Christopher Barclay
48254426 - crbarcla
CS 165
6/7/19

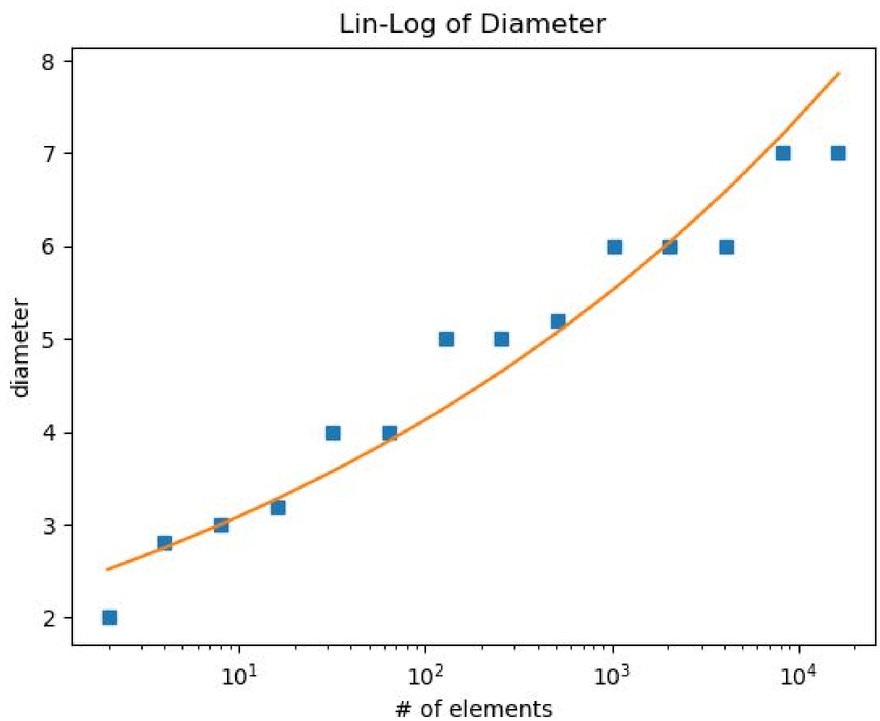
Project 3 Graph Algorithms

Introduction

This project focused on implementing graph algorithms. The project covers graph creation, specifically Barabasi-Albert (BA), and graph analysis. These graphs can be used to represent many things in the real world. The BA model generates random scale-free networks which is used in a lot of social media related solutions. The analysis of interest are graph diameter, clustering coefficient, and degree distribution. Clustering coefficient and diameter have been analysed for generated graphs of size n {2, 4, 8, 16, 32, …, 8192}. Each n was averaged over 5 reps. Degree distributions were taken on generated graphs of size n {1000, 10000, 100000} only once without averaging. The algorithms' analysis is represented with lin-log, lin-lin, and log-log plots. Clustering coefficient and diameter on lin-log, and degree distributions on lin-lin and log-log. The language used for this project was C++, specifically the 2011 standard.

Graph Diameter

The graph diameter is the length of the longest path in the graph. It can be used to analyse the graph and how it is performing. To determine the diameter of the graph, you first find the shortest path between all pairs of vertices and select the distance of the longest one. Graph diameter as a function of n is plotted below on a semi log plot (lin-log).



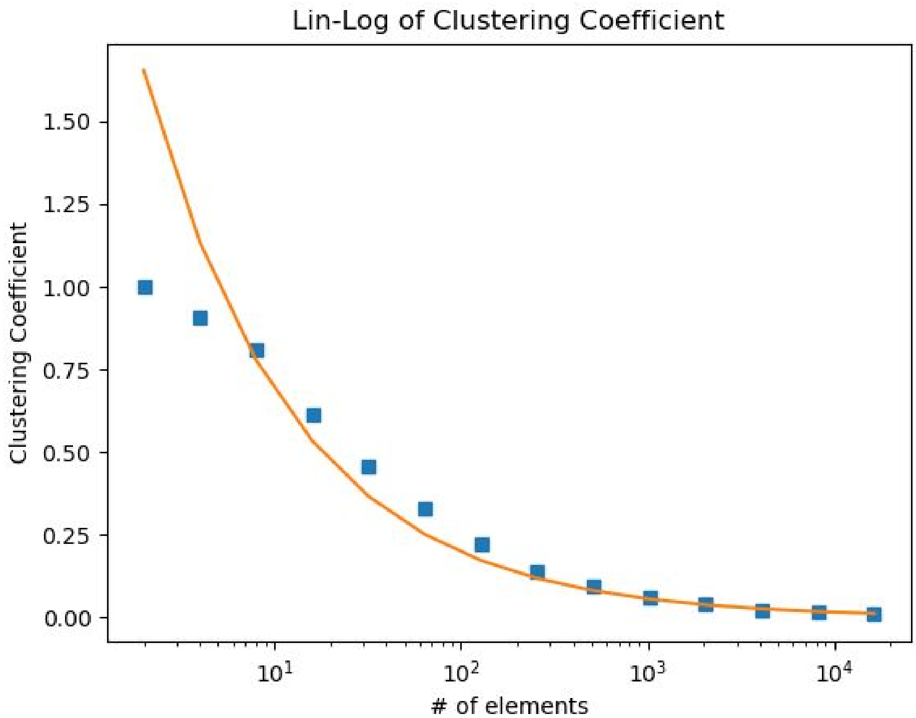
Diameter

n	2	4	8	16	32	64	128	256	512	1024	2048	4096	8192	16384
D	2	2.8	3	3.2	4	4	5	5	5.2	6	6	6	7	7

The diameter increases as a function of n. It grows proportional to n.

Clustering-Coefficient

The clustering-coefficient determines the relationship of vertices and how they form triplets, open or closed. It can be used to analyse the graph and how it forms clusters in triangles. To determine the clustering-coefficient of the graph, you take the ratio of closed triplets and all triplets. Another formula is three times the number of triangles divided by the number of length 2 paths. Graph clustering coefficient as a function of n is plotted below on a semi log plot (lin-log).



Diameter

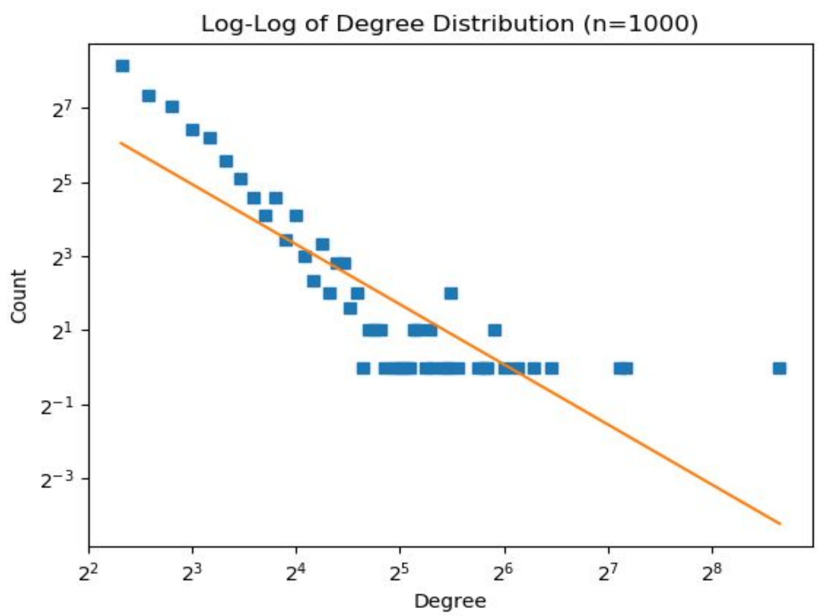
n	2	4	8	16	32	64	128	256	512	1024	2048	4096	8192	16384
D	1	0.9	0.81	0.615	0.457	0.327	0.22	0.14	0.09	0.058	0.038	0.023	0.015	0.009

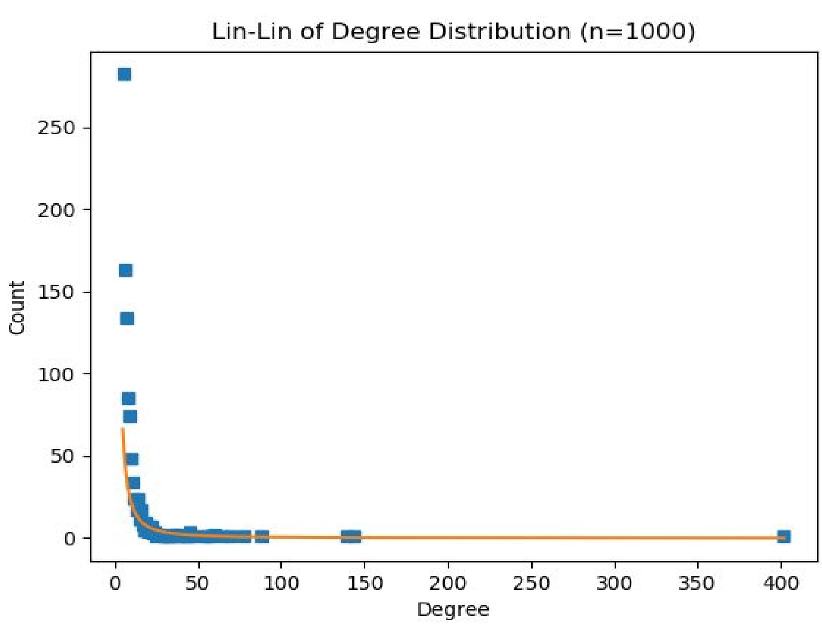
The clustering coefficient decreases as a function of n. It grows logarithmically to n.

Degree Distribution

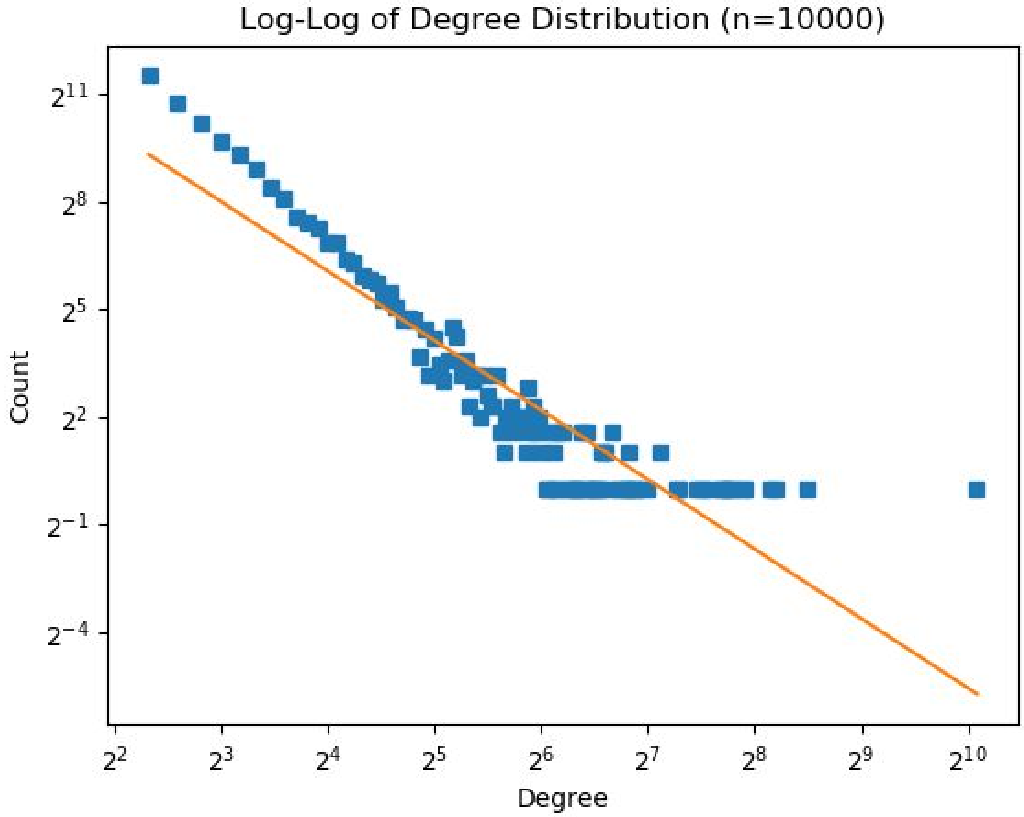
Degree distribution determines the number of vertices for each degree in the graph. The degree of a node is the number of edges connecting it to other vertices. This is very useful for analyzing networks and finding strong and weak connections. Most real world networks have low degrees for most nodes and high degree for fewer nodes sometimes called “hubs”. In order to find the degree distribution you map vertice edge counts to a map with key degree, and value vertice count or probability depending. The linear graphs seem to provide more useful information. Both the log graphs and the linear graphs show the count decreasing as a function of n. The linear graphs provide a better picture and show there is no power law and that is decreases logarithmically. The lower degrees have much higher counts as expected.

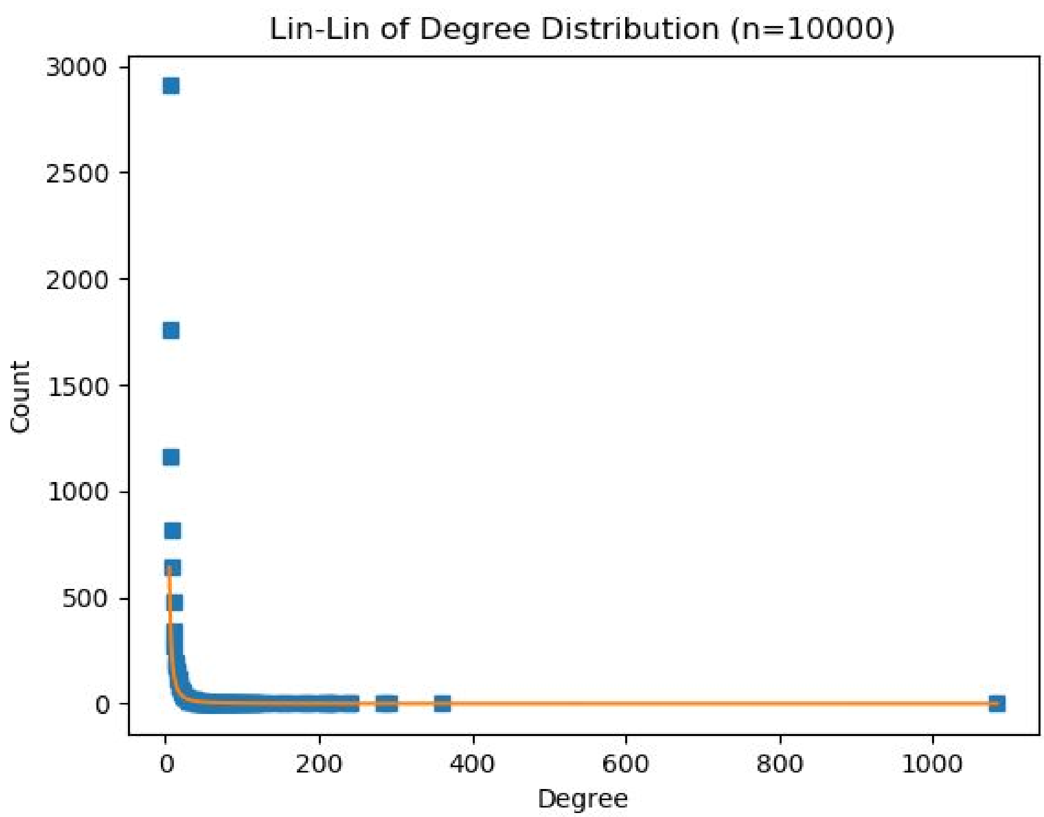
N = 1000



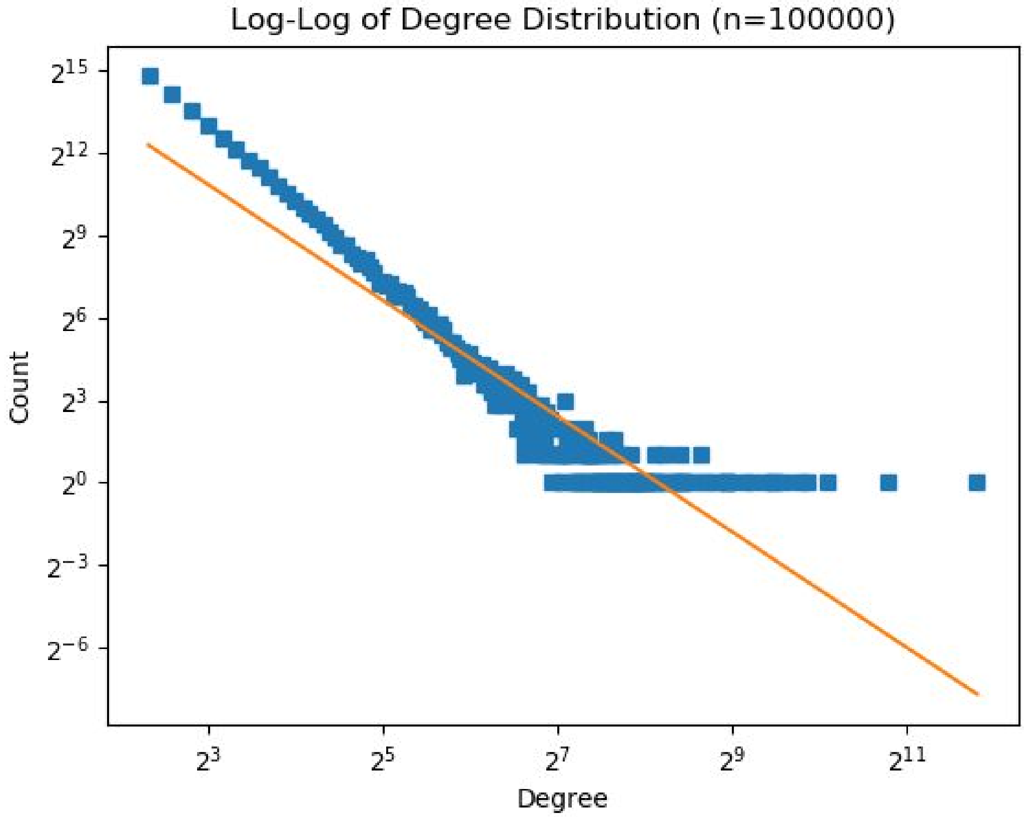


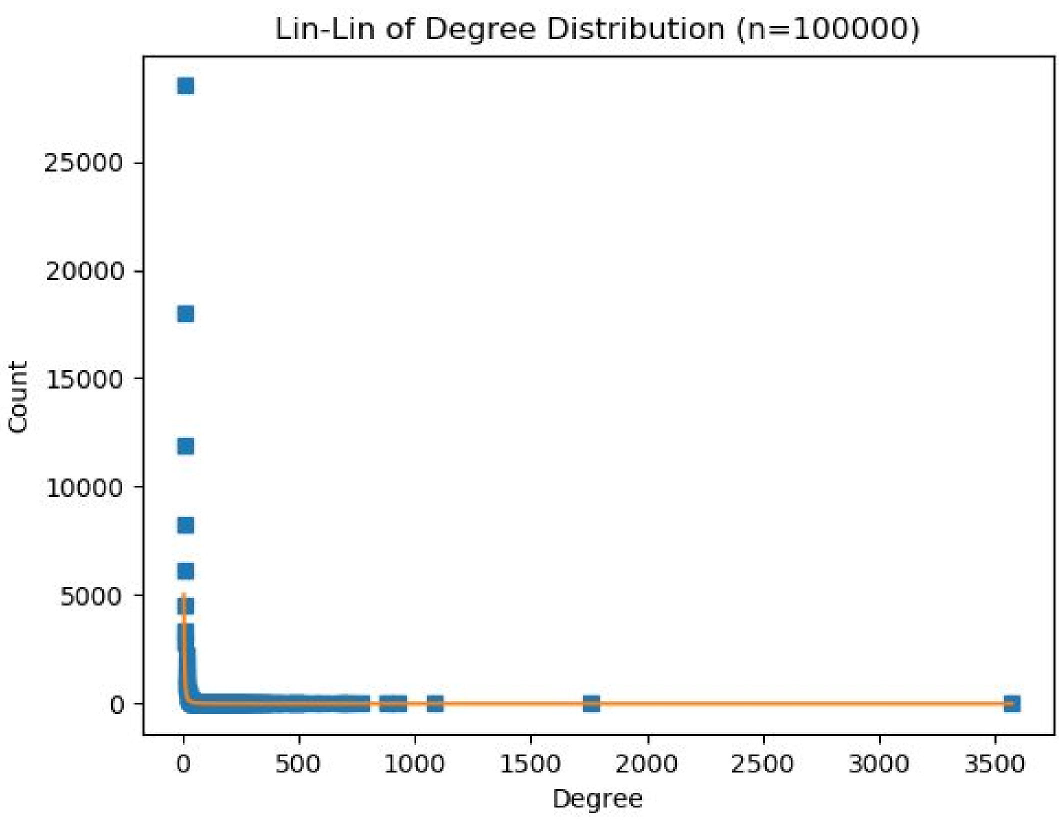
N = 10000


Log-Log of Degree Distribution (n=10000)
Count
Degree
2^{11}
2^{8}
2^{5}
2^{2}
2^{-1}
2^{-4}
2^{2} 2^{3} 2^{4} 2^{5} 2^{6} 2^{7} 2^{8} 2^{9} 2^{10}


Lin-Lin of Degree Distribution (n=10000)
Count
Degree
3000
2500
2000
1500
1000
500
0
0 200 400 600 800 1000

N = 100000


Log-Log of Degree Distribution (n=100000)
Count
Degree
2^{15}
2^{12}
2^{9}
2^{6}
2^{3}
2^{0}
2^{-3}
2^{-6}
2^{3}
2^{5}
2^{7}
2^{9}
2^{11}


Lin-Lin of Degree Distribution (n=100000)
Count
Degree
25000
20000
15000
10000
5000
0
0
500
1000
1500
2000
2500
3000
3500

Conclusion

All algorithms discussed provide helpful information about graphs and their characteristics. The BA model can create useful scale-free networks for use in many real world applications. The diameter can put in perspective the size of the network. Clustering-coefficient can show the occurrence of clusters and how vertices congregate. Degree distributions are useful in calculating probabilities of vertice degrees. All of these algorithms are useful in the creation, analysis, and modification of graph based networks, specifically scale-free ones.