Network Topology and Tie Strength in a Primary School Network

Robin Beura |robinbeura@gmail.com

Objective of the paper:

The objective of the paper is to portray and understand contacts between youngsters at school that would help evaluate the transmission chances of respiratory diseases and recognize circumstances inside schools where the danger of transmission is higher.

An exact portrayal and comprehension of the contacts between youngsters at the school level would assist with evaluating the transmission chances of respiratory diseases, and to recognize the circumstances during school days where the danger of transmission is higher.

Data Collection method:

A closeness detecting framework dependent on radio-frequency identification devices (RFID) was deployed in a French primary school, and utilized it to gather, in an unsupervised way, time-resolved information on the face-to-face vicinity of youngsters and educators.

Data on face-to-face interactions were gathered on Thursday, October first and Friday, October second 2009. The data recorded 77,602 contact occasions between 242 people (232 kids and 10 educators). In this setting, every kid has on normal 323 contacts each day with 47 different youngsters, prompting a normal day by day association season of 176 minutes.

About the network:

The network provided is an undirected edge list. The nodes contain the attributes name, label, class name, gender. The edge contains attributes source, target, weight and count. The weights are removed for network edges with weights that are less than the median value of the original primary school network, and thus allows us to visualize the remaining edge weights as standardized edge "widths".

An initial check on the nodes and edges suggests that there are 238 nodes (i.e., the children in the primary school) and 3125 undirected edges of varying weights(i.e., the interactions between the children and teachers).

After network is also a simplified one and all edges are strongly connected.

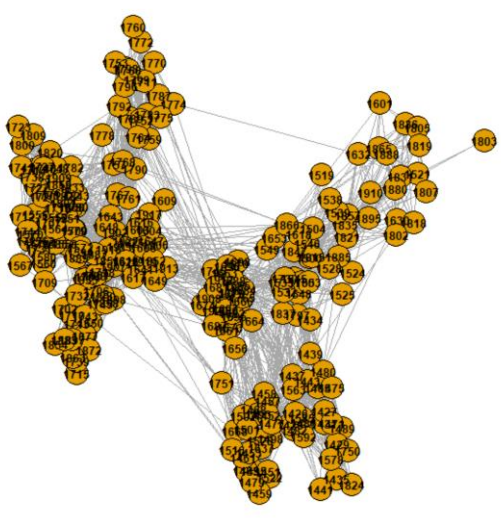

Figure 1: Fruchterman-Reingold layout representation of network

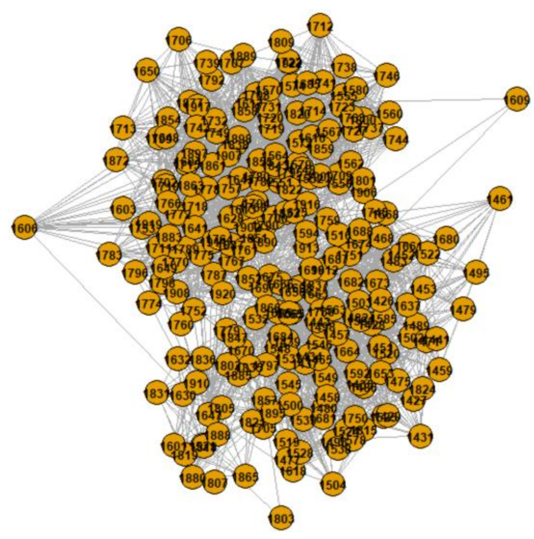

Figure 2: Kamada-Kawai layout representation of network

Basic network statistics:

Table 1: Attributes of the Primary School Network

Attributes	Values
Edges count	3125
Nodes count	238
Simple graph	True
Connected graph	True
Strong cluster	238
Weak cluster	0
Reciprocity	1
Transitivity	0.5330348
Mean distance	2.481686
Diameter (weights = NA)	5

Table 1 provides some basic statistics of the primary school network. Here we can observe that the graph is connected. We can also see that the edges are strongly connected from Figure 2 thus when we tried plotting the weakly connected plot nothing showed up.

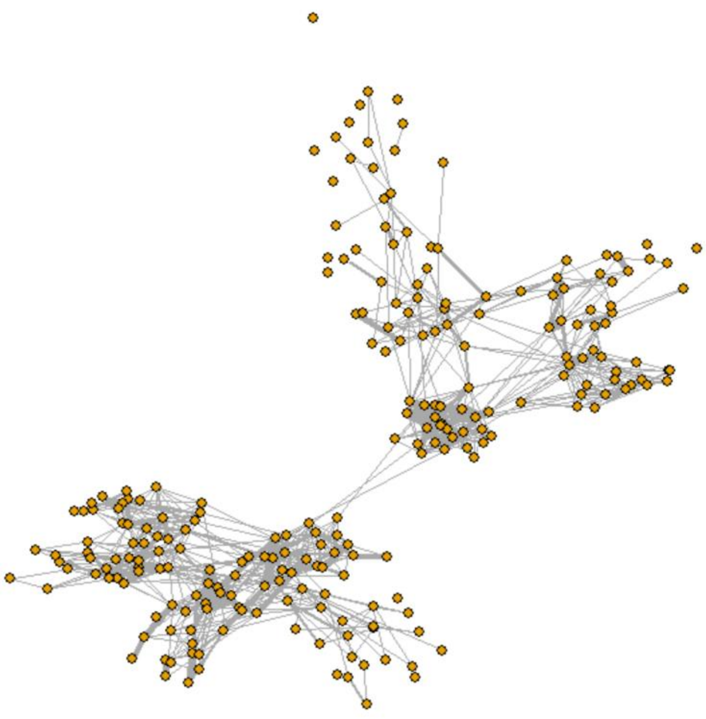

Figure 3: The network is strongly connected.

Thus there also turns out to be 238 strong cluster(viz., also the total number of nodes), and 0 weak clusters. The **transitivity** – also known as clustering coefficient, measures that probability that adjacent nodes of a network are connected - here is 0.0533 or 5.33%, which suggests that there is a grouping possiblity for children maybe in the same class or with the teachers. Also, with high **reciprocity** (i.e., 1) which measures the propensity of each edge to be a mutual edge i.e., this observation may not be that useful in the undirected network case but could have been a lot more useful in directed network. It provides information likelihood of vertices in a directed network to be mutually linked.

Average path length – calculates the length of all the shortest paths from or to the vertices in the network– is 2.482 and In undirected graphs, we can observe that the **diameter** – an index measuring the topological length or extent of a graph by counting the number of edges in the shortest path between the most distant vertices – is 5 which means if we neglect the weights we can reach from one end of a network to the other side by walking over 5 edges.

If we recall **cliques** are a maximal complete subgraph of a given network—i.e., a gathering of individuals where everyone is associated straightforwardly to every other person. "Maximal" implies that no different nodes can be added to the clique without making it less connected. Figure 4 shows one among the 8 cliques of size 18.

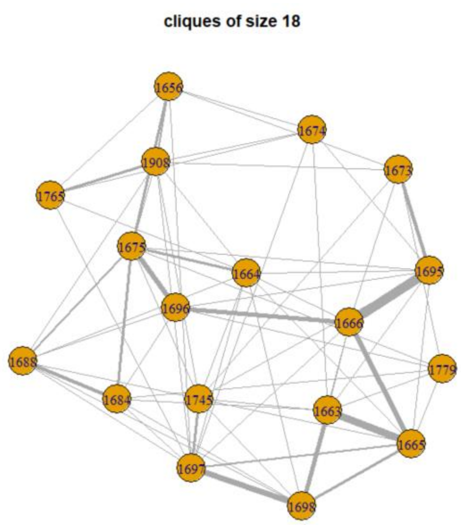

Figure 4: Shows one out of 8 cliques of size 18

Maximal Cliques	Length
11	2
47	3
122	4
190	5
277	6
268	7
296	8
189	9
147	10
83	11
54	12
50	13
38	14
14	15
28	16
20	17
8	18

Table 2: All the cliques and their sizes

For our Primary School network a census of this sort reflects that there are 11 cliques of size two and 47 cliques of size three, followed by 122 cliques of size four and the largest being 8 cliques of size 18. From Figure 4. We observe that node 1666, 1665, 1695 has maximum interactions oweing to the width of the edges they have. They may be the most influencial ones in this subgraph.

Degree centrality measures the counts how many neighbors a node has, in an undirected network this type of measure gives us the information about how influencial a node is. So, in our Primary School network, nodes 1594, 1866, 1851 have highest degree centrality which is 1758.7061, 1667.18 and 1614.95

respectively evident from Table 3, suggesting that these nodes are most has most connection and are most influencial. The surprising part is that these nodes belongs to childrens not teachers. This is important because here children are having more interactions to other children than teachers.

We checked the nodes belonging to teachers thus they are 1668, 1709, 1521, 1653, 1650, 1745, 1746, 1824, 1753, 1852 and none are present with high degree centrality.

Nodes	Degree Centrality
1594	1758.7061
1866	1667.1786
1851	1614.9481
1675	1435.3619
1698	1383.4341
1761	1317.5707
1665	1205.5005
1790	1062.0591
1751	934.439
1855	888.6279

Table 3: Degree centrality of nodes in descending order

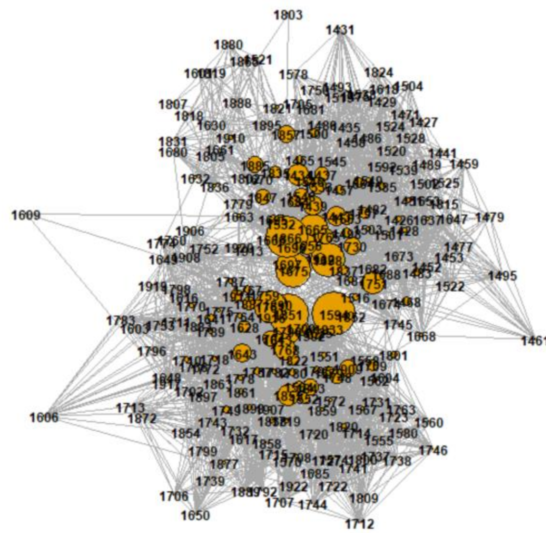

Figure 5: Nodes with high degree centrality shown

For understanding the domain where children and teachers are actively participating, detecting the communities become important. We chose walk trap algorithm for this task . Walk trap algorithm – which tries to find densely connected subgraphs, also called communities in a graph via random walks. The idea is that short random walks tend to stay in the same community.

The communities can be easily seen on the graph and the related cluster details can be seen in the table on the right side.

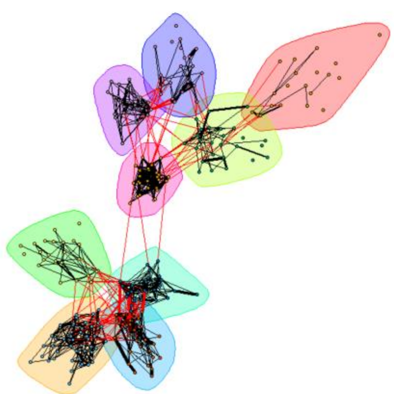

Figure 6: Community detection using Walktrap algorithm

	Class_Name										
Cluster	1A	1B	2A	2B	3A	3B	4A	4B	5A	5B	Teachers
1	0	0	0	0	0	0	0	19	0	0	1
2	0	0	0	0	23	21	0	0	0	0	2
3	0	0	0	0	0	0	21	3	0	1	1
4	23	0	0	0	0	0	0	0	0	0	1
5	0	0	23	0	0	0	0	0	0	0	1
6	0	0	0	26	0	0	0	0	0	0	1
7	0	0	0	0	0	0	0	0	0	22	1
8	0	0	0	0	0	0	0	0	21	0	1
9	0	25	0	0	0	0	0	0	0	0	1

Table 4: Community detection results based on class_name

The table above shows that the students in 3A, 3B are assigned to the same cluster like 4A and 4B. This shows that there is interactions between children in class 3A and 3B, 4A and 4B.

The figure below shows that the weight of the edge which signifies the strength of the ties between the nodes is directly proportional to the edge betweenness. This in turn in accordance with global efficiency hypothesis which states the relationship between tie strength and betweenness is positive.

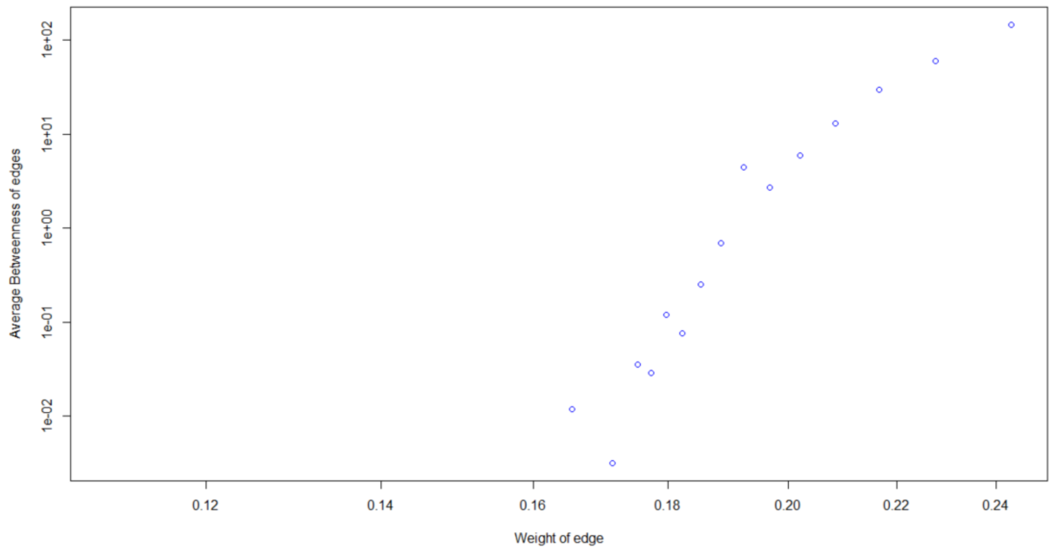

Figure 7: Shows the relationship between the Weight of edge and degree betweenness of edges on Primary School Network

The figure below shows nodes degree against transitivity or cluster coefficient. As mentioned on the Onella et al. (2007) paper that they observed the local relationship between the network topology and tie strength effects on global information diffusion. Both strong and weak ties have relatively insignificant effect on diffusion of information.

1) The high degree nodes have less clustering coefficient suggesting that they are influential nodes and thus the information diffusion would be high in that case. The only disadvantage here is that they do not have high significance.
2) The lower degree nodes have less airtime for any interaction and hence less information diffusion. Therefore, like high degree nodes it has little significance in information diffusion.

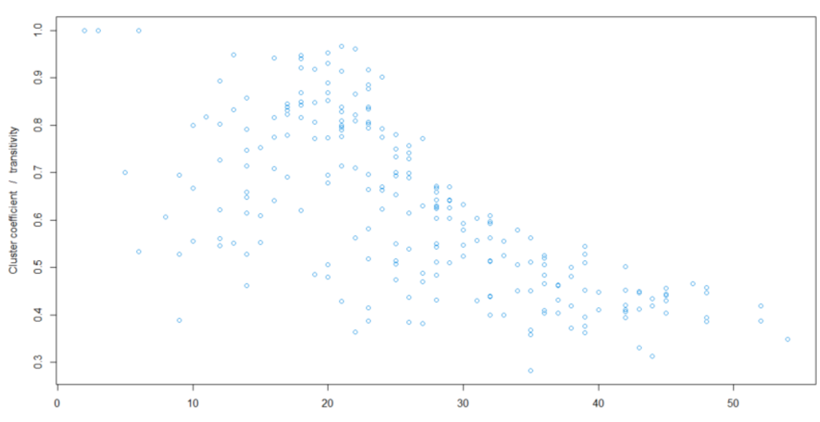

Figure 8: Shows the relationship between the degree and transitivity on Primary School Network

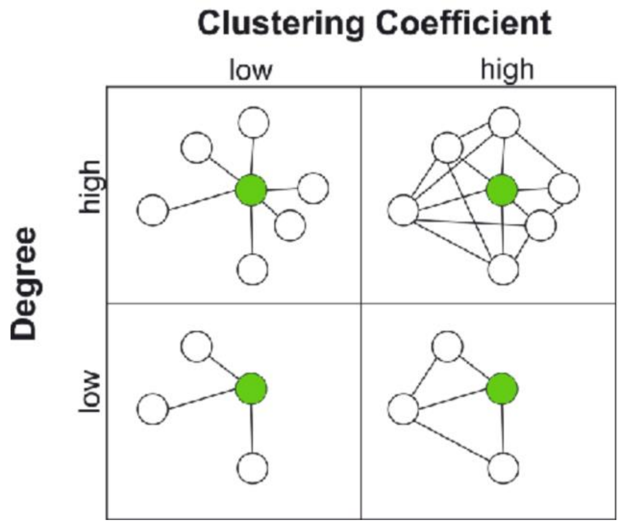

Figure 9: Explains the relationship between degree and clustering coefficient.

Reference:

[1] https://www.researchgate.net/figure/Representation-of-the-clustering-coefficient-and-degree-Networks-have-nodes-with_fig5_261995698

[2] https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0023176#s4

[3] https://www.pnas.org/content/104/18/7332