Social Media Analysis Assignment

Anudit Nagar (E18CSE024)

Video Link - https://drive.google.com/drive/folders/1lkAGSMHtt-VwtVyYPiGCpiCbnd4tih6?usp=sharing

Property (a) — Number of nodes (n)			
Dataset 1 (15s.csv)	145	Dataset 2 (15.csv)	1258
ER Model	145	ER Model	1258
BA Model	145	BA Model	1258

Number of vertices of the graph. Same for Dataset and respective equivalent ER model graphs and BA model graphs.

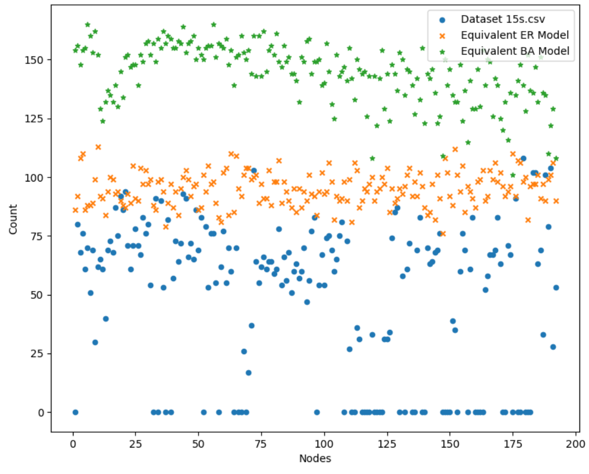
Property (b) — Number of edges (m)			
Dataset 1 (15s.csv)	2512	Dataset 2 (15.csv)	7682
ER Model	196	ER Model	15969
BA Model	1350	BA Model	12480

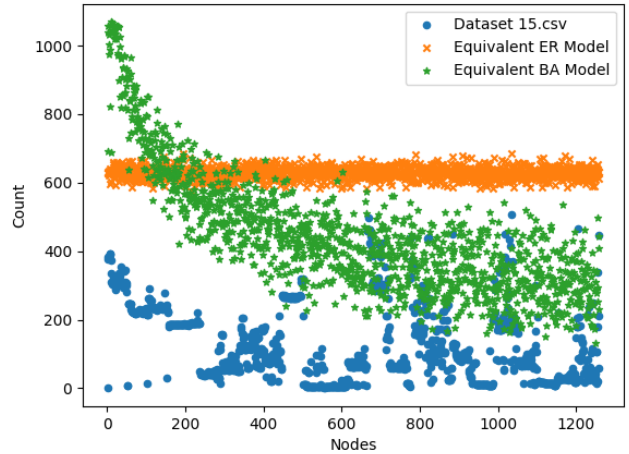
A line joining two nodes is an edge. Dataset1 is larger than Dataset2. Equivalent Erdős–Rényi model contains low number of edges due to low probability of connection and Barabási–Albert model has a lot of edges due to high connectivity.

Property (c) — Number of triangles			
Dataset 1 (15s.csv)	88	Dataset 2 (15.csv)	548
ER Model	137	ER Model	1258
BA Model	145	BA Model	1258

Number of triangles in the graph is equal to trace(A^3) / 6. Where trace(A) is the sum of the elements on the main diagonal of matrix A. The Barabási–Albert model contains more triangles compared to Erdős–Rényi model and the Dataset.

Property (c) — 3 node path graphs





Graph of Paths with 3 nodes depicting properties similar to the number of triangles shown above.

Property (d) — Maximum degree of a node			
Dataset 1 (15s.csv)	78	Dataset 2 (15.csv)	253
ER Model	9	ER Model	41
BA Model	68	BA Model	172

The Maximum Degree of a node is the degree of the vertex with the greatest number of edges incident to it. The BA model has a much larger Maximum degree compared to ER but the dataset provided has the largest.

Property (e) — Average degree			
Dataset 1 (15s.csv)	17.324	Dataset 2 (15.csv)	6.107
ER Model	1.431	ER Model	12.694
BA Model	9.310	BA Model	9.921

The average degree of a graph is used to measure the number of edges compared to the number of nodes. To do this we simply divide the summation of all nodes' degree by the total number of nodes. The Dataset1 hash the largest average degree and ER Model for Dataset 1 has the lowest.

Property (f) — Size of Largest connected component or Giant Component			
Dataset 1 (15s.csv)	145	Dataset 2 (15.csv)	308
ER Model	8	ER Model	127
BA Model	15	BA Model	125

A component also called a connected component, of a graph is a subgraph in which any two vertices are connected to each other by paths, and which is connected to no additional vertices in the supergraph. Datasets have a considerably larger Components as compared to ER or BA Models.

Property (g) — Diameter			
Dataset 1 (15s.csv)	6	Dataset 2 (15.csv)	15
ER Model	5	ER Model	11
BA Model	10	BA Model	14

The diameter of graph is the maximum distance between the pair of vertices. It can also be defined as the maximal distance between the pair of vertices. Dataset2 has the largest diameter and the trend continues with It's ER and BA model representations.

Property (h) — Power law exponent			
Dataset 1 (15s.csv)	2.1	Dataset 2 (15.csv)	2.33

A scale-free network is a network whose degree distribution follows a power law. Exponent is the constant by which these values vary. ER model is not having scale free phenomenon and BA model has a fixed exponent.

Property (i) — Average Clustering coefficient			
Dataset 1 (15s.csv)	0.393	Dataset 2 (15.csv)	0.038
ER Model	0.003	ER Model	0.01
BA Model	0.106	BA Model	0.026

The neighborhood of a node, u, is the set of nodes that are connected to u. If every node in the neighborhood of u is connected to every other node in the neighborhood of u, then the neighborhood of u is complete and will have a clustering coefficient of 1. If no nodes in the neighborhood of u are connected, then the clustering coefficient will be 0. The clustering coefficient, along with the mean shortest path, can indicate a "small-world" effect. For the clustering coefficient to be meaningful it should be significantly higher than in version of the network where all the edges have been shuffled.

Property (j) — Algebraic Connectivity			
Dataset 1 (15s.csv)	0.8475578	Dataset 2 (15.csv)	0.395625
ER Model	2.96919783	ER Model	8.3075925
BA Model	7.27700926	BA Model	7.4429649

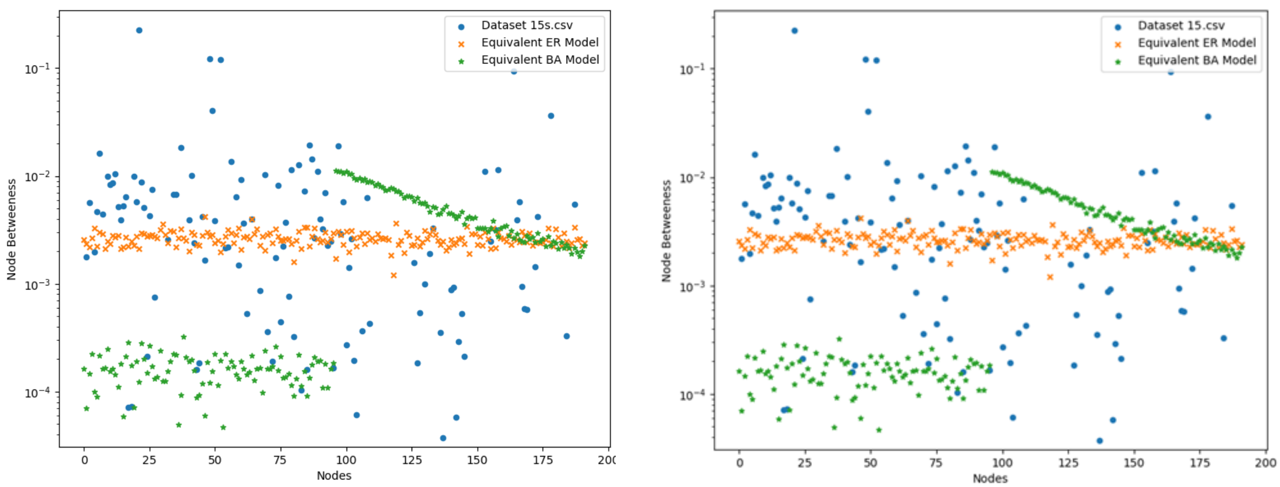
The algebraic connectivity (also known as Fiedler value or Fiedler eigenvalue) of a graph G is the second-smallest eigenvalue (counting multiple eigenvalues separately) of the Laplacian matrix of G.[1] This eigenvalue is greater than 0 if and only if G is a connected graph. This is a corollary to the fact that the number of times 0 appears as an eigenvalue in the Laplacian is the number of connected components in the graph. The magnitude of this value reflects how well connected the overall graph is. We see that is connectivity is very large in BA model which is also depicted in the visualizations below.

Property (k) — Average Path length			
Dataset 1 (15s.csv)	2.020	Dataset 2 (15.csv)	6.002
ER Model	1.952	ER Model	3.137
BA Model	2.226	BA Model	3.380

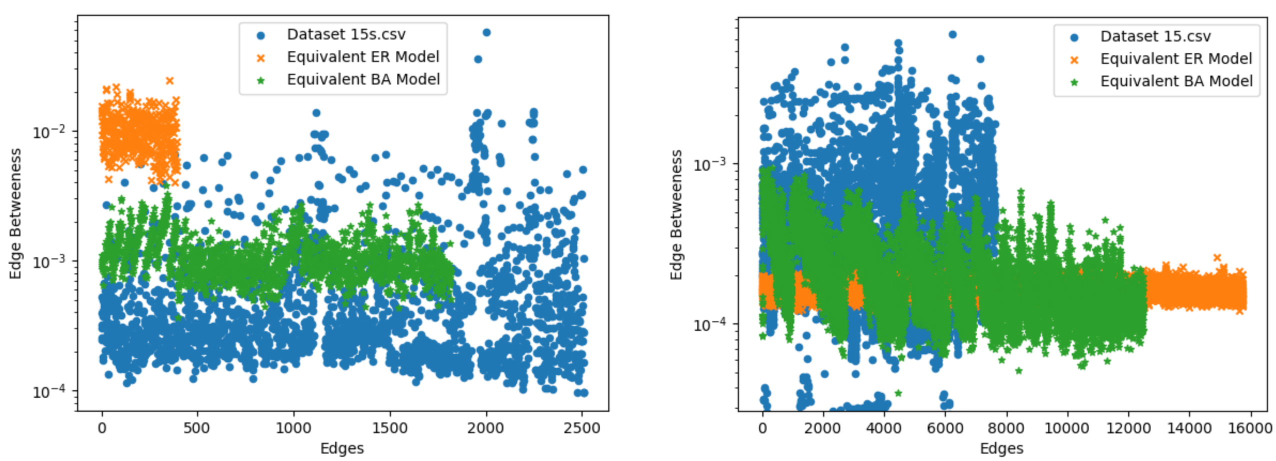
The average path length in the graph – average distance between any couples of nodes may also highlight properties of a given graph. Dataset2 has a considerably longer average path length compared to Dataset1.

Betweenness centrality finds wide application in network theory; it represents the degree to which nodes stand between each other. For example, in a telecommunications network, a node with higher betweenness centrality would have more control over the network, because more information will pass through that node. We see two types of betweenness, node betweenness and edge betweenness below.

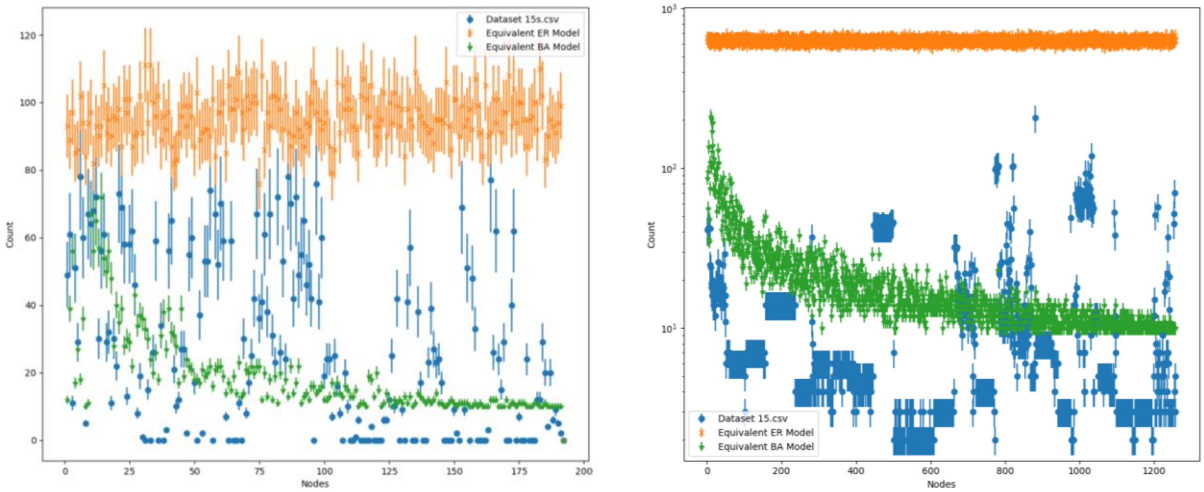
Property (l) — Node betweenness distribution



Property (m) — Edge betweenness distribution

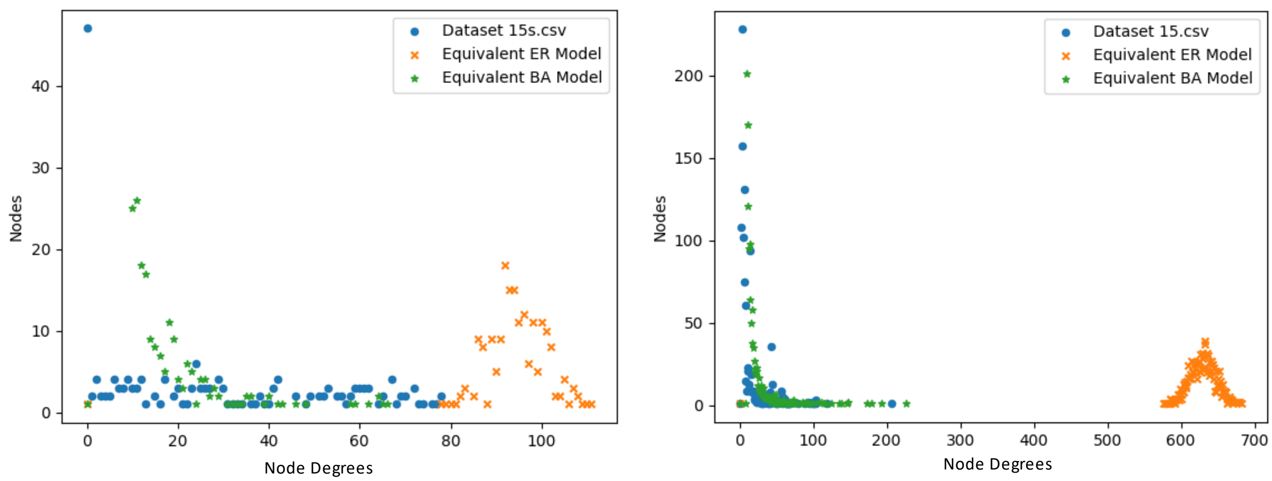


Property (n) — Dispersion with respect to each of the nodes’ degree i.e. Standard deviation



A dispersion graph shows the range of a set of data and illustrates whether data groups are dispersed or not and thrier magnitudes.

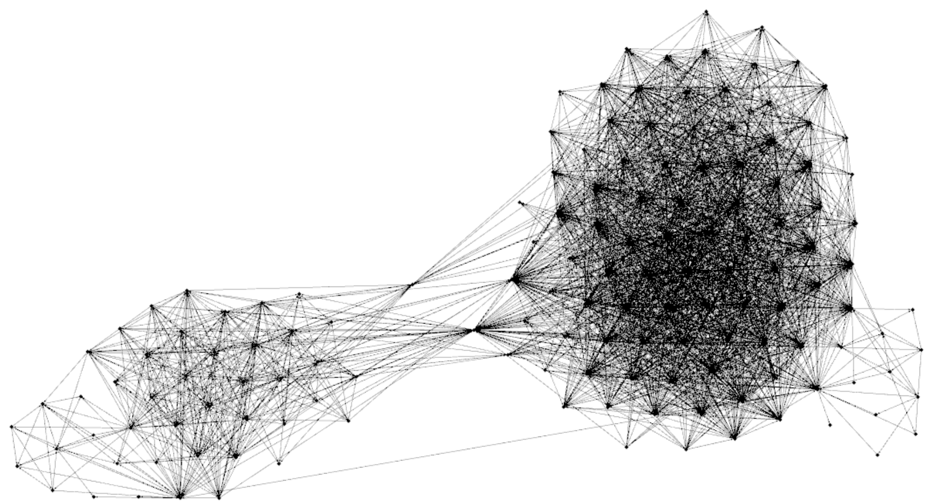
Property (o) — Generate the degree distribution



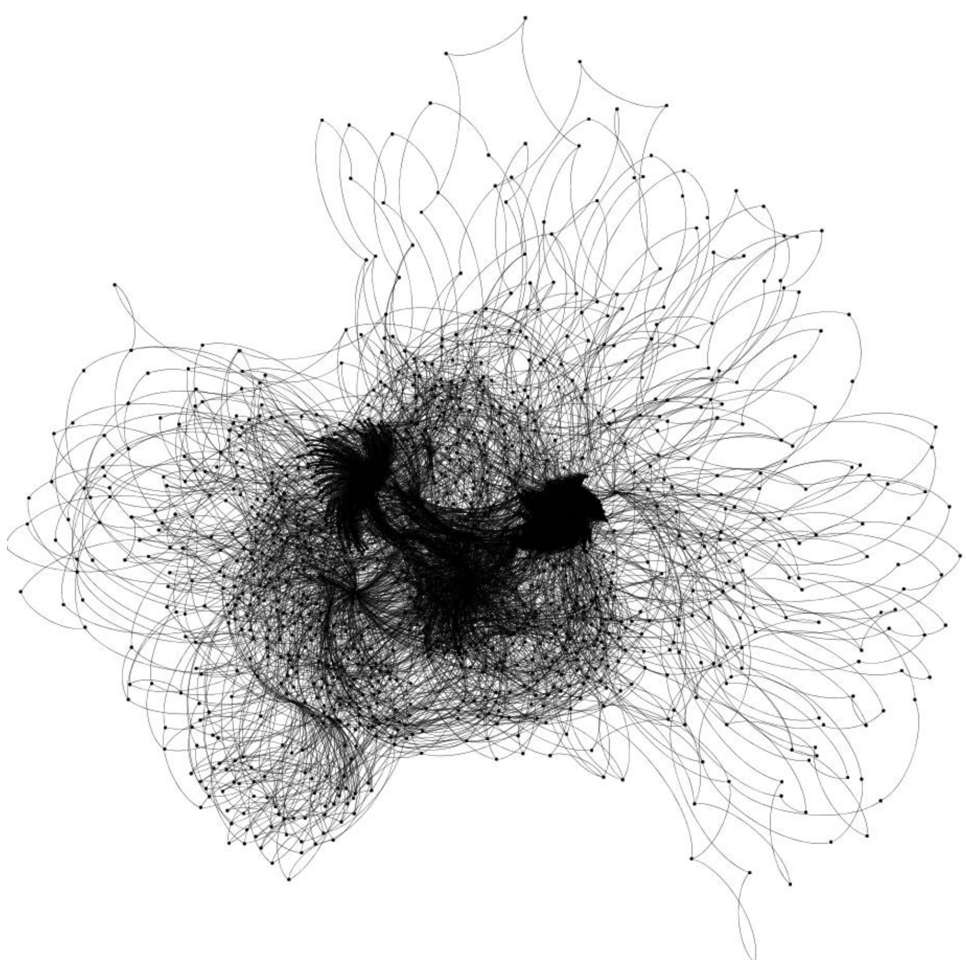
Degree distribution is the degree compared the number of nodes with that degree. The distribution of the datasets are somewhat similar to the ER Model. The Degree distribution shows the normally distributed ER model and highly connected BA Models.

Property (m) — Visualize the Graph

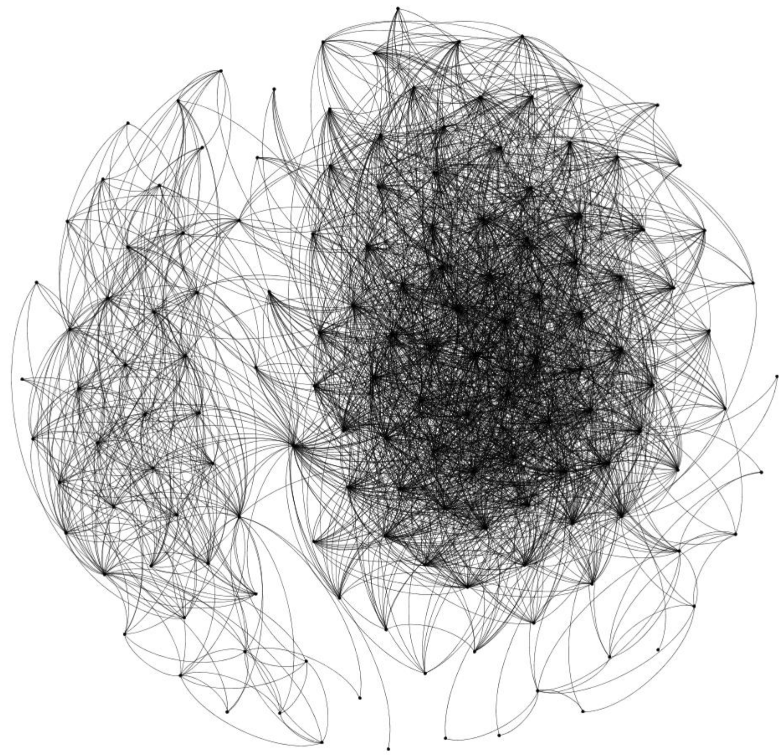
Dataset1 – 15s.csv



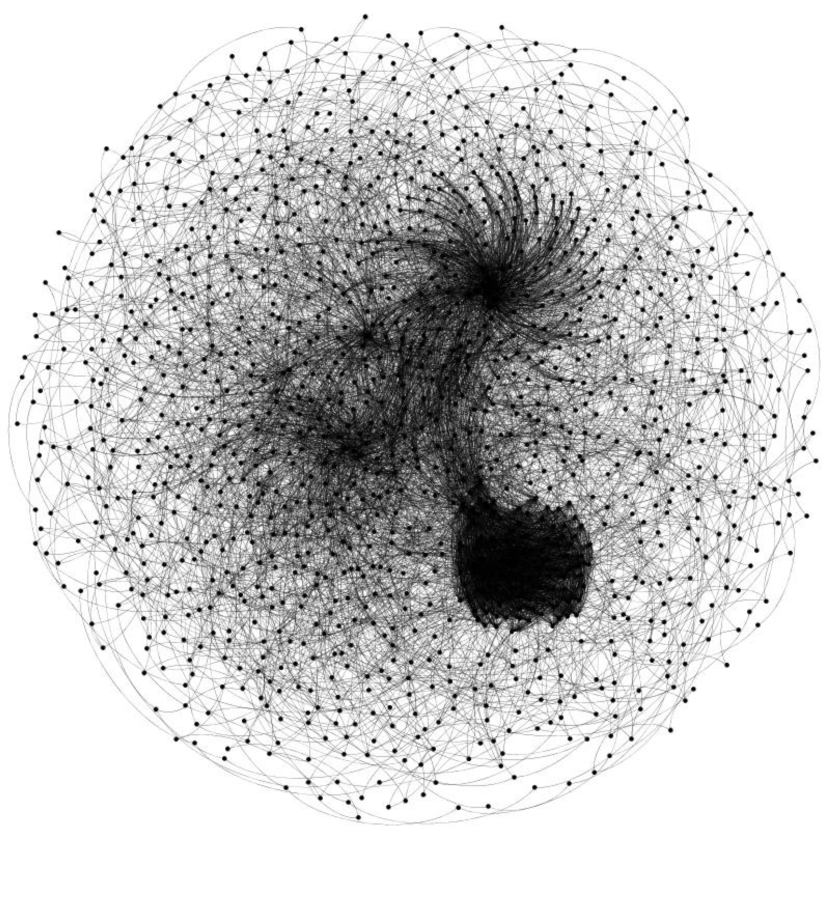
Dataset2 - 15.csv



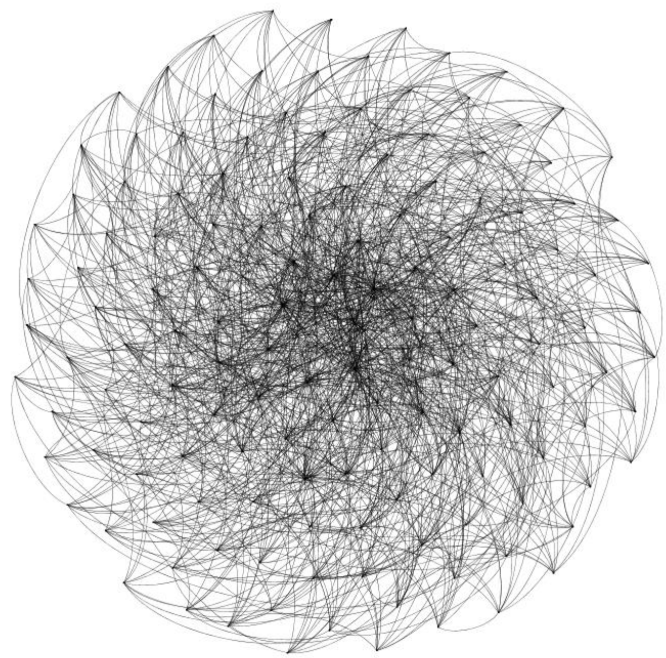
15s.csv – ER Graph



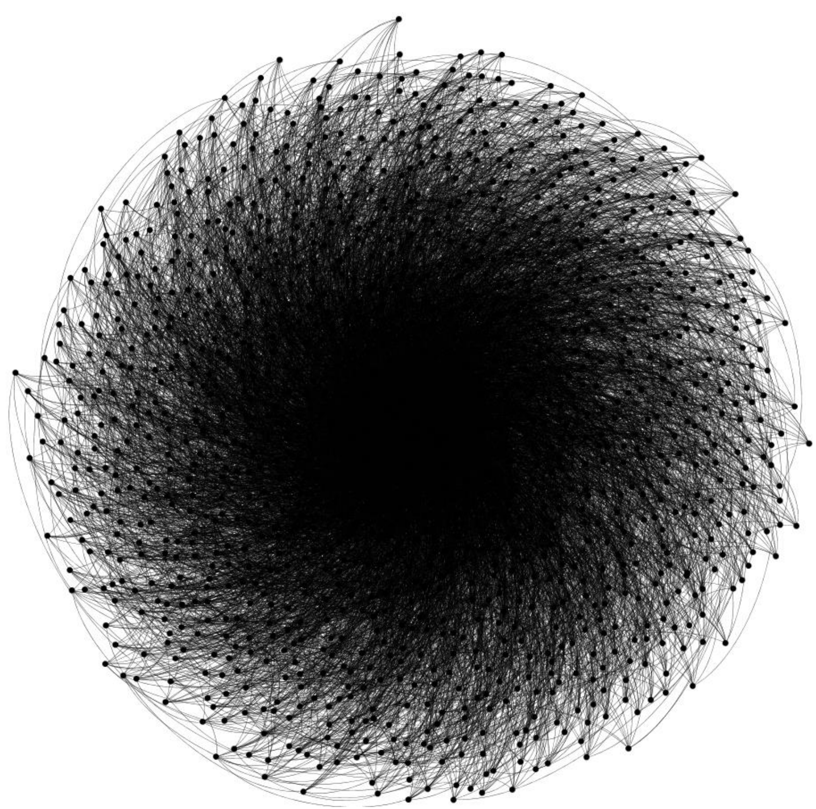
15.csv – ER Graph



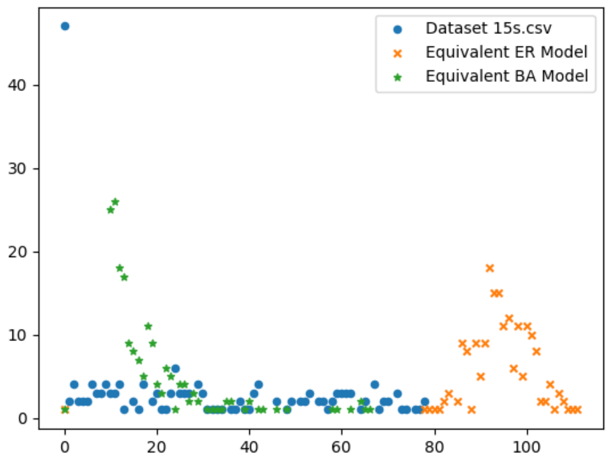
15s.csv – BA Graph

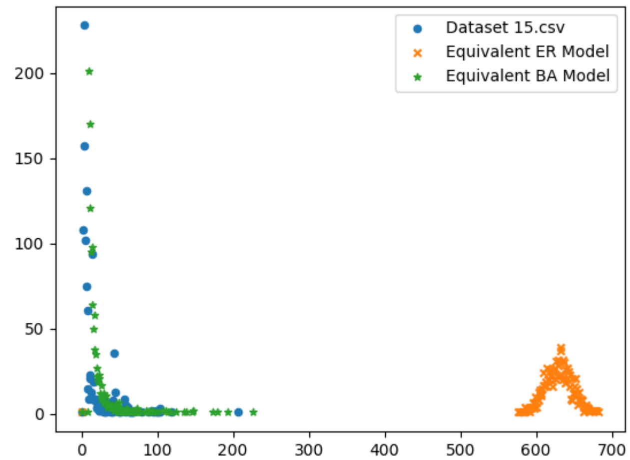


15.csv – BA Graph



Property (q) — Probability distribution





For the length of the shortest paths pd with respect to d such that x-axis is degree and varies as d=1,2,3,... and pd is calculated for each such shortest distance.

Property (r) — Spectral Radius			
Dataset 1 (15s.csv)	3	Dataset 2 (15.csv)	7.4
ER Model	2.5	ER Model	5.5
BA Model	5	BA Model	7

For a graph G, we denote by ρ(G) the largest eigenvalue of A(G) and call it the spectral radius of G. BA model has a high spectral radius as compared to ER Model.

Property (s) — Assortativity (determined with the help of Pearson Correlation Coefficient)			
Dataset 1 (15s.csv)	0.2597	Dataset 2 (15.csv)	-0.06
ER Model	-0.1116	ER Model	0.00517
BA Model	-0.0245	BA Model	-0.0245

Assortativity, or assortative mixing is a preference for a network's nodes to attach to others that are similar in some way. In Dataset1 nodes have a higher tendency to connect to each other while BA Model has the least tendency.