Rufeng Ma, 661681840

Road network of Chicago region

This is the road transportation network of the Chicago region (USA). Nodes are transportation nodes (1,467 vertices), and edges are connections (1,298 edges). The data source is from the website which was updating by astabler on GitHub. The link is here www.bgu.ac.il/~bargera/tntp/. They collected multiple transportation networks including Anaheim Network, Barcelona Network, Austin Network, Chicago network, and more. I downloaded the TSV file from KONECT. This database was created by KONECT, in April 2017. All edges are undirected, and edges are unweighted. Study these transportation networks can help people solve the most basic problems in transportation research, like traffic assignment problem.

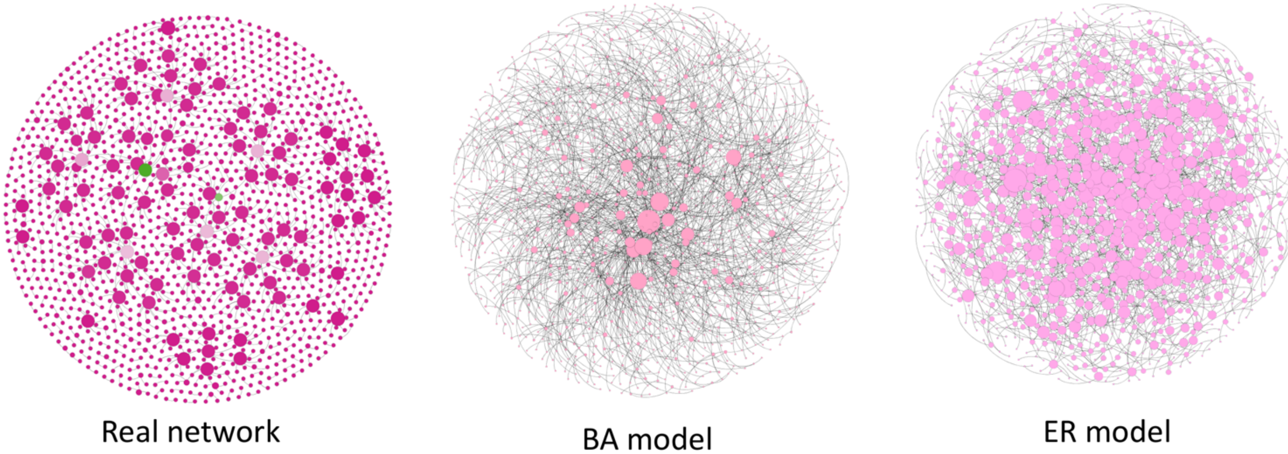


Figure 1. Road networks of Chicago region. Here are 3 networks, one is the real network, data collected from Chicago region, the other 2 are BA model and ER model network. BA model based on the minimum degree of the real network. When I was doing this, I process the real network in the Gephi software first, then can get the minimum degree and the node number. Then I used the code Professor Szymanski provided on the class to get 2 txt file contains edges of nodes. The ER network based on the average degree of the nodes in the real network, then we can get the similar number of nodes. Node size are ranking based the degree of each node. But I did not control the minimum and the maximum node size are the same in the 3 networks. That is the part I can improve. And also the edges thickness I also need to control them to make the comparison clearer. Layout was Fruchterman Reingold. The gravity is 20.0, area is 10000.0, and the speed is 1.0. I used the statistics functions in the Gephi. They are average degree, network diameter, connected components, average clustering coefficient, and average path length. The length distribution of shortest paths I cannot find in Gephi. But the embedded function for NetworkX was easy to be found via google. So for this part I used NetworkX to get the length of shortest paths. Please see the attachment. Following is the average of the degree, the length distribution of the shortest paths, the clustering coefficient distribution, the betweenness centrality distribution and the connected components size distribution. See table 1, it contains these average data for real network,

BA and ER model. Table 2. It contains the variances of each set of data for real network, BA and ER model.

Network	Avg. of degree	Avg. of clustering coefficient distribution	Avg. of betweenness centrality distribution	Avg. of connected components size distribution
Real network	1.76959782	0	1154.94206	20.77164281
BA	2	0.08907764	5227.07362	0
ER	2.08510638	0	4732.73159	4.402618658

Table 1. Average of properties

Network	Var. of degree	Var. of clustering coefficient distribution	Var. of betweenness centrality distribution	Var. of connected components size distribution
Real network	6.45301778	0	107876240	1789.91985
BA	2	3.31E-08	1234961521	0
ER	2.08510638	0	59457348.3	160.532278

Table 2. Variance of properties

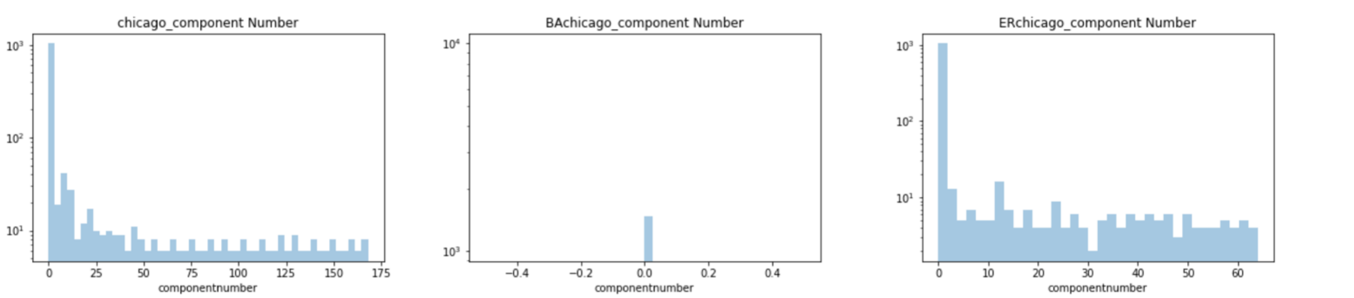


Figure 2. The component number of 3 networks for Chicago road.

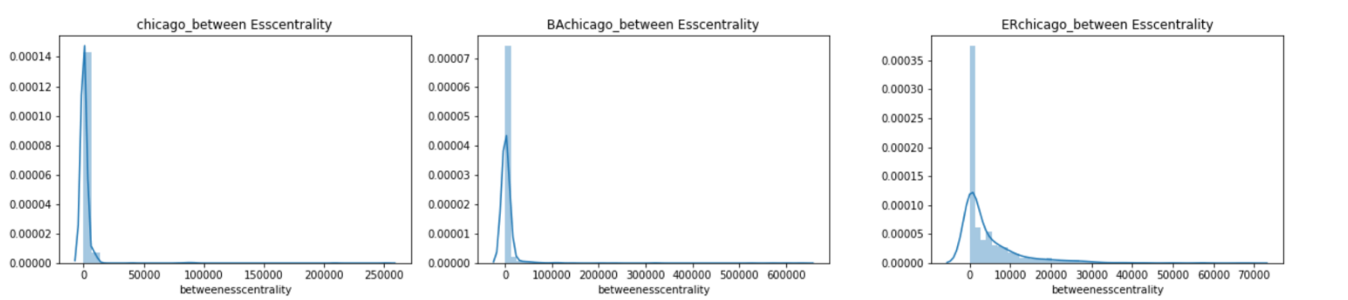


Figure 3. The betweenness centrality of 3 networks for Chicago road.

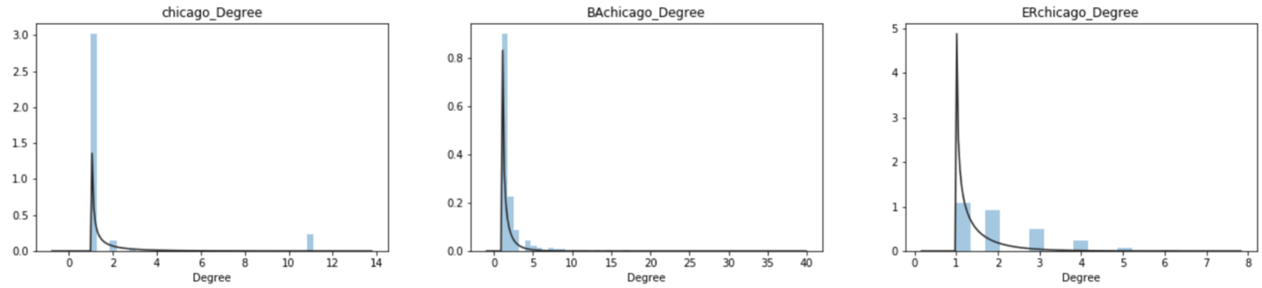
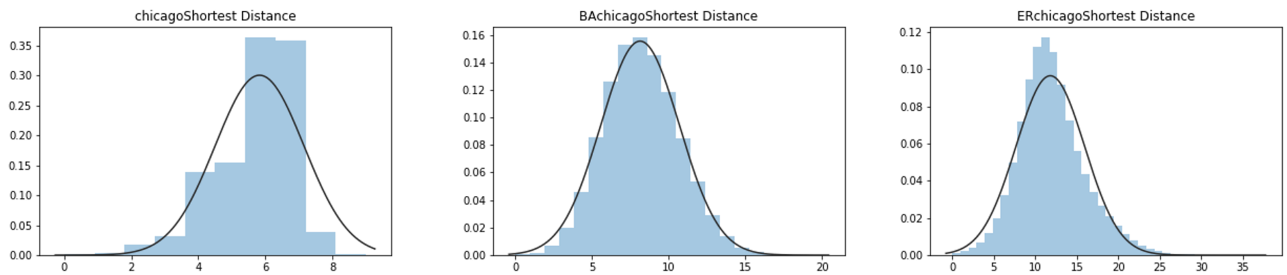


Figure 4. The degree of each nodes of 3 networks for Chicago road.



The length distribution of shortest path for 3 Chicago road networks

For question 4 and 5.

For the Chicago road, the most important nodes are the nodes who have the highest degree numbers. Which means that is the main node can connect multiple roads. Which is a cross. And also if the node always on the shortest path that also means it is an important point. The betweenness centrality high node is another important node. This is a scale free network.

Hamsterster friendship network

This is the network contains friendship between user of the website hamsterster.com.

Nodes are users (1,858 vertices), and edges are friendships (12,534 edges). The data was collected by KONECT, in April 2017. http://konect.uni-koblenz.de/networks/petster-friendships-hamster.

This is interesting that people can know each other because they have the same kind of pets. I found a paper talked about catster, dogster and hamsterster. Because individual owners may own both cats and dogs, and the set might can connect to hamsterster. All edges are undirected, and edges are unweighted. Study the hamsterster can help us understand how people know each other from a hobby network.

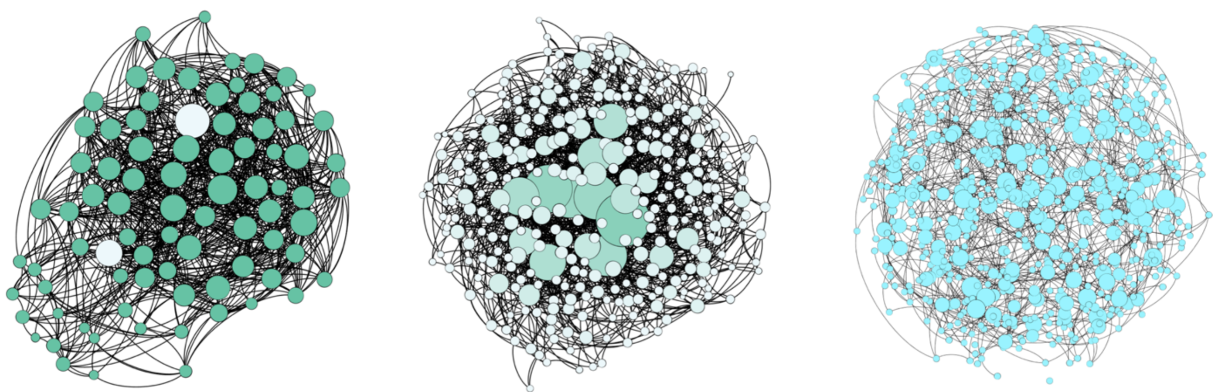


Figure 5. Hamsterster friendship network. Here are 3 networks, one is the real network, data collected from the website Hamsterster, the other 2 are BA model and ER model network. BA model based on the minimum degree of the real network. When I was doing this, I process the real network in the Gephi software first, then can get the minimum degree and the node number. Then I used the code Professor Szymanski provided on the class to get 2 txt file contains edges of nodes. The ER network based on the average degree of the nodes in the real network, then we can get the similar number of nodes. Node size are ranking based the degree of each node. But I did not control the minimum and the maximum node size are the same in the 3 networks. That is the part I can improve. And also the edges thickness I also need to control them to make the comparison clearer. Layout was ForceAtlas2. The tolerance (speed) is 1.0, the approximation is 1.2, the gravity is 1.0, and the scaling is 2.0. I filted the network using node degree. I used the statistics functions in the Gephi. They are average degree, network diameter, connected components, average clustering coefficient, and average path length. The length distribution of shortest paths I cannot find in Gephi. But the embedded function for NetworkX was easy to be found via google. So for this part I used NetworkX to get the length of shortest paths. Please see the attachment. Following is the average of the degree, the length distribution of the shortest paths, the

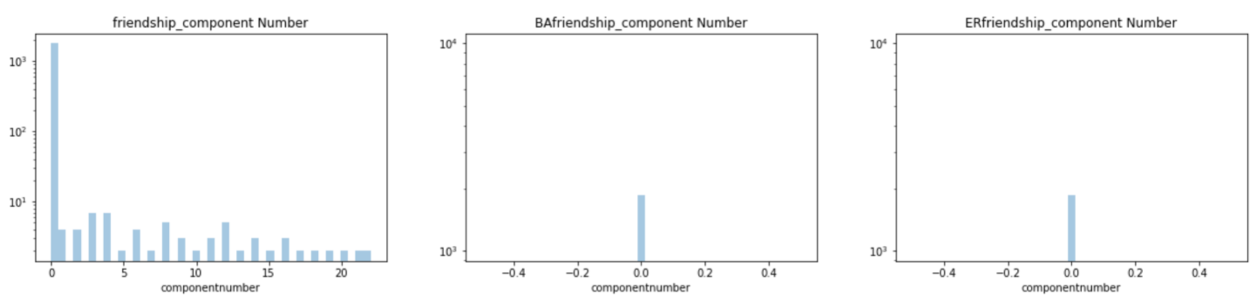
clustering coefficient distribution, the betweenness centrality distribution and the connected components size distribution. See table 3, it contains these average data for real network, BA and ER model. Table 4. It contains the variances of each set of data for real network, BA and ER model.

Network	Avg. of degree	Avg. of clustering coefficient distribution	Avg. of betweenness centrality distribution	Avg. of connected components size distribution
Real network	13.4919268	0.1413864	2108.8972	0.360064586
BA	10	0.02711327	2001.56405	0
ER	13.0086114	0.00746424	2051.14801	0

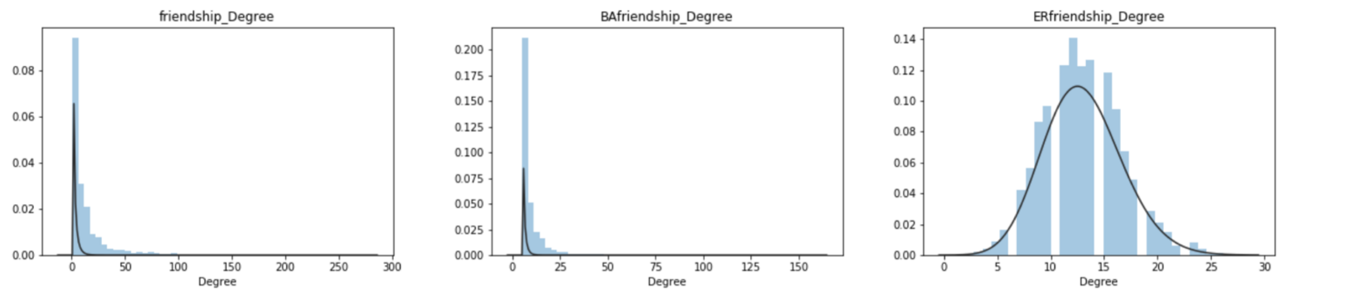
Table 3. Average of properties

Network	Var. of degree	Var. of clustering coefficient distribution	Var. of betweenness centrality distribution	Var. of connected components size distribution
Real network	430.014205	0.04209211	49087262.6	4.75719806
BA	130.97469	0.00228543	73911553.6	0
ER	13.5152731	0.00013842	1309371.92	0

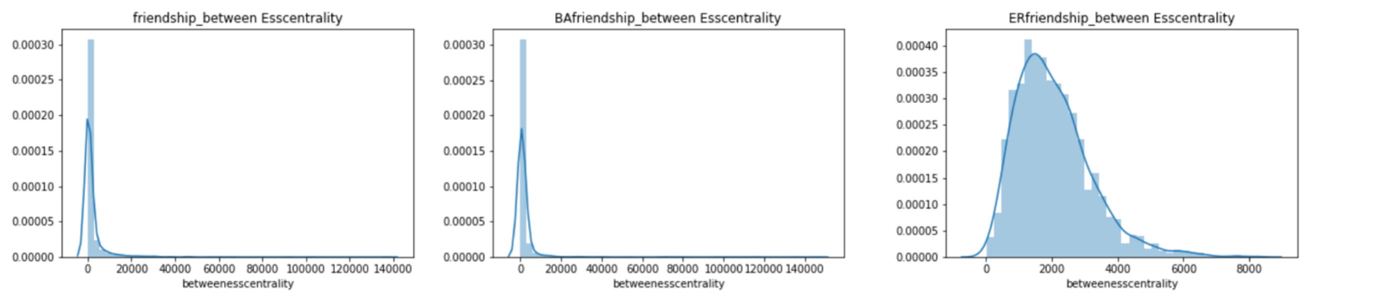
Table 4. Variance of properties



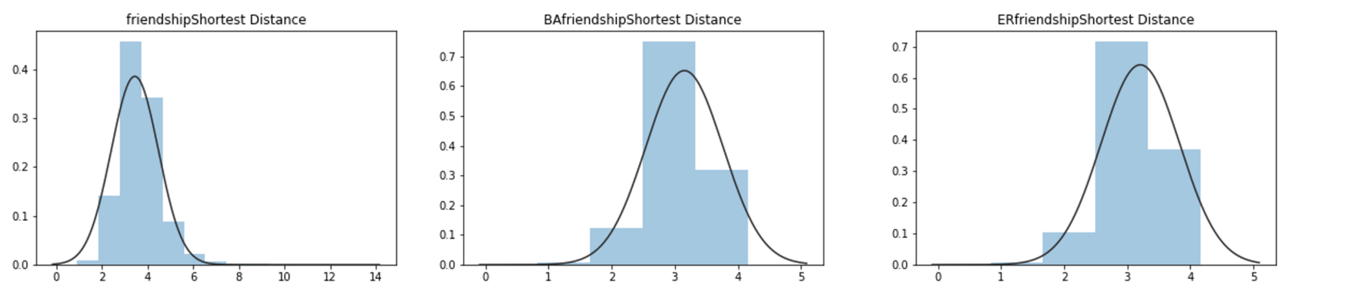
The component number of friendship network



The degree of friendship network



The betweenness centrality of friendship network.



The length distribution of shortest path for 3 friendship networks

Discussion:

For question 4 and 5.

For the friendship networks, the most important nodes are the nodes who have the highest degree numbers. Which is a cross. And also if the node always on the shortest path that also means it is an important point. The betweenness centrality high node is another important node. This is a scale free network.

Facebook social circles

The dataset consists of ‘circles’ (‘friends lists’) from Facebook. Facebook data was collected from survey participants using the Facebook app. Facebook data has been anonymized by replacing the facebook ids for each user with a value. Because this dataset is giant for my computer, I deleted some notes and edges. Finally, the final version of the network contains 1,570 nodes and 64,920 edges. Compared with the previous 2 cases, I found the social networks has way more connections between people. I think this is why the website like Facebook called social networks, because the purpose of people having these websites’ account is to make connections. From their description, I know the edges are unweighted and undirected, so I did not consider the in- and out- degrees.

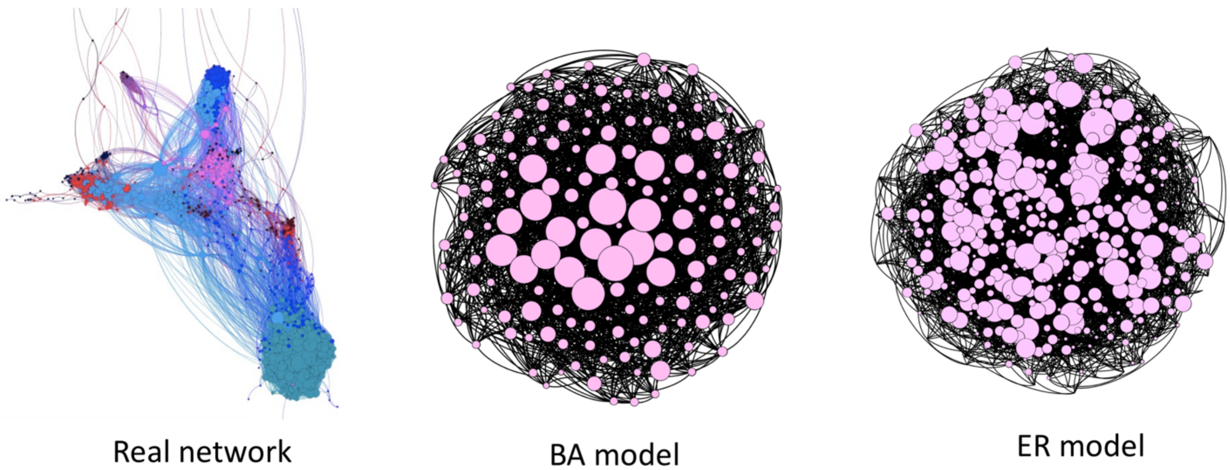


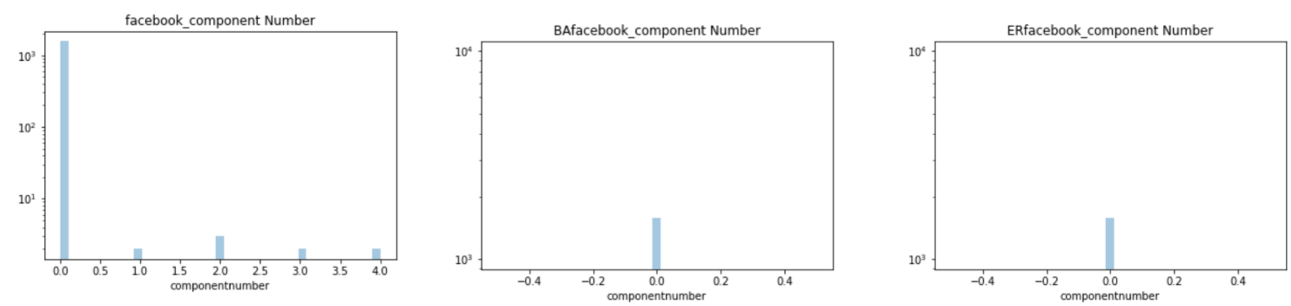
Figure 3. Facebook social circle. Here are 3 networks, one is the real network, data collected from the Facebook App, the other 2 are BA model and ER model network. BA model based on the minimum degree of the real network. When I was doing this, I process the real network in the Gephi software first, then can get the minimum degree and the node number. Then I used the code Professor Szymanski provided on the class to get 2 txt file contains edges of nodes. The ER network based on the average degree of the nodes in the real network, then we can get the similar number of nodes. Node size are ranking based the degree of each node. But I did not control the minimum and the maximum node size are the same in the 3 networks. That is the part I can improve. And also the edges thickness I also need to control them to make the comparison clearer. Layout was ForceAtlas2. The tolerance (speed) is 1.0, the approximation is 1.2, the gravity is 1.0, and the scaling is 2.0. I filted the network using node degree. I used the statistics functions in the Gephi. They are average degree, network diameter, connected components, average clustering coefficient, and average path length. The length distribution of shortest paths I cannot find in Gephi. But the embedded function for NetworkX was easy to be found via google. So for this part I used NetworkX to get the length of shortest paths. Please see the attachment. Following is the average of the degree, the length distribution of the shortest paths, the clustering coefficient distribution, the betweenness centrality distribution and the connected components size distribution. See table 5, it contains these average data for real network, BA and ER model. Table 6. It contains the variances of each set of data for real network, BA and ER model.

Network	Avg. of degree	Avg. of clustering coefficient distribution	Avg. of betweenness centrality distribution	Avg. of connected components size distribution
Real network	41.2496815	0.52438326	2618.8293	0.014012739
BA	40	0.08907764	996.93949	0
ER	39.2649682	0.02503906	1051.36752	0

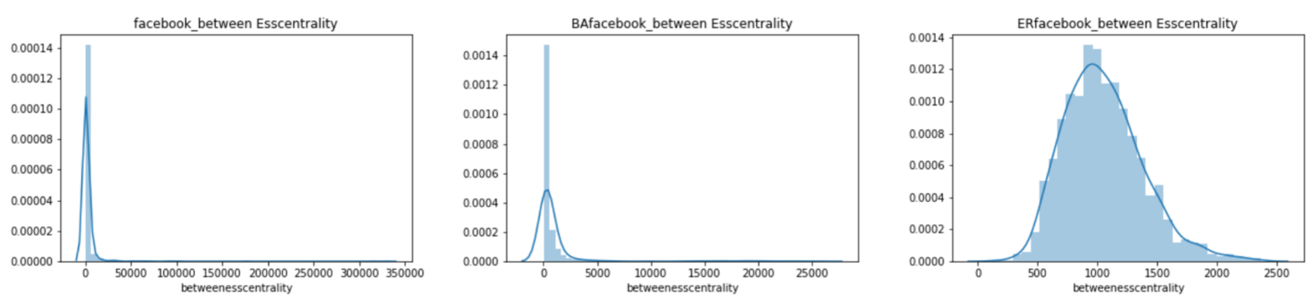
Table 3. Average of properties

Network	Var. of degree	Var. of clustering coefficient distribution	Var. of betweenness centrality distribution	Var. of connected components size distribution
Real network	1841.7069	0.05058068	200331141	0.04059383
BA	1398.0204	0.00121736	8824832.36	0
ER	37.951417	3.54E-05	108268.824	0

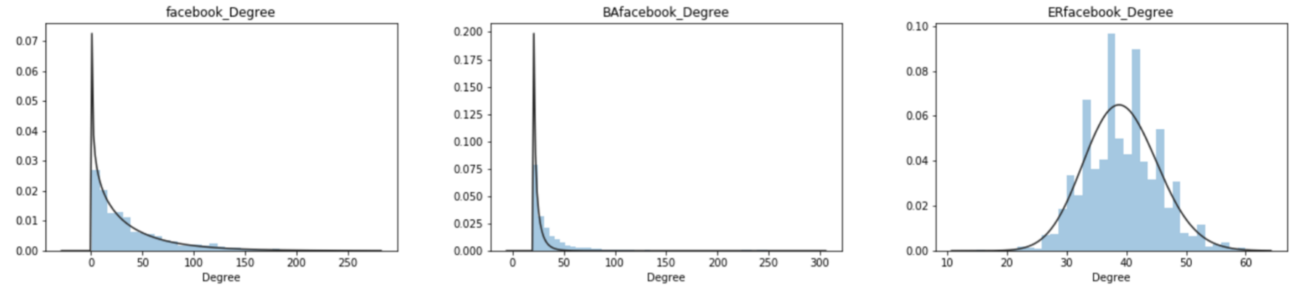
Table 4. Variance of properties



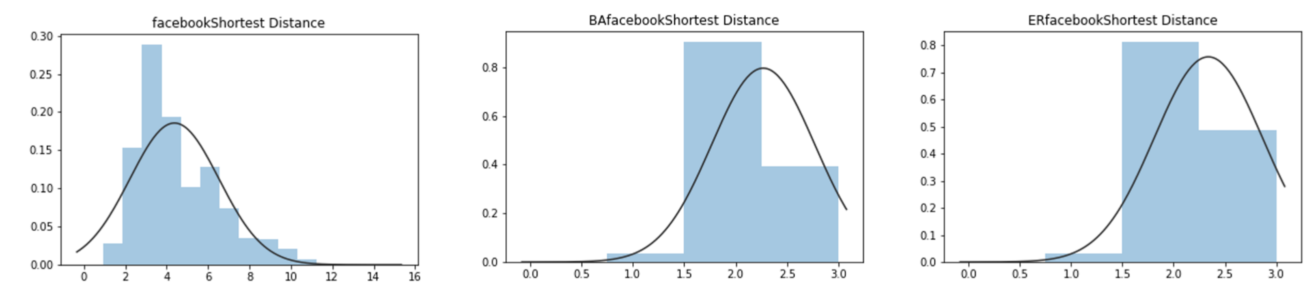
The component number of Facebook network.



The betweenness centrality of Facebook network.



The degree of Facebook networks.



The length distribution of shortest path for facebook networks.

Discussion:

For question 4 and 5.

For the facebook networks, the most important nodes are the nodes who have the highest degree numbers. Which is a cross. And also if the node always on the shortest path that also means it is an important point. The betweenness centrality high node is another important node. This is a scale free network.

Other:

I used Python for global network properties calculation. The node vice properties are exported from Gephi into .csv files. Additionally, the shorted path is calculated using Python networkX package. Then I used Python Pandas to read the .csv files into the DataFrame for mean and variance calculations. The corresponding histograms are plotted using the Matplotlib package and histogram fitting is done using Scipy stat package. All the code is included in the Jupyter notebook file.

All python code can be run on Windows laptop. And all code has been attached as attachements.