Università degli Studi di Padova

Department of Information Engineering

Master degree in ICT for Internet and Multimedia

Course: Network Science

Homework 1

Davide Carta

ID: 1210702

December 20^{th} 2019

This homework aims at extracting relevant analytics from a network. For this task MATLAB was used.

About the dataset

The chosen dataset represent a collaboration network from the e-print arXiv and covers scientific collaborations between authors papers submitted to General Relativity and Quantum Cosmology category. If an author *i* co-authored a paper with author *j*, the graph contains a undirected edge from *i* to *j*. If the paper is co-authored by *k* authors this generates a completely connected (sub)graph on *k* nodes. The data covers papers in the period from January 1993 to April 2003 (124 months). It begins within a few months of the inception of the arXiv, and thus represents essentially the complete history of its GR-QC section. This dataset was taken from https://snap.stanford.edu/data/ca-GrQc.html .

Network overview

The network is undirected, meaning that we don't distinguish the direction of an edge and the in degree of each node corresponds to the out degree. This results in a symmetric adjacency matrix which can be observed below

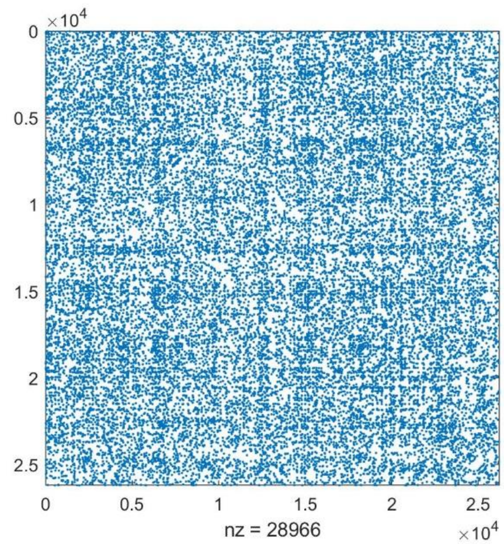

***Figure 1**. Adjacency Matrix*

The main features of this network are collected in the following table:

Nodes	5242
Edges	14483
Min degree	1
Max degree	81

***Table 1**. Main features of the network*

1. Degree distribution, γ estimation and higher moments

The average degree of this network is found to be {<**k**> = 5.528} ∈ (1, ln(N)) indicating that the network is in the *supercritical regime*, which allows to state that a giant component exists. For what concerns the degree distribution I expect a scale-free network with a power-law distribution, as most real networks show these properties. Below the degree distribution plots:

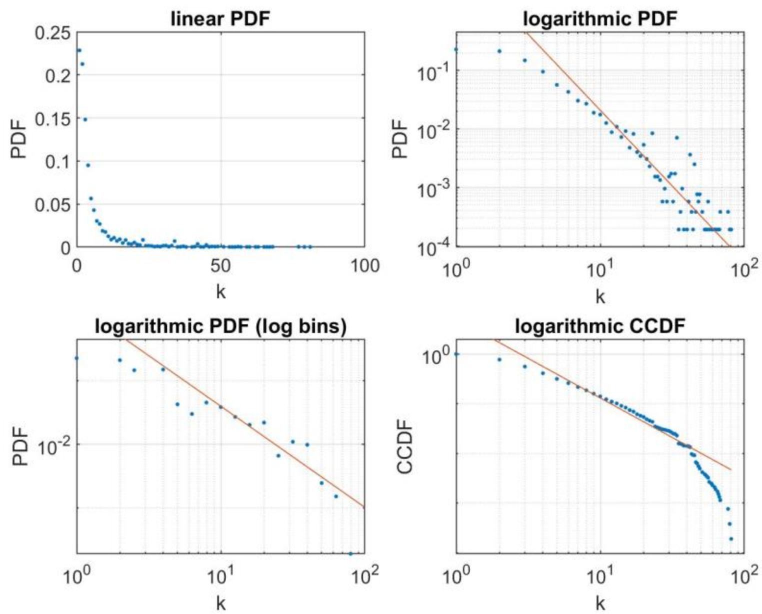

Figure 2. *Degree PDF*

The power-law exponent is obtained performing the Maximum Likelihood estimation, setting K_{min} = 8 (empirically derived from the bottom left plot above) which yields a value $\boldsymbol{\gamma_{ML}}$= 2.592 showing that this network exhibits the *ultra-small world* property. A table is now built for the values of higher moments of the degree distribution.

Variance	62.705
Skewness	3.831
Kurtosis	22.128

Table 2. *Higher degree moments*

The variance is very large as expected from the theory since γ_{ML} ∈ (2,3). The skewness tells how symmetric the distribution is around the average while the kurtosis describes the shape of the distribution.

2. Distance distribution and diameter

At this section the distance between any possible pair of nodes was computed. The diameter is nothing but the maximum of these values and measures 17.

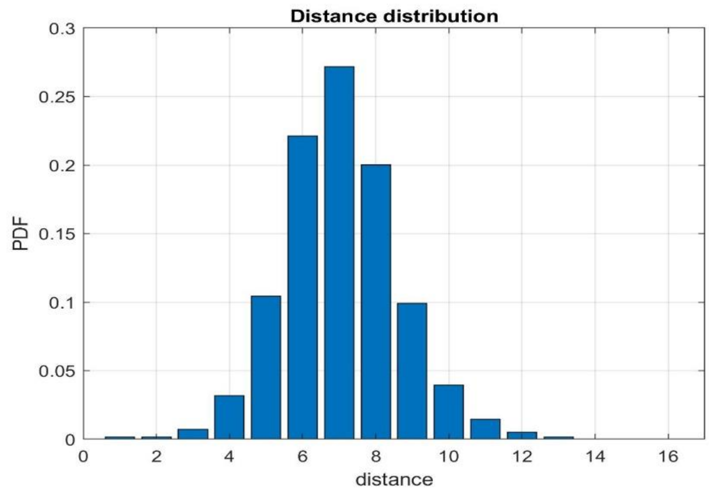

Figure 3. *Distance distribution*

3. Clustering coefficient and its distribution

The clustering coefficient measures the network's local link density. The formula used to compute it is:

$$C_i = \frac{2L_i}{K_i(K_i-1)}$$

where L_i represents the number of links between the K_i neighbors of node i. The average value found is <**C**> = 0.105. Below the plot of the distribution.

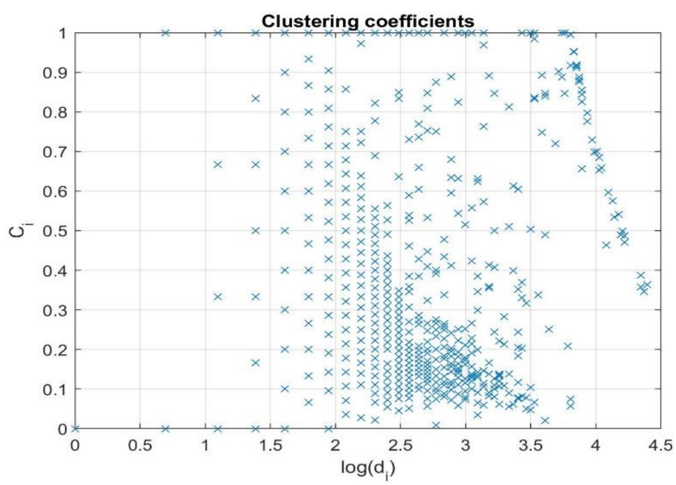

Figure 4. *Clustering coefficient distribution*

From the plot and the average value found, I can state that the network is *weakly connected.*

4. Nearest neighbor degree, structural and natural cutoff, assortativity and assortativity under randomization

The network used for this homework, represents a scientific collaboration, so an assortative behavior is expected. Below, the plot that prove this.

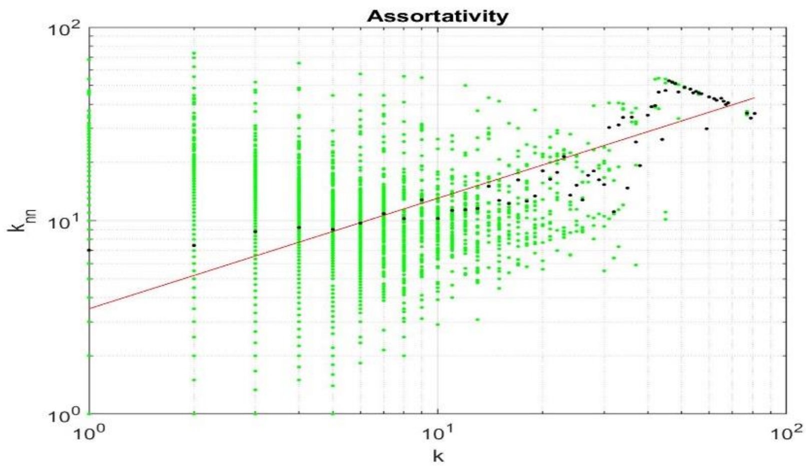

Figure 5. *Assortativity*

In assortative networks high degree nodes tend to connect more to high degree nodes, which in my case means that scientific papers with lots of collaborations are more likely to cite and be cited by other significative collaboration task forces. The assortativity factor found is: **μ** = 0.57.

Recalling the previously estimated value of γ, the *structural* and *natural cutoff* were computed as follows:

$$K_s = (< K > N)^{\frac{1}{2}}$$

$$K_{max} = N^{\frac{1}{1-\gamma}}$$

whose values are:

K_s	170.23
K_{max}	216.95

Table 3. *Structural and Natural cutoff*

Since $K_{max} > K_s$ the network presents the *structural disassortativity* property.

At this point, the following test was performed. The network was randomly rewired preserving the degree of each node. Subsequently the assortativity was recomputed. The plot showing the comparison between the assortativity of the original network and the assortativity under randomization is shown below.

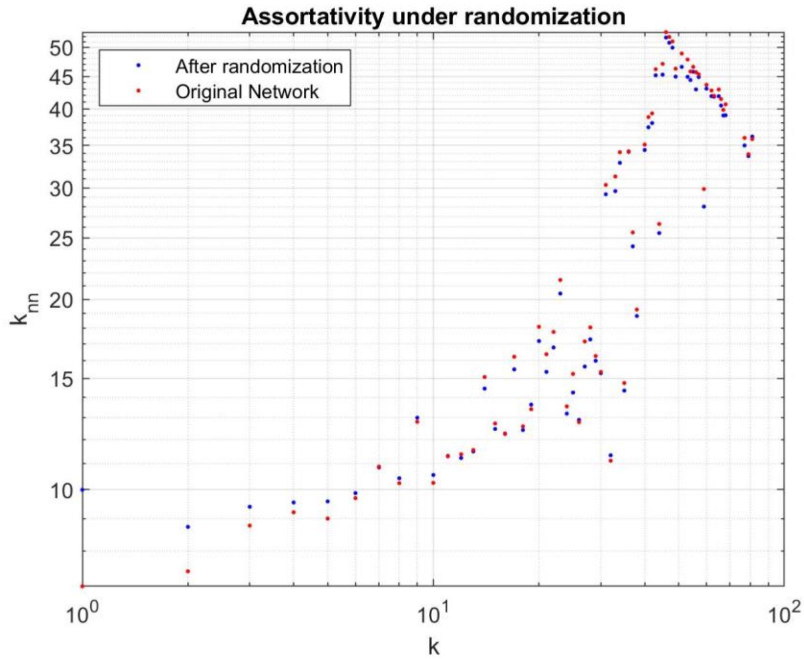

Figure 6. *Assortativity under randomization*

As the plot shows, the structural disassortativity behavior is present both in the original and the randomized network. The new assortativity factor found is: **μ** = 0.52.

5. Robustness to random failures and attacks

The *inhomogeneity ratio* and the *critical threshold,* computed as*:*

$$k = \frac{<k^2>}{<k>}$$

$$f_c = 1 - \frac{1}{k-1}$$

whose values are:

k	11.34
f_c	0.903

Table 4. *Structural and Natural cutoff*

tell us how the network behaves in the following tests. More specifically, when the ratio f_c of nodes is removed from the network, the giant component vanishes.
Robustness under random failures is tested randomly removing a node at each iteration and observing how the size of the giant component vanishes. Robustness to attacks instead, is tested assuming the attacker has some prior information on our network

thus, targeting the node with the highest degree, at each iteration. The comparison of the two tests is shown in the plot below.

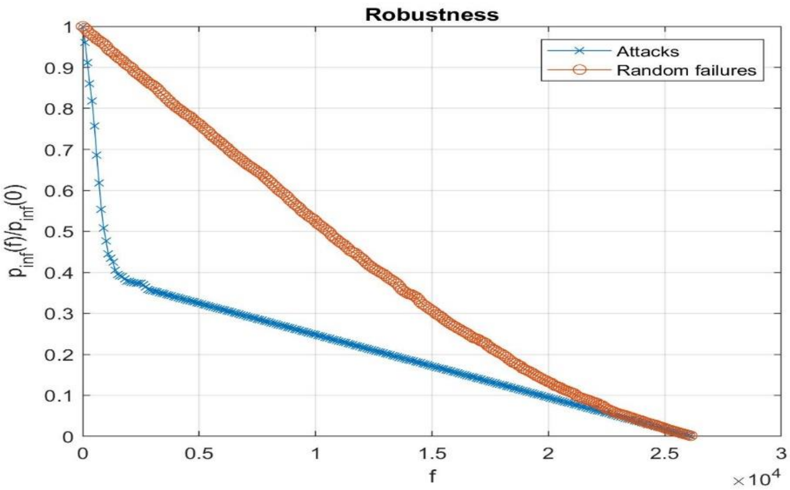

***Figure 7**. Robustness*

As can be seen, the network is robust if random failures occur, while it's not when a malicious attack is performed on it.