Social Network Analysis Report: Who Talks to Whom

GROUP G

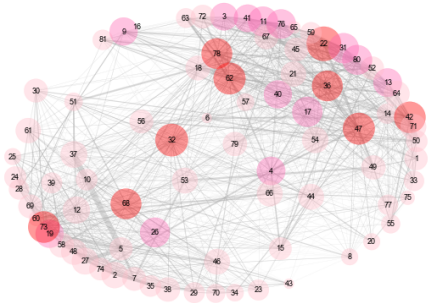
A. OVERVIEW

The "Who-Talks-to-Whom" data contains 81 nodes, each stands for a student in the MSc Business Analytics 2017-18 course. There are two sheets: "Sent" and "Received" which respectively represents the number of 1-to-1 conversation a student started (sent out) and participated (received) within the week 17-24 Nov 2017.

The "Sent" matrix and the transpose of "Received" matrix should be identical in theory, but due to errors from factors such as different definition of "conversation" and false memory, the two matrices in the data provided is slightly different. Therefore, we take the average number of the "Sent" and transposed "Received" matrix as the final total conversations happened within the week.

From the raw data, it is noticeable that the "Sent" and "Received" matrix are very similar - among the 6561 pairs of students conversation, only 273 conversation (around 4%) in "Received" have a different number from "Sent". It shows that the conversations between the students are quite balanced - between student A and B, the number of conversation started by A is very close to the number by B. It also reduces the information we could obtain from the comparison of "Sent" and "Received", e.g. main information provider & receiver in the course.

Total Conversation Network



From the graph "Total Conversation Network", the whole class seems to split up to two groups with one more "talkative" (upper right group) and one more "silent" (lower left group). The two groups are connected quite tightly by several socially active students (No. 32, 68, 62, …) who are represented by big red dots - they are the ones who has talked to the most number of people (top 10% talkative). To learn more about the social network in our course, we could look at some centrality measures.

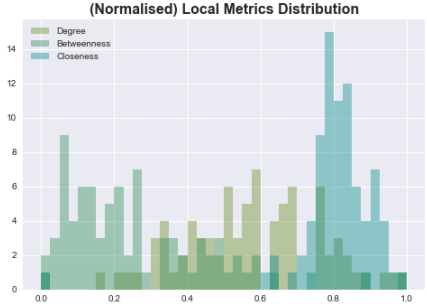
B. CENTRALITY MEASURES

The centrality measures help us to identify the relatively important (or influential) students in the course. We look at five measures which give the "importance score" based on different factors.

1. Local Metrics:

We use three local centrality measures to explore the network: degree centrality, betweenness centrality, and closeness centrality. They all focus on the ego network of certain nodes.

(Normalised) Local Metrics Distribution



Degree centrality defines how important a student is purely based on the number of students he/she talked to during the week. The olive colour histogram in the above graph shows the normalised distribution of the degree centrality of the students. It seems to follow a normal distribution: students who talked with many people and who barely talked are the minority of the class.

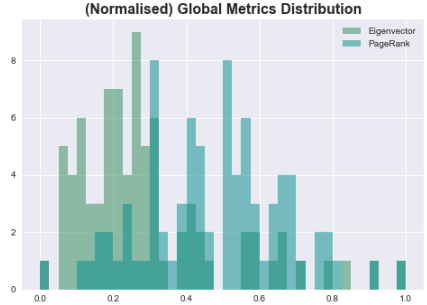
Betweenness centrality shows who act as bridges between students and could influence the information flow around the class. Most students' betweenness centrality lie in the range (0, 0.02], and there are only 3 students who have the number around 0.03. Those students (No.32, 22, 62) would affect the information transmission stronger than others if they were removed.

Closeness centrality scores students based on the sum of "shortest path" between a certain student to all other students - high-closeness centrality students would be able to influence others or spread news more quickly. The numbers of the vast majorities lie in between [0.45, 0.70], except for student No.16 whose closeness centrality is nearly 0. It is hard for this student in general to quickly obtain or spread information within the class.

2. Global Metrics:

We use two global centrality measures: EigenCentrality and PageRank. These measures not only look at the ego network of each node, but also take node's neighbours into consideration when scoring the node's importance.

(Normalised) Global Metrics Distribution



EigenCentrality measures the importance of a student based on how well connected he/she is and how well his/her connections are, and so on through the whole course. PageRank uses similar way but also takes direction and weight into account - the higher amount and frequency of incoming conversations, the more influential of the person. From the histogram, PageRank scores are more evenly distributed than EigenCentrality. Some students who talk to lots of students who are also "talkative" may have much higher EigenCentrality, but their influence to others are also spread and thus weakened. Therefore we have PageRank more concentrated than EigenCentrality. The top 3 highest PageRank students are: No.45, 22 and 18. It is interesting that only No.22 is a "big red node": No.45 and 18 are not noticeably talkative to everyone in the class; however, they have very thick links with others students, for example, No.18 frequently communicated with No.51 who belongs to the lower left group which increases the range of his/her opinion's influence. These are the students who hold wide-reaching influence in the class and are good candidates of "student leader" or student committee members.

C. CLUSTERING

(Local) Clustering coefficient measures the likelihood of a student's connections talking to each other, which reflects how clustered our class is. The calculation result shows that more than 90% students has a clustering coefficient between 0.5 and 0.75 and the average clustering coefficient equals 0.61. This modest number reflects the situation that BA class is a newly formed community where members are not completely familiar and connected to all other person, but the past 3 months already helped people to build relationship to a certain amount of classmates. More networking and team-building chances should be created for students to reach out to the classmates that they have not communicated much with, for example, the students from "the other group". These weak-ties would be very helpful in terms of career and personal development.