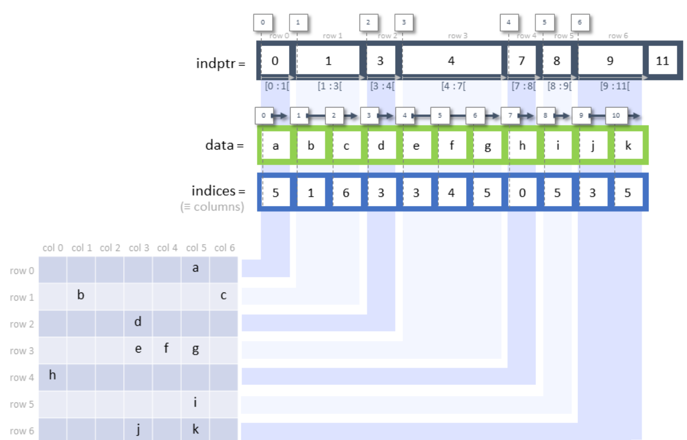
Orkhan Bayramli
Network Science
Homework 1

Introduction

The data for Homework 1 has been chosen from Sandford open datasets. Data is the sequence of snapshots of the Gnutella peer-to-peer file sharing network from August 2002. There are total of 9 snapshots of Gnutella network collected in August 2002. Nodes represent hosts in the Gnutella network topology and edges represent connections between the Gnutella hosts. In general, there are 10876 nodes and 39994 edges. The network is directed.

Average degree

A key property of each node is its *degree*, representing the number of links it has to other nodes. In this particular dataset, degree of node is the number of connections that one node has to another one. In order to find it, the dataset has been converted into sparse matrix. Sparse matrix is an efficient way of storing and illustrating the dataset because of its memory-efficient storing technique. Since the real networks are sparse, most of the matrix will be containing 0s. However, in this case, only the links which connects nodes will be stored in a special way. Following image shows the explanation of a function which was used in the coding part:



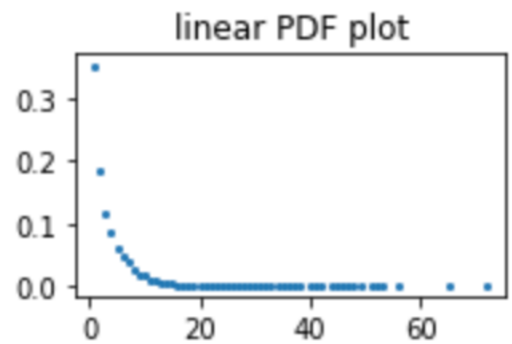
Another property of is the **average degree** of a network which is calculated by the following formula:

$$< k > = \frac{1}{N}\sum_{i=1}^{N} k_i$$

k_i is the degree of node i. N is the number of nodes in the network. The sum of all degrees is divided by the number of nodes. In our case, the average degree became 3,67. Our network corresponds to the supercritical regime in which the <k> is bigger than 1 as most of the real networks correspond.

Power law and degree distribution

Following picture has been captured from the output of coding part. The graph of probability density function shows that we have high degree nodes much bigger than 3,67. In other words, the hubs are present in our network having higher degrees. The graph corresponds to the power law distribution rather than Poisson.

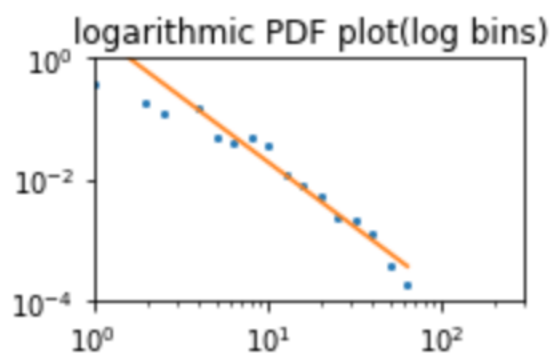


In the random network, we observe that the degrees of nodes follow Poisson distribution which has bell shaped graph meaning most nodes will be possessing average number of degrees. However, a scale-free network is a network whose degree distribution follows a power law as mentioned above.

$$p(k) = Ck^{-\gamma}$$

Gamma is the degree exponent. By the help of, following formula, we can find it from the network. In our case, the gamma is equal to 3,14.

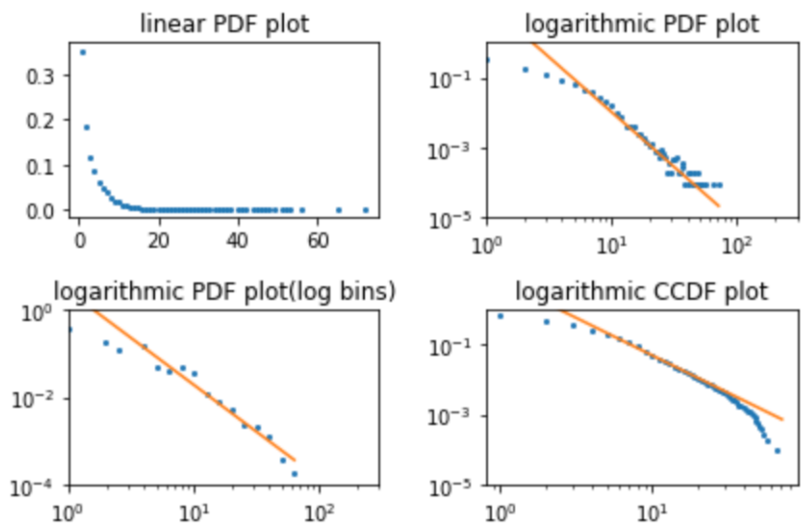
$$\log p_k \sim -\gamma \log k$$



C is the normalizing factor and is calculated in following way.

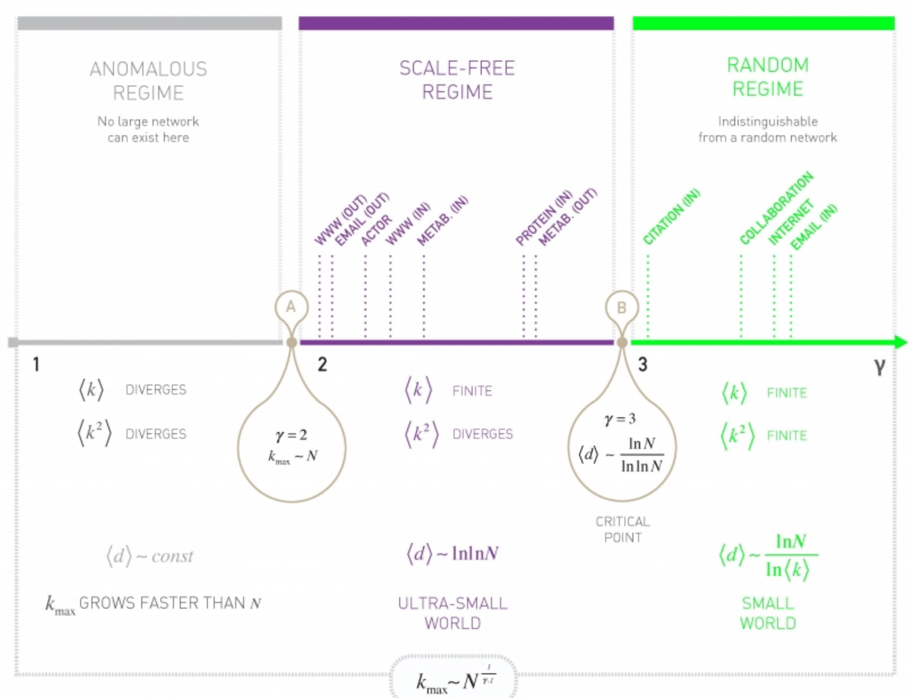
$$C = (\gamma - 1)\, k_{min}^{\gamma-1}$$

The full result including ccdf is as follows:



Diameter of network

As stated above, the gamma for our network is 3.14 (> 3). In this case, while hubs continue to be present, for $\gamma > 3$ they are not sufficiently large and numerous to have a significant impact on the distance between the nodes. Scale-free networks with large γ are hard to distinguish from a random network. That's why, diameter of a network which is the longest shortest path of network is almost equal to the average distance ($d_{max} = < d >$). The following picture illustrates the gamma dependent properties of scale-free networks including average degree. In our case, average distance is 7.14.



Clustering coefficient

Clustering coefficient C_i measures the density of links in node i's immediate neighborhood. The range of C_i is between 0 and 1, 0 means poor and 1 means dense connection. The average clustering coefficient is found in the following way:

$$< C > = \frac{< k >}{N}$$

In our network, the average clustering coefficient became 0.0034 meaning that the nodes in the network is loosely connected which can be explained through the presence of hubs.

Assortativity

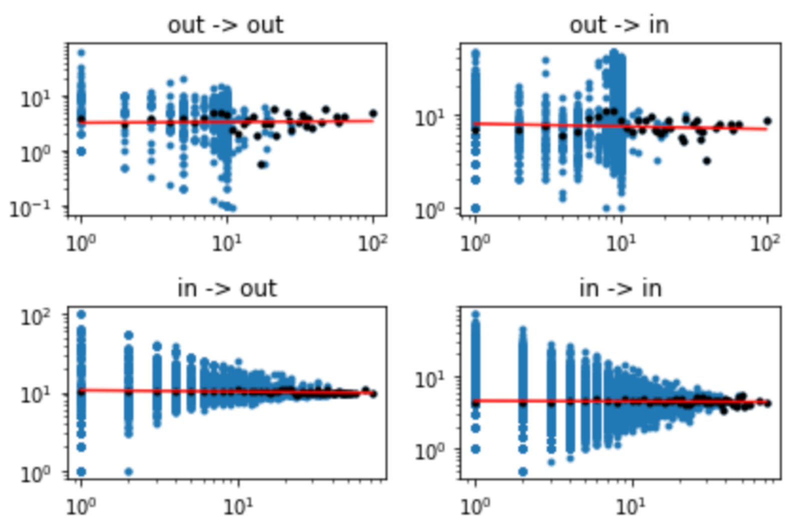
Assortative networks are the networks whose nodes tend to connect to the similar same degree nodes. In other words, hubs avoid the small nodes or hubs incline to connect to the hubs. Assortativty of a network is found by the average degree of its neighbors.

$$k_{nn}(k_i) = \frac{1}{k_i} \sum_{j=1}^{N} A_{ij} k_j$$

The following picture depicts the relationship between $k_{nn}(k_i)$ and k_i. If we take into account that:

$$k_{nn}(k) = ak^{\mu}$$

Then, by calculating the correlation exponents we can state that our network is neutral which is typical to the random network. It is no chance that, according to the degree exponent – gamma, our network was slightly tend to the random network.



Correlation exponents varied in the range of -0.028 < μ < 0.015. To measure correlations in directed networks we must take into account that each node i is characterized by an incoming k_i^{in} and an outgoing k_i^{out} degree.

Structural cut-off

For random networks and scale-free networks with $\gamma \geq 3$ the exponent of k_{max} ($k_{max} \sim N^{1/\gamma-1}$) is smaller than 1/2, hence k_{max} is always smaller than k_s. In other words, the node size at which the structural cutoff turns on exceeds the size of the biggest hub. Consequently, we have no nodes for which $E_{kk'} > 1$. For these networks we do not have a conflict between degree correlations and the simple network requirement.