Inferring Social Activities with Mobile Sensor Networks

Trinh Minh Tri Do
Idiap Research Institute
Switzerland
do@idiap.ch

Kyriaki Kalimeri
Fondazione Bruno Kessler
Italy
kalimeri@fbk.eu

Bruno Lepri
Fondazione Bruno Kessler
Italy
lepri@fbk.eu

Fabio Pianesi
Fondazione Bruno Kessler
Italy
pianesi@fbk.eu

Daniel Gatica-Perez
Idiap Research Institute
EPFL
Switzerland
gatica@idiap.ch

ABSTRACT

While our daily activities usually involve interactions with others, the current methods on activity recognition do not often exploit the relationship between social interactions and human activity. This paper addresses the problem of interpreting social activity from human interactions captured by mobile sensing networks. Our first goal is to discover different social activities such as chatting with friends from interaction logs and then characterize them by the set of people involved, and the time and location of the occurring event. Our second goal is to perform automatic labeling of the discovered activities using predefined semantic labels such as coffee breaks, weekly meetings, or random discussions. Our analysis was conducted on a real-life interaction network sensed with Bluetooth and infrared sensors of about fifty subjects who carried sociometric badges over 6 weeks. We show that the proposed system reliably recognized coffee breaks with 99% accuracy, while weekly meetings were recognized with 88% accuracy.

Categories and Subject Descriptors

H.5.m [**Information Interfaces and Presentation**]: Miscellaneous

Keywords

face-to-face interaction; mobile sensing; activity; social computing

1. INTRODUCTION

Most of our daily activities involve interactions with others as humans are social by nature. Human activity, besides being related to location and time [21], and physical motion [3], also relates to social interactions such as having lunch with colleagues, dining with friends, and traveling with family. Yet, the relationship between human activity and social interactions has not been thoroughly explored.

One of the challenges of relating social interaction to activity comes from the nature of the collected interaction data. Online, communication, and citation networks are examples of large-scale interaction networks that have advanced the state of the art in social network analysis [1, 13, 2]. However, the analysis on these data sources is usually limited to static, aggregated patterns since these networks do not correspond to daily life interactions. The introduction of people-centric sensing has opened the opportunity to collect and analyze daily social interaction, with various sensing methods such as Bluetooth proximity or WiFi [10].

Bluetooth has been used as a platform for social interaction and communication among groups of people [14]. By analyzing the Bluetooth proximity network, a recently proposed instance of probabilistic topic model, called GroupUs, was able to discover multiple group activities from timestamped social interaction links [8]. While the discovered activity topics were relevant and could be interpreted by experts, the automatic transformation from discovered activity topics to semantically meaningful labels (e.g., meeting, chatting, eating) is still an open issue.

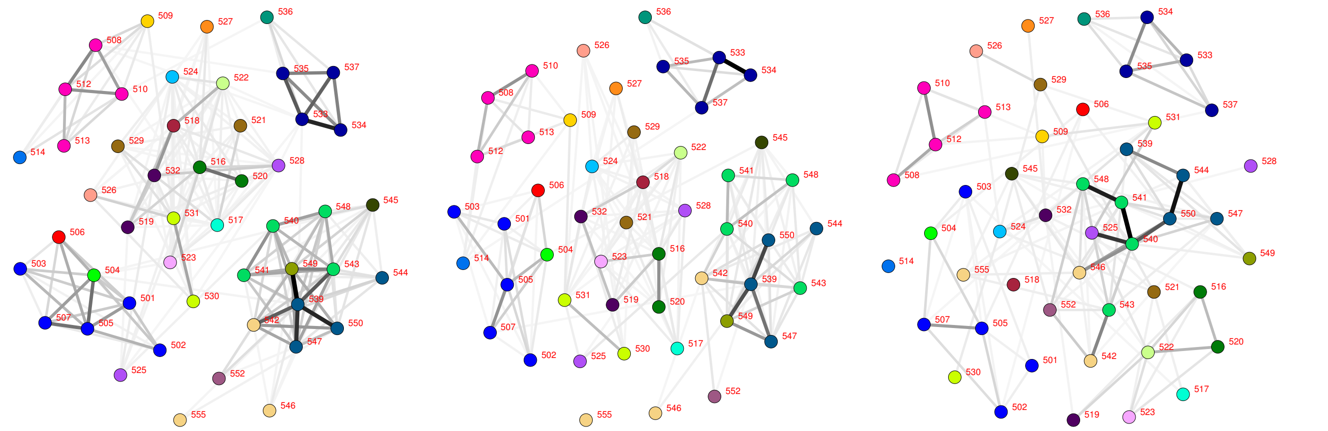
This work was inspired from the above work and our first goal is to generalize the idea to the case of multimodal interaction data. Specifically, our study was conducted on data collected from a real organization over several weeks with Bluetooth (BT) and infrared sensors (IR), where interaction links were localized by their proximity to fixed stations. Our research questions were then closely related to typical multimodal research questions: (a) How to exploit multiple types of interaction sensors, in particular BT and IR sensors?, (b) How to use location data to enhance the recognition of social interaction patterns?, and (c) How well does the method perform on this specific multimodal dataset?

Our second goal was to build a fully automatic system which senses social interactions, discovers emerging activity topics, then labels the discovered activity topics with a set of predefined activity labels. We believe that such a system has potential applications in context-aware interfaces (e.g., change the ringtone when the user enters a meeting) or individual and social behavior analysis (e.g., the dependency between social activity and mood). To achieve this, we proposed a supervised learning framework where activity topics were represented as feature vectors and a random forest was used to learn and infer if a given activity topic corresponded to a specific activity label.

This paper makes three contributions. First, we extend a recent algorithm to handle spatio-temporal context of interactions and allowed the model to work with both BT and IR data. Second, we introduce an automatic labeling method for the set of discovered activity topics. This component provides a fully automatic framework for sensing and interpreting human interaction. Finally, our paper presents a case study of interactions in a real organization, in

ICMI'13, December 9–13, 2013, Sydney, NSW, Australia

ACM 978-1-4503-2129-7/13/12 ...$15.00.
http://dx.doi.org/10.1145/2522848.2522894.



(a) Raw Bluetooth hits (b) Strong Bluetooth hits (c) Infrared hits

Figure 1: In-person interaction networks captured by Bluetooth and Infrared sensors. Nodes are colored by the office numbers.

which we provide a detailed analysis on sensing quality and how well actual activities can be recognized.

The rest of the paper is organized as follows. Section 2 discusses related work in the context of human activity and social interaction analysis. Section 3 presents the corpus on which our analysis was conducted. Then, the proposed extension of the GroupUs algorithm is presented in Section 4. Section 5 presents our findings on manual interpretation and automatic labeling of discovered activity topics. Finally, Section 6 provides concluding remarks.

2. RELATED WORK

Research on human activity has focused on individual activity inferred from a variety of data types such as video and audio, accelerometer, and indoor location [22, 6, 16]. While the number of sensors for capturing individual activity keeps increasing over time, another direction is to exploit social media for studying human activity. For example, Noulas et al. collected check-in data from Foursquare, a location-based social network which connects online social networks with the physical world, and showed that human activity varies within the course of a day or a week [18]. Using data from Twitter, Golder and Macy revealed the dependencies between mood and physical/social activities [11].

Our study on the relationship between human activity and social interactions was inspired from an emerging body of work that is investigating the possibilities of analyzing human behavior using mobile sensors. At present, Bluetooth and Wi-Fi networks allow the collection of data on specific structural and temporal aspects of social interaction, offering ways to approximate social interaction as spatial proximity or co-location of wearable devices, e.g., by means of Bluetooth hits [17, 9]. These means, however, do not always yield good proxies to the social interactions occurring among the individuals. Mobile phone traces suffer from the same problem: they can be used to model human mobility with the great advantage of easily scaling up to millions of individuals [12], but offer only rough approximations to social interaction in terms of spatial co-location. Cattuto et al. proposed a framework for monitoring social interactions that tries to reconcile scalability and resolution by means of inexpensive, active RFID devices [5]. Another strategy for behavioral data collection is to resort to image and video processing based on cameras placed in the environment [7]. This approach provides rich datasets that are, in turn, computationally complex: they require line-of-sight access to the monitored spaces and people, and specific effort for equipping the relevant physical spaces. By using sociometric badges, some previous studies revealed important insights into organizational processes, such as the impact of communications on the performance of teams [20], or the relationship between features captured by sociometric badges, employee self-perceptions (from surveys) and productivity [19].

Finally, unsupervised learning approaches have been used for human activity discovery in the past, both for individual activities and social activities. As an example, Vahdatpour et al. find recurrent patterns from multidimensional time-series given by multiple wearable sensors [23]. More recently, Do et al. proposed topic models for capturing group interaction patterns from Bluetooth proximity networks [8]. While these previous studies focus on activity discovery, this paper also considers the automatic labeling task for the discovered activities.

3. SOCIOMETRIC BADGES CORPUS

The aim of our study is to investigate behavioral patterns within organizational environments. The SocioMetric Badges Corpus [15], which is used in this work, is a multimodal corpus collected in a research institute over a six-week consecutive period, involving a population of 54 subjects (46 subjects that belong to four computer science research groups and 7 subjects of the IT department), during their working hours. Sociometric Badge sensors were employed for this study; these sensors are equipped with accelerometers, microphones, bluetooth, and infrared sensors which capture body movements, prosodic speech features, collocation, and face-to-face interactions, respectively [20]. For the purposes of our study we have exploited the data provided by the Infrared and Bluetooth sensors.

3.1 Data collection

Organizational Information. The subjects involved in the study belong to two distinct categories of employees; the administrative group and the research units. The population of the study consisted of 54 subjects (49 males, 5 females) with a mean of 36.8 years of age and standard deviation of 8.6 years. Due to technical issues, the Bluetooth data was missing for 5 subjects. This study is then conducted on the data of 49 subjects.

Bluetooth Data. Bluetooth detections can be used as a coarse indicator of proximity between devices. Radio signal strength indicator (RSSI) is a measure of the signal strength between transmitting and receiving devices. The range of RSSI values for the radio transceiver in the badge is (-128, 127). All sociometric badges broadcast their ID every five seconds using a 2.4 GHz transceiver (TR_{radio} = 12 transmissions per minute). Figure 1(a) shows the Bluetooth proximity network where nodes correspond to subjects (colored by office number) and the strengths of ties correspond to the number of Bluetooth hits. The sensed network is relatively dense and highly affected by the locations of subjects' office. This is not surprising given that the physical range of Bluetooth is around 10 meters (Class 2), meaning that a large proportion of BT hits do not correspond to an actual face-to-face interaction.

Infrared data. Infrared transmissions are used to detect face-to-face interactions between people. In order for a badge to be detected through an IR sensor, two individuals must have a direct line of sight and the receiving badge's IR must be within the transmitting badge's IR signal cone of height $h \leq 1$ meter and radius $r \leq htan\theta$, where $\theta = \pm 15^o$ degrees. The infrared transmission rate (TR_{ir}) was set to 1Hz. The accumulated IR network is visualized in Figure 1(c) in the same way as the one for the BT network. Since IR hits corresponds well to face-to-face interactions, the IR network is less influenced by desk locations.

Localizing interaction data. Seventeen BT devices were additionally used as fixed stations at key locations in order to infer the location of subjects and their interactions. These points were the restaurant, the cafeteria, and the coffee machines, as well as the meeting and seminar rooms at the organization. The BT devices used for localization have been grouped in four broader categories called *meeting rooms, admin meeting room, restaurant,* and *cafeteria*. For each BT or IR hit with the sociometric badges (called interaction hit), we find the nearest BT hit between the observer and one of the fixed stations (called localization hit), and then add the found location to the interaction hit if the time difference between two hits is less than one minute. At the end, the locations of BT or IR hits among subjects belong to either one of the four categories above or a special category called *others*, in case there is no localization hit to any fixed station during the interaction.

3.2 BT or IR for interaction sensing?

While IR hits imply actual interaction between two people, the strict detection conditions (a direct line of sight and limited angles) mean that the device may fail to capture actual interaction in several situations such as group meeting (e.g., people sit around a big table) or when two interlocutors look at the same object (e.g., screen, board). In practice, we found that IR data is very sparse and its density is about only 2% of the BT data set.

As an alternative to IR, BT proximity can be used as a reliable method to sense face-to-face interaction with low false negative rate. When using Bluetooth proximity data, the challenge is how to reduce its high false positive detection rate, which comes from its relatively long range compared to the face-to-face interaction.

With these points in mind, we chose to combine both IR and BT data, for which we only keep BT hits with strong signal strength (high RSSI value). In our experiment, a RSSI value greater or equal than -80 is considered as strong, which is generally produced when the distance of the two devices is less than 5 meters (which is sufficient to detect interactions among people in a meeting room of medium size). Figure 2 shows the distributions of RSSI values of BT hits when there is an IR hit (that is, when there is a face-to-face interaction) and when there is not (i.e., we are not sure if there is a real interaction). For the given threshold, the plot shows that 88% of IR hits can be captured by strong BT hits, and 24% of BT hits were classified as strong. The accumulated network constructed from strong BT hits is showed in Figure 1(b).

4. SOCIAL ACTIVITY DISCOVERY WITH TOPIC MODEL

Topic models are widely used for text analysis to find topics from text corpora and to summarize documents based on the set of topics. For example, latent Dirichlet allocation (LDA) takes as input

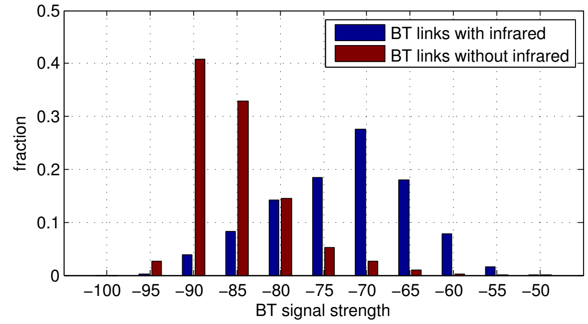


Figure 2: BT signal strength distributions change significantly depending on whether the face-to-face interaction is captured by IR or not.

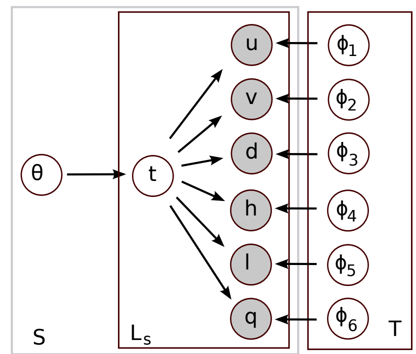


Figure 3: Graphical model of the extended GroupUs for handling spatio-temporal context and other additional attributes of interaction links.

a corpus of documents, represented as bags of words, and outputs the set of discovered topics, characterized by their most frequent words [4]. In LDA, each word is assumed to be generated from a latent topic, and each document can be summarized by a multinomial distribution over topics.

The LDA framework can be applied to human interaction mining where a *word* is an interaction between two individuals and a *document* consists of the interactions captured within a time interval, as in the GroupUs algorithm [8]. While this algorithm works on BT links with temporal context, we extend it in this paper to handle spatio-temporal links sensed with BT and IR.

Data representation. The recording period is divided into time slice of W minutes, where W should be large enough so that face-to-face interactions can be captured reliably by at least BT or IR. Each link i contains the following attributes:
- u_i: the head of the link for the observer.
- v_i: the head of the link for the observed person.
- d_i: day-of-week when the link was observed.
- h_i: time-of-day when the link was observed.
- l_i: the location where the link was observed.
- q_i: the type of the link (BT or IR).
- s_i: the identifier of the time slice that the link belongs to. $s_i \in \{1..S\}$ where S is the total number of time slices.

In the data, we had $d_i \in \{Mon, ..., Fri\}$ and $h_i \in \{8am..6pm\}$ according to the data collection setting. The location variable can be one of the five location categories described in Section 3.

Probabilistic model. Our extended GroupUs model is illustrated in Figure 3 where shaded nodes correspond to observed variables, latent variables t correspond to hidden activity topics of links, and θ and ϕ are model parameters. We use a plate representation where each node corresponds to a set of random variables, whose size is given by the capital letter in the corner. S stands for the number of slices, L_s stands for the number of links in slice s, and T is the number of activity topics. Compared to GroupUs, the proposed model has similar structure but contains more observed variables (location l and data type q). Its generative process is then similar to the original algorithm:

Initialization:

Draw distribution $\theta_s \sim Dirichlet(\boldsymbol{\alpha})$ for each slice s.
Draw distribution $\phi_t \sim Dirichlet(\boldsymbol{\beta})$ for each activity topic t.

For each link of the slice s**:**

Draw an activity topic $t|s \sim Multinomial(\theta_s)$.
Draw a first person $u|t \sim Multinomial(\phi_{1t})$.
Draw a second person $v|t \sim Multinomial(\phi_{2t})$.
Draw a day of week $d|t \sim Multinomial(\phi_{3t})$.
Draw a time of day $h|t \sim Multinomial(\phi_{4t})$.
Draw a location $l|t \sim Multinomial(\phi_{5t})$.
Draw a data type $q|t \sim Binomial(\phi_{6t})$.

The model parameter ϕ is key for the interpretation of topics as it encodes the conditional distributions of observations given activity topics. ϕ_{1t} and ϕ_{2t} characterize who were active members of the group activity corresponding to activity topic t. ϕ_{3t} and ϕ_{4t} reveal when the activity happened, while ϕ_{5t} indicate where the activity was usually observed. Finally, ϕ_{6t} indicates which type of link is usually observed in the activity topic t. The joint probability of observed and unobserved variables can be written by:

$$\begin{aligned}&P(\mathbf{u},\mathbf{v},\mathbf{d},\mathbf{h},\mathbf{l},\mathbf{q},\mathbf{s},\mathbf{t};\boldsymbol{\alpha},\boldsymbol{\beta})\\ &= \int_{\theta,\phi} P(\mathbf{u},\mathbf{v},\mathbf{d},\mathbf{h},\mathbf{l},\mathbf{q},\mathbf{s},\mathbf{t},\theta,\phi;\boldsymbol{\alpha},\boldsymbol{\beta})\partial\theta\partial\phi\\ &= \int_{\theta} P(\mathbf{t}|\theta)P(\theta;\boldsymbol{\alpha})\partial\theta \int_{\phi} P(\mathbf{u},\mathbf{v},\mathbf{d},\mathbf{h},\mathbf{l},\mathbf{q}|\mathbf{t},\phi)P(\phi;\boldsymbol{\beta})\partial\phi\end{aligned} \tag{1}$$

where the integration over model parameters θ and ϕ can be computed efficiently since we use conjugate priors for all link attributes. The model can be learned by collapsed Gibbs sampling which samples the posterior distribution $P(\mathbf{t}|\mathbf{u},\mathbf{v},\mathbf{d},\mathbf{h},\mathbf{l},\mathbf{q}\ \boldsymbol{\alpha},\boldsymbol{\beta})$ and provides estimates of θ and ϕ.

The output of the extended GroupUs algorithm consists of the learned parameters (θ and ϕ) and the vector of topic assignments $\mathbf{t}$. As discussed earlier, each activity topic t can be characterized by the corresponding distributions over people (parameterized by ϕ_{1t} and ϕ_{2t}), time (ϕ_{3t} and ϕ_{4t}), locations (ϕ_{5t}), and link types (ϕ_{6t}). However, the topic assignment $\mathbf{t}$ is also helpful for interpreting the discovered activity topics. Basically, this vector $\mathbf{t}$ assigns each link in the dataset to a discovered topic, so that we know exactly the set of links for a given topic t. Based on this, we can extract additional information about the activity topic, such as visualizing the “activeness” of the topic over time, find who were involved at a specific time, or compute the duration of the activity at a specific time.

Post-processing of discovered activity topics. One limitation of GroupUs and topic models in general lies in the number of activity topics T that need to be specified. A too small value of T results in topics which corresponds to multiple activities, and a large value of T wil produce some topics with similar patterns.

We remark that while a person might have multiple activities with another person, activities usually differ in sets of involved people and locations. We thus propose a post-processing process which merges similar activity topics based on who is involved in the activity and where the observed interactions took place. Interestingly, the GroupUs algorithm provides an efficient way to extract the set of active members from the distribution over people, ϕ_{1t} and ϕ_{2t} [8]. For the location, we can simply use the most likely location ϕ_{5t} to define the dominant location of the activity topic.

After the merging step based on active members and dominant location, the model structure is unchanged but the number of topics may be reduced. The sampled topic assignment $\mathbf{t}$ can be transformed to the merged model based on the mapping from the set of original topics to the set of merged topics. Then, the model parameter θ and ϕ of the merged model can be re-estimated based on the topic assignment according to the merged model.

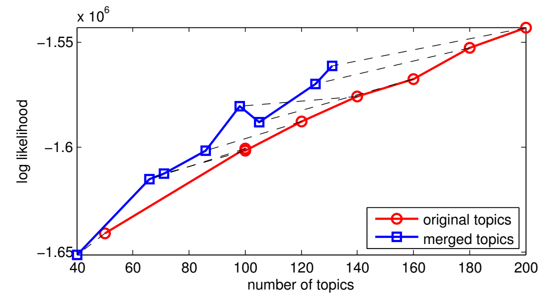


Figure 4: Log likelihood as a function of number of activity topics. Dashed line segments connect results before and after merging. The merging step reduce significantly the number of topics while keeping likelihood value reasonable compared to the original model.

5. RESULTS

We set the slice duration W to 5 minutes, resulting in 3670 non-empty slices for the 6 weeks of data. Overall, there are 160,000 dyadic links, where the distribution links over the 5 location categories *meeting rooms, admin meeting room, restaurant,cafeteria,* and *others* are 14.7%, 10.8%, 7.6%, 8.6%, and 58.2% respectively. We tried multiple values for the number of topic T, varying from 50 to 200, to study the behavior of the model with respect to this hyper parameter. For each value of T, the algorithm outputs a set of original activity topics and a corresponding set of merged ones.

Figure 4 shows the log likelihood of the data as a function of the number of topics T. The dashed line segments connect results before and after merging topics. As can be seen, the merging step reduces significantly the number of activity topics while keeping the likelihood at a reasonable level. Since the merging step is quite efficient, we continue the study with merged activity topics instead of the original ones. The main advantage is to reduce the number of activity topics to be annotated.

The remaining part of this section is dedicated to the interpretation of the discovered topics. In Subsection 5.1, we provide an subjective validation of the set of discovered topics and highlight our findings. In Subsection 5.2, we go further by studying the possibility of automatic interpretation of the set of discovered topics into predefined categories, thus providing a fully automatic framework for human interaction sensing.

5.1 Manual interpretation of topics

The goals of manual interpretation are to validate if the discovered topics are meaningful and to understand which kind of activity can be captured by the extended GroupUs algorithm. We got the annotation done for 6 sets of merged topics, which were outputted by the extended GroupUs algorithm with various settings of T. The 6 sets consists of 40, 66, 86, 98, 131, and 71 merged topics and they were numbered from 3 to 8 respectively (see Table 1).

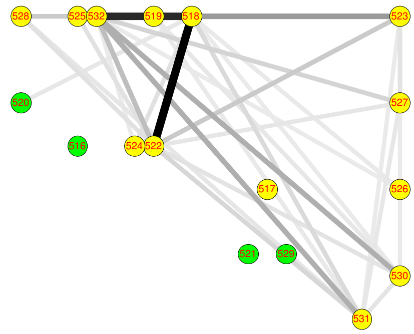
Annotation process: Three people from the organization where the dataset was collected accepted to participate in the annotation experiment. Therefore, these annotators in principle personally know the subjects, their roles, the relationships between them, and also the main activities in the organization.

The visualization of each activity topic is built from the outputs of the extended GroupUs algorithm as in Figure 5. Each topic can be described directly with the learned parameter ϕ, which contains the distribution of links over people, location, and time. We do not include the distribution over link types (BT or IR) as these distribution usually is dominated by BT. Besides the plots coming from ϕ, we also use the assignment of each interaction link to an activity topic (the vector $\mathbf{t}$) for more complex visualization such as computing the number of interactions between two given participants for a given topic. At the end, there are six different plots for each activity topic t which help the annotator to recognize the activity:
⋄ *Interaction network:* The network of people for the considered activity topic; active members are highlighted in a different color (yellow). Nodes' positions correspond to the actual desk locations of subjects to simplify the identification of subject.
⋄ *Time:* The distribution of links over the weekly calendar. While this can be given by the learned parameter ϕ in an factorized form ($\phi_{3t}\ \phi_{4t}{}^{\top}$), we decide to estimate the weekly calendar distribution based on $\mathbf{t}$ to have a more accurate estimation.
⋄ *Location:* The distribution of links over locations where they were observed is given by ϕ_{5t}.
⋄ *Interaction over time:* The distribution of links over 5-min time slices. This plot is helpful to know the number of occurrences of the activity and the number of links when the activity happened.
⋄ *Duration:* The distribution of time slices where the activity topic t happened over the duration of the activity.
⋄ *Number of involved people:* Distribution of time slices where the activity topic t happened over the number of people who were involved (having a link assigned to activity topic t in the time slice).

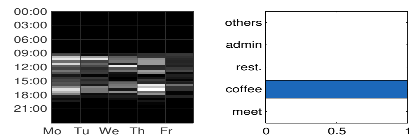
For each discovered activity topic, the annotator gave a relevance score from 1 (Not relevant) to 4 (Very relevant). The relevance score is a subjective assessment of the annotator about the discovered activity topic, for which we do not have a specific definition. We gave some hints to the annotators by suggesting several dimensions to look at: *is it meaningful? is it easy to interpret? does it correspond to an actual activity? is it a specific activity or does it correspond to multiples activities?*. If a topic is meaningful, easy to interpret, and corresponds to an actual social activity, then ideally it should get the highest score of 4. If some dimensions are not satisfied, then the relevance score is lowered. Note that low relevance scores can come both from the sensing framework (e.g., proximity links do not correspond to interaction) and the discovery method.

Each annotator was also asked to assign each presented activity topic to one or multiple predefined semantic labels. A collaborative annotation session was done online by annotators and authors to decide on the labels for social activity. From this pilot analysis, we agreed that the set of discovered activity topics could be explained by seven labels:
⋄ *dailyNearbyOffice:* false positive face-to-face interactions from people who sit very close to each other;
⋄ *randomChatWork:* work related discussion between people who work together;
⋄ *randomChatNonWork:* non-work interaction (but not a coffee break) between friends such as chatting, walking together, smoking together;
⋄ *coffeeBreak:* coffee breaks near a coffee machine or the cafeteria;
⋄ *restaurant:* having lunch in the restaurant area;
⋄ *irregularMeeting:* meeting of a group of people, which does not always happens at the same time and the same day of the week.
⋄ *weeklyMeeting:* happening every week at the same time and day.

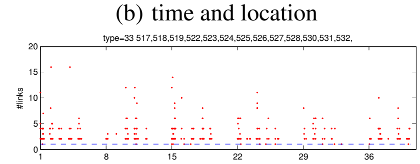
Note that if the discovered activity topic seems to include multiple activities (for example, *randomChatNonWork* and *CoffeeBreak* together), then the annotator can assign multiple labels. All annotations were done on a shared Google spreadsheet, each set corresponding to one separate sheet. Each annotator was the responsible person of one or multiple sets and needed to finish the assigned sets. Furthermore, other annotators could look at his/her annotations and comment. The objective was to find an agreement between annotators rather than averaging annotation from multiple annotators.



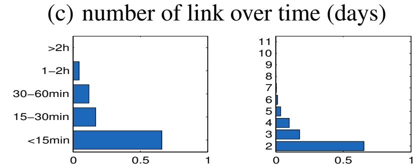
(a) Network visualization



(b) time and location



(c) number of link over time (days)



(d) duration and number of people

Figure 5: Example of topic visualization for the annotation process. Various information such as time, location, and duration are shown for activity topic 33 in set 3 of the experiment.

Discovered activity topics. Examples of discovered topics are given in Table 1 for which we report the experience of the annotator when she/he looked at the visualization of the activity topic. The "Code" field corresponds to the set and the ID number of the discovered activity topic (e.g., set4-33 means topic 33 in set4). The "Observation" field is our text description for the activity topic in terms of who, when, and where. These information can be easily extracted from the activity topic visualization (Fig. 5). The "Annotation" field corresponds to the annotated label(s) and relevance score. Finally, we report the annotator comments about the discovered activity topic in italic.

We selected a number of activity topics to demonstrate what kind of activities were discovered. Topic 33 in set4 corresponds to coffee break interactions of 13 people in several offices, its visualization is shown in Figure 5. From the plots, we can extract that this particular activity topic involved 13 people (yellow nodes in the network plot), could happen any time, and took place at the coffee machine. The timeline slot in Figure 5(c) show that this activity happened every working day of the week, and the distributions in Figure 5(d) shows that the events of this activity are usually short (lest than 15minutes) and that people usually come in group of 2 or 3 people.

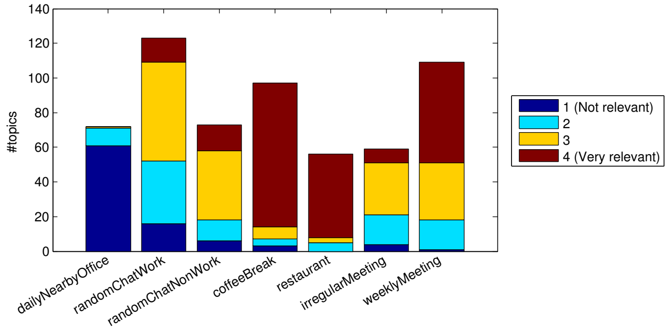


Figure 6: Relevance scores of activity topics in different categories of activity. The results were aggregated from all the sets.

These findings are summarized in Table 1 and they correspond to the comment from the annotator.

Due to space constraints, we put the visualization of other activity topics in a separate supplementary document, and only report the summarization of these activity topics in Table 1. Topic 30 of set6 is an example of "unreal" interaction between four people coming from a small office. The annotator can recognize this activity topic as a *dailyNearbyOffice* since the interaction links occur almost everyday and lasts several hours. Topic 57 in set6 is an example of interaction between friends and between colleagues. The annotator thinks that it is a mix of *randomChatWork* and *randomChatNonWork* since these kind of social activity usually involve only 2 or 3 people; and the durations of the interaction are usually short (quick chat) but can also be relatively long (probably work related discussion). Topic 43 of set6 involve 7 people who works together in several projects. Based on the duration and relationship between these people, the annotator labeled it as *randomChatWork*.

In some cases, the discovered activity topic does not correspond to a single activity label. We found that *randomChatNonWork* is usually confused with *randomChatWork* since their patterns about number of people, time, and location are very similar. The main difference between these two labels lies in the duration of interaction of event and the relationship between people. This information, however, is not exploited by the model so that the discovered activity topics are not discriminative between the two activity labels. Topic 5 of set6 is a typical example of the confusion between *randomChatWork* and *randomChatNonWork*. These are interactions between two people from the same room, which are relatively short but it is unclear if the discussion is work-related. Similarly, topic 57 of set6 is another example of confusion, but involve more people, and some of them are friends.

The three remaining examples are dedicated to *restaurant* (set3-29), *irregularMeeting* (set6-12), and *weeklyMeeting* (set4-18). While the patterns of these examples of *restaurant* and *weeklyMeeting* are clean (i.e., with the highest relevance score), the chosen activity topics of *irregularMeeting* was not easy to interpret for the annotators. More specifically, these topics correspond to meetings of one hour, but the number and composition of attendees varied significantly from one meeting to another, making it hard to recognize in which context these meetings were held (for which projects, held by which group of people, etc.)

Subjective evaluation. Overall, 16% of the topics get relevance score 1 (Not relevant), 16% of them get score 2 (OK - but not so good), 25% get score 3 (Good), and the remaining 44% get score 4 (Very relevant). After a close inspection, we found clear patterns on the dependencies between relevance score and activity label. On

Table 1: Examples of discovered activity topics, their annotations, and comments from annotators.

<table>
<tr><th>Code</th><th>Observation</th><th>Annotation</th></tr>
<tr><td>set4-33 (Fig. 5)</td><td>Who: 13 people
When: can be any time, most popular around 10am and 4pm
Where: coffee machine</td><td>coffeeBreak
relevance: 4</td></tr>
<tr><td colspan="3">"This activity topic corresponds to interactions during coffee breaks; people usually came in group of 2 or 3; lasted around 15 minutes; this activity topic is relevant and it reflects the actual activity."</td></tr>
<tr><td>set6-30</td><td>Who: 4 people
When: can be any time of the day
Where: not available (no connection to any fixed station)</td><td>dailyNearbyOffice
relevance: 1</td></tr>
<tr><td colspan="3">"Happened almost everyday from 10am to 7pm with a break at lunch; this activity topic does not seem to correspond to real face-to-face discussion since these 4 people came from a small office and the activity happened almost everyday and lasts several hours. Not a relevant activity."</td></tr>
<tr><td>set6-43</td><td>Who: 7 people
When: any time, usually after lunch
Where: not available</td><td>randomChatWork
relevance: 4</td></tr>
<tr><td colspan="3">"Two or three people discuss together; usually short but can be long up to 2 hours; these people work together thus the topic is very relevant; these seven people come from different offices."</td></tr>
<tr><td>set6-5</td><td>Who: 2 people
When: can be any time of the day
Where: not available</td><td>randomChatWork,
randomChatNonWork
relevance: 3</td></tr>
<tr><td colspan="3">"Interaction between two colleagues from the same office; usually last less than 15 minutes but it is not clear if the discussion is work-related since these two people also work together."</td></tr>
<tr><td>set6-57</td><td>Who: 7 people
When: can be any time of the day
Where: not available</td><td>randomChatWork,
randomChatNonWork
relevance: 3</td></tr>
<tr><td colspan="3">"Irregular interactions involving 2 or 3 people; interaction durations are usually less than 15 minutes but can be up to 2 hours; can be a discussion about work or non-work chatting since some people are friends; The seven people come from different offices."</td></tr>
<tr><td>set3-29</td><td>Who: 4 people
When: between 12:00 and 14:00
Where: restaurant area</td><td>restaurant
relevance: 4</td></tr>
<tr><td colspan="3">"These 4 people (sitting in the same office) usually come together to have lunch, lasts usually one hour."</td></tr>
<tr><td>set6-12</td><td>Who: 8 people
When: some Mondays and Tuesdays, 5pm-6pm
Where: usually at meeting rooms</td><td>irregularMeeting
relevance: 2</td></tr>
<tr><td colspan="3">"It is clear that this activity topic corresponds to one-hour meetings; the 8 people actually work together; the number of attendees vary from 5 to 10, make it hard to interpret the activity topic."</td></tr>
<tr><td>set4-18</td><td>Who: 11 people
When: every Monday at 12pm-1pm
Where: not available</td><td>weeklyMeeting
relevance: 4</td></tr>
<tr><td colspan="3">"Group meeting where members are supposed to participate every week; lasting one hour; happened every Monday in the 6 week period."</td></tr>
</table>

Table 2: Feature extraction for automatic labeling of discovered activity topics.

Group	Description
Location	Distribution of links over locations.
Time	Distributions over time-of-day and day-of-week.
People	Number of active members and distribution of time slices over the number of involved people.
Duration	Distribution of time slices over durations.
Link	Distribution of number of links per time slice.

one hand, most topics with low relevance score were annotated as *dailyNearbyOffice*, corresponding to groups of people whose desks are very close. While the captured interaction links did not correspond to real interaction due to sensing (proximity does not always mean interaction), the model was able to regroup them in several activity topics. We see in the next section that there is the possibility to recognize these "unreal" interactions automatically. On the other hand, coffee breaks, lunch in the restaurant, and weekly meetings were well discovered by the model. The relevance of the remaining activity topics were dominated by score 3 and 2, corresponding to Good and OK, respectively. As stated earlier, low relevance scores may come from sensing quality and/or the discovery method. Our analysis confirmed that low relevance scores come mainly from the sensing part (false positive interaction links) while the activity discovery algorithm performs relatively well.

5.2 Automatic topic labeling

In the previous section, we showed that many of the discovered activity topics correspond to actual group activities, and they can be interpreted by people who know the organization well. As the next step, it would be practical to have a system that discovers multiple activity topics and recognizes automatically their semantic meaning such as weekly meetings or coffee breaks. We formalize the recognition problem in a supervised learning framework where data points correspond to discovered activity topics and the ground-truth is a whether an activity topic correspond to a given label.

Feature extraction. For the automatic labeling task, each discovered activity topic t is represented by a vector of 49 features belonging to six groups as shown in Table 2. The set of location features consists of the distribution over the five categories of location, which is encoded by ϕ_{5t}. The entropy of these multinomial distributions is also used as a feature. Time features come from the distribution over day of the week, ϕ_{3t}, and time of the day, ϕ_{ht}, as well as their entropy. People features are computed based on the detected number of people involved in the activity topic. We used the number of active members (given by the extended GroupUs algorithm) and the distribution over number of attendees (i.e., fraction of active time slices in which there are exactly X people being involved in topic t, with X varying from 2 to 49), as showed in Figure 5(d,right). Similarly, duration features are computed as the fraction of active time slices of topic t which are parts of an activity lasting Y minutes (Figure 5(d,left)). Finally, for link features, we compute the fraction of time slices of topic t which has Z interaction links being assigned to the topic.

Automatic labeling performance. The labeling task is to predict if a discovered topic corresponds to an activity label. Thus, for the seven activity labels, we have seven binary classification tasks. We use a random forest classifier with 1000 trees, and all results are obtained with a 10-fold cross-validation process on 492 annotated activity topics. To see how important each set of features is, we ran the random forest for each feature group separately, then, starting with the best feature group in terms of average accuracy, we progressively added the remaining feature groups to the random forest system and evaluate the labeling performance.

Table 5.2 reports automatic labeling accuracy for the seven activity labels. We used a majority class predictor as a baseline and compare its performance with nine random forest systems with various feature configurations. The *average* row, corresponding to the average performance over seven activity labels, shows that Duration (**D**), Time (**T**), and Location (**Lo**) features were relatively predictive of activity labels, while Link (**Li**)and People (**P**)features were less so. While location and social activity are closely related in practice, the location was not the most important information since, in our setting, the location data were coarse (five categories), noisy, and not always available. Note, however, that the location feature group was the most discriminant for *coffeeBreak* and *restaurant*, which were reliably recognized by random forest systems. In contrast, the random forest system with all features did not outperform the baseline performance for *randomChatNonWork*. This issue may come from the challenge to distinguish between *randomChatNonWork* and *randomChatWork* without knowing the actual relationship between people. This assumption is confirmed by the classification results of the special activity label *randomChat* which merges *randomChatWork* and *randomChatNonWork*. The reduction in the number of errors (between Baseline and RF with all features) for *randomChatNonWork*, *randomChatWork*, and *randomChat* were respectively 1% (72 vs 73) , 33% (82 vs 123), and 46% (80 vs 149).

Comparing the two random forest systems with and without location features (**D+T** and **D+T+Lo** respectively) we found that their performance is comparable for many activity labels except *randomChat* and *coffeeBreak*. To verify if the location information is key to distinguish these two categories, we apply the merging technique again to have a new label called *randomChatCoffee*. The comparable performance of **D+T** and **D+T+Lo** systems for *randomChatCoffee* suggests that location information is only helpful to distinguish between *randomChat* and *coffeeBreak*, while the classification performance for other activity categories does not depend significantly on the location information.

6. CONCLUSION

This study opens the possibility to infer social activity categories from interaction between people. We have proposed a variant of topic model which can discover various activity topics from timestamped, localized social interactions being sensed by BT and IR sensors. We also showed how to assign these discovered topics to common sense activity labels, thus enabling an automatic framework for sensing and inferring activity from mobile sensors. The analysis was conducted on a real interaction data set in an organization, and the method was validated by a few of their members. The subjective evaluation results suggest that many discovered activity topic are highly relevant, while false positive interactions were not filtered completely from the data set. Interestingly, these sensing "noise" were grouped in several activity topics, thus they may be filtered by the probabilistic method. Finally, our experiment on automatic labeling demonstrates that some activities can be recognized reliably, even when the location of the interaction is not available. Furthermore, the results show that a multi-feature approach for classification outperformed the single cue approach for the activity labels.

As future work, we are interested in applying the framework to characterize the relationship between social activity and personal states such as mood. Our framework can be extended to incorporate more features such as speech or body activity. Another direc-

Table 3: Accuracy (%) of random forest with different groups of features. Baseline (B) was a majority class predictor. Feature groups are represented by their initials. Best feature group and best feature combinations are highlighted in bold for each row.

		RF with single feature group					RF with multiple feature groups			
Category	B	**D**	**T**	**Lo**	**Li**	**P**	**D+T**	**D+T+Lo**	**D+T+Lo+Li**	all features
dailyNearbyOffice	85.4	**90.7**	89.4	89.0	85.2	84.1	**89.6**	89.0	**89.6**	89.0
randomChatWork	75.0	**79.7**	79.1	71.5	72.2	74.8	80.9	**83.3**	82.3	**83.3**
randomChatNonWork	85.2	**85.0**	84.8	79.7	83.7	84.6	84.8	**85.6**	85.2	85.4
coffeeBreak	80.3	92.7	88.6	**97.8**	82.7	78.0	94.1	98.6	98.6	**99.0**
restaurant	88.6	95.9	97.4	**99.8**	89.4	88.6	98.2	**98.8**	**98.8**	**98.8**
irregularMeeting	88.0	87.4	**87.8**	86.6	87.8	86.8	87.2	89.0	**89.2**	**89.2**
weeklyMeeting	77.8	86.2	**86.4**	70.7	81.3	75.0	88.4	**88.0**	**88.0**	87.8
average	82.9	**88.2**	87.6	85.0	83.2	81.7	89.0	90.3	90.2	**90.4**
randomChat	69.7	**78.9**	77.2	68.3	66.7	66.3	79.7	83.3	82.7	**83.7**
randomChatCoffee	52.4	**85.8**	82.5	69.3	69.3	63.2	85.0	**85.6**	85.2	85.4

tion that we want to investigate is to compare our global network analysis with ego network analysis to understand the advantages and shortcomings of each approach.

Acknowledgments

This research was funded by the SNSF SONVB project and the Marie Curie-COFUND-7th Framework PERSI project. We thanks Ailbhe Finnerty for help with the annotation, and Sociometric Solutions and Prof. Alex Pentland (MIT) for providing sociometric badges.

7. REFERENCES

[1] R. Albert, H. Jeong, and A.-L. Barabási. Internet: Diameter of the world-wide web. *Nature*, 401(6749):130–131, 1999.

[2] L. Backstrom, E. Sun, and C. Marlow. Find me if you can: improving geographical prediction with social and spatial proximity. In *Proc. WWW*, pages 61–70, 2010.

[3] L. Bao and S. S. Intille. Activity recognition from user-annotated acceleration data. In *Pervasive Computing*, pages 1–17. Springer, 2004.

[4] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022, 2003.

[5] 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*, 5(7):e11596, 2010.

[6] X. Chen and J. Yang. Towards monitoring human activities using an omnidirectional camera. In *Proc. ICMI*, pages 423–428, 2002.

[7] M. Cristani, R. Raghavendra, A. Del Bue, and V. Murino. Human behavior analysis in video surveillance: a social signal processing perspective. *Neurocomputing*, 2012.

[8] T. M. T. Do and D. Gatica-Perez. Human interaction discovery in smartphone proximity networks. *Personal and Ubiquitous Computing*, 17(3):413–431, 2013.

[9] W. Dong, B. Lepri, and A. S. Pentland. Modeling the co-evolution of behaviors and social relationships using mobile phone data. In *Proc. MUM*, pages 134–143, 2011.

[10] N. Eagle and A. Pentland. Reality mining: sensing complex social systems. *Personal and ubiquitous computing*, 10(4):255–268, 2006.

[11] S. A. Golder and M. W. Macy. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science*, 333(6051):1878–1881, 2011.

[12] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, 2008.

[13] C. A. Hidalgo and C. Rodriguez-Sickert. The dynamics of a mobile phone network. *Physica A: Statistical Mechanics and its Applications*, 387(12):3017–3024, 2008.

[14] B. Lavelle, D. Byrne, G. J. Jones, and A. F. Smeaton. Bluetooth friendly names: bringing classic hci questions into the mobile space. In *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI... but not as we know it-Volume 2*, pages 63–66. British Computer Society, 2007.

[15] B. Lepri, J. Staiano, G. Rigato, K. Kalimeri, A. Finnerty, F. Pianesi, N. Sebe, and A. Pentland. The sociometric badges corpus: A multilevel behavioral dataset for social behavior in complex organizations. In *SocialCom/PASSAT*, pages 623–628, 2012.

[16] C.-H. Lu and L.-C. Fu. Robust location-aware activity recognition using wireless sensor network in an attentive home. *Automation Science and Engineering, IEEE Transactions on*, 6(4):598–609, 2009.

[17] A. Madan, M. Cebrian, D. Lazer, and A. Pentland. Social sensing for epidemiological behavior change. In *Proc. UbiComp*, pages 291–300, 2010.

[18] A. Noulas, S. Scellato, C. Mascolo, and M. Pontil. An empirical study of geographic user activity patterns in foursquare. *ICWSMâĂŽ11*, 2011.

[19] D. O. Olguın, P. A. Gloor, and A. S. Pentland. Capturing individual and group behavior with wearable sensors. *Proc. AAAI Spring Symposium*, 9, 2009.

[20] D. Olguin-Olguin, B. N. Waber, T. Kim, A. Mohan, K. Ara, and A. Pentland. Sensible organizations: technology and methodology for automatically measuring organizational behavior. *Trans. Sys. Man Cyber. Part B*, 39(1):43–55, 2009.

[21] S. Phithakkitnukoon, T. Horanont, G. Di Lorenzo, R. Shibasaki, and C. Ratti. Activity-aware map: Identifying human daily activity pattern using mobile phone data. In *Human Behavior Understanding*, pages 14–25. Springer, 2010.

[22] Y. Song, L.-P. Morency, and R. Davis. Multimodal human behavior analysis: learning correlation and interaction across modalities. In *Proc. ICMI*, pages 27–30, 2012.

[23] A. Vahdatpour, N. Amini, and M. Sarrafzadeh. Toward unsupervised activity discovery using multi-dimensional motif detection in time series. In *Proc. IJCAI*, pages 1261–1266. Morgan Kaufmann Publishers Inc., 2009.