

Übung: Statistik in Netzwerken

1.) Pearson Korrelation (Linearer Zusammenhang)

Sie lesen von der Behauptung, dass Geld ein wichtiger Faktor für einen hohen sozialen Status ist. Sie möchten dies mit Hilfe von Statistik und dem Wissen aus Sozialer Netzwerkanalyse überprüfen. Im Modul Soziale Netzwerkanalyse wurde erwähnt, dass das Ansehen (Prestige) zu den Social Forces zählt, wieso man sich mit jemanden verbindet. Deshalb haben Sie sich entschieden, Freunde nach ihrem Vermögen zu fragen. Gehen Sie davon aus, dass ihre Freunde ehrlich zu ihnen sind.

Datensatz	Vermögen (in CHF 1'000)	Anzahl Freunde auf Facebook
1	50	100
2	54	120
3	30	210
4	66	90
5	41	120
6	11	10
7	20	55
8	100	200

a.) Berechnen Sie den Pearson Korrelationskoeffizienten für den Datensatz, einmal mit dem Ausreisser und einmal ohne den Ausreisser. Nehmen Sie am besten Excel zur Hilfe, wie im Script angegeben.

Pearson Korrelations-Koeffizient mit Ausreisser:

Pearson Korrelations-Koeffizient ohne Ausreisser:

c.) Sie haben einen starken linearen Zusammenhang zwischen Vermögen und den Anzahl Freunden auf Facebook gefunden. Ist diese allgemein gültig? Begründen Sie.

2.) Spearman Korrelation (Rangkorrelation)

Gegeben ist ein Datensatz¹, wo der IQ und die Anzahl Stunden, die eine Person vor dem TV verbringt, gegenüberstellt:

Datensatz	IQ	Stunden vor dem TV(pro Woche)
1	106	7
2	86	0
3	100	27
4	101	50
5	99	28
6	103	29
7	97	20
8	113	12
9	112	6
10	110	17

Berechnen Sie den Rangkorrelations-Wert mit der im Script angegebenen Formel:

$r_s =$

3.) Definition von Hypothesentests

Gegeben sei ein Co-Authorship Netzwerk, bei welchem abgebildet ist, welcher Wissenschaftler zusammen mit welchen anderen Personen Publikationen geschrieben hat.

Es handelt sich dabei um ein One-Mode Netzwerk mit den Attributen:

Knotenattribute:	Kantenattribute
Name des Autors	Anzahl gemeinsame Publikationen
Universität	
Anzahl Auszeichnungen	
Durchschnittliche Anzahl Veröffentlichungen / Jahr	

Zudem haben sie ein zweites Netzwerk, in welchem dieselben Autoren die Knoten darstellen und die Kanten die privaten Freundschaftsbeziehungen.

Definieren Sie für den gegebenen Anwendungsfall drei Hypothesentests:

- Einen Monadic Hypothesentest: (Ein Knoten-Attribut soll dabei eine Netzwerk-Metrik sein)
- Einen Dyadic Hypothesentest
- Einen Mixed Monadic-Dyadic Hypothesentest. (Ein Knoten-Attribut soll dabei die Basis für ein neues Netzwerk sein.

¹ Quelle: https://en.wikipedia.org/wiki/Spearman%27s_rank_correlation_coefficient, 28.07.2019

4.) Permutations-Tests & QAP in Python

Gehen Sie zu den Jupyter Notebooks und dann in den Unterordner zum Thema Statistik in Netzwerken. Da hat es einen Ordner «Exercises» und für die Lösungen einen Ordner «Solutions». Im Ordner «original source» ist der Quellcode der verwendeten QAP Library mit Beispiel-Notebook. In diesen Beispiel-Notebooks ist ersichtlich, wie QAP angewendet wird.

Zu den Aufgaben:

- a.) Permutations-Test: Öffnen Sie im Ordner «Exercises» das Notebook «Florentine_Permutation». Es geht darum, dass in einem Netzwerk mit Hilfe von Permutations-Tests für verschiedene Zentralitätsmasse untersucht wird, ob ein signifikanter Zusammenhang zwischen dem Vermögen und dem Zentralitätsmass existiert. Ziemlich vieles ist bereits umgesetzt. Das Einzige, was Sie noch implementieren müssen ist die Methode zur Berechnung der initialen Korrelation wie auch des p-Wertes (eine Methode, welche ein Tupel zurückgibt). Die zu implementierende Zelle beginnt mit TODO und am Schluss in der Resultat-Zelle ist ebenfalls nochmals ein TODO, wo es darum geht, das Resultat zu interpretieren. Wenn Sie die fehlende Methode implementiert haben, können Sie das Notebook ausführen.
- b.) QAP Test: Öffnen Sie im Ordner «Exercises» das Notebook «Schoolclass_QAP». Hier geht es darum zu untersuchen, ob die Klassenzugehörigkeit signifikant mit dem Kommunikationsverhalten zwischen Schulkindern korreliert. Die Umwandlung von der Knoten-Attributs «Klassenzugehörigkeit» ist schon als Kante in einem neuen Graphen modelliert. Auch sind beide Adjazenzmatrizen bereits als Numpy-Array vorhanden. Sie müssen lediglich noch die drei letzten Zellen ergänzen, wo es darum geht, QAP auszuführen, das Resultat anzuzeigen und zu interpretieren.