CSC 591 - ADBI
Unity ID - djgadre
Assignment 4 - Network Properties in Spark GraphFrames

Note - I have included script for checking power law in file degree.py as a function named checkPowerLaw(). The function will be called when degree.py is run and it will generate a graph of counts against degree for analyzing whether degree distribution follows power law.

Degree Distribution

1. Are generated random graphs scale free?
   ● The 4 random graphs generated using networkx are not scale free. The csv's having their degree distribution is added in the submission.
   ● Also, the plot of counts against degrees does not follow the distribution of power law. The plots are added in the submission as gnm1_plot.png, gnm2_plot.png, gnp1_plot.png, gnp2_plot.csv

2. Are the stanford graphs scale free?
   ● The provided stanford graphs are scale free. The csv's having their degree distribution are added in the submission.
   ● Also, the plot of counts against degrees follows the distribution of power law. The plots are added in the submission as amazon.graph.small_plot.png, amazon.graph.large_plot.png, etc.

Centrality

1. Rank of nodes from highest to lowest closeness centrality

```
+---+-------------------+
| id|      closeness    |
+---+-------------------+
|  F| 0.07142857142857142|
|  C| 0.07142857142857142|
```

```
| H| 0.06666666666666667|
| D| 0.06666666666666667|
| B|0.058823529411764705|
| E|0.058823529411764705|
| G| 0.05555555555555555|
| A| 0.05555555555555555|
| I|0.047619047619047616|
| J|0.034482758620689655|
+---+-------------------+
```

2. Best Candidates to hold data?
   Machines C and F would be best candidates to hold data as they have highest closeness centrality with other nodes.

Articulation Points

1. Members to be targeted to best disrupt communication?

   Members to be targeted to disrupt communication are the articulation points:
   Mohamed Atta
   Usman Bandukra
   Mamoun Darkazanli
   Essid Sami Ben Khemais
   Djamal Beghal
   Nawaf Alhazmi
   Raed Hijazi