ECE 18-755: Networks in the Real World

HW1 Solutions

Problem 1 (10 pts)

a) Consider 10 isolated nodes. For each pair of nodes, throw a fair die and connect them if the number on the die is 1. Describe the graph you obtained. Is it connected or not? What is the average degree and the degree distribution? Are there any cycles?

Now repeat the experiment and connect nodes if the number on the die is 1 or 2. How is the graph different from the previous case?

ANS: The random graph obtained by linking two nodes only if the die outcome is 1 is given below Fig.1.a): The random graph in Fig.1.a) is not connected.

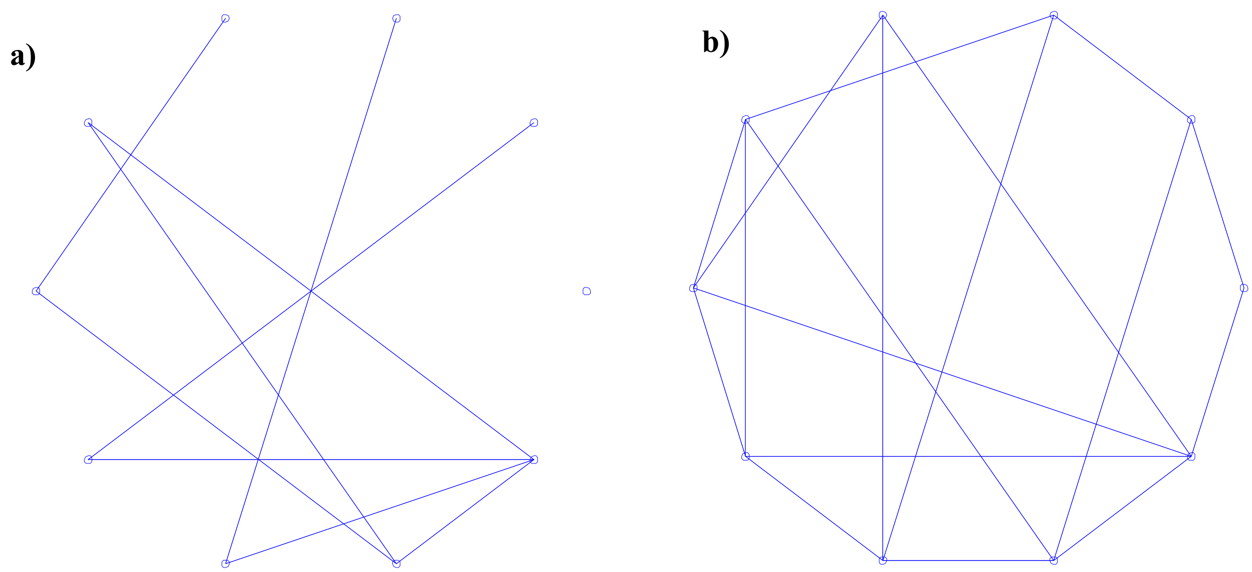


Figure 1. a) The random graph obtained by linking two random nodes only if the outcome of throwing a die is 1. b) The random graph obtained by linking two random nodes if the outcome of the die is either 1 or 2.

The random graph in Fig.1.b is connected. The average degree of the random graphs in Fig.1.a and Fig.1.b is 0.9 and 1.8, respectively. Both graphs contain cycles.

Degree distribution:

p = 1/6

	$P(d)$
$d = 0$	0.1
$d = 1$	0.3
$d = 2$	0.4
$d = 3$	0.1
$d = 4$	0.1

p = 1/3

	$P(d)$
$d = 2$	0.1
$d = 3$	0.3
$d = 4$	0.5
$d = 5$	0.1

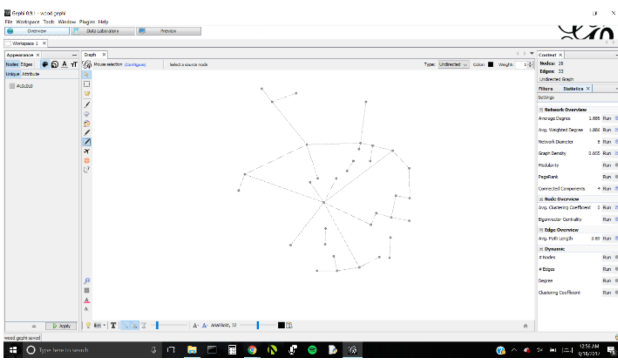
b) Install the following software tool: Gephi: http://gephi.org/. Use it to draw and analyze the graphs you obtained in part a) (i.e., report the number of unreachable pairs of nodes, average distance among reachable pairs of nodes, network diameter, clustering coefficient, etc.). Provide a screenshot of your work.

ANS: Gephi graphs with feature analysis

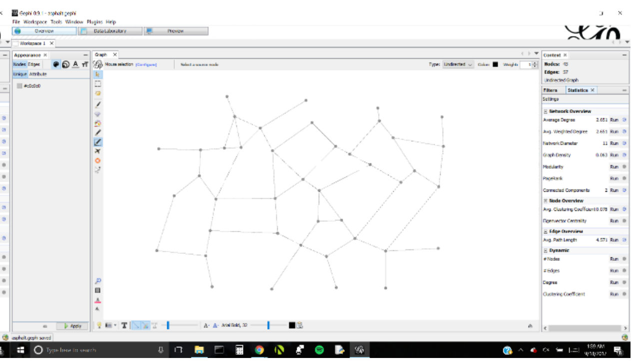
c) Now build and analyze a few real world networks. More specifically, we will look into how microstructures of different materials can affect the network features of crack lines and compare how these propagate through different materials.

ANS:

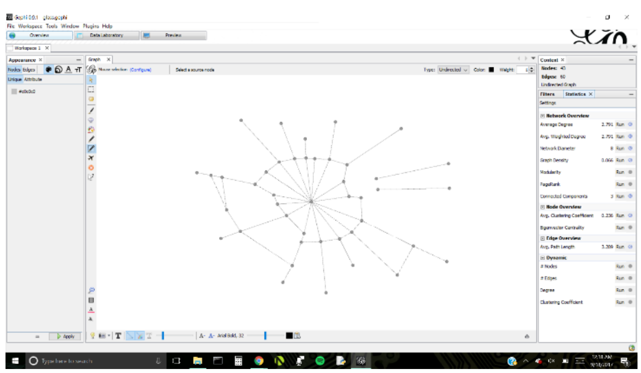
Wood



Asphalt



Glass



Discuss the characteristics of the materials and their effects on the network properties. E.g. The higher average degree in glass due to its characteristic to splinter, a higher average path length in asphalt, which breaks in chunks together, and a higher number of weakly connected nodes (separate small cracks) in wood due to its anisotropic properties.

Problem 2 (15 pts)

Consider the network of loans (as directed arcs) between 16 Renaissance Florentine families given in the text file *Renaissance.net*[1].

a) Using the graph visualizing tool of your choice plot the corresponding graph of the file *Renaissance.net*. (For instance, you can use Gephi but you can also hand draw it.) What is the *IN* and *OUT* degree distribution of the network?

ANS:

Nodes ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
IN	1	3	2	3	3	1	4	1	6	1	3	0	3	2	4	3
OUT	1	3	2	3	3	1	4	1	6	1	3	0	3	2	4	3

The corresponding graph is given below:

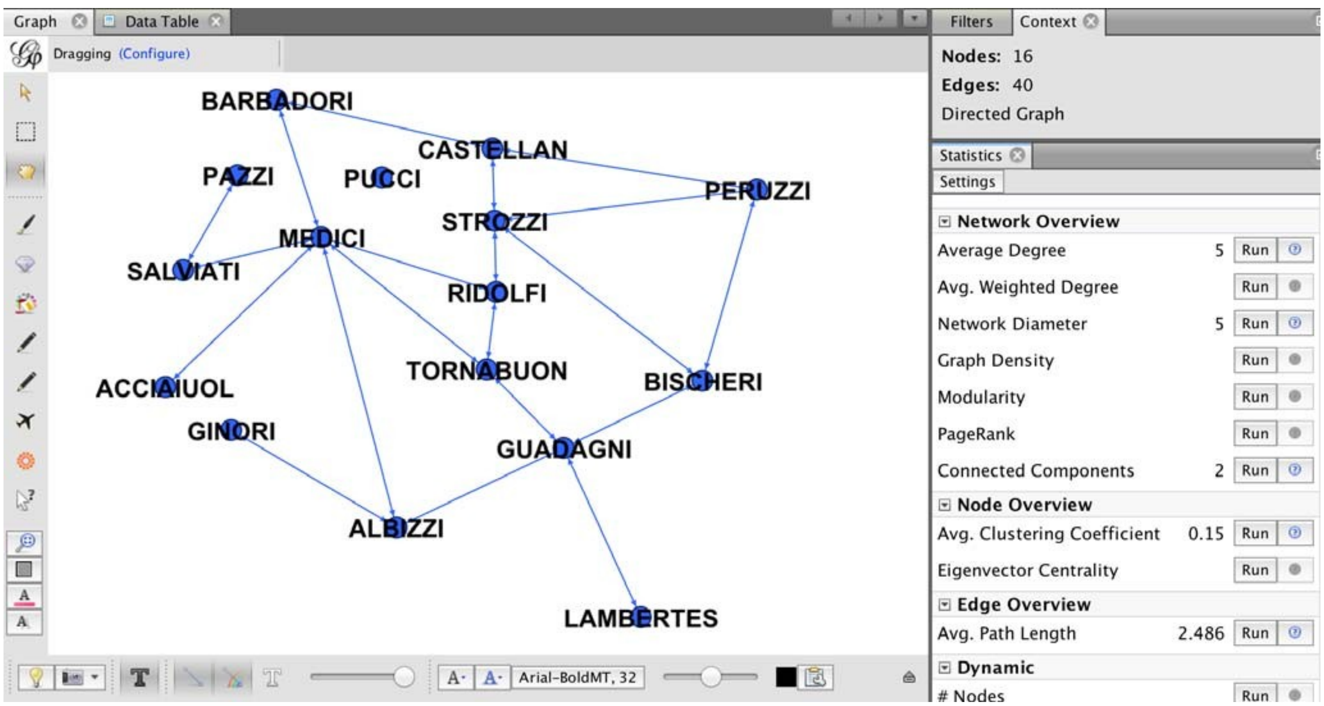


Figure 2. The graph corresponding to the loans among 16 Renaissance Florentine families using Gephi.

b) What is the average distance between the nodes of the largest strongly connected component? How does this compare to the network diameter?

ANS: If we do not exclude node 12 (Pucci), then the average distance between nodes is infinite. Excluding node 12 we obtain a strongly connected component with an average distance of 2.48 hops. If we do not exclude node 12 (Pucci), then the network diameter is infinite. Excluding node 12 we obtain a graph with a network diameter of 5. Note that the network diameter is defined as the maximum shortest path between any two nodes of the graph.

. 1. The *Renaissance.net* file contains a network of 16 nodes and 38 directed arcs. Note that the file is written such that the Gephi software can read it and display the corresponding network.

c) What nodes are the most important in terms of total degree and betweenness? Illustrate one case where these two statistics don't agree and explain why.

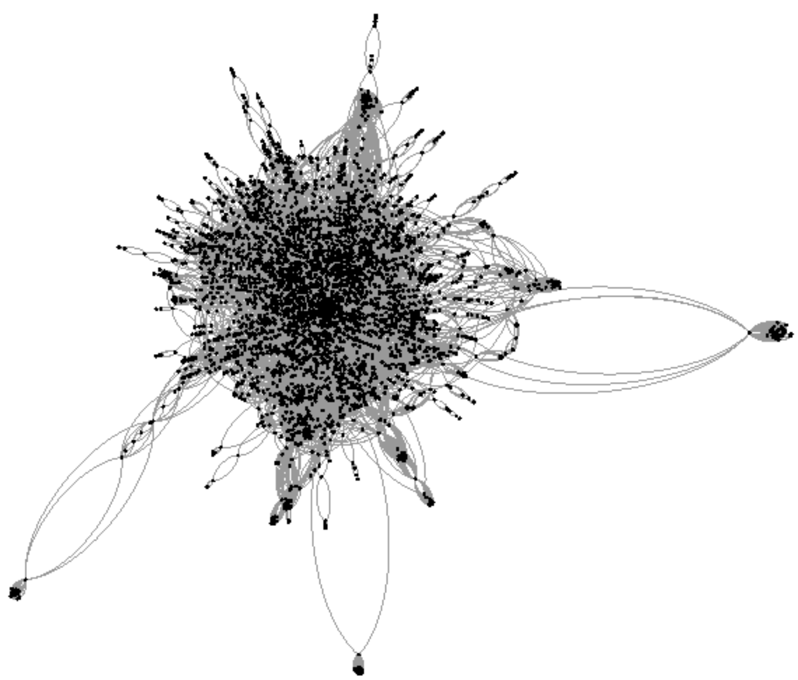
ANS: Node 9 (Medici) has the highest degree (12) and betweenness (95). While both Guadagni and Strozzi have total degrees of 8, Guadagni has a higher betweenness (46 vs 18). This is because if Guadagni is removed, nobody can reach to Lambertes anymore, whereas removing Strozzi has no such an effect. In this sense, betweenness measures importance better than degree.

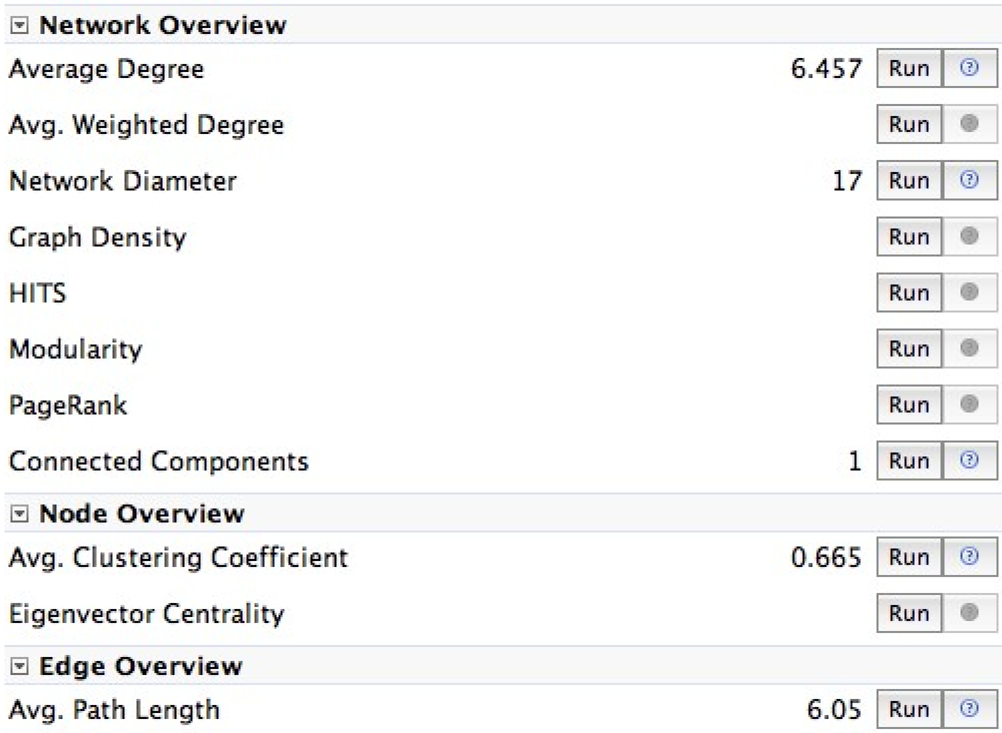
Problem 3 (20 pts)

The file "arXiv_lcc.gml" is an undirected network of scientific collaborations between authors who submitted to General Relativity and Quantum Cosmology. Use the Gephi network tool to:

a) Visualize and calculate the average degree, average path length, diameter, and average clustering coefficient of this network.

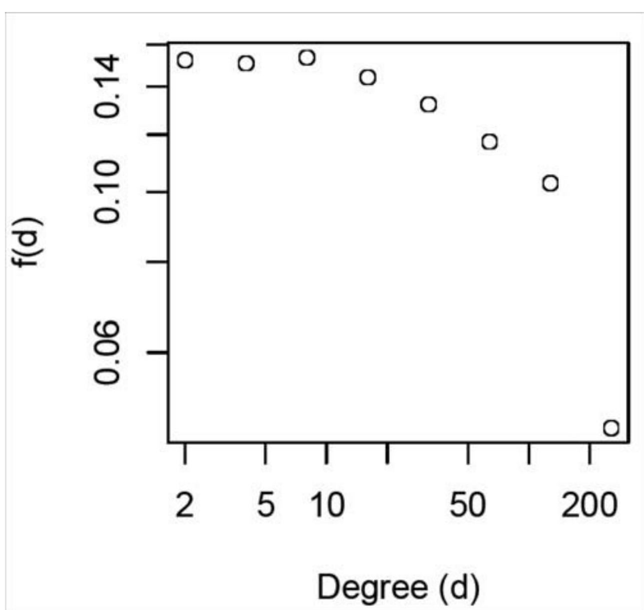
ANS:

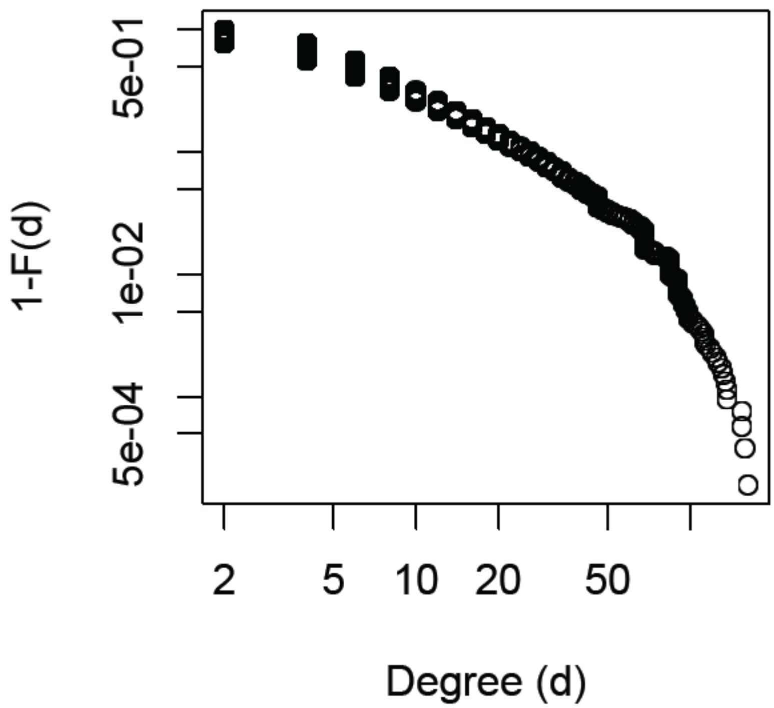




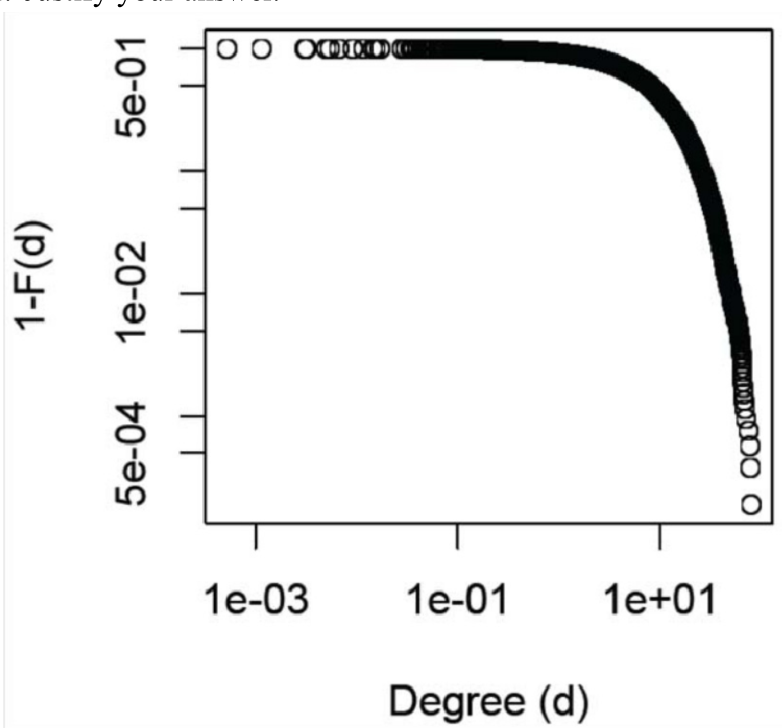
Network Overview	
Average Degree	6.457
Avg. Weighted Degree	
Network Diameter	17
Graph Density	
HITS	
Modularity	
PageRank	
Connected Components	1
Node Overview	
Avg. Clustering Coefficient	0.665
Eigenvector Centrality	
Edge Overview	
Avg. Path Length	6.05

b) Plot the degree distribution using both probability density function (pdf) and rank frequency plot (both in log-log scale).





c) Generate 5000 exponentially-distributed numbers with a mean identical to the average degree. Plot the rank frequency plot in log-log scale and contrast it to the result of b). What kind of network is this collaboration network? Justify your answer.



ANS: This is a scale free networks with an exponential cutoff.

Problem 4 (25 pts)

Based on the concepts of characteristic path length and clustering coefficient learned in class, study the impact of increased randomness in the network structure using a regular lattice with $N = 100$ nodes and connectivity $k = 4$ (*i.e.*, each node has 4 directed links to its neighbors as described below). More precisely, to introduce some randomness into the regular lattice proceed as follows:

- Consider first a regular lattice where nodes are distributed uniformly over a circle; each node is directly connected to its first and second immediate neighbors (that is, $k = 4$ as discussed in class). Note, however, that the circle is not divided into 16 parts as in the lecture notes but instead in 100 pieces.
- For each node and the edge connecting it to its first and second order neighbors (clockwise), re-wire this edge with a probability p to a new node chosen uniformly at random over the entire circle. If the two nodes are already connected, then leave the edge as it was. Stop iterating when each node was considered exactly once.

Using a logarithmic scale, vary the probabilities p between 0.0001 and 1 and plot on the same graph the normalized values of the characteristic path length and clustering coefficient (that is $L(p)/L(0)$ and $C(p)/C(0)$ similar to what we discussed in class). Comment the results.

ANS: By increasing the randomness into the regular lattice we induce small world effects. By applying the above rewiring procedure with increasing probability p (increased randomness) we obtain a graph with a smaller characteristic path length $L(p)$; the clustering $C(p)$ stays pretty stable until p gets closer to 0.1 after which is falls pretty fast.

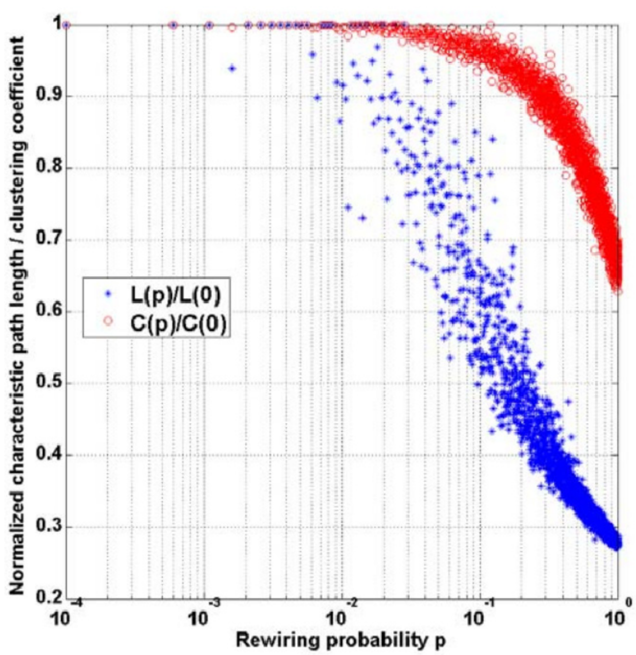


Figure 3. The plot of normalized characteristic path length and normalized clustering coefficient as a function of the increased randomness in the network.

Problem 5 (30 pts)

Based on what we have learned in class, study the characteristics of an Erdos-Renyi (ER) graph and a Scale-Free (SF) network generated as described below.

Scale-free: Given $N = 1000$ nodes and a preferential attachment mechanism for linking nodes construct a SF graph as follows:

- Initially, choose $m = 5$ nodes and randomly connect them. Make sure that you do not add self-loops and multiple edges between two nodes.
- At each iteration, pick a node from the remaining (N - m) nodes and connect the new node to at most m nodes with a certain probability. More precisely, the probability that the new node i will be connected to a node j is proportional with its degree k_j; this probability can be written as $P(i \rightarrow j) \sim k_j \,/\, \sum_j k_j$. Note that we consider only undirected links so if i and j are connected, then both entries $A(i,j)$ and $A(j,i)$ in the adjacency[1] matrix are set to 1. Make sure you do not add self-loops to the generated graph.
- Continue this process until all the initial nodes are linked at least once to the growing network.

Erdos-Renyi: Given $N = 1000$ isolated nodes, a probability $p = 0.05$ of connecting two randomly selected nodes, and a maximum number of edges equal to the number of edges in the SF graph, generate the ER graph as follows:

- At each iteration pick two random nodes and if they are not already connected, then link them with probability p.
- Continue this process until the maximum number of iterations is reached.

Save the two adjacency matrices into "ER.txt" and "SF.txt" files, respectively. Answer the following questions:

a) Compute the degree distribution of the two networks and explain their properties (*e.g.*, maximum and minimum degree, existence of hubs).

ANS: The maximum, mean and minimum degree in the ER graph are as follows: 25, 11.936 and 3 respectively. The maximum, mean and minimum degree in the SF graph are as follows: 124, 11.937 and 6, respectively.

b) Estimate and plot the empirical PDF of the degree distributions for the two networks. Use any of the software tools to display the two networks.

---

1. The adjacency matrix A of a finite undirected graph of N vertices and E edges is a matrix of size NxN, where each row corresponds to a distinct node and the non-zero entries in a row denote an edge between two nodes. For instance, if $A(i,j)$=1 then it means that there is an undirected edge between nodes i and j. Note that we do not allow self-loops in these graphs so the diagonal elements of $A(i,i) = 0$. Moreover, the sum of all the elements in the upper triangular part of the adjacency matrix A should equal the total number of edges E in the graph.

The solution should look like this:

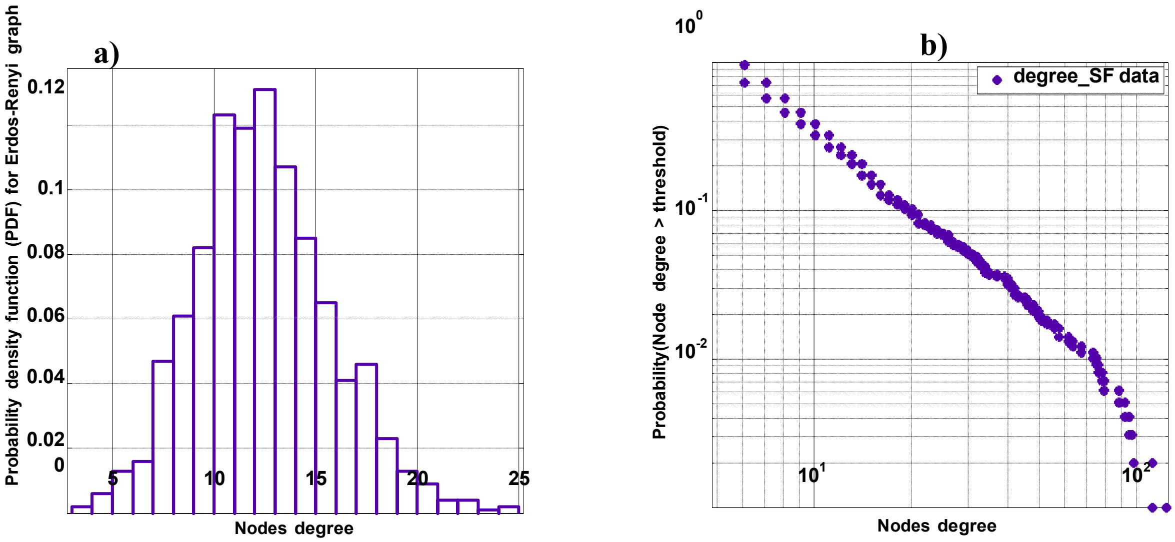


Figure 4. a) Probability density function of the degree for ER graph consisting of 1000 nodes and 5968 edges where the probability of connecting two randomly nodes was set to 0.05. b) The probability to find a node whose degree exceeds a certain threshold for a SF graph with 1000 nodes where each newly linked node is linked to at most *m*=5 nodes from the initial network.

c) Compute the clustering coefficient for the two networks.

ANS: The clustering coefficient for the ER graph is 0.0042380. The clustering coefficient for the SF network is 0.0144762.

d) For both graphs remove now 2% of the nodes with the highest degree. Recompute the clustering coefficient for both graphs and compare the results with the previously computed coefficients. Explain what do you observe.

ANS: By removing 20 most connected nodes (*i.e.*, 2% of the nodes with highest degree), the clustering coefficient for the ER graph becomes 0.0041516. As one can observe, the clustering coefficient does not change much with the procedure of removing the highest connected nodes. In other words, we can say that the ER networks are resilient to directed attacks.

In contrast, by removing 20 most connected nodes, the clustering coefficient for the SF network becomes 0.0053985. As one can see, a targeted attack (i.e., removing the highest connected 2% nodes) has a significant impact on the SF network as its clustering coefficient decreases from 0.0144762 to 0.0053985.