant correlation between Instagram influencers in the same Audience Country and their follower counts. Instead, the average distance between nodes varies quite a bit across different countries. For example, during one trial I got 2.87 as the average distance between nodes in the United States – indicating a level of clustering or connectedness amongst Instagram influencers in the U.S. With this in mind, the average follower count in the US is quite high at over 16,000,000 so there are many high-profile influencers with large followings. By comparison, when running a trial for Brazil, the average distance between nodes was slightly lower at 2.67, but the average follower count is around 3,000,000. Based off the results, perhaps

there are other factors such as content quality, engagement rate, and influencer niche that play a larger role in determining a higher follower count for an Instagram influencer.

As demonstrated by this project, by exploring the relationship between Instagram influencer follower counts and their location we can gain insights into how audiences interact with content on Instagram which could be useful for social media marketing strategies and learning about human behavior online. Thus, these results, provide an indication of the level of influence Instagram celebrities have in each respective country, as well as the popularity of the individual accounts.