

Freshmen Social Network

Yun-Tsz Tsai, Nino Janashia

Introduction

This report analyzes the evolution of a friendship network among sociology freshmen at the University of Groningen during 1996-1997. Using network data collected over five time points, this study examines relationships, node-level attributes (e.g., gender, program, smoking behavior), and key network metrics. The analysis applies social network analysis techniques to uncover patterns and insights about the students' social interactions and dynamics over time (for more details see <https://www.stats.ox.ac.uk/~snijders/siena/>).

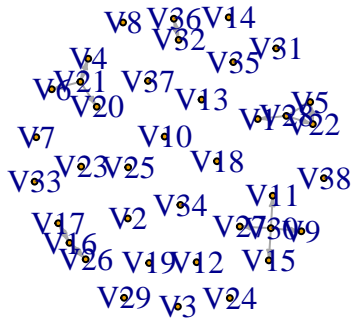
Data handling

The original data (a) comes in the format of a matrix, (b) is a relatively small data set with 38 nodes, and therefore we decided to keep it the way it is and analyze the network using adjacency matrix instead of a more memory-efficient edge list. In addition, the original data contains label that divided the ties between individual into 5 categories, with 1~3 represent positive relationship, 4 represents neutral, and 5 represents negative ties. After some explorations, for period 4, we only keep the positive ties for the sake of easier interpretation.

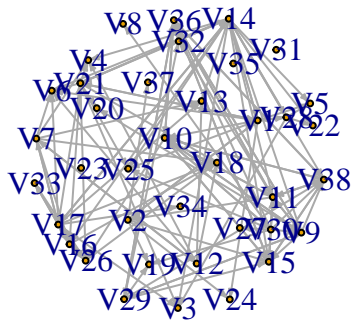
Since the analysis aims to examine the differences of the friendship network at different periods of time, we do not remove any isolated nodes for comparison.

Initial exploration of the network

Friendship Network at Time 1



Friendship Network at Time 4



Descriptive data for the network

The in-degree and out-degree are identical for both networks at time 1 and time 4. However, most figures at time 4 are much higher than those at time 1, except for reciprocity. This might be because as the friendship network matured, individuals became more selective and do not reciprocate as much compared to Time 1. The density increases slightly from 0.01 to 0.09, meaning that after a certain period of time, individuals in this network have more positive relationships with one another. However, the density is still low in general.

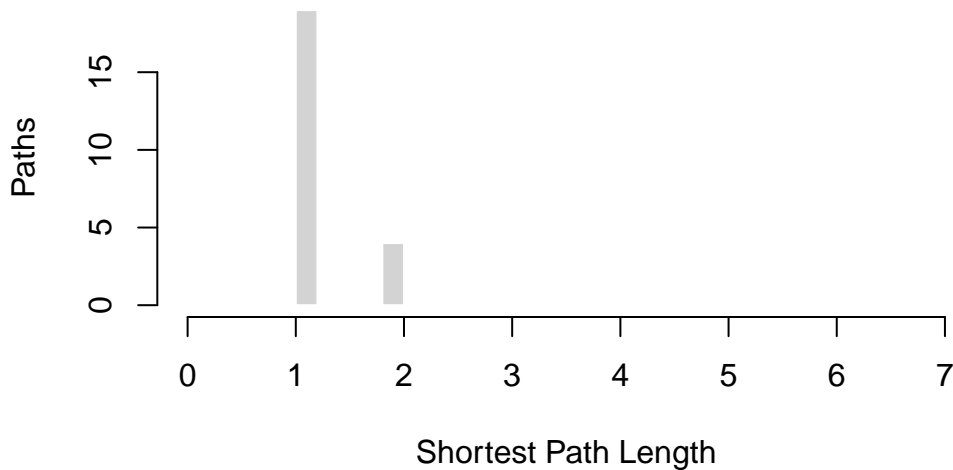
```
# A tibble: 7 x 3
  ~~~~~ Time_1 Time_4
  <I<chr>> <dbl> <dbl>
1 in degree (mean)    0.5    3.39
2 in degree (sd)      0.65    2.33
3 out degree (mean)   0.5    3.39
```

4 out degree (sd)	0.98	4.23
5 density	0.01	0.09
6 reciprocity	0.53	0.43
7 transitivity	0.2	0.42

There is no differences between the in-degree and out-degree shortest paths in terms of distribution for both Time 1 and Time 4. However, the reason for 1 being the shortest and most frequent length of path at Time 1 is that there is not many edges then and individuals are mostly closely connected to each other. At Time 4, individuals nominated more individuals whom they have a positive relationship with, which increased the density of the network as well as the approachability of individuals. Thus, we see the distribution of shortest paths deviates more compared to that of Time 1.

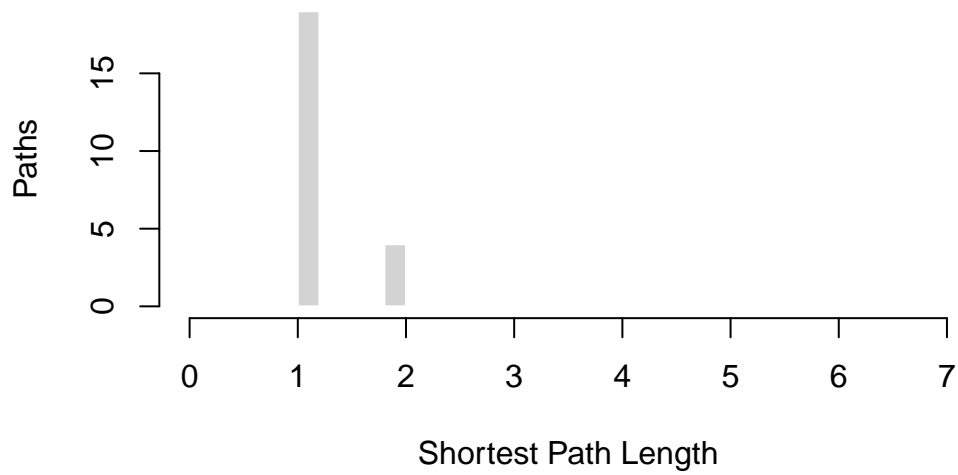
Shortest Path Lengths

at Time 1 –in degree



Shortest Path Lengths

at Time 1 – out degree



Shortest Path Lengths

at Time 4 –in degree

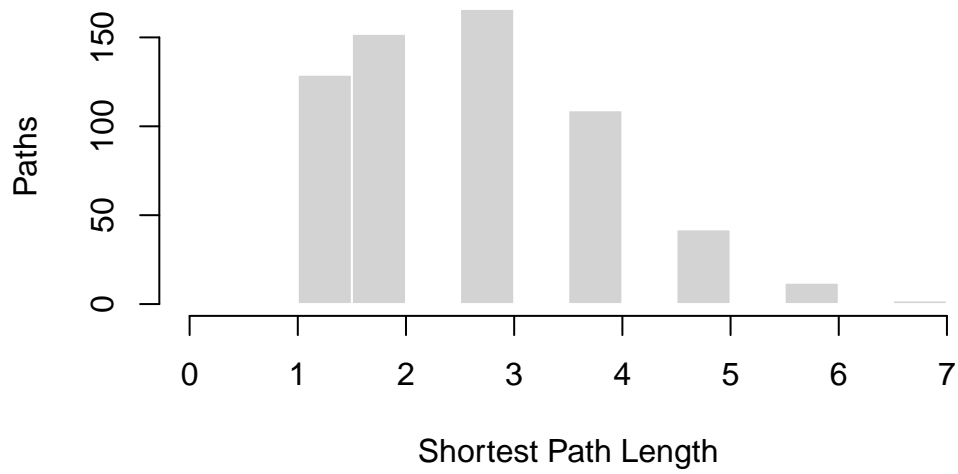
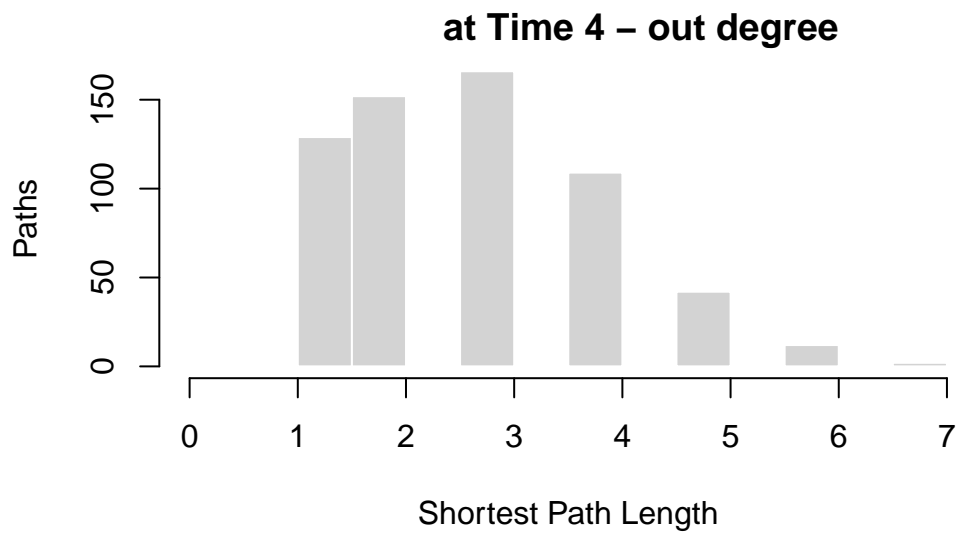


Table 1: Descriptive data
for nodes level
variables

		N
gender	Female	24
	Male	14
program	2- year	15
	4- year	23
smoke	No	20
	Yes	18
drug	No	22
	Yes	16

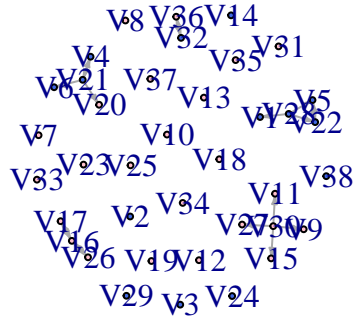
Shortest Path Lengths



Visualization of node attributes

Friendship Network at Time 1

blue: male, pink: female



Friendship Network at Time 4

blue: male, pink: female



Assortativity

EX index at Time 1 = -0.404288

EX index at Time 4 = -0.1461519

It looks like as time went by, the tendency of assortativity decreases slightly, with both boys and girls nominating the opposite sex more compared to Time 1. This shift suggests that while cross-gender friendships remained common, students started forming more gender-homophilous clusters overtime

As for the measurements for robustness and vulnerability of the network, one can use the number of nodes in a network as the denominator and see if removing the node from the graph

causes any noticeable impact (e.g. a significant decrease of network density) on the original network structure. If so, increases the numerator by 1 and repeat the process after examining all the nodes.