SDS Autumn 2019, M2, Assignment 1

Daniel S. Hain (dsh@business.aau.dk) & Roman Jurowetzki (roman@business.aau.dk)

Introduction

In your first assignment in M2, you will replicate a well known network analysis, with different data and some twists. The data is to be found at: https://github.com/SDS-AAU/SDS-master/tree/master/00_data/network_krackhard (Hint: You neet to download the `raw` data)

Data: What do I get?

Background

Let the fun begin. You will analyze network datacollected from the managers of a high-tec company. This dataset, originating from the paper below, is widely used in research on organizational networks. Time to give it a shot as well.

Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

The company manufactured high-tech equipment on the west coast of the United States and had just over 100 employees with 21 managers. Each manager was asked to whom do you go to for advice and who is your friend, to whom do you report was taken from company documents.

Description

The dataset includes 4 files - 3xKrack-High-Tec and 1x High-Tec-Attributes.

Krack-High-Tec includes the following three 21x3 text matrices:

- ADVICE, directed, binary
- FRIENDSHIP, directed, binary
- REPORTS_TO, directed, binary

Column 1 contains the ID of the **ego** (from where the edge starts), and column 2 the **alter** (to which the edge goes). Column 3 indicates the presence (=1) or absence (=0) of an edge.

High-Tec-Attributes includes one 21x4 valued matrix.

- ID: Numeric ID of the manager
- AGE: The managers age (in years)
- TENURE: The length of service or tenure (in years)
- LEVEL: The level in the corporate hierarchy (coded 1,2 and 3; 1 = CEO, 2 = Vice President, 3 = manager)
- DEPT: The department (coded 1,2,3,4 with the CEO in department 0, ie not in a department)

Instructions: What shall I do?

1. Create a network

- Generate network objects for the companies organizational structure (reports to), friendship, advice
- This networks are generated from the corresponding edgelists
- Also attach node characteristics from the corresponding nodelist

2. Analysis

Make a little analysis on:

A: Network level characteristics

Find the overal network level of:

- Density
- Transistivity (Clustering Coefficient)
- Reciprocity

... for the different networks. Describe and interpret the results. Answer the following questions:

- Are relationships like friendship and advice giving usually reciprocal?
- Are friends of your friends also your friends?
- Are the employees generally more likely to be in a friendship or advice-seeking relationship?

B: Node level characteristics

Likewise, find out:

- Who is most popular in the networks. Who is the most wanted friend, and advice giver?
- Are managers in higher hirarchy more popular as friend, and advice giver?

C: Relational Characteristics

Answer the following questions:

- Are managers from the same 1. department, or on the same 2. hirarchy, 3. age, or 4. tenuere more likely to become friends or give advice? (hint: assortiativity related)
- Are friends more likely to give each others advice?

3. Aggregated Networks

Reconstruct the advice and friendship network on the aggregated level of departments, where nodes represent departments and edges the number of cross departmental friendships/advice relationships.

4. Visualization

Everything goes. Show us some pretty and informative plots. Choose what to plot, and how, on your own. Interpret the results and share some insights.

Deliverables

Please submit a PDF or HTML version of your notebook on peergrade.io (if you submit HTML, please zip it before - large embedded HTMLs from cause crashing when oppened directly in peergrade). In adittion, feel free to include a colab link (non-mandatory)). Please make sure it runs without errors and others can access it (i.e. own test in “anonymous” setting in your browser) .

This notebook should:

- It should solve the questions in an straightforward and elegant way.
- It should contain enough explanations to enable your fellow students (or others on a similar level of knowledge) to clearly understand what you are

doing, why, what is the outcome, how to interpret it, and how to reconstruct the exercise. Be specific and understandable, but brief.

Further process and dates

- You will receive an upload link on peergrade.io with concrete instructions.
- The notebook upload is also due 15.10.2020, 11:55pm. Delays are not accepted.
- After the upload deadline, you will recieve an invitation to peergrade your fellows' exams on peergrade.io. You will be asked for the evaluation of 3 peer-assignments is part of the assignment and mandatory.

Hints

General

- Keep in mind, you are looking in all cases at directed and unweighted networks.
- To create networks on a higher lecel of aggregation, you want to aggregate the edgelist the corresponding level before creating a network.

R

- Use `fread` from the `data.table` package to conveniently read the files
- There are packages and procedures for node and edge aggregation in `R` on graph objects, but the easiest way would be to use `dplyr` to already perform aggregations on the dataframes before creating a graph object.

Python

- Use Pandas for aggregations (if needed) and then create new graphs
- Use the `Networkx` documentation/reference (e.g. assortativity) to look up the functionality of algorithms that have been covered in the lectures but not explicitly in Python examples or DC courses.