1. Zachary's Karate Club

Calculate relevant centralities for the nodes in the Karate Club network. Partition the network into clusters*. Visualize the network so that the important nodes are highlighted#. Do you find two partitions centered on the administrator (node 34) and the instructor (1)? Node 9 was the only one that Zachary's analysis failed to properly place. Does your analysis say that node 9 should stay with the administrator or the instructor? The original article notes that number 9 was close to receiving his/her blackbelt and suggests that this might have influenced his/her decision.

* See *kernighan_lin_bisection* or *girvan_newman*
\# For examples, see https://networkx.github.io/documentation/latest/auto_examples/index.html
You might try using *draw_kamada_kawai* or *spring_layout*

Figure 1 presents the Karate Club network. The size of the nodes corresponds to the degree of the node. Hence, the bigger the node the more important and influential the node is. As shown, nodes 34 (degree=17) and node 1 (degree=16) are shown to be more important than the rest. Node 33 (degree=12) comes in third place.

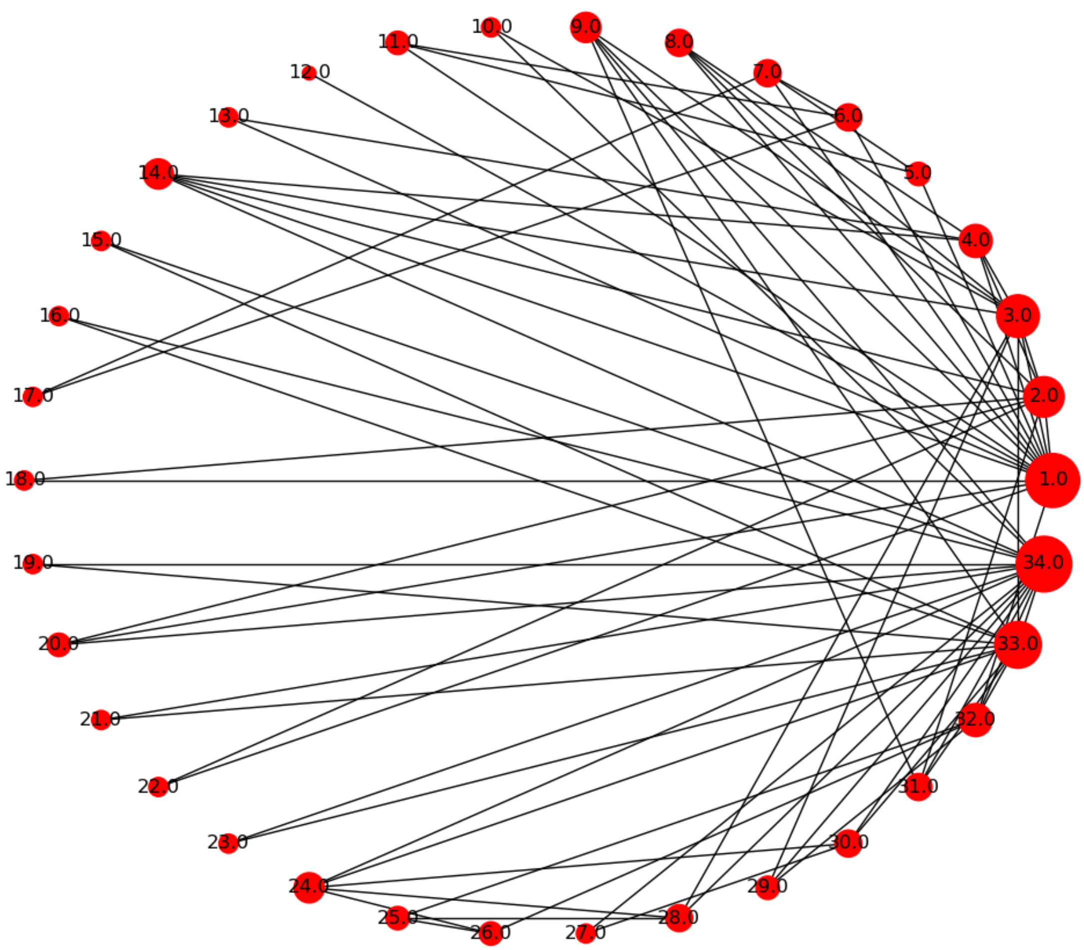


Figure 1. Graphical Representation of the Karate Club network.

Type: Graph
Number of edges: 78
Number of nodes: 34
Average degree: 4.5882
Network diameter = 5
Network density = 0.13903743315508021
Network average clustering coefficient = 0.5706384782076823
Network average shortest path length= 2.408199643493761

The following presents the relevant centralities for the nodes in the Karate Club network. Nodes are sorted by each centrality metric in descending order. The administrator (node 34) of the Karate Club is the highest in degree, eigenvector and page rank centrality metrics, second in betweenness and third in closeness centrality metrics. The instructor (node 1) ranks highest in betweenness and closeness centrality metrics, and comes second in degree, eigenvector and page rank centrality metrics. The administrator and the instructor are influential and important in the club in different ways, depending on the centrality metric being measured. See lists below:

```
Nodes sorted by degree centrality:
('34.0', 0.5151515151515151)
('1.0', 0.48484848484848486)
('33.0', 0.36363636363636365)
('3.0', 0.30303030303030304)
('2.0', 0.2727272727272727)
('4.0', 0.18181818181818182)
('32.0', 0.18181818181818182)
('9.0', 0.15151515151515152)
('14.0', 0.15151515151515152)
('24.0', 0.15151515151515152)
('6.0', 0.12121212121212122)
('7.0', 0.12121212121212122)
('8.0', 0.12121212121212122)
('28.0', 0.12121212121212122)
('30.0', 0.12121212121212122)
('31.0', 0.12121212121212122)
('5.0', 0.09090909090909091)
('11.0', 0.09090909090909091)
('20.0', 0.09090909090909091)
('25.0', 0.09090909090909091)
('26.0', 0.09090909090909091)
('29.0', 0.09090909090909091)
('10.0', 0.06060606060606061)
('13.0', 0.06060606060606061)
('15.0', 0.06060606060606061)
('16.0', 0.06060606060606061)
('17.0', 0.06060606060606061)
('18.0', 0.06060606060606061)
('19.0', 0.06060606060606061)
('21.0', 0.06060606060606061)
('22.0', 0.06060606060606061)
('23.0', 0.06060606060606061)
('27.0', 0.06060606060606061)
('12.0', 0.030303030303030304)

Nodes sorted by betweenness centrality:
('1.0', 0.43763528138528146)
('34.0', 0.30407497594997596)
('33.0', 0.145247113997114)
('3.0', 0.14365680615680618)
('32.0', 0.13827561327561325)
('9.0', 0.05592682780182781)
('2.0', 0.053936688311688304)
('14.0', 0.04586339586339586)
('20.0', 0.03247504810004811)
('6.0', 0.02998737373737374)
('7.0', 0.029987373737373736)
('28.0', 0.02233345358345358)
('24.0', 0.017613636363636363)
('31.0', 0.014411976911976909)
('4.0', 0.011909271284271283)
```

```
('26.0', 0.0038404882154882154)
('30.0', 0.0029220779220779218)
('25.0', 0.0022095959595959595)
('29.0', 0.0017947330447330447)
('10.0', 0.0008477633477633478)
('5.0', 0.0006313131313131313)
('11.0', 0.0006313131313131313)
('8.0', 0.0)
('12.0', 0.0)
('13.0', 0.0)
('15.0', 0.0)
('16.0', 0.0)
('17.0', 0.0)
('18.0', 0.0)
('19.0', 0.0)
('21.0', 0.0)
('22.0', 0.0)
('23.0', 0.0)
('27.0', 0.0)
```

Nodes sorted by closeness centrality:

```
('1.0', 0.5689655172413793)
('3.0', 0.559322033898305)
('34.0', 0.55)
('32.0', 0.5409836065573771)
('9.0', 0.515625)
('14.0', 0.515625)
('33.0', 0.515625)
('20.0', 0.5)
('2.0', 0.4852941176470588)
('4.0', 0.4647887323943662)
('28.0', 0.4583333333333333)
('31.0', 0.4583333333333333)
('29.0', 0.4520547945205479)
('8.0', 0.44)
('10.0', 0.4342105263157895)
('24.0', 0.39285714285714285)
('6.0', 0.38372093023255816)
('7.0', 0.38372093023255816)
('30.0', 0.38372093023255816)
('5.0', 0.3793103448275862)
('11.0', 0.3793103448275862)
('18.0', 0.375)
('22.0', 0.375)
('25.0', 0.375)
('26.0', 0.375)
('13.0', 0.3707865168539326)
('15.0', 0.3707865168539326)
('16.0', 0.3707865168539326)
('19.0', 0.3707865168539326)
('21.0', 0.3707865168539326)
('23.0', 0.3707865168539326)
('12.0', 0.36666666666666664)
('27.0', 0.3626373626373626)
('17.0', 0.28448275862068967)
```

Nodes sorted by eigenvector centrality:

```
('34.0', 0.373371213013235)
('1.0', 0.3554834941851943)
('3.0', 0.31718938996844476)
('33.0', 0.3086510477336959)
('2.0', 0.2659538704545025)
('9.0', 0.2274050914716605)
('14.0', 0.22646969838808148)
```

```
('4.0', 0.2111740783205706)
('32.0', 0.19103626979791702)
('31.0', 0.17476027834493085)
('8.0', 0.17095511498035434)
('24.0', 0.15012328691726787)
('20.0', 0.14791134007618667)
('30.0', 0.13496528673866567)
('28.0', 0.13347932684333308)
('29.0', 0.13107925627221215)
('10.0', 0.10267519030637758)
('15.0', 0.10140627846270832)
('16.0', 0.10140627846270832)
('19.0', 0.10140627846270832)
('21.0', 0.10140627846270832)
('23.0', 0.10140627846270832)
('18.0', 0.09239675666845953)
('22.0', 0.09239675666845953)
('13.0', 0.08425192086558088)
('6.0', 0.07948057788594247)
('7.0', 0.07948057788594247)
('5.0', 0.07596645881657382)
('11.0', 0.07596645881657381)
('27.0', 0.07558192219009324)
('26.0', 0.05920820250279008)
('25.0', 0.05705373563802805)
('12.0', 0.05285416945233648)
('17.0', 0.023634794260596875)
```

Nodes sorted by page rank:

```
('34.0', 0.1009179167487121)
('1.0', 0.09700181758983709)
('33.0', 0.07169213006588289)
('3.0', 0.057078423047636745)
('2.0', 0.05287839103742701)
('32.0', 0.03715663592267942)
('4.0', 0.03586064322306479)
('24.0', 0.03152091531163228)
('9.0', 0.029765339186167028)
('14.0', 0.029536314977202986)
('6.0', 0.02911334166344221)
('7.0', 0.02911334166344221)
('30.0', 0.02628726283711208)
('28.0', 0.025638803528350497)
('31.0', 0.02458933653429248)
('8.0', 0.024490758039509182)
('5.0', 0.021979406974834498)
('11.0', 0.021979406974834498)
('25.0', 0.021075455001162945)
('26.0', 0.021005628174745786)
('20.0', 0.019604416711937293)
('29.0', 0.01957296050943854)
('17.0', 0.016785378110253487)
('27.0', 0.015043395360629753)
('13.0', 0.014645186487916191)
('18.0', 0.014558859774243493)
('22.0', 0.014558859774243493)
('15.0', 0.014535161524273825)
('16.0', 0.014535161524273825)
('19.0', 0.014535161524273825)
('21.0', 0.014535161524273825)
('23.0', 0.014535161524273825)
('10.0', 0.014308950284462801)
('12.0', 0.009564916863537148)
```

The Karate Club network was partitioned into two clusters using two methods: the Kernighan_lin_bisection[1] and the *girvan_newman* algorithms.[2] Both algorithms resulted in two clusters with the administrator (node 34) and the instructor (node 1) as the center of each. In both cases, node 9 stays in the cluster with administrator (node 34).

Figures 2 and 3 below presents the clusters resulted from Kernighan_lin_bisection algorithm. The algorithm bisects the nodes into two disjoint subsets of the same size using the minimum edge cut approach. Ten edges were cut as a result. Figures 4 and 5 below presents the partitions resulted from Girvan Newman algorithm. With this algorithm, the cluster with the administrator (node 34) is slighter bigger than the one with the instructor (node 1). Girvan Newman algorithm removes one edge at a time in the descending order of betweenness centrality. Ten edges from the most betweenness were removed.

The original article shows that node 9 has a stronger (higher weight) relationship with node 34 (the administrator/John A.) than with node 1 (the instructor/Mr. Hi). See Figure 3 "Quantified Matrix of Relative Strengths of the Relationships in the Karate Club: The Matrix C" on page 462. My analysis using either Kernighan_lin_bisection or the *girvan_newman* algorithms on the provided, unweighted network data shows that node 9 stays with node 34. This agrees with run result from the *NETFLOW* presented in the original article (page 465). The NETFLOW results show that node 9 sides with the Sink (the administrator/node 34/John A.). However according to the original report, node 9 joined the instructor's (Mr. Hi's/node 1's) club after the Karate Club split. So, all three algorithms failed to correctly predict node 9's decision. The original article notes that node 9 was close to receiving his/her black belt and suggests that this might have influenced his/her decision. After karate club split, it makes sense that node 9 chose to join Mr. Hi's private karate studio so that he could continue his practice with the same instructor till he got his black belt. However, this decision factor was not part of the network data and hence would not have an effect in the result of the partition by any of the three algorithms.

---

[1] The input to the algorithm is an undirected graph $G = (V,E)$ with vertex set V, edge set E, and (optionally) numerical weights on the edges in E. The goal of the algorithm is to partition V into two disjoint subsets A and B of equal (or nearly equal) size, in a way that minimizes the sum T of the weights of the subset of edges that cross from A to B. If the graph is unweighted, then instead the goal is to minimize the number of crossing edges; this is equivalent to assigning weight one to each edge. (Source: Wikipedia.org; link: https://en.wikipedia.org/wiki/Kernighan%E2%80%93Lin_algorithm)

[2] The Girvan–Newman algorithm detects communities by progressively removing edges from the original network. The connected components of the remaining network are the communities. Instead of trying to construct a measure that tells us which edges are the most central to communities, the Girvan–Newman algorithm focuses on edges that are most likely "between" communities. (Source: Wikipedia.org; link: https://en.wikipedia.org/wiki/Girvan%E2%80%93Newman_algorithm)

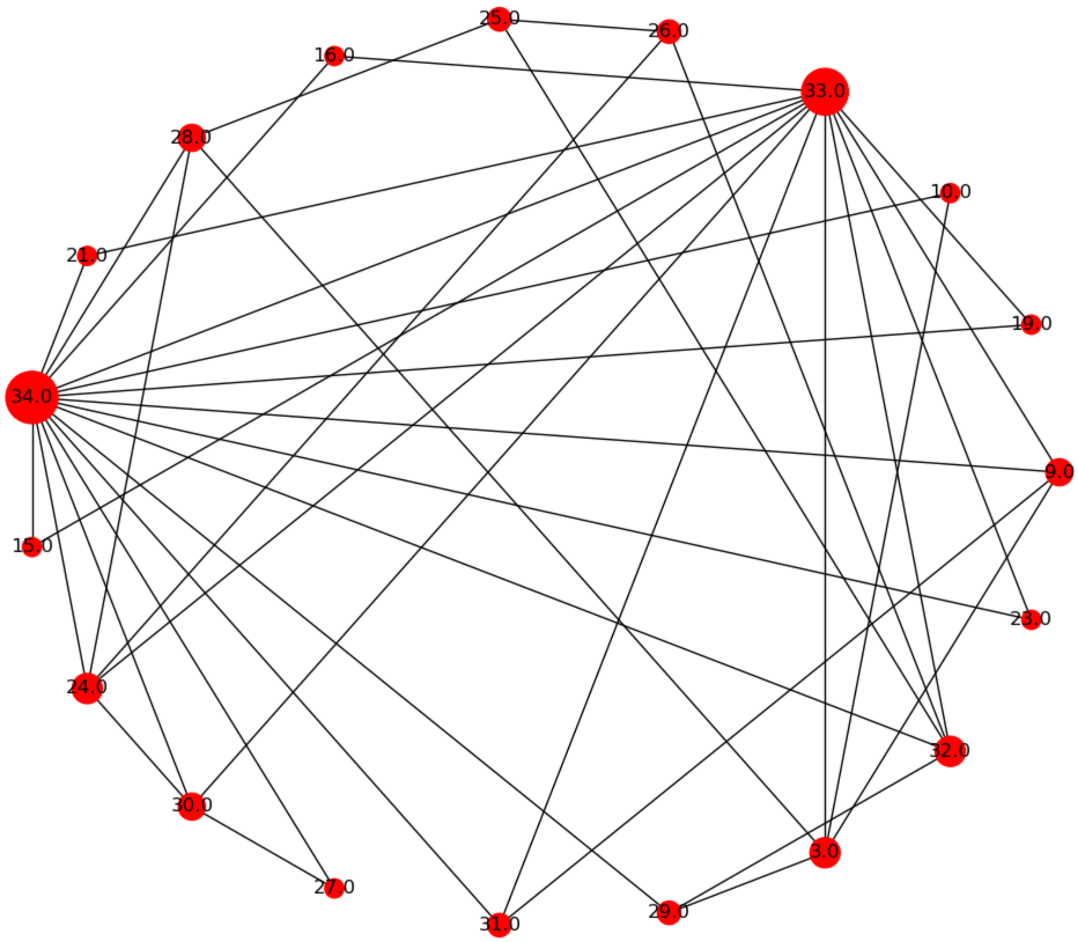


Figure 2. Bisected Cluster Containing Node 34, the Administrator and Node 9.

Using Kernighan_lin_bisection algorithm:
Set 1 = {'24.0', '15.0', '21.0', '30.0', '29.0', **'34.0'**, '33.0', '32.0', '31.0', '25.0', **'9.0'**, '27.0', '16.0', '26.0', '19.0', '23.0', '28.0'}
Type: SubGraph
Number of nodes: 17
Number of edges: 34
Average degree: 4.0000
Network diameter = 3
Network density = 0.25
Network average clustering coefficient = 0.6911578617460971
Network average shortest path length= 1.875

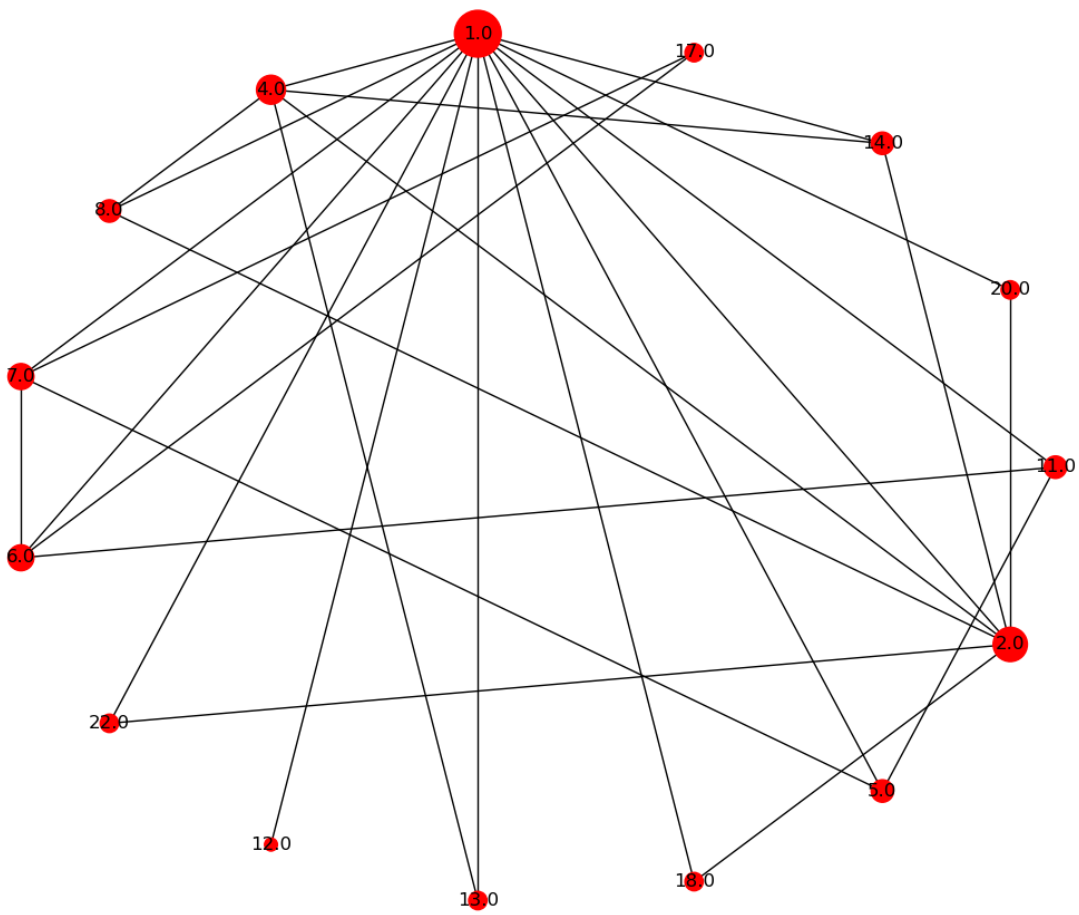


Figure 3. Bisected Cluster Containing Node 1, the Instructor.

Using Kernighan_lin_bisection algorithm:
Set 2 = {'13.0', '5.0', '7.0', '8.0', '11.0', '3.0', '17.0', '22.0', '18.0', '10.0', '2.0', '12.0', '20.0', '6.0', '4.0', '14.0', **'1.0'**}
Type: SubGraph
Number of nodes: 17
Number of edges: 34
Average degree: 4.0000
Network diameter = 4
Network density = 0.25
Network average clustering coefficient = 0.6597285067873303
Network average shortest path length= 1.9044117647058822

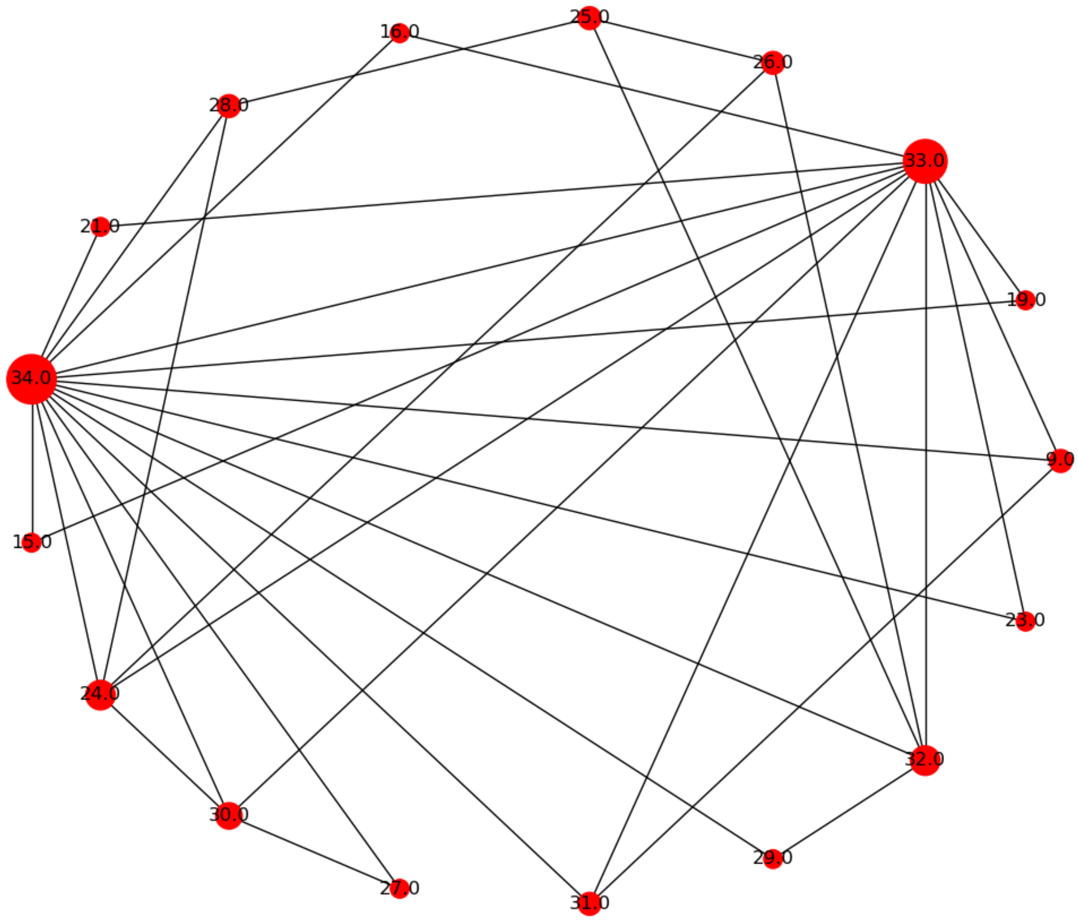


Figure 4. Girvan Newman Algorithm:
Cluster Containing Node 34, the Administrator and Node 9

Using girvan_newman algorithm:
Community 2 = {'24.0', '21.0', '32.0', '15.0', '31.0', '30.0', '10.0', '25.0', **'9.0'**, '27.0', '29.0', **'34.0'**, '26.0', '16.0', '3.0', '33.0', '19.0', '23.0', '28.0'}
Type: SubGraph
Number of nodes: 19
Number of edges: 40
Average degree: 4.2105
Network diameter = 3
Network density = 0.23391812865497075
Network average clustering coefficient = 0.5599908863066758
Network average shortest path length= 1.8888888888888888

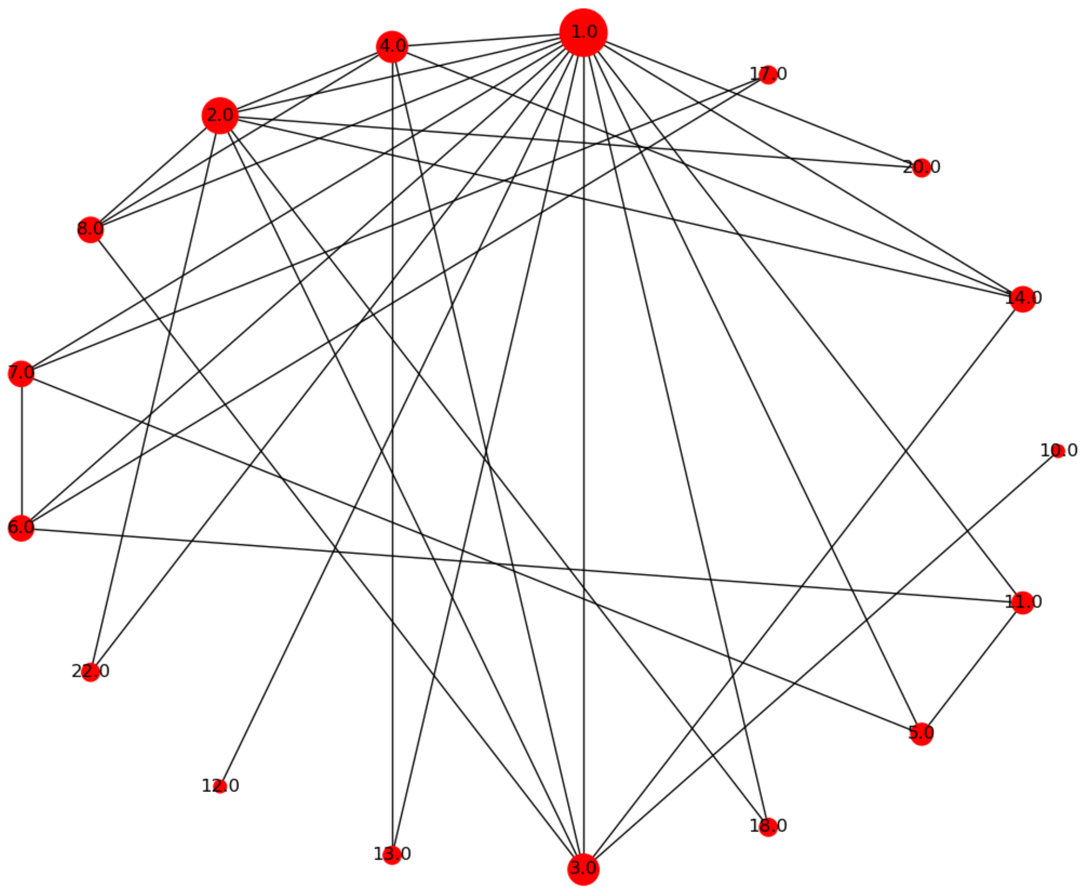


Figure 5. Girvan Newman Algorithm: Cluster Containing Node 1, the Instructor.

Using girvan_newman algorithm:
Community 1 = {'13.0', '22.0', '5.0', '18.0', '7.0', '2.0', '20.0', '12.0', '8.0', '6.0', '4.0', '11.0', '14.0', **'1.0'**, '17.0'}
Type: SubGraph
Number of nodes: 15
Number of edges: 28
Average degree: 3.7333
Network diameter = 3
Network density = 0.26666666666666666
Network average clustering coefficient = 0.6987301587301586
Network average shortest path length= 1.819047619047619

2. Trump's Personal & Business Relationships

Identify the 10 most influential nodes in the network using i) degree, ii) page rank, and iii) betweenness centrality. Visualize the network.

You can extract the nodes and links from the original data file or simply import the graphML file.

You can read more information about the network here: https://www.buzzfeed.com/johntemplon/help-us-map-trumpworld?utm_term=.vaJO0mwvX0#.ifDlvxpM1v

Figure 1a presents the graph that shows Trump and 1513 entities that he has direct or indirect personal and business relationship. Figures 1b through 1d show progressively zoomed in portion of the complete graph. This network has a low average clustering coefficient of 0.236, and low average shortest path length of 3.800. Most of the entities do not have a relationship among themselves. This can be clearly seen in the outer rings of the graph that the nodes are not connected.

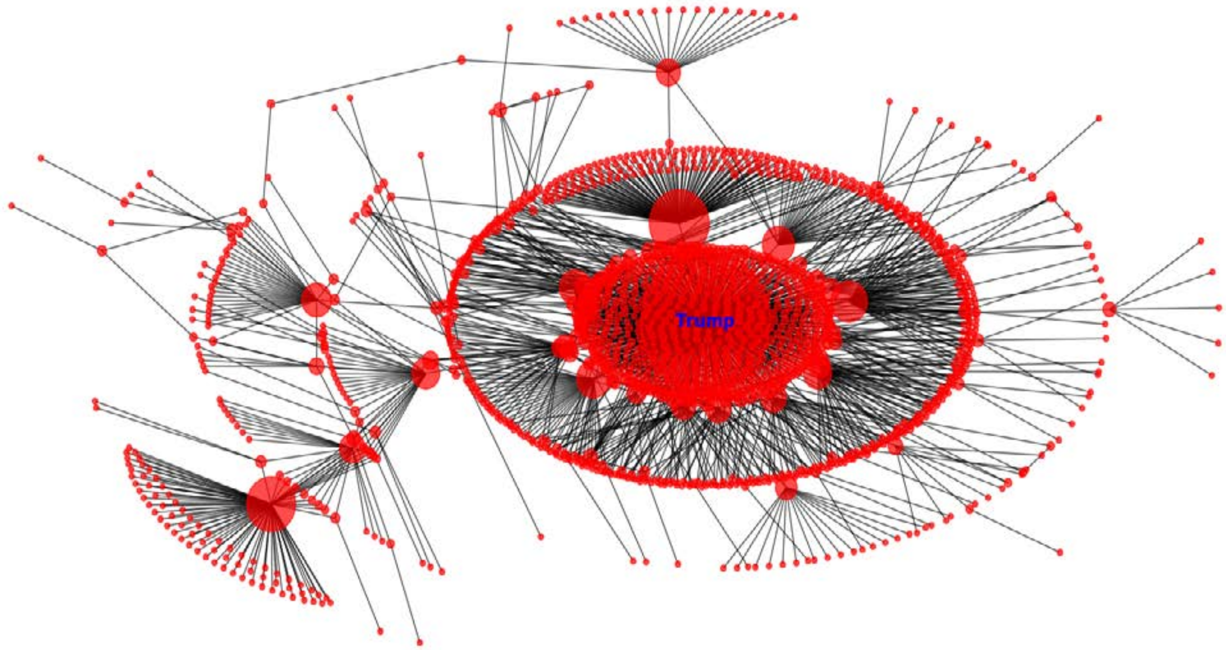


Figure 1a. Graph Showing Trump's Personal and Business Relationships

```
Graph name: Trump World
Type: Graph
Number of nodes: 1514
Number of edges: 1857
Average degree:   2.4531
Average clustering coefficient = 0.23631932409465756
Average shortest path length = 3.7999862049817477
```

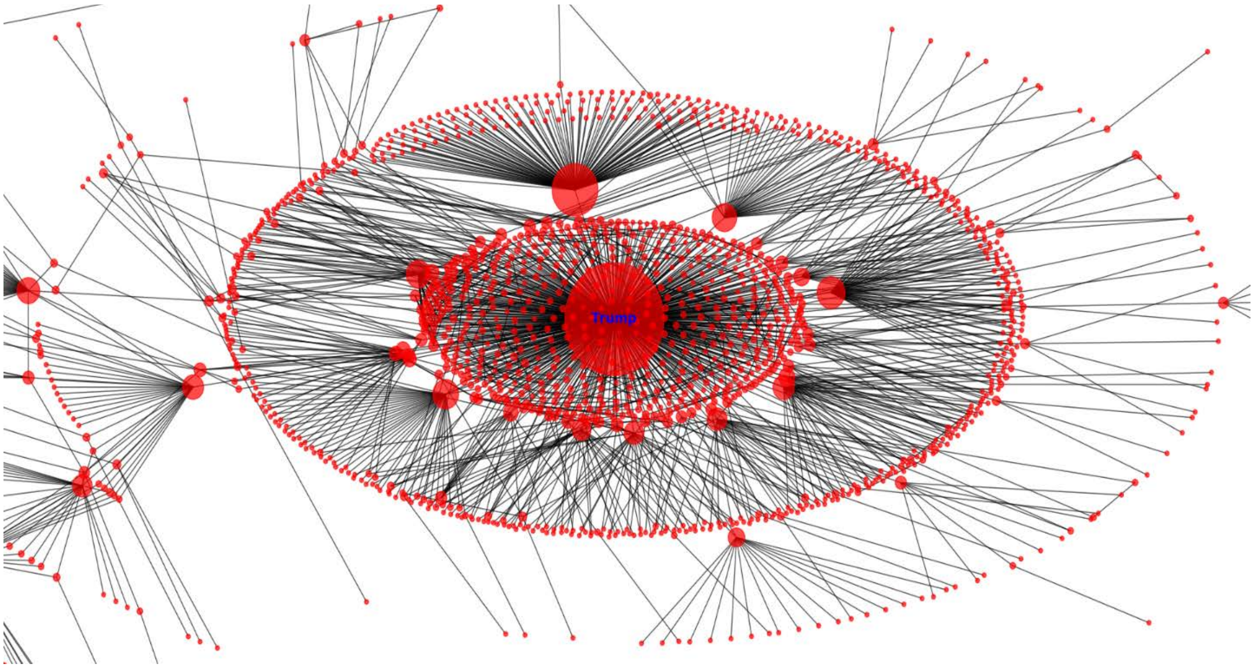


Figure 1b. Graph Showing Trump's Personal and Business Relationships – Zoomed In 3rd Ring

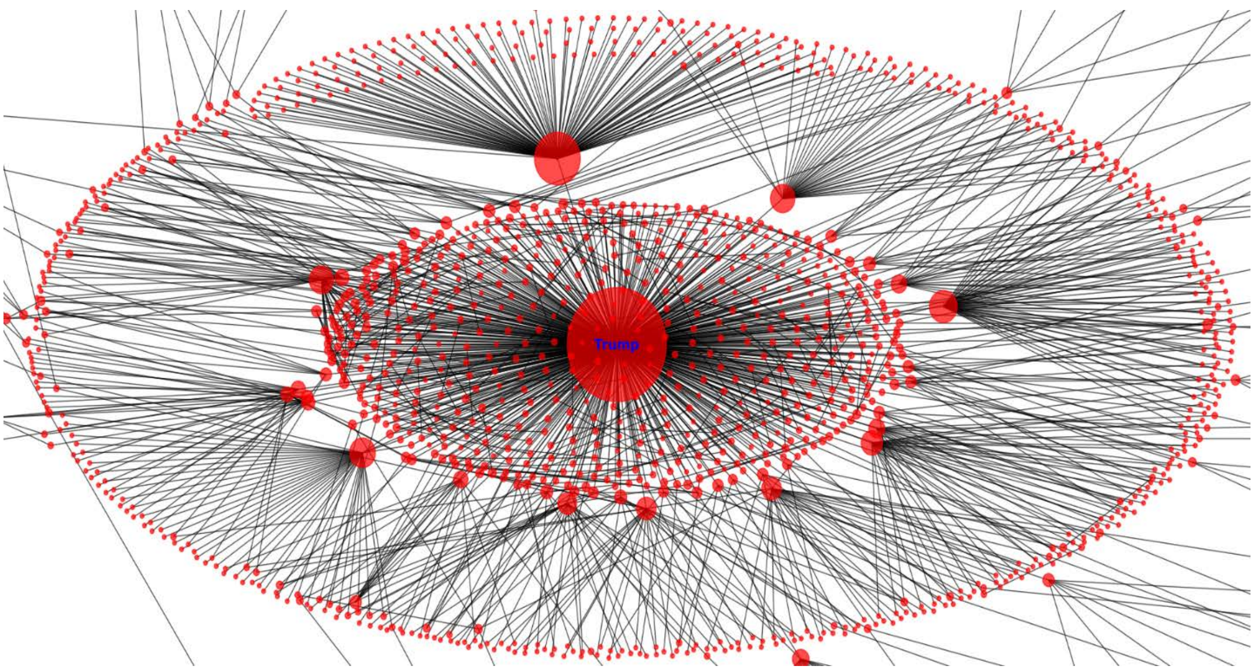


Figure 1c. Graph Showing Trump's Personal and Business Relationships – Zoomed In 2nd Ring

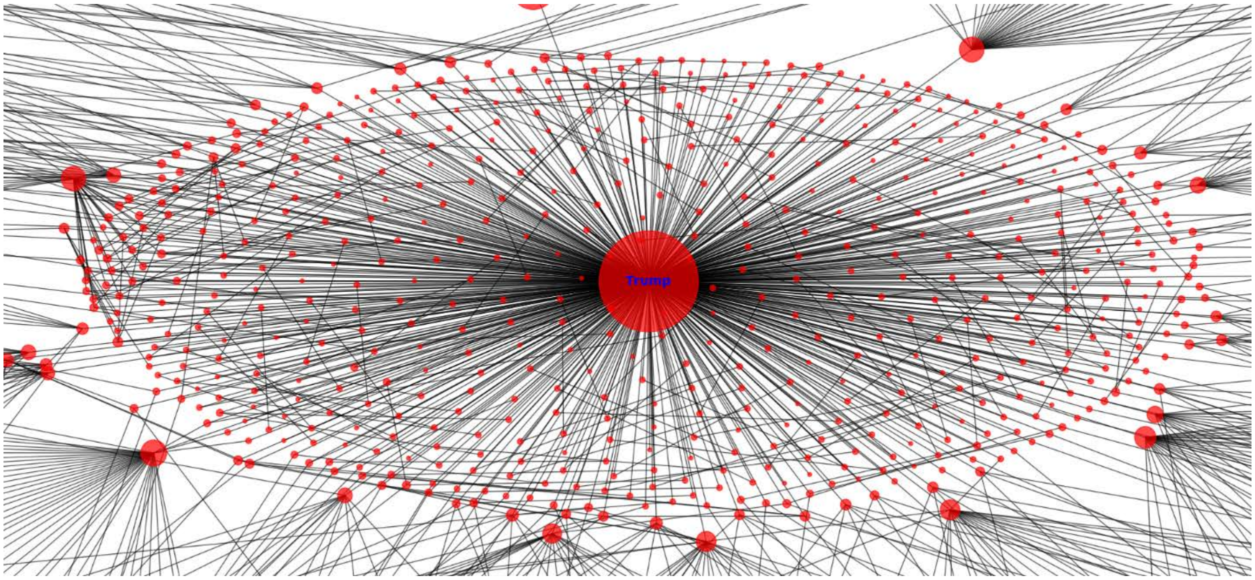


Figure 1d. Graph Showing Trump's Personal and Business Relationships – Zoomed In the Center

Figure 2 shows Trump and 605 entities who have direct personal and business relationship with Trump. The average clustering coefficient of this subgraph is 0.595 while the average shortest path length is 1.995, which shows that the sub-network is a high clustered small world. As shown in the Figure 2, most of the entities are interconnected in some degrees.

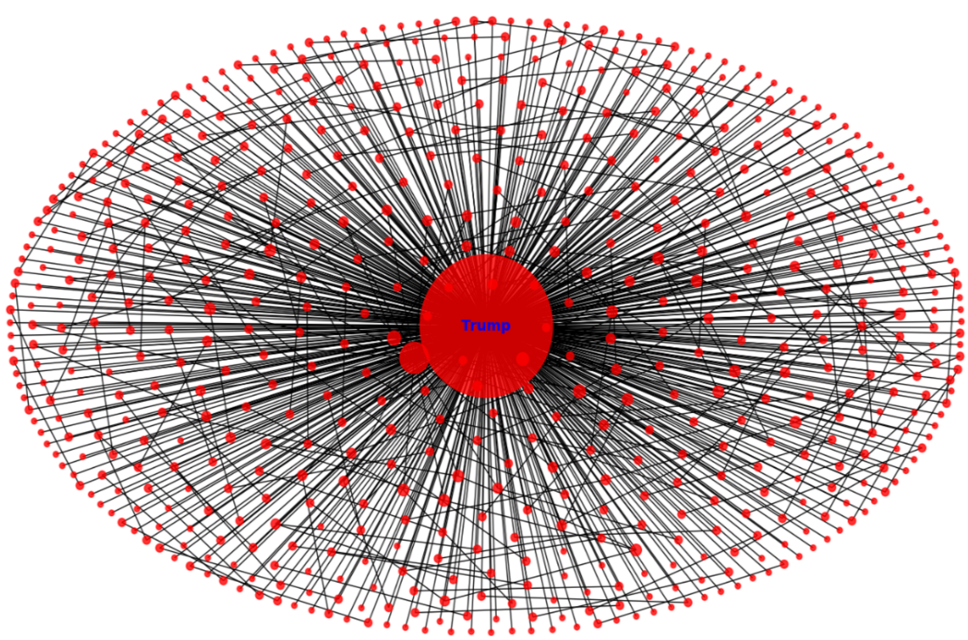


Figure 2. Subgraph: Trump and 605 Entities with Direct Personal and Business Relationships

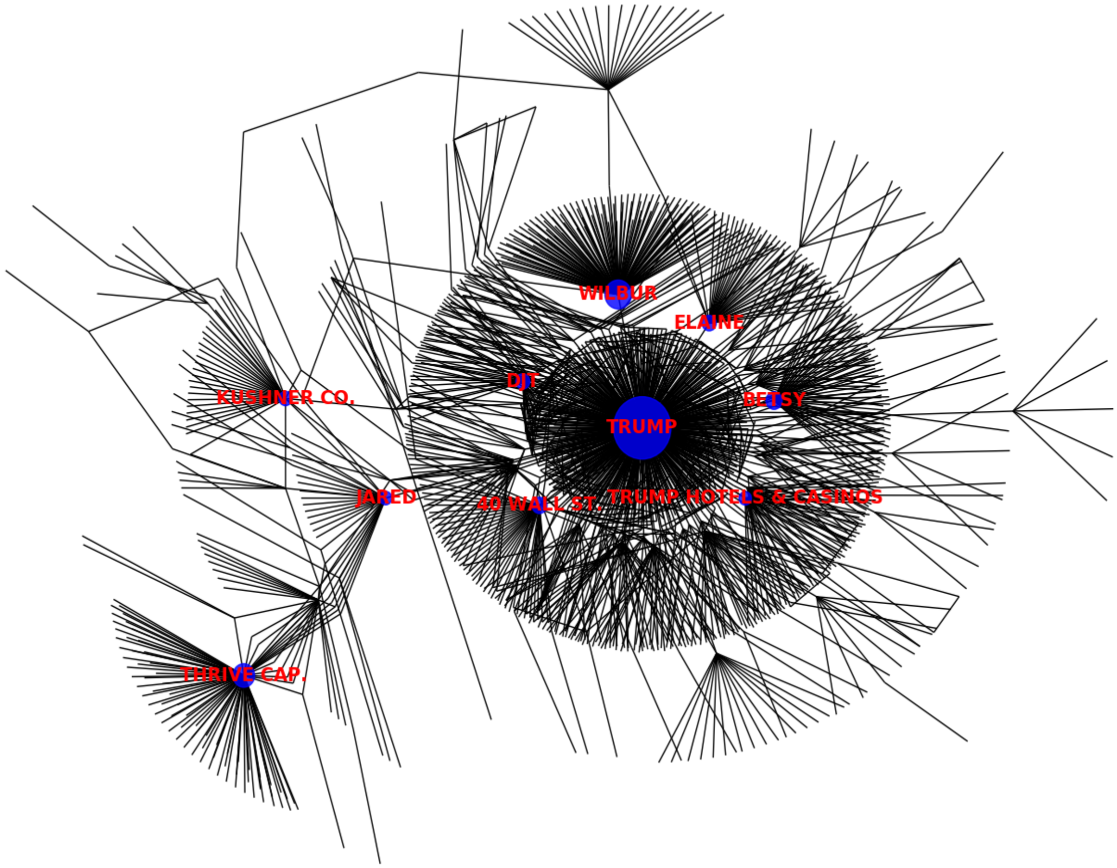
```
Type: SubGraph
Number of nodes: 605      Number of edges: 847      Average degree:   2.8000
Average clustering coefficient = 0.5945555310618618
Average shortest path length = 1.995364238410596
```

From the original network, top 10 most influential entities are identified in terms of degree centrality, page rank and betweeness centrality (see Table 1). Figures 3, 5 and 7 show the positions of these top 10 entities in the network. Note that while these top 10 entities are influential in the network in these centrality, most of them are not directly related to others (Figures 4, 6 and 8). Trump is the only entity that has a high degree of 5 or 6 in each of the subgraph. The same 10 entities rank top 10 in both degree centrality and page rank, but in different orders. Ivanka Trump and Goldman Sachs (underlined below) made it in the list for top 10 betweenness in place of Trump Hotels and Casino Resorts and DJT Holdings, LLC.

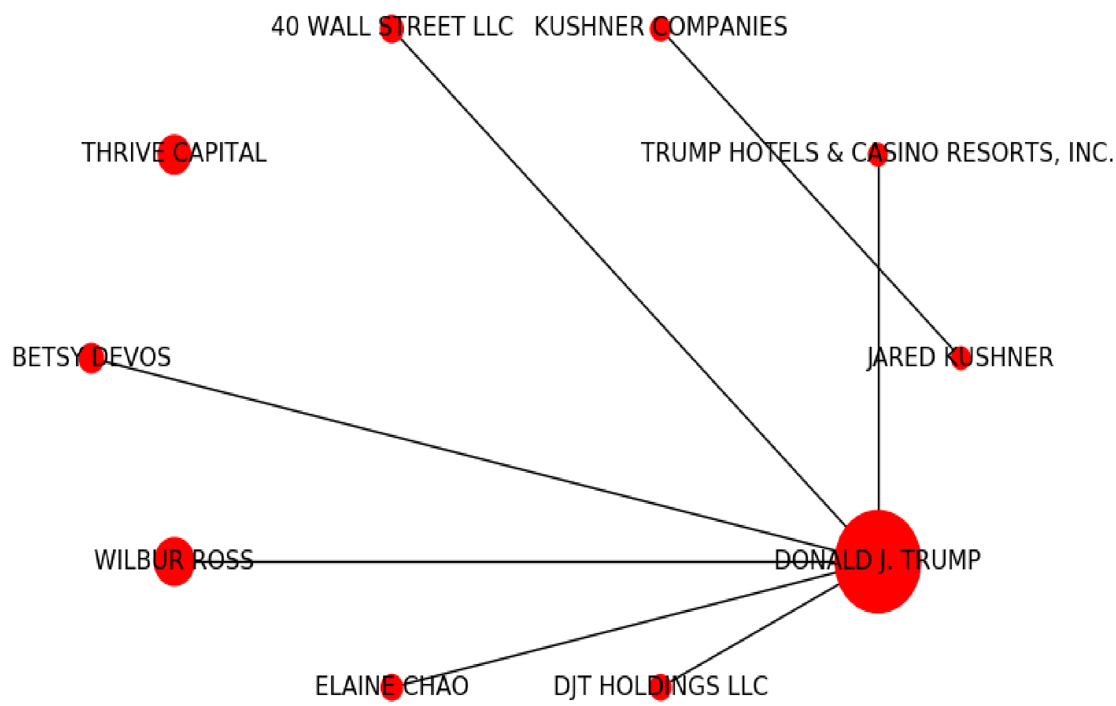
Table 1. Top 10 most Influential Entities in the Trump World

Top 10 by degree centrality:	**Top 10 by page rank:**	**Top 10 by betweenness centrality:**
DONALD J. TRUMP, 0.3992	DONALD J. TRUMP, 0.1301	DONALD J. TRUMP, 0.9452
WILBUR ROSS, 0.0853	WILBUR ROSS, 0.0385	WILBUR ROSS, 0.1733
THRIVE CAPITAL, 0.0568	THRIVE CAPITAL, 0.0232	<u>IVANKA TRUMP, 0.1488</u>
BETSY DEVOS, 0.0317	BETSY DEVOS, 0.0140	JARED KUSHNER, 0.1456
40 WALL STREET LLC, 0.0271	ELAINE CHAO, 0.0107	THRIVE CAPITAL, 0.0908
ELAINE CHAO, 0.0245	40 WALL STREET LLC, 0.0105	BETSY DEVOS, 0.0650
DJT HOLDINGS LLC, 0.0238	KUSHNER COMPANIES, 0.0085	ELAINE CHAO, 0.0566
KUSHNER COMPANIES, 0.0212	TRUMP HOTELS & CASINO RESORTS, INC., 0.0082	<u>GOLDMAN SACHS, 0.0516</u>
JARED KUSHNER, 0.0185	JARED KUSHNER, 0.0069	40 WALL STREET LLC, 0.0471
TRUMP HOTELS & CASINO RESORTS, INC., 0.0185	DJT HOLDINGS LLC, 0.0065	KUSHNER COMPANIES, 0.04687

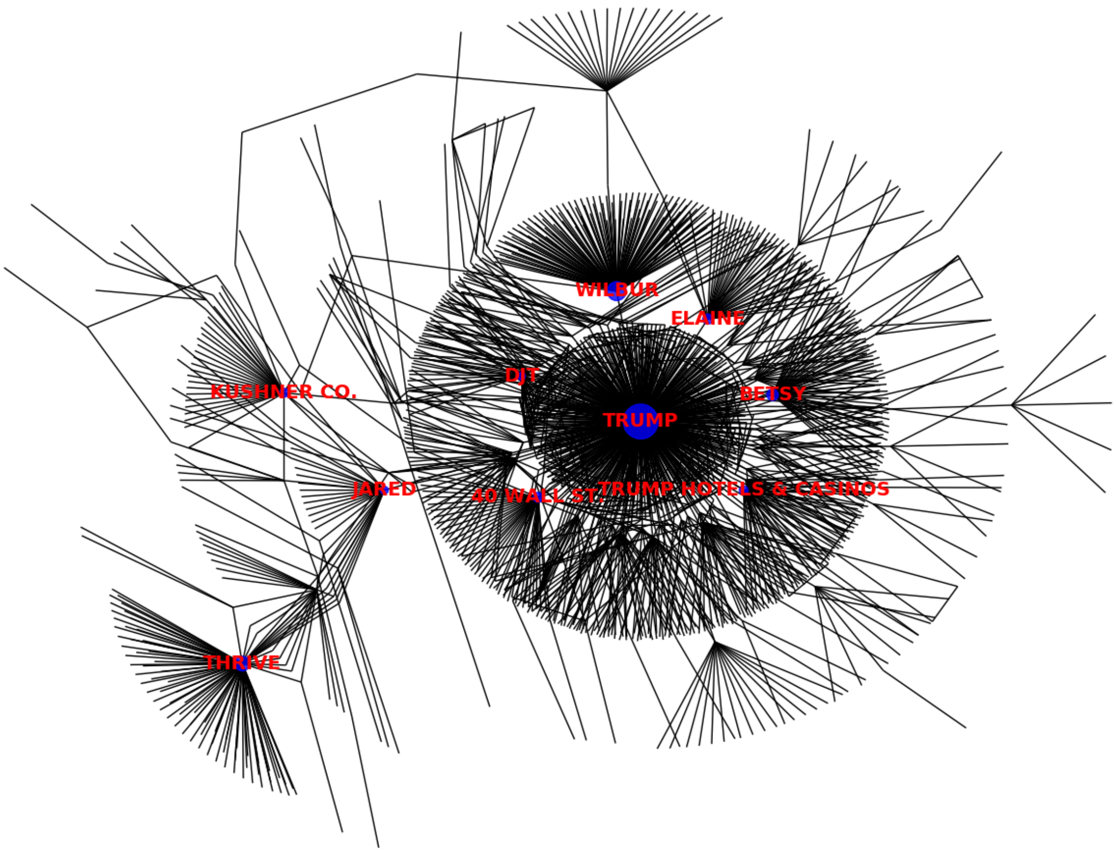
Figures 3. Original Graph with Top 10 Influential Nodes by Degree Centrality




```
Nodes sorted by degree centrality:
('DONALD J. TRUMP', 0.3992068737607402)
('WILBUR ROSS', 0.085261070720423)
('THRIVE CAPITAL', 0.05684071381361533)
('BETSY DEVOS', 0.03172504957038995)
('40 WALL STREET LLC', 0.027098479841137475)
('ELAINE CHAO', 0.024454725710508923)
('DJT HOLDINGS LLC', 0.023793787177792465)
('KUSHNER COMPANIES', 0.021150033046926635)
('JARED KUSHNER', 0.018506278916060805)
('TRUMP HOTELS & CASINO RESORTS, INC.', 0.018506278916060805)
```



Figures 4. Subgraph: Top 10 Influential Nodes by Degree Centrality



Figures 5. Original Graph with Top 10 Influential Nodes by Page Rank

```
Nodes sorted by page rank:
('DONALD J. TRUMP', 0.13013174597007712)
('WILBUR ROSS', 0.03847738345722221)
('THRIVE CAPITAL', 0.023248572406997688)
('BETSY DEVOS', 0.013962852313231746)
('ELAINE CHAO', 0.010702630459283678)
('40 WALL STREET LLC', 0.010487462008664914)
('KUSHNER COMPANIES', 0.008484919781768241)
('TRUMP HOTELS & CASINO RESORTS, INC.',
0.008166537779596172)
('JARED KUSHNER', 0.006884235829750 88)
('DJT HOLDINGS LLC', 0.006530631471853009)
```

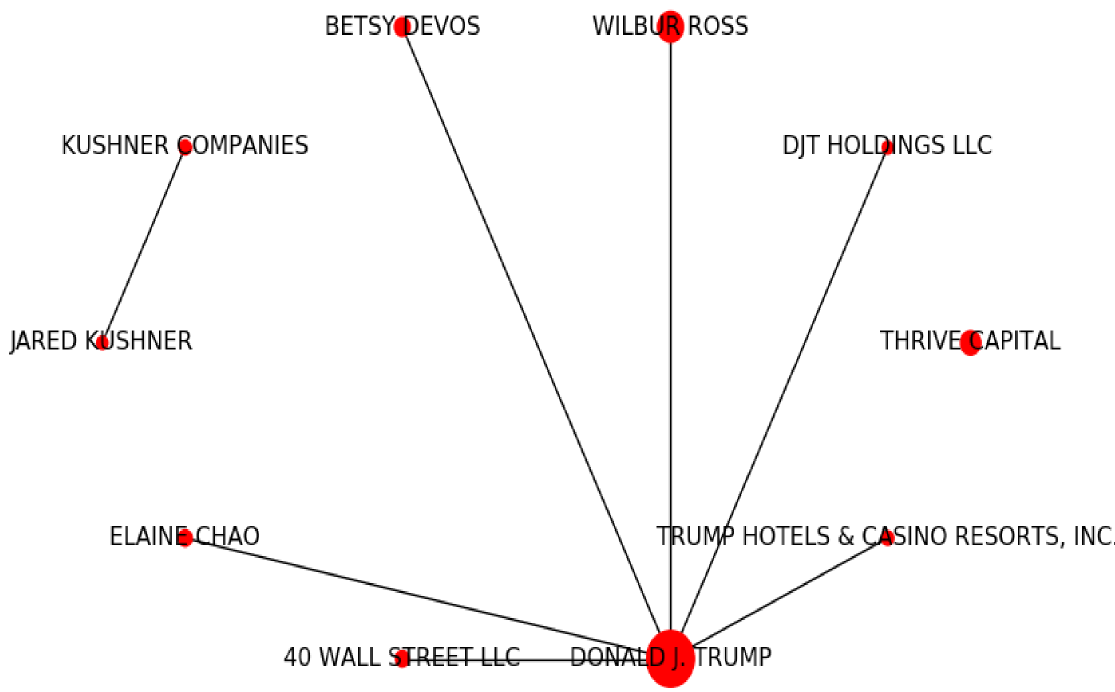
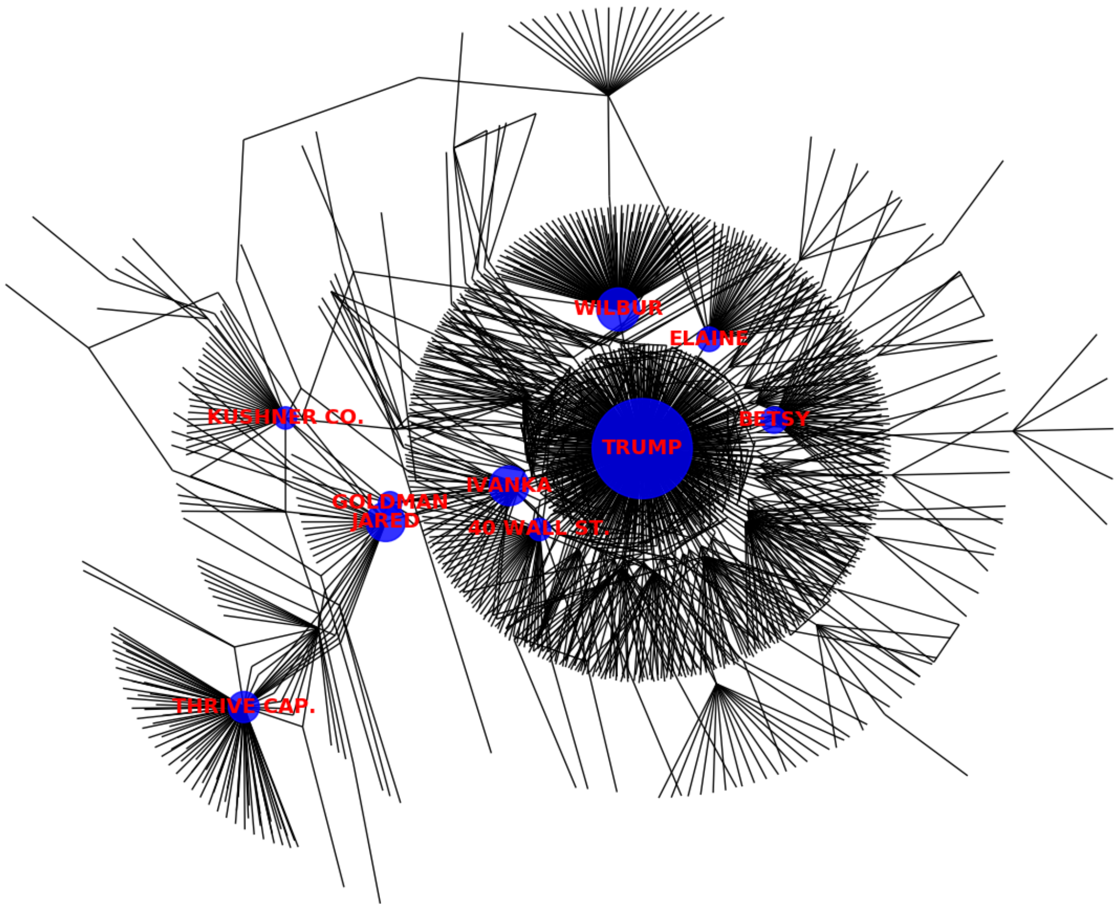


Figure 6. Subgraph: Top 10 Influential Nodes by Page Rank



Figures 7. Original Graph with Top 10 Influential Nodes by Betweenness Centrality

```
Nodes sorted by betweenness centrality:
('DONALD J. TRUMP', 0.9452682400221283)
('WILBUR ROSS', 0.17326048223771057)
('IVANKA TRUMP', 0.14881006807417613)
('JARED KUSHNER', 0.14557025982876173)
('THRIVE CAPITAL', 0.09080055451227514)
('BETSY DEVOS', 0.06501443690251797)
('ELAINE CHAO', 0.05655362364570866)
('GOLDMAN SACHS', 0.05162062543111188)
('40 WALL STREET LLC', 0.04710498535660623)
('KUSHNER COMPANIES', 0.046866141283716606)
```

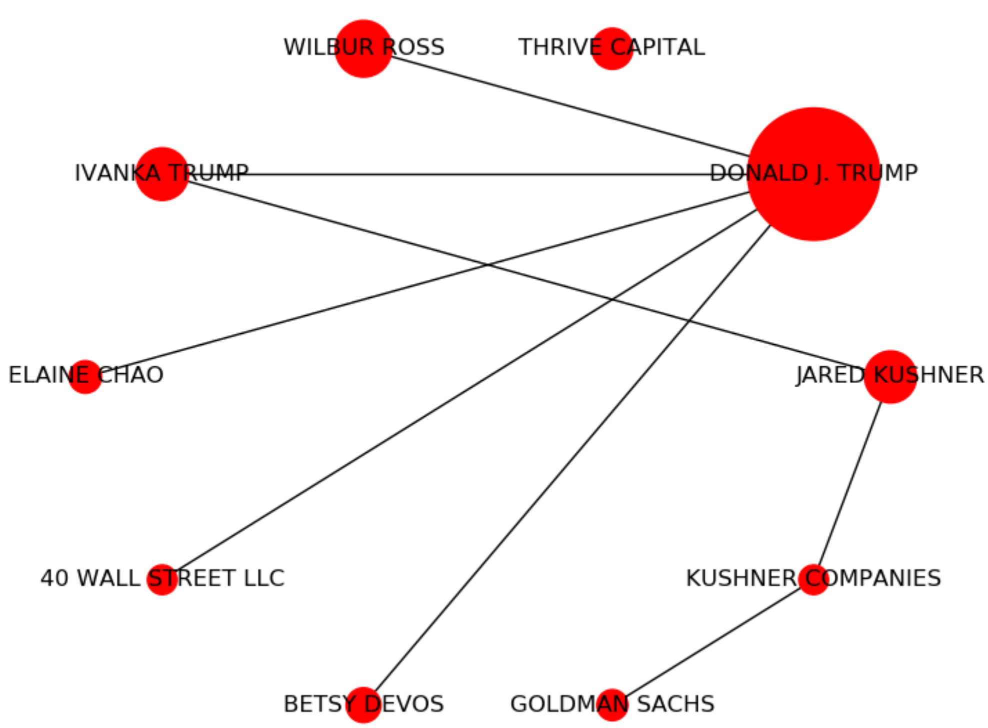


Figure 8. Subgraph: Top 10 Influential Nodes by Betweeness Centrality

3. Watts Strogatz Model

Random networks, in which nodes are connected randomly, are small worlds but not clustered. Watts-Strogatz (WS) proposed a network model that is both clustered and a small world, like real networks to some extent. They started off with the idea that people know their neighbors, so they connect each node with their *n* nearest neighbors.

This produces a clustered network, but it is a large world. They randomly rewired some of the links. WS discovered that rewiring only *a few* links is enough to decrease the average distance between nodes, i.e. make it a small world, while not affecting the clustering considerably.

Investigate this claim by WS. Start with nearest neighbor network (provided) of 100 nodes, where each node connects to its 6 nearest neighbors (similar to Fig. 1).

a. Calculate the average clustering coefficient and average path length of this network (C_{ave} and d_{ave})
b. Choose one link randomly and rewire one end to a different, randomly chosen, node
c. Calculate C_{ave} and d_{ave} again
d. Repeat parts b and c until all links have been rewired
e. Perform this simulation a sufficient number of times* to obtain representative averages of C_{ave} and d_{ave} for each number of rewired edges.
f. Plot these averages as a function of the number of rewired edges. Argue the claim made by WS.

* How can you tell if you've completed a sufficient number of simulations? This is a good question to consider.

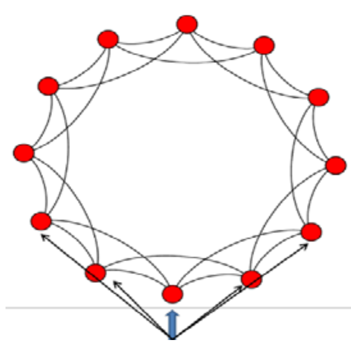


Figure 1. Each node connects with nearest neighbors. This produces a clustered but large-world network

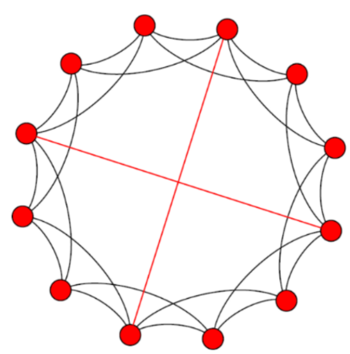


Figure 2. Network with rewired links (red). This is a clustered, small world network. (Note, red links in image were added not rewired

Introduction and Procedure

This project started with a regular graph with 100 nodes, each connected to 6 nearest neighbors, 3 on each side. A random node from a random edge was then selected to rewire to a random non-neighbor. This results in a slightly different graph. This continued till all edges were randomly selected and rewired. Eventually, a random graph was resulted. See depiction in Figure 1. Each time after an edge was rewired, two global graph metrics were calculated to evaluate the effect of edge rewiring: a) the average clustering coefficient, and b) the average shortest path length (distance) between all paired nodes. Clustering coefficient, C_{ave}, of a node is a measure of how connected the node's neighbors. To calculate the average shortest path length (distance), d_{ave}, of the graph, the graph must be connected, which means that each node must be connected to at least one other node. Average shortest path length of the graph is number of edges in the shortest path between two nodes, averaged over all pairs of nodes. This process was repeated for three set of 40 trials.

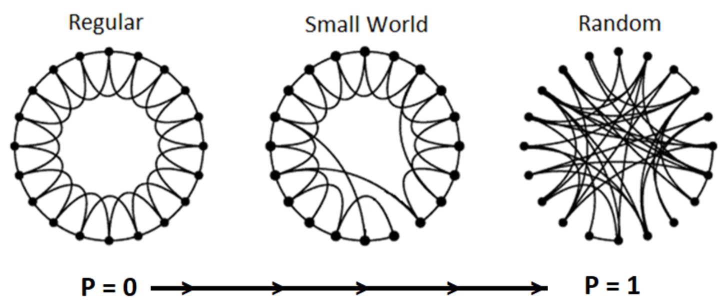


Figure 1. Regular Graph, Small World and a Random Graph

The Watts-Strogatz (WS) algorithm generates a regular graph consisting n nodes, each connected to k closest neighbors. The WS algorithm then progressively goes through one edge at a time rewiring it at random with p probability till all edges are gone through. Watts-Strogatz claims that rewiring only *a few* links is enough to decrease the average distance between nodes, i.e. make it a small world, while not affecting the clustering considerably.

Observation and Conclusion

My founding agrees with Watts-Strogatz's claim. After a few edges (2 to 8 edges) are rewired in each of the 120 trials, the graph changed from a regular graph to a small world graph. The graph has an average distance between all paired nodes, d_{ave}, less than 6 and the average clustering coefficient, C_{ave}, was not affected much.

Figure 2 below visualizes how rewiring of each random edge alters the graph slightly differently in the beginning. The randomized process resulted in a different random graph in each of the trials A through D. Some of the new wires reached nodes farther away, while some reached closer non-neighbors. A long-range rewired edge would significantly reduce the shortest path length (distance) between the nodes connected by the new edge. The shortest path (d(i) = 1) between the nodes connected by the i^{th} rewired edge went directly across the ring instead of the original series of hops. Such rewiring also reduced the distance between the ***nodes near the new edge***, since the neighboring nodes on each side could also utilize the edge to get to the neighboring nodes on other side of the ring. Hence the effect on the average distance, d_{ave}, between paired nodes of the graph amplifies even with one long-range rewired edge.

On the other hand, the clustering coefficient, C(i), of node i measures how much its friends are connected. To the node that was let go because an edge was rewired, it lost only one friend out of the 6 original neighbors. But this did not alter how the node's remaining friends are connected at all. On the other hand, to the node that acquired a new neighbor, its clustering coefficient was reduced since the new neighbor was not connected to any of the original 6 neighbors. So the rewiring only affected the C(i) of one of the originally connected nodes. Thus, the effect on the average clustering coefficient, C_{ave}, of the graph as a whole is not as significant as on the average shortest path length or distance, d_{ave}, of the graph as a whole. See Table 1 for a listing and changes of the metrics for these trials. Observed variance of metrics from trial to trial is further discussed at the end of this report.

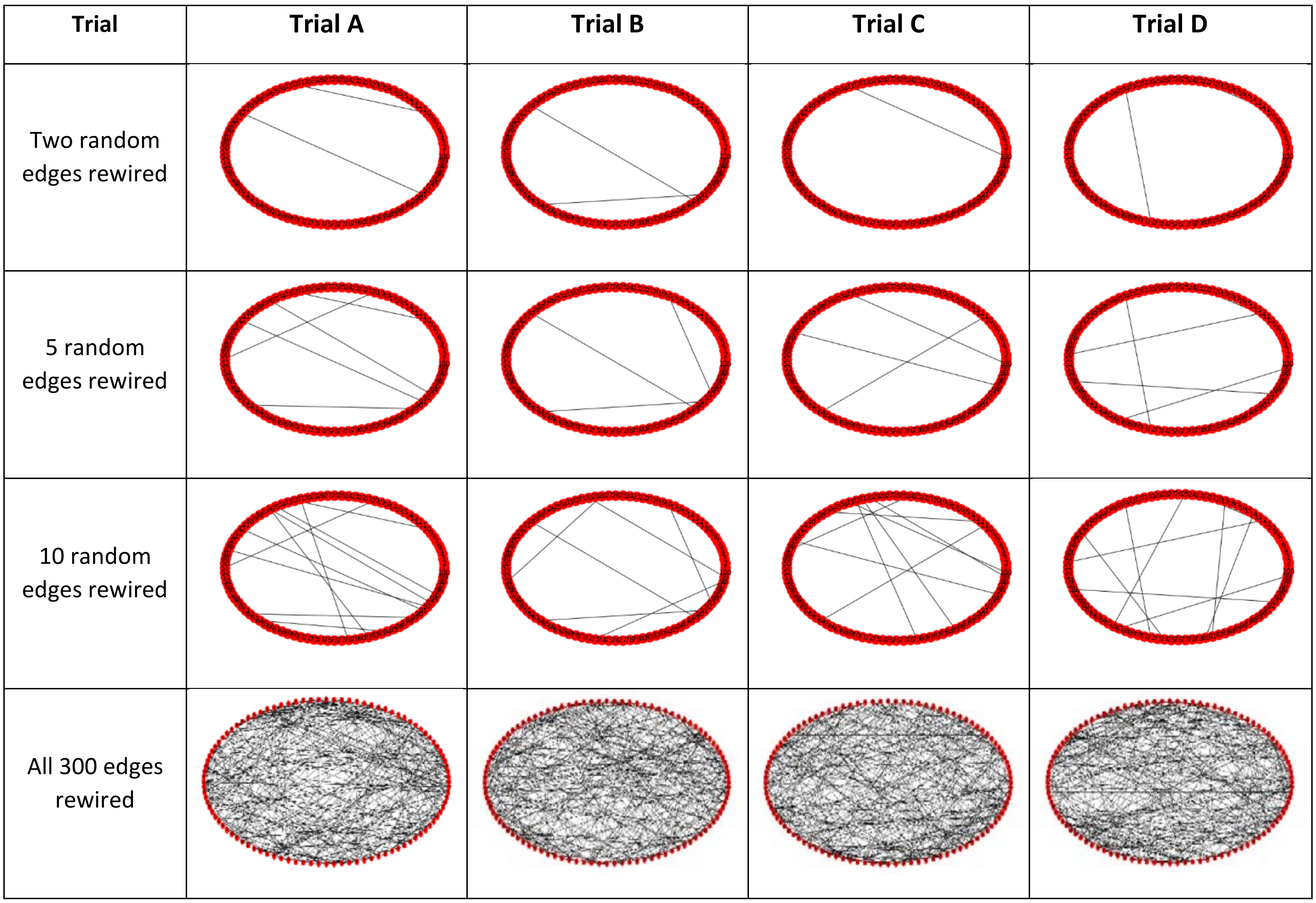


Figure 2. Visualization of Changes to Graph for Trials A To D

*Table 1. Trials A -D: Average Clustering Coefficient (**C_{ave}**) and Average Distance (**d_{ave}**) of Graph*

	Trial A		**Trial B**		**Trial C**		**Trial D**	
# of Rewired Edges	**C_{ave}**	**d_{ave}**	**C_{ave}**	**d_{ave}**	**C_{ave}**	**d_{ave}**	**C_{ave}**	**d_{ave}**
1	0.6000	8.7576	0.6000	8.7576	0.6000	8.7576	0.6000	8.7576
2	0.5943	7.7311	0.5943	7.1277	0.5943	7.3511	0.5966	8.3061
3	0.5886	6.7968	0.5862	6.2887	0.5888	7.1313	0.5909	6.8657
4	0.5807	6.5655	0.5784	6.1394	0.5850	6.6323	**0.5829**	**5.9784**
5	**0.5750**	**5.8851**	**0.5703**	**5.4354**	0.5817	6.4966	0.5771	5.6291
6	0.5693	5.3747	0.5702	5.4202	**0.5736**	**5.5673**	0.5714	5.1598
7	0.5648	5.2697	0.5645	5.2630	0.5702	5.3392	0.5636	4.9970
8	0.5596	4.9036	0.5575	5.2343	0.5664	5.1826	0.5579	4.7337
9	0.5556	4.8699	0.5529	5.2226	0.5609	5.1341	0.5551	4.6921
10	0.5476	4.7026	0.5495	4.9913	0.5523	4.7899	0.5471	4.6125
11	0.5401	4.6558	0.5420	4.7085	0.5481	4.7297	0.5411	4.4412
12	0.5387	4.5438	0.5364	4.6089	0.5401	4.6592	0.5392	4.4119
13	0.5390	4.5315	0.5317	4.6044	0.5339	4.5400	0.5358	4.3463
14	0.5340	4.3885	0.5273	4.5465	0.5288	4.4671	0.5299	4.2200
15	0.5300	4.3754	0.5213	4.4859	0.5239	4.4475	0.5223	4.1877
16	0.5257	4.3087	0.5181	4.3246	0.5212	4.2214	0.5144	4.1295
17	0.5208	4.2954	0.5147	4.2560	0.5143	4.1784	0.5119	4.0855
18	0.5164	4.2002	0.5114	4.2356	0.5085	4.0747	0.5085	4.0232
19	0.5138	4.1582	0.5030	4.1461	0.5006	4.0339	0.5053	4.0063
20	0.5114	4.1143	0.5009	4.0305	0.4929	4.0119	0.5020	4.0026
21	0.5070	4.0768	0.5032	4.0368	0.4900	3.9149	0.5000	3.9780
22	0.5000	3.9622	0.4977	3.9943	0.4854	3.8822	0.4950	3.9053
23	0.4970	3.9349	0.4904	3.9622	0.4826	3.8275	0.4904	3.8721
24	0.4925	3.9079	0.4823	3.9267	0.4798	3.8040	0.4869	3.8566
25	0.4868	3.8469	0.4775	3.8265	0.4733	3.7503	0.4813	3.8137

Data Analysis and Results

The rewiring process was repeated for 40 trials and the mean C_{ave} and mean d_{ave} were plotted as a function of the number of rewired edges. Figure 3 presents the normalized $\overline{C}_{ave}$ and $\overline{d}_{ave}$ values on the y-axis against the number of rewired edges on the x-axis in a semi-logarithmic scale. The plot of the mean average distance $\overline{d}_{ave}$ drops considerably at the beginning of the process when only a few edges are rewired, while the plot of mean average clustering coefficient $\overline{C}_{ave}$ declines on more gradual slope. For $\overline{d}_{ave}$ to drop below 6, which is 0.68 of the original average shortest path length 8.76, thus making the graph a small world, it only took as few as 5 edges to be rewired. At that point, $\overline{C}_{ave}$ was still well above 0.9 of the original average clustering coefficient. See green dotted line in Figure 3 showing the average metrics when 5 edges were rewired.

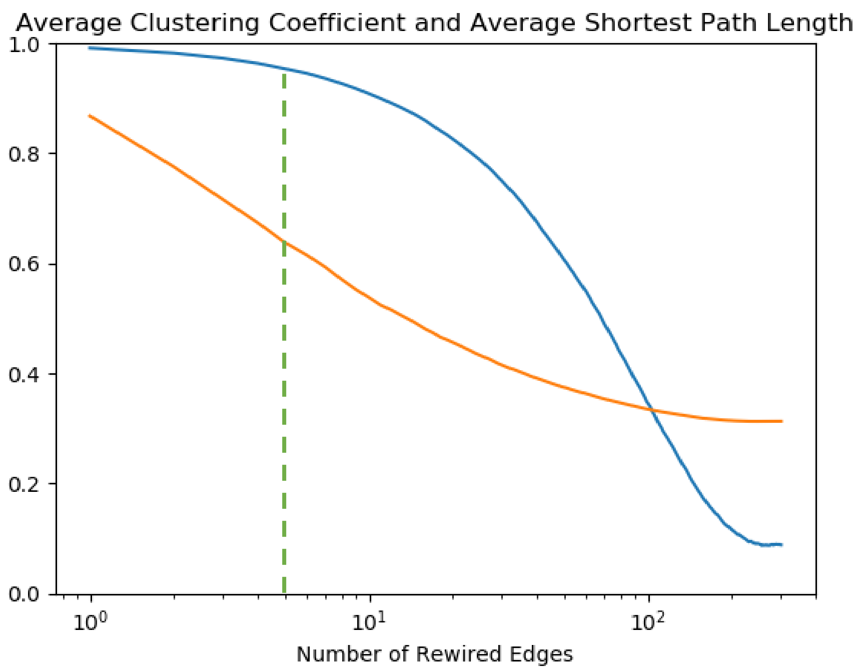


Figure 3. Mean Average Shortest Path Length $\overline{d}_{ave}$ and Mean Average Clustering Coefficient C_{ave} of Graph as a Function of the Number of Rewired Edges

Variance from Trial to Trial and Conclusion

Because of random chance, a rewired edge may be short-range, long-range or somewhere between. So, the effect of rewiring to the average distance, d_{ave}, varies from trials to trials each time an edge is rewired, as shown in Table 1 above for Trials A through D. So it varied from 4 to 6 edge rewiring to change the graph to a small world.

Table 2 below shows the average results over 10, 20, 40, 80 and 120 trials. It shows that it took 20 trials and up for the variance of the mean of the average distance, d_{ave}, to stay below 0.2. And it shows consistently that 5 edge rewiring was enough to change the original regular graph to a small-world with the mean average distance $\overline{d}_{ave}$ below 6. So, I conclude that 20 trials are sufficient with variance below 0.2 for purpose of this exercise. Additionally, Table 3 shows that the results over 20, 40, 80 and 120 trials consistently reduced the mean average distance $\overline{d}_{ave}$ of the graph by more than 0.68 of the original d_{ave} to below 6 and the mean average clustering coefficient $\overline{C}_{ave}$ well above 0.95 of the original C_{ave}, of the original graph.

Table 2. Mean Average Clustering Coefficient ($\bar{C}_{ave}$) and Mean Average Shortest Path Length ($\bar{d}_{ave}$) of Graph

	From 10 trials			From 20 trials			From 40 trials			From 80 trials			From 120 trials		
# of Rewired	Mean C_{ave}	Mean d_{ave}	Variance d_{ave}	Mean C_{ave}	Mean d_{ave}	Variance d_{ave}	Mean C_{ave}	Mean d_{ave}	Variance d_{ave}	Mean C_{ave}	Mean d_{ave}	Variance d_{ave}	Mean C_{ave}	Mean d_{ave}	Variance d_{ave}
0	0.6000	8.7576	0.0000	0.6000	8.7576	0.0000	0.6000	8.7576	0.0000	0.6000	8.7576	0.0000	0.6000	8.7576	0.0000
1	0.5946	7.9639	0.2295	0.5940	7.8303	0.2511	0.5943	7.7950	0.2487	0.5945	7.7770	0.2436	0.5945	7.7603	0.2404
2	0.5881	7.2795	0.1811	0.5880	7.2729	0.2694	0.5887	7.1059	0.2753	0.5890	6.9991	0.3505	0.5889	7.0211	0.3411
3	0.5828	6.5783	0.3498	0.5827	6.5773	0.3292	0.5828	6.4656	0.2675	0.5834	6.4088	0.2863	0.5833	6.4466	0.2927
4	0.5759	6.1321	0.2240	0.5767	6.0955	0.2152	0.5771	6.0481	0.2021	0.5779	6.0322	0.2086	0.5778	6.0588	0.2002
5	**0.5708**	**5.7009**	0.2223	**0.5721**	**5.6843**	0.1458	**0.5722**	**5.7314**	0.1644	**0.5727**	**5.7075**	0.1595	**0.5729**	**5.7227**	0.1636
6	0.5648	5.5190	0.2626	0.5666	5.4471	0.1826	0.5670	5.4688	0.1411	0.5674	5.4218	0.1086	0.5676	5.4299	0.1011
7	0.5600	5.3847	0.2588	0.5613	5.2347	0.1639	0.5617	5.2206	0.1082	0.5624	5.1891	0.0905	0.5625	5.2117	0.0896
8	0.5571	5.1796	0.1005	0.5568	5.0561	0.0813	0.5567	5.0299	0.0753	0.5574	5.0155	0.0699	0.5573	5.0292	0.0666
9	0.5536	4.9458	0.0735	0.5519	4.8562	0.0575	0.5518	4.8747	0.0600	0.5522	4.8529	0.0623	0.5523	4.8735	0.0604
10	0.5483	4.8314	0.0688	0.5467	4.7374	0.0546	0.5464	4.7490	0.0493	0.5469	4.7213	0.0510	0.5471	4.7403	0.0506
11	0.5433	4.7149	0.0707	0.5417	4.6515	0.0508	0.5410	4.6417	0.0496	0.5420	4.6084	0.0434	0.5422	4.6270	0.0459
12	0.5378	4.5926	0.0552	0.5364	4.5469	0.0393	0.5360	4.5337	0.0370	0.5368	4.5006	0.0335	0.5372	4.5128	0.0324
13	0.5327	4.4627	0.0438	0.5316	4.4316	0.0282	0.5309	4.4189	0.0296	0.5313	4.4079	0.0263	0.5321	4.4286	0.0299
14	0.5274	4.3240	0.0205	0.5259	4.3198	0.0147	0.5255	4.3308	0.0202	0.5261	4.3240	0.0215	0.5269	4.3398	0.0244
15	0.5225	4.2232	0.0164	0.5210	4.2312	0.0113	0.5203	4.2436	0.0156	0.5211	4.2438	0.0180	0.5220	4.2663	0.0221
16	0.5172	4.1662	0.0120	0.5159	4.1696	0.0098	0.5153	4.1803	0.0143	0.5163	4.1777	0.0156	0.5172	4.2001	0.0202
17	0.5124	4.1182	0.0148	0.5104	4.1167	0.0135	0.5094	4.1272	0.0146	0.5111	4.1220	0.0145	0.5120	4.1409	0.0181
18	0.5072	4.0756	0.0156	0.5050	4.0688	0.0144	0.5043	4.0767	0.0157	0.5062	4.0704	0.0147	0.5072	4.0868	0.0183
19	0.5019	4.0397	0.0127	0.4995	4.0249	0.0136	0.4994	4.0283	0.0131	0.5011	4.0243	0.0135	0.5019	4.0356	0.0142
20	0.4978	3.9888	0.0175	0.4954	3.9734	0.0164	0.4951	3.9868	0.0145	0.4966	3.9811	0.0150	0.4972	3.9897	0.0143
21	0.4930	3.9249	0.0142	0.4904	3.9165	0.0134	0.4898	3.9299	0.0119	0.4913	3.9257	0.0112	0.4923	3.9384	0.0110
22	0.4883	3.8912	0.0117	0.4862	3.8886	0.0124	0.4853	3.8902	0.0117	0.4865	3.8837	0.0102	0.4874	3.8973	0.0105
23	0.4829	3.8465	0.0138	0.4805	3.8486	0.0126	0.4803	3.8595	0.0115	0.4818	3.8501	0.0099	0.4828	3.8615	0.0097
24	0.4780	3.8027	0.0105	0.4760	3.8040	0.0086	0.4753	3.8166	0.0095	0.4766	3.8114	0.0091	0.4777	3.8249	0.0089
25	0.4732	3.7541	0.0091	0.4711	3.7658	0.0073	0.4706	3.7815	0.0084	0.4717	3.7775	0.0082	0.4729	3.7919	0.0081

Table 3. Normalized Average Clustering Coefficient, Cave(i)/Cave(0),and Average Shortest Path Length, dave(i)/dave(0), of Graph

	From 20 trials		From 40 trials		From 80 trials		From 120 trials	
# of Rewired Edges	$C_{ave}(i)/C_{ave}(0)$	$d_{ave}(i)/d_{ave}(0)$	$C_{ave}(i)/C_{ave}(0)$	$d_{ave}(i)/d_{ave}(0)$	$C_{ave}(i)/C_{ave}(0)$	$d_{ave}(i)/d_{ave}(0)$	$C_{ave}(i)/C_{ave}(0)$	$d_{ave}(i)/d_{ave}(0)$
1	0.9901	0.8941	0.9904	0.8901	0.9908	0.8880	0.9909	0.8861
2	0.9800	0.8305	0.9812	0.8114	0.9816	0.7992	0.9815	0.8017
3	0.9711	0.7510	0.9713	0.7383	0.9723	0.7318	0.9722	0.7361
4	0.9612	0.6960	0.9618	0.6906	0.9631	0.6888	0.9630	0.6918
5	0.9534	0.6491	0.9537	0.6544	0.9546	0.6517	0.9549	0.6535
6	0.9444	0.6220	0.9449	0.6245	0.9457	0.6191	0.9459	0.6200
7	0.9354	0.5977	0.9362	0.5961	0.9374	0.5925	0.9375	0.5951
8	0.9280	0.5773	0.9279	0.5744	0.9289	0.5727	0.9289	0.5743
9	0.9198	0.5545	0.9197	0.5566	0.9204	0.5541	0.9204	0.5565
10	0.9111	0.5409	0.9106	0.5423	0.9115	0.5391	0.9118	0.5413
11	0.9029	0.5311	0.9017	0.5300	0.9033	0.5262	0.9036	0.5283
12	0.8940	0.5192	0.8934	0.5177	0.8946	0.5139	0.8953	0.5153
13	0.8861	0.5060	0.8848	0.5046	0.8855	0.5033	0.8868	0.5057
14	0.8765	0.4933	0.8758	0.4945	0.8768	0.4937	0.8781	0.4955
15	0.8684	0.4831	0.8672	0.4846	0.8686	0.4846	0.8700	0.4872
16	0.8599	0.4761	0.8588	0.4773	0.8605	0.4770	0.8620	0.4796
17	0.8507	0.4701	0.8490	0.4713	0.8518	0.4707	0.8534	0.4728
18	0.8417	0.4646	0.8406	0.4655	0.8437	0.4648	0.8454	0.4667
19	0.8326	0.4596	0.8323	0.4600	0.8351	0.4595	0.8365	0.4608
20	0.8256	0.4537	0.8251	0.4552	0.8276	0.4546	0.8287	0.4556
21	0.8174	0.4472	0.8163	0.4487	0.8189	0.4483	0.8205	0.4497
22	0.8103	0.4440	0.8088	0.4442	0.8108	0.4435	0.8123	0.4450
23	0.8009	0.4395	0.8006	0.4407	0.8029	0.4396	0.8046	0.4409
24	0.7933	0.4344	0.7922	0.4358	0.7943	0.4352	0.7961	0.4367
25	0.7852	0.4300	0.7843	0.4318	0.7861	0.4313	0.7881	0.4330