Social Media Network Analysis

Saad Sharif

February 20, 2026

1 Introduction

This assignment will aim to analyse the characteristics of social networks across three online platforms (Instagram, Facebook and Twitter). First, we will gather initial statistics on the network data to understand their characteristics. Next we will aim to identify communities within each network using methods such as Girvan-Newman and Kernighan-lin, and compare these results. Finally, we will plot our results from the community detection and conclude our results from these analyses.

The network data used in this analysis were obtained from a publicly available dataset [1]. In our datasets, we have 1000 people and can see whether there is a relationship between each pair of people based on if their intersecting cell has a 0 or a 1. A relationship between two people may exist in one platform but not in the others. Therefore we will analyse these three networks and compare our findings.

The goal of this report is to compare structural properties of social media platform networks. We will investigate how differences in connectivity and density influence the results of community detection algorithms.

2 Network Metrics

2.1 Degree Distribution

One of the first things we can analyse is the degree distribution across the three platforms. The goal of this is to understand how the connections are distributed among the users.

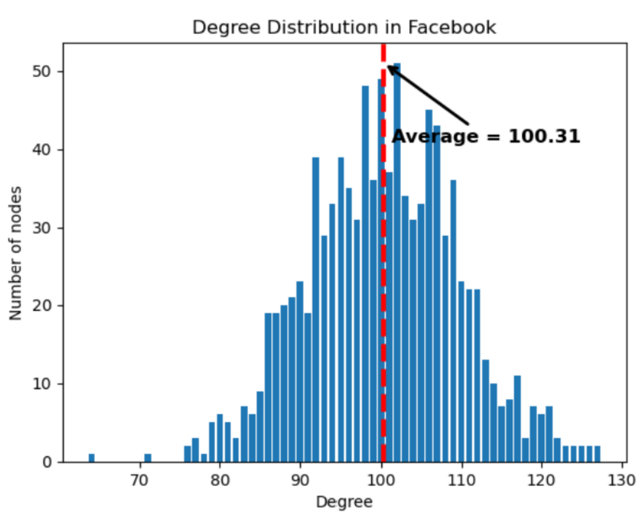

Figure 1: Degree Distribution in Facebook

First we analyse the degree distibution in the network obtained from Facebook. Here we can see the degree ranges from 60 to 130 and average of 100.31.

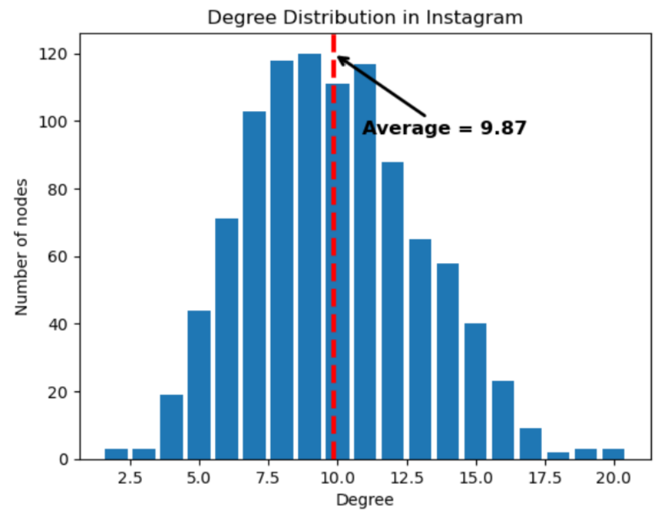

Figure 2: Degree Distribution in Instagram

For Instagram, we have a significantly lower range for the degrees which is from 0 to 20, with the average being 9.87.

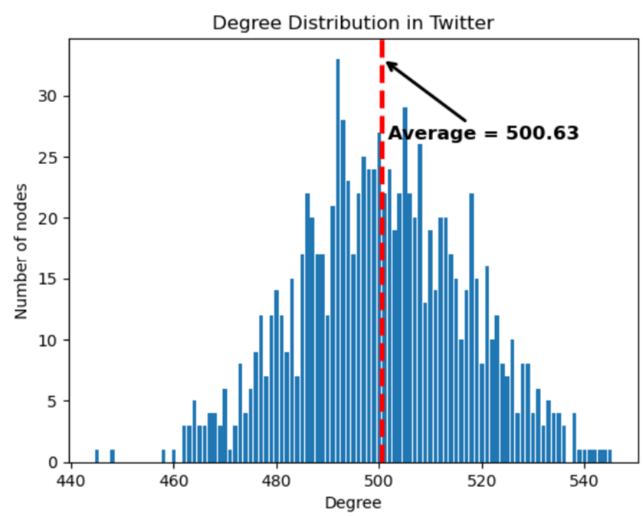

Figure 3: Degree Distribution in Twitter

Finally, for Twitter we have the largest range for the degree values which span 440 to 600, with the average being around 500.63.

We can understand from these charts that Twitter is a significantly more connected network where the nodes have more relationships with each other, whilst Facebook is moderately connected and Instagram is a very sparse network with very few connections between the nodes.

2.2 Graph Density

Graph density is another metric that is useful to observe as it tells us how connected a network is in comparison to the maximum number of connections it could contain. Given this metric is a fraction it can range from 0 which represents almost no connections, to 1 which represents a network where every node is connected to each other.

We can see in **Figure 4** that the Graph Density for Twitter is the highest at 0.50, followed by Facebook at 0.10 and Instagram with the least value at 0.01 all to 2dp.

2.3 Global Clustering Coefficient

This metric tells us the likelihood of nodes to form tightly connected groups, where a higher value indicates neighbours of a node are likely to also be connected.

Again we can see in **Figure 4** that Twitter has the highest value of 0.50, followed by Facebook at 0.10 and then Instagram at 0.01 all to 2dp. This again tells us Twitter is more likely to have neighbours of a node to be connected in comparison to the Facebook and Instagram networks.

2.4 Average Shortest path length

average number of edges needed to reach one node from another using the shorter path possible. Here we can see that Instagram has the highest at 3.27, followed by Facebook at 1.90 and then Twitter with the least at 1.50 all to 2dp.

	Network Type	Graph Density	Global Clustering (Transitivity)	Average Shortest Path Length
0	Facebook	0.100406	0.100290	1.899632
1	Twitter	0.501131	0.501201	1.498869
2	Instagram	0.009876	0.008399	3.273137

Figure 4: Network table metrics

3 Community Detection

Now we move on to Community Detection algorithms and how to analyse the communities found with useful metrics.

3.1 Girvan Newman

One algorithm for detecting communities is known as Girvan Newman. The idea behind this method is that edges that appear to be connecting communities, often referred to as bridges, are aimed to be removed first. The algorithm first computes "edge betweenness" for every edge which is the proportion of shortest paths that pass through this edge in comparison to all shortest paths between various nodes. Once these have been computed, the edges with the highest "betweenness" are removed in order to weaken connections between communities.

Edge betweenness is recalculated again and the process is repeated as communities become more evident. At each split, modularity, which tells us how good a networks community structure is, is computed and the partition with the highest modularity is kept.

3.2 Kernighan Lin

Another algorithm is known as Kernighan Lin which works by dividing a network into two equal sized groups while minimising the number of edges between the two groups. It does this by finding the best pair of nodes to swap between two groups if this results in an improvement in the cut, which is the number of connections between the two groups.

We can apply Kernighan Lin recursively on each community until it no longer improves the modularity constraints.

3.3 Metrics

Once communities have been identified using either Girvan Newman or Kernighan Lin algorithm, we can analyse them individually using the following metrics:

Size: This metric tells us the size of each community in terms of number of nodes in them. This would tell how balanced or imbalanced the communities are that are found in a network by size.

Internal Edges: Number of edges inside the community where more edges would indicate stronger connectivity while less edges would mean a weaker connectivity within a community.

Density: This tells us the number of edges that exist in a community as a fraction of the total number of all possible internal connections the community can happen.

Avg Degree: The average number of connections the nodes have in a community.

Clustering: As previously discussed, this metric tells us the likelihood of nodes to form tightly connected groups.

Cut Edges: Number of edges that connect a community to nodes outside the community.

Conductance: This measures the proportion of edges to nodes outside a community, to the total edges from the nodes within the community.

4 Results

When applying the community detection algorithms, we only used the top 100 nodes ranked by degree counts, as this would focus on the main influential people.

4.1 Girvan Newman Facebook Graph

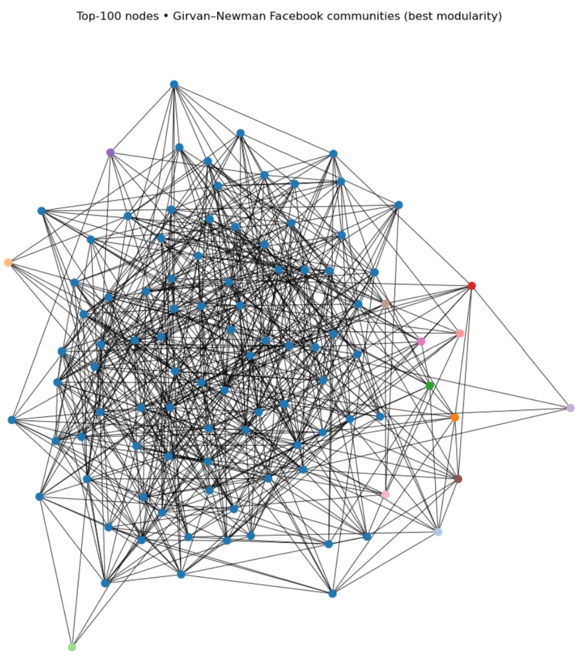

Figure 5: Girvan-Newman Facebook Graph

	community	size	internal_edges	density	avg_degree	clustering	cut_edges	conductance
0	0	87	581	0.155306	13.356322	0.155864	84	0.763636
1	1	1	0	0.000000	0.000000	0.000000	10	1.000000
2	2	1	0	0.000000	0.000000	0.000000	10	1.000000
3	3	1	0	0.000000	0.000000	0.000000	7	1.000000
4	4	1	0	0.000000	0.000000	0.000000	10	1.000000
5	5	1	0	0.000000	0.000000	0.000000	4	1.000000
6	6	1	0	0.000000	0.000000	0.000000	11	1.000000
7	7	1	0	0.000000	0.000000	0.000000	8	1.000000
8	8	1	0	0.000000	0.000000	0.000000	8	1.000000
9	9	1	0	0.000000	0.000000	0.000000	4	1.000000
10	10	1	0	0.000000	0.000000	0.000000	11	1.000000
11	11	1	0	0.000000	0.000000	0.000000	10	1.000000
12	12	1	0	0.000000	0.000000	0.000000	10	1.000000
13	13	1	0	0.000000	0.000000	0.000000	7	1.000000

Figure 6: Girvan-Newman Facebook Metrics

After applying the Girvan Newman algorithm to the Facebook Graph with only the top 100 nodes by degree number, we get the results in **Figure 5**. We identified 14 communities, but these results are essentially one community with 87 nodes and 13 other communities consisting of just a single node. The main dominating community has a density of 0.16 and clustering coefficient of 0.16.

4.2 Girvan Newman Twitter Graph

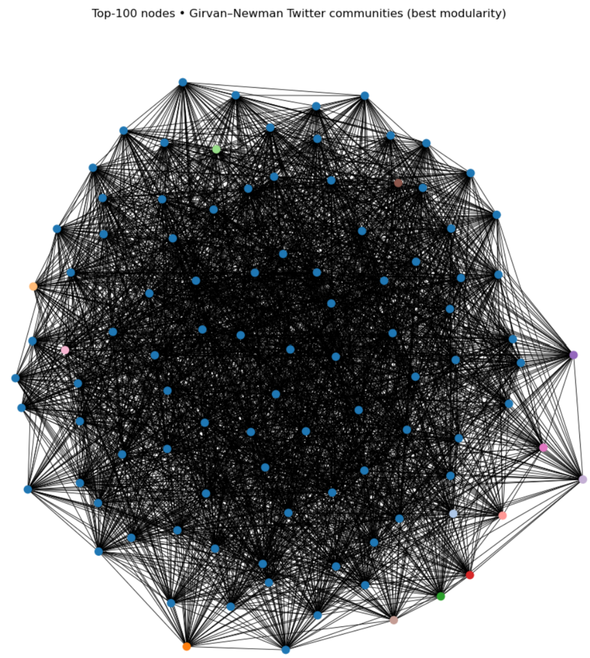

Figure 7: Girvan-Newman Twitter Graph

	community	size	internal_edges	density	avg_degree	clustering	cut_edges	conductance
0	0	87	2183	0.583534	50.183908	0.581474	530	0.863192
1	1	1	0	0.000000	0.000000	0.000000	47	1.000000
2	2	1	0	0.000000	0.000000	0.000000	49	1.000000
3	3	1	0	0.000000	0.000000	0.000000	44	1.000000
4	4	1	0	0.000000	0.000000	0.000000	50	1.000000
5	5	1	0	0.000000	0.000000	0.000000	49	1.000000
6	6	1	0	0.000000	0.000000	0.000000	46	1.000000
7	7	1	0	0.000000	0.000000	0.000000	45	1.000000
8	8	1	0	0.000000	0.000000	0.000000	51	1.000000
9	9	1	0	0.000000	0.000000	0.000000	42	1.000000
10	10	1	0	0.000000	0.000000	0.000000	47	1.000000
11	11	1	0	0.000000	0.000000	0.000000	48	1.000000
12	12	1	0	0.000000	0.000000	0.000000	47	1.000000
13	13	1	0	0.000000	0.000000	0.000000	49	1.000000

Figure 8: Girvan-Newman Twitter Metrics

Applying Girvan Newman to Twitter also yields a similar result as seen in **Figure 7** where one Community consists of 87 nodes and the other 13 nodes become their own independent communities. The main dominating community has a density of 0.58 and clustering coefficient of 0.58 indicating a stronger interconnectedness in the network in comparison to Facebook.

4.3 Girvan Newman Instagram Graph

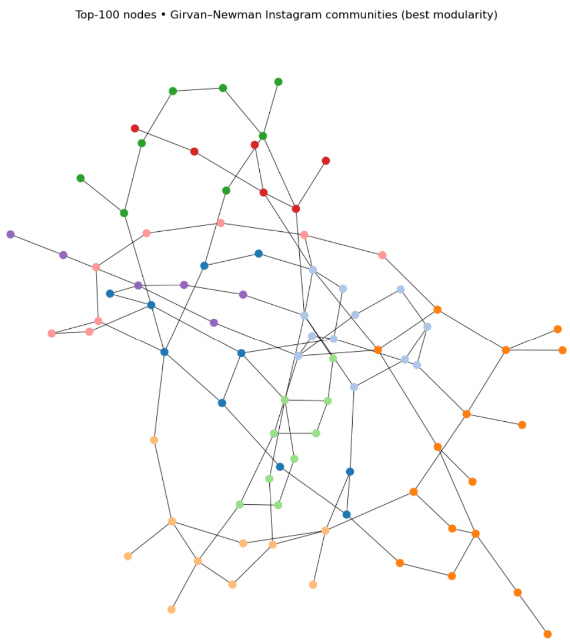

Figure 9: Girvan-Newman Instagram Graph

	community	size	internal_edges	density	avg_degree	clustering	cut_edges	conductance
0	0	10	10	0.222222	2.000000	0.0	12	0.375000
1	1	12	13	0.196970	2.166667	0.0	14	0.350000
2	2	16	16	0.133333	2.000000	0.0	6	0.157895
3	3	10	10	0.222222	2.000000	0.0	5	0.200000
4	4	8	7	0.250000	1.750000	0.0	3	0.176471
5	5	9	9	0.250000	2.000000	0.0	6	0.250000
6	6	6	5	0.333333	1.666667	0.0	3	0.230769
7	7	8	7	0.250000	1.750000	0.0	4	0.222222
8	8	6	5	0.333333	1.666667	0.0	3	0.230769

Figure 10: Girvan-Newman Instagram Metrics

Applying Girvan Newman to the Instagrams top 100 nodes, we get the resulting network graph in **Figure 9**. Here the algorithm identified 9 communities which are ranging in sizes from 6 to 16 nodes. Here we can see that Girvan Newman found a more balanced partition of the network into communities with similar densities and conductance values, which highlights how the algorithm performance is influenced by the overall sparsity of the entire network.

4.4 Kernighan Lin Facebook Graph

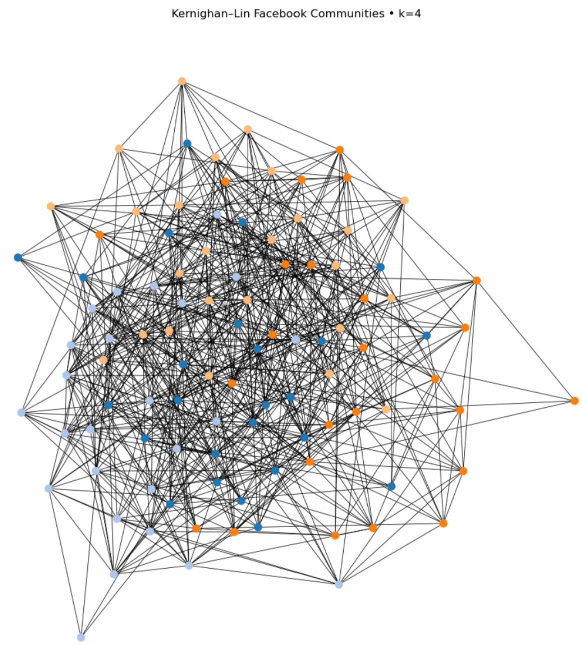

Figure 11: Kernighan Lin Facebook Graph

	community	size	internal_edges	density	avg_degree	clustering	cut_edges	conductance
0	0	25	80	0.266667	6.40	0.238635	190	0.542857
1	1	25	79	0.263333	6.32	0.238581	182	0.535294
2	2	25	75	0.250000	6.00	0.266020	173	0.535604
3	3	25	81	0.270000	6.48	0.229365	181	0.527697

Figure 12: Kernighan Lin Facebook Metrics

When applying Kernighan Lin to the Facebook top 100 network, we receive four communities of size 25 nodes, as seen in **Figure 11**. Their internal edges range from 75 to 81 and the average degree ranges from 6 to 6.48. The conductance values are approximately 0.53-0.54. This is a significantly lower number of communities found in comparison to the results of the Girvan Newman algorithm but no one community dominates the others in size, indicating a more balanced outcome.

4.5 Kernighan Lin Twitter Graph

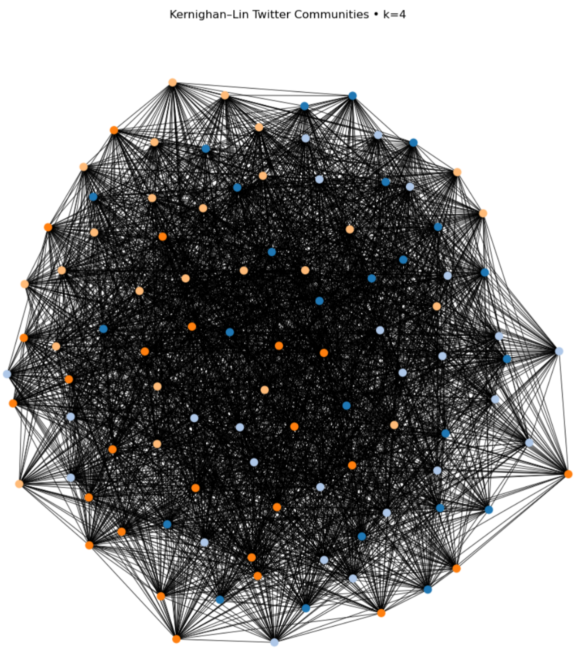

Figure 13: Kernighan Lin Twitter Graph

	community	size	internal_edges	density	avg_degree	clustering	cut_edges	conductance
0	0	25	218	0.726667	17.44	0.722879	925	0.679647
1	1	25	221	0.736667	17.68	0.723096	957	0.684060
2	2	25	219	0.730000	17.52	0.721907	929	0.679590
3	3	25	220	0.733333	17.60	0.720282	943	0.681851

Figure 14: Kernighan Lin Twitter Metrics

Similarly, the Kernighan Lin algorithm applied to the Twitter top 100 network in **Figure 13** also results in 4 communities identified each with a size of 25 nodes. The count of internal edges range from 218 to 221 and average degree ranges from 17.44 to 17.68. The conductance values are slightly higher than that of the Facebook network as they around 0.68 indicating stronger sense of communities in Twitter in comparison to Facebook.

4.6 Kernighan Lin Instagram Graph

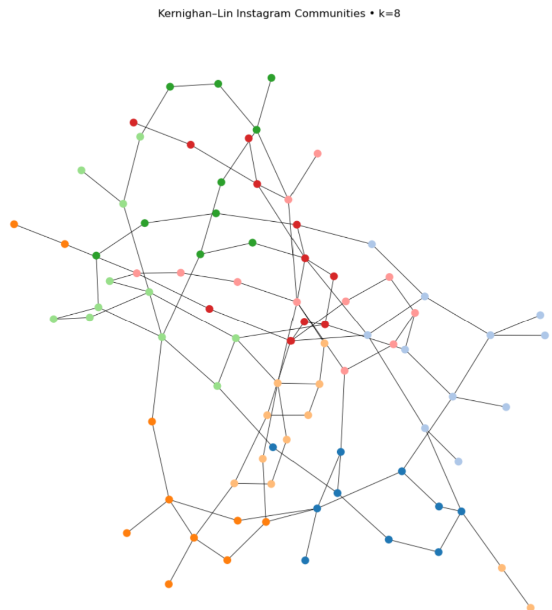

Figure 15: Kernighan Lin Instagram Graph

	community	size	internal_edges	density	avg_degree	clustering	cut_edges	conductance
0	0	10	10	0.222222	2.000000	0.0	7	0.259259
1	1	11	10	0.181818	1.818182	0.0	7	0.259259
2	2	10	8	0.177778	1.600000	0.0	6	0.272727
3	3	11	10	0.181818	1.818182	0.0	7	0.259259
4	4	10	8	0.177778	1.600000	0.0	6	0.272727
5	5	11	12	0.218182	2.181818	0.0	8	0.250000
6	6	11	10	0.181818	1.818182	0.0	12	0.375000
7	7	11	10	0.181818	1.818182	0.0	11	0.354839

Figure 16: Kernighan Lin Instagram Metrics

Finally, for the Instagram top 100 network, we have found 8 communities using the Kernighan Lin algorithm, each having between 10 to 11 nodes in size as seen in **Figure 15**. The internal edges range from 8 to 12 and average degree from 1.6 to 2.2. The conductance figures are considerably lower but vary more as they range from 0.25 to 0.38 indicating a weaker sense of communities found using this algorithm in comparison to the Twitter and Facebook networks.

5 Conclusion

In this report we have analysed the networks of three different platforms, Facebook, Twitter and Instagram. We identified the top 100 nodes in each platform that have the most connections in the network and applied two community detection algorithms, Girvan Newman and Kernighan Lin.

We have found that Twitter is the most densely connected network due to high average degree, short average path lengths and strong clustering coefficients. Due to these properties, we find that using Girvan Newman algorithm, a major community is detected because of the interconnectedness of the network.

On the other hand, Instagram had very sparse connectivity and a low clustering coefficient with also longer shortest path lengths. This meant that using both Girvan Newman and Kernighan Lin algorithms we could identify many smaller communities of similar sizes.

We also observe some differences between the algorithms Girvan Newman and Kernighan Lin, where Girvan Newman is very sensitive to edge density and performs best where there are clear bridges existing in the network. In contrast to this, we can see Kernighan Lin returns a more balanced result in the sizes of communities in both dense and sparse networks.

6 References

[1] Kaggle, "Huawei Social Network Data," https://www.kaggle.com/datasets/huawei-social-network-data