Project : Network Properties in spark

unity id : Svengal

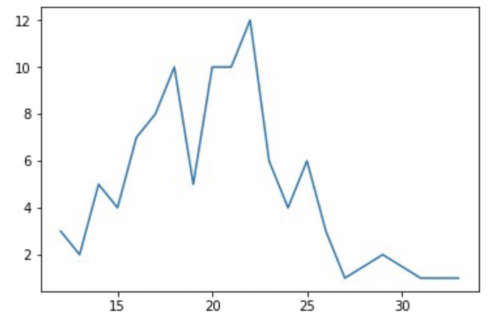
Degree Distribution:

1. Generate a few random graphs. You can do this using networkx's random graph generators. Do the random graphs you tested appear to be scale free?

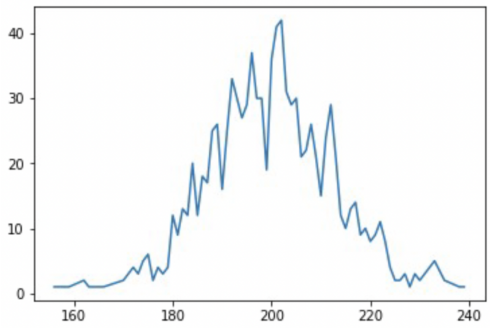
To find whether it is scale free or not we should find whether they follow power law or not. If log-log plot of the distribution is linear then they are scale free.

On running degree.py we get plot distribution for 4 random graphs.

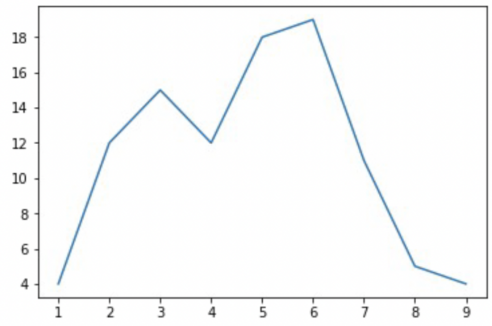
Gnm1.csv



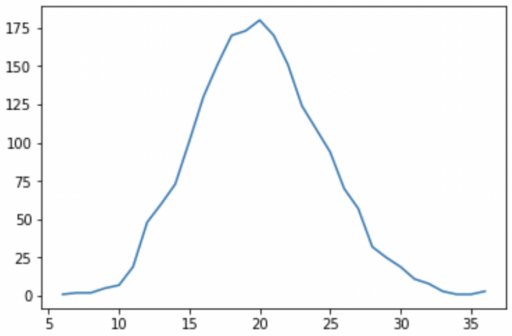
Gnm2.csv:



Gnp1.csv



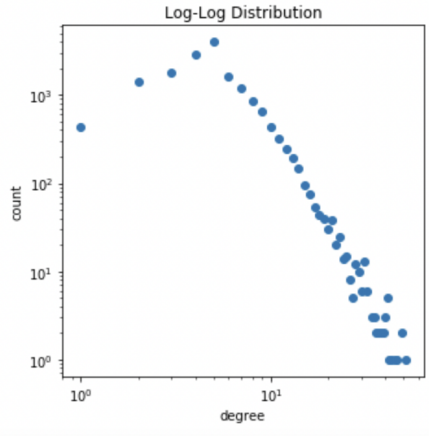
Gnp2.csv

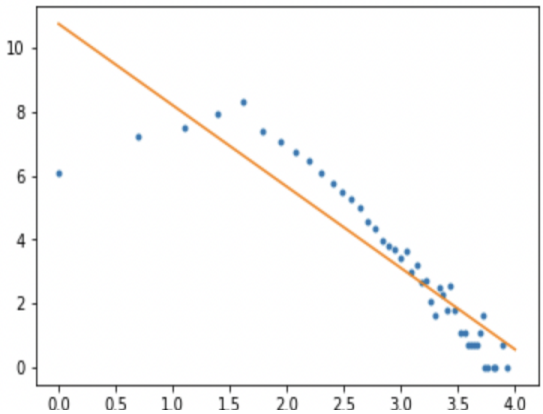


After generating the log-log distribution of these random graphs,they are not scale free because they don't follow power law.

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

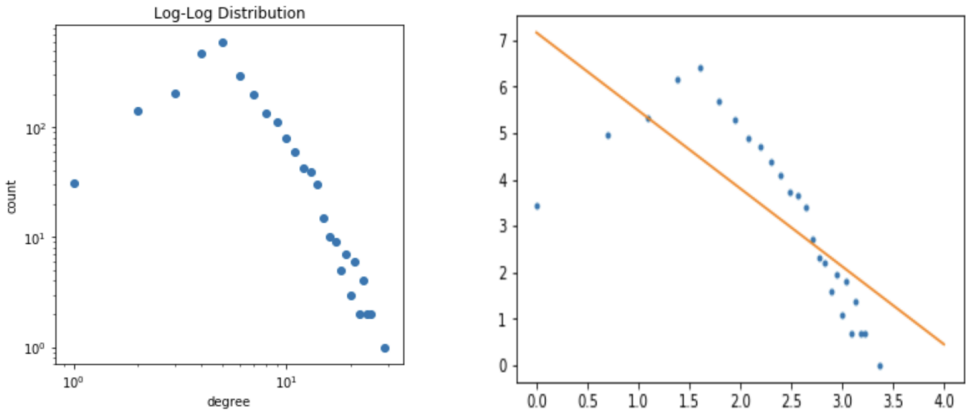
Amazon large :





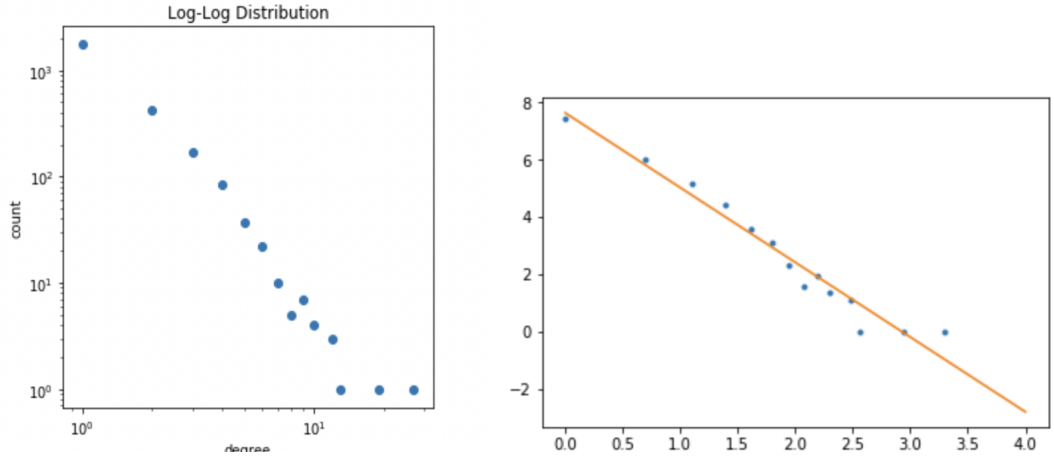
A straight line with a negative slope is constructed to estimate the log-log value of the above graph. Hence amazon large is scale free.

Amazon small :



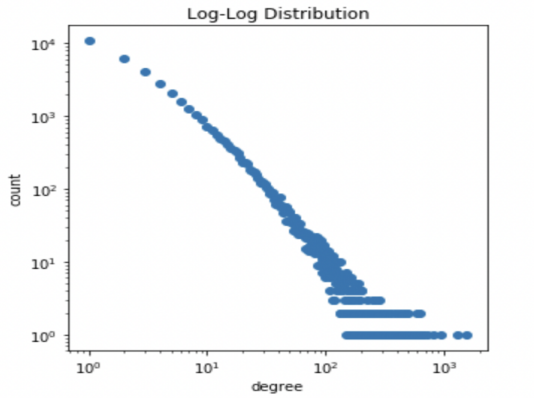
A straight line with a negative slope is constructed to estimate the log-log value of the above graph. Hence amazon small is scale free.

Youtube small:



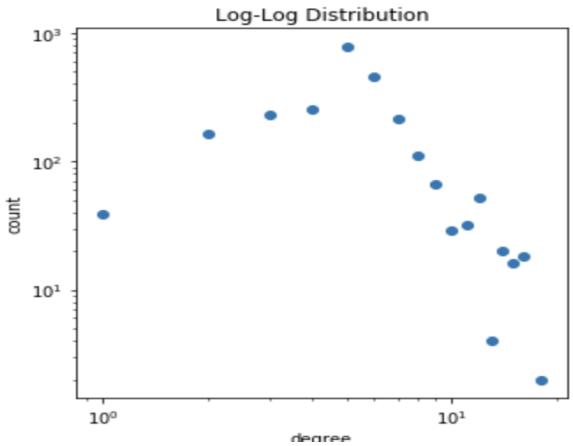
A straight line with a negative slope is constructed to estimate the log-log value of the above graph. Hence youtube short is scale free.

Youtube large:



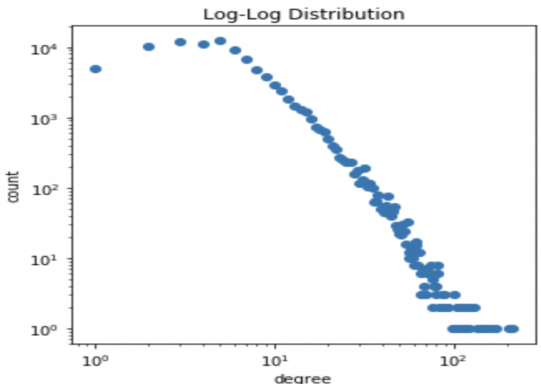
A straight line with a negative slope can be constructed to estimate the log-log value of the above graph. Hence youtube large is scale free.

Dblp short:



It is not scale free as log log distribution cn't not reasonably fit with a single line.

Dblp large:



It can be estimated with a straight line, so it is scale free.

Centrality:

1. **Rank the nodes from highest to lowest closeness centrality.**

   In the centrality.py we find the shortest path of a vertex to all of its neighbors and we find the inverse of the sum of all distances to its neighbors which gives us the closeness of a vertex. We can say that, more the closer neighbors more the closeness.

```
+---+-------------------+
|key|           closeness|
+---+-------------------+
|  F| 0.07142857142857142|
|  C| 0.07142857142857142|
|  H| 0.06666666666666667|
|  D| 0.06666666666666667|
|  B|0.05882352941176471|
|  E|0.05882352941176471|
|  G| 0.05555555555555555|
|  A| 0.05555555555555555|
|  I|0.047619047619047616|
|  J|0.034482758620689655|
+---+-------------------+
```

2. **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?**

   According to the ranking of closeness, C and F are the best vertices to act as centralized data.

Articulation points:

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

Upon their removal of these following terrorists from the network, the connected components in the graph increases and this is used to disrupt communication in the organization.

id	articulation
Mohamed Atta	1
Usman Bandukra	1
Mamoun Darkazanli	1
Essid Sami Ben Khemais	1
Djamal Beghal	1
Nawaf Alhazmi	1
Raed Hijazi	1