Detecting Social Interactions through Commercial Mobile Devices

Michele Girolami, Fabio Mavilia
CNR-ISTI, Pisa, Italy
e-mail: michele.girolami@isti.cnr.it,
fabio.mavilia@isti.cnr.it

Franca Delmastro, Emanuele Distefano
CNR-IIT, Pisa, Italy
e-mail: franca.delmastro@iit.cnr.it,
emanuele.distefano@iit.cnr.it

***Abstract*—Social interactions represent an important factor in the human society and it presents different issues depending on the user category involved. In this paper, we present technological issues of using commercial mobile devices of the users to detect social interactions. Then, we propose a solution based on Bluetooth wearable Tags, minimally invasive and low-cost. This solution is based on the analysis of the RSSI emitted by BLE beacon messages and received by the user personal platform. We collected such information by exploiting commercial smartphones during a calibration campaign. To this purpose, we recruited volunteer students from the high school who mimic a number of interactions with class-mates. We compare the results of our algorithm with respect a diary of the interactions, giving us an overall accuracy of 81% and *F*-Score measure of 84%.**

Keywords—Bluetooth low energy; Proximity detection; Social interactions

I. INTRODUCTION

Human interactions are governed by the explicit willingness of establishing meaningful social relationships. A tie between two subjects is defined as a combination of the amount of time, the emotional intensity, the intimacy, and the reciprocal services that characterize the tie itself [1]. More clearly, each tie is a link between subjects, and its strength depends on several factors [2], such as the frequency of their interaction, the intimacy level, and the affinity of the subjects involved.

The automatic detection of social interactions is an emerging research field, which helps revealing complex dynamics of the society with high resolution. Specifically, understanding the way people interact can improve the work organization [4, 9], the monitoring of the social dynamics of specific users' categories (i.e., teenagers or elderly), and help understanding the spread of transmissible infectious diseases [5 - 8].

This work investigates the possibility of detecting social interactions among students of a high school by using their personal mobile devices and unobtrusive sensing technologies through BLE technology. To this aim, we propose a novel algorithm called SocializeME Detector (SME-D), designed to analyze BLE beacons and to detect social ties on a temporal scale. In addition, we present a preliminary calibration campaign of the algorithm based on real experiments conducted with students of I.T.I.S. E. Fermi high-school located in Lucca, Italy. We built a dataset of interactions obtained by reproducing accurate tests combined with a diary of the ground-truth of such interactions.

Our contribution differs from previous studies on this field for several aspects. Firstly, we investigate the possibility of using commercial smartphones to advertise and to collect BLE beacons demonstrating that, currently, such approach is not feasible due to the heterogeneous implementation of BLE firmware in different versions of mobile OS (both Android and iOS). We present the main issues encountered on this study to provide additional feedbacks to mobile devices producers. Most of the solutions presented in the literature are based on customized sensing units [4-8], or they provide a limited test set on commercial devices [3]. Differently, we rely on commercial devices, leading us to consider several drawbacks. For example, we experienced a remarkable and variable loss rate of the expected beacons; the heterogeneity of Bluetooth chipset, causes significant differences in the quality of the received signals; variations related to the *wearing* position of the smartphones (e.g., in front or back pocket, in a hand, or on a desk) impact on the detection accuracy. The combination of such factors makes our calibration campaign a representative case-study of the hidden complexities behind the detection of human interactions. Then, we compare the performance of SME-D against a fine-grained ground-truth, collected by keeping track of the start and end time of each meeting we reproduced. Therefore, we are able to assess the performance of SME-D by measuring precision, recall, accuracy and F-score metrics on a granular time scale.

Experimental results show different performance of the considered tests. Specifically, the results we obtained on specific tests show accuracy values ranging from 72% to 87% and F-score values ranging from 75% to 89%. By combining all the tests, the overall results report an accuracy of about 80% and F-score of 84%. The presented results demonstrate the effectiveness of SME-D algorithm as well as the quality of the dataset collected for calibration purposes. Section II reports innovative works in the field of human social interaction, while Section III describes the SME framework. Section IV describes the SME-D algorithm, and Section V and VI describe the data collection campaign and the results obtained.

II. RELATED WORK

The SocioPattern research project is a proximity sensing system based on RFID active tags [5 - 8]. The research group, along the years, collected many useful datasets reproducing dynamics of the interactions in different settings. All the datasets are obtained with the architecture presented in [1-5], which consists of wearable RFID sensors emitting a signal with a predefined power level, in order to locate humans in the range of 1 to 1.5 meters. RFID sensors are tuned to detect only face-to-face meetings. The interactions happening far from the recording stations are not recorded. SocioPattern have been used to monitor the evolution of the interactions of students at a high school. They recruited about 300 students during two campaigns in order to investigate and compare patterns of encounters and their temporal features. A similar work is presented in [8], in which the authors present an in-depth study of dynamics of the interaction in primary school. Differently from our approach, the environment is provisioned with a number of recording units that, generally, are installed on the ceiling.

A different approach is followed by the MIT Human Dynamics Laboratory Works leaded by prof. Pentland. Authors of [4] present the design of a smart badge able to capture what authors define as *honest signals*. Namely, those signals emitted by humans as an explicit willingness to interact. The smart badge can detect proximity and voice activity (talkativeness) in different contexts. Combining such signals, the authors are able to accurately detect meetings and to study the efficacy of the interactions in working environments. The solution proposed is highly efficient, however, it relies on ad-hoc sensing units.

The authors of the Copenhagen Networks Study [10-11] analyze the interactions of people by using a Funf-based logging application [12]. The application captures multiple signals, including WiFi scans, locations and Bluetooth scans. In this study, all the people were provisioned with a specific device model (Samsung Galaxy Nexus in 2012 and LG Nexus 4 in 2013) in order to reduce incompatibility issues but, at the same time, limiting the possibility to extend the sensing campaign. Under this respect, it is worth to notice that our approach relies on commercial and heterogeneous devices.

Recently, another work has been presented including the use of commercial mobile devices, specifically Android Wear and Tizen smartwatches, to detect proximity interactions [3]. The authors present results related to the use of BLE advertising and scan operations implemented on a customized device (developed by the authors) and on two commercial smartwatches. The work presents interesting results in a working environment involving 35 people. However, the authors did not investigate the technological issues related to the heterogeneity of available commercial devices that represents a real limitation to a large scale deployment of this type of systems in real environments.

Fig. 1 BLE Tags used for our experiments.

III. THE SOCIALIZEME FRAMEWORK

The first version of SocializeMe included a mobile application developed for Android and iOS smartphones. The app was designed to be completely transparent to the user, being able to run in background by maintaining BLE scan and advertise operations active, in order to collect information about the devices in proximity. Unfortunately, we encountered several limitations to the app implementation in the two development environments. On the one hand, iOS does not allow scan/advertise operations from an app in background. Since SocializeME would like to be used in a minimal invasive way for the final user, we cannot require an exclusive use of the personal mobile device during the monitoring operations. On the other hand, we discovered that not all Android devices (considering Android 5.0 and Bluetooth 4.0 as minimal versions) support BLE advertising even if supporting the BLE peripheral mode. Through the experimental campaign with the students, we have been able to test more than 15 different devices and 42% of them did not support the application requirements. In Table I, we present a summary of the tested commercial devices.

TABLE I. SUMMERY OF THE TESTED DEVICES

Device model	Android version	Bluetooth version	BLE Advertising
Samsung Galaxy S7 Edge	6 and 7	4.2	✓
Samsung Galaxy S6 Edge	7	4.1	✓
Samsung Galaxy S7	7	4.2	✓
Sony e5823	7	4.0	✓
Xiaomi mi5	7	4.2	✓
Xiaomi mi4i	5.0.2	4.1	✓
Asus Zenfone 2	5.0.0	4.0	✓
Motorola Nexus 6	7.0	4.1	✓
LG Nexus 5X	7.1	4.2	✓
Honor 8	7/7.1	4.1	✓
OnePlus	7.1.2	4	✗
Huawei P8 Lite	5	4	✗
Huawei P9 Lite	6.0.0	4.1	✗
Samsung Galaxy S5	5.1.1	4	✗
Sony Xperia M2	5.1.1	4.0	✗

In order to support a large-scale testbed in the school environment, we decided to move to an alternative solution. We decided to use commercial BLE Tags produced by Global Tag[1],

[1] http://www.global-tag.com

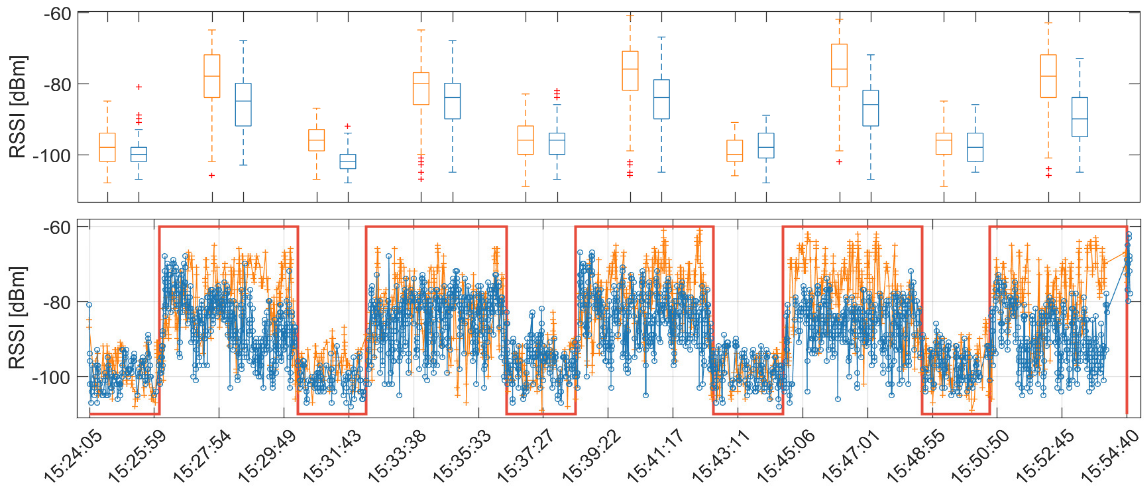

Fig. 2 Example of raw data and statyistical distribution of RSSI from a dyad.

as shown in Figure 1, since they are low cost devices, easy-to-configure and fully compliant with Bluetooth v4 protocol stack. Tags support both the Eddystone and iBeacon beaconing protocol. The advertisement rate and the transmit power can be tuned, ranging from 1 *Hz* to 10 *Hz* and from -23 *dBm* to 4 *dBm*, respectively. We maintained SocializeME app as collector of beacon signals. In addition, the students can observe the list of detected neighbors and the execution time through a simple GUI. The app locally stores the collected information and sends it to a remote server in presence of Internet connectivity.

IV. THE SME-D ALGORITHM

SocializeME app collects, for each user, a time series of the BLE beacon received from tags in proximity. The design of our algorithm has been inspired by the analysis of the BLE signals received in terms of Received Signal Strength Indicator (RSSI). Figure 2 reports a meaningful example, during which two volunteers (A and B) had 5 face-to-face meetings lasting for 4 minutes, and interleaved by a 2 minutes' pause. During the meeting, the volunteers lay at a distance ranging from 1 to 1.5 meters. The lower part of Figure 2 shows the raw data from both devices, A is shown in blue color, while B in orange color and

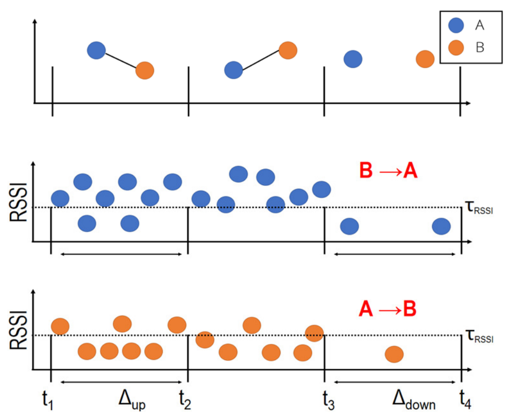

Fig. 3 The SME-D algorithm.

the ground truth of the interactions, namely the time intervals during which A and B were effective in touch (reported as a red stair-line in Figure 2).

A first observation concerns the asymmetry of the two devices. Specifically, we observe that the distribution of the RSSI estimated by A differs from the distribution of B. The upper part of Figure 2 shows a box-plot for each device and for each time interval identifying an interaction. Each box shows the median, the 25th and 75th percentile as well as the minimum and maximum RSSI observed. We can note that, even if the two volunteers stayed in the same position during the face-to-face meeting, their distributions differ. In this example, beacons received by B are generally estimated with a higher RSSI than that of beacons received by A. We repeated the same experiments by using two identical mobile devices and we still noticed a difference in the RSSI distribution.

We measure the Coefficient of Variation $CV=\sigma/|\mu|$ of the RSSI distributions (both of A and B). It indicates the stability of the distribution. Values of $CV < 1$ mean that the distribution has a low variance while with $CV > 1$ the distribution has a high variance. For what concerns the example reported in Figure 2, we experience values of $CV < 1$ ranging from 0.079 to 0.092 and from 0.104 to 0.119 for beacons received by A and B and vice-versa. We observe similar results also for all the experiments executed. More precisely, we measure that most of the meetings are *stable*, with values of CV always below 0.2.

Finally, we assess the beacon loss rate for both devices and, similarly to CV, we experience different values for the two devices. In this specific case, A and B report a 49% and 37% loss rate, respectively. This example is a representative case for most of the tests we analyzed during the calibration campaign (see Section V), highlighting the impact of RSSI analysis on the detection of social interactions. However, additional factors affect the overall *quality* and *quantity* of the beacons received. For example, the body orientation of the volunteers, the presence of other people in the nearby, electromagnetic interferences as well as the presence of physical obstacles in the environment.

Therefore, we consider two main parameters in the definition of SME-D algorithm: *(i)* the percentage of received beacons, and *(ii)* the minimum RSSI value to consider the received message as a valid indicator of physical proximity.

The goal of SME-D is to accurately estimate the start and the end time of all the interactions between all the possible dyads in a group of people (i.e., each pair of devices in proximity). SME-D processes the raw data collected from all the devices, and it reports a time series of their interactions.

For each dyad, SME-D analyzes the interactions in both directions: A→B, by analyzing beacons received by A, and B→A. Specifically, we consider a sliding time-window of duration Δ_{up}, during which we evaluate the following *opening condition*:

- to receive at least $p\%$ of the expected beacons;
- the RSSI of the received beacons is greater or equal the threshold value τ_{rssi}.

We consider that a meeting occurs in that time window if at least one of the two directions verifies the opening condition.

Once the meeting is detected, it holds until the *closing condition* is detected: the time interval between the last received beacon with RSSI $\geq \tau_{rssi}$, is grater or equal to Δ_{down}.

Figure 3 shows a graphical representation of the SME-D algorithm for the dyad A and B. In this example, the direction B→A verifies the opening condition and SME-D detects a meeting between A and B. On the contrary, after time t_3, the RSSI of the beacons received in both directions are below the threshold τ_{rssi} for at least a Δ_{down} of time. Therefore, the closing condition verifies and SME-D no longer detects a meeting between A and B.

V. The Data Collection Campaign

This research is conducted in the framework of SocializeME project, aimed at studying social dynamics among students of a high school. Social inclusion and interactions might represent an important issue in adolescent age, and the high school is the reference institution where students spend most of their time and where they establish both positive and negative relationships. In addition, teachers and the school headmaster can receive important feedback from the analysis of students' social interactions, by improving the lessons organization or monitoring some school areas used during breaks or out of the lessons. A large-scale data collection campaign is scheduled to start on January 2018 for roughly 4 months. However, in order to calibrate the SME-D algorithm and to test the proposed solution, we conducted some preliminary experiments with a limited set of students.

The calibration campaign consists of several round of tests during which we recruit volunteer students from different classes. All the volunteer's parents signed a written informed consent agreeing on the possibility of analyzing data collected for statistical purposes in an anonymized form. We deliver to the students a test plan in which we describe the tests to be done as well as the report to be filled in for the ground truth. The test plan is organized in three phases:

- Installation: the students install, check and test the SocializeME app on their mobile devices. The goal is to report bugs, anomalies or incompatibility issues with their personal smartphone.
- Tests: the students execute each run by strictly following the test plan instructions.
- Report: the students fill the ground-truth diary reporting the time intervals during which they effectively are engaged in a social interaction.

We included three kinds of tests in the test plan, according to common ways of using and *wearing* a smartphone:

- Test 1: standing face-to-face with smartphones placed in the front pocket;
- Test 2: sitting with smartphones placed in the front pocket;
- Test 3: standing face-to-face with smartphones placed in the back pocket.

Each of the test has been repeated for 5 runs. Each of the runs spans for 6 minutes, of which 2 minutes of non-interaction and 4 minute of interaction.

After each test, students fill in the ground-truth diary, in which they report: start time of the interaction, end time of the interaction and, if any, remarks about the test. At the end of the calibration campaign, we collected 300 interaction tests from 20 different dyads.

VI. Experimental Results and SME-D Calibration

Before reporting the experimental results, we describe the evaluation metrics we used in the data analysis. Then, we present the performance of each type of test described in Section V (test 1, 2 and 3), and the aggregated performance results. Specifically, we highlight how SME-D parameters p and τ_{rssi} affect the accuracy of the detection, and the distribution of the beacon loss. We configure the BLE Tags with the emission power set to -6 *dbm* and a 5 *Hz* advertisement rate. Tags emit iBeacon messages containing only the ID of the student. This configuration has been selected from a previous experimental campaign used to estimate the battery life of BLE Tags, which resulted to be more than one month with this setting.

A. Evalution Metrics

In order to assess the performance of SME-D, we compare the result of our algorithm with the ground truth. Specifically, we measure the following quantities: true positive (TP), true negative (TN), false positive (FP) and false negative (FN). These quantities assess the number of right/wrong answers of the algorithm with respect to the number of observations in the ground truth (i.e., the existence or not of an interaction for a specific dyad). Then, given the confusion matrix, we consider the following metrics:

$$accuracy = \frac{TP + TN}{\mathrm{TP + TN + FP + FN}}$$

which assesses the proportion of correct answers of SME-D with respect to the total amount of observations. We also consider the *F*-score, which combines both precision *P*=TP/(TP+FP) and recall *R*=TP/(TP+FN), as follows:

$$F - score = 2 * \frac{P * R}{P + R}$$

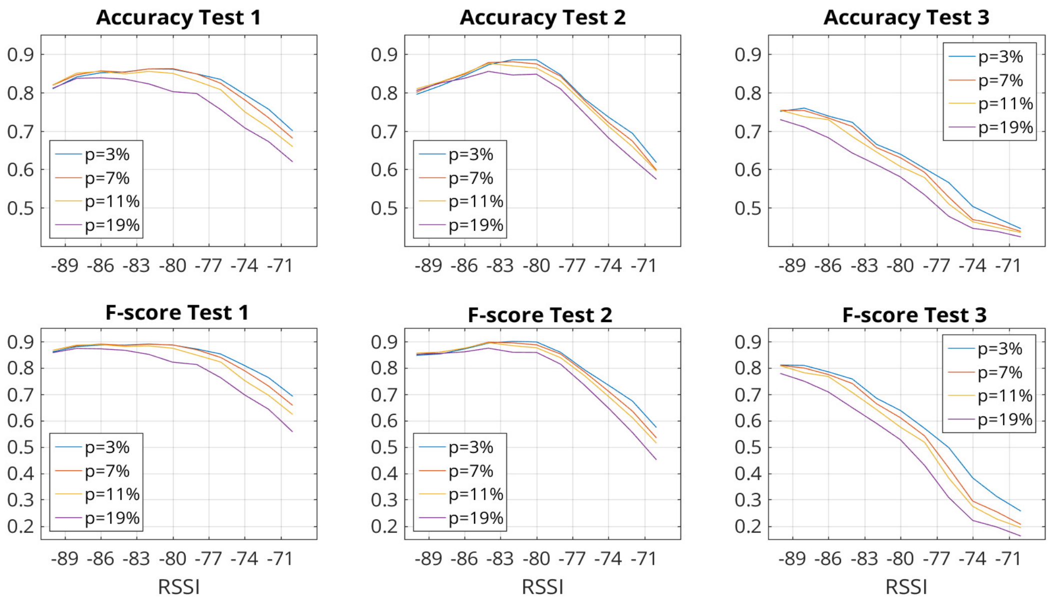

Fig. 4 Accuracy and F-score for each of the tests.

When *F*-score = 1, SME-D obtains perfect precision (*P* = 1) and perfect recall (R=1). Finally, we analyze the distribution of the packet loss rate, expressed as the ratio between the expected number of beacons and the received beacons.

B. Experimental Results

We start our analysis by considering the two SME-D parameters, namely p and τ_{rssi}. They establish the number of valid beacons to be considered for the opening and closing conditions. For each test, we measure accuracy and F-score while varying both p and τ_{rssi}. Figure 4 reports the overall results. Here we present results obtained with p equal to 3%, 7%, 11% and 19%, and τ_{rssi} in the range -90 *dbm*, - 70 *dbm*. We experience that values of p and τ_{rssi} outside those ranges do not affect positively the overall metrics.

The first observation is that the accuracy of test 1 and 2 have a similar trend. The accuracy increases as τ_{rssi} increases to reach the maximum value around a specific value, which varies from -82 *dbm* for test 1, to -84 *dbm* for test 2. After such threshold, the accuracy decreases. The increases of τ_{rssi} (from lower values to higher values) as a double effect on the positive answers of the algorithm (TP and FP). Both TP and FP increases with τ_{rssi}. This depends on the fact that beacons received with a low power (i.e., emitted by devices not in proximity) still contribute to verify the opening condition. Differently, the variation of the percentage p has the effect of increasing the number of required beacons to be considered for the opening condition. The higher p, the higher number of beacons is expected. As a result, increasing values of p reduce the number of both FP and TP.

Results for test 3 differ from test 1 and 2. During test 3, the dyads stay face-to-face with their smartphones placed in the back pocket. In this case, the body attenuation is generally stronger than that of test 1 and 2. As a result, the overall metrics are highly affected. Specifically, we do not observe any significant local maximum that can indicate a possible optimization point. Differently, both accuracy and F-Score decrease as τ_{rssi} increases.

We also study the beacon loss rate for test 3. As observed in Section III, devices record a number of beacons far lower than the expected one. For each tests, we report in Figure 5 the distribution of the beacon loss rate. Results are obtained by considering the beacon loss rate of all the 20 dyads during all the 300 meetings. For what concerns test 1 and 2 the median ranges from 55% and 50% respectively while, as expected, for test 3 it is 62% (the worst case).

Since we cannot assume a priori the way the students use and wear their smartphones, we combine the results of all the tests, providing an overall performance assessment of SME-D. Figure 6 shows both accuracy and F-score. The result trends is generally similar to test 1 and 2. Both accuracy and F-score

TABLE II. CONFUSION MATRIX FOR EACH OF THE TESTS.

	Test 1				*Test 2*				*Test 3*			
	True conditions		*Acc. [%]*	*F-score [%]*	*True conditions*		*Acc. [%]*	*F-score [%]*	*True conditions*		*Acc. [%]*	*F-score [%]*
	Int.	*No-Int.*			*Int.*	*No-Int.*			*Int.*	*No-Int.*		
SME-D: Interaction	**788**	132	85.32	88.54	**741**	89	87.19	89.28	**605**	124	72.21	75.67
SME-D: No-Interaction	72	**398**			89	**471**			265	**406**		

TABLE III. CUMULATIVE CONFUSION MATRIX.

	True conditions		Accuracy [%]	F-score [%]
	Interaction	*No Interaction*		
SME-D: Interaction	2134	345	81.56	84.7
SME-D: No-Interaction	426	1275		

increase to a maximum value after which they decrease. The optimal tuning is obtained for the following values:

$$\hat{p} = 3\%, \hat{\tau} = -84\ dbm$$

which provide an overall accuracy of 81.56% and F-Score of 84.7%. With this configuration, we report on Table II the confusion matrix for each test, in terms of TP, TN, FP, FN, accuracy and F-score. Values confirm what observed in Figure 4, where the accuracy is greater than 85% and *F*-Score greater than 88% for test 1 and test 2, while test 3 is highly affected by errors, with an accuracy of 72.21% and *F*-Score of 75.67%. Finally, we report on Table III the cumulative confusion matrix obtained by combining each matrix of Table II. We observe an overall accuracy of 81.56% and *F*-Score of 84.7%.

VII. COCLUSIONS

Interactions among people represent complex events to be detected and studied. This work, proposes an unobtrusive approach to capture such events by exploiting commercial mobile devices. We first describe the SocializeME framework designed to collect BLE beacons emitted by mobile devices. We report our experience concerning the possibility of using commercial smartphones both to advertise and to record beacons in an unobtrusive way. Then, we exploit BLE commercial tags and we analyze BLE beacons to infer proximity among subjects and, in turn, to infer an interaction among them. To this purpose, we also present the SME-D algorithm designed to detect social interactions among people with high temporal resolution. In order to validate SME-D, we organized a calibration campaign with students from the high-school with the goal of reproducing face-to-face meetings. We reproduced the meetings by following different patterns, so that to obtain a realistic case-study. Our experimental results report an accuracy and F-Score metric of about 80% and 84% respectively, even with a remarkable beacon loss rate. The results obtained provide us a solid framework for an in-depth analysis of the dynamics of the human interactions. To this end,

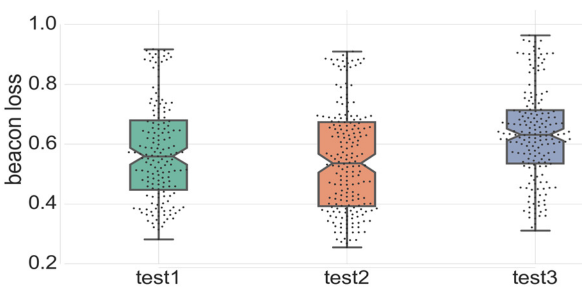

Fig. 5 Distribution of the beacon loss rate across the tests

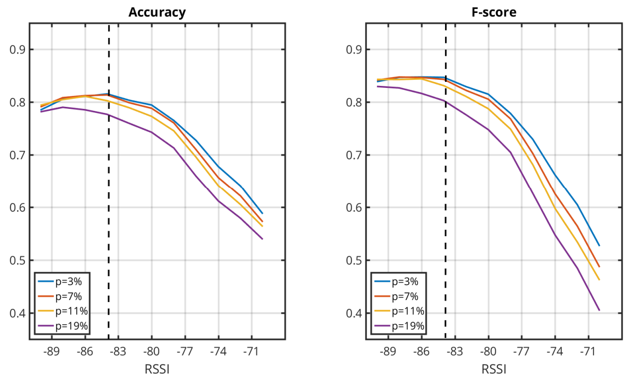

Fig. 6 Accuracy and F-score for all the tests.

we scheduled an extensive data collection campaign of about 4 month starting from January 2018 and involving about 80 students from the high-school.

ACKNOWLEDGMENT

This work is partially supported by the SocializeME project funded by "Fondazione Cassa di Risparmio di Lucca", a collaboration between CNR-ISTI/IIT and Polo Scientifico Tecnico Professionale E. Fermi – G. Girogi.

REFERENCES

[1] Granovetter, M. "The Strength of Weak Ties: A Network Theory Revisited". Sociological Theory. 1: 201–233, 1983.

[2] Marsden, P. V., and K. E. Campbell. "Measuring Tie Strength." Social forces vol. 63, no. 2 pp. 482-501, Dec. 1984.

[3] A. Montanari, S. Nawaz, C. Mascolo and K. Sailer, "A Study of Bluetooth Low Energy performance for human proximity detection in the workplace," in proceeding of PerCom 2017, Kona, HI, pp. 90-99.

[4] J.P Onnela, B. N. Waber, A.Pentland, S. Schnorf and D. Lazer, "Using sociometers to quantify social interaction patterns", Scientific Reports 4, 5604 (2014)

[5] Barrat A, Fournet J (2014), "Contact patterns among high school students". PLoS ONE 9(9):e107,878

[6] C. Cattuto, W. van den Broeck, A. Barrat, V. Colizza, J. F. Pinton, and A. Vespignani, "Dynamics of person-to-person interactions from distributed RFID sensor networks," PLoS One, vol. 5, no. 7, pp. 1–9, 2010.

[7] R. Mastrandrea, J. Fournet, and A. Barrat, "Contact patterns in a high school: A comparison between data collected using wearable sensors, contact diaries and friendship surveys," PLoS One, vol. 10, no. 9,1–26, 2015.

[8] J. Stehlé et al. "High-resolution measurements of face-to-face contact patterns in a primary school," PLoS One, vol. 6, no. 8, 2011.

[9] A. Almaatouq, L. Radaelli, A. Pentland, and E. Shmueli, "Are You Your Friends' Friend? Poor Perception of Friendship Ties Limits the Ability to Promote Behavioral Change.," PLoS One, vol. 11, no. 3, p. e0151588, Jan. 2016.

[10] V. Sekara and S. Lehmann, "The strength of friendship ties in proximity sensor data," PLoS One, vol. 9, no. 7, pp. 1–14, 2014.

[11] A. Stopczynski, V. Sekara, P. Sapiezynski, A. Cuttone, M. M. Madsen, J. E. Larsen, and S. Lehmann, "Measuring large-scale social networks with high resolution," PLoS One, vol. 9, no. 4, 2014.

[12] Aharony N, Pan W, Ip C, Khayal I, Pentland A (2011) Social fMRI: Investigating and shaping social mechanisms in the real world. Pervasive and Mobile Computing 7: 643–659