

Social Network Analysis

Matthieu Cisel

November 2021

1 Goals

In this class, students will learn the various concepts surrounding network analysis – vertices, nodes, centrality, etc. – and how to design graphs to represent these networks. In terms of activity, learners will engage in an ongoing research on social networks in Academia, and more specifically, on the composition of various types of juries. In the first steps of the learning unit, students have to provide a few insights into a set of a few dozens of juries, using the R programming language

2 Introduction

The assignment corresponds this time to a slideshow including all the graphs that are asked here. Both the aesthetics and the structure (titles) of the PPT will be central to your grade. Avoid putting too much text (more than 4 sentences) per slide. You can intertwine text slides and graph slides.

You have been given access to a database on recruitment juries constituted by two students of the class during their 2020 internship. It is constituted of the names of each member of the jury of Maître de Conférences (MCF, equivalent of an assistant professor) for four consequent years, from 2017 to 2020. The goal of this learning unit is to realize data visualizations in order to get a better understanding of the social networks that exist within higher education. While the expectations of the first assignment are rather precise, the scaffolds will fade at each assignment, in order to give you as much agency as possible. Document and organize your code in a proper manner, you may be graded on that specifically. Submit the PDF version (with graphs) of your R Notebook.

Follow the Network analysis with R class in Datacamp. Submit the certificate in Teams.

3 Simple SNA

Create a subset of the dataset and represent a bipartite and a multipartite graph of internal members, and the same thing external members (do not mix years) :

1. 2017, internal members
2. 2017, external members
3. 2018, internal members
4. 2018, external members
5. 2019, internal members
6. 2019, external members
7. 2020, internal members
8. 2020, external members

Use `gganimate` (in R, or any Python equivalent) to create a GIF-based succession of graphs for the years 2017 to 2020, for external members only. Make an interactive graph of one year (internal members) with `networkD3` (or the Python equivalent).

Explain what is the difference between a bipartite or a multipartite graph. Explain the difference between an edge list and a adjacency matrix. Describe the graphs that you created, and interpret it after a conversation with your instructor regarding how recruitment juries work.

Apply a community detection approach to detect communities in one of the graphs with only external members.

Create a last graph where all years are pooled together, but only for external members. What is the problem with this graph ?

Choose the relevant dataset to compute the following indicators. Identify the top ten members (create a table per indicator) with the highest values for the following indicators : degree, strength, betweenness. Create the plots but this time, the width of the nodes should depend upon the value of the centrality indicators.

Explain how these indicators are built. What is the difference between node and edge betweenness. Make a graph where you plot the 3 edges with the highest betweenness.

4 Sankey diagram and Chord plot

Go on R or Python graph gallery, first copy the code from the chord plot, then from the Sankey plot. Just edit the names of the variables / labels so that the graphs will reflect an interesting phenomenon in higher education (the phenomenon of your choice). Create an interpretation of your graph and a quizz along the description to test if the reader has understood your interpretation.

5 Association Rule Mining

Follow the class "Market Basket Analysis (R or Python)" in Datacamp. Submit the certificate in teams.

Use the TEL.domain.mat dataset on discipline tags (mathematics, computer science) on more than 100.000 PhD manuscripts. This time, we will use approaches from the market basket analysis toolkit.

Explain the concept of lift in your own words. Apply a market basket analysis to plot the association rules between different disciplines. Create all of the graphs suggested in the Datacamp class that seem relevant in this case. Describe one of the graphs that you created, and try to interpret it.