Social Network Analysis

Date : 02/12/2021

By : Sarvesh Meenowa

Introduction

- Understanding the social networks that exist within higher education, specifically in recruitment juries.
- Perform data vizualisation on internal and external jury members 2017
- Interpretation of the results
- Exploring centrality indicators
- Difference between:
  - Edge betweenness and node betweenness
  - Edge list and adjacency matrix
  - Bipartite and multipartite graphs
- Lift metric

Methodology

TOOLS :

- R version 4.0.3
- Scraped data from theses.fr
- Igraph package
- NetworkD3 package

Techniques:

- Interactive Network Visualization using R
- Network Analysis Manipulation using R
- Market basket Analysis(MBA)

Results I - Internal Jury 2017

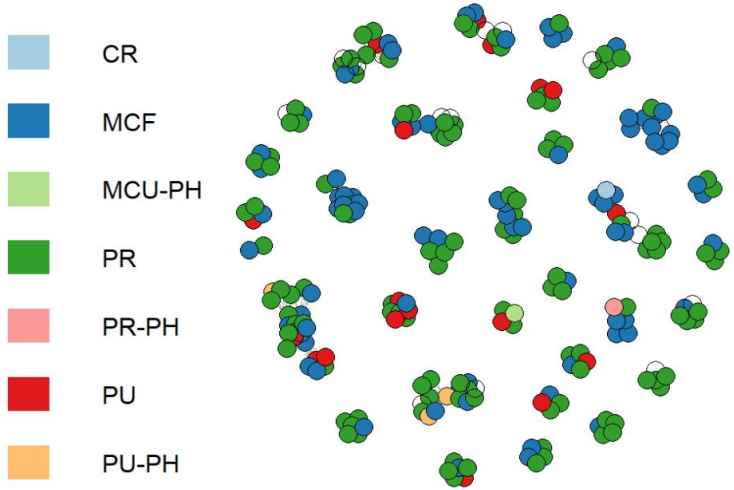

Figure 1 : Multipartite graph of internal jury members from the year 2017 where the nodes are colored by the levels of the jury.

Results II - External jury 2017

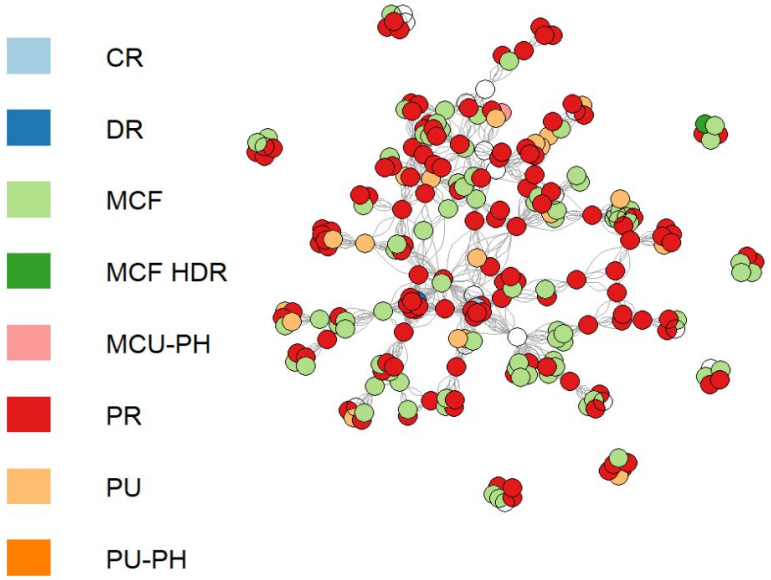

Figure 2 : Multipartite graph of external jury members from the year 2017 where the nodes are colored by the levels of the jury.

Results III - Node centrality indicators

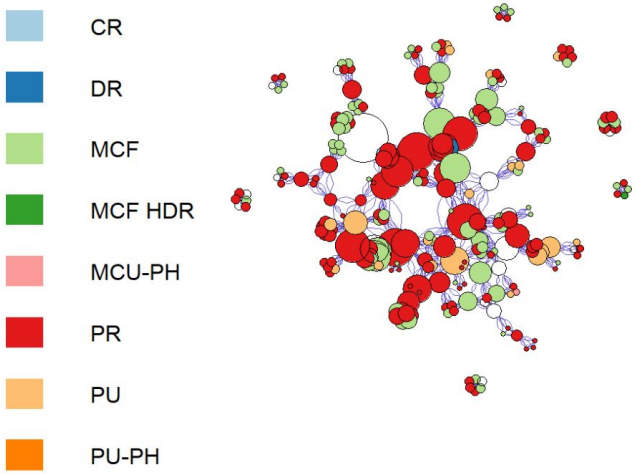

Figure 3.1 : Multipartite graph for external jury members 2017 where node size depends on degree centrality.

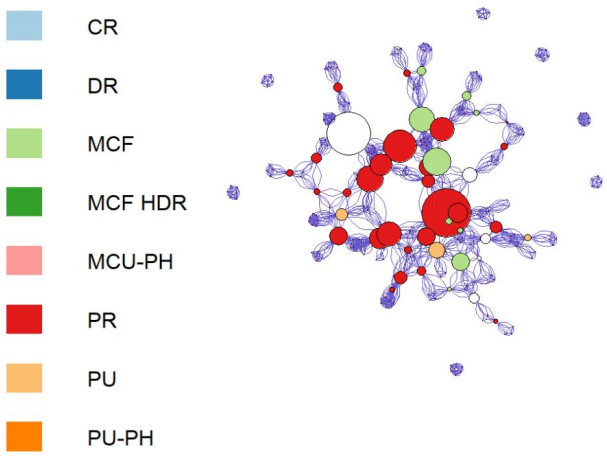

Figure 3.2 : Multipartite graph for external jury members 2017 where node size depends on betweenness centrality.

Results IV -Degree centrality table for external members 2017

Names	Degree	Level
Denise ORANGE RAVACHOL	52	PR
Caroline DESOMBRE	40	MCF
Fabienne BRIERE-GUENOUN	40	MCF
Dominique BERGER	38	PR
Nicole BIAGIOLI	36	PR
Bruno GARNIER	36	PR
Eric FLAVIER	34	PR
Pierre Périer	34	PR
Nassira HEDJERASSI	32	PR
Line NUMA BOCAGE	30	PR

Key:
PR : Professeurs des universités
MCF: Maitre de conférence

Table 1 : External jury members 2017 with their corresponding levels arranged in descending order of degree centrality.

Discussion I - Internal Jury 2017

- From Figure 1, smaller networks inside the internal multipartite graph represent small institutions
- One small network ⇒ has at least 1 PR(professeurs des universités) ⇒ highest ranked jury position
- Bigger networks/more interconnections ⇒ Bigger institutions / possibly a jury with high centralities / could also be partner institutions

Discussion II - Graphs and tables interpretation

- From Figure 2, for external jury members, they move from one institution to different institutions
    - The multipartite graph is therefore more connected ⇒ more links
    - Mostly red nodes ⇒ PR level ⇒ mostly high ranked professors move more between institutions

Discussion III - Graphs and tables interpretation

- From Figure 3.1 and 3.2, bigger size nodes ⇒ jury members with high centralities
- Mostly red nodes ⇒ jury members with PR(professeurs des universités) level
    - Higher ranked members ⇒ higher centralities/more relations ⇒ main actors of the network
- Also from Table 1 ⇒ external jury members with highest centralities ⇒ mostly PR level

Explain what is the difference between a bipartite or a multipartite graph.

- Graphs with 2, 3 or multiple « kinds » of nodes
  - 2 kinds of nodes, it’s a Bipartite Graph (or 2-partite)
    - Example : Social Network : Persons < - > Companies
  - 3 kinds of nodes it’s a 3-partite Graph (Multipartite , k > 2)
    - Example : Movie Network : Actors < - > Movies < - > Movie Companies
  - k kinds of nodes it’s a k-partite Graph

Explaining the difference between an edge list and a adjacency matrix

- **Adjacency matrix:** $A_{ij} = 1$ if and only if there is a link from j to i.
- **Edge list :** keeps track of all the nodes and edges forming list of pairs (from source to target, sometimes with weight)
- Example : Consider the network in Figure 4 :

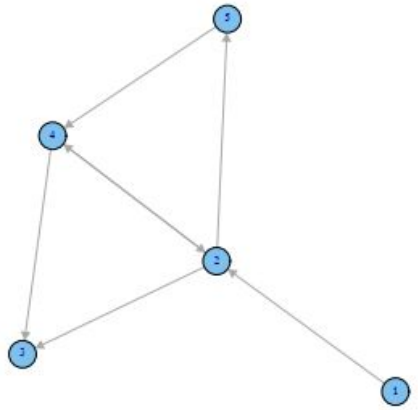

Figure 4 : Directed network example

Adjacency matrix:

$$\begin{bmatrix} & 1 & 2 & 3 & 4 & 5 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 & 1 & 0 \\ 3 & 0 & 1 & 0 & 1 & 0 \\ 4 & 0 & 1 & 0 & 0 & 1 \\ 5 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

Edge list(unweighted):

(1,2)
(2,3)
(2,4)
(4,2)
(2,5)
(4,3)
(5,4)

Explaining how Centrality indicators are built

- **Degree centrality :** number of edges connected to a given node
  - **Consider centrality of matrix A :**

$$C_D(j) = \sum_{j=1}^{n} A_{ij}$$

- **Betweenness Centrality :** number of geodesic paths (shortest paths) that go through a given node
  - **where σ(k, ℓ, i)is the number of shortest paths from nodes k to ℓ that go through node i, and σ(k, ℓ)is the number of shortest paths from nodes k to ℓ.**

$$B_i \triangleq \sum_{\substack{k \subset \mathcal{V} \\ k \neq i}} \sum_{\substack{\ell \subset \mathcal{V} \\ \ell \neq i,k}} \frac{\sigma(k, \ell, i)}{\sigma(k, \ell)},$$

- **Strength centrality :** sum of the weights of edges connected to the node

What is the difference between node and edge betweenness?

- **Edge betweenness** : the number of the shortest paths that go through an edge in a graph or network (Girvan and Newman 2002)

- **Node betweenness** : the number of the shortest paths that go through a node in a graph or network

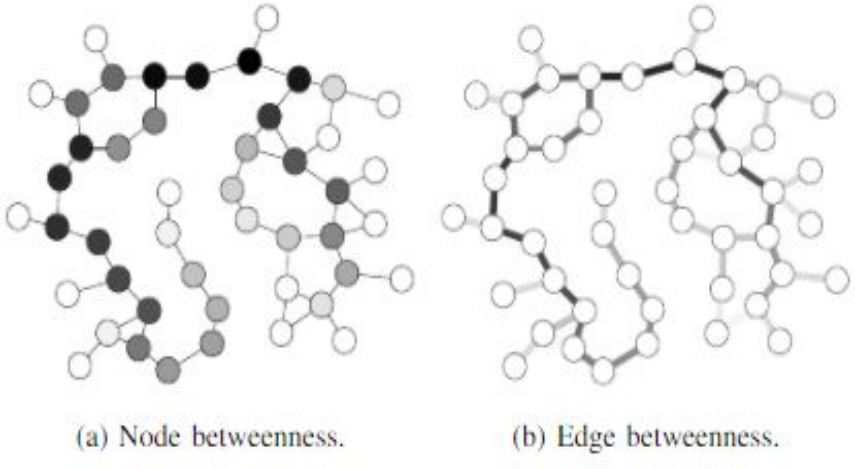

Figure 5 : Edge and Node betweenness

What is lift ?

- Metric for assessing the relationship between items.
- **Numerator**: Proportion of transactions that contain X and Y.
- **Denominator**: Proportion if X and Y assigned randomly and independently.

$$Lift(X \rightarrow Y) = \frac{Support(X \rightarrow Y)}{Support(X)Support(Y)}$$

- ***Lift* = 1** $\Rightarrow$ no relationship between X and Y (i.e., X and Y together only by chance)
- ***Lift* > 1** $\Rightarrow$ *X has increased the probability that Y will occur on this transaction*
- ***Lift* < 1** $\Rightarrow$ *X has decreased the probability that Y will occur on this transaction*

References

M, Girvan. *Girvan, M. & Newman, M. E. J. Community Structure in ...* https://www.researchgate.net/publication/11313118_Girvan_M_Newman_M_E_J_Community_structure_in_social_and_biological_networks_Proc_Natl_Acad_Sci_USA_99_7821-7826.

Kassambara. “Network Analysis and Manipulation Using R.” *STHDA*, 28 Nov. 2017, http://www.sthda.com/english/articles/33-social-network-analysis/136-network-analysis-and-manipulation-using-r/.

Shizuka, Dai. “Measuring Networks Part 1: Centrality and Global Measures.” *Dshizuka.github.io*, https://dshizuka.github.io/networkanalysis/04_measuring.html.

Ognyanova, Katya. *Network Analysis with R and Igraph: Netsci X Tutorial*. 27 May 2020, https://kateto.net/networks-r-igraph.