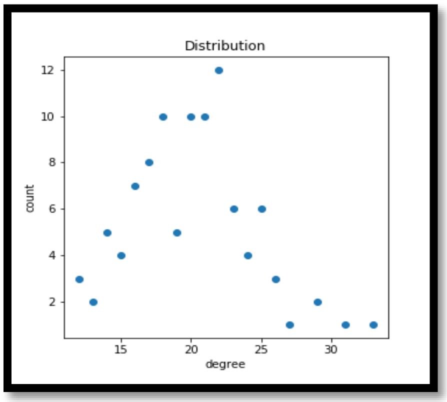
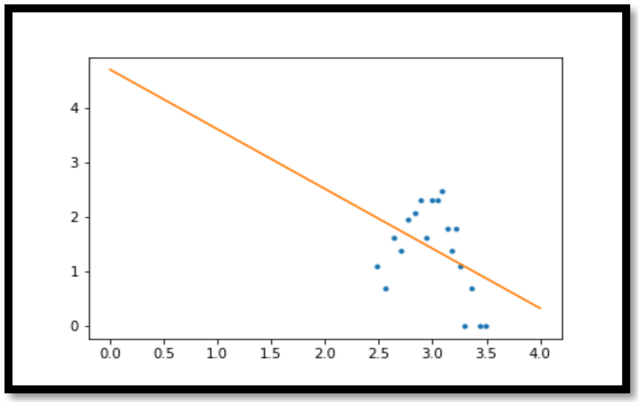
Network Properties in Spark

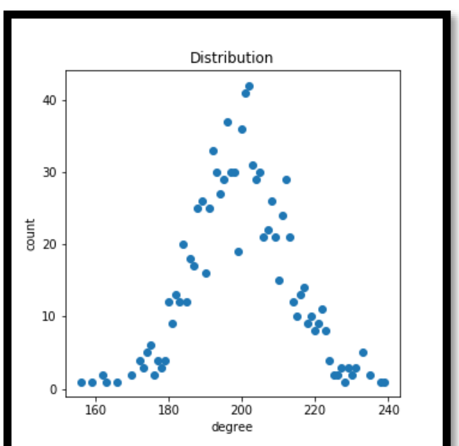
Problem 1: Degree Distribution

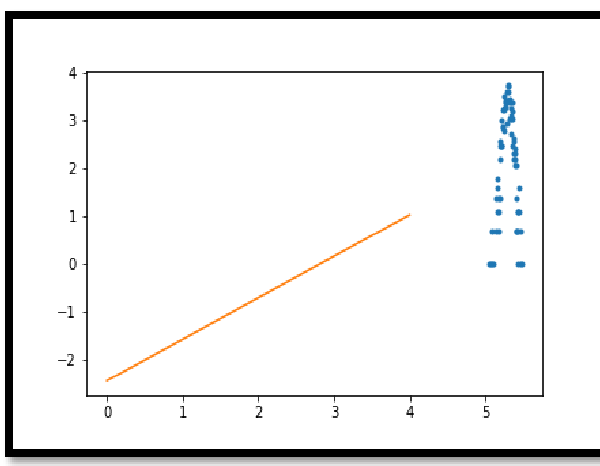
1. Generate a few random graphs. You can do this using networkx's random graph generators. Do the random graphs you tested appear to be scale free? (Include degree distribution with your answer).
   The graphs are fitted to log distribution and their slopes are as follows:
   Graph 1 : GNM1
   Coefficients : [-1.09361927 4.69994133]
   Residuals : [9.9345787]
   The log distribution cannot be fitted with a straight line. Therefore, distribution doesn't follow power law.





   Graph 2 : GNM2
   Coefficients : [0.86460813 -2.42745382]
   Residuals : [93.11191765]
   The log distribution cannot be fitted with a straight line. Therefore, distribution doesn't follow power law.



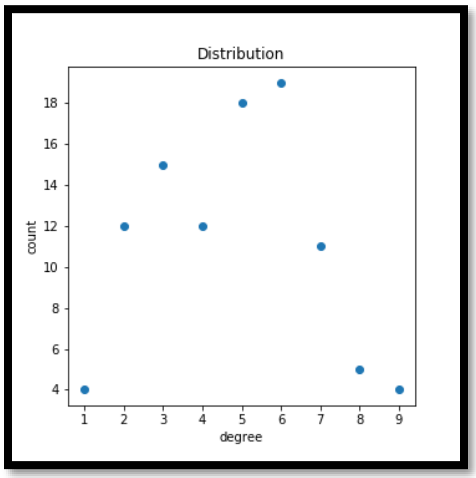


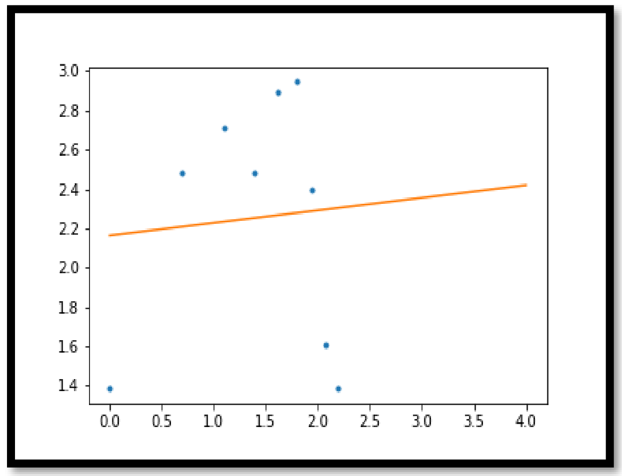
Graph 3 : **GNP1**

Coefficients : [0.06379652 2.16398712]

Residuals : [3.11961529]

The log distribution cannot be fitted with a straight line. Therefore, distribution doesn't follow power law.



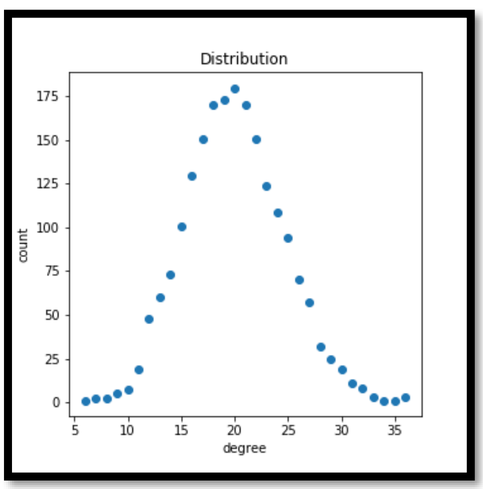


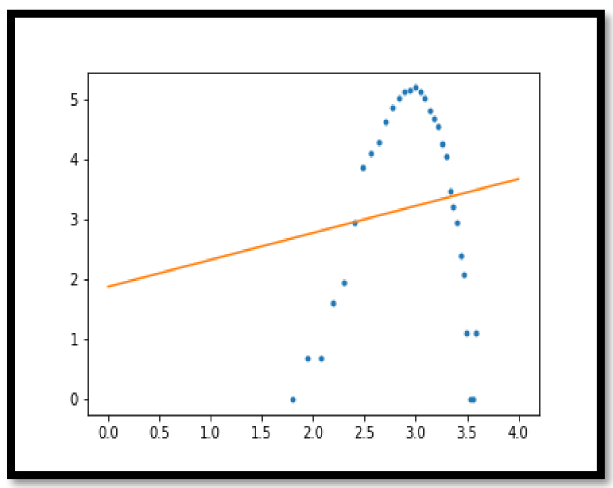
Graph 4 : **GNP2**

Coefficients : [0.44868707 1.87506857]

Residuals : [94.32739034]

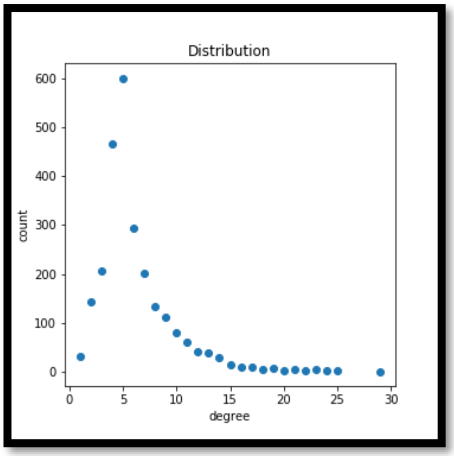
The log distribution cannot be fitted with a straight line. Therefore, distribution doesn't follow power law.

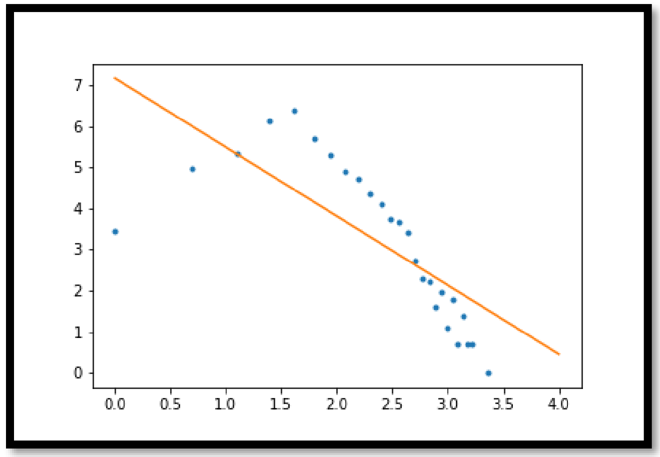




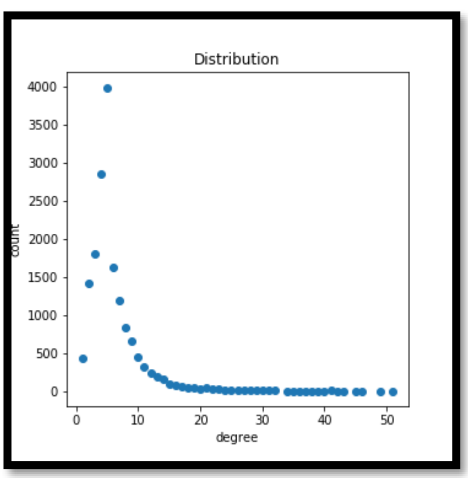
All the randomly generated graphs are not scale-free.

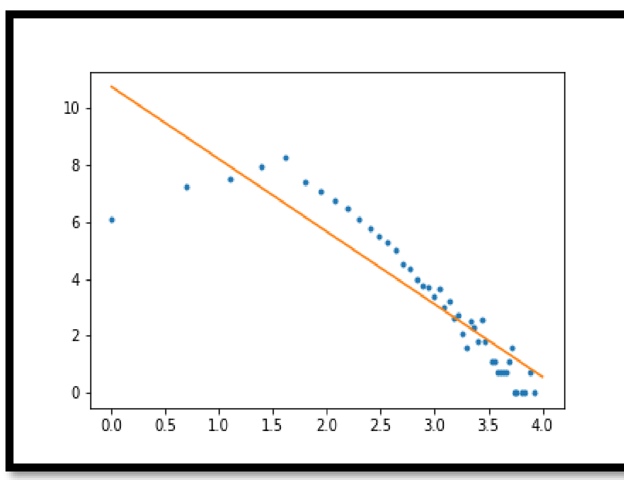
2. Do the Stanford graphs provided to you appear to be scale free?
   Graph 1 : Amazon Small
   Coefficients : [-1.67835309 7.16281337]
   Residuals : [39.90749226]
   The log distribution can be fitted with a straight line. Therefore, distribution does follow power law.



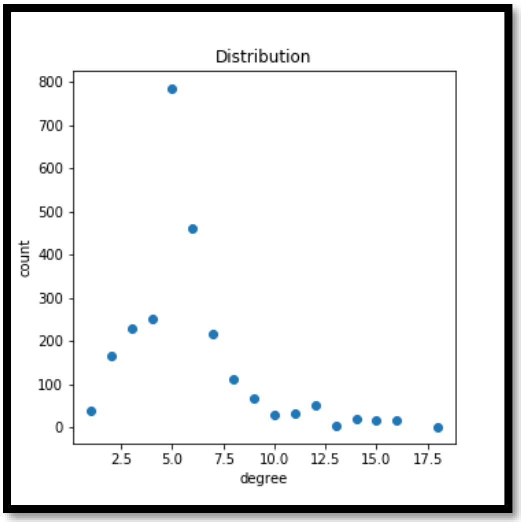


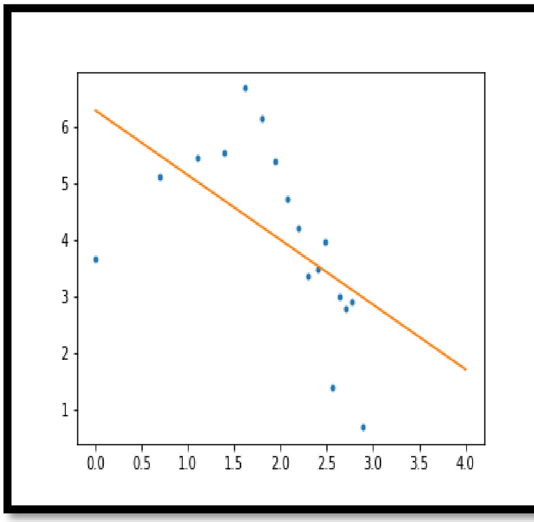
   Graph 2 : Amazon Large
   Coefficients : [-2.54809941 10.75019659]
   Residuals : [53.42809018]
   Similar to amazon small
   The log distribution can be fitted with a straight line. Therefore, distribution does follow power law.



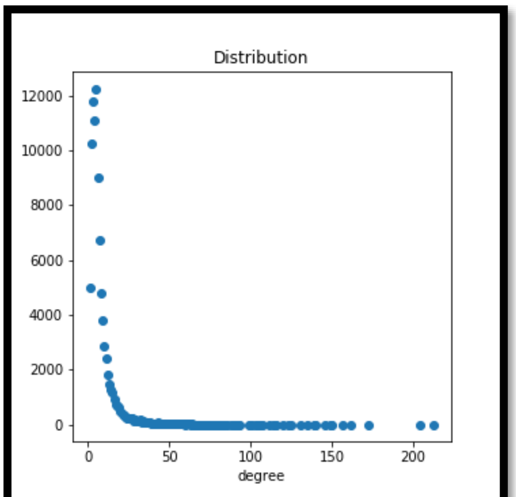


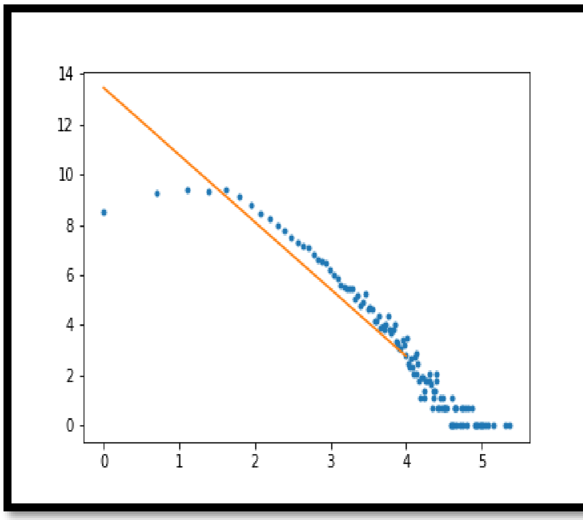
Graph 1 : **DBLP Small**
Coefficients : [-1.14109288 6.27327106]
Residuals : [28.71080164]
The log distribution can be fitted with a straight line. Therefore, distribution kindof does follow power law. It is scattered and has more noise because it tries to fit the skew.



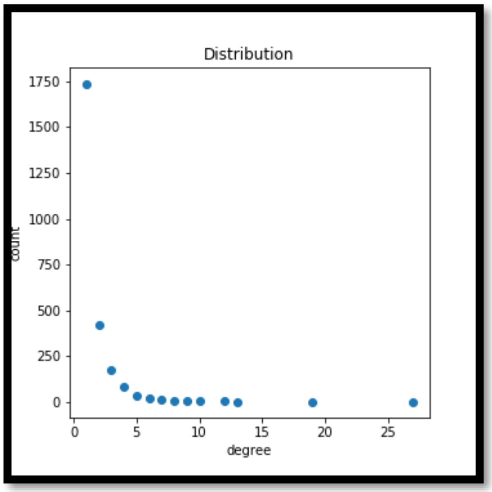


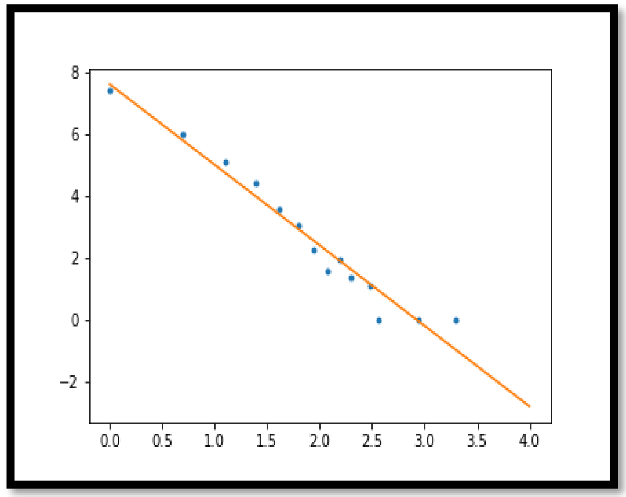
Graph 2 : **DBLP Large**
Coefficients : [-2.67185045 13.44832764]
Residuals : [69.92866355]
The log distribution can be fitted with a straight line. Therefore, distribution does follow power law. It looks like an ideal power distribution.



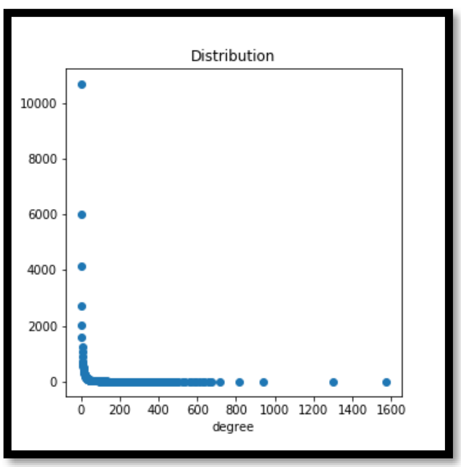


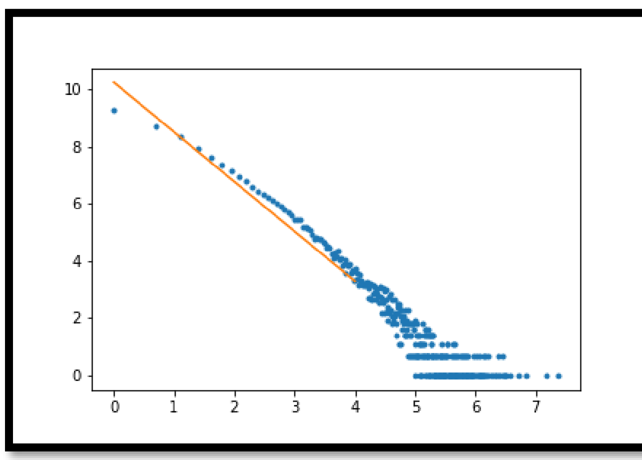
Graph 1 : **Youtube** Small
Coefficients : [-2.60472814 7.63227776]
Residuals : [2.72996778]
The log distribution can be fitted with a straight line. Therefore, distribution does follow power law. It just has less datapoints.





Graph 2 : **Youtube** Large
Coefficients : [-1.73318997 10.22937638]
Residuals : [115.5071705]
The log distribution can be fitted with a straight line. Therefore, distribution does follow power law. It looks like an ideal power distribution.





The Stanford graphs provided appear to be scale free.

Problem 2: Centrality

1. Rank the nodes from highest to lowest closeness centrality

id	closeness
C	0.071429
F	0.071429
D	0.066667
H	0.066667
B	0.058824
E	0.058824
A	0.055556
G	0.055556
I	0.047619
J	0.034483

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.

   C and F would be the best candidates to hold this data on the machines having few hops to the access this data.

Problem 3 : Articulation Points

1. In this example, which members should have been targeted to best disrupt communication in the organization?
   From the initial connected components of 3, the following the following members increase the connected components therefore the following members should be targeted to disrupt communication in the organization.

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