Social Computing – Project 1

Results for the social network data provided to us as part of the data set yielded the following results for the ego nodes and overall data about the graph itself.

The screenshot above shows the details after running prog_1.py which used SNAP to analyze the data and provide the statistics. First a undirected graph is built using the data in the combined data set text file and then methods available in SNAP to calculate the Betweenness Centrality for each of the ego nodes and the clustering coefficient for the graph.

Hypothesis 1: A node with higher degree would belong to more number of circles

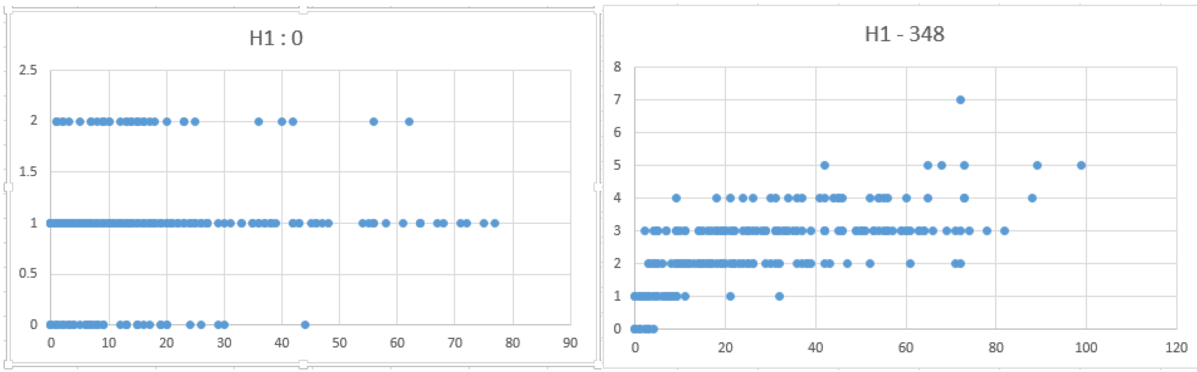

Figure 1a and 1b: Scatter Plots for the data associated with ego nodes 0 and 348, the plots are read as number of circles (y-axis) to number of nodes for the ego nodes (x-axis).

From figure 1a and 1b we see there is no real correlation between the numbers of circles a node is part off to the number of nodes it is connected to. For example for ego node 0, we have multiple nodes which are part of 2 circles but have a varying range of node degree (0 – 62), similarly for node 348 we have the node degree ranging from 5 to as much as 81 for the same circle count of 3. From the above graphs, with simple analysis we can conclude by saying the hypothesis is **false**.

Hypothesis 2: A circle with a higher density would have a greater degree of intersection with other circles

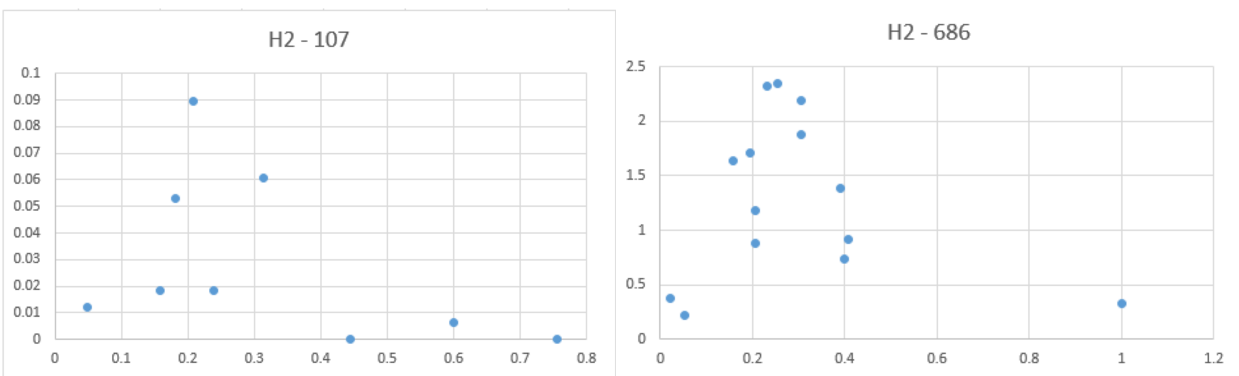

Figure 2a and 2b: Scatter Plots for the data associated with circles for ego nodes 107 and 686, the plots are read as density of the edges within a circle (x-axis) and degree of intersection with other circles (y-axis)

From figures 2a and 2b, we see there is no proof that high degree of intersection with other circles stems from the fact that the density of the circle is high. We have a case for ego node 686 where a circle with density 1, actually has a low degree of intersection with other circles while circles with moderate density actually have a higher intersection factor with other circles. From the simple example we can conclude that the hypothesis does not hold and is **false**.

Hypothesis 3: In a network, nodes with more common circles have more common features as compared to nodes with less common circles

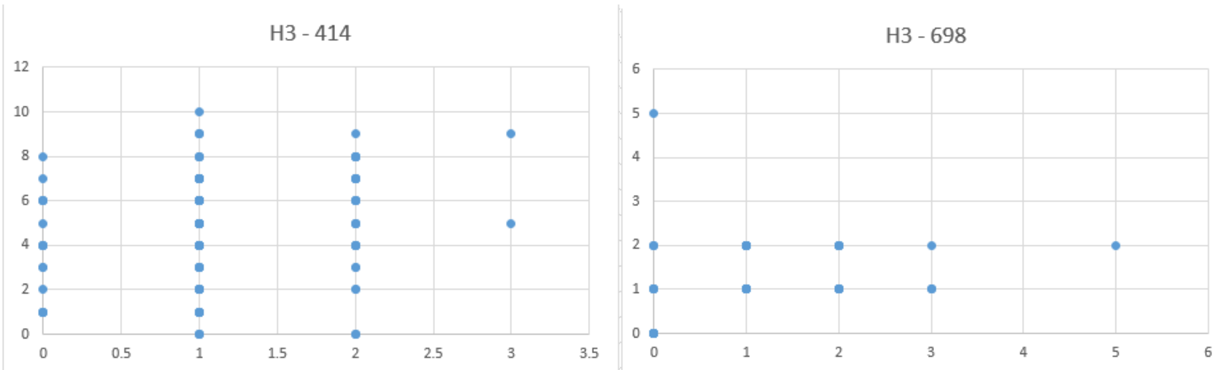

Figure 3a and 3b: Scatter plots for data associated with ego nodes 414 and 698 and nodes that have common circles, the plots are read as number of common features between a node and the ego node for which we have the data (y-axis) to the number of common circles they are part of (x-axis)

From figures 3a and 3b, we see there is no direct correlation between number of common features and number of common circles, we have multiple nodes in the data for both ego nodes 414 and 698 that have many common features with no common circles and many common circles with no common features (in 414 we have example of 2 common circles with 0 common features and 2 common circles and 11 common features). With this simple analysis it is easy to claim that the hypothesis 3 is not valid and hence **false**.