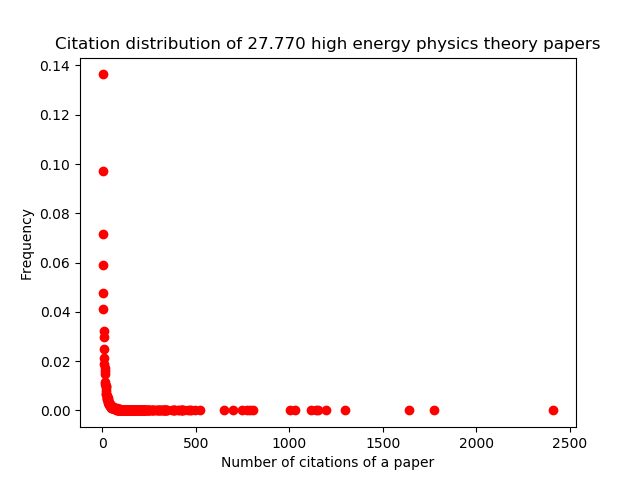
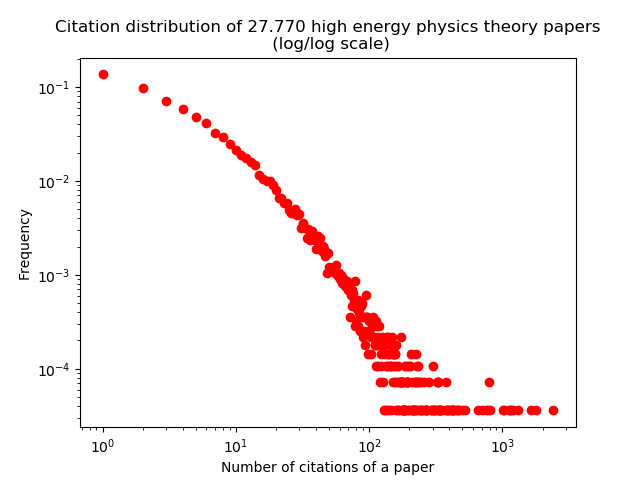
Analysis of Citation Graphs for physics papers

**Question 1: Visuals for the in-degree distribution of a citation graph**

Citation distribution of physics paper in linear/linear and log/log-scales:



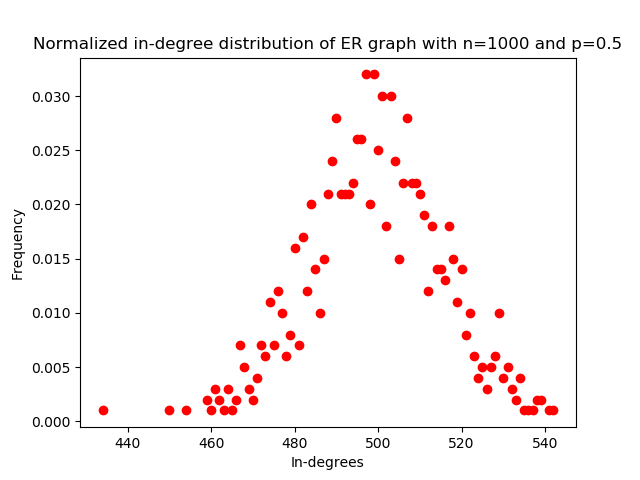
****

**Question 2: ER-algorithm for directed graphs**

1. The expected value of the in-degree is the same for every node in an ER-graph, since the algorithm does not distinguish between nodes. Every node is treated the same way.

The expected value of the in-degree of a node N is easy to calculate: There are n-1 other nodes in the graph, each of which has probability p of having their head in N. Therefore we expect N to have in-degree equal to (n-1)\*p.

1. The expected in-degree distribution for an ER-graph is a binomial distribution with parameters n-1 and p. For sufficiently high number of nodes, the in-degree distution of an ER-graph looks like a normal distribution (i.e. Gauss curve) with mean value = (n-1)\*p. For example, with n=1000 and p=0.5, the distribution of a typical ER graph looks like this (in linear scales):



Note that the minimal and maximal in-degree numbers only differ by about 100.

1. The shape of the in-degree distribution plot for ER does definitely not look like the shape of in-degres distribution for the citation graph. This should be obvious from the above graphs.

The in-degree distribution for citation graphs is approximately linear in log/log-scales, which implies that it behaves like a decreasing power function. In practical terms, this means that the vast majority of physics papers get very few (if any) citations, while there a some outlier paper that get very many citations (these outlies are rare, but not extremely rare, so the in-degree numbers in practice spans several orders of magnitude). As the ER distribution shows, this I NOT the product of a random process (some physics paper are quite simply very important (or ‘hyped’), while most papers get very little attention).

**Question 3:**

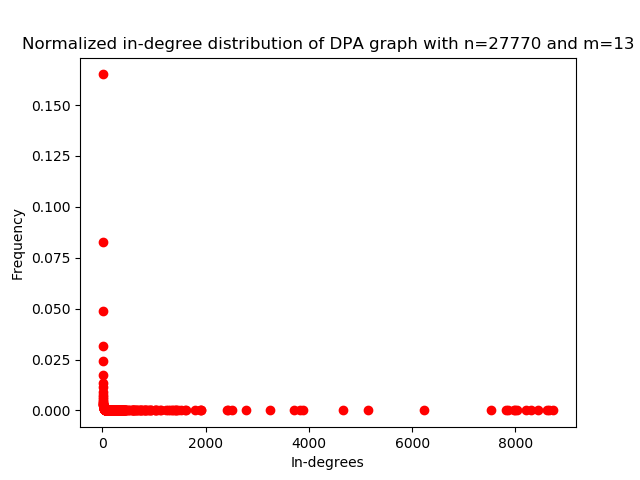
In my implementation of the DPA-graph I will use values:

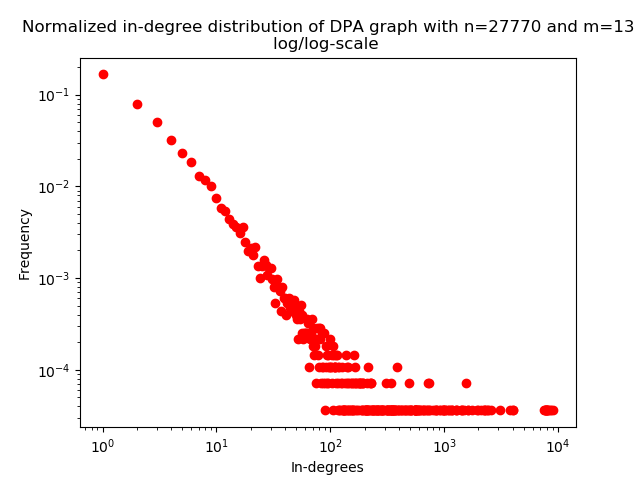
n = 27770 (equal to the number of journals/nodes in the citation graph)

m = 13 (roughly equal to the average out-degree in the citation graph)

**Question 4:**

Here is the normalized in-degree distribution of the DPA graph in linear/linear scale and in log/log scale.





**Question 5**

Interestingly, the DPA graph and the citation (popularity) graph are strikingly similar. The share the same power distribution shape. It is a rich gets richer phenomemon: The more connected you are, the more likely your network is to grow. In the DPA graph, this is is clear from the specifics of the algorithm. Note also, that random events “in the beginning” can have huge effects in the long run.