Facebook Social Circle Dataset Community Detection Analysis

Dataset Overview

The Facebook Social Circle Dataset was used for this analysis to analyze community detection within the network. The network itself can be described as an undirected network that is compiled of 4,039 nodes and 88,235 edges, where nodes represent Facebooks profiles and edges represent friendship connections. Each component of the network is referred to as a community. The remainder of this analysis will take a deeper look into the communities of this dataset using the following methods: modularity, graph clustering, hierarchical clustering, and k-clique communities.

Community Detection

In general, network interaction provides rich information about the relationship between entities. A community can be defined as a set of nodes between which the interactions are "relatively" frequent. Communities are also known as groups, subgroups, modules, or clusters. Furthermore, Facebook friends within the given network can be broken down into groups, where the groups can be categorized as communities. Community detection can then be identified as formalizing strong groups based on network properties. Therefore, given a node, the output of the community detection within the Facebook network is a community membership of "some" nodes. After further analyzing, 13 communities were found within the Facebook network and it can be confirmed that the network follows hierarchy-centric community detection. The community detection within the Facebook Social Circle network can be further analyzed through assessing the modularity of both the entire network and the top three communities.

Modularity of Entire Network

Good communities within a network can be characterized as having high modularity. In addition, maximizing modularity measure is a way of doing graph clustering and hierarchical graph clustering. The Facebook Social Circle network as a whole has a rather high modularity of approximately 0.83. Therefore, it can be assumed that the Facebook network has dense connections between the nodes within the network and has a strong ability to divide the network into groups or communities. Figure 1 provides a visualization for the modularity breakdown of the network in the form of a Gephi graph where each color represents a modularity class.
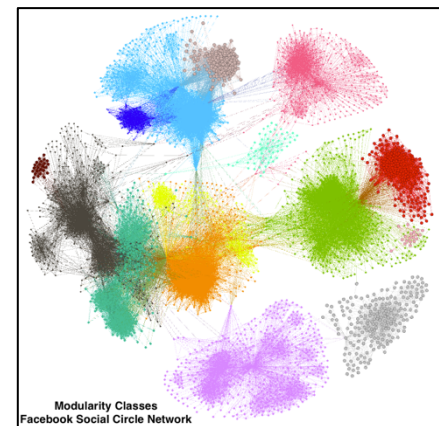

Modularity Classes
Facebook Social Circle Network

Figure 1

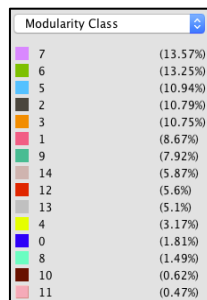| Modularity Class	
7	(13.57%)
6	(13.25%)
5	(10.94%)
2	(10.79%)
3	(10.75%)
1	(8.67%)
9	(7.92%)
14	(5.87%)
12	(5.6%)
13	(5.1%)
4	(3.17%)
0	(1.81%)
8	(1.49%)
10	(0.62%)
11	(0.47%)

Figure 2

As displayed in Figure 2, Modularity Class 7 has the highest modularity percentage within the network of 13.57%. It can be implied that Modularity Class 7 has very dense friendship connections between the Facebook profiles in comparison to the remaining modularity classes. This can further be proven by the analogous modularity class labeled in purple in Figure 1. From the visualization, we can see how the modularity class rankings compare to the color-coded nodes. For example, Modularity Class 10 has a smaller modularity in comparison to Modularity Class 5 and the difference in ranking can be seen between the aqua and brown node clusters.

The Facebook Social Circle network can be further characterized as having hierarchical modularity where the network can be broken down from the top down (also known as divisive hierarchical clustering) into two partition levels. In addition, this hierarchical modularity starts with a full node and breaks it into networks.

('partition at level', 0, 'is', {0: 0, 1: 1, 2: 2, 3: 3, 4: 4, 5: 5, 6: 6, 7: 7, 8: 8, 9: 9, 10: 10, 11: 11, 12: 12, 13: 13, 14: 14, 15: 10, 16: 15, 17: 0, 18: 16, 19: 9, 20: 17, 21: 18, 22: 5, 23: 19, 24: 20, 25: 21, 26: 22, 27: 23, 28: 4, 29: 24, 30: 25, 31: 26, 32: 27, 33: 24, 34: 28, 35: 29, 36: 22, 37: 7, 38: 25, 39: 22, 40: 30, 41: 31, 42: 32, 43: 33, 44: 34, 45: 11, 46: 13, 47: 35, 48: 36, 49: 37, 50: 38, 51: 39, 52: 40, 53: 41, 54: 40, 55: 27, 56: 42, 57: 4 3, 58: 44, 59: 0, 60: 34, 61: 45, 62: 5, 63: 46, 64: 47, 65: 48, 66: 29, 67: 7, 68: 42, 69: 49, 70: 50, 71: 51, 72: 1 3, 73: 36, 74: 52, 75: 53, 76: 54, 77: 55, 78: 56, 79: 50, 80: 57, 81: 12, 82: 1, 83: 56, 84: 58, 85: 59, 86: 60, 87: 8, 88: 26, 89: 39, 90: 35, 91: 44, 92: 21, 93: 61, 94: 62, 95: 42, 96: 45, 97: 9, 98: 43, 99: 32})
('partition at level', 1, 'is', {0: 0, 1: 1, 2: 2, 3: 3, 4: 4, 5: 1, 6: 5, 7: 4, 8: 6, 9: 7, 10: 8, 11: 9, 12: 6, 13: 7, 14: 10, 15: 8, 16: 11, 17: 0, 18: 12, 19: 7, 20: 13, 21: 14, 22: 1, 23: 15, 24: 16, 25: 17, 26: 18, 27: 19, 28: 4, 29: 1, 30: 20, 31: 21, 32: 22, 33: 1, 34: 23, 35: 24, 36: 18, 37: 4, 38: 20, 39: 18, 40: 25, 41: 26, 42: 27, 43: 28, 44: 29, 45: 9, 46: 7, 47: 30, 48: 6, 49: 31, 50: 32, 51: 27, 52: 17, 53: 33, 54: 17, 55: 22, 56: 29, 57: 22, 58: 34, 59: 0, 60: 29, 61: 9, 62: 1, 63: 35, 64: 36, 65: 37, 66: 24, 67: 4, 68: 29, 69: 38, 70: 29, 71: 39, 72: 7, 73: 6, 74: 40, 75: 41, 76: 42, 77: 43, 78: 0, 79: 29, 80: 44, 81: 6, 82: 1, 83: 0, 84: 45, 85: 46, 86: 47, 87: 6, 88: 21, 89: 2 7, 90: 30, 91: 34, 92: 17, 93: 48, 94: 49, 95: 29, 96: 9, 97: 7, 98: 22, 99: 27})

Figure 3

Level 0 is the first partition, which contains the smallest communities and has a modularity of approximately 0.73, while level 1 contains larger communities with a modularity of approximately 0.80. In essence, the higher the partition level, the larger the communities. Furthermore, both partition levels consist of 99 nodes, where the keys are the nodes and the values are the group it belongs to. Plotting the dendrogram is quite tedious, but a breakdown of the partition levels can be found in Figure 3. In addition to the entire network, the top 3 communities within the network and their modularity can be identified and analyzed based on the best partition of the graph.

Modularity of Top 3 Communities

After analyzing the best partition and communities within the Facebook Social Circle network, the top 3 communities within the network can be identified as node 13, node 27, and node 3. These three nodes have the highest group membership within the network. It can be implied that the Facebook profiles are connected to the greatest amount of profiles in comparison to the remaining communities within the network. An analysis of each community can be found below.

Top Community - Node 13

Facebook profile 13 can be categorized as the overall top community within the network with a group membership of 7. The average degree amongst the community is approximately 1.6667, while the highest number of cliques within the community is 2. Furthermore, the density of the community is approximately 0.33 and the average shortest path among the community is 2.33 hops. The

Betweenness Centrality	Closeness Centrality	Degree Centrality
(42, 0.6)	(42, 0.5556)	(70, 0.4)
(63, 0.6001)	(63, 0.5556)	(42, 0.4)
(70, 0.4)	(70, 0.4545)	(20, 0.4)
(20, 0.4)	(20, 0.4545)	(63, 0.4)
(76, 0)	(76, 0.3333)	(76, 0.2)
(14, 0)	(14, 0.3333)	(14, 0.2)

Table 1

community has a diameter of 5 and a modularity of approximately 0.3. In addition, the betweenness, closeness, and degree centrality rankings within this community can be found in Table 1. It can be further implied that node 70 has the higher number of friendship connections, while nodes 42 and 63 are the more central nodes.

In addition, the k-cliques found in this community can be classified as [20, 70, 42, 76, 14, 63] and the max cliques can be identified as [[42, 20], [42, 63], [20, 14], [70, 76], [70,63]]. It can be confirmed that the largest geodesic distance between the nodes in the maximal subgraph is <= k. Furthermore, it can be implied that nodes 70, 42, 20, and 63 have the highest number of cliques with 2 cliques existing for each of the nodes.

Top Community 2 – Node 27

Facebook profile ID 27 can be categorized as the second overall top community within the network with a group membership of 7. The average degree amongst the community is approximately 1.6667 and the highest number of cliques within the community is 2. In addition, the density of this community is approximately 0.33, while the average shortest path among the community is 2.3333 hops. The

Betweenness Centrality	Closeness Centrality	Degree Centrality
(66, 0.6001)	(66, 0.5556)	(66, 0.4)
(54, 0.6001)	(54, 0.5556)	(72, 0.4)
(72, 0.4)	(72, 0.4545)	(44, 0.4)
(44, 0.4)	(44, 0.4545)	(54, 0.4)
(33, 0.0)	(33, 0.3333)	(33, 0.2)
(83, 0.0)	(83, 0.3333)	(83, 0.2)

Table 2

community has a diameter of 5 and a modularity of approximately 0.26. Furthermore, the betweenness, closeness, and degree centrality rankings of this community can be found in Table 2. It can be concluded that nodes 66, 72, 44, and 54 have an equivalent amount of friendship connections within the community, while nodes 66 and 54 are the more central nodes within the community.

In addition, the k-cliques found in this community can be classified as [33, 66, 83, 54, 72, 44] and the max cliques can be identified as [[33, 44], [66, 72], [66, 54], [83, 72], [44, 54]]. It can be further implied that 66, 72, 44, and 52 have the highest number of cliques with 2 cliques existing for each of the nodes and the largest geodesic distance between the nodes in the maximal subgraph is <= k.

Top Community 3 – Node 3

Facebook profile ID 3 can be categorized as the third overall top community within the network with a group membership of 6. The average degree amongst the community is approximately 1.6, while the highest number of cliques within the community is 2. Furthermore, the density of the community is approximately 0.4 and the average shortest path

Betweenness Centrality	Closeness Centrality	Degree Centrality
(35, 0.6666)	(35, 0.6666)	(35, 0.5)
(15, 0.5)	(15, 0.5714)	(47, 0.5)
(47, 0.5)	(14, 0.5714)	(15, 0.5)
(3, 0)	(3, 0.4)	(3, 0.25)
(62, 0.0)	(62, 0.4)	(62, 0.25)

Table 3

among the community is 2 hops. The community has a modularity of approximately 0.21875 and a diameter of 4. In addition, the betweenness, closeness, and degree centrality rankings within this community can be found in Table 3. According to Table 3, nodes 35, 47, and 15 have an equivalent amount of the largest friendship connections within the community, while nodes 35 and 15 can be characterized as the more central nodes.

In addition, the k-cliques found in this community can be classified as [3, 35, 15, 62, 47] where the largest geodesic distance between the nodes in the maximal subgraph is <= k. The max cliques can be identified as [[15, 35], [15, 3], [62, 47], [47, 35]] and it can be further implied that nodes 15, 35, and 47 have the highest number of cliques within the community with 2 cliques existing for each of the nodes.

With these findings in mind relating to the top 3 communities within the network, several analyses can be made. Nodes 13 and 27 have the highest degree amongst the top 3 communities, which implies that these nodes are connected to the greatest amount of profiles within their community. In addition, although nodes 13 and 27 have the highest degree value, node 3 has the greatest amount of density within its community. Therefore, it can be speculated that Facebook profile ID 3 has a larger portion of potential friendship connections in the network

that are actual friendship connections. Furthermore, all three of the communities have a maximum number of cliques of 2.

In addition to the analysis of the top 3 communities within the network, a hierarchical structure for each of the 3 communities can be found below. According to the analysis, nodes 13 and 27 have a structure of 6 nodes and 5 paths, while node 3 has a structure of 5 nodes and 4 paths.
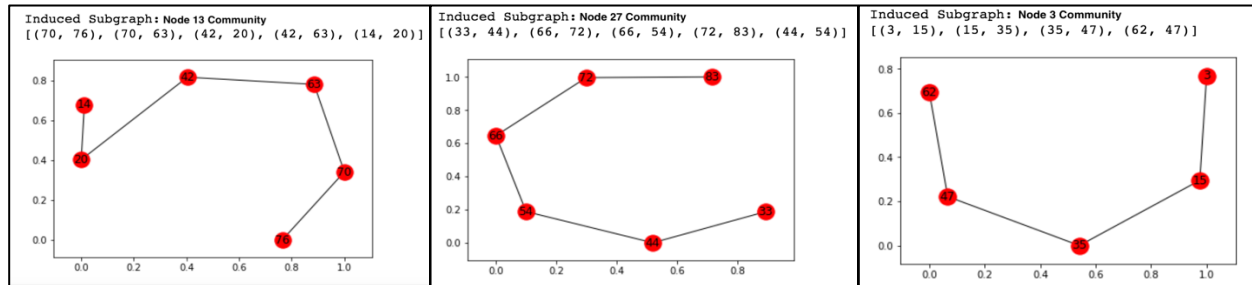
Furthermore, the partition levels for each of the communities are: Node 13 - {20: 1, 70: 0, 42: 1, 76: 0, 14: 1, 63: 0}, Node 27 - {33: 0, 66: 1, 83: 2, 54: 1, 72: 2, 44: 0}, Node 3 - {35: 0, 3: 1, 47: 0, 62: 0, 15: 1}. As previously mentioned, node 3 has a denser structure, while nodes 13 and 27 have a structure with a larger diameter.

Overall Analysis of Communities

Overall, the Facebook Social Circle network is compiled of 13 communities and a network characterized by high modularity, where the nodes in the community are reachable with the same or similar distances. The network can be broken down into multiple partition levels, but for the purpose of this analysis, the dendrogram provides a breakdown of 2 partition levels where 99 nodes exist within each level. In addition, the network reflects dense and very tight-knit friendship connections. From the analysis, it can be concluded that the top 3 communities within the network are Facebook profile 13, 27, and 3. In addition, each of the top 3 communities have maximal cliques of 2. Lastly, a small k-clique analysis of a 2-clique can be found in Figure 4, where every node in the community is reachable within the same distance of 2 steps. Unlike a k-club, the k-clique is allowed to go outside of the subgraph to reach if necessary as well.

```
G=nx.erdos_renyi_graph(100, 0.01)
G.add_edges_from(G.edges())
c = list(nx.k_clique_communities(G, 2))
print(c[0])

frozenset([1, 2, 4, 6, 71, 8, 79, 75, 76, 66, 78, 15,
16, 17, 19, 21, 86, 87, 89, 27, 92, 93, 96, 34, 99, 4
1, 43, 53, 57, 58, 63])
```

Figure 4