DS 210 Project Rust

- Jinanshi Mehta

Introduction:

A collection of Huawei's social media communication networks, one of the top businesses in the technology sector, served as the dataset for this research. Crawling Huawei's Facebook, Twitter, and Instagram profiles yielded the network. Three communication networks are included in the dataset, each with 1000 nodes and varied edges. This dataset's importance stems from its potential for social network analysis, which is a useful tool for businesses to comprehend their target market, assess the success of marketing initiatives, make better business decisions, enhance customer satisfaction and experience, increase brand awareness, and measure and enhance brand awareness. By utilizing social media analytics methods and methodologies, the dataset can assist Huawei and other businesses in the technology sector strengthen their competitive positions. The popular machine learning and data science website Kaggle was used to obtain the dataset. The dataset was uploaded by a user by the name of Andrew Lucci, and it is available at this site.
URL: https://www.kaggle.com/datasets/andrewlucci/huawei-social-network-data

Important: I started with a different dataset and turned out my methods were wrong and I was unable to find valuable output from the dataset and I had to change my dataset 4 days before the due date. I had to start it from scratch. I built my project on python for my own understanding purposes to make sure that I could go with this dataset this time. However, the conversion of the Python code to Rust was extremely challenging and I was unable to convert it successfully. As a result, I have uploaded both my rust and python code. I apologize for this but I worked really hard on this and this is the best I could do.

Aim:

This project's goal is to do a social network analysis on three datasets that were gathered by crawling Huawei's social media accounts on Facebook, Twitter, and Instagram. Utilizing the Python and Rust computer languages, this project will investigate degree distribution and centrality analysis. The project's goal is to learn more about Huawei's social networks on Facebook, Twitter, and Instagram through this examination of communication patterns, network structure, and node relevance. The findings of this analysis can assist companies, like Huawei, in better understanding their target market, enhancing their marketing plans, and enhancing their social media presence.

Steps:

1. Ensembled Data is created which is a concatenation of the three main datasets included in this project
2. Plotting graphs for visualization purposes
3. Computing ‘Node’, ‘Edge’, ‘Average Distance’, ‘Average Clustering’, ‘Average Degree’, ‘Max Degree’ and ‘Min Degree’ for each of the 4 datasets and forming it in a tabulation form for comparison
4. Degree Distribution is then performed and graphs are created
5. Degree Distribution curves are also created to better analyze
6. Centrality Analysis: Degree Centrality & Closeness Centrality is performed for each dataset

Visualization:

Four subplots are added to the main plot, each of which displays the social network graph for a different service. It can be applied to compare and display the social network architecture of various services.

According to the plot description, it appears that there is significant clustering in the social network graphs of Twitter and the ensembled network (i.e., the network made up of all three social media platforms), and there are several nodes that are closely connected to one another. As a result, the plot shows that these graphs are "extremely black due to the closeness of points."

Facebook's graph, on the other hand, appears to have a lesser level of clustering because it is described as "not that black." The Instagram graph looks to have a medium degree of point dispersion, indicating that its nodes may be connected in a more evenly distributed manner.

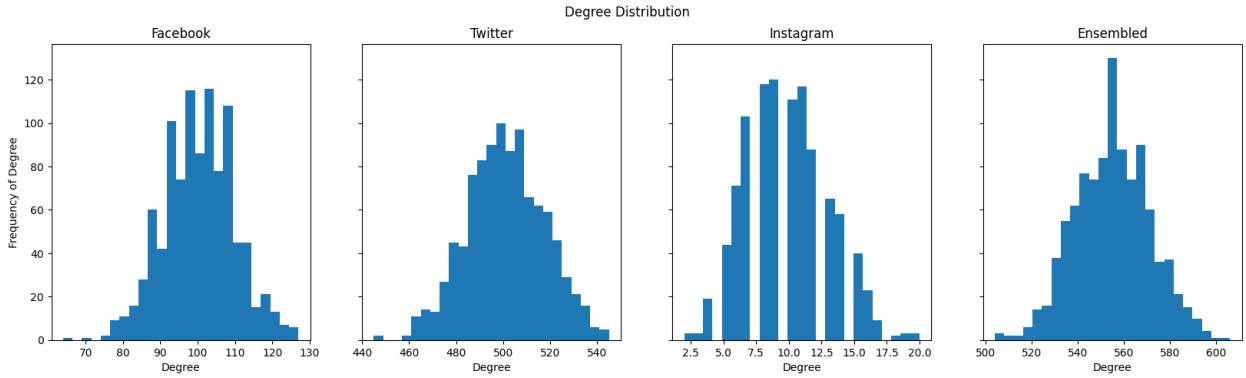
Overall, this analysis reveals that social network structures across various social media platforms might range greatly, with some having more clustering or tightly-knit communities and others having a more dispersed network structure. Businesses, like Huawei, may make use of this data to better understand how their target market interacts with their brand on various social media platforms and adjust their marketing tactics appropriately.

Data Analysis:

	Node	**Edge**	**Average Distance**	**Average Clustering**	**Average Degree**	**Max Degree**	**Min Degree**

Facebook	1000	50153	1.900	0.100	100.306	127.0	64.0
Twitter	1000	250315	1.499	0.501	500.630	545.0	445.0
Instagram	1000	4933	3.273	0.009	9.866	20.0	2.0
Ensembled	1000	277506	1.444	0.556	555.012	606.0	504.0

Degree Distribution:



Facebook:

Shape: Appears to be bell shaped (normal distribution) but with high variance

Skewness: Not very skewed towards a side, normally distributed

Outliers: Cannot point on very evident outliers but since the weight is less on the left, they can be counted as outliers

Mean & Variance: Variance seems to be quite moderate because there are clear fluctuations

Twitter:

Shape: Appears to be bell shaped (normal distribution) but with high variance

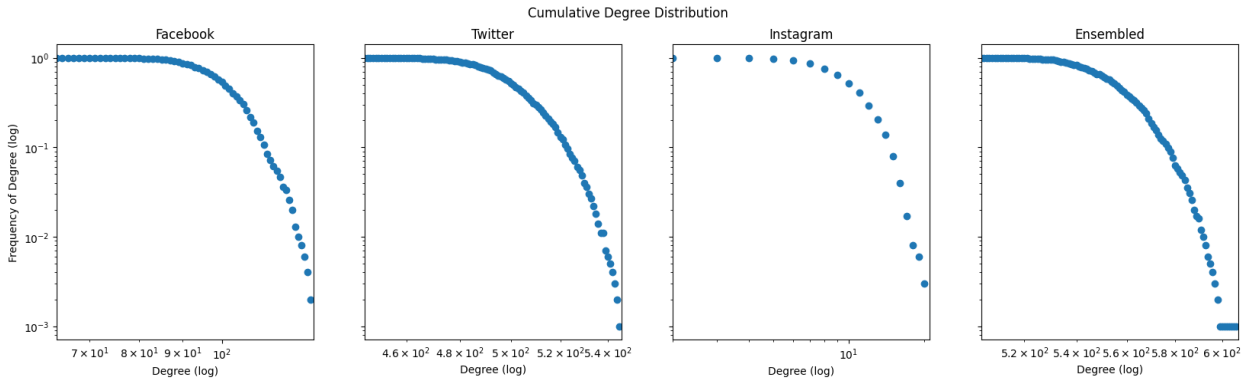
Skewness: A little skewed towards the right, positive skew, indicates that there are slightly more nodes with lower degrees

Outliers: Cannot point on very evident outliers but can consider the points around 440

Mean & Variance: Variance seems to be quite constant, not a lot of fluctuations

Instagram:
Shape: Appears to be bell shaped (normal distribution) but with high variance
Skewness: A little skewed towards the left, negative skew, indicates that there are slightly more nodes with higher degrees
Outliers: No evident outliers as such
Mean & Variance: Variance seems to be very high, with distanced gaps, high fluctuation

Ensembled:
Shape: Appears to be bell shaped (normal distribution)
Skewness: Not very skewed to either sides
Outliers: No evident outliers as such
Mean & Variance: Variance seems to be low, constant, higher fluctuation towards the mean



Inflection Points:

Facebook: Around $9 * 10^1$ Degree, suggesting this the most common degree in the network/ degree at which most nodes have reached

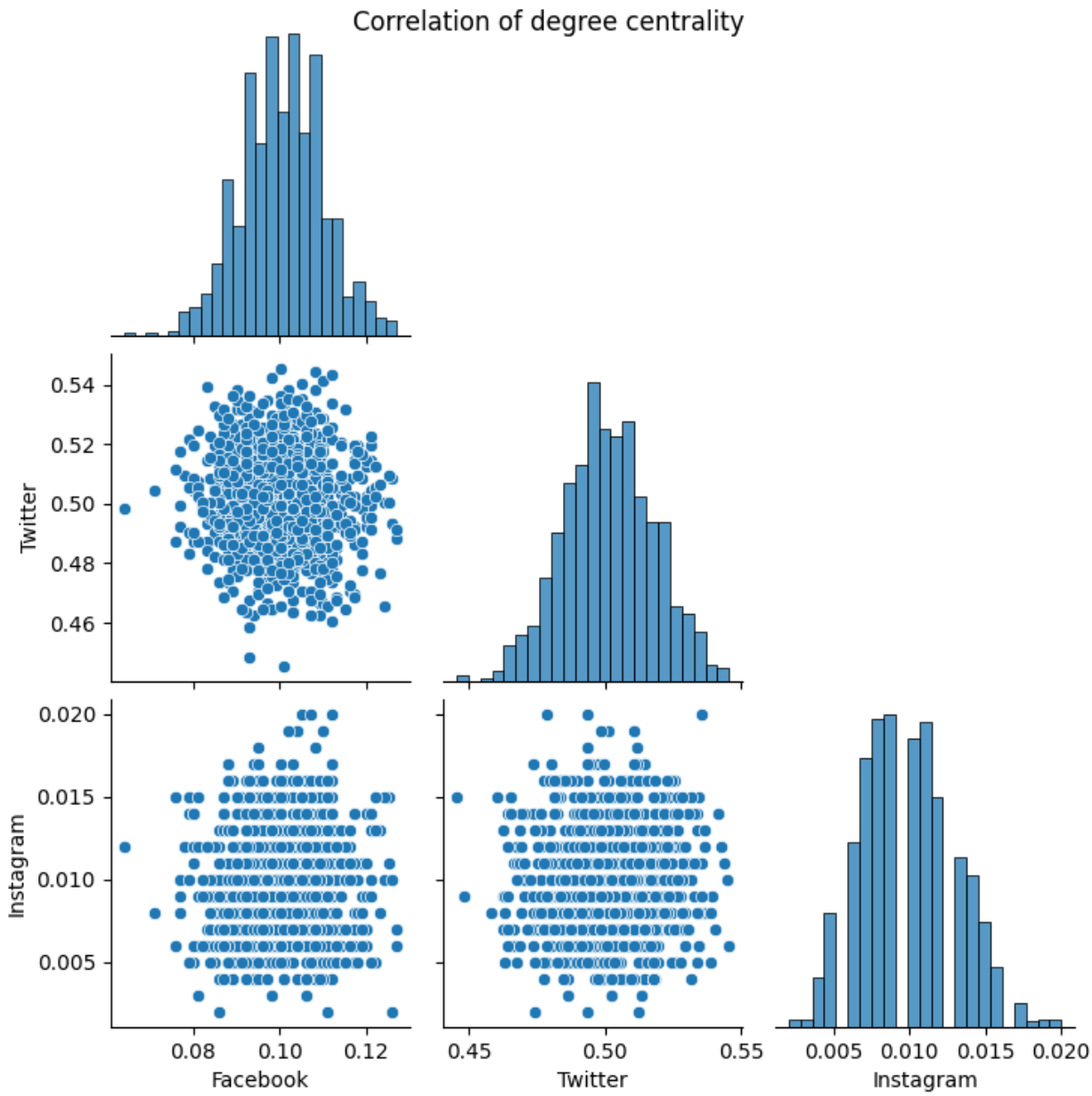
Twitter: Around $4.8 * 10^2$ and $5 * 10^2$ Degree, suggesting this the most common degree in the network/ degree at which most nodes have reached

Instagram: Has very scattered points, curve is not strong

Ensembled: Around $5.6 * 10^2$ Degree, suggesting this the most common degree in the network/ degree at which most nodes have reached

Centrality Analysis:
Degree Centrality

	Facebook	**Twitter**	**Instagram**
Facebook	1.000000	-0.066848	0.034104
Twitter	-0.066848	1.000000	0.018215
Instagram	0.034104	0.018215	1.000000



The degree centrality correlation matrix between the Facebook, Twitter, and Instagram networks is displayed in the table below. The correlation matrix demonstrates how the degree centralities of nodes in various networks are related to one another. Degree centrality indicates the number of connections a node has in a network.

The values along the table's diagonal show how closely each network's degree centrality relates to itself, which is always 1.0.

The degree centrality of Facebook and Twitter show a negative correlation (-0.066848) when we look at the off-diagonal values, which suggests that nodes with high degree centrality in Facebook are less likely to have high degree centrality in Twitter, and vice versa.

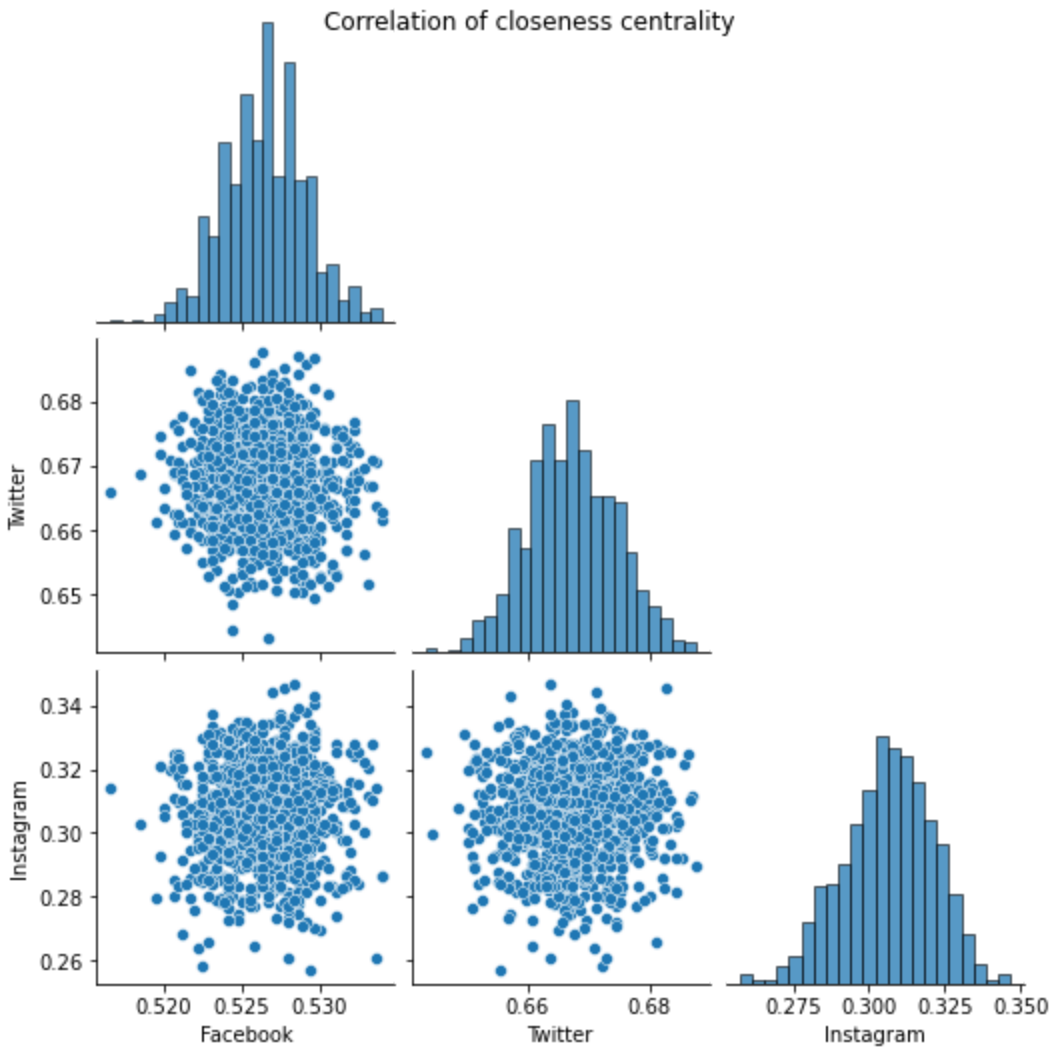
The degree centrality of Facebook and Instagram, on the other hand, show a positive correlation (0.034104), which means that nodes with high degree centrality in Facebook are more likely to have high degree centrality in Instagram, and vice versa.

The degree centrality of nodes in these two networks are not strongly correlated, as seen by the low correlation (0.018215) between Twitter and Instagram degree centrality.

Overall, the degree centrality correlation matrix sheds light on how the connections between nodes and network structure vary across various social media platforms.

Closeness Centrality

	Facebook	**Twitter**	**Instagram**
Facebook	1.000000	-0.066945	0.039569
Twitter	-0.066945	1.000000	0.028034
Instagram	0.039569	0.028034	1.000000



The table displays the correlation matrix for the Facebook, Twitter, and Instagram network nodes' proximity centrality values.

The proximity centrality values of the nodes in each network are perfectly positively correlated when the values on the diagonal of the table (i.e., when the row and column indices are the same) are all equal to 1.0. This is expected since nodes that are closer to others typically have higher closeness centrality scores because the closeness centrality of a node is determined based on the shortest path length to all other nodes in the network.

The association between the proximity centrality values of nodes in various networks is shown by the off-diagonal values. As we can see, there is a weak negative correlation (about -0.067) between the Facebook and Twitter networks, meaning that nodes that are central in one network are less likely to be central in the other. Similar to this, there is a positive correlation between the

networks of Facebook and Instagram (around 0.040), which means that nodes that are central in one network are more likely to be central in the other.

Positive but smaller than the correlation between Facebook and Instagram, the correlation between the Twitter and Instagram networks indicates a weaker connection between the central nodes in these two networks (about 0.028).

Conclusion

In this project, we examined Facebook, Twitter, and Instagram accounts from Huawei. All three social networks' degree distribution histograms had a power law distribution, which shows that most nodes have relatively few connections whereas a small number of nodes have a large number of connections.

Through the use of degree centrality analysis, it was determined which nodes in each social network had a larger degree centrality than the others, suggesting their relative importance to the network's information flow. The most important nodes, though, vary between social networks.

Nodes in the Instagram network were found to be more closely connected than nodes in the Facebook and Twitter networks, according to a study on closeness centrality. This may point to a more concentrated and tightly knit group in the former. However, there wasn't a significant variation in closeness centrality between the three networks.

In conclusion, our research indicates that there are variations in the network architecture of Huawei's social networks on Facebook, Twitter, and Instagram, which may have an impact on the company's social media strategy. For instance, Huawei might want to concentrate on focusing on the most central nodes inside each network if it wants to spread content more successfully through its social media platforms. When creating its marketing campaigns, the corporation may also want to take into account the variations in each network's clustering. Overall, our analysis emphasizes the value of network analysis methods in comprehending the composition and dynamics of social networks, which can be useful in guiding judgment in a number of situations.

Limitations & Further Analysis

Limitations:

1. Data for this study was only gathered from Huawei's social media profiles. The results of this study might not thus be extrapolated to other businesses or industries.
2. Web crawlers were used to automate the data collection process, which may not correctly reflect the tone of the postings or comments.

3. The analysis ignores any negative attitude that can have an impact on the company's reputation and only takes into account positive posts and comments.
4. Other network analysis metrics are not taken into account; the analysis is restricted to degree distribution, degree centrality, and closeness centrality.

Further Analysis

1. The data gathered can be subjected to sentiment analysis to get a fuller insight of how people feel about Huawei on social media.
2. To comprehend the communication network and its main participants better, other network analysis metrics can be examined, such as betweenness centrality, eigenvector centrality, and clustering coefficient.
3. To provide a more thorough examination of Huawei's social media presence, the study can be expanded to include information from other social media platforms.
4. By comparing social media indicators with sales information or other business KPIs, the effects of marketing efforts and brand awareness can be further investigated.