Heidelberg University
Institute of Computer Science
Database Systems Research Group

Lecture: Complex Network Analysis

Prof. Dr. Michael Gertz

Assignment 4
Scale-Free Networks

`https://github.com/nilskre/CNA_assignments`

Team Member: Patrick Günther, 3660886,
Applied Computer Science
rh269@stud.uni-heidelberg.de

Team Member: Felix Hausberger, 3661293,
Applied Computer Science
eb260@stud.uni-heidelberg.de

Team Member: Nils Krehl, 3664130,
Applied Computer Science
pu268@stud.uni-heidelberg.de

1 Problem 4-1 Power laws

Consider the *degree distribution functions* p_k of the following two undirected networks:

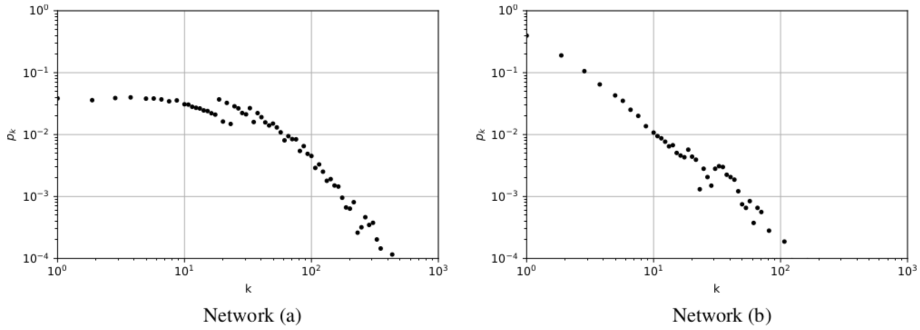

Figure 1: Degree distribution of two networks.

1. One of these networks is approximately scale-free, the other is not. Identify the scale-free network and explain how you came to your conclusion.

   Network b is is scale free. Figure 2 shows the log-log plots of the degree distribution of the two networks a and b. For scale-free networks the degree distribution forms a straight line across the diagonal $p_k \sim k^{-\gamma}$. This is given for network b.

2. A particular network is believed to have a degree distribution that follows a power law. A random sample of nodes is taken and their degrees are measured. The degrees of the first twenty nodes with degrees 10 or greater are:

   16 17 10 26 13 14 28 45 10 12

   12 10 136 16 25 36 12 14 22 10

   For $K_{min} = 10$, estimate the exponent γ of the power law using the estimation method presented in the lecture.

   Hint: you only have to estimate the exponent (see Slide 4-33)!

   Calculate the error σ of your estimation using the equation $\sigma = \frac{\gamma-1}{\sqrt{N}}$

   The following solution is based on Slide 4-28 Estimating the degree exponent. From step 1. (Estimate value of the degree exponent corresponding to K_{min}) following formula is used:

$$\gamma = 1 + N\left[\sum_{i=0}^{N}(ln(\frac{k_i}{K_{min}-\frac{1}{2}}))\right]^{-1} \tag{1}$$

Now the values $N = 20$, k_i from the list above and $K_{min} = 10$ are inserted:

$$\gamma = 1 + 20\left[ln(\frac{16}{20-\frac{1}{2}}) + ln(\frac{17}{20-\frac{1}{2}}) + ln(...) + ...\right]^{-1}$$

$$\gamma \approx 2,531$$

Based on N and γ the error σ is calculated:

$$\sigma = \frac{\gamma - 1}{\sqrt{N}} = \frac{2,531 - 1}{\sqrt{20}} = 0,34$$

Now, one would normally iteratively increase K_{min} from $K_{min} = k_{min}$ to $K_{min} = k_{max}$ and then choose K_{min} such that it minimizes the error σ. Then the resulting power law distribution would fit best to the real distribution. But as the task description states to execute this procedure only for $K_{min} = 10$, we stop here.

2 Problem 4-2 Friendship Paradox

Let p_k be the probability of a randomly chosen node to have the degree k.

If we instead randomly choose an edge, let $q_k = \frac{1}{C} \cdot k \cdot p_k$ be the probability that a node at one of its ends has degree k, where $\frac{1}{C}$ is a normalization factor.

1. Assuming that the network has a power-law degree distribution with $2 < \gamma < 3$, minimum degree k_{min}, and maximum degree k_{max}, show how to derive the normalization factor C.

   Our solution:

   As every probability distribution must add up to 100%, we derive the following equation and solve for C.

$$1 = \int_{k_{min}}^{k_{max}} q(k)\, dk = \int_{k_{min}}^{k_{max}} \frac{1}{C} \cdot k \cdot k^{-\gamma}\, dk$$

$$\Leftrightarrow C = \int_{k_{min}}^{k_{max}} k^{1-\gamma}\, dk = [\frac{1}{2-\gamma} k^{2-\gamma}]_{k_{min}}^{k_{max}} = \frac{k_{max}^{2-\gamma} - k_{min}^{2-\gamma}}{2-\gamma} \tag{2}$$

Solution from the tutorial:

Based on equation 7.3 from slide 7-9 $q_k = \frac{k*p_k}{<k>}$, the parameter C is estimated:

$$q_k = \frac{k * p_k}{\sum_{k'} k' * p_{k'}} = \frac{k * p_k}{< k >} = \frac{k * p_k}{C} \tag{3}$$

Explanation of the equation:

$k * p_k$: A randomly selected edge can connect to a node with k stubs in exactly k different ways. The probability to choose such a node with degree k is p_k.

$\sum_{k'} k' * p_{k'}$: A randomly selected edge can also connect to nodes with other degrees. The sum builds the basic set (dt.: Grundmenge) of the probability calculation to normalize the probability to one.

2. q_k is also the probability that a randomly chosen node has a neighbor with degree k. Show how to compute the average degree of the neighbors of a randomly chosen node.

   Our solution:

   With $C = \frac{k_{max}^{2-\gamma} - k_{min}^{2-\gamma}}{2-\gamma}$ we receive

$$q_k = \frac{(2-\gamma) k^{1-\gamma}}{k_{max}^{2-\gamma} - k_{min}^{2-\gamma}} \tag{4}$$

   We then calculate the 1^{st} moment of the probability distribution q_k:

$$\langle k_{neighbors} \rangle = \int_{k_{min}}^{k_{max}} k \cdot q(k)\, dk = \int_{k_{min}}^{k_{max}} \frac{(2-\gamma) k^{2-\gamma}}{k_{max}^{2-\gamma} - k_{min}^{2-\gamma}}\, dk = \frac{(2-\gamma)}{k_{max}^{2-\gamma} - k_{min}^{2-\gamma}} \frac{1}{(3-\gamma)} [k^{3-\gamma}]_{k_{min}}^{k_{max}} = \frac{(2-\gamma)}{k_{max}^{2-\gamma} - k_{min}^{2-\gamma}} \frac{k_{max}^{3-\gamma} - k_{min}^{3-\gamma}}{(3-\gamma)} \tag{5}$$

Solution from the tutorial:

The average degree of a nodes' neighborhood $< k >$ can be calculated as follows (given in equation 7.9 on slide 7-12):

$$< k_F > = \sum_k k * q_k \tag{6}$$

Now the equation for q_k from the previous subtask is inserted, which results in the following equation (as given in equation 7.9 on slide 7-12):

$$< k_F > = \sum_k k * \frac{k * p_k}{< k >} = \frac{< k^2 >}{< k >} \tag{7}$$

Furthermore σ_k^2 is given from equation 4.16 on slide 4-19

$$\sigma_k^2 = < k^2 > - < k >^2 \tag{8}$$

Move $< k >^2$ to the other side by addition:

$$< k^2 > = < k >^2 + \sigma_k^2$$

Divide by $< k >$:

$$< k_F > = < k > + \frac{\sigma_k^2}{< k >}$$

3. Given a power-law degree distribution network, with $N = 10^4$, $\gamma = 2.3$, $k_{min} = 1$ and $k_{max} = 1000$. Compute the average degree of the neighbors of a randomly chosen node and compare this to $\langle k \rangle$.

   Our solution:

   We now insert $\gamma = 2.3$, $k_{min} = 1$ and $k_{max} = 1000$ into Equation 10 and receive the average degree of for neighbors:

$$\langle k_{neighbors} \rangle = \frac{(2 - 2.3)}{1000^{2-2.3} - 1^{2-2.3}} \frac{1000^{3-2.3} - 1^{3-2.3}}{(3 - 2.3)} \approx 61.23 \tag{9}$$

   If where compare this to the average degree of the node itself using formula (4.9) and (4.17) from the slides of the lecture:

$$\langle k \rangle = (\gamma - 1) k_{min}^{\gamma-1} \frac{k_{max}^{2-\gamma} - k_{min}^{2-\gamma}}{2-\gamma} =$$
$$(2.3-1)1^{2.3-1} \frac{1000^{2-2.3} - 1^{2-2.3}}{2-2.3} \approx 3.79 \tag{10}$$

Solution from the tutorial:

The n-th moment of a degree distribution can be calculated by equation 4.17 from slide 4-20:

$$< k^n >= C * \frac{k_{max}^{n-\gamma+1} - k_{min}^{n-\gamma+1}}{n-\gamma+1} \tag{11}$$

The normalization condition C is calculated as given in equation 4.9 on slide 4-10

$$C = (\gamma - 1) * k_{min}^{\gamma-1} = (2,3-1) * 1^{2,3-1} = 1,3 \tag{12}$$

Insert all known values and calculate:

$$< k^2 >= 1,3 * \frac{1000^{2-2,3+1} - 1^{2-2,3+1}}{2-2,3+1} = 231,94 \tag{13}$$

Use $< k >$ from equation 10:

$$< k >= 3,79 \tag{14}$$

Insert $< k^2 >$ and $< k >$ into equation 7:

$$< k_F >= \frac{< k^2 >}{< k >} = \frac{231,94}{3,79} = 61,234 \tag{15}$$

4. Try to explain the paradox of subtask 3, i.e., explain why on average, the neighbors of a node have more neighbors than the node itself?

   This is known phenomenom from sociology that can be explained as individuals tend less to be friends with someone who has very few friends. Also one might argue that most persons are friends with at least on "hub" in society, a friend with very much other friends. This "hub" pushes the average friendship degree of ones friends even further away from ones own friendship degree (like an "outlier" in the sample).

3 Problem 4-3 Network Measures of Real Graphs

See *problem_3.ipynb*.

1. The diameter of the largest connected component is 15.
2. The node with the highest degree hast the ID 2332 and a degree of 1098.
3. The number of triangles in the network is 3501542.
4. The global clustering coefficient of the network is approximately 0.15.
5. The power-law exponent of the degree distribution is approximately 0.79..., which is wrong. It should be arround 1.43
6. See the following figure:

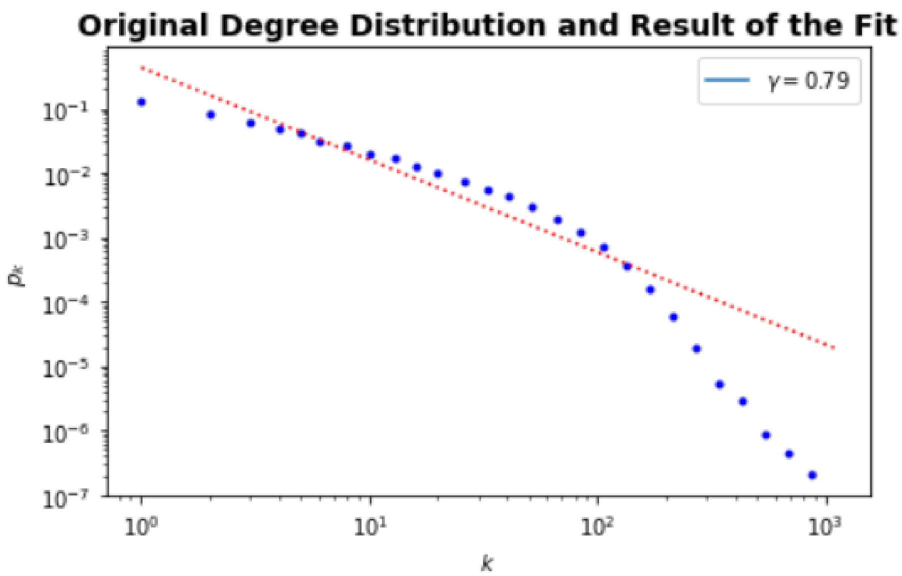

Figure 2: Real and fitted degree distribution of facebook links.