

Social Network Analysis

Do students want to work with their best friends?

Eddie Lin

2018/03/11

Project Description

Using social networks analysis(SNA) to understand the the social dynamic for middle school students. As there seems to be a norm that we should never work with our best friend, I want to know if we can see that life wisdom reflected among middle school students. Specifically, my interested questions are as follows:

1. Is there a pattern when students choose who is their good friend in class?
2. Do students tend to work with their good friends in class?
3. Other findings such as students of highest popularity, isolated students,etc.

Tools & Data

- R
- R packages: igraph
- Data source: *M. Vickers and S. Chan(1981) Representing Classroom Social Structure*

Data Description

The data were collected by Vickers & Chan from 29 seventh grade students in a school in Victoria, Australia. Students were asked to nominate their classmates on a number questions:

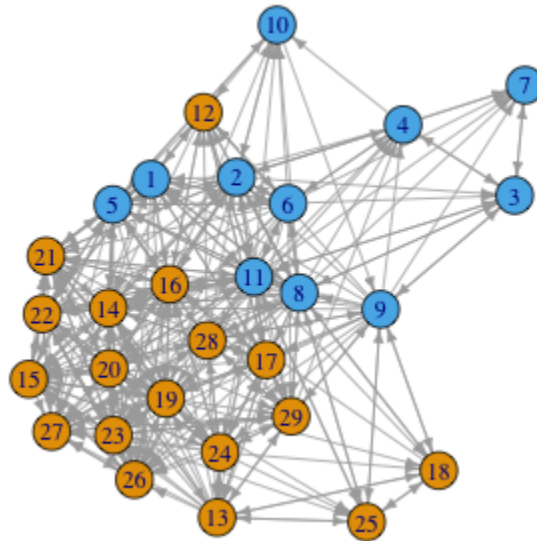
1. Who do you get on with in the class?
2. Who are your best friends in the class?
3. Who would you prefer to work with?

Preliminary Network Analysis

In SNA, there are several methods to compute centrality to evaluate who is "important" in a network. However, I will begin by visualizing the networks for all 3 questions without using centrality measure.

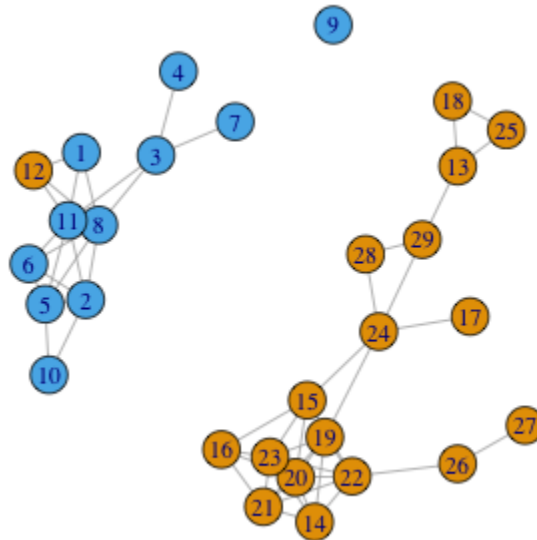
Who do you get on with in class ?

- male students(blue nodes)/ female students(yellow nodes)
- the edge (line) width is weighed by the total of votes received by a student (node)



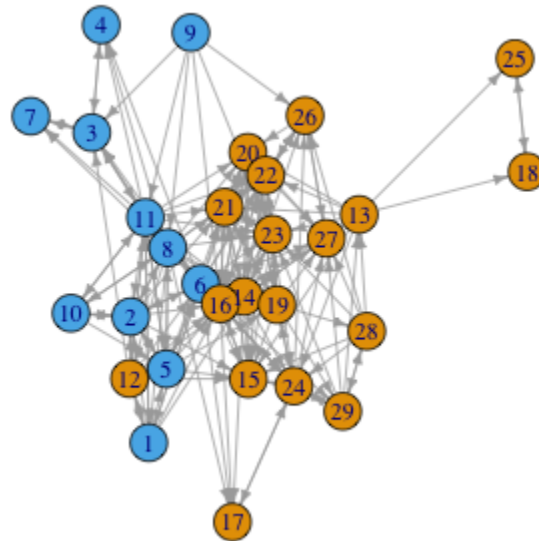
Who is your best friend in class ?

- male students(blue nodes)/ female students(yellow nodes)
- the edge (line) width is weighed by the total of votes received by a student (node)
- for this measure, I use "mutual vote", meaning the person a student votes has to vote back to form an mutual edge



Who do you prefer to work with in class ?

- male students(blue nodes)/ female students(yellow nodes)
- the edge (line) width is weighed by the total of votes received by a student (node)



Findings

- there seems to have different network formations across 3 questions, suggesting students are opinionated differently towards these questions
- the network of best friend voting is a lot sparse than the other 2 questions, because counting only the mutual votes and perhaps more conservative voting due to the nature of that question
- the voting results are pretty distinctly divided by gender
- to answer our research questions and see the networks more clearly, I will use different centrality measures to investigate the networks further

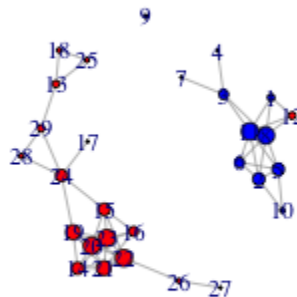
Centrality in SNA

There are different kinds of centrality measure and it depends on the purpose and the question one is to analyze. Here is a useful tutorial video that gives a brief explanation about different centrality measures. In the current project, I will use degree centrality and betweenness centrality and focus on the **best friend** and **who to prefer to work with** question.

Degree centrality: who is your best friend?

- it seems that students **8,11,22,23** are the winners
- here, the winners mean the students who get the most "**mutual votes**"
- no surprise, the result is still pretty gender-distinct (w/ 1 exception in the male student group)

Degree Centrality: Best Friend(mutual)

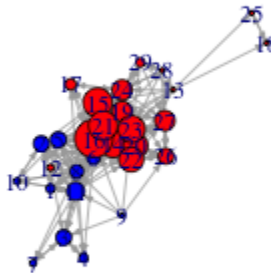


```
## [1] 9 4 7 17 27 10 18 25 26 28 1 12 13 29 3 6 16 2 5 14 15 24 19
## [24] 21 8 11 20 22 23
```

Degree centrality: who to prefer to work with?

- studnet **21, 16** are the ones that most people want to work with
- the result is based on the total of "**received votes**" that one has got
- so far, we can see that "**most popular students**" do not match "**people I want to work with the most**"

Degree Centrality: Who to Prefer to Work With



```
## [1] 9 13 18 25 28 7 10 12 1 4 29 17 6 8 2 3 5 26 11 19 24 27 20
## [24] 22 14 23 15 21 16
```

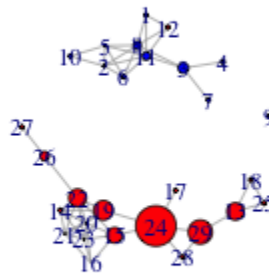
Betweenness centrality: the plays as the "bridge" in class?

- unlike degree centrality that measures the total of votes (roughly speaking), betweenness centrality measures who takes the critical position in the network
- we can imagine a person with high betweenness centrality is like a very important bridge that lots of people have to go through it to get to others (shortest distance)
- so it is not so much of "**the popular vote**" but more like "**the gate keeper**" to be the winner of this measure

Betweenness centrality: who is your best friend?

- studnet 8, 11, 29, 24 are the ones having the highest between centrality in 2 different gender groups
- this result for male students matches that from using degree centrality

Betweenness Centrality: Best Friend(mutual)

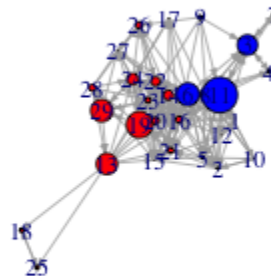


```
## [1] 1 4 6 7 9 10 12 14 17 18 25 27 28 16 21 20 23 2 5 8 11 26 3
## [24] 15 13 22 19 29 24
```

Betweenness centrality: who to prefer to work with?

- **11, 19** seem to be the "**bridge**" for students in this class
- again, this result shows that your best friend may not be the person you want to work with the most

Betweenness Centrality: Who to Prefer to Work With



```
## [1] 4 7 9 18 25 10 1 27 12 17 15 2 5 23 21 28 16 26 20 22 14 24 8
## [24] 3 13 29 6 19 11
```

Think beyond the numbers ...

It is great that we've got some results, but it is worth taking a pause to think about what do these actually mean in reality. Now by using degree and betweenness centrality, it is known that **the person you like is not the person you want to work with**, but why will these two centrality measures give different winners? Or, which centrality measure should we use or should we just use both?

Wait, what are our research questions again ???

- Although there may not be a perfect answer, but if we think about the definitions of **degree** and **betweenness** centrality, we can assume that using **degree** makes more sense in answering the research questions. As the original paper asked the 7th grade students to vote based on different questions, we do care about who get the most votes rather than who is like a bridge in this class
- the **betweenness** measure is useful, too. Especially when we want to know who is the gate keeper in a company that lots of messages/ decisions have to go through that person (may not mean that person is a high-level manager). Or in the technical world, if a node of internet network is very essential (when it is down, all down), then that node will have very high betweenness.
- so back in our case, a person that stands on a very important (bridge) spot may not mean the (s)he get the most votes on the 2 questions (though it might in other cases). So I will opt for the result produced by using **degree centrality**

Findings & Summary

- Is there a pattern when students choose who is their good friend in class?
 - **Yes, gender seems to be a pretty salient factor that separates the class, meaning there is no "the best friend" for the whole class, but there are "best friends" for students of two different genders**
- Do students tend to work with their good friends in class?
 - **No, they don't. This result seems to fit the norm of "you never find your colleague as your best friend"**
- Other findings such as students of highest popularity, isolated students, opinions leaders, etc.
 - **We can see the most popular students in class, but we can also see there is an isolated student that no one favors him with any vote as best friend (but he votes for others). The class manager may want to see about this student to see how he is doing in class.**

Limitations & Suggestions

No data analytics is perfect, I came up with a few thoughts and make some suggestions in the followings:

1. When students give votes without a standardized limit, this may affect the result of the "best friend" measure. If a student votes for 20 other students while another votes only for 2, then the odds of forming a mutual vote could be affected.
2. The lack of other student data such as past performance, test scores, demographics, etc, limits the scope of data analytics. We can further investigate how students bound with each other and how to create better student work group if additional student data is given.
3. There are other useful measures in SNA such as dyad, triad, cliques, and cutpoint that can take this or similar research further, but these measure will be more meaningful if we knew more about students' background or if we have a more specific question. For instance, how do we team up students in a group according to their skill levels for collaborative learning.