Proc. ICCCN'18, Hangzhou, China, July 30-August 2, 2018

Interaction Patterns in a Multilayer Social Network

Ashwin Bahulkar
Department of Computer Science
Rensselaer Polytechnic Institute
Troy, NY, USA
bahula@rpi.edu

Boleslaw K. Szymanski
Department of Computer Science
Rensselaer Polytechnic Institute
Troy, NY, USA
szymab@rpi.edu

***Abstract*—In this paper, we study the interaction patterns among university students whose interactions are recorded in an aligned multi-layer social network. One layer of this network represents the smartphone communications, including calls and text messages. The other layer represents face-to-face interactions. We analyze this multi-layer network to find whether communication and face-to-face interactions are correlated, and what impact the aging, that is the growing over time familiarity of each node with its peers, has on students' interaction patterns. We also investigate to what extent the academic year structure and external events, such as holidays, affect the network and the interactions between the nodes, and how the individual's communication pattern profile varies as a function of the node's degree and intensity of its interactions. The results that we obtained shed a light on how students' interaction patterns are impacted by the structure of a social network, its age, and social profiles of its nodes.**

***Index Terms*—multi-layer network, interaction patterns, face-to-face interactions, social network aging**

I. Introduction

In this paper, we study the NetSense dataset to understand the nature of interactions among university students represented by its nodes. The network embedded in the NetSense dataset is node-aligned and multi-layered. The two layers represent different kinds of student interactions. One layer represents smartphone communications between the nodes, including calls and text messages exchanged between the students. The second layer represents face-to-face interactions among students implied by their close proximity as measured by the Bluetooth interfaces of their smartphones. The uniqueness of this network is that we can observe the evolution of the network from the beginning of social group formation among freshmen arriving at the university for the first time. This means that most of the ties in the network are created from scratch since most of the students who joined the NetSense project did not know each other before arriving on the university campus. The interaction data is recorded from the time the students enter the university, and we use the first two years of this data in our analyses.

This research was supported in part by the Army Research Laboratory under Cooperative Agreement Number W911NF-09-2-0053 (the ARL Network Science CTA), and by the Office of Naval Research (ONR) Grant No. N00014-15-1-2640. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies either expressed or implied of the Army Research Laboratory, or the U.S. Government.

Our paper focuses on the following questions. How strongly correlated are the numbers of face-to-face interactions for a pair of nodes with this pair communication volumes? Does the social network aging, that is the accumulating over time familiarity among students on campus, impact the structure of network layers? How do the academic year structure and the external events affect the network and the interactions between the nodes? And finally, what communication profiles can be discovered from the interaction data?

The rest of the paper is organized as follows. In the next section, we briefly describe the NetSense dataset that we used in our study. Section III provides brief summary of related literature. Next, in Section IV, we report measurements of how strongly face-to-face interactions are correlated with communications. Section V explores various aspects of the evolution of the interactions between individuals in the social network. We look at how aging of the network affects the degree of nodes and the number of interactions with different groups of contacts. In Section VI, we discuss how the academic calendar and external events, such as holidays, affect the network and specifically how they influence the interactions between nodes. Section VII identifies the different kinds of communication profiles which can be distinguished in the network based on the degree and volume of interactions of nodes. Finally, in Section VIII, we offer conclusions based on our results.

II. The Dataset

In this section, we briefly introduce the NetSense data [1] and the multi-layer network derived from it. The NetSense project whose data is used here collected information from the smartphone of the university students participating in this project. At the beginning of the Fall semester in 2011, 200 incoming freshmen were participating in this project. Over 150 of them still participated in the Spring of 2013. The participants received free smartphones with unlimited voice and text plans. The project collected time-stamped logs of communication records for all participants. This data contains information on the date, time and duration for all calls and the date, time and character length for all text messages. Data for the first five semesters (lasting from the Fall of 2011 to the Spring of 2013) of the project is used in this paper. Data for the semesters after the Spring of 2013 was available, but we did not use that data for our analyses because of the low participation of students in the NetSense project in that period. Edges formed out of communication interactions constitute the communication layer of the network. We treat each school semester as a snapshot of one network instance. An edge exists between two nodes in a snapshot if a single call or a text message has been exchanged between them during the semester.

In addition to the described above communication layer, we also infer the face-to-face interactions between the project participants based on the Bluetooth interactions of their smartphones. The Bluetooth interactions are time-stamped as well. In addition, every Bluetooth interaction is associated with a signal strength value, which is called the received signal strength indicator, or RSSI. The RSSI values can vary between 0 dB to -120 dB, a higher value implying higher quality of signal. From all the Bluetooth interactions, we use only those which are the most likely to be face-to-face. As mentioned in [2], interactions with RSSI values above -65 dB are more likely to be face-to-face than interactions with lower RSSI values. Thus, we use only the interactions which have an RSSI value above -65 dB. Edges formed out of Bluetooth interactions form the face-to-face interaction layer in the network.

Table I lists the size of the two layers of the network for each snapshot. Clearly, the Summer semester of 2012 has very few nodes participating in the network, which is caused by the small number of undergraduate students staying on campus during the summers.

The dataset also contains surveys conducted at the beginning of every semester asking the students about their significant contacts, who could be either on or off the university campus. The surveys also ask the students about their behavioral traits. We would be using this layer only sporadically here, since much of this information has been already used and the results reported in the literature.

TABLE I
NUMBER OF NODES AND EDGES IN THE COMMUNICATION AND THE FACE-TO-FACE INTERACTION LAYERS OF THE NETWORK

Semester	Number of Nodes	Communication layer	Bluetooth layer
1 (Fall 2011)	189	509	5663
2 (Spring 2012)	185	443	4434
3 (Summer 2012)	92	104	582
4 (Fall 2012)	171	344	2537
5 (Spring 2013)	153	300	1500

III. Related Work

The detailed discussion of the NetSense project design and execution is provided in [1]. A few studies focused on co-evolution of the NetSense network layers and some of them also study the data usage by the network nodes. Reference [3] focuses on data usage by different categories of users, while in [4] the authors categorize communication links into different types and study evolution of these types over 18 months of NetSense data. Our previous work in [5] discusses how the network layers in NetSense co-evolve and to what extent the formation and dissolution of edges in one layer can be predicted by the edge status in another. Reference [6] analysis edge evolution based on the personal preferences of the students connected by this edge. In contrast, here we look at how the interaction behavior changes with time and which interaction behavioral patterns exist in different layers. Reference [7] measures the maturity of the social network using the different types of entropies computed over the communication patterns of nodes. Here, we measure the age of social network by the level of concentration of face-to-face meetings and communication volumes between the node and a small number of the most engaging peers of this node.

IV. Assessing the Strength of Correlation between Communication and Face-to-Face Interactions

We want to find how strongly correlated are face-to-face interactions with communications based on phone calls and messages. For this purpose, for each edge in the network, we draw a log-log plots of the number of face-to-face interactions against the number of calls as seen in Fig. 1 and against the number of messages shown in Fig. 2. We find that they are indeed highly correlated. We observe that edges with large number of face-to-face interactions tend to have higher volumes of communication as well. There are some edges which do not have any communication between them, but have some number of face-to-face interactions. We do not plot these edges, based on evidence that such edges are likely to have mostly random collocations [5].

An interesting observation here is that edges with very large numbers of face-to-face interactions do not have the highest volume of communication between them. In fact such volume is often lower than for edges with medium numbers of face-to-face interactions. This tendency could be the result of those edges connecting students sharing dormitory rooms, which certainly makes high volume of smartphone calls unlikely.

We also calculate the Pearson coefficient of correlation between the number of face-to-face interactions and the numbers of calls made and messages sent. We report the results for each semester in Table II. As seen in the table, except the Summer 2012 semester, which is atypical because very few students stay on campus during summers, the Pearson coefficient varies between 0.49 and 0.62 indicating the existence of significant correlation between the face-to-face contacts and the communications.

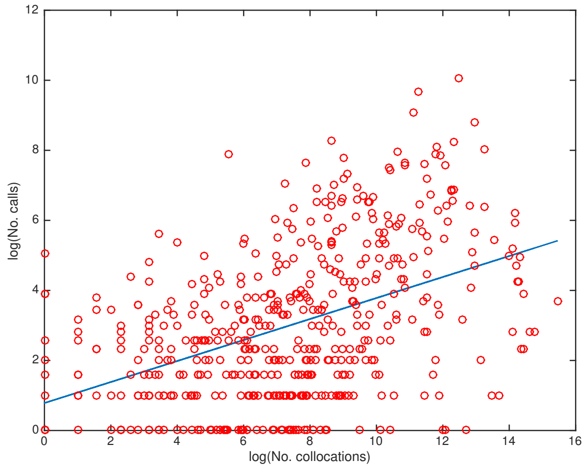

Fig. 1. The log-log plot showing for each edge the number of calls versus the number of collocations across this edge. The blue line is the least squares regression line.

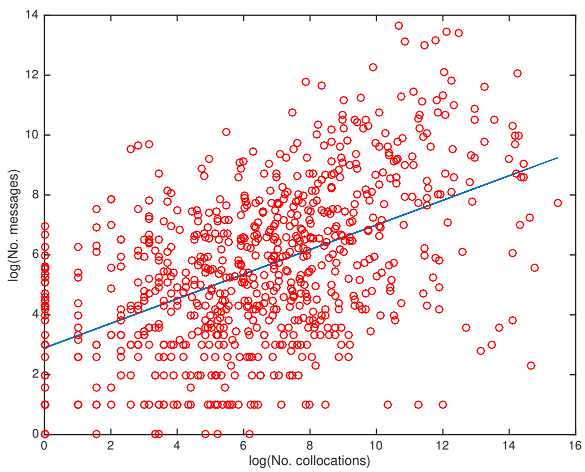

Fig. 2. The log-log plot showing for each edge the number of messages versus the number of collocations. The blue line is the least squares regression line.

V. The Process of Social Network Aging

In this section, we study changes in the nature of interactions caused by the network aging over time, which is result of students accumulating the increasing familiarity with other students. In the following subsections, we discuss how the degrees of nodes change and how interactions of each individual with contacts change.

A. *Time Evolution of Node Degrees*

We first examine what are the changes in the degrees of nodes over time. We want to find if nodes become more selective in contacting others as time passes and the network stabilizes. We plot the node degrees for every week over two years and find that node degrees measured over a week drop

TABLE II

The Pearson coefficient of correlation between the log of numbers of face-to-face interactions and the logs of numbers of calls and messages

Semester	No. calls	No. messages
1	0.51	0.49
2	0.56	0.51
3	0.22	0.34
4	0.52	0.62
5	0.49	0.60

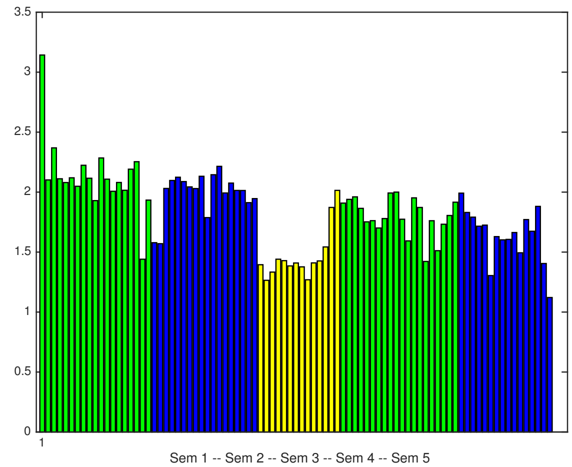

Fig. 3. The average degree of nodes per week. The colors differentiate the different semesters. The green bars denote the Fall semesters of 2011 and 2012, the blue bars denote the Spring semesters of 2012 and 2013 and the yellow bars denote the Summer semester of 2012.

on average from 2.2 to 1.7 by the end of the second year. Figure 3 illustrates the change in degree over time.

B. *Change of the Communication Volume per Contact*

While degrees drop gradually over time, number of calls and messages per contact do not necessarily drop in the same manner. We plot the number of calls made and messages sent by each node to its contacts averaged over all nodes for every week. Figures 4 and 5 illustrate the results. The averages of numbers of calls made by each node per its contacts are highest in the first semester. They drop a little in the second semester, but there are still several peaks in the plot in later times. The significant rise in the number of calls and messages per contact is evident in the summer of 2012. This is the time when few nodes in the network are active, and they tend to exchange significantly higher numbers of calls and messages among themselves than in the regular semesters. We also see peaks in the numbers spread over several weeks in the following semesters of Fall 2012 and Spring 2013. The observation is that node degrees and interactions within the network do not go hand in hand, and this indicates that merely the topology of this network is not strictly a reflection of the activity underlying the network structure. The network might

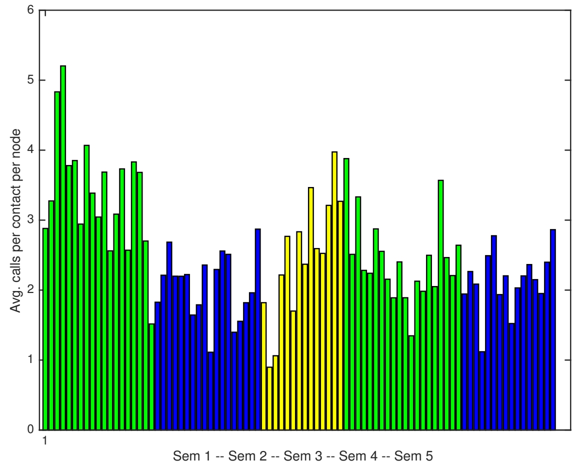

Fig. 4. The number of calls made per contact in every week.

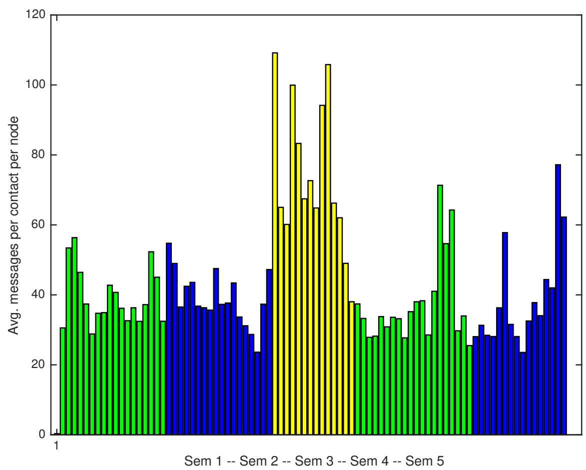

Fig. 5. The number of messages sent per contact in every week.

get sparser as time goes, however the strength of the edges does not necessarily follow this trend.

C. Changing Levels of Interactions with Top Contacts

Here, we want to measure the evolution of node interactions with their top contacts versus the interactions with the other contacts. To explore this, we look at the fraction of all interactions that every node has with contacts that are the node's top n contacts, measured in terms of fractions of messages, fractions of minutes spent on phone calls and fractions of face-to-face interactions that these nodes receive. We refer to this value as Fraction of Top Contact Interactions, $FTCI$. In addition to this, we also look at how many contacts cover the top $x\%$ of interactions of the node. We refer to the result as Number of Top Volume Contacts, $NTVC$. We compute these indicator setting n = 3 for $FTCI$ and x = 0.9 for $NTVC$. The results for each regular semester for $1 - FTCI$ and $NTVC$ are shown in Tables III and IV (the atypical summer semester was removed because of small number of students staying on campus for summers). The observed drop in both of these metrics indicates that over time the number of contacts receiving overwhelming majority of interactions narrows down. At the same time the fraction of interactions with the established top contacts increases over time. Both trends indicate that over time nodes increase their selectivity of with whom to interact.

TABLE III
THE FRACTION OF INTERACTIONS OUTSIDE OF THE TOP CONTACTS $1 - FTCI$ FOR THE MINUTES OF CALLS, AND THE NUMBERS OF MESSAGES AND OF BLUETOOTH INTERACTIONS FOR FALL AND SPRING SEMESTERS

Semester	Calls	Messages	Bluetooth interactions
1	0.018	0.053	0.314
2	0.013	0.046	0.244
4	0.003	0.026	0.232
5	0.025	0.021	0.189

TABLE IV
THE NUMBER OF TOP VOLUME CONTACTS $NTVC$ FOR THE MINUTES OF CALLS, AND THE NUMBERS OF MESSAGES AND OF BLUETOOTH INTERACTIONS FOR FALL AND SPRING SEMESTERS

Semester	Calls	Messages	Bluetooth interactions
1	1.83	2.43	12.9
2	1.63	2.32	9.22
4	1.44	2.02	8.21
5	1.45	2.02	5.95

VI. The Impact of Academic Calendar and External Events on Node Behavior Profiles and the Network Structure

In this section, we explore how the volumes of interactions are influenced by the academic calendar and events external to the network. We study the overall fluctuations of the volume of communications in the network to find if there are any time periods when this volume significantly changes. Once found, we look at these time periods to find if there are any possible external factors or events which might be affecting the interactions. We compute the numbers of calls made by every node in each week of each semester and plot the average of these numbers for all nodes across each time period as shown in Figures 6 and 7. We observe a peak in communication volume at the beginning of each semester and a drop around the Thanksgiving week of each year, which is around the end of November. We see a sharp rise in the communication volume before the end of every semester. We also see a drop around the winter break time, which is around mid-December to mid-January. An interesting observation here is that while the average number of calls per node drop significantly in the summer, messages per node rise sharply in the summer, indicating a switch of the preferred form of communication from calls to text messages.

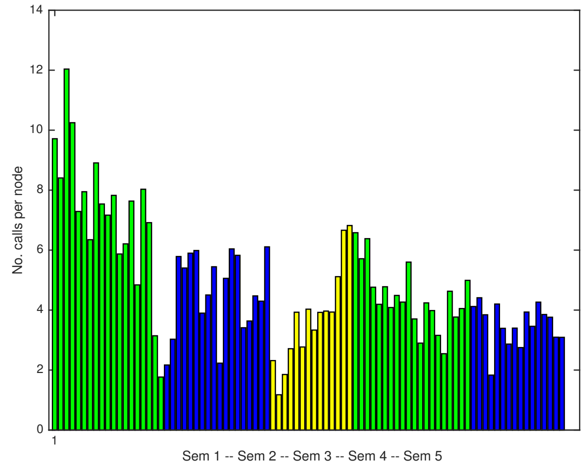

Fig. 6. The average number of calls made by each node in every week.

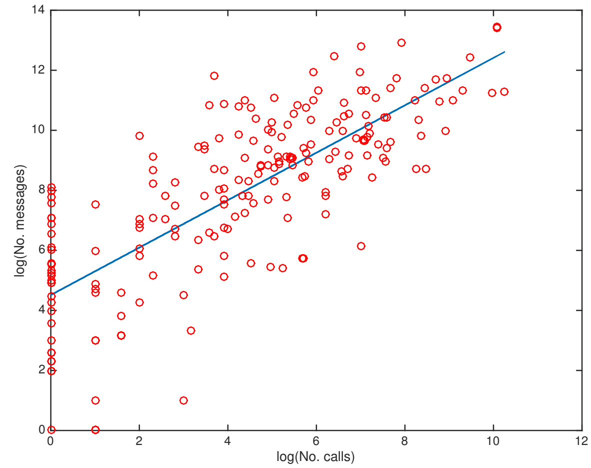

Fig. 8. The log-log plot showing for each edge the number pf calls made and the number of messages sent across this edge. The blue line is the least squares regression line.

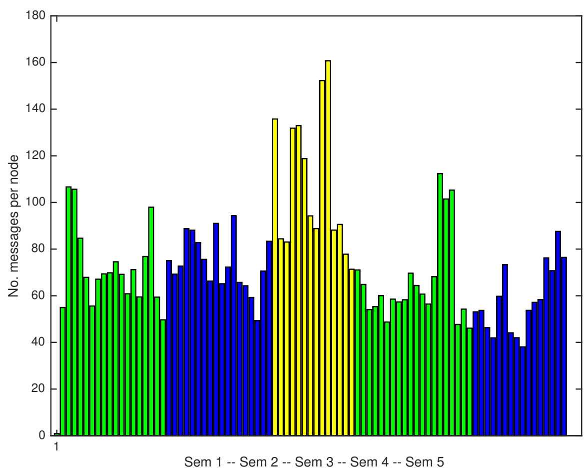

Fig. 7. The average number of messages sent by each node in every week.

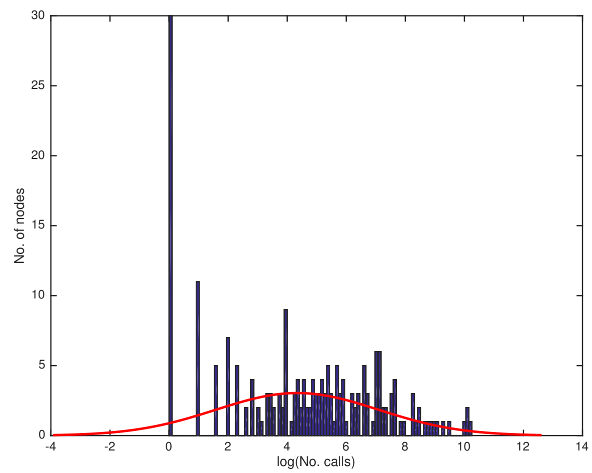

Fig. 9. The distribution of the number of nodes as a function of the log of number of calls made by the node. The red curve fits the Normal Distribution to the data.

VII. Node Profiles

In this section we identify the different kinds of node interaction profiles based on the interaction volumes and the degrees of nodes.

A. Correlation between Calls and Messages

In this subsection we look at how strongly the total number of calls made and messages sent by nodes are correlated with each other. Figure 8 illustrate this correlation for the first semester, and since all semesters plots have similar shape, those plots are not repeated here for the sake of saving space. We observe a good correlation between the number of calls and the number of messages made by nodes. The Pearson coefficients for correlation between the number of calls and the number of messages are consistently around 0.8 for all the semesters.

B. Distribution of Calls and Messages among Nodes

We plot here the distribution of the numbers of nodes as a function of the log of number of calls made and messages sent by each node. Figures 9 and 10 illustrate these distributions for one of the semesters, the rest of the semesters have a similar distribution so are not repeated here for the sake of saving space. We observe that the plot is close to the normal distribution when the nodes which make zero calls and send zero messages are excluded because the lack of communications indicates random character of the collocations.

C. Volume of Interactions versus Node Degree

We want to find here if there is any relationship between node degree and the total volume of interaction with each

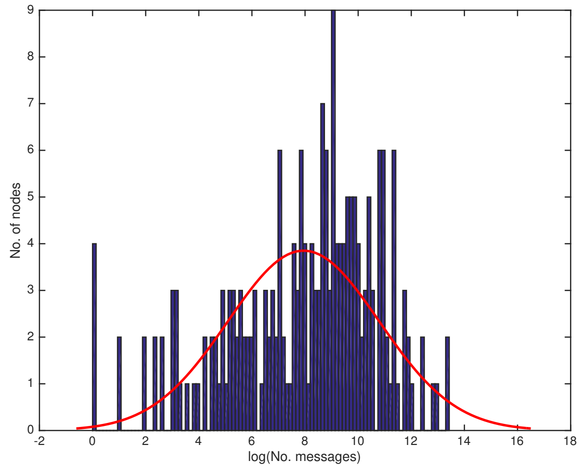

Fig. 10. The distribution of the number of nodes as a function of the log of number of messages sent by the node. The red curve fits the Normal Distribution to the data.

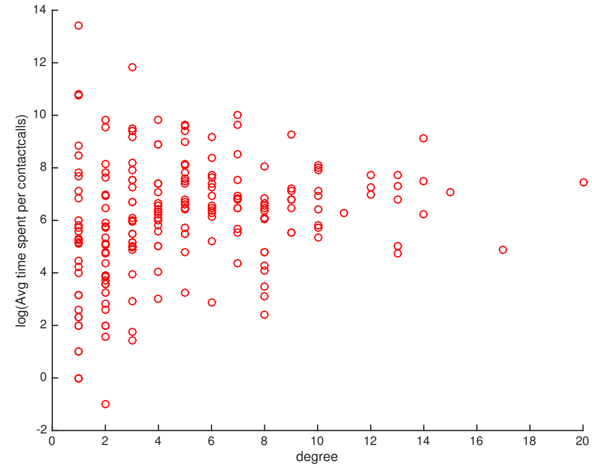

Fig. 11. The plot showing for each node the log of number of minutes spent per contact versus the node degree during the first semester.

contact. We plot the average volume of communication with each node represented by the sum of total number of minutes spent on calls and the number of messages as a function of the degree of a node. Figure 11 shows this plot for the first semester. The plots for the remaining semesters are similar, so they are not included here for the sake of saving space. The figure shows that there is a variation in the behavior of individuals with low degrees; some seem to spend a lot of time with each contact, while others seem to have few interactions with each of their contacts. However, we observe that as the degree of nodes increase, the nodes are much less likely to spend on average very large amounts of time with their contacts. However, they still seem to spend significant amounts of time with each contact.

We also plot the average number of face-to-face interactions with each contact versus the degree of a node. Figure 12 shows the plots for the first semester, the plots for all the semesters are similar, so they are not included here for the sake of saving space. Unlike the communication layer where node's high degree is not associated with the high amount of interactions, high degree nodes in the face-to-face interaction layer tend to have large number of collocations with each contact. This makes clear the difference in the nature of communication and face-to-face interactions. While there is a limit of how much a high-degree node can interact with his contacts through communication, such limit does not exist or is high for the face-to-face interactions even for the high-degree nodes in this layer.

D. Communication Profiles of Individuals Making Large Number of Calls

Here, we measure how much nodes differ in terms of volume of interactions with their contacts. A large variance in such volume, measured by the standard deviation in the number of interactions per contact, might reveal different objectives of communication. A large variance would imply having a few contacts with a large volume of communication, and the remaining contacts involved with low volume of communication, revealing a strong preferences in personal and professional/functional contacts, and low uniformity in relations. In contrast, low variance might suggest no biases towards professional or personal contacts, resulting in uniformity in relations.

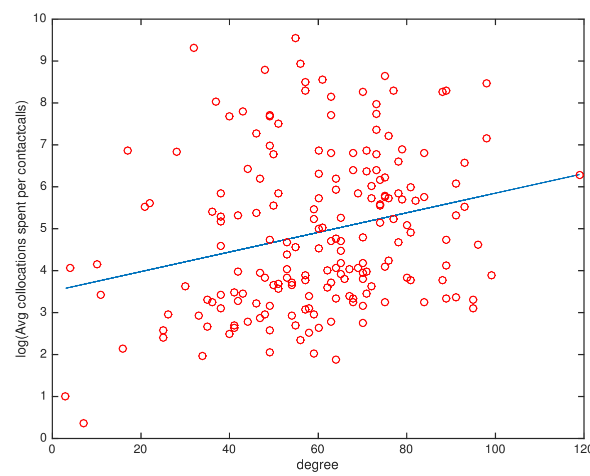

Fig. 12. The plot showing for each node the log of number of face-to-face interactions per each contact versus the node degree during the first semester. The blue line is the least squares regression line.

We first plot the degree of a node against its average standard deviation of the number of calls per contact. We find

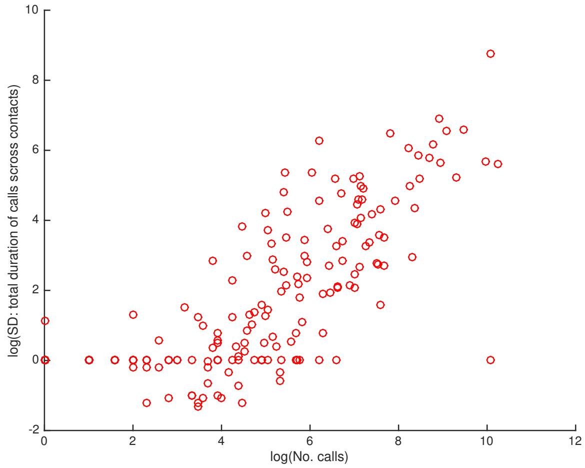

Fig. 13. The log-log plot of the standard deviation of the number of calls per contact as a function of the total number of call for each node.

that there is no strong correlation between the two. However, we find a strong correlation between the total number of calls made by a node with the standard deviation of its number of calls per contact. Figure 13 illustrates this plot. We find that nodes making a large number of calls, in general, tend to have a large standard deviation, which means that such nodes have different kinds of relationships with their contacts, differentiating between few high volume contacts and the large number of low volume contacts.

E. Dependence of Network Structure Surrounding a Node on Its Extrovert or Introvert Behavior

We explore the differences in behavioral patterns between people with introvert and extrovert personalities. We conjecture that those with introverted personalities may make fewer contacts but have stronger and more intimate interactions with their contacts. We define behavioral introverts as those having a higher than average ratio of collocation with nominated contacts to collocation with non-nominated contacts, and behavioral extroverts as those having this ratio lower than average ratio across the entire population of participants in NetSense project. We compare the number of interactions of both types, collocations and communications, and the degree of behavioral introverts and extroverts. We also verify how many of these behavioral introverts and extroverts declare themselves as such in the NetSense surveys. The results are presented in Table V. The results show that behavioral introverts tend to have more interactions, both in terms of communications and collocations with their contacts, as compared to behavioral extroverts. On the other hand, introverts tend to have lower degrees in both layers of the networks, as compared to extroverts. We also verify our labeling of introverts and extroverts based on collocation behavior with the self-reported traits of nodes.

TABLE V
DIFFERENCE IN COMMUNICATION PATTERNS BETWEEN INTROVERTS AND EXTROVERTS

Semester	Introverts	Extroverts
Number of Collocations Per Contact	400	100
Number of calls per contact	22	18
Degree: Bluetooth network	30.1	45.2
Degree: communication network	5.1	6.2
Percentage of self declared introverts	26	15

VIII. CONCLUSIONS

Our results indicate that communication and face-to-face interactions are highly correlated, with the exception of nodes with large number of face-to-face interactions which tend to be associated with lower communication volumes then predicted by the linear regression. We also discover that there is a significant change in the interactions of nodes as the network ages, with larger and larger fraction of interactions involving smaller and smaller numbers of preferred individuals. Our results demonstrate that the structure of the academic calendar and its features such as exams, the beginning and the end of semesters, as well as external events such as holidays affect significantly the interaction between the students. We also find differences in the interaction patterns of nodes based on their degrees and the numbers of calls they make. For example, we find that differences in interaction patterns of nodes with high degrees compared to nodes with low degrees are that the former have only moderate number of communication interactions with each node, while the latter maintain high communication volume with all peers.

Our results highlight the changing behavior of students as time passes, in a real life social network. While the study has its limitations given the size of the study and the bias for a particular age-group, we still are able to get an insight into how interaction patterns change over a time between people in a rapidly evolving, closed network which has developed from scratch. The work also sheds lights upon the different kinds of interaction profiles which can exist in a network. Using just the degree and the number of interactions of a node is sufficient to predict several behavioral aspects of the node.

REFERENCES

[1] Striegel, A., Liu, S., Meng, L., Poellabauer, C., Hachen, D., Lizardo, O., "Lessons learned from the netsense smartphone study," ACM SIGCOMM Computer Communication Review, vol. 43, 4, pp. 51–56, 2013.

[2] Shu, L., Yingxin, J., Striegel, A. "Face-to-Face Proximity Estimation Using Bluetooth On Smartphones," IEEE Transactions on Mobile Computing, pp. 811–823, 2003.

[3] Meng, L., Liu, S., Striegel, A., "Analyzing the impact of proximity, location, and personality on smartphone usage," Proc. IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), pp. 293–298, 2014.

[4] Meng L., Milenkovi T., Striegel A., “Systematic Dynamic and Heterogeneous Analysis of Rich Social Network Data, Complex Networks V, Studies in Computational Intelligence,“ vol 549, 2014.

[5] Bahulkar, A., Szymanski, B.K., Chan, K., Lizardo, O. “Coevolution of a multilayer node-aligned network whose layers represent different social relations,” Computational Social Networks, vol. 4, 1, 2017.

[6] Bahulkar, A., Szymanski, B.K., Chawla, N., Lizardo, O., Chan, K. “Influence of Personal Preferences on Link Dynamics in Social Networks, Complexity,” vol. 2017 (2017).

[7] Kulisiewicz, M., Kazienko, P., Szymanski, B.K., Michalski, R.,“Entropy Measures of Human Communication Dynamics,” ArXiv preprint arXiv:1801.04528, 2018.