Christina Hill
2.28.18

Facebook Social Circle Dataset Community Detection Analysis

Dataset Overview

The Facebook Social Circle Dataset was used for this analysis to further 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

As previously mentioned, this analysis will focus on analyzing community detection within the network. In general, network interaction provides rich information about the relationship between entities. With this in mind, 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. 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. 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.

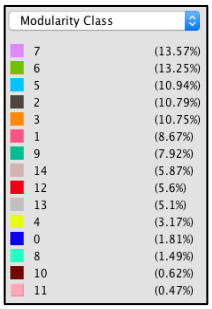


Figure 1

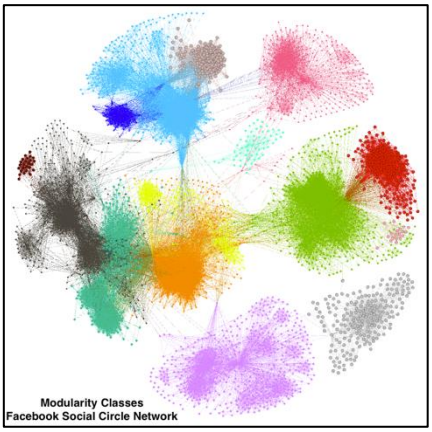


Figure 2

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.

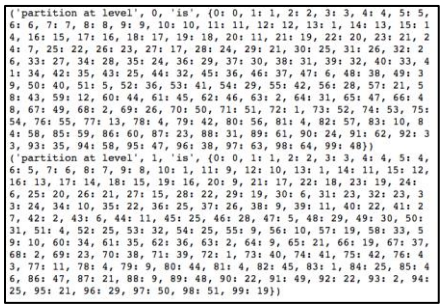


Figure 3

The Facebook Social Circle network can be further characterized as having hierarchical modularity where the network can be broken down from the top down into two partition levels. Level 0 is the first partition, which contains the smallest communities, while level 1 contains larger communities. 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. A breakdown of the partition levels can be found in Figure 3. Through the information gathered from the partition levels, the top 3 communities within the network can be identified and analyzed.

Modularity of Top 3 Communities

After analyzing the partition levels and communities within the Facebook Social Circle network, the top 3 communities within the network can be identified as node 8, node 4, and node 0. 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 1

Facebook profile ID 8 can be categorized as the overall top community within the network with a group membership of 9. The average degree amongst the community is approximately 1.75, while the highest number of cliques within the community is 2. Furthermore, the density of the community is approximately 0.25 and the average shortest path among the community is 3 hops. The modularity of the community is approximately 0.81. 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 66 has the higher number of friendship connections, while node 42 and 54 respectively are the more central nodes.

Betweenness Centrality	Closeness Centrality	Degree Centrality
(42, 0.5714)	(42, 0.4375)	(66, 0.2857)
(54, 0.5714)	(54, 0.4375)	(71, 0.2857)
(66, 0.4761)	(66, 0.3888)	(42, 0.2857)
(26, 0.4761)	(26, 0.3888)	(21, 0.2857)
(71, 0.2857)	(71, 0.3181)	(54, 0.2857)
(21, 0.2857)	(21, 0.3181)	(26, 0.2857)
(9, 0.0)	(9, 0.25)	(9, 0.1428)
(81, 0.0)	(81, 0.25)	(81, 0.1428)

Table 1

In addition, the k-cliques found in this community can be classified as [frozenset([81, 66, 21, 54, 71, 9, 42, 26])] and the max cliques can be identified as [[81, 71], [66, 54], [66, 71], [21, 9], [21, 26], [42, 26], [42, 54]]. It can be further implied that nodes 66, 71, 42, 21, 54, and 26 have the highest number of cliques with 2 cliques existing for each of the nodes. From these findings, it can be concluded that the largest geodesic distance between the nodes within the community are <=k.

Top Community 2

Betweenness Centrality	Closeness Centrality	Degree Centrality
(48, 0.6001)	(48, 0.5556)	(46, 0.4)
(55, 0.6001)	(55, 0.5556)	(48, 0.4)
(46, 0.4)	(46, 0.4545)	(55, 0.4)
(62, 0.4)	(62, 0.4545)	(62, 0.4)
(5, 0.0)	(5, 0.3333)	(80, 0.2)
(80, 0.0)	(80, 0.3333)	(5, 0.2)

Table 2

Facebook profile ID 4 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 1.667 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 2.33 hops. The modularity of the community is approximately 0.88. Furthermore, the betweenness, closeness, and degree centrality rankings within this community can be found in Table 2. It can be concluded that nodes 46, 48, 55, and 62 have an equivalent amount of friendship connections within the community, while nodes 48 and 45 are the more central nodes within the community.

In addition, the k-cliques found in this community can be classified as [frozenset([80, 48, 46, 55, 62, 5])] and the max cliques can be identified as [[80, 62], [48, 46], [48, 55], [5, 46], [62, 55]]. It can be further implied that the largest geodesic distance between the nodes within the community are <=k, while nodes 48, 46, 55, and 62 have the highest number of cliques with 2 cliques existing for each of the 4 nodes.

Community 3

Betweenness Centrality	Closeness Centrality	Degree Centrality
(3, 0.6001)	(3, 0.5556)	(0, 0.4)
(59, 0.6001)	(59, 0.5556)	(3, 0.4)
(0, 0.4)	(0, 0.4545)	(59, 0.4)
(61, 0.4)	(61, 0.4545)	(61, 0.4)
(1, 0.0)	(1, 0.3333)	(1, 0.2)
(72, 0.0)	(72, 0.3333)	(72, 0.2)

Table 3

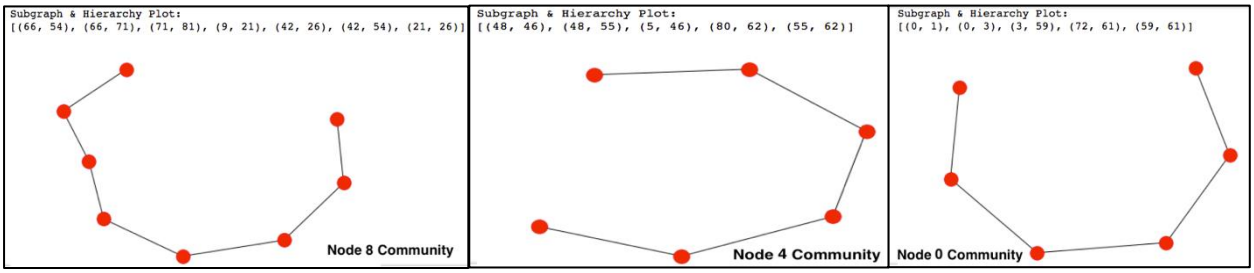
Facebook profile ID 0 can be categorized as the third overall top community within the network with a group membership of 7 as well. The average degree amongst the community is approximately 1.6667, while the highest number of cliques within the community is 2. Furthermore, the density of the community is approximately 0.33 and the average shortest path among the community is 2.33 hops. The modularity of the community is approximately 0.89. In addition, the betweenness, closeness, and degree centrality rankings within this community can be found in Table 3. According to Table 3, nodes 0, 3, 59, and 61 have an equivalent amount of the largest friendship connections within the community, while nodes 3 and 59 can be characterized as the more central nodes.

In addition, the k-cliques found in this community can be classified as [frozenset ([0, 1, 3, 72, 59, 61])] and the max cliques can be identified as [[0, 1], [0, 3], [72, 61], [59, 3], [59, 61]]. It can be further implied that nodes 0, 3, 59, and 61 have the highest number of cliques with 2 cliques existing for each of the 4 nodes.

With these findings in mind relating to the top 3 communities within the network, several analyses can be made. Node 8 has the highest degree amongst the top 3 communities, which implies that node 8 is connected to the greatest amount of profiles within their community. In addition, although node 8 has the highest degree value, nodes 4 and 0 have the greatest

amount of density within their community. Therefore, it can be speculated that Facebook profile ID 4 and Facebook profile ID 0 have 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, which confirms that none of the nodes within the community had a geodesic distance between the nodes that was <=k.

In addition to the analysis of the top 3 communities within the network, a hierarchical structure of the 3 communities can be found below.



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 further broken down into 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 8, 4, and 0. In addition, each of the top 3 communities have maximal cliques of 2.