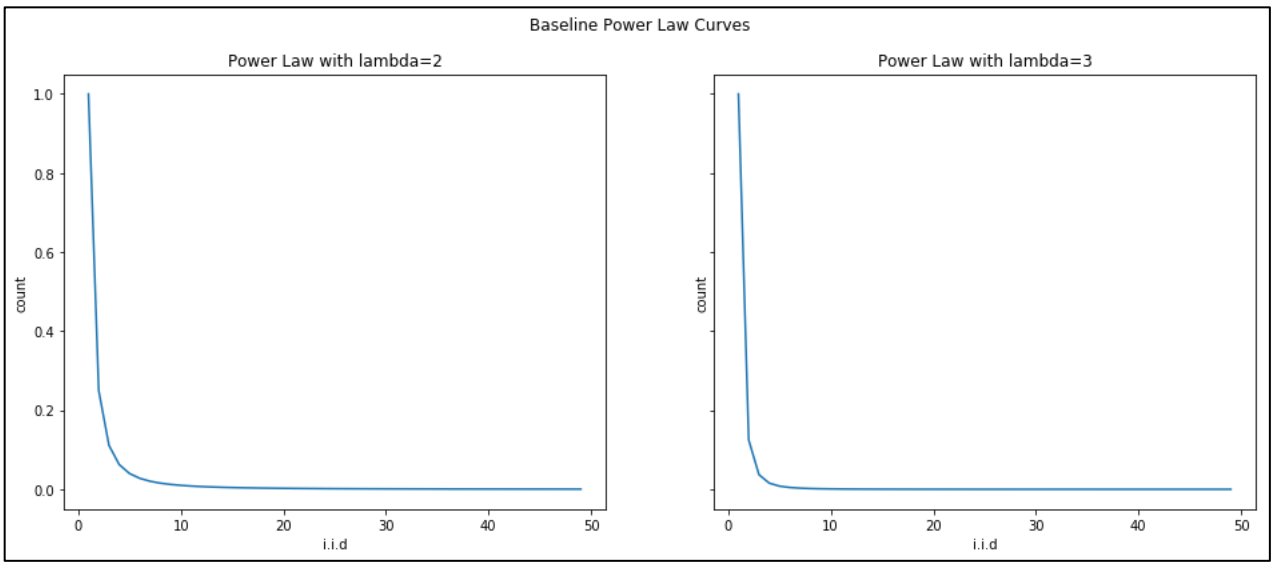
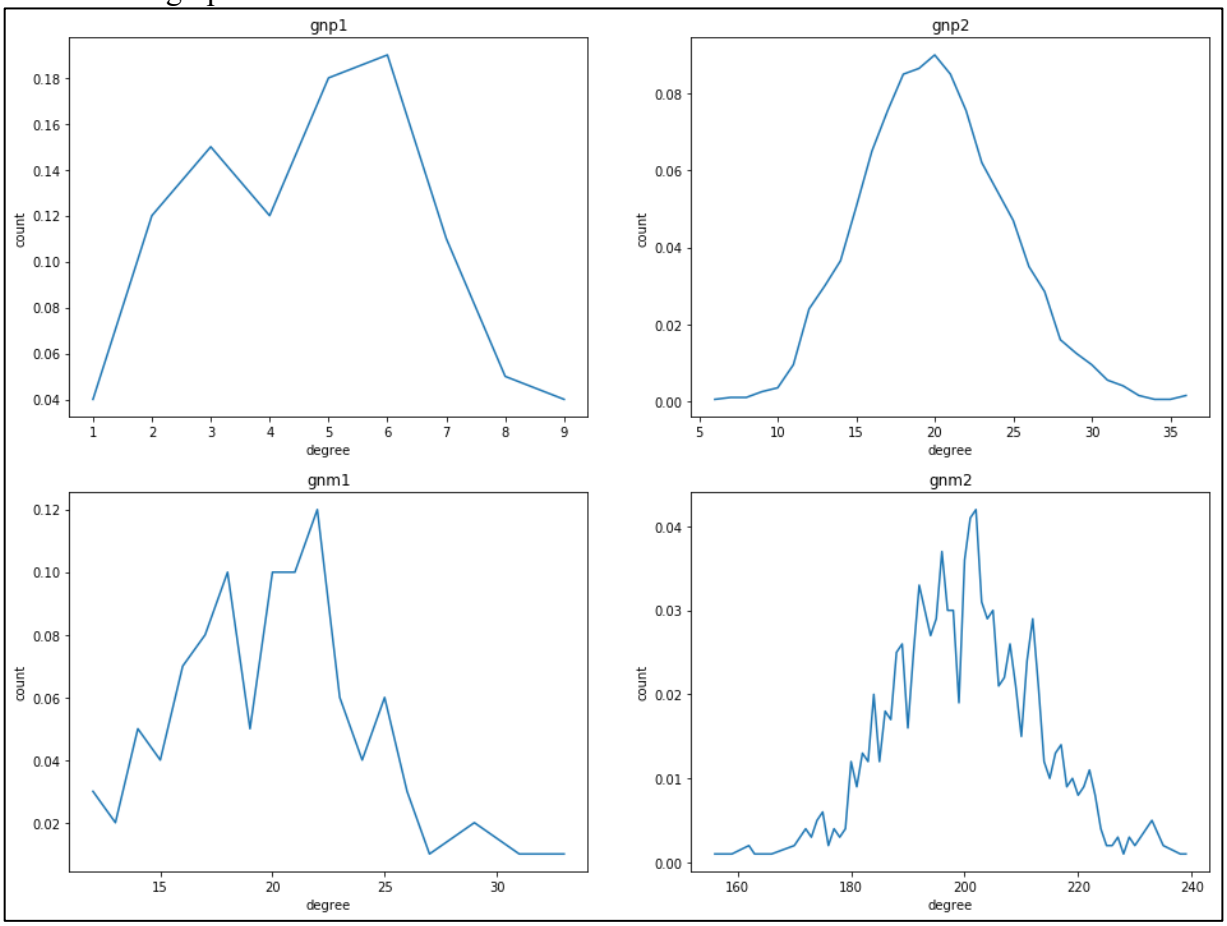
Network Properties with Apache Spark

Student ID – 200208218
Name – Aman Chauhan
Unity ID – achauha3

Degree Distribution

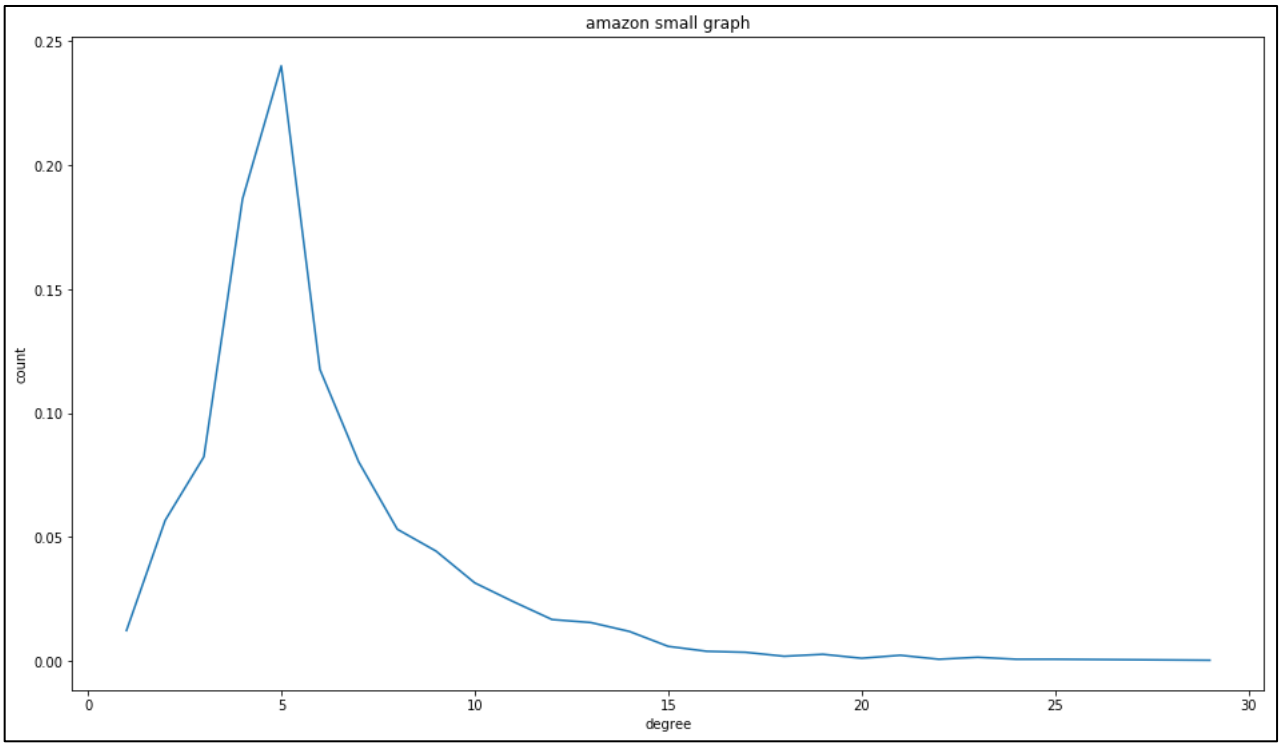


1. **Do the random graphs appear to be scale free?**
   Distribution of degrees for random graphs shows a Normal Distribution.
   Hence these graphs aren't scale free.



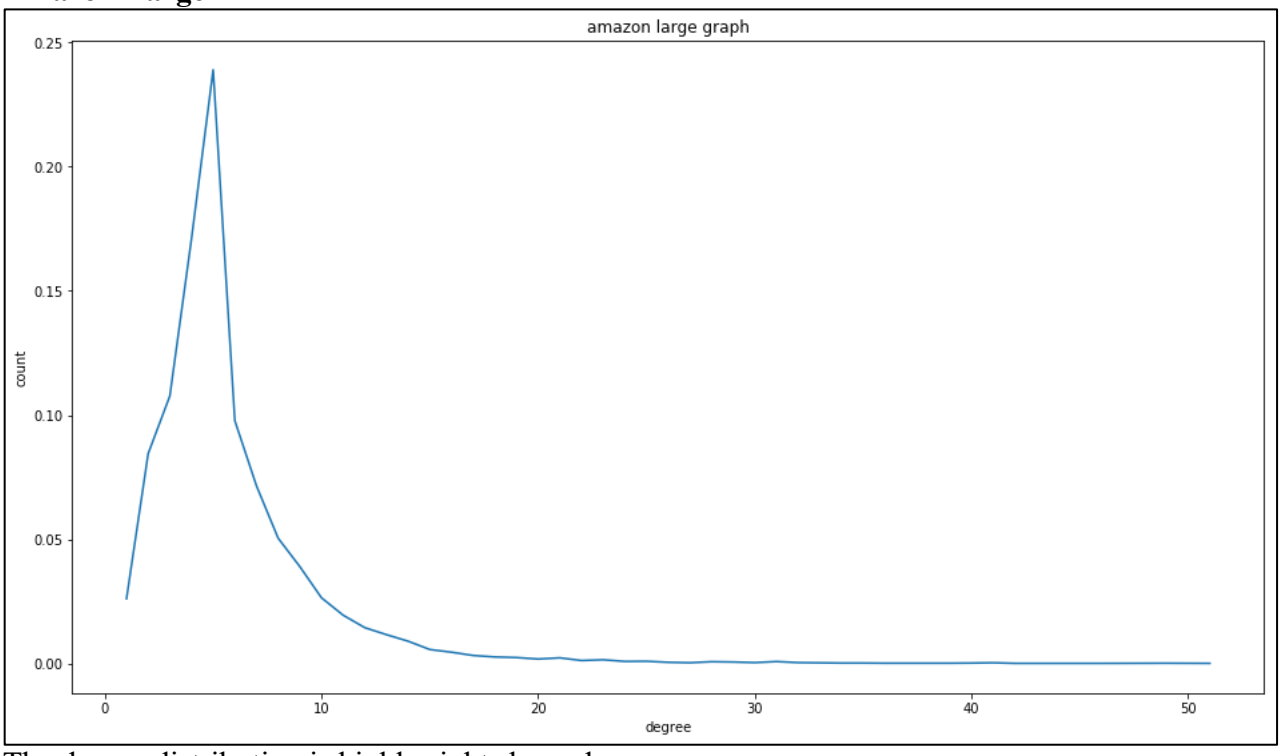
2. Do the Stanford graphs appear to be scale free?

Amazon Small



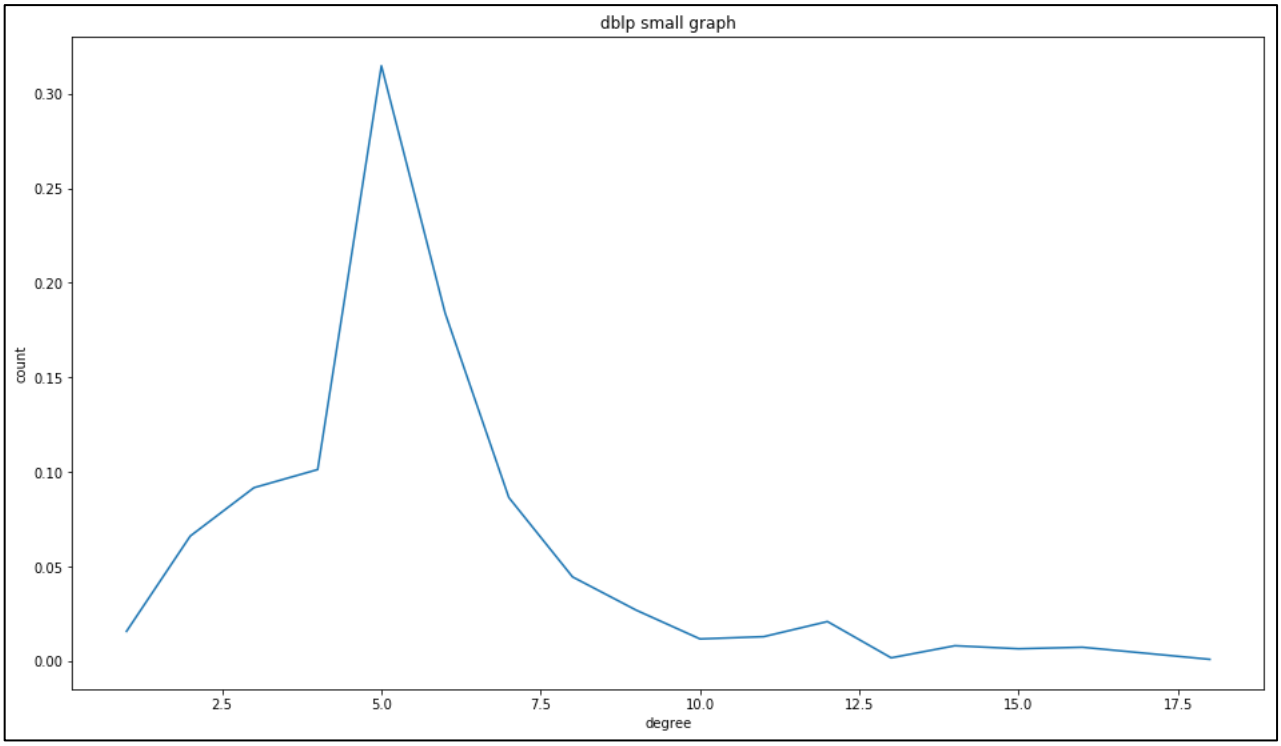
The degree distribution is highly right skewed.
We can say that after degree 5, the graph appears to be scale free as its distribution seems to follow power law.

Amazon Large



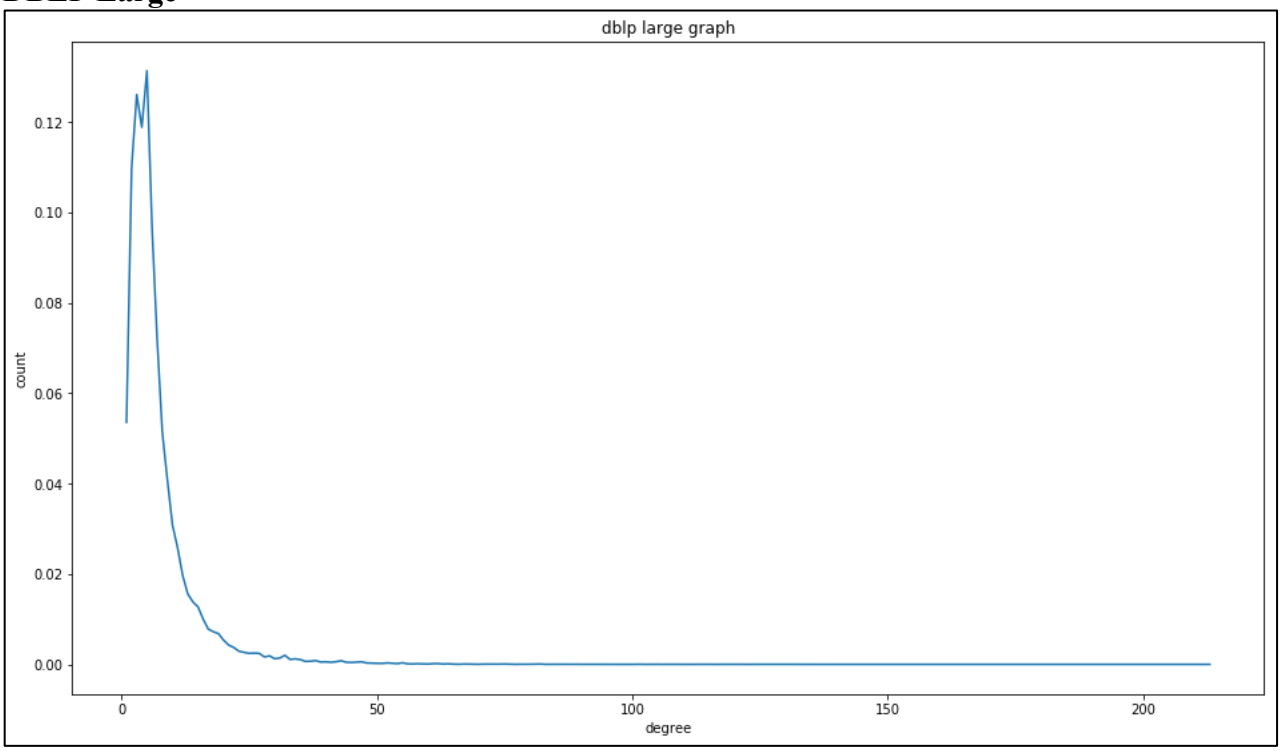
The degree distribution is highly right skewed.
We can say that after degree 5, the graph appears to be scale free as its distribution seems to follow power law.

DBLP Small



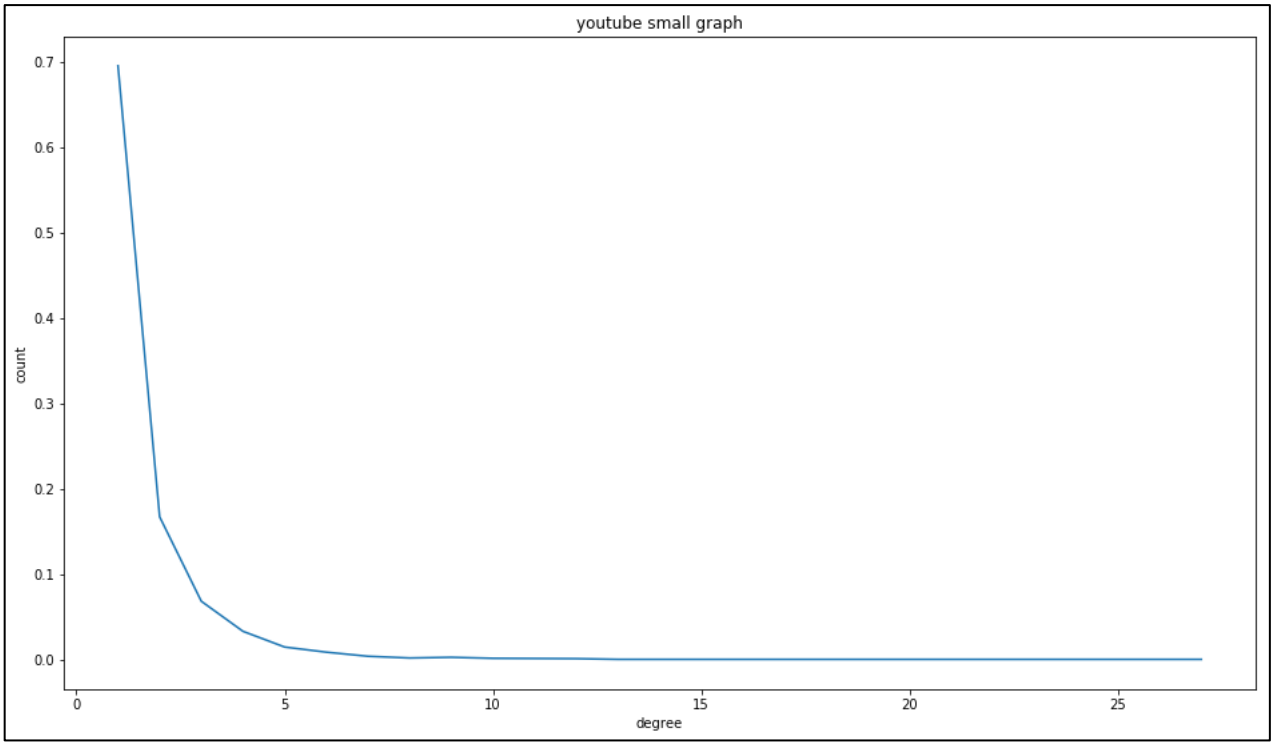
The degree distribution is right skewed.
We can say that the graph does not appear to be scale free as its distribution does not follow power law.

DBLP Large



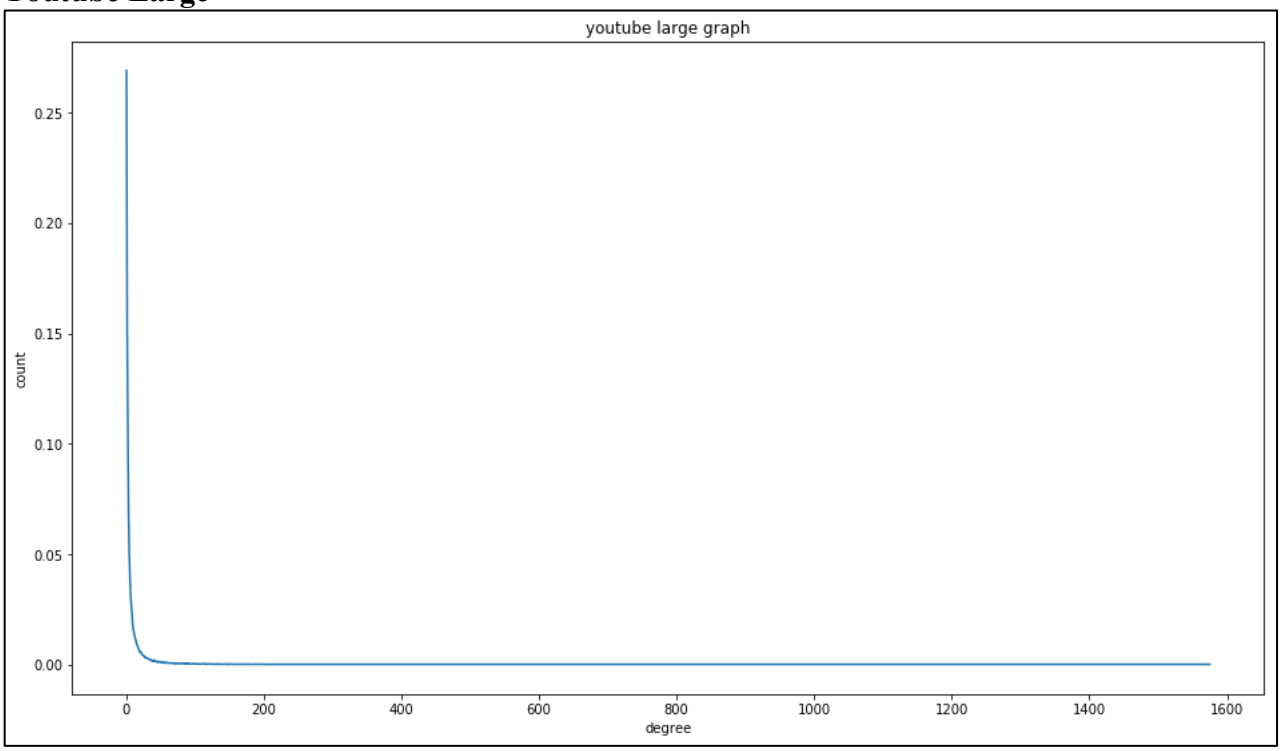
Apart from the first few degrees, this distribution clearly follows the power law.
Hence this graph is scale free.

Youtube Small



This distribution clearly follows the power law.
Hence this graph is scale free.

Youtube Large



This distribution clearly follows the power law.
Hence this graph is scale free.

Centrality

3. Rank the nodes from highest to lowest closeness centrality.

id	closeness
C	0.07142857
F	0.07142857
D	0.06666667
H	0.06666667
B	0.05882353
E	0.05882353
A	0.05555556
G	0.05555556
I	0.04761905
J	0.03448276

4. Suppose we had some centralized data that would sit on one machine but would be shared with all computers on the network. Which two machines would be the best candidates to hold this data based on other machines having few hops to access this data?
Machine C and F will be the best candidates, as they have the highest closeness centrality. This means the sum of shortest paths to these nodes from all other nodes is the least. Hence, the number of hops other machines would have to make to access this data will be the least.

Articulation Points

5. In this example, which members should have been targeted to best disrupt communication in the organization?

id	components
Mohamed Atta	5
Usman Bandukra	4
Mamoun Darkazanli	4
Essid Sami Ben Khemais	6
Djamal Beghal	6
Nawaf Alhazmi	4
Raed Hijazi	4

Essid Sami Ben Khemais, Djamal Beghal and Mohamed Atta created the greatest number of connected components when they were removed from the graph. This indicates that if these members of the organization are targeted, it will best disrupt the communication in the organization.