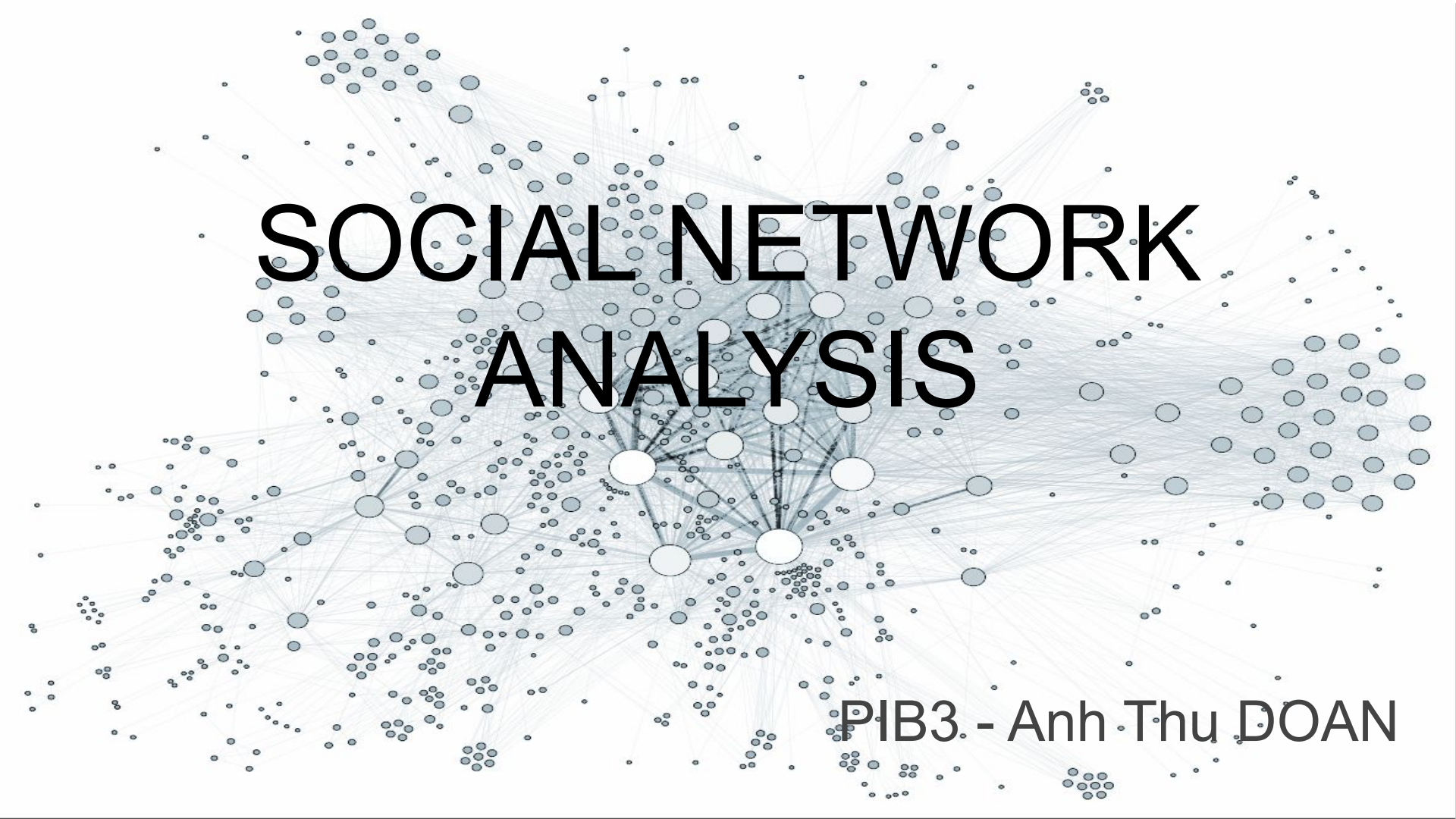


SOCIAL NETWORK ANALYSIS

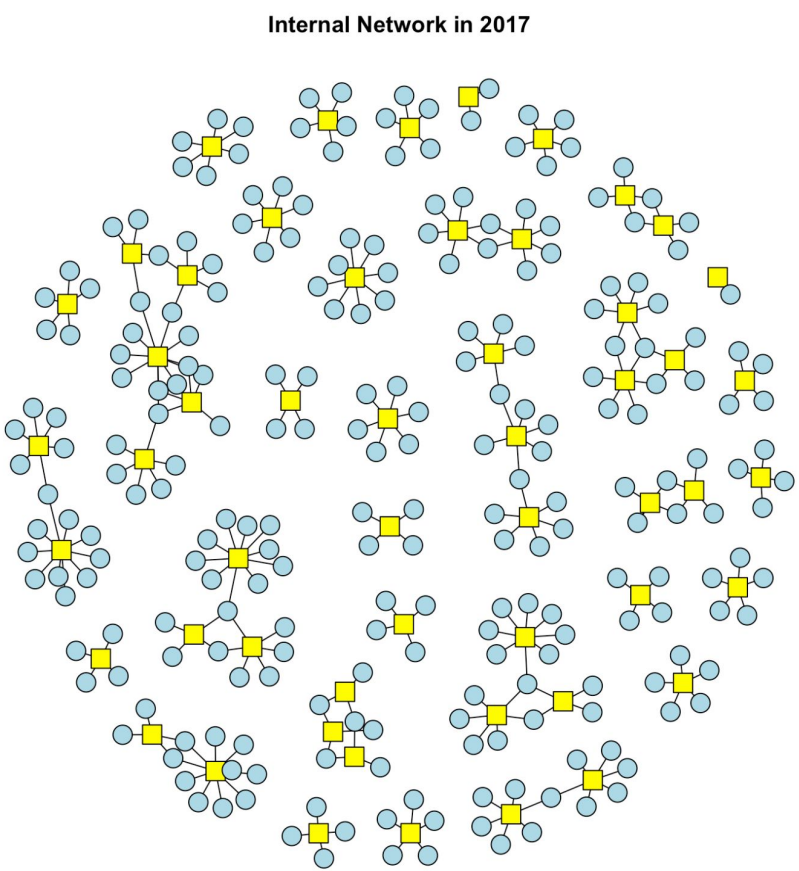
PIB3 - Anh Thu DOAN

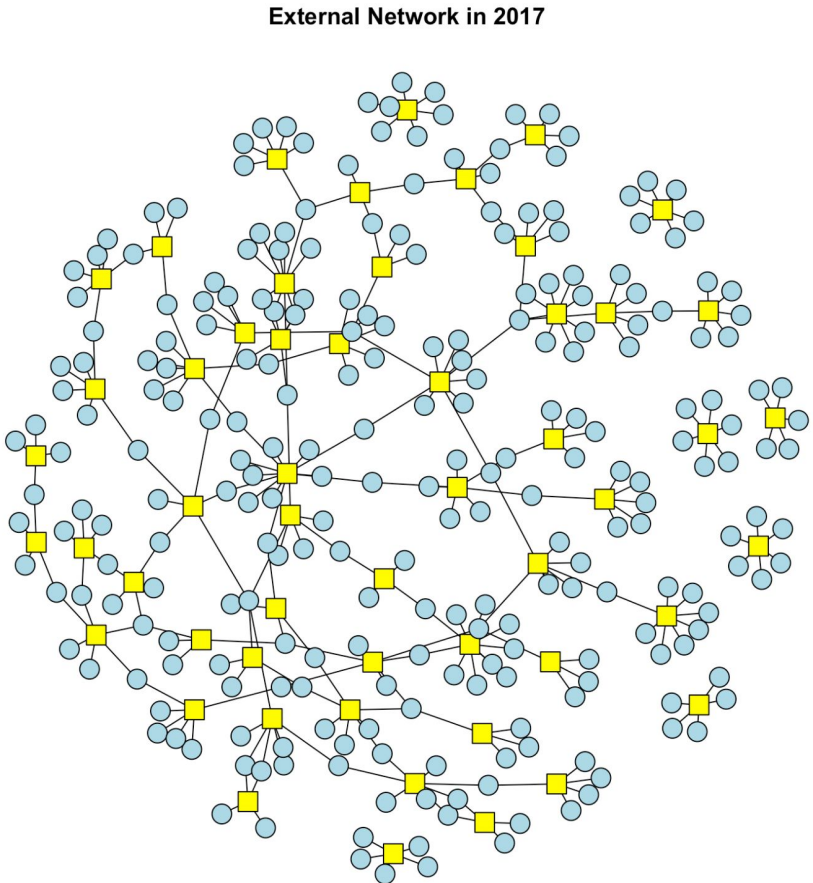
Introduction

- The data is constituted of each member of the jury of Maître de Conferences (MCF) from 2017 to 2020.

- To understand a community by mapping the relationships that connect them as a network and then drawing out key individuals, groups, or associations between the individuals.

Simple SNA





Interpretation

- Internal network was created by small cluster which only connected together with the same number of post.

- In external network was also gather by small cluster but not only the same number of post connected together but also with the different number of post

Bipartite graph VS Multipartite graph

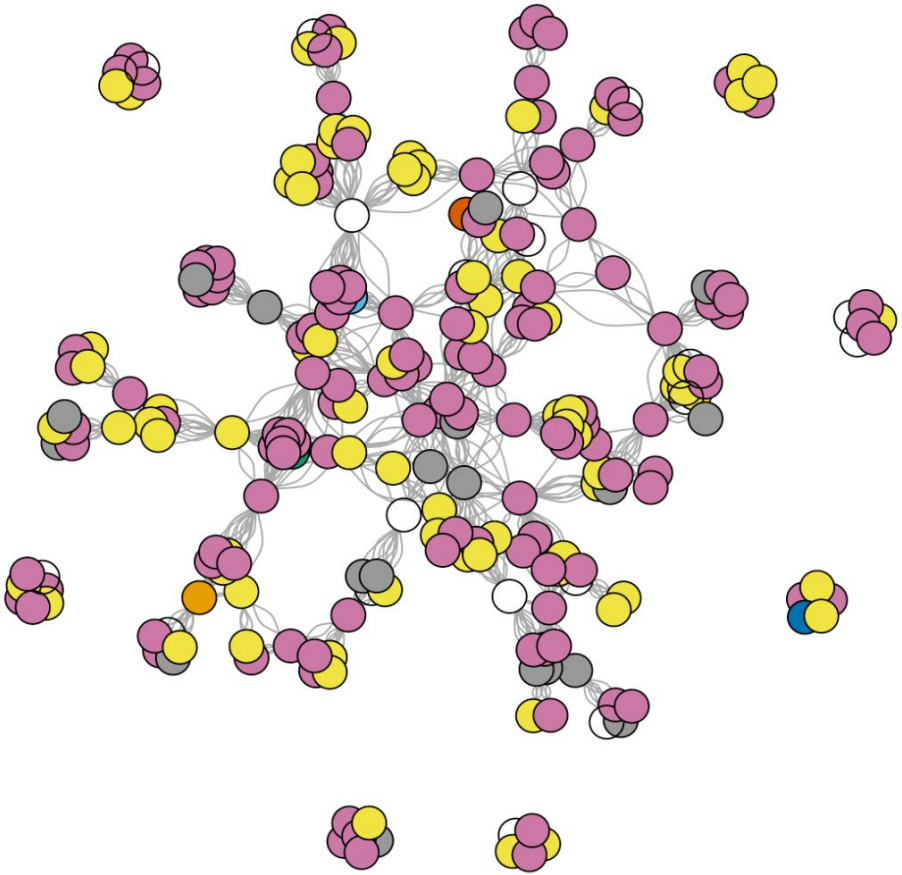
(k-partite graph is a graph whose vertices are or can be partitioned into k different independent sets)

Bipartite graph

- When k = 2 these are the bipartite graphs
- A graph whose vertices can be divided into two disjoint and independent sets

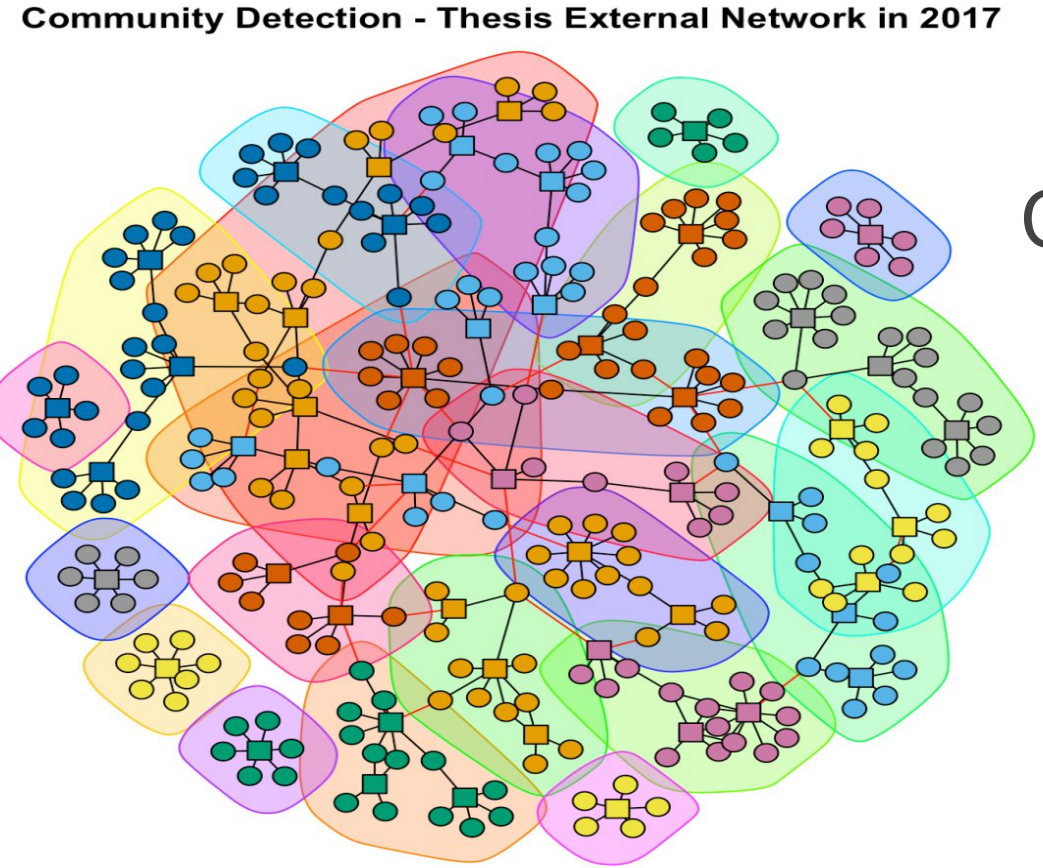
Multipartite graph

- When k > 2 the graph can be called as the multipartite graphs
- A graph whose vertices can be divided into more than two disjoint and independent sets



Edge list Vs Adjacency matrix

- An edge list is a list (or array) of all the E edge in a graph. Each lists are one of the simplest representations of a graph
- An adjacency matrix is a matrix that represents exactly which vertices/ nodes in a graph have edges between them.

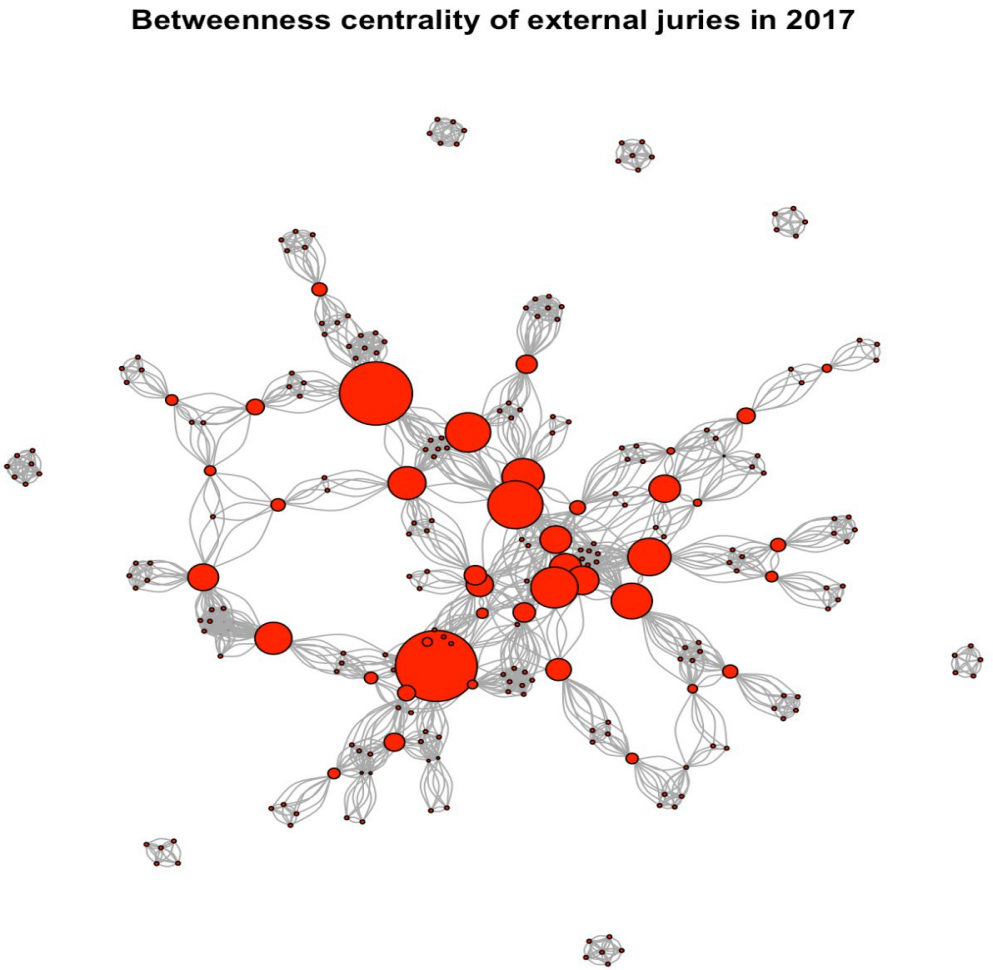


Community Detection

- Every jury is not closed on itself
- There is overlapping between the communities

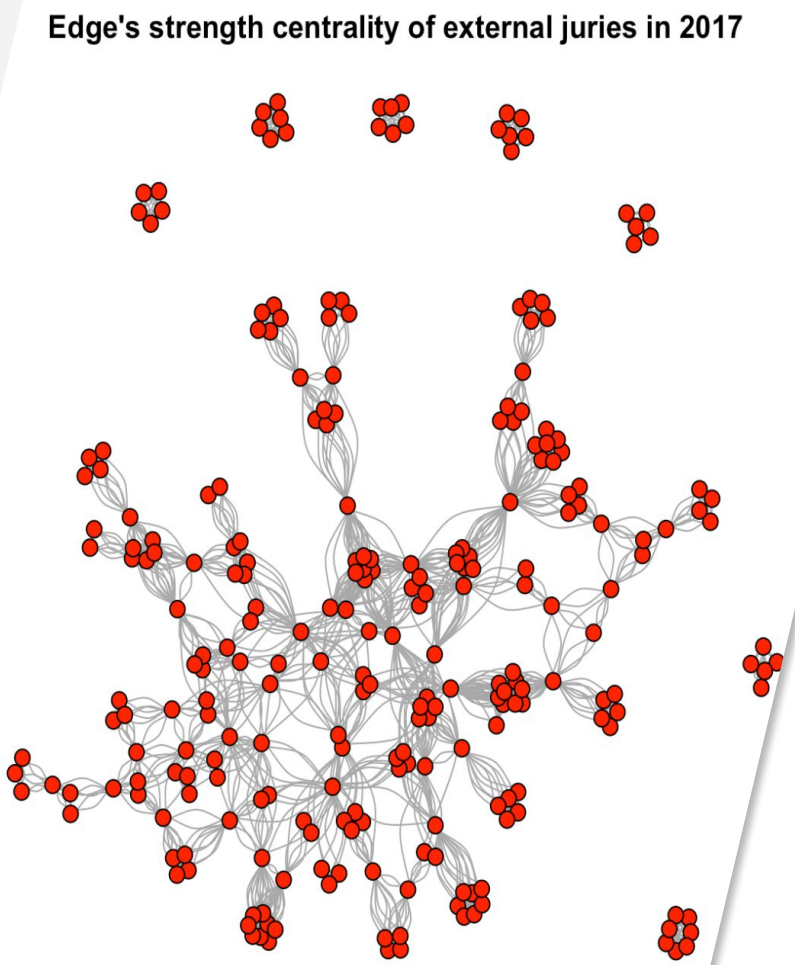
Betweenness centrality

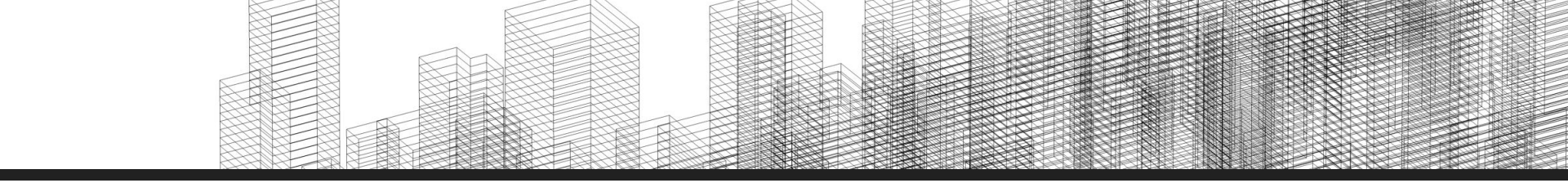
- The bigger the node higher the betweenness centrality.
- This means, that on average the members of the juries with high node betweenness centrality have participated in many other smaller neighbour juries.



Edge Strength

- Most of the nodes with high edge strength are centered in the middle of the social network as those juries corresponds to members of juries that participated in other juries represented by the neighbor nodes.



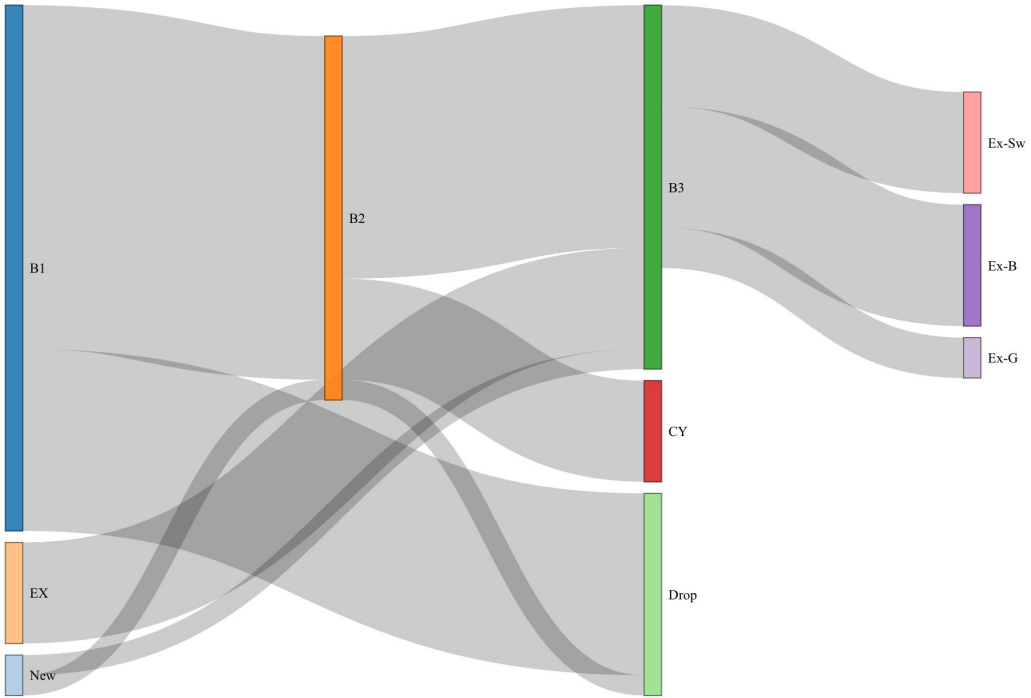


Sankey Diagram & Chord Plot

- The dataset was used in this part is created based on the changing of number of student in Ygrec Bachelor Program

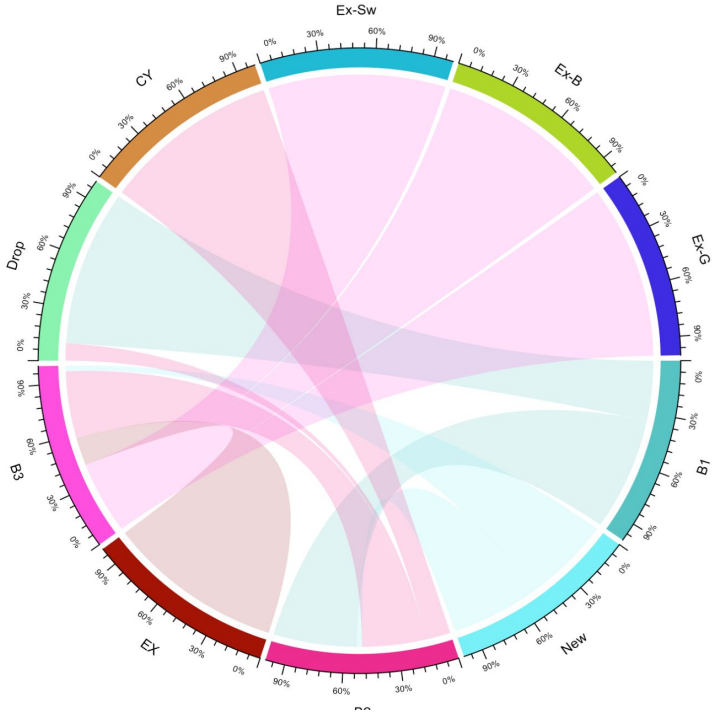
- Methodology: Sankey diagram & Chord plot

Sankey diagram



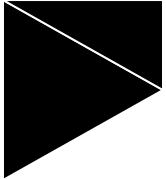
- Around 35% of students dropped the program after the first year.
- Also, 30% of the second-year students changed to the engineering program.

Chord Plot



How many percentages of students drop out of the program from B2 in the total dropped students?

A. 20%
B. 50%
C. 10%



Association Rule Mining

- There are 3 key metrics to consider when evaluating association rules:
    1. Support
    2. Confidence
    3. Lift

LIFT

- No direction (lift{X,Y} is always equal to the lift{Y,X})
- Lift{X,Y} = lift{Y,X} = support{X,Y} / (support{X} * support{Y})
- If lift = 1; implies no relationship between X and Y (i.e., X and Y occur together only by chance)
- If lift > 1; implies that there is a positive relationship between X and Y (i.e., X and Y occur together more often than random)
- If lift < 1; implies that there is a negative relationship between X and Y (i.e., X and Y occur together less often than random)