Social Networks & Recommendation Systems

IV. Network metrics.

Grzegorz Siudem

Warsaw University of Technology

Warsaw University of Technology

MSc program in Data Science has been developed
as a part of task 10 of the project
„NERW PW. Science - Education - Development - Cooperation”
co-funded by European Union from European Social Fund.

Project

Power law in the real data - case study

Excercise 1.

Import the data, draw their histogram.

Power law in the real data - case study

Excercise 1.

Import the data, draw their histogram.

Excercise 2.

Correct the scale of the plot's axes to double logarithmic.

Power law in the real data - case study

Excercise 1.

Import the data, draw their histogram.

Excercise 2.

Correct the scale of the plot's axes to double logarithmic.

Excercise 3.

Apply logarithmic binning as well.

Power law in the real data - case study

Excercise 1.

Import the data, draw their histogram.

Excercise 2.

Correct the scale of the plot's axes to double logarithmic.

Excercise 3.

Apply logarithmic binning as well.

Excercise 4.

Compute and draw the survival function.

Power law in the real data - case study

Excercise 1.

Import the data, draw their histogram.

Excercise 2.

Correct the scale of the plot's axes to double logarithmic.

Excercise 3.

Apply logarithmic binning as well.

Excercise 4.

Compute and draw the survival function.

Excercise 5.

Which chart is the most readable? Which is most robust for the noise?

Power law in the real data - case study continuation

Excercise 6.

Use the linear regression to the previous plots to determine the parameter α estimator.

Power law in the real data - case study continuation

Excercise 6.

Use the linear regression to the previous plots to determine the parameter α estimator.

Excercise 7.

Compute (formula and value for our data) MLE estimator for α with aknown $x_{\min}$ assuming following distribution

$$\mathcal{P}(x) = \frac{\alpha - 1}{x_{\min}} \left(\frac{x}{x_{\min}} \right)^{-\alpha}.$$

Power law in the real data - case study continuation

Excercise 6.

Use the linear regression to the previous plots to determine the parameter α estimator.

Excercise 7.

Compute (formula and value for our data) MLE estimator for α with aknown $x_{\min}$ assuming following distribution

$$\mathcal{P}(x) = \frac{\alpha - 1}{x_{\min}} \left(\frac{x}{x_{\min}} \right)^{-\alpha}.$$

Excercise 8.

How will the result change if we do not know $x_{\min}$?

Power law in the real data - case study continuation

Excercise 6.

Use the linear regression to the previous plots to determine the parameter α estimator.

Excercise 7.

Compute (formula and value for our data) MLE estimator for α with aknown $x_{\min}$ assuming following distribution

$$\mathcal{P}(x) = \frac{\alpha - 1}{x_{\min}} \left(\frac{x}{x_{\min}} \right)^{-\alpha}.$$

Excercise 8.

How will the result change if we do not know $x_{\min}$?

Warning!

Excercises 1-8 in total are worth 1*P* for the project.

Case study – summary

P4.0 Excercises 1-8. [1P]

P4.1 Collect obtained scripts for fat-tailed data analysis in one file. We will use them again and again. [1P]

Nearest neighbors degree in practice

P4.2 Prove (writing formulas) that for uncorrelated network $\langle k \rangle_{nn}(k_i) = \frac{\langle k^2 \rangle}{\langle k \rangle}$. [1P]

P4.3 Determine (empirically) the dependence of the average degree of the nearest neighbor on the degree of the vertex for selected real or artificial networks. [1P]

P4.4 Check how random edge switching affects the result of the previous task. [1P]

Correlation coefficient in practice

P4.5 Make the derivation omitted on the lecture slide. [1P]

P4.6 Find the correlation coefficient for the network from task P4.3-P4..4 (before and after edge switching). [1P]

What real networks are?

Let's check!

Network's name	$\leftrightarrow$?	N	E	$\langle k \rangle$	α	ℓ	r
...	...	...	...	...	...	...	...

P4.7 Let's fil the above table with metrics od selected real networks, use the built-in functions for counting the parameters. Compare the results with the literature. [2.5P]

Erdős numer – empirical data

Erdős Number Project

`oakland.edu/enp/`

P4.8 Analyse the data on the ENP website. Draw histograms of the Erdős number for the Nobel prize and the Fields medal laureates. [1P]

Inspiration for the final project?

Bacon number

Oracle of Bacon

`oracleofbacon.org`

P4.9 Check the Bacon number of selected actors. Draw a histogram of Bacon number among Oscar winners. Who will find the actor with the highest Bacon number? [1P]

Inspiration for the final project?

Warsaw University of Technology

MSc program in Data Science has been developed
as a part of task 10 of the project
„NERW PW. Science - Education - Development - Cooperation”
co-funded by European Union from European Social Fund.