Social Network Analytics

In this Project the Onnela et al. (2007) mobile communication network and the primary school social interaction network are compared. We are basically looking to see which principal these networks follow. The global efficiency or the triadic closure theories.

I explored the following concepts in this report to find out which theory our networks follow:

- Community detection
- Homophily
- Tie strength (strong and weak ties)
- Edge Betweenness
- Neighborhood Overlap
- Closeness, betweenness, and eigen centrality

1-Primary School Network

1.1-Community detection algorithm in Primary School Network

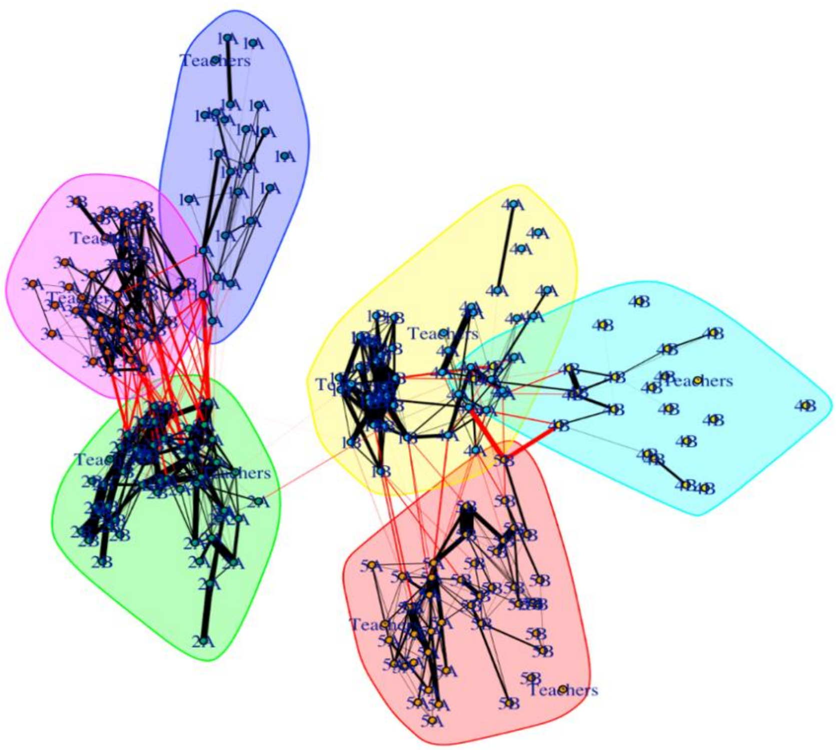

Figure1- The community detection by walktrap algorithm

wc.member \ stud.class	1A	1B	2A	2B	3A	3B	4A	4B	5A	5B	Teachers
1	0	0	0	0	0	0	0	0	21	23	2
2	0	25	0	0	0	0	21	0	0	0	2
3	0	0	23	26	0	0	0	0	0	0	2
4	0	0	0	0	0	0	0	22	0	0	1
5	23	0	0	0	0	0	0	0	0	0	1
6	0	0	0	0	23	21	0	0	0	0	2

Table 1- The distribution of each grade and classes in each community

The figure 1 is shows six communities in the Primary school network. I used multiple algorithms such as fast greedy, walktrap, spinglass, label propagation, and girvan-newman. In my opinion the walktrap algorithm works better because it does not divide the main communities to more sub-communities. As you can see in the table-1 most of students who are in the same grade are in the same community. In addition, the figure-1 indicates that each grade classes have created a local community and the students interact more with each other rather than other grade's students. This graph made me think that there might be a Homophily in this network. So, I tested homophily theory in this network.

1.2- Homophily

Category(class)	Number of Nodes	p	q	2pq	Number of Edges connecting to out of the class/total edges	Homophily
1A	23	23/228= 0.1	0.9	0.18	164/3125=0.05	YES
1B	25	0.11	0.89	0.20	0.08	YES
2A	23	0.1	0.9	0.18	0.07	YES
2B	26	0.11	0.89	0.20	0.08	YES

Table-2 Homophily

In Table.1, I calculated the homophily detection formula taught in the class. As you can see, the result completely indicates the Homophily in the primary school network. As Triadic Closure Principle is driven by intrinsic network structure and social context (Homophily), I ddecided to explore whether the primary school's interaction network is following the Triadic Closure Principle or not. So, I needed to answer these questions:

1- How does Betweenness correspond with tie strength?
2- How does betweenness correspond with neighborhood overlap?
3- How does neighborhood overlap correspond with tie strength?

1.3- Tie Strength, Neighborhood Overlap, Edge Betweenness

Firstly, I applied the professor's codes and removed the edges containing the weight lower than median(60). Then, I computed the edge betweenness based on the edge.betweenness function in the igraph library. As there were no function to calculate the neighborhood overlap, I wrote my own codes to calculate the neighborhood overlap:

```
ST <- read.table('edge2.csv', sep = ',')

S = ST[,1]
T = ST[,2]

OVERLAP = as.data.frame(matrix(nrow = length(S),ncol = 2))

for (i in 2:length(S)) {
  A = ST[i,1]
  B = ST[i,2]
  Connected_to_A = unique(ST[ST[,1] == A, 2])
  Connected_to_B = unique(ST[ST[,1] == B, 2])
  Union_AB = union(Connected_to_A, Connected_to_B)
  Intersect_AB = intersect(Connected_to_A, Connected_to_B)
  NAME = paste(as.character(A), " -> ", as.character(B))
  OVERLAP[i, 1] = NAME
  OVERLAP[i, 2] = length(Intersect_AB)/length(Union_AB)
}
```

Figure2- The self-written Neighborhood Overlap function

To calculate the tie strength, I read the Onnela et al. (2007) and Stehle et al. (2011) articles. Onnela assumed the weight of relationship between two individuals the the value tie strength (and it was the total time of mobile call between two individuals). Stehle et al. (2011) did not mentioned directly how they calculated the tie strength. Thus, I applied the Onnela et al. (2007)'s method to calculate the tie strength. Finally, I required to the measured values be normalized and I normalized them in range of [0,1].

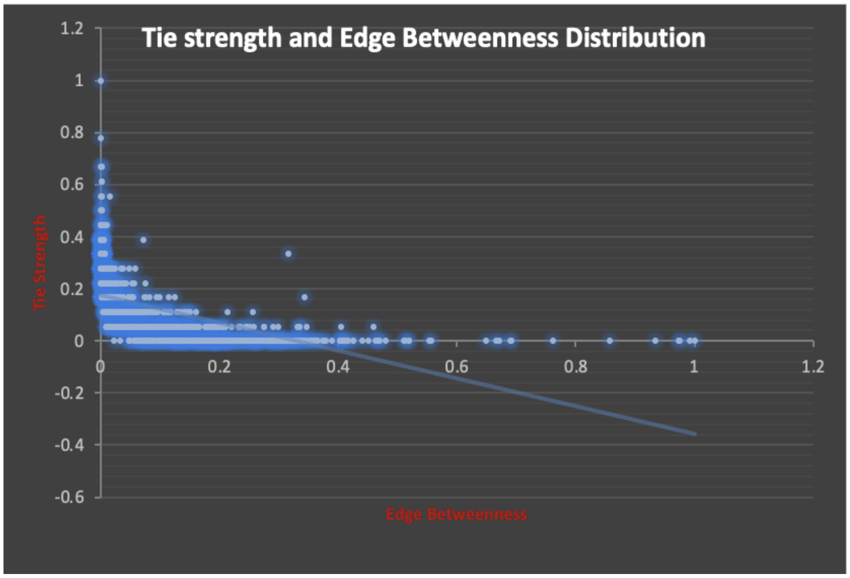

Figure3- The Tie Strength and Edge Betweenness relationship

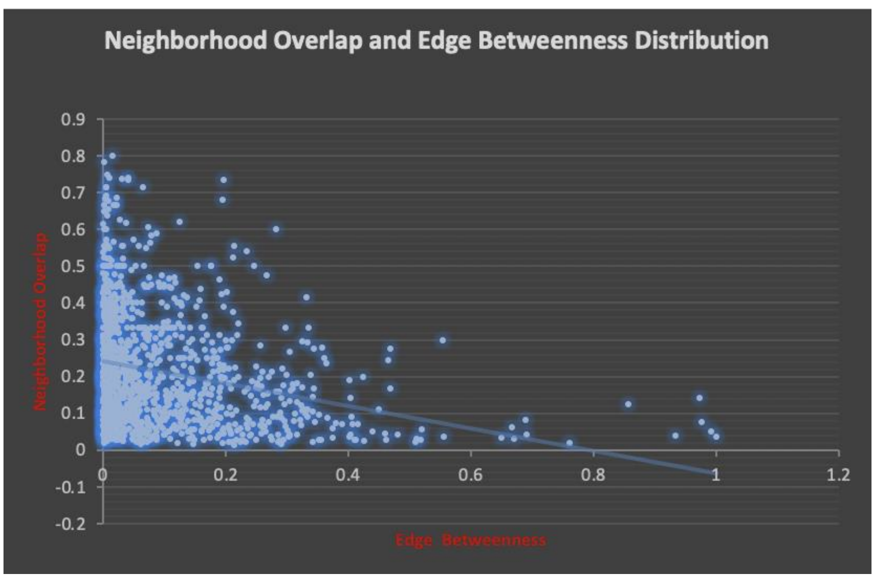

Figure4- The Neighborhood Overlap and Edge Betweenness relationship

The figure-3 displays that the tie strength and edge betweenness has an inverse relationship. I draw a regression line to show this inverse relationship. Based on figure-4, we can clearly see that the neighborhood overlap and edge betweenness has an inverse relationship too.

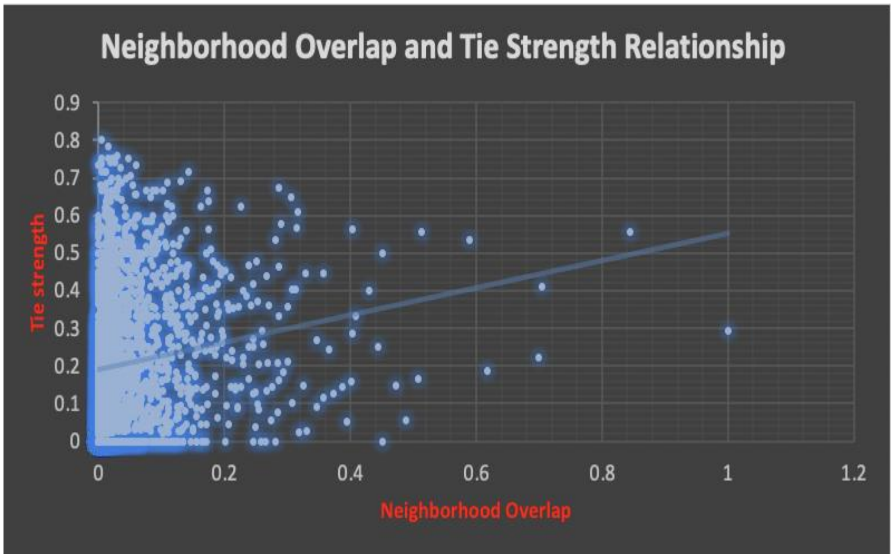

Figure5- The Neighborhood Overlap and Tie strength Relationship

Figure-5 represents a positive relationship between Tie strength and Neighborhood overlap but the data points are more scattered in this plot and do not show s strong positive relationship.

All the above evidence proves that the Primary school network like most of social networks is following the Triadic Closure Principle (weak ties). Based on the below figure we can explain the features of a network which is following the Triadic Closure Principal.

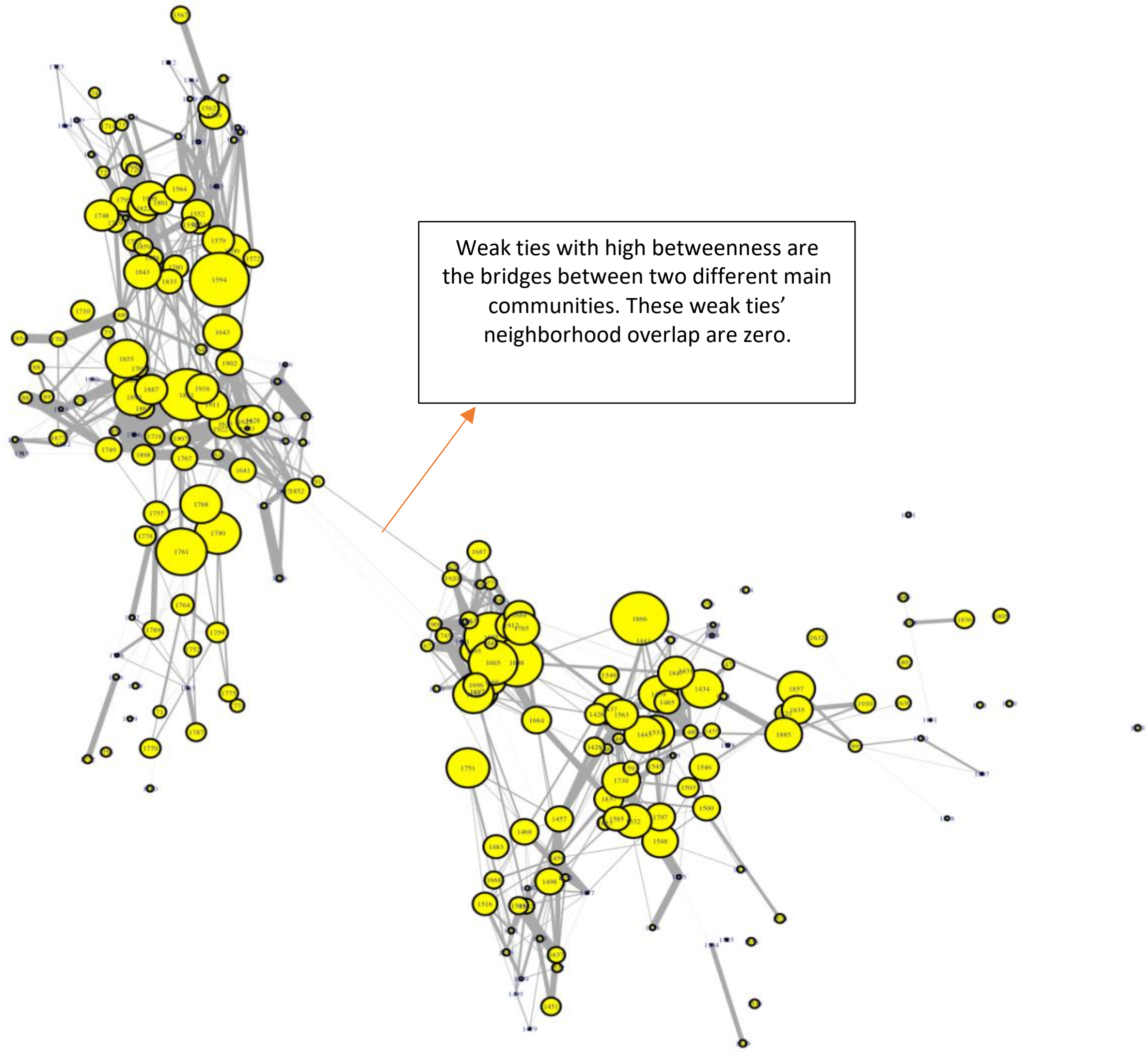

Figure6- The Primary School plot based on weighted edges and the different sizes of nodes based on

In figure 6, the nodes size is based on their betweenness centrality, and the edges width is based on is based on the weight. In this plot we can see three weak ties which are playing the role of a bridge between two main communities. These weak ties play an important role to transfer the information from one main community to another big community. Weak ties appear to be crucial for maintaining the network's structural integrity, but strong ties play an important role in maintaining local communities. Onnela(2007). Strong ties are prevalent in local communities and their edge betweenness is low. The removal of these strong ties has low impact on the network's overall integrity. The edges' weights and betweenness centrality are negatively correlated in the primary school network. These findings were proven in the scatter previous plots too and the following tables:

	Neighborhood Overlap	Edge Betweenness
Edge 1665-1761: Weak tie (bridge)	0	723.5
Edge 1439-1890: Weak tie (bridge)	0	573.2
Edge 1700-1866: Weak tie (bridge)	0	569.1

Table-3 Weakest ties neighborhood Overlap and Edge Betweenness

	Neighborhood Overlap	Edge Betweenness
Edge 1522-1502: Strongest tie located in a local community	0.8	0

Table-4 Strongest tie neighborhood Overlap and Edge Betweenness

1.4- Nodes' Betweenness and Closeness Centrality

I also computed the Betweenness centrality and closeness centrality for the Primary School network's nodes:

	closeness	betweenness
1594	0.011630617	1758.7061
1866	0.011687217	1667.1786
1851	0.011843713	1614.9481
1675	0.011227130	1435.3619
1698	0.012196393	1383.4341
1761	0.010478291	1317.5707

Table-5 Highest Betweenness and Closeness Centrality

The table 5 displays the nodes with highest betweenness and closeness centrality. These nodes are among the local communities and have some strong relationship with the nodes in their communities. In order to calculate the closeness centrality I replaced the weights with the inverse of the weights, because in social networks a node's closeness centrality has a negative relationship with its edges' weights.

2- Onnela et al. (2007) mobile communication network:

2.1- mobile communication network community detection

`

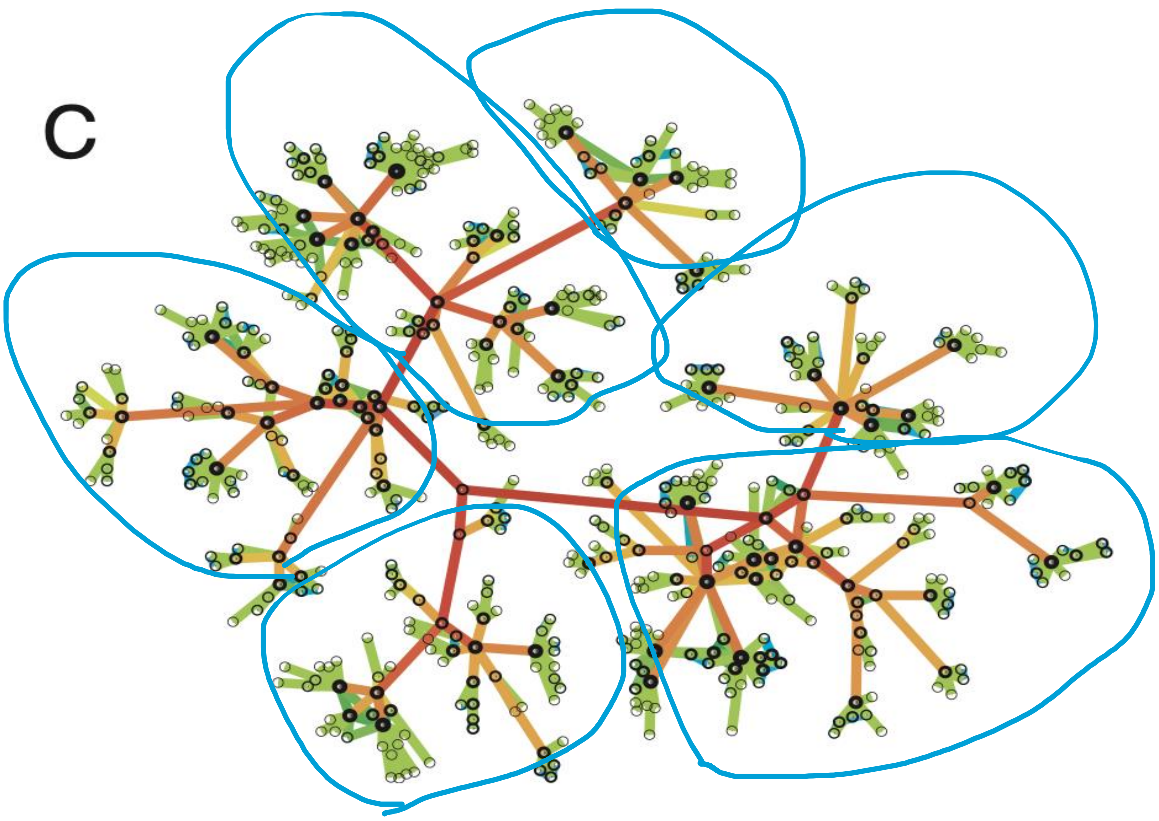

Figure7- The mobile network community detection

In the Figure-7 the Onnela et al. (2007) show different local communities. "The links connecting communities have high values (red), whereas the links within the communities have low values (green)." (Onnela et al, 2007, P.3)

2.1- Tie Strength, Neighborhood Overlap, Edge Betweenness

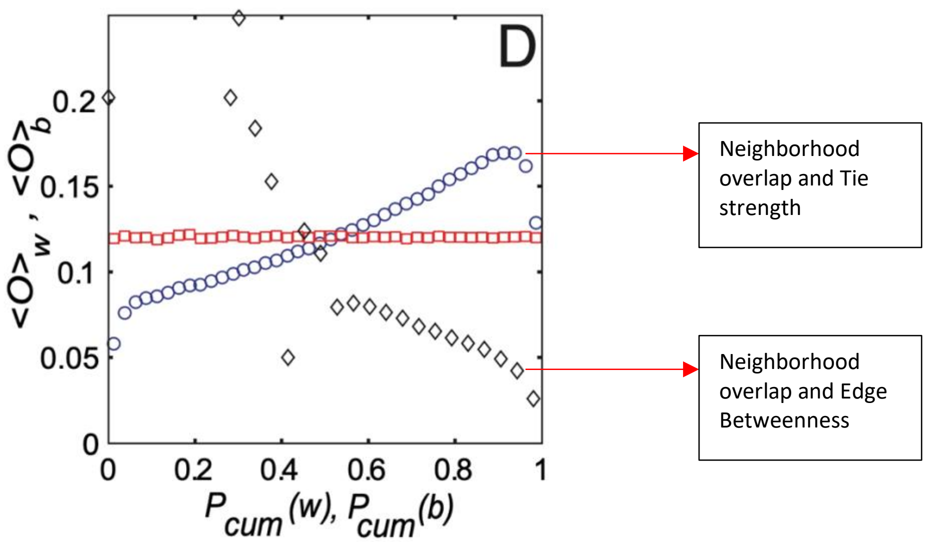

Figure8- The relationship between the Neighborhood Overlap and Tie strength and the Neighborhood Overlap and Edge Betweenness

In this plot the Onnela et al. (2007) displays that the Neighborhood Overlap and Tie strength has a direct and positive relationship. However and the Neighborhood Overlap and Edge Betweenness has inverse relationship.

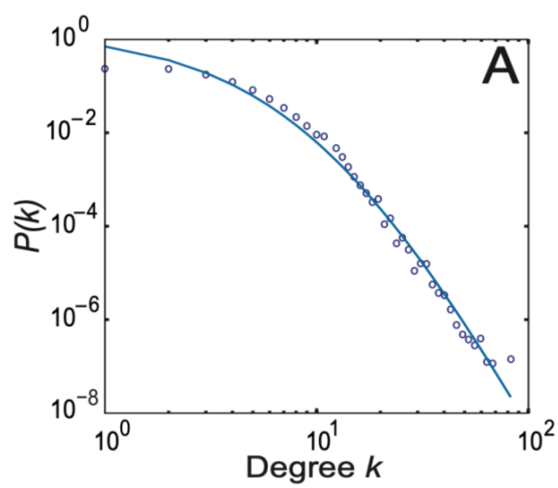

Figure9- The Distribution of Degree Centrality

Figure-9 displays that some people are calling few people and some of the individuals calling many people. Onnela et al. (2007)

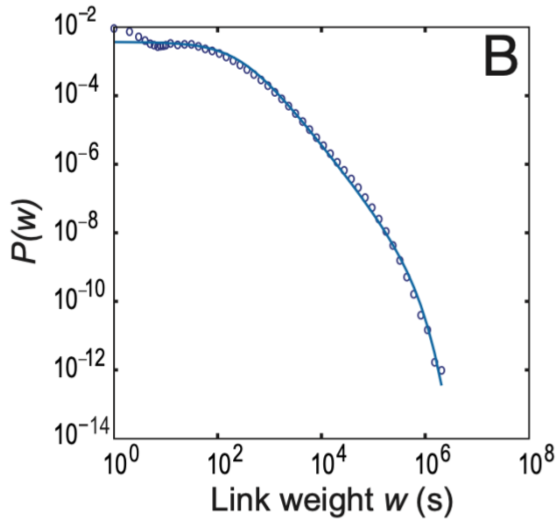

Figure10- The Distribution of weights

Figure-10 shows that some people's call duration is low but some individuals' call duration is low. The Onnela et al. (2007) assumed the call duration as weights in the network.

Conclusion

The main purpose of theses analysis is to determine which principal these networks follow. Based on the Onnela et al. (2007) survey we have already known these two principles two explain the networks observations:

1-Global efficiency principle which means "the tie strengths are optimized to maximize the overall flow in the network" (Onnela et al, 2007, P.4).

2-Dyadic hypothesis which indicates that "the strength of a particular tie depends only on the nature of the relationship between two individuals and is thus independent of the network surrounding the tie." (Onnela et al, 2007, P.4).

However, the Onnela et al. (2007) show that tie strengths correlate with the local network structure around the tie, and both the dyadic hypothesis and the global efficiency principle are unable to account for the empirical observations, so the triadic closure principle is the best theory to explain the Onnela et al. (2007) survey observation. My analysis on primary school network and the Onnela et al. (2007) analysis on mobile communication network answered similarly to these questions:

1-How does Betweenness correspond with tie strength? They have inverse relationship.

2-How does betweenness correspond with neighborhood overlap? They have inverse relationship.

3-How does neighborhood overlap correspond with tie strength? They have direct relationship.

So, based on these findings both the primary school and Onnela et al. (2007) mobile communication network is following the triadic closure principle.

References

Kolaczyk, Eric D., and Gábor Csárdi. *Statistical analysis of network data with R*. Vol. 65. Springer, 2014. (KC)

Onnela, J-P., Jari Saramäki, Jorkki Hyvönen, György Szabó, David Lazer, Kimmo Kaski, János Kertész, and A-L. Barabási. (2007). Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences*, *104*(18), 7332-7336.

Stehle´ J, Voirin N, Barrat A, Cattuto C, Isella L, et al. (2011) High-Resolution Measurements of Face-to-Face Contact Patterns in a Primary School. PLoS ONE 6(8): e23176. doi:10.1371/journal.pone.0023176