Production and Network Measurements

Table of Contents

Figures

1 Introduction

This project aims to study the results of the metrics of two existing networks. Then, we'll create new networks based on some specific properties and also comment on metric and graph results.

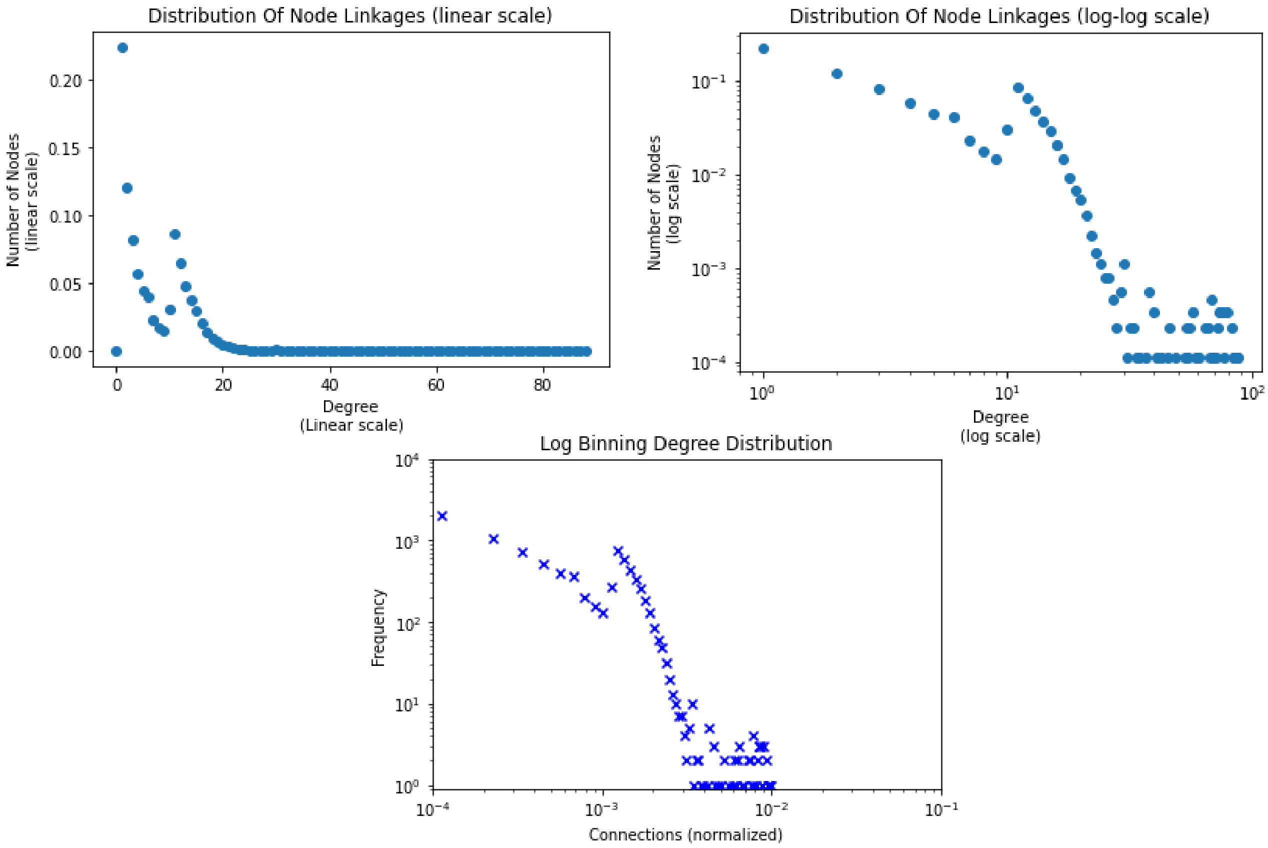
2 Commenting on the Results

2.1 Gnutella p2p network

Gnutella is a peer-to-peer network, with 8,846 nodes and 31,839 edges. The fact that there are no **self-loops of** nodes is obvious, because a directed network like this, nodes represent hosts in the network that communicate with each other, and therefore there could be no self-loops on identical computer nodes. The **reciprocity** is 0, which suggests that there are no reciprocal hosts between them. The **sink** nodes are 4,996 and are those that only receive data to the other hosts in the p2p network, while 118 **source** hosts only send data to the others. The fact that the **Average Degree** is 7.19 means that on average each node is connected to 7.19 nodes, which suggests a relatively sparse network with many nodes not having multiple connections. The **diameter** was calculated based on the longest shortest path and was 22. This could perhaps be considered large if the network represented a social or collaboration network. In this case, however, I believe that for a network representing real computers, which may be quite far away from each other, a diameter of 22 is relatively small. The value of 0.0036 for the **Average Clustering Coefficient** indicates that the neighbors are not particularly connected to each other, which does not enhance the robustness of the network. As for the **Global Clustering Coefficient**, its value is 0.0037 signaling a sparse network and not highly clustered. The fact that the size of the largest **Strongly Connected Component** of the network is 13,453 means that there is a group of nodes in the network that are highly connected to each other and can reach each other quickly through a path of directed edges. Similarly, a value of the largest **Weakly Connected Component of** 31,837 in a directed network indicates that there may be nodes with only incoming or outgoing edges, leading to disconnected groups of nodes.

In general, the Gnutella network, although directed, does not represent any particular relationship between its nodes, but rather represents hosts that are connected to each other (with some randomness) to share some files.

Figure 2.1: Distribution of the degree of Gnutella p2p Network nodes

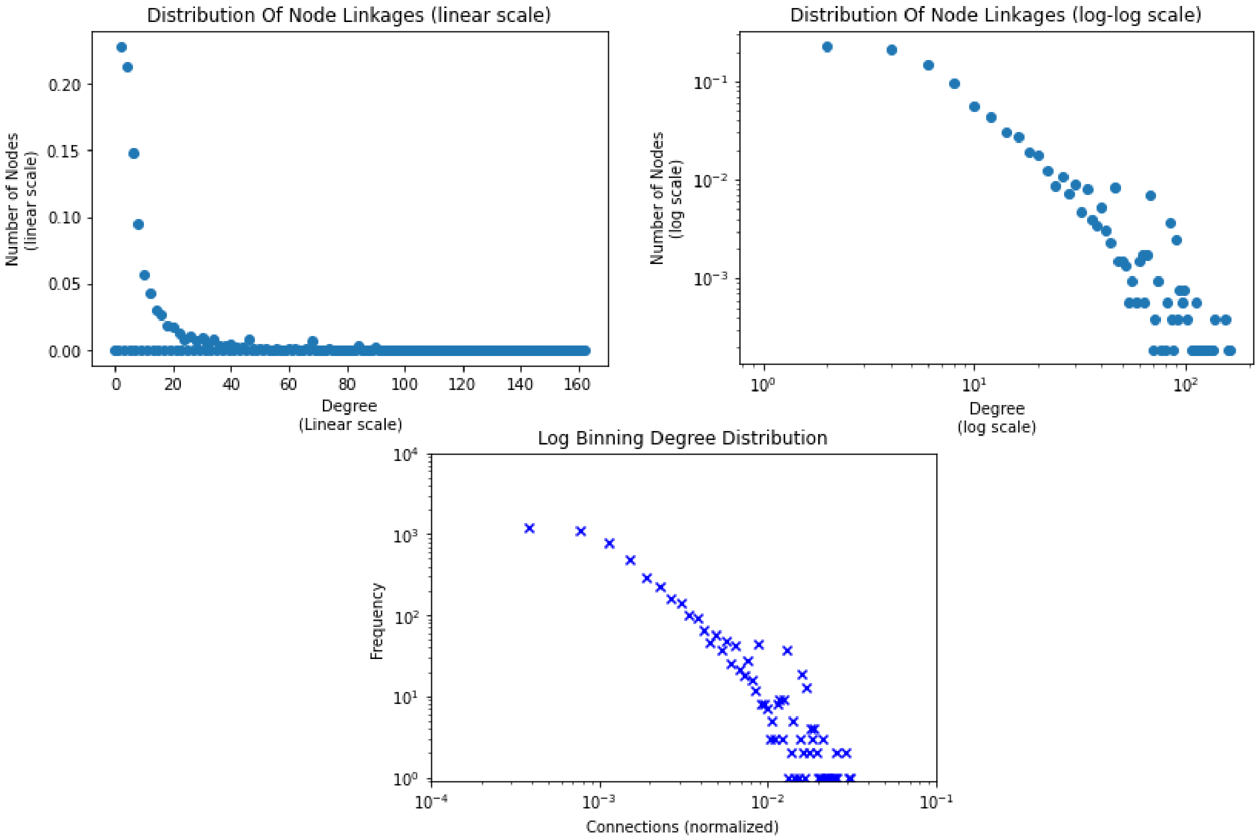


Annotation: In the above graph it can be seen that there are a very large number of nodes which have a low degree, while a few nodes have a high degree. This is means that the nodes follow the scale-free distribution, i.e. they follow the power-law rule. As for the graph with the log-log scale grade distribution, all the points are not exactly along a line but the tail looks like a straight line which suggests that the grade distributions follow the power-law distribution.

2.2 GR-QC network

GR-QC is a network containing collaborations between scientific authors, with 5,242 nodes and 28,980 edges. Unlike the previous network, here the metrics analysis proves that there are many mutual edges (28,968), and the fact that there are no sink and source nodes indicates that every author in the network has collaborated with at least one other author. This is an expected feature for an author collaboration network, because each author usually collaborates with others to publish papers. The high value of Clustering Coefficients suggests that authors in the field of General Relativity and Quantum Cosmology tend to collaborate with other authors who also collaborate with each other.

Figure 2.2: Distribution of the GR-QC Network node rank



Comment**:** as in the previous network, that nodes seem to follow the power-law distribution. In the graph with the log-log scale point distribution, there may be some noise, but it also looks like a straight line.

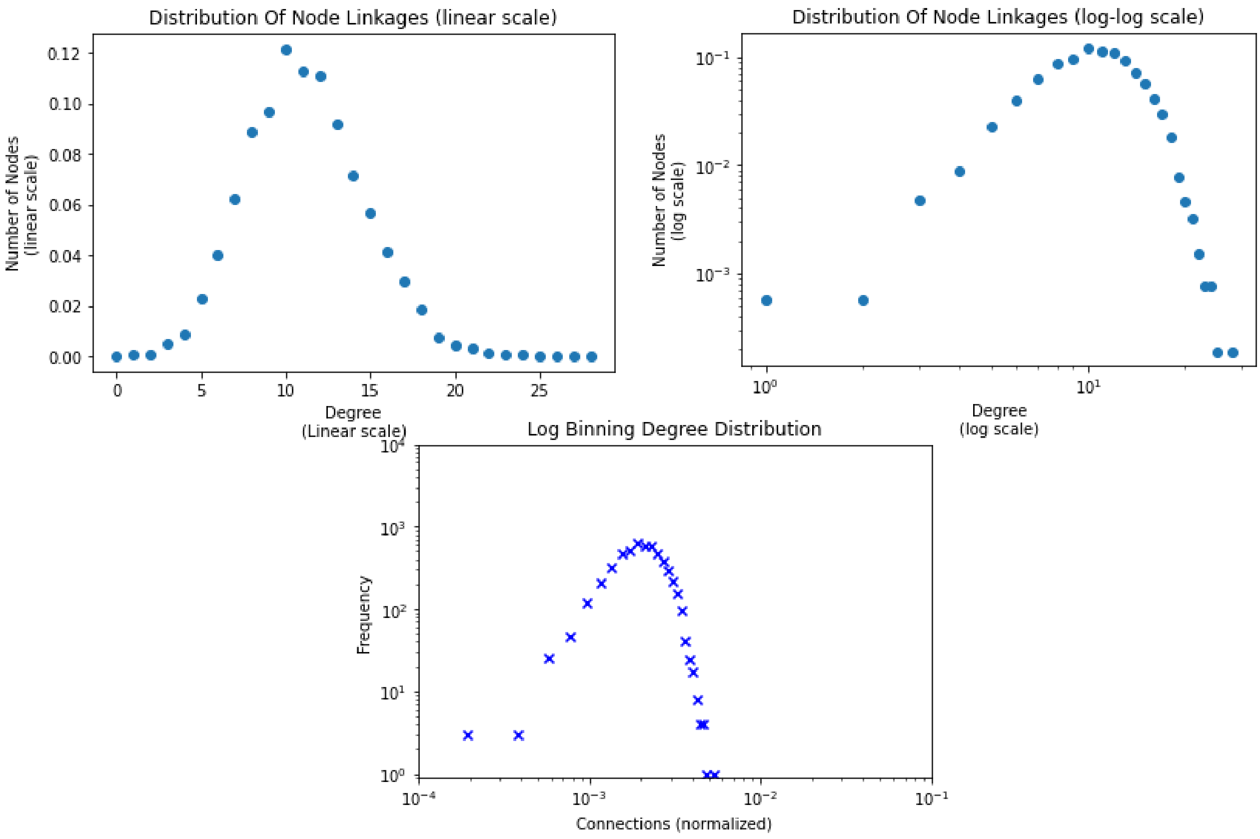
2.3 Production of Synthetic Networks

At this point it should be noted that for the next four synthetic networks the number of nodes was fixed at 5200. Still, it was requested to choose such parameters during their generation that they have properties as close as possible to 2° . The choice of parameters was made after testing the values of the p, k and m parameters present in some of them. Finally, a fixed seed was set so that the number of edges was always constant so that the same metrics could be easily recalculated.

2.4 Erdős-Renyi Network

To generate this undirected network, a value of 0.002174 (edge generation probability) was chosen for p. This results in a network of 5,200 nodes and 28,984 edges, which is close enough to what is needed. In the case of a random Erdos-Renyi network, nodes are connected randomly, meaning that there is no specific community structure. However, the small diameter suggests that the network is likely to be efficient in terms of information transmission.

Figure 2.3: Distribution of the degree of Erdős-Renyi Network nodes

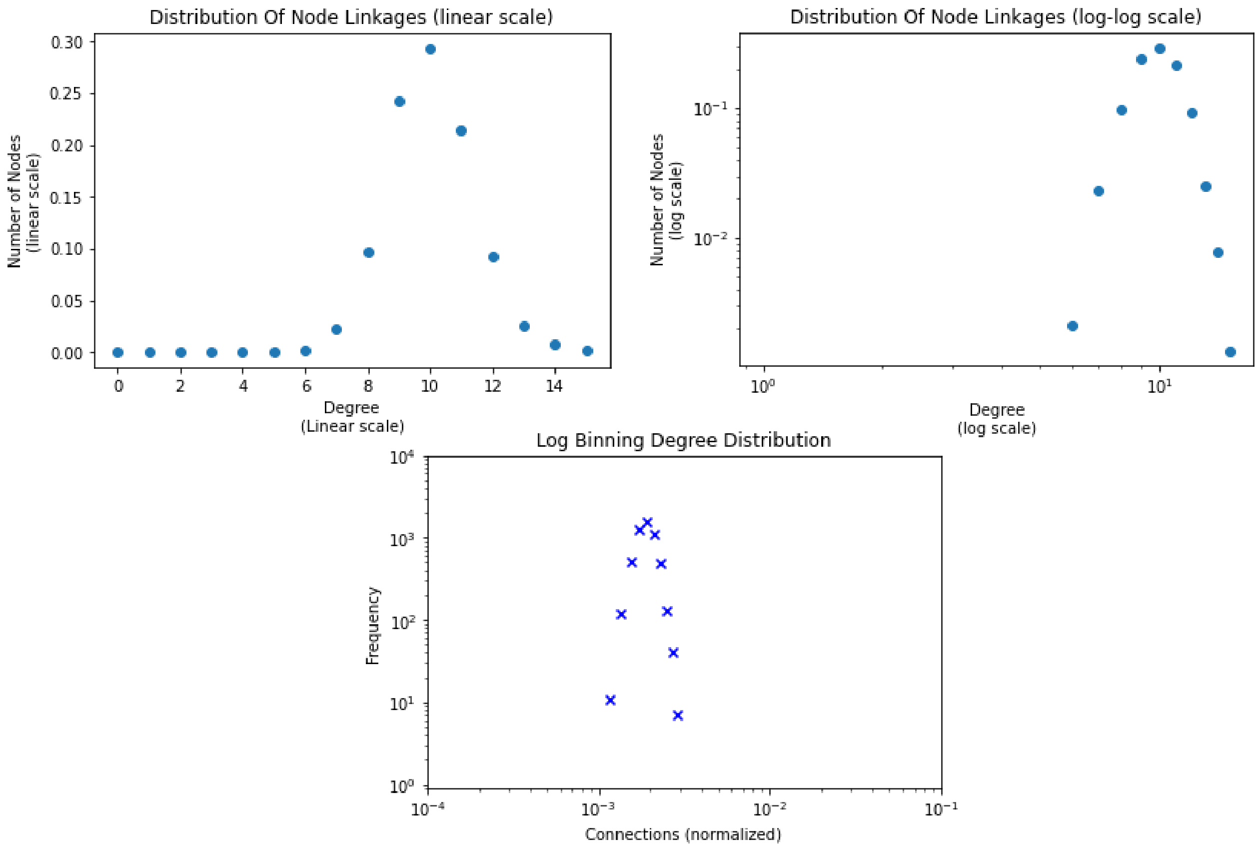


***Comment*:** The above graphs show that this network follows the Poisson distribution, since most nodes are concentrated in one point, representing their degree. As evidenced by the metrics results, the clustering is low and there are no central nodes in the network.

2.5 Watts & Strogatz Network

To generate this undirected network, values of 0.2 and 11 were chosen for p and k respectively. This results in a network of 5,200 nodes and 26,000 edges. A first look at the metric results for this network we can see that the authors were right that with this random network there is significant clustering and transitivity, while enhancing the small-world effect.

Figure 2.4: Distribution of the degree of nodes in the Watts & Strogatz Network

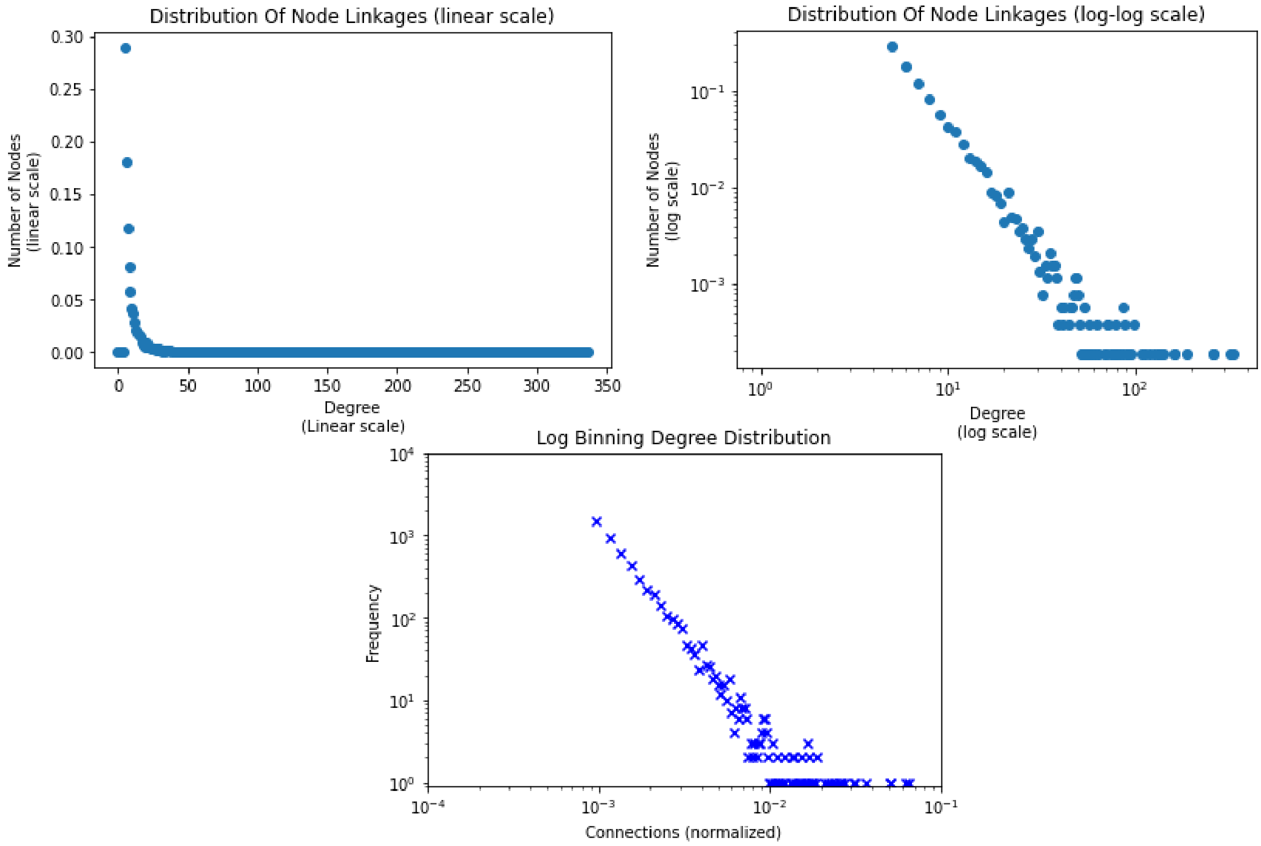


Comment**:** Just like the previous network, the graphs show that this network follows the Poisson distribution, since the largest number of nodes is concentrated in one point.

2.6 Barabasi-Albert network (Undirected)

To generate this undirected network, a value of 5 was chosen for m. This results in a network of 5,200 nodes and 25,975 edges. The truth is that the metric results are not close to the 2º network, but as shown below the node grades follow the power-low distribution.

Figure 2.5: Distribution of the degree of the Barabasi-Albert Undirected Network

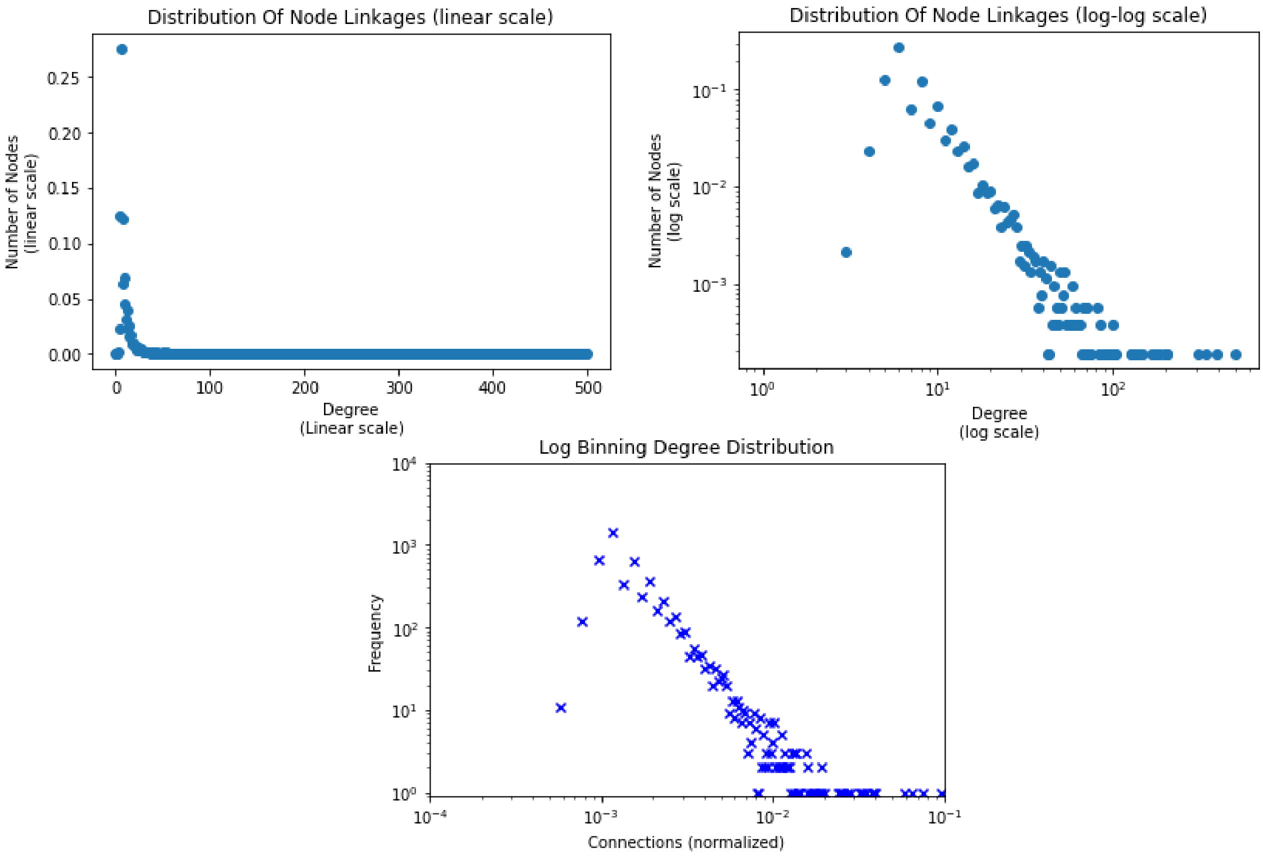


***Comment*:** Figure 2.5 clearly shows that the grades of the nodes in this network follow the power-law distribution.

2.7 Barabasi-Albert Network (Directed)

For the generation of this undirected network, a value of m = 3 was chosen and it is different from the previous network because we then convert it to a directed network and therefore the number of reciprocated edges is increased. This results in a network of 5,200 nodes and 28,980 edges. As shown in the table in the excel file, the metrics results of this generated network are the closest to 2^{o} compared to the others.

Figure 2.6: Distribution of the degree of the Barabasi-Albert Directed Network



***Comment*:** Figure 2.6 shows that the network node grades follow the power-law distribution, with some data mainly in the tail being noise. For better visualization of the graph, we could use larger bins at higher degrees.