Student Voting Networks
PLSC 57200 Proposal

Reid McIlroy-Young

February 5, 2018

When not attending class one of my jobs is as the a TA to Prof, James Evans' Computational Content Analysis class. One of the new things we are trying this year is to have the students vote for a question about the each of reading to be asked in class. These votes are on a public forum (github.com/Computational-Content-Analysis-2018 if you are interested). These votes are public and the students are required to pose at least three questions and vote at least five times. This gives me a longitudinal sample of the student's affinities.

For my project I hope to look at how the voting network changes over time and how voting habits relate to the participants relationships. I already have a data collection system setup and have used it to generate some preliminary network diagrams (figures 2, 3, 4, layout is a simple force-directed model and size is based on degree) and network information (table 1). I suspect that popularity will be the dominant force for centrality and can already see that certain actors tend to be more central across weeks than others, e.g. *HyunkuKwon* or *Sunnyjooey* are near the top (for a loose definition of top) in most. Also of interest is to see if there is much clique balancing and other more local phenomena. One additional source of data I hope to add is the friendship network of the class, I have permission to distribute a survey, but am still working on the logistics.

As I already have a good amount of the data, that aspect of the analysis will not be a problem. But there are still likely to be some issues, first since I am looking at votes, all edges have a weight of one so filtering or aggregating them is tricky. Second, not all students are present every week and some may vote randomly. This, since the voting and question publishing take place simultaneously there are likely to be temporal effects as people can't vote on questions that haven't been published yet. I have already done some analysis of the distribution of votes, figure 1, and the distribution does appear to have some of those effects, but not be entirely dominated. These issues are both solvable if enough students are good actors so I believe I will be able to get around them. A final note about this analysis, the forum GitHub hosts millions of other discussions, most of which are public. If I have time I will try to apply the methodology I develop for the the students votes to a few other communities.

Lesson Date	Number of Votes	Number of People	Mean in Degree	Median in Degree
2018−1−5	52	18	2.89	2
2018−1−12	192	35	5.49	6
2018−1−19	182	28	6.50	6
2018−1−26	146	28	5.21	4
2018−2−2	101	26	3.88	3
2018−2−9	0	0	0	0
2018−2−16	0	0	0	0
2018−2−23	0	0	0	0
2018−3−2	0	0	0	0

Table 1: *Basic network information, note that for week 1 there was only a day given to students and questions were not required*
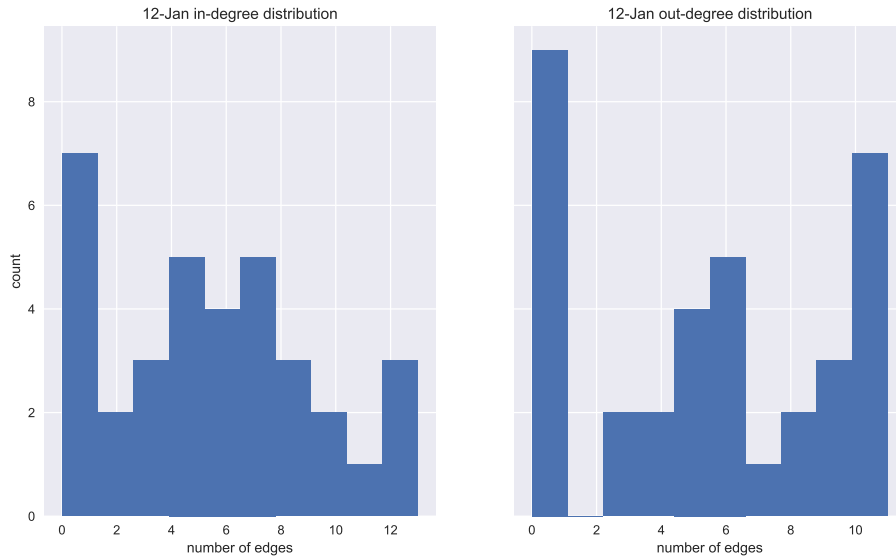

Figure 1: *Histogram of in and out degree distributions for January 12*
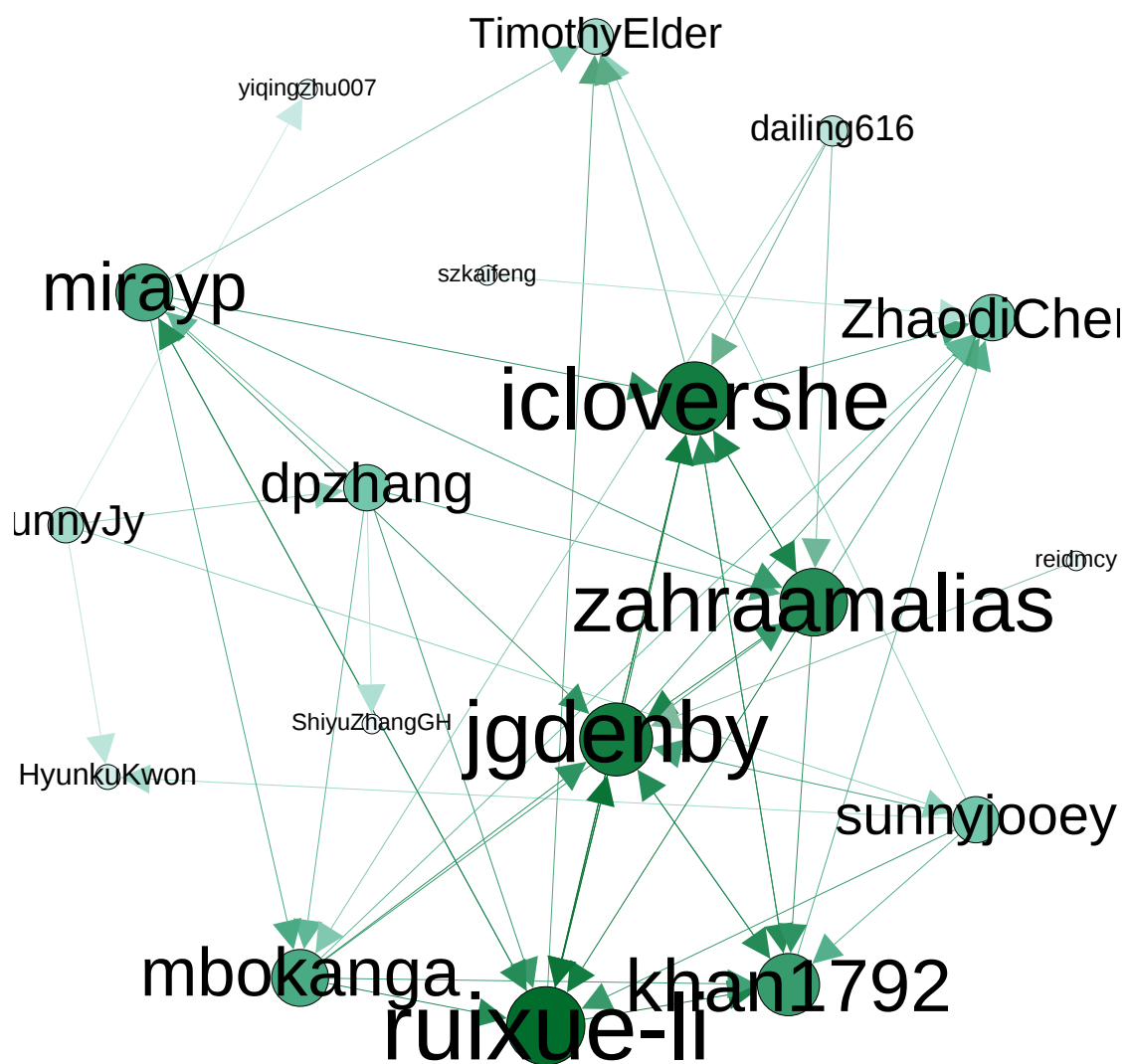
Figure 2: *Preliminary visualization of the January 5 network, note this week gave students the shortest window to comment* 3

Figure 3: *Preliminary visualization of the January 12 network*

Figure 4: *Preliminary visualization of the January 26 network*