

# **Final Project Report**

**By: Jonathan Hafez**

For this Final Project, I analyzed a dataset containing data about blue-verified pages on Facebook, which are represented as nodes in an undirected graph with connections representing mutual likes among a group of blue-verified users. I used the BFS (Breadth-First Search) algorithm to find the shortest path between the different usernames and their respective "liked" pages, allowing me to understand the relationships between the different Facebook pages and how they are connected through mutual likes.

To make the analysis more efficient, I selected a random sample of 1000 lines every time the code was run. I then calculated the mean distance between the selected combinations by summing the distances found and dividing by the number of trials (1000). This allowed me to estimate the sampling distribution of the mean and make inferences about the population from which the samples were drawn, even though I was not using the complete dataset (which contained 52,311 edges).

I ran my code 10 times and obtained the following values for the average distance between 1000 nodes: {5.336, 5.26, 5.365, 5.323, 5.323, 5.291, 5.329, 5.292, 5.36, 5.268}. These values are relatively close to the value of 5.3, indicating that, despite the apparent lack of connection between different blue verified pages on Facebook, one could say that it is possible to reach any desired blue verified page by making a connection with just five other pages. Overall, this analysis allowed me to gain insights into the connections between different blue verified pages on Facebook, and it helped me understand how they are related through mutual likes.

In conclusion, the analysis of the data on blue-verified pages on Facebook revealed that there are significant connections between different pages through mutual likes. By using the BFS algorithm and sampling a random subset of the data, I efficiently analyzed a large dataset and estimated the sampling distribution of the mean distance between different combinations of pages. The analysis results, which were obtained by running the code 10 times and calculating the average distance between 1000 nodes, showed that it is possible to reach any desired blue verified page by connecting with just five other pages.

The use of statistical techniques, such as sampling and calculating the mean, and the using sophisticated algorithms, like BFS, allowed for the efficient analysis of a large dataset and the estimation of the sampling distribution of the mean distance between different combinations of pages. These techniques helped to draw meaningful conclusions from the data and provided a better understanding of the relationships within the network of blue-verified pages on Facebook. Overall, the analysis revealed the significance of connections between different pages through mutual likes and provided valuable insights into the connections between different users on social media platforms.