

Social Network Analysis

Project

You will analyze a social network with help of the software Gephi. You are being provide by the already scrapped data from YouTube.

Please Note: You have to submit your project individually. Submit your project by 17TH June 2020.

Task 1

Please attach your database that contains both of your YouTube channels in Gephi format. This means that you have to transform the two .csv files into a single one with only three columns.

Task 2

Attach a screenshot of your "Overview" tab in Gephi, which shows your network after you ran the "Yifan Hu" Layout algorithm.

Task 3

Take a screen shot of the Data Table of Edges

Task 4

Calculate the average Degree of your network. Display and analyze all three resulting network measures:

- A. Degree
- B. In-Degree
- C. Out-Degree

Take screenshot of your network

Answer the following questions:

1. What is the difference between them?
2. How many categories do you get for each?
3. Can you make sense of the numbers it indicates the number of degree per category for each of the three measures? Why or why not?

Task 5

1. How many nodes (videos) are shared by both YouTube channels? Count them or calculate them.
2. Calculate the network Modularity and take screen shot of the network.

Task 6

Calculate the "Undirected Closeness Centrality" for your network, through "Average Path Length" (attach network screen shot) and then answer the questions:

1. How many groups of nodes do you get?
2. Please interpret the different groups. Which nodes are part of which group and why?
3. Calculate the "directed Closeness Centrality" for your network, through "Average Path Length"(Also attach screenshot)

Task 7

Calculate PageRank for your network, a special version of Eigenvector Centrality. Then answer the following questions: (attach screen shot of the network)

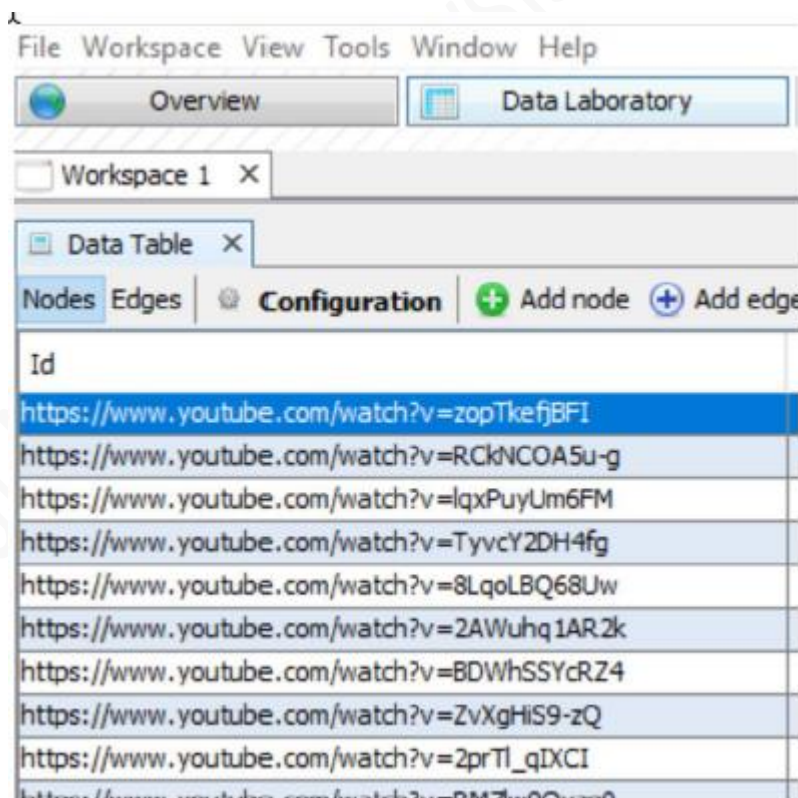
1. How many groups of nodes do you get for PageRank?
2. What do they measure?
3. Is this useful?

Attach Screen shot of each if any changes done in network

Task 8

Please attach a screenshot of your "Data Laboratory" tab, now at the end, after you have done the preceding analysis.

Go to "Data Table > Nodes" (not Edges) and make sure that the "Id" column is completely readable (not cut off to its right):



1. Take a screenshot of the full size "Data Laboratory" window (not just the part shown in this excerpt screen shot).
2. What does it mean by weight in Data Table in edges Tab?

Task 9

Attach a screenshot of your "Overview" tab in Gephi, which shows your network after you ran the "Force Atlas" Layout algorithm with repulsion strength of 20000.

THE END