NANYANG TECHNOLOGICAL UNIVERSITY

CZ4071 Network Science
Individual Assignment

Submitted by:

Andrew Koh Jin Jie
U1522383H

School of Computer Science & Engineering

Date of submission: 9th April 2018

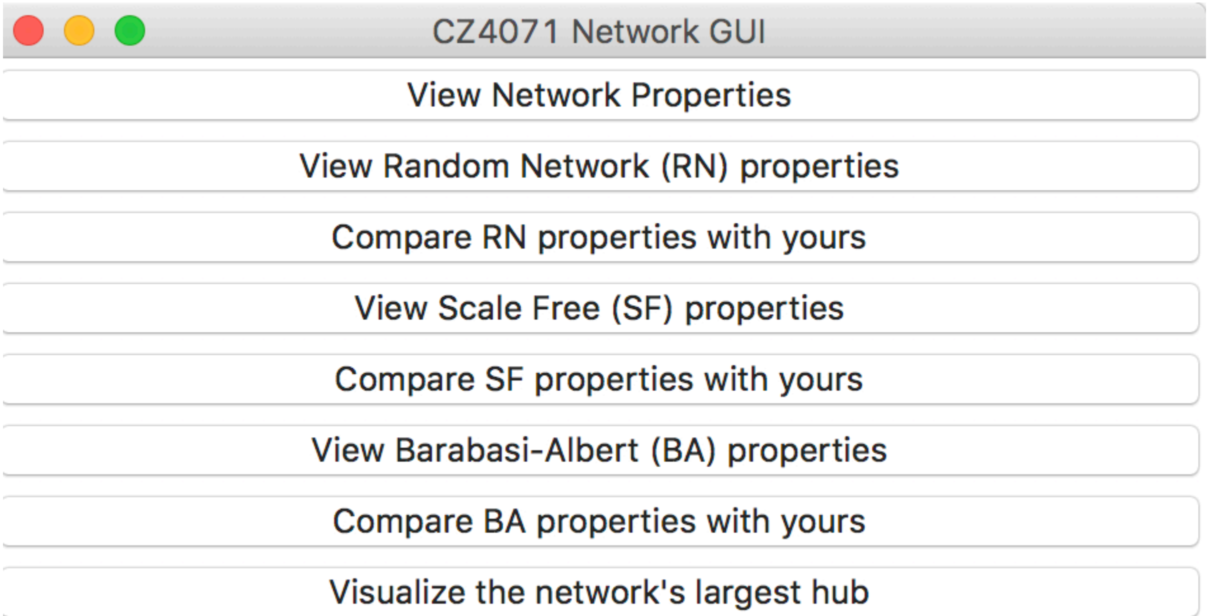
Using the program

The GUI is intuitive, simple and easy to use. First, when the command line is typed in, the user will be asked to choose a csv file to load and read. There are sample csv files to read and try out. The format of the csv file is each row corresponding to an edge. The first and second column are the vertex numbers of the both side of the edge

```
Choose csv file to be analysed
```

Then, the program will take a minute or so to run the backend computations. It may or may not be faster or slower depending on the user's processing power.

When it is done, the main menu will appear.



View your Network's Properties

Clicking on "View Network Properties" will display this screen.

```
Network degree distribution for k =  1 is approximately 0.042730187814933575
Network degree distribution for k =  2 is approximately 0.04976637654603756
Network degree distribution for k =  3 is approximately 0.049344938158497484
Network degree distribution for k =  4 is approximately 0.04863032524049473

Network average path length for N =  54575is  3.3350154838473238

Network average clustering coefficient for N =  54575 is  0.13647174362435638

Network average degree for N =  54575 is  18.257553825011453
```

Quit

This interface provides the information of several useful properties that are often considered when analyzing a network.

However, using the other comparison buttons are recommended in order to give the user a better understanding of the network.

View Random Network/Scale Free/Barabasi-Albert Properties

Clicking on "View Random Network Properties" or "View Scale Free Properties" or "View Barabasi-Albert Properties" will display this screen.

```
RN degree distribution for k = 1 is expected to be 0.0
RN degree distribution for k = 2 is expected to be 0.0
RN degree distribution for k = 3 is expected to be 0.0
RN degree distribution for k = 4 is expected to be 0.0
RN average path length for N = 47540is expected to be  1.0687929049512799

RN average clustering coefficient for N = 47540 is expected to be  0.4999894825410181

RN average degree for N = 47540 is expected to be  23769.5
```

Quit

```
SF degree distribution for k = 1 is expected to be 1.5
SF degree distribution for k = 2 is expected to be 0.26516504294495535
SF degree distribution for k = 3 is expected to be 0.09622504486493761
SF degree distribution for k = 4 is expected to be 0.046875
SF average path length for N = 47540 is expected to be  5.8616683143308901

SF size of largest hub for N = 47540 is expected to be  1312.3191055305542

SF average degree for N = 47540 is expected to be  23769.5
```

Quit

```
BA degree distribution for k = 1 is expected to be 1.0
BA degree distribution for k = 2 is expected to be 0.125
BA degree distribution for k = 3 is expected to be 0.037037037037037035
BA degree distribution for k = 4 is expected to be 0.015625
BA average path length for N = 47540is expected to be  4.5312062034700133

BA diameter for N = 47540is expected to be  6.9809680692455904

BA average clustering coefficient for N = 47540 is expected to be  0.00030494951212812717
```

Quit

These values are computed using the theoretical formulas that the respective types of networks of the same size are projected to have. These values will allow the user to check against his network's and decide how similar his network is to the theoretical models.

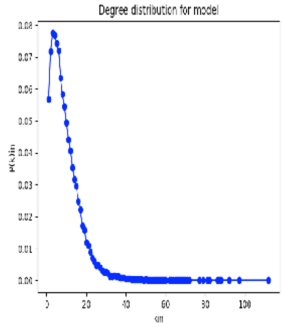
Again, using the compare buttons will provide a more comprehensive idea and analysis.

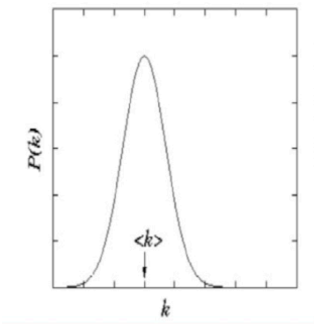
Comparing Random Network/Scale Free/Barabasi-Albert Properties with yours

Clicking on “Comparing Random Network Properties” or “Comparing Scale Free Properties” or “Comparing Barabasi-Albert Properties” will display this screen.

```
numnodes = Number of Nodes
apl = Average Path Length
acc= Average Clustering Coeff
avgdeg= Average degree
```

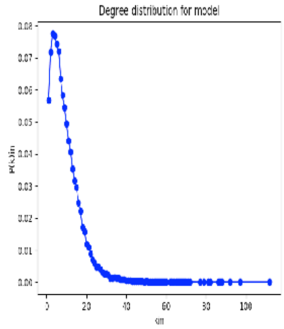
	Network	RN
numnodes	47540	47540
apl	4.6842357692793133	1.0687929049512799
acc	0.11618233863494011	0.4999894825410181
avgdeg	9.376861590239798	23769.5

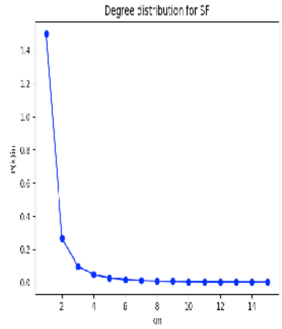





```
numnodes = Number of Nodes
apl = Average Path Length
Ldeg= Largest degree
avgdeg= Average degree
```

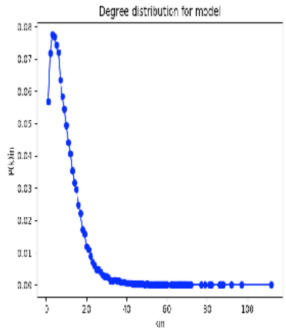
	Model	SF
numnodes	47540	47540
apl	4.6842357692793133	5.8616683143308901
Ldeg	112	1312.3191055305542
avgdeg	9.376861590239798	23769.5

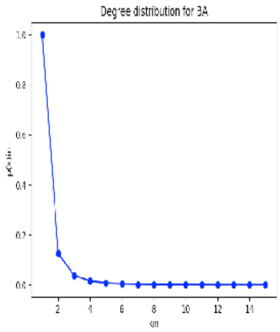





```
numnodes = Number of Nodes
apl = Average Path Length
acc = Average Clustering Coeff
```

	Network	BA
numnodes	47540	47540
apl	4.6842357692793133	4.5312062034700133
acc	0.11618233863494011	0.00030494951212812717





These buttons are an extension the above functions. It allows graphical comparison of degree distribution. As you can see from the above sample images, the input network's degree distribution follows that of a scale-free and barabasi-albert model instead of a random network.

Challenges

There were various challenges that surfaced while creating the program.

Iterative algorithms to calculate properties running through a large network will create a huge need for processing power. As a result, depending on the size of the network, the program might take a long time to run (1-3min). As the size of the network increases, it is unpractical to keep the program running.

Visualizing the network also requires a large processing power. To save time, only the network's largest hub will be visualized.

Another problem was that matplotlib, the python graph plotting library, was not fine-tuned and would not scale for large networks. As a result, the graph drawn would be unclear and stretched out. To mitigate this problem, only significant values are drawn. This allows for a much clearer graph which serves as a better comparison

Lastly, computing degree distribution for random networks would result in Overflow as the value would get too big. Hence, we only compute the first 5 values for comparison. For the graph, we use a static graph and provide the value of <k>.