FACE-TO-FACE CONTACT PATTERS IN A PRIMARY SCHOOL

By Anuj Verma

INTRODUCTION

There is not much information available on mixing patterns of children in school environments. Describing and understanding contacts between children at school would help quantify the transmission opportunities of respiratory infections and identify situations within schools where the risk of transmission is higher. This data consists of the measurements carried out in a French school (6–12 years children). The data was collected in a time-resolved face-to-face proximity of children and teachers using a proximity-sensing infrastructure based on radio frequency identification devices. This project uses social network analysis to examine the patterns of face-to-face contacts among the primary school students.

DATA

The data used in this project was acquired from www.sociopatterns.org under the title 'Primary School – Cumulative Networks' and can be obtained from here in GEXF files.

The dataset comprises of two weighted networks corresponding to the two days of study of face-to-face proximity between students and teachers. Each day have a separate network where nodes are individuals and edges represent the face-to-face interactions between these individuals.

Nodes have two attributes: classname, which indicates class of individual. There are 10 classes (1A, 1B, 2A, 2B, 3A, 3B, 4A, 4B, 5A, 5B) associated to students. Teachers are assigned to 'Teachers' class. The second attribute is gender, which indicates gender of the individual. All teachers are assigned an unknown value for gender. Edges have two attributes: duration, which indicates the cumulative time spent by individuals A and B in face-to-face proximity in a day and count, which represents the number of times two individuals established a contact during a day.

DATA FILTERING

The data on primary school cumulative networks is available in GEXF format which can be read by gephi but not igraph. To convert this into igraph readable format, the data was first loaded in gephi and then exported to igraph readable format graphml without any changes on the original data.

The data contains for day 1 of the study contains 236 nodes and edges 5899 and data for day 2 consists of 238 nodes and 5539 edges. For simplicity of the network, edges between individuals who spent total contact time less than 2 minutes (120 secs) during the day were removed. This action gives us only the strong contacts between the individuals. Edges between individuals who connected twice or less were removed. Furthermore, the teacher nodes were removed from the graph and the duration of contact was assigned as new edge weight.

The filtered network contains 222 nodes and 1558 edges for day 1 and 222 nodes and 1558 edges for day 2.

ANALYSIS

The analysis focuses on identifying the mixing patterns of primary school students of same grade and students of different grades. This analysis can be helpful in controlling the transmission of some contagious diseases in an event of occurrence.

The patterns of contacts between children are analyzed through the number of contacts between each pair and the duration of these contacts.

Figure 1 displays the comparison of these contacts based on A) contact duration and B) frequency of contacts for both days. As we can see, the patterns are very similar for day 1 and day 2. A large number of the contact durations are about 1minute long. The blue vertical line is at 2 minutes and we can see approximately 1700 (1000 + ~700) edges with contact duration less than 2 minutes. Less and less distinct student interactions happen for more time. We can see the distinct pairs of individuals interacting more than 100 minutes. Similar story for number of contacts made by two individuals. We see a very high count for number of contacts less than 2 (blue vertical line) and very low count for more contacts between the pairs.

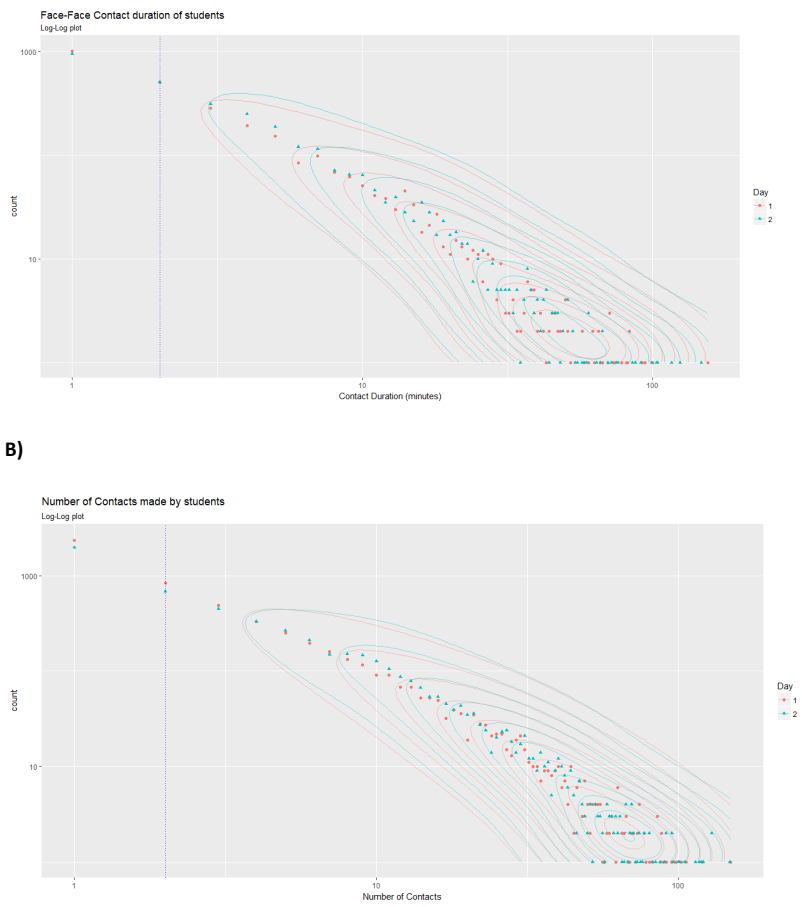

Figure 1: Log-Log plot of distribution of data based on A) contact duration and B) number of

The density for both of these graphs shows that students ended up spending most of their time in school with less than 10 fellow students. We decided that the contact durations up to 2 mins and up to 2 number of contacts between individual pairs will be filtered out from our networks of both days.

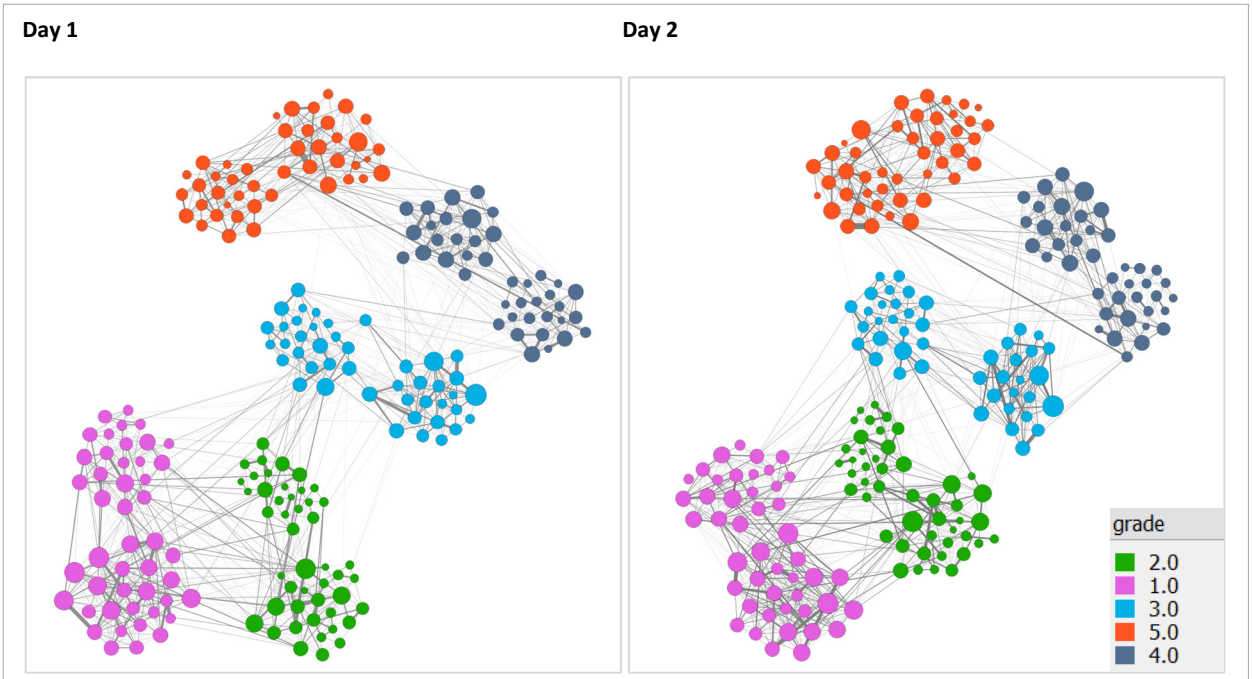

Figure 2: Network visualizations for Day 1 and Day 2 graphs. Different node colors represent different grades of students for both graphs.

The Figure 2 shows visualizations for day 1 and day 2 created from the filtered networks. Both days network contains 222 nodes and 1558 edges. The nodes represent the individuals which consist of children from grades 1st, 2nd, 3rd, 4th and 5th. The color of nodes represents different grades (listed in the Figure 2). The nodes with more connections (degree) have larger size than the nodes with lesser connections. The edges between two nodes represent the face-to-face interaction between two individuals and are weighted by the total contact duration of two individuals during the day. The figure highlights the mixing patterns between the children of different classes and how the children preferentially attach to classmates. The classes between same grade are connected more as compared to the classes of different grades. Grades 1 and 2 have minimal contacts to grades 4 and 5.

Assortativity

The assortativity between different grades is shown in the bar chart. It is clear that both the days show homophily characteristics which means that the students of similar grades mix well with each other as compared to the students of different grades. The assortativity factor is highest between 1st-4th, 1st-5th, 2nd-4th and 2nd-5th grades. This implies that junior grades and senior grade students have minimum contacts with each other. From the bar chart, it is clear that this network have assortative characteristics based on the grades in primary school.

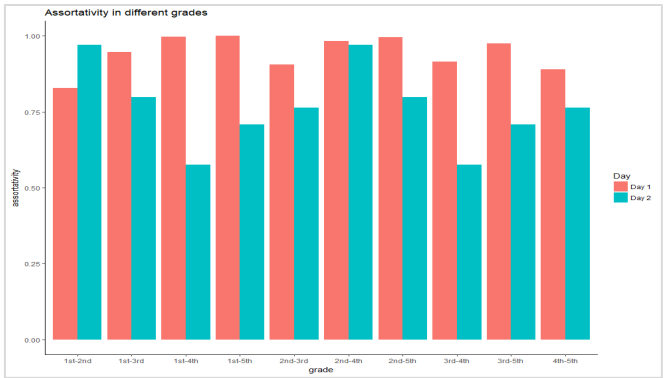

Figure 3: Inter grade assortativity is high for both days but day1 network shows more homophily characteristics overall.

This theory can be bolstered by Conditional Uniform Graph (CUG) test which will help us determine whether the grade based assortative nature of our network is a unique property of this network or this can be found in any

random network. The CUG test observed p-value 1 for this network attribute (Figure 4) which means that the probability of a random network having assortativity characteristics by grade is none. This makes sense since random networks only calculate probability of an edge which is never based on property. Now we have a proof

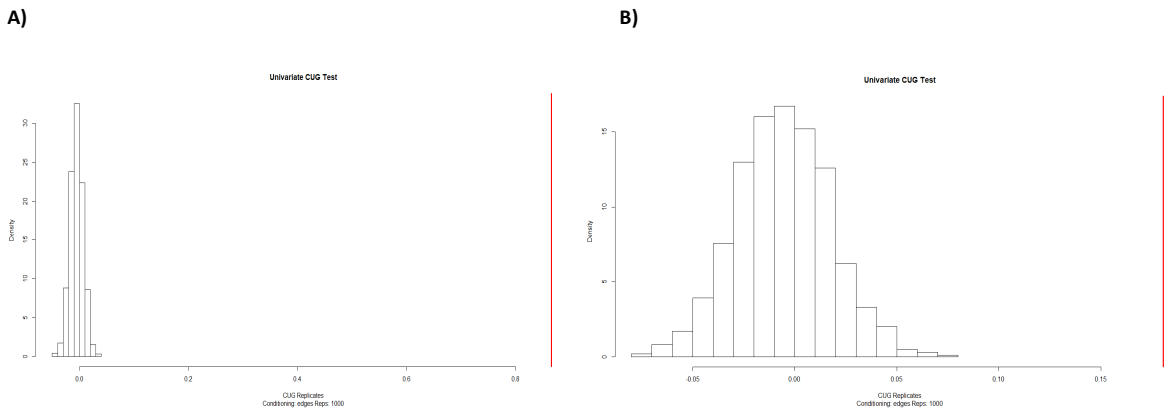

Figure 4: A) CUG test for assortativity by grade and **B)** for assortativity by gender based on randomly generated networks.

that random networks are unable to show the characteristic of homophily by grade attribute. In addition to this, it would be interesting to see if our network is assortative by gender of students. The CUG test will support this theory by again giving a p-value 1. This means that it is improbable for random network associations to generate a network characteristic similar to a network with gender assortativity. Figure 4 shows the CUG test output for both A) assortativity by grades and, B) assortativity by gender attributes of the nodes.

To confirm the CUG results, assortativity can be examined by conducting Quadratic Assignment Procedure (QAP) test. Instead of generating random networks, this approach takes a fixed network and attempts to re-arrange the nodes to determine the probability based on the arrangement of those node types in the network. The grade and gender attributes can again be tested for their significance using QAP tests.

The Figure 5 shows the results of QAP test for grade and gender characteristics. These tests support the results

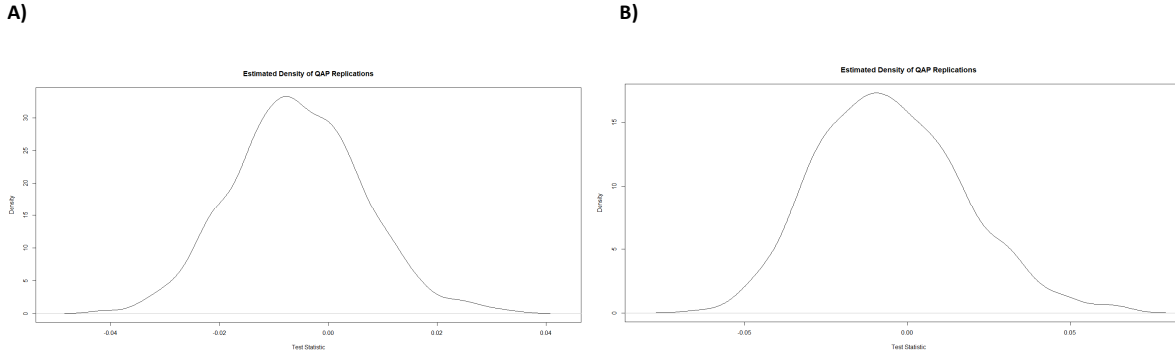

Figure 5: A) QAP test for assortativity by grade and **B)** for assortativity by gender.

produced by CUG test with estimated p-values for both the graphs as 1. This means that even after 1000 iterations of re-arranging the nodes, QAP was not able to generate network properties similar to the observed network. Now we can say that the attribute values associated to our network are significant.

CONCLUSION

The above results show that children mix preferentially with children within their grades. One reason could be the age factor since students in same grade will be of similar age. This homophily could also be due to the fact that children of same grade study together and have same schedule. The interactions between different grade students could have happened between the lunch times or in some common rooms like restrooms or recreational activity rooms. It is also observed that students from junior grades (1 and 2) have minimal interaction to senior grades (4 and 5). This is an important finding since it can be helpful in advising the authorities aiming to control the propagation of communicable diseases. It can be suggested that in case of a communicable disease breakout closing only a few classes rather than closing whole school could be a good alternative solution. The school timings for different grades can also play an important role. For example, changing the timing of the grades from morning to evening could help isolating the grades in case of a disease breakout.

LIMITATIONS

Some of the limitations of this project include:

The study was conducted only for short period of time (only for two days). We cannot draw any conclusions on what happens at longer time periods.

The study only shows the closed proximity of children but there are few diseases which spread only in case of physical contacts. The devices installed does not detect the physical contacts. This may alter conclusions for some of the diseases.