Lab 8

1. For the 1st data set available on the LMS, the characteristics of this dataset is as following

This is a gene fusion network. Nodes are genes, and two nodes are connected if the two genes have been observed to have fused during the emergence of cancer. Here First column represents ID of source node and 2nd column represents ID of destination node.

This is a unipartite undirected network and it is unweighted, lacks multi-edges, and does not contain loops with respect to each of the nodes. Import the above dataset in a code and investigate the following properties:

1. Number of nodes (n)
2. Number of edges (m)
3. Number of triangles
4. Maximum degree of a node
5. Average degree
6. Size of Largest connected component
7. Diameter
8. Power law exponent
9. Average Clustering coefficient
10. Algebraic Connectivity
11. Average Path length
12. Node betweenness distribution
13. Edge Betweenness Distribution
14. Dispersion with respect to each of the nodes’ degree i.e. Standard deviation.
15. Generate the degree distribution
16. Visualize the graph
17. Probability distribution for the length of the shortest paths pd with respect to d such that x-axis will be degree and it will vary as d=1,2,3,… and pd is to be calculated for each such shortest distance.
18. Spectral radius

Compare the properties of this dataset with a random network (ER graph) of equal number of nodes such that probability of having an edge in between any two nodes of this networks is obtained by

P= where nC2 represents nC2 and m represents the number of edges in the given dataset.

Investigate the above properties of the given dataset and ER graph. To ascertain the properties of ER graph, run 20 iterations for both graphs. Compare the properties of both. Plot the properties for these two types of graphs) with respect to property (l), (m), (n), (o), (q) such that there is a distinct plot for each of these properties.