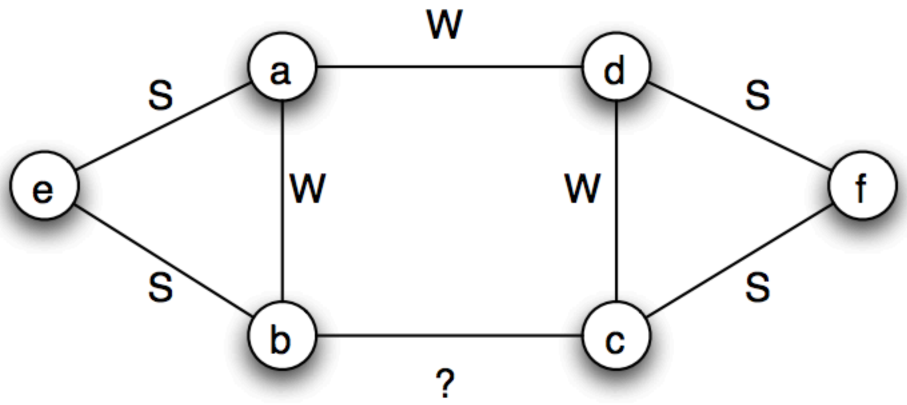
CS286 – Social Network Analysis
HW-1 Submission

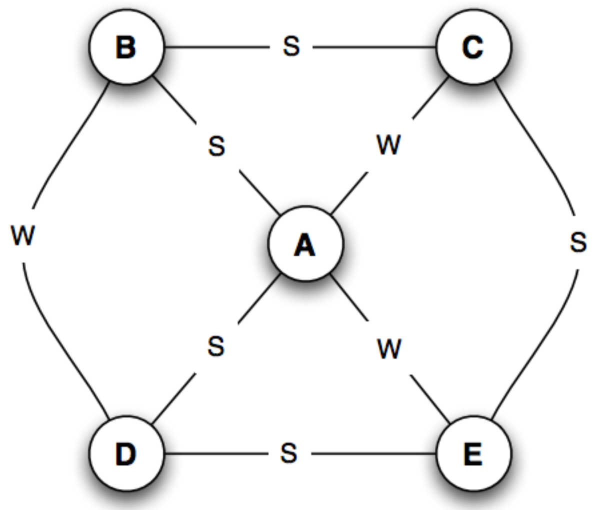
Ujjawal Garg (SJSU ID: 011917334)
ujjawal.garg@sjsu.edu

Sol 1 a).



B--C must be a Weak edge because if it was to be a strong edge, then the Strong Triadic Closure assumption would be violated by nodes B and C. In case of node B, since it has strong ties to both E & C, there must also exist an edge between E & C (whether weak or strong does not matter). But, this is not the case. Similar argument can be made for node C also. Thus, B--C must be weak edge.

Sol 1 b).

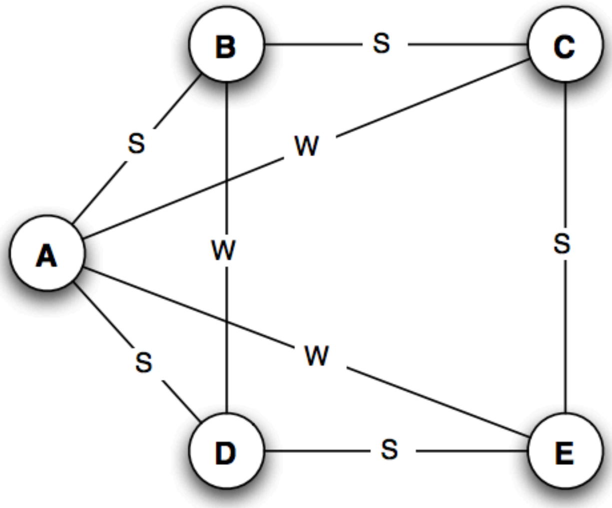


There are 5 Strong edges in the graph. We need to check ${}^{5}C_{2}$ (=10) cases to verify the Strong Triadic Closure (STC) property. Out of these 10, edges we only consider the cases where there is a common edge between the two edges. We are left with 5 cases to check.

Case 1: Consider connections B--A and B--C. Both are strong ties and an edge exists between nodes A & C. So, the Strong Triadic Closure property is not violated
Case 2: Consider connections D--A and D--E. Both are strong ties and an edge exists between nodes A & E. So, the Strong Triadic Closure property is not violated
Case 3: Consider connections C--B and C--E. Both are strong ties, but there is no edge between B and E. So, **node C violates the Strong Triadic Closure property**.
Case 4: Consider connections E--D and E--C. Both are strong ties, but there is no edge between D and C. So, **node E violates the Strong Triadic Closure property**.
Case 5: Consider connections A--B and A--D. Both are strong ties and an edge exists between nodes B & D. So, the Strong Triadic Closure property is not violated

We find that only the node E & node C violate Strong Triadic Closure property. Thus, other do not violate the Strong Triadic Closure and thus **nodes A, B, D must satisfy the Strong Triadic Closure.**

Sol 1 c).

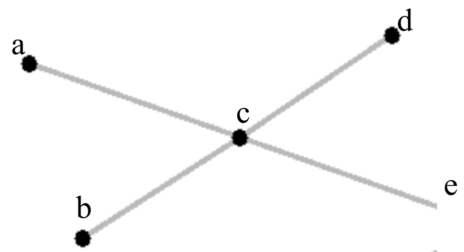


This figure is exactly same as above figure, just position of nodes is different. Again, there are 5 Strong edges in the graph. We need to check ${}^5C_2$ (=10) cases to verify the Strong Triadic Closure (STC) property. Out of these 10, edges we only consider the cases where there is a common edge between the two edges. We are left with 5 cases to check.

Case 1: Consider connections B--A and B--C. Both are strong ties and an edge exists between nodes A & C. So, the Strong Triadic Closure property is not violated
Case 2: Consider connections D--A and D--E. Both are strong ties and an edge exists between nodes A & E. So, the Strong Triadic Closure property is not violated
Case 3: Consider connections C--B and C--E. Both are strong ties, but there is no edge between B and E. So, **node C violates the Strong Triadic Closure property**.
Case 4: Consider connections E--D and E--C. Both are strong ties, but there is no edge between D and C. So, **node E violates the Strong Triadic Closure property**.
Case 5: Consider connections A--B and A--D. Both are strong ties and an edge exists between nodes B & D. So, the Strong Triadic Closure property is not violated

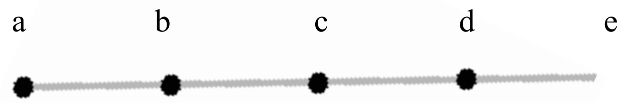
We find that only the node E & node C violate Strong Triadic Closure property. Thus, other **nodes A, B, D automatically satisfy the Strong Triadic Closure.**

Sol 3 a).



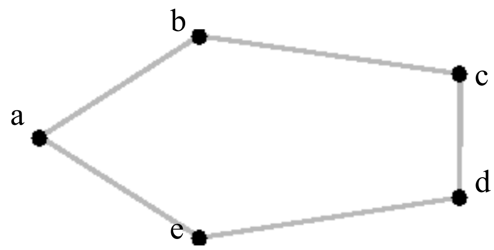
Node	Degree		Closeness		Between-ness	
	Actual	Normalized	Actual	Normalized	Actual	Normalized
a	1	0.25	0.14	0	0	0
b	1	0.25	0.14	0	0	0
c	4	1	0.25	6	6	1
d	1	0.25	0.14	0	0	0
e	1	0.25	0.14	0	0	0

Sol 3 b)



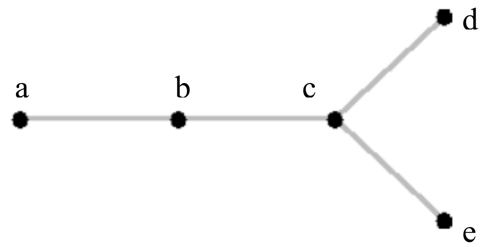
Node	Degree		Closeness		Between-ness	
	Actual	Normalized	Actual	Normalized	Actual	Normalized
a	1	0.25	0.1	0.4	0	0
b	2	0.5	0.14	0.57	3	0.5
c	2	0.25	0.167	0.67	4	0.67
d	2	0.25	0.14	0.57	3	0.5
e	1	0.25	0.1	0.4	0	0

Sol 3 c)



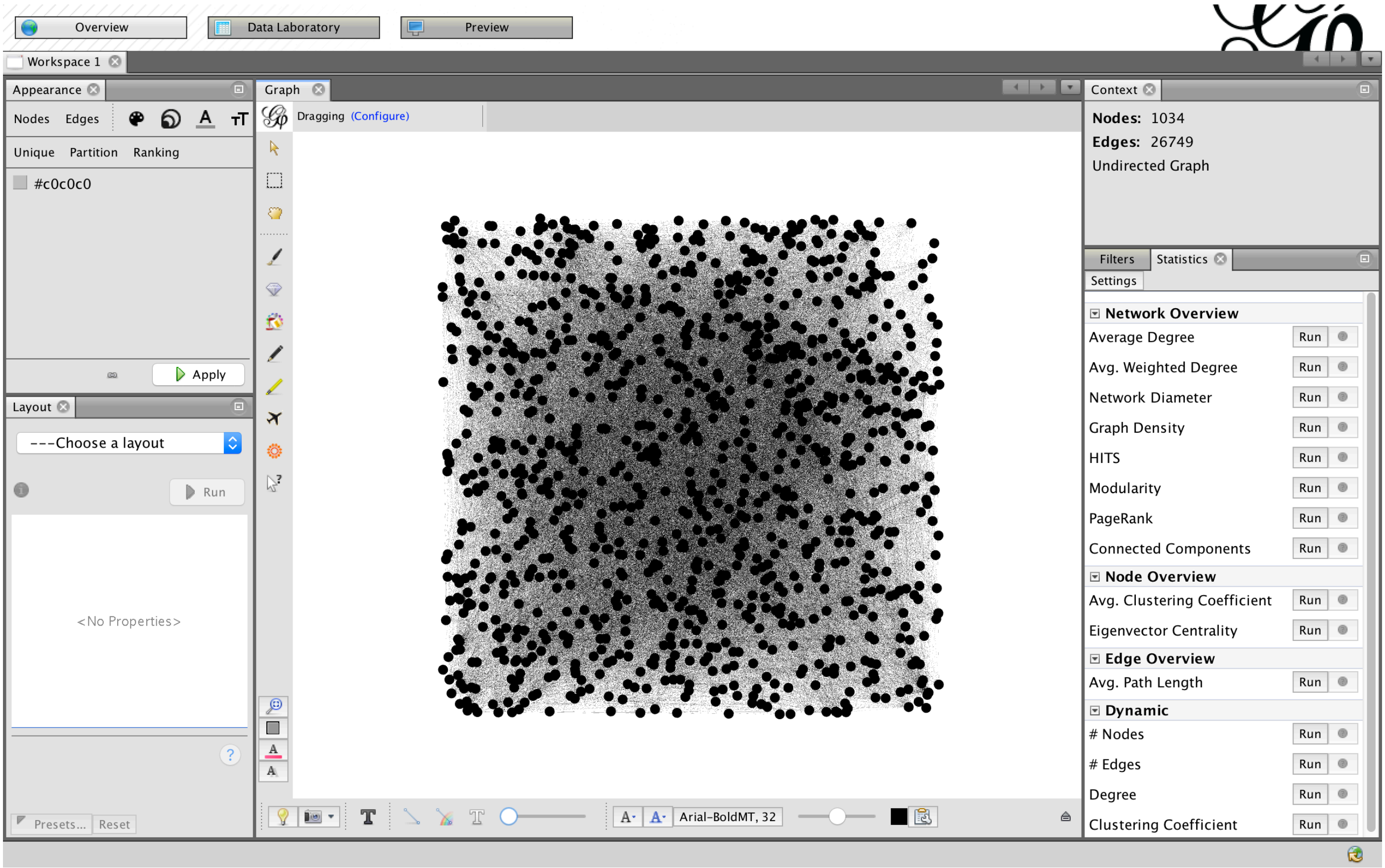
Node	Degree		Closeness		Between-ness	
	Actual	Normalized	Actual	Normalized	Actual	Normalized
a	2	0.5	0.167	0.67	1	0.166
b	2	0.5	0.167	0.67	1	0.166
c	2	0.5	0.167	0.67	1	0.166
d	2	0.5	0.167	0.67	1	0.166
e	2	0.5	0.167	0.67	1	0.166

Sol 4 d)



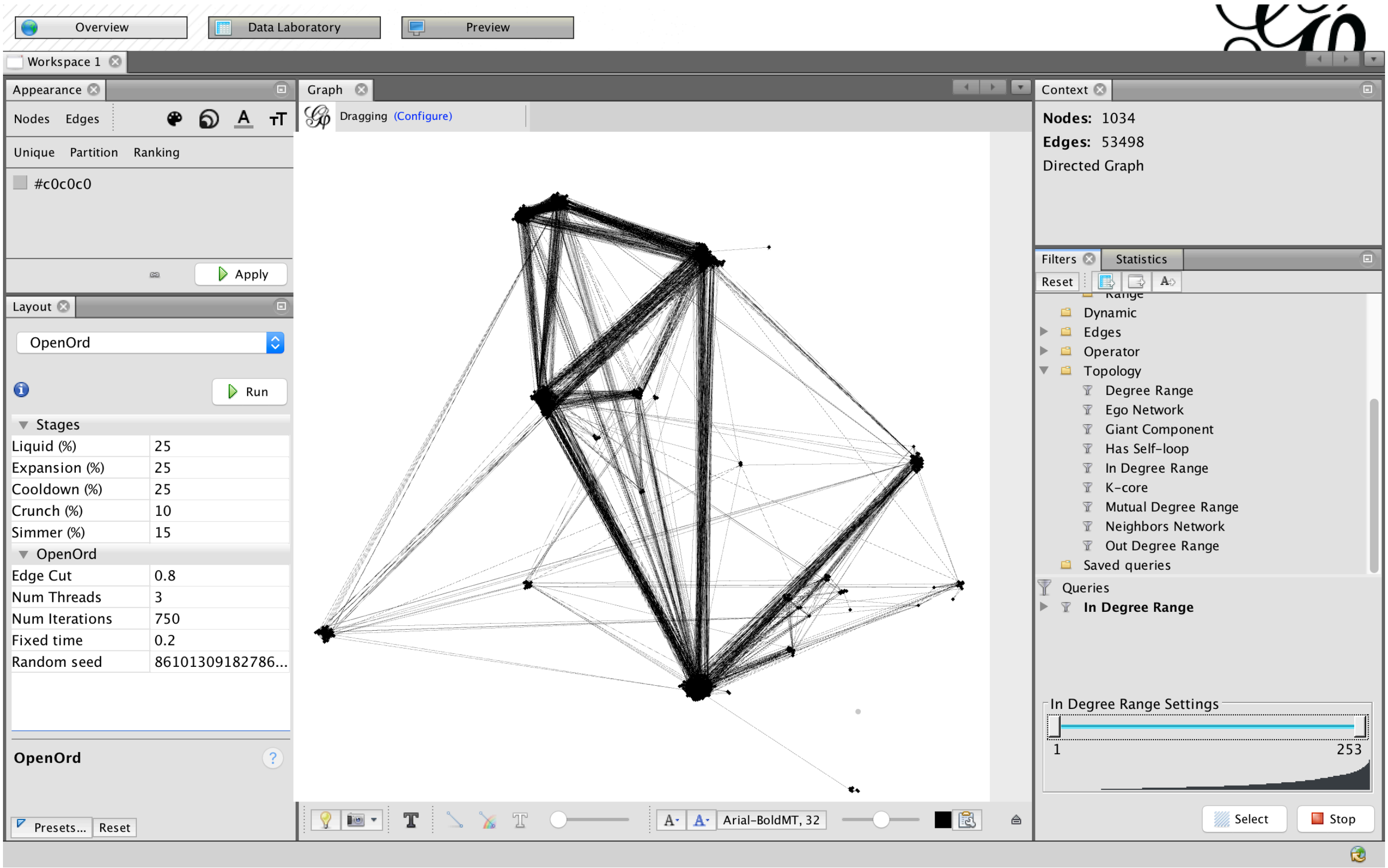
Node	Degree		Closeness		Between-ness	
	Actual	Normalized	Actual	Normalized	Actual	Normalized
A	1	0.25	0.11	0.44	0	0
B	2	0.50	0.167	0.67	3	0.5
C	3	0.75	0.2	0.8	4	0.66
D	1	0.25	0.125	0.5	0	0
E	1	0.25	0.125	0.5	0	0

Sol 2.
I used the Stanford Facebook dataset (specifically 107.edges file) and the Gephi tool for the analysis.

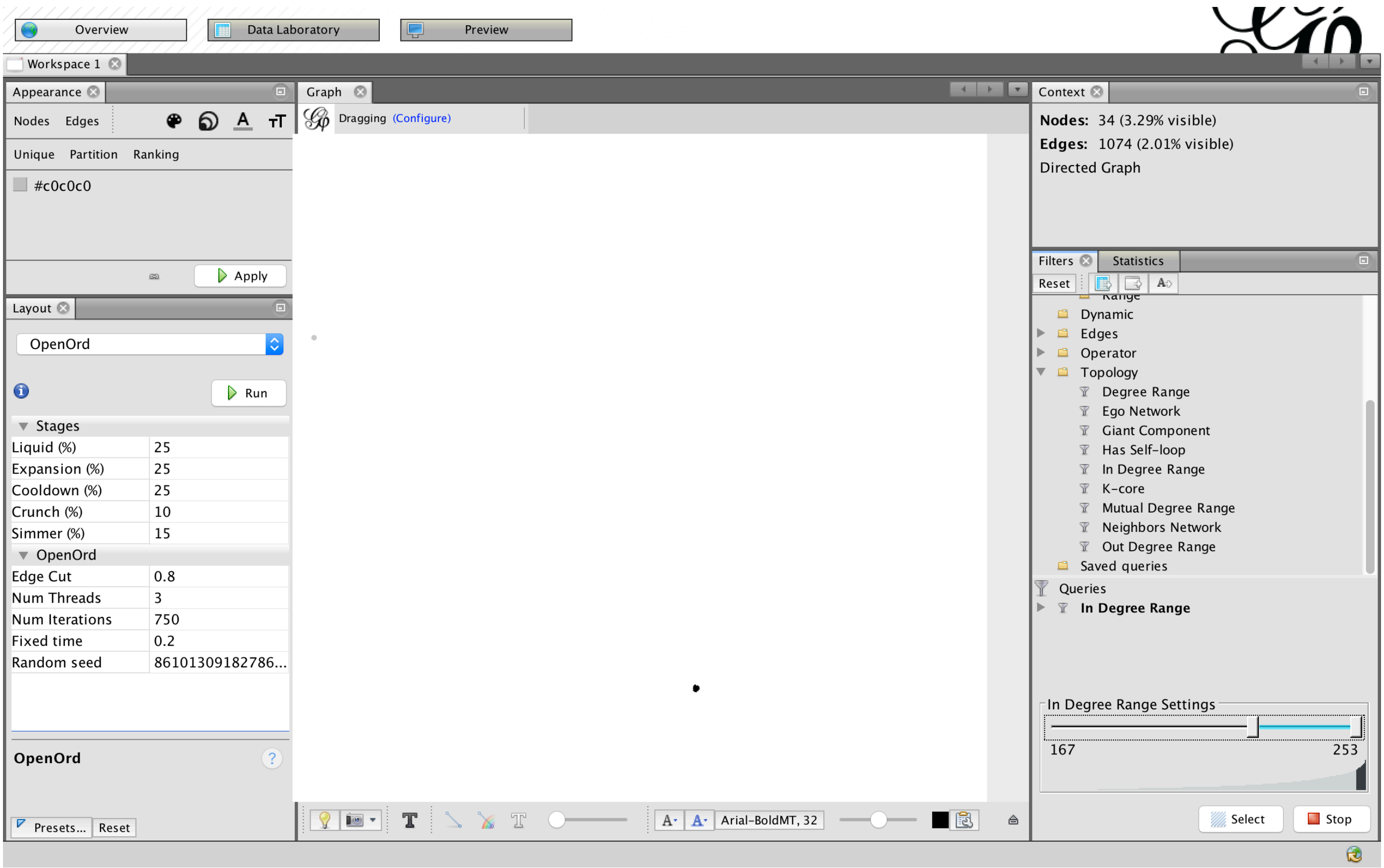


Graph Statistics:
# of Nodes: 1034
# of Edges: 53948
Avg Clustering Coefficient: 0.534
Average Degree: 51.739

Hypothesis 1: Popular people tend to form a cluster.
Using the Frutcherman-Reingold algorithm (through OpenOrd layout) clustering of the graph was done.

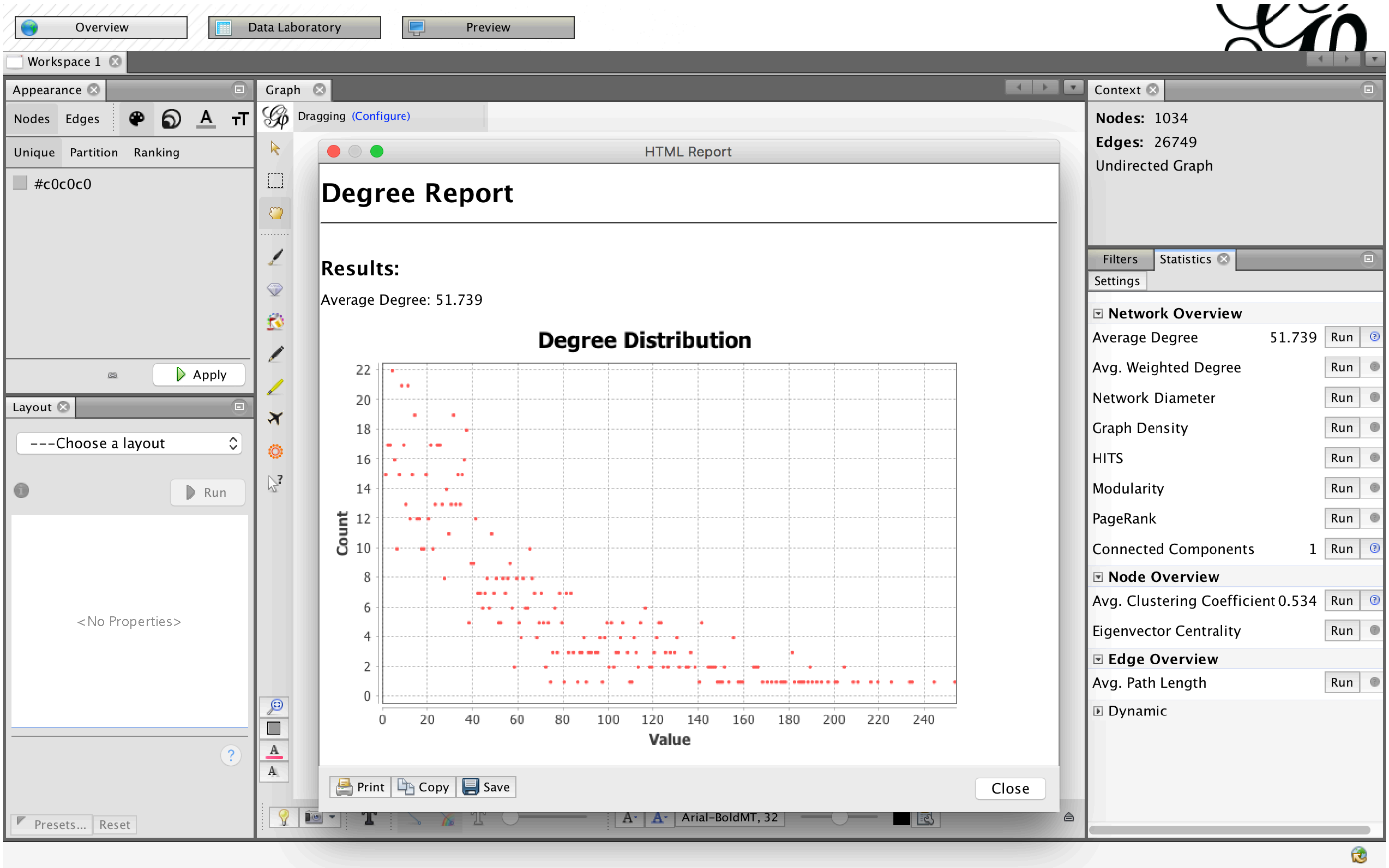


Then the graph was filtered based on degrees between 167 and 253(max).

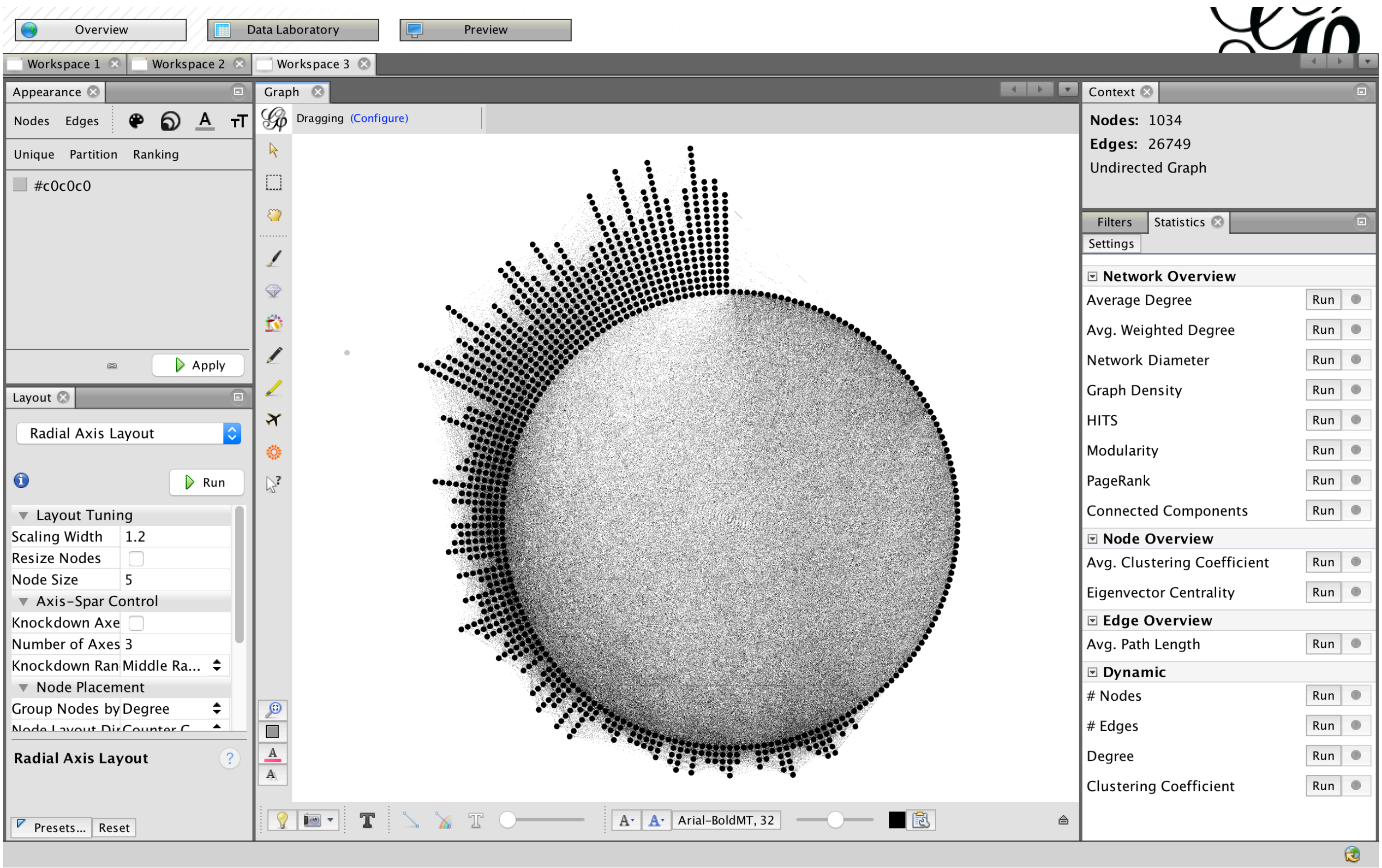


As we can see, the people with large no. of connections, (degree >167) belong to the same cluster. So, this hypothesis was correct.

Hypothesis 2: Degree vs No. of nodes is a decreasing curve.

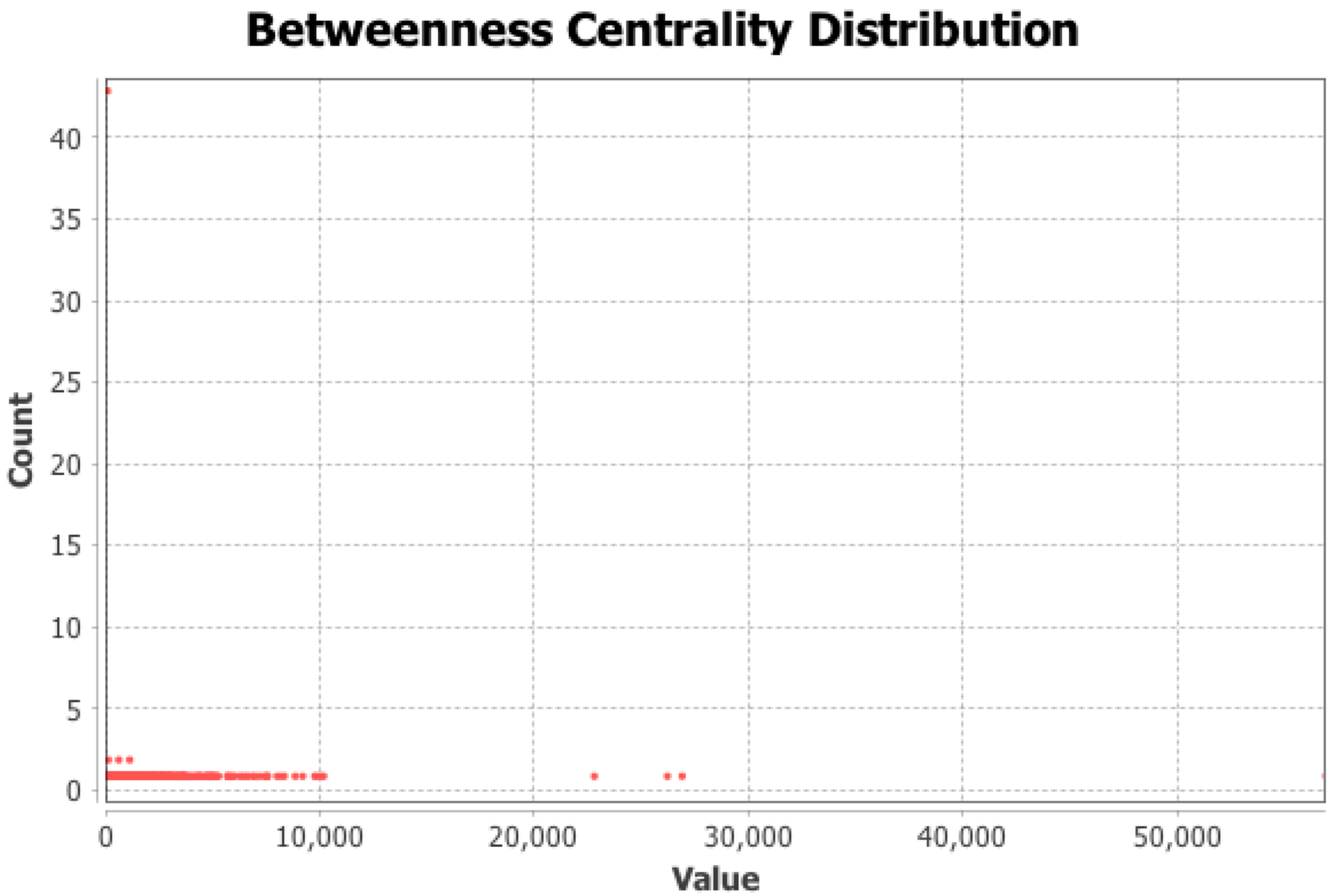


As we can see, this hypothesis holds. We also observe that degree 4 has the max count, with 22 people. This can also be visualized using Radial Axis Layout based on degrees of nodes.



Experiment 1: Now let's consider a scenario where we want to market a product. We want to identify the key users that we want to give our product for free so that they can share their review to their connections. Our product is quite costly, so we want to minimize the count of these users. Also, we consider that the information shared by these users will only reach their ego network, i.e. only connections with depth = 1 will be affected.

First, let's find the betweenness centrality of the nodes.

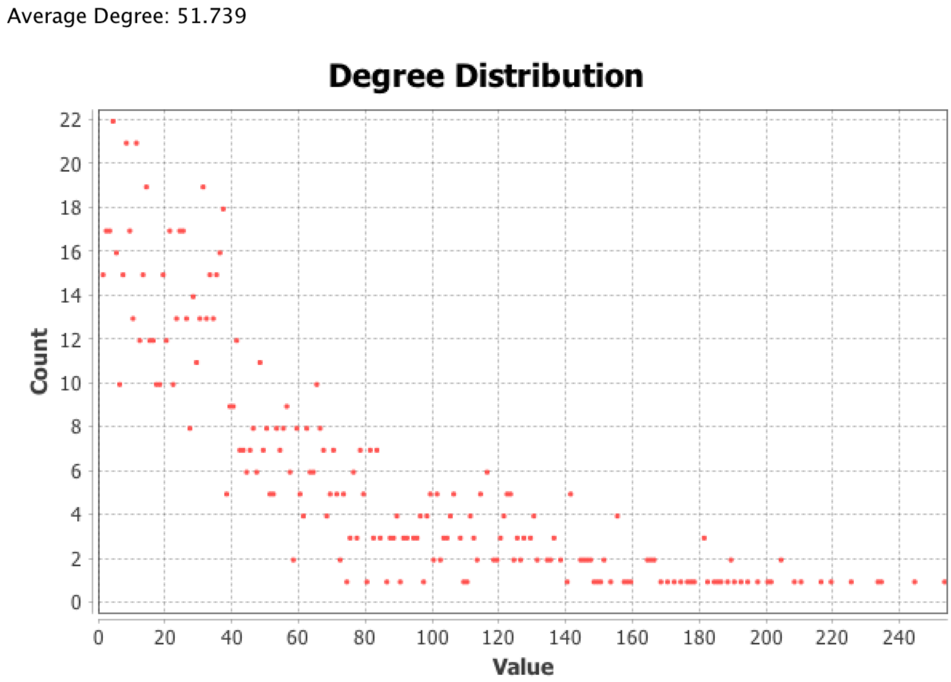


It can be observed that very few nodes (only 5) have betweenness centrality greater than 10000. These nodes are Node 917, 483, 1086, 1584, 1334. Creating a UNION of ego network of these 5 nodes (with depth 1) gives a network of 501 nodes. Thus, **these five people alone are directly connected to 50% of the network.** If we increase the depth level to 2, we can cover 925 nodes (almost 90%!!) of the network.

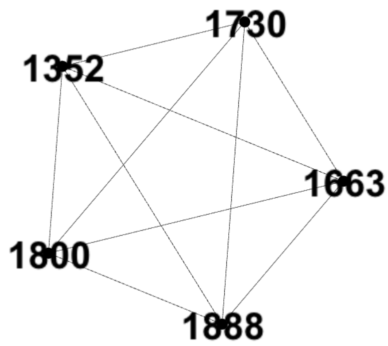
These were the individual contributions of each node for depth=1:
Node 917 = 130 nodes (betweenness=26093)
Node 483 = 167 nodes (betweenness=22692)
Node 1086 = 205 nodes (betweenness=56798)
Node 1334 = 100 (betweenness=10071)
Node 1584 = 211 (betweenness=26785)

Now, we can further try to find the NOT of these 501 nodes that we have reached, and see what insight we can have on the remaining nodes. But let's try to compare this result against the degree centrality.

The network has following degree distribution. As seen during Hypothesis 2 validation, there are very few nodes with large degree values.



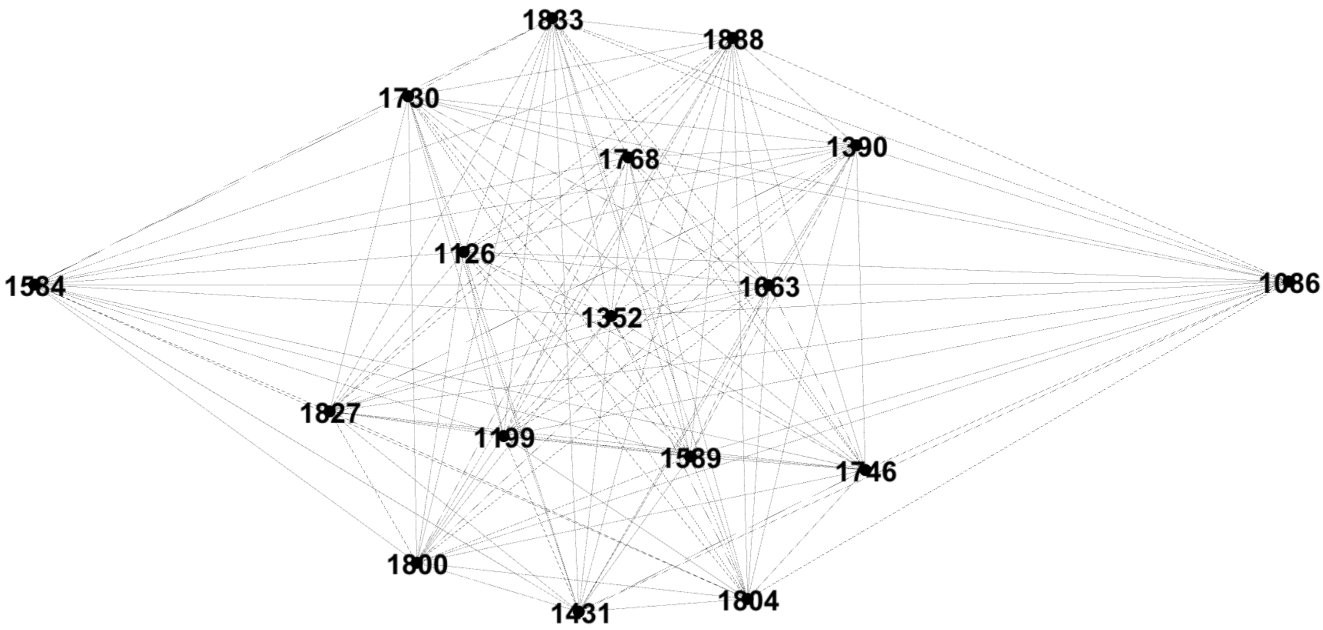
Let's consider top 5 nodes (with max degrees). We observe that these 5 nodes are fully connected.



UNION of ego network of these nodes gives 348 nodes, roughly 1/3rd of the total nodes. The **degree centrality actually turns out to be a poor metric choice for our problem statement.** Even after increasing the depth level to 2, we only reach 636 nodes, roughly 60% of the network (much worse than 90% achieved by betweeenness metric). One possible explanation for this could be related to Hypothesis 1. Since these nodes are actually part of a cluster, they share a lot of friends among each other.

These are the individual contributions of each node:
Node 1888 = 253 nodes (max degree)
Node 1800 = 245 nodes
Node 1663 = 235 nodes
Node 1352 = 234 nodes
Node 1730 = 226 nodes

Observation: One interesting observation during this experiment was also related to Hypothesis 1. The top nodes with large degrees not only form a cluster, they are actually fully connected. In fact, among the top 17 nodes there was only one edge missing (between node 1584 and 1086).



The fact that these two nodes are not connected can signify that they have opposite opinions or views and they might have a different kind of friends. So, one might assume that these two would be a good choice for our key users. But, that’s actually not the case. The UNION of their ego network gives only 286 nodes (~27%), which is not significantly higher than their individual contributions (211 nodes & 205 nodes respectively).