# Facebook Social Circle Dataset Analysis

## Dataset Information Summary

The Facebook Social Circle Dataset information that was used during this analysis is a dataset that consists of "circles" (also known as "friends lists") from Facebook. The Facebook data was compiled from survey participants using the Facebook app. In addition, the dataset included node features (also known as Facebook profiles), circles, and ego networks (Leskovec, 2012). The remainder of this analysis will take a deeper look into the dataset's connectivity, degree distribution, centrality, and the small world notion associated with the network.

## Defining The Network

The Facebook Social Circle network is defined by friendship relationships. As provided in Figure 1, the network itself can be described as an undirected network that is compiled of 4,039 nodes and 88,234 edges. Within the network, the nodes represent Facebook profiles, while the edges refer to friendship connections. When an edge connects two nodes, a friendship connection has formed between the two Facebook profiles. With this in mind, the friendship connection is therefore characterized as undirected and bidirectional where if Person A and Person B are friends, then Person B and Person A are friends as well. The connection between the network's nodes and edges can be further analyzed by assessing the network's connectivity.

```
In [30]: print nx.info(G)
         print nx.is_directed(G)

         Name:
         Type: Graph
         Number of nodes: 4039
         Number of edges: 88234
         Average degree:  43.6910
         False
```

*Figure 1*

## Connectivity

As an undirected network, the Facebook Social Circle network does not consist of weakly connected components or strongly connected components. Both component types generally exist only within directed networks. This can be further proven because each Facebook profile (node) within the component cannot be reached from every other Facebook profile in the component by following directed friendship connections (links). In addition, every Facebook profile cannot be reached from every other Facebook profile by following friendship connections in either direction. Therefore, the network consists of zero disconnected components and exactly one connected component, where two Facebook profiles are connected to each other through friendship connections, and are not connected to additional Facebook profiles. In addition to the network's connectivity, the degree of the network can be analyzed to correspond with friendship connections.

## Degree

The degree of a node within a network can be defined as the number of adjacent edges associated with the node. The degrees found within the Facebook Social Circle network directly correlates with the number of friends a person has on their Facebook profile. Based on the degrees found within the network, we can sort the degrees based on various scenarios. For example, in Table 1, the network nodes and their corresponding degrees are listed in descending order. This is an exact representation of the top five Facebook profiles within the network that have the highest number of friends. From Table 1, we can see that Facebook profile 107 has a total of 1,045 friends and the highest number of friends within the network.

| TOP 5 Facebook Profiles | |
|---|---|
| **Facebook Profile ID** | **# of Friends** |
| **107** | 1,045 |
| **1684** | 792 |
| **1912** | 755 |
| **3437** | 547 |
| **0** | 347 |

*Table 1*

Aside from assessing which users have the highest number or friends, we are able to further analyze the network degrees to evaluate which Facebook users have the most influence within the network. Nodes with higher degrees represent Facebook profiles that have a greater influence on their network, while nodes with lower degrees represent Facebook profiles who have considerably less influence within their network. For example, if we were to reference Table 1, one could make an educated assumption that based on the number of friends associated with Facebook profile ID 0 vs Facebook Profile ID 1912, that Facebook profile ID 1912 has a greater influence and reach in the network because they have 408 more friends than Facebook profile ID 0.

In addition to examining the amount of influence within the Facebook Social Circle network, the average degree of the network can provide vital information. The average degree of the network can be classified as the average number of friends each Facebook profile has within the network, or the number of friendship connections (edges) that are incident on a Facebook profile (node). As previously displayed in Figure 1, the average degree of the network is approximately 43.6910. From this approximation, we are able to confidently assume that out of the 4,039 Facebook profiles that are present within the network, each Facebook profile has an average of roughly 43 friends. The degree of the nodes can be further analyzed for centrality.

**Centrality Overview**

Although centrality has varying definitions, for this specific network, we will define centrality as the measure of influence and access a Facebook profile has in the network. Within the Facebook Social Circle Network and as displayed in Figure 2, node 107 has the overall highest degree centrality, closeness centrality, and betweenness centrality. According to these findings, it can be assumed that Facebook profile ID 107 is a very important person who is connected to the highest number of friends and is the main person who functions as a broker within the network. A deeper analysis of each centrality can be found below.

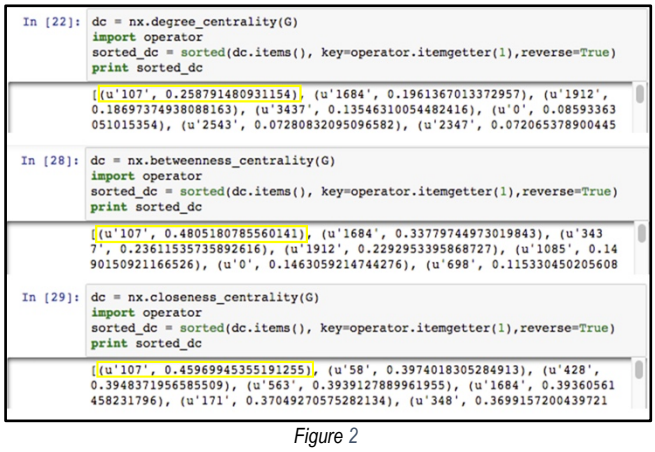


*Figure 2*

**Degree Centrality**

The degree centrality of the network can be classified as, “He who has many friends, is the most important.” In other words, a Facebook profile that has a high number of friendship connections, is considered to be a very central, visible, and important person. As the node with the highest degree centrality of approximately 0.2587, Facebook profile ID 107 can be defined as the most important and central person within the network. This analysis can further be confirmed by the data previously represented in Table 1 where Facebook profile ID 107 had the highest number of friends amongst the network. Based on node 107’s top degree centrality, the assumption can be made that node 107 has access to a lot of friends; associates with various types of people; and those friendship connections in the physical world are being emulated in the online Facebook environment.

**Closeness Centrality**

Aside from assessing node degrees, the closeness centrality represented in the network considers the lengths of the shortest paths between a node and the remaining nodes within the

network. From the Facebook network perspective, closeness centrality can be characterized as: the more central or number of friends a person has, the closer that person is to all the other Facebook profile IDs within the network. As previously displayed in Figure 2, node 107 has the highest closeness centrality of approximately 0.4596 amongst 4,039 Facebook profiles. From these findings, it can be deduced that as Facebook profile ID 107 grows closer (distance-wise) to the other Facebook profiles, its centrality will increase because the path lengths are becoming shorter between Facebook profile ID 107 and the remaining Facebook profiles. Likewise, it can be stated that as Facebook profile ID 348 grows more distant from the other Facebook profiles, its centrality will decrease.

**Betweenness Centrality**

Unlike closeness centrality, betweenness centrality considers whether a node (Facebook profile) is along the path or is not along the path within a network. The betweenness centrality also considers how a Facebook profile is functioning as a broker and how many other Facebook profiles have to go through that specific Facebook profile. For example, the friendship connection between Facebook profile ID 563 and Facebook profile ID 428 could depend on Facebook profiles that are on the path between those two Facebook profiles.

With this in mind, Figure 2 provides statistical information that node 107 has the highest betweenness centrality of approximately 0.4805 within the network. It can be concluded that node 107 has the highest amount of profiles that have to go through it. This implies that a higher betweenness centrality within the Facebook network is dependent on if a profile has friendship connections between many other profiles along the path. In essence, the more important and higher the number of friends a profile has, the higher the betweenness centrality will be for that specific profile. The betweenness centrality helps assess which Facebook profiles can function as a broker of information between varying Facebook profile groups within the network and who has top influential power. Therefore, it could be concluded that node 1684 and node 343 would be better candidates for brokering information than node 0 and node 1698.

**Small World Notion**

Considering the summation of findings that have been presented thus far within this analysis, an examination of the small world notion should be considered for the Facebook Social Circle network. Within a small world network, most nodes are not neighbors of one another, but the neighbors of any given node are likely to be neighbors of each other and most nodes can be reached from every other node by a small number of hops or steps (Wikipedia, n.d.). From the Facebook network perspective, the notion of a small world network would hypothetically be: Facebook profiles that are not linked through a friendship connection, are linked by a short chain of associates. To further analyze the network's small world notion, the network would have to consist of the two characteristics of a small world network: high clustering and low average shortest path.

Networks that are highly clustered reflect the proven notion that people you know, also know each other. A snippet of the clustering coefficient for each node within the network can be found in descending order in Figure 3. According to the data displayed concerning each nodes'

```
In [4]: dc = nx.clustering(G)
        import operator
        sorted_dc = sorted(dc.items(), key=operator.itemgetter(1),reverse=True)
        print sorted_dc

[(u'4025', 1.0), (u'4028', 1.0), (u'4029', 1.0), (u'3990', 1.0), (u'293', 1.0), (u'273', 1.
0), (u'279', 1.0), (u'581', 1.0), (u'2441', 1.0), (u'102', 1.0), (u'90', 1.0), (u'2860', 1.
0), (u'1755', 1.0), (u'1176', 1.0), (u'401', 1.0), (u'454', 1.0), (u'4033', 1.0), (u'4032',
1.0), (u'4034', 1.0), (u'4036', 1.0), (u'379', 1.0), (u'377', 1.0), (u'393', 1.0), (u'374
6', 1.0), (u'3744', 1.0), (u'244', 1.0), (u'241', 1.0), (u'652', 1.0), (u'2691', 1.0), (u'1
967', 1.0), (u'1762', 1.0), (u'1546', 1.0), (u'699', 1.0), (u'1627', 1.0), (u'1383', 1.0),
(u'3582', 1.0), (u'1253', 1.0), (u'256', 1.0), (u'255', 1.0), (u'2204', 1.0), (u'160', 1.
0), (u'3689', 1.0), (u'3688', 1.0), (u'3686', 1.0), (u'110', 1.0), (u'2517', 1.0), (u'273
6', 1.0), (u'802', 1.0), (u'1776', 1.0), (u'321', 1.0), (u'327', 1.0), (u'1558', 1.0), (u'1
395', 1.0), (u'3813', 1.0), (u'3811', 1.0), (u'220', 1.0), (u'2219', 1.0), (u'154', 1.0),
(u'2431', 1.0), (u'2505', 1.0), (u'2722', 1.0), (u'2721', 1.0), (u'2855', 1.0), (u'2857',
1.0), (u'1394', 1.0), (u'2980', 1.0), (u'1016', 1.0), (u'1233', 1.0), (u'306', 1.0), (u'380
8', 1.0), (u'3807', 1.0), (u'234', 1.0), (u'233', 1.0), (u'145', 1.0), (u'2671', 1.0), (u'1
486', 1.0), (u'1276', 1.0), (u'476', 1.0), (u'1224', 1.0), (u'2885', 1.0), (u'3879', 1.0),
(u'3875', 1.0), (u'201', 1.0), (u'205', 1.0), (u'207', 1.0), (u'228', 1.0), (u'3653', 1.0),
(u'2704', 1.0), (u'3317', 1.0), (u'785', 1.0), (u'468', 1.0), (u'153', 1.0), (u'3003', 1.
0), (u'1586', 1.0), (u'1372', 1.0), (u'3911', 1.0), (u'3910', 1.0), (u'3147', 1.0), (u'21
6', 1.0), (u'218', 1.0), (u'3489', 1.0), (u'63', 1.0), (u'1854', 1.0), (u'1856', 1.0), (u'9
10', 1.0), (u'3467', 1.0), (u'750', 1.0), (u'1027', 1.0), (u'1034', 1.0), (u'467', 1.0),
(u'78', 1.0), (u'485', 1.0), (u'3922', 1.0), (u'3992', 1.0), (u'3854', 1.0), (u'195', 1.0),

In [5]: print nx.average_clustering(G)

0.60554671862
```

*Figure 3*

clustering coefficient, it can be implied that a clustering coefficient of 1 represents friendship connections where all of a Facebook profile's friends know each other. After analyzing each nodes' clustering coefficient, it was found that a sizeable percentage of the nodes had a clustering coefficient of 1. In addition, after taking a further look at the nodes that were towards the bottom of the list of clustering coefficients, it was discovered that clustering coefficients varied between 0 and 0.9883. Along with these findings, the average clustering coefficient for the network was calculated to be approximately 0.6055, which demonstrates relatively high clustering. As a result, it can be implied that the Facebook network consists of high and heavily clustered groups with a dense amount of friendship connections.

Given 2 nodes (Facebook profiles), the shortest path length is calculated by the minimal number of links that lead from node A to node B or from node B to node A. Figure 4 shows a small snippet of the shortest paths found within the Facebook Social Circle network and the number of hops present for node 801. As displayed, there are quite a few links that exist between node 801 and another node. After further calculations, it was discovered that the average shortest path from one node to another node is approximately 3.6925, which is a relatively low average shortest path. This implies that on average, it takes roughly 3.6 links to "hop" from one Facebook profile to another.

```
christinahill — python — 81×47
>>> dc = nx. shortest_path_length(G)
import operator
sorted_dc = sorted(dc.items(),key=operator.itemgetter(1),reverse=True)
print sorted_dc
>>> import operator
>>> sorted_dc = sorted(dc.items(),key=operator.itemgetter(1),reverse=True)
>>> print sorted_dc
[(u'801', {u'1718': 6, u'2031': 7, u'1200': 6, u'2602': 7, u'4026': 8, u'1869': 6
, u'4024': 8, u'4025': 8, u'4022': 8, u'4023': 8, u'4020': 8, u'1868': 6, u'4028'
: 8, u'4029': 8, u'344': 7, u'345': 7, u'346': 7, u'347': 7, u'340': 7, u'341': 7
, u'342': 7, u'343': 7, u'3997': 8, u'3996': 8, u'3995': 8, u'3994': 8, u'348': 5
, u'349': 6, u'3991': 8, u'3990': 8, u'2318': 7, u'2319': 7, u'2316': 7, u'2317':
 7, u'2314': 7, u'2315': 7, u'2312': 7, u'2313': 7, u'2310': 7, u'2311': 7, u'298
': 7, u'299': 7, u'296': 7, u'297': 7, u'294': 7, u'295': 7, u'292': 7, u'293': 7
, u'290': 7, u'291': 7, u'3773': 4, u'3772': 4, u'3771': 4, u'3770': 4, u'3777':
4, u'3776': 4, u'3775': 4, u'3774': 4, u'2147': 7, u'3779': 4, u'3778': 4, u'270'
: 7, u'271': 7, u'272': 7, u'1138': 6, u'274': 7, u'275': 7, u'276': 7, u'277': 7
, u'278': 7, u'279': 7, u'1780': 6, u'2146': 7, u'2268': 7, u'2269': 7, u'1132':
6, u'2262': 7, u'2263': 7, u'2260': 7, u'2261': 7, u'2266': 7, u'2267': 7, u'2264
': 7, u'2265': 7, u'2442': 7, u'2443': 7, u'2440': 7, u'2441': 7, u'1781': 6, u'2
447': 7, u'2444': 7, u'2445': 7, u'2448': 7, u'2449': 7, u'2189': 7, u'108': 7, u
'109': 7, u'102': 7, u'103': 7, u'100': 7, u'101': 7, u'106': 7, u'107': 5, u'104
': 7, u'105': 7, u'2046': 7, u'2047': 7, u'2044': 7, u'2045': 7, u'2042': 7, u'20
43': 7, u'2040': 7, u'2041': 7, u'2048': 7, u'2511': 7, u'1212': 6, u'2839': 5, u

In [6]: print nx.average_shortest_path_length(G)
        3.6925068497
```

*Figure 4*

With these findings, it can be implied that the network is compiled of a low average shortest path and high clustering. This further confirms the notion that the Facebook Social Circle network fulfills the small world notion.

**Conclusion**

In conclusion, the Facebook Social Circle network is an undirected network that establishes friendship relationships through friendship connections between Facebook profiles. The network reflects a very dense and tightly knit group of friends and friendship connections. It can be implied that the degree intensity of a Facebook profile is dependent on the amount of friendship connections that are associated with the profile. From the analysis, it was concluded that node 107 is the most valuable person within the network who had the greatest amount of access to friends and influence within the network. As a network that follows the small world notion, it can be assumed that Facebook profiles that are not linked through a friendship connection, will inevitably be linked by a short chain of associates.

**Works Cited**

Leskovec, J. (2012). Social circles: Facebok. Retrieved from SNAP: https://snap.stanford.edu/data/egonets-Facebook.html

Wikipedia. (n.d.). Small-world network. Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Small-world_network