

Contextual grouping: discovering real-life interaction types from longitudinal Bluetooth data

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

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

Abstract—By exploiting built-in sensors, mobile smartphone have become attractive options for large-scale sensing of human behavior as well as social interaction. In this paper, we present a new probabilistic model to analyze longitudinal dynamic social networks created by the physical proximity of people sensed continuously by the phone Bluetooth sensors. A new probabilistic model is proposed in order to jointly infer emergent grouping modes of the community together with their temporal context. We present experimental results on a Bluetooth proximity network sensed with mobile smart-phones over 9 months of continuous real-life, and show the effectiveness of our method.

I. INTRODUCTION

Social network analysis has gained much interest and became a key technique in modern sociology [28]. Social interaction data can take on several forms ranging from online collaboration to face-to-face interaction, and the types of analysis that can be performed depend much on the ability to gather accurate interaction data. This can vary from small amounts of data collected manually via questionnaires in the early days of social science [25] to complex networks of millions of nodes collected via digital systems such as phone communication and social media [17], [18].

Mobile phones -the most ubiquitous device in history- are rapidly emerging as sensing platforms of daily life activities and events, including social interaction. Built in phones and other mobile devices, Bluetooth (BT) is a reasonable approximation for sensing face-to-face social interaction [5]. Bluetooth devices can tell if two people carrying them are in proximity with high probability. While imperfect, Bluetooth-based proximity as an onboard capability on phones offers important technical advantages such as low battery consumption and the ability to work in both indoor and outdoor environments. Furthermore, people are often willing to share their Bluetooth information by setting their phones BT device to discoverable mode, and in practice many users do so on a regular basis.

A simple application of Bluetooth for social sensing is to consider the number of discovered nearby devices as an measurement of the density of the environment [22]. At a public place, one would observe many nearby devices from many unknown people. A different challenge is that of discovering the recurrent patterns of interaction with people in our social network (work colleagues, family members) and the context (temporal, spatial) in which they occur in real life. Many of

these interactions take place over multiple time scales and multiple groups: we might have breakfast and dinners with our family every day, meet our collaborators twice a week, our boss twice a month, and our sports teammates every sunday. The robust discovery of real-life social interaction types therefore calls, on one hand, for methods that are able to handle uncertainty on a principled way, and on the other, for longitudinal data of long enough support to learn all these types of effects.

In this work, we are interested in analyzing proximity interaction in order to discover a temporally grounded social context, such as being at a group meeting or being at home with family members. Our goal is to discover automatically different grouping modes of a real community, and the times when these modes occur. A key challenge lies in the stochasticity of the data which results in high uncertainty on how to group people based on their interactions. We address this problem with a probabilistic approach and propose a graphical model for analyzing pairwise interaction. Our model allows: (1) finding emergent grouping modes of a proximity-based real social network, (2) assigning grouping modes to the set of interactions at a given time, and (3) predicting pairwise interactions at a given time. This paper makes three contributions:

- We introduce a new model for group interaction discovery from proximity data, based on a probabilistic graphical model formulation, that can handle stochasticity.
- We conduct our analysis on a new large-scale interaction data set and report statistics of quality and robustness on Bluetooth proximity data in real-life conditions.
- We show that from Bluetooth proximity, our method can infer different grouping modes, and assign group memberships to the individuals who best conform them. We objectively validate our framework by studying the predictive performance of the method on holdout data, and evaluate how the model recover simulated events.

The paper is organized as follows. Section II reviews related works on Bluetooth data analysis and group discovery on social network. An overview of our framework is presented in III. Data representation and our method for group interaction discovery are presented in Section IV and V. Section VI presents our data collection framework and fundamental

statistics of the data. Experimental results are presented in Section VII. Finally, we draw conclusions in Section VIII.

II. RELATED WORK

The idea of using Bluetooth as proximity sensors has been well studied for some time [26], [22], [19]. One could consider the list of detected devices of a Bluetooth scan in order to infer nearby Bluetooth devices. Some works aim at using richer proximity information by exploiting low level signals. As an example, the system in [3] is designed for estimating distances between wireless devices with an average distance error of only 1.9 meter in their condition.

Bluetooth proximity has many applications in practice, such as an indicator of the density of the environment [22]. More elaborately, Terry et al. [26] look for proximity patterns over time in order to infer shared interests. The work in [6] uses Bluetooth proximity as a ground-truth for validating positioning system based on GSM network.

Most noticeably, Bluetooth proximity data can also be used for building a social network and for discovering social context. Clauset et al [4] present an analysis on Bluetooth proximity network recored over 9 months showing that some strong signals can be captured such as periodicity. In [5], Eagle et al. proposed to use Bluetooth and phone calls data to define pairwise links between people and in this way infer friendship networks, as an alternative to questionnaire-based, self-reported data. Related to our work, Mardenfeld et al. [16] recently proposed an algorithm for group discovery which is based on fully connected components of the Bluetooth proximity network. This method, however, has a number of drawbacks such as sensitivity to noise, it is likely to miss very large groups (such as a lecture at an auditorium), and has a complexity that grows exponentially in group size.

Our work can be positioned within the emerging body of work of reality mining [7], [24], [12] which aims at analyzing human behavior at large-scale using mobile phones as sensors of activity. Some works address face-to-face interaction discovery by using dedicated mobile devices (including microphones) other than phones [10], [29], [21]. While these dedicated devices provide a definite advantage over Bluetooth to sense the actual interaction in terms of spatial resolution, and use voice and infrared sensors, they need to be worn in specific conditions to work in practice. Furthermore, they represent an additional device that many people might not be willing to carry in real life.

In data mining and machine learning, there has been substantial research in detecting latent groupings based on network connectivity. People are typically grouped based on their interactions with others in a network [13], [20], [2]. Some methods have been proposed to discover groups of communities which are mainly based on discovering block structure from interaction. Stochastic block structure models [14], [1] aim at finding groups for each individual in a given network. Fu et al. [8] extended these models to dynamic networks by allowing model parameters (interaction probabilities) to change over the global state of the network. The state is

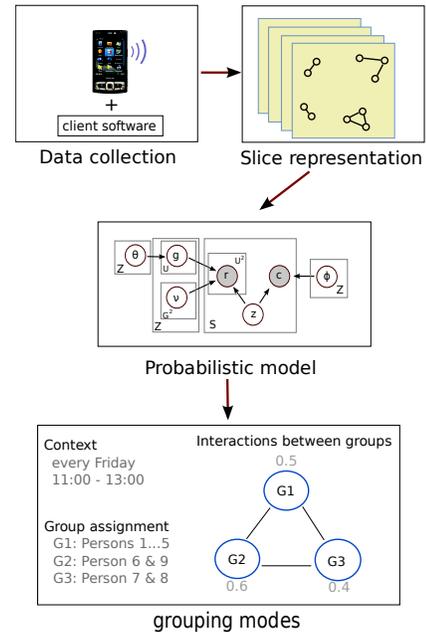


Fig. 1. Overview of our method.

determined by the time of the slice, and the temporal effect is modeled by introducing smoothing between parameters in two consecutive states. Our work is also based on the block structure of interaction, however, in order to deal with the dynamic of the network, we consider different interaction modes rather than one single mode that changes over time. Finally, our work is related to the group-topic model by Wang et al. [27], which had the same goal of group discovery but in a very different context (text analysis). Mathematically, the main difference between the two models is that our work represents an improvement on the modeling of interaction, allowing favor homogeneity between observations of a discovered mode of interaction.

III. OVERVIEW OF OUR WORK

Our approach to discover group interaction patterns is described in Figure 1. Its principal components are:

- **Data collection:** At the low level, a software application is installed into smartphones, which are then distributed to a set of volunteers who have a real social connection for gathering data. Longitudinal data needs to be collected over months of real life to capture actual events in the life of a community. Data from the set of users are uploaded daily to a central server for analysis.
- **Data representation:** The raw data is transformed into a time-slice based representation, consisting of an interaction graph between users and the temporal context of the slice.
- **Group discovery model:** We propose the use of a probabilistic graphical model to deal with stochastic interaction data. Latent variables are introduced in the model in order to capture emergent patterns from the interaction data.
- **Output:** Our method outputs different grouping modes

from interaction data. Each grouping mode is characterized by (1) the temporal context, (2) the group assignment for each of the users in the population, and (3) the probabilities of intra-group interaction and inter-group interaction, which resume the interaction probabilities between users in the population in a compact matrix.

The blocks of our approach are explained in the following three sections.

IV. DATA REPRESENTATION

We consider the Bluetooth raw interaction data recorded continuously in real-life over a population of users, with the task of discovering emergent group interactions within the population. We conduct our analysis with a slice-based approach, where all Bluetooth-detected interactions within a short period (e.g. 10 minutes) are grouped together, forming a slice of the dynamic network (see Figure 2). Each slice is characterized by a binary relational matrix indicating the existence of interaction between users. Furthermore, the time of the interaction is also key to understand the type of group interaction, hence, we include the temporal information in the description of the slice. Formally, a slice s is represented by:

- \mathbf{r}_s : the binary relation matrix. i.e., $r_{su} = 1$ if user u sees user v at any moment, regardless of the duration, during time-slice s , and $r_{su} = 0$ otherwise. Note that multiple times u sees v are not counted and also that \mathbf{r}_s can be a symmetric or asymmetric matrix.
- c_s : the temporal context of the slice, a discrete value that describes the corresponding time of the day and day of the week. It always corresponds to one entry of the $24 \text{ hours} \times 7 \text{ days}$ grid of a weekly calendar.

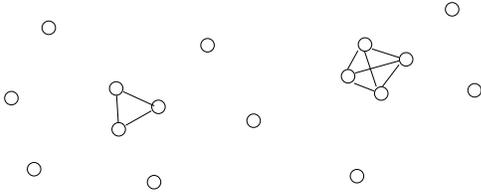


Fig. 2. A snapshot of proximity network for a given time slice.

V. GROUPING MODE DETECTION

A key challenge for mining from BT proximity lies in the fact the data is noisy, which may be due to technical limitations of the sensor as well as the presence of real “noise”. Considering a weekly face-to-face group meeting interaction as an example, even if all members attended the meeting, it could happen that some Bluetooth links between members could be lost due to sensor failures. Furthermore, one or more members of the group could be absent from the group meeting for a few times, or could have forgotten their phone at their home or office.

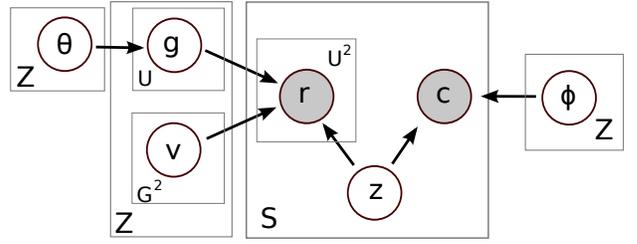


Fig. 3. Graphical model for grouping mode discovery.

TABLE I
NOTATIONS USED IN THIS PAPER.

| | |
|--|---|
| Observation and index: | |
| S | number of slices in training data. |
| U | number of people. |
| s | slice index. |
| u, v | persons. |
| r_{su} | relation between u and v in slice s . |
| c_s | temporal context of slice s . |
| Hyper-parameters and model parameters: | |
| Z | the number of grouping modes. |
| G | the (maximum) number of groups. |
| z_s | the grouping mode of slice s . |
| g_{zu} | the group assigned to user u in grouping mode z . |
| θ_{zg} | probability of observing the group g in mode z . |
| ϕ_{zc} | probability of observing the context c in mode z . |
| ν_{zgh} | interaction probability between group g and h in mode z . |
| α, β | parameters of the Dirichlet prior distribution for $\theta_z, \phi_z \forall z$. |
| γ_{gh} | parameters of the Beta prior distribution for $\nu_{zgh} \forall z$. |

A. Our probabilistic model

In order to deal with the stochasticity of the data, we propose a probabilistic approach where observations are represented by random variables. Assuming that there are various grouping modes in which a community’s members form groups, e.g. attending group meeting, having lunch together, etc., we introduce some latent variables in order to capture these emergent group interaction patterns from the observations. These latent variables are characterized by model parameters that can be learned by fitting the model to the data. To facilitate reading, the notations used in this paper are described in Table I. Only most basic terms are listed, and we denote structured terms of multiple elements by bold symbols. For instance, c_s is a scalar quantity, and $\mathbf{c} = (c_s)_{s=1..S}$ is a vector.

Our probabilistic graphical model is illustrated in Figure 3, where observed random variables r and c are represented by shaded nodes. Let S be the total number of slices, and z the latent variable that corresponds to the *grouping mode* index of the slice, and that can take values from 1 to Z , meaning that we want to capture Z grouping modes. Each grouping mode is characterized by three elements:

- Group assignment g : a variable that represents the mapping from the set of users to a small number of groups G .
- Bernoulli distributions (with parameter ν): the probability of having interaction between groups, both for intra-group interactions and inter-group interactions.

- Context distribution (with parameter ϕ): the probability of observing the grouping mode in specific temporal contexts.

In this model, G is a hyper-parameter defining the maximum number of groups in each grouping mode. In other words, there are G available slots for assigning people, and people in the same slot are considered to be a group. For a given grouping mode z , we assume that the pairwise interaction probability between people depends only on their groups, modeled by a Bernoulli distribution with parameter ν_{zgh} where g and h are the two involved groups. Finally, each grouping mode typically occurs in some specific temporal context, and the model also takes this into account during the learning step. The generative process for the set of interaction slices is defined as follows:

1) Initialization:

For each grouping mode z :

$$\theta_z \sim \text{Dirichlet}(\alpha), \phi_z \sim \text{Dirichlet}(\beta).$$

For each pair of groups (g, h) : $\nu_{zgh} \sim \text{Beta}(\gamma_{gh})$

For each grouping mode z :

$$\text{For each person } u: g_{zu} \sim \text{Multinomial}(\theta_z)$$

2) For each slice s :

$$z \sim \text{Uniform}(\frac{1}{Z}), c \sim \text{Multinomial}(\phi_z).$$

For each pair of people (u, v) , $u \neq v$

$$r_{suv} \sim \text{Bernoulli}(\nu_{zgh}) \text{ with } g = g_{zu} \text{ and } h = g_{zv}$$

In our model, we use Dirichlet prior distributions (with parameters α and β) for model parameters θ and ϕ , respectively, and Beta distributions (with parameters γ) for model parameters ν . The Dirichlet distribution is the conjugate prior of the Multinomial, and the Beta distribution is the conjugate prior of the Bernoulli, which are chosen for algebraic convenience. Importantly, the Beta prior distribution is a key for group discovery since it directly influences the probability of interaction between groups. For instance, we can tune these parameters γ so that people from the same group are expected to have high probability of interaction (Section VII gives more detail on this). Note also that the model always allows missing interaction since we do not have any hard constraint on fully connected groups, which is the main key of probabilistic approaches.

Let $(\mathbf{r}, \mathbf{c}) = (\mathbf{r}_s, c_s)_{s=1..S}$ be the set of observed relational matrices and contexts, $\mathbf{z} = (z_s)_{s=1..S}$ be the mode assignment for each slice, and $\mathbf{g} = (g_{zu})_{z=1..Z, u=1..U}$ be the group assignment for each user u in each grouping mode z . The joint probability of observations (\mathbf{r}, \mathbf{c}) , grouping mode assignments \mathbf{z} , and group assignments \mathbf{g} can be obtained by integrating over model parameters θ, ϕ, ν :

$$\begin{aligned} & P(\mathbf{c}, \mathbf{r}, \mathbf{z}, \mathbf{g}; \alpha, \beta, \gamma) \\ &= P(\mathbf{z}) \times P(\mathbf{c}|\mathbf{z}; \alpha) \times P(\mathbf{g}|\mathbf{z}; \beta) \times P(\mathbf{r}|\mathbf{z}, \mathbf{g}; \gamma) \\ &= P(\mathbf{z}) \int_{\phi} P(\mathbf{c}|\mathbf{z}, \phi) P(\phi|\alpha) d\phi \\ & \int_{\theta} P(\mathbf{g}|\mathbf{z}, \theta) P(\theta|\beta) d\theta \int_{\nu} P(\mathbf{r}|\mathbf{z}, \mathbf{g}; \nu) P(\nu|\gamma) d\nu \\ &= \frac{1}{Z^S} \prod_z \frac{B(\alpha + \mathbf{n}_z)}{B(\alpha)} \prod_z \frac{B(\beta + \mathbf{m}_z)}{B(\beta)} \prod_{z, g, h} \frac{B(\gamma_{gh} + \mathbf{q}_{zgh})}{B(\gamma_{gh})} \end{aligned} \quad (1)$$

where $B(\cdot)$ is the Beta function, \mathbf{n}_z is a context count vector of grouping mode z , \mathbf{m}_z is a G -dimensional group count vector of z , and \mathbf{q}_{zgh} is a two-dimensional vector counting how many times the pair of groups (g, h) have interactions and how many times they do not have interactions. Note that these counting vectors are actually functions over group and grouping mode assignments (\mathbf{g}, \mathbf{z}) , and we drop the function notation to simplify the presentation. Formally, given a current assignment \mathbf{z}, \mathbf{g} , the above quantities can be computed as:

$$\begin{aligned} n_{zc} &= \sum_{s: z_s=z} \mathbf{1}(c_s = c) \\ m_{zg} &= \sum_u \mathbf{1}(g_{zu} = g) \\ q_{zgh1} &= \sum_{s: z_s=z} \sum_{u: g_{zu}=g} \sum_{v \neq u: g_{zv}=h} \mathbf{1}(r_{suv} = 1) \\ q_{zgh2} &= \sum_{s: z_s=z} \sum_{u: g_{zu}=g} \sum_{v \neq u: g_{zv}=h} \mathbf{1}(r_{suv} = 0) \end{aligned} \quad (2)$$

where $\mathbf{1}(\cdot)$ is the indicator function.

Note that the integration over hidden parameters θ, ϕ, ν in Eq. (1) can be computed efficiently since we use conjugate priors in each elementary distributions. To simplify the discussion, we refer the derivation of Eq. (1) to appendix A.

B. Inference and parameter estimation

The proposed probabilistic model defines relations between observed variables (pairwise interaction and temporal context) and latent variables (grouping mode). These relations are parameterized by θ, ϕ, ν as discussed in the previous section. Discovering the group modes is the process of learning model parameters from observed data.

The problem of learning optimum model parameters in our model is intractable in general. However, a wide variety of approximation techniques can be used, including Laplace approximation, variational approximation, and Markov chain Monte Carlo (MCMC) [9]. In this work, we learn the model using collapsed Gibbs sampling (inspired by [11]), which samples the posterior distribution $P(\mathbf{z}, \mathbf{g}|\mathbf{c}, \mathbf{r}; \alpha, \beta, \gamma)$ via two conditional distributions:

- $P(z_s | \mathbf{g}, \mathbf{z}_{\neg s}, \mathbf{c}, \mathbf{r}; \alpha, \beta, \gamma)$: the conditional probabilities of the mode assignments for a slice s given all other assignments.
- $P(g_{zu} | \mathbf{g}_{\neg zu}, \mathbf{z}, \mathbf{c}, \mathbf{r}; \alpha, \beta, \gamma)$: the conditional probabilities of the group assignments for user u in mode z given all other assignments.

These probabilities can be computed based on the fact that they are proportional to the joint probability in Eq. (1). By eliminating invariant terms with respect to the variable to be sampled, we can show that:

$$\begin{aligned} P(z_s = z | \mathbf{z}_{\neg s}, \mathbf{g}, \cdot) &\propto \frac{\alpha + n_{zc}^{\neg s}}{\sum_c (\alpha + n_{zc}^{\neg s})} \prod_{g, h} \frac{B(\gamma + \mathbf{q}_{zgh})}{B(\gamma + \mathbf{q}_{zgh}^{\neg s})} \\ P(g_{zu} = g | \mathbf{g}_{\neg zu}, \mathbf{z}, \cdot) &\propto (\beta + m_{zg} - 1) \prod_h \frac{B(\gamma_{gh} + \mathbf{q}_{zgh})}{B(\gamma_{gh} + \mathbf{q}_{zgh}^{\neg u})} \end{aligned} \quad (3)$$

where $n_{zc}^{\neg s}$ and $\mathbf{q}_{zgh}^{\neg s}$ are the counts without considering slice s . Similarly, $\mathbf{q}_{zgh}^{\neg u}$ is the group interaction count without considering user u in mode z . The probability of assigning a mode z to slice s depends on its compatibility with the group assignment and the set of slices in that mode. The derivation of the above sampling equation is detailed in appendix B.

Given the group assignment for each person in each mode, and the mode assignments for all slices, we estimate the model parameters as follows:

$$\begin{aligned} \nu_{zgh} &= \frac{\gamma_{zgh1} + q_{zgh1}}{\gamma_{zgh1} + \gamma_{zgh2} + q_{zgh1} + q_{zgh2}} \\ \phi_{zc} &= \frac{\alpha + n_{zc}}{\sum_{c'} (\alpha + n_{zc'})} \\ \theta_{zg} &= \frac{\beta + m_{zg}}{\sum_h (\beta + m_{zh})} \end{aligned} \quad (4)$$

The learning algorithm is described in Algorithm 1. The algorithm starts with random group assignment \mathbf{g} for each grouping mode, and random mode assignments \mathbf{z} for the set of slices. Then, these assignments are resampled iteratively until convergence via collapsed Gibbs sampling. We maintain the counts $n_{zc}, m_{zg}, q_{zgh1}, q_{zgh2}$ over iterations, which are updated after each sampling step so that each iteration requires only a few computations. Note that the equations on lines 7 and 13 of Algorithm 1 are equivalent to the sampling equations in Eq. (3), since the counts were decreased just before the sampling step. After the sampling process, the algorithm outputs the group assignment for each mode, the mode assignment for each slice, as well as estimates of the parameters ν, ϕ, θ .

C. Relation to existing models

The model proposed in this paper is closely related to the group-topic model [27], which was applied for text-based analysis of political data. In our setting, relationships (given by phone sensor data) represent the main source of information, while temporal context provides additional information with respect to time and date when interactions occur. Mathematically, our model represents an improvement over the group-topic model on the way relationships are modeled, by sharing the hidden parameter ν across time slices. This allows to favor assignment of grouping modes with high homogeneity between slices that are assigned to the same mode.

VI. LONGITUDINAL PROXIMITY DATA

Mining real-life social interaction calls for longitudinal data of long enough support to discover recurrent patterns. We present in this section the data collection framework and an analysis on the data quality.

A. Data collection

We use a subset of the large-scale smartphone data set introduced in [15], which used Nokia N95 phones to collect data. A software client was installed in the phones in order to record Bluetooth data on a 24/7 basis, with the only restriction of having to recharge the phone once a day (typically done during nights). The mobile phone scans to detect nearby Bluetooth devices every 1-3 minutes, storing the logs in the phone's memory. The log data are then uploaded daily to a server, via a user-defined wifi connection.

The Bluetooth data we used corresponds to 9 months of real life of a set of 30 volunteer users who have a real social connection. 21 of the users are colleagues working at a mid-size organization, and the remaining 9 users are

family members from these people. Users carried their device as their actual (and only) phone and therefore used them in real conditions. All data has been anonymized. The data was recorded from October 2009 to July 2010 and corresponds to about 1.7 million non-empty Bluetooth scans.

B. Data quality analysis

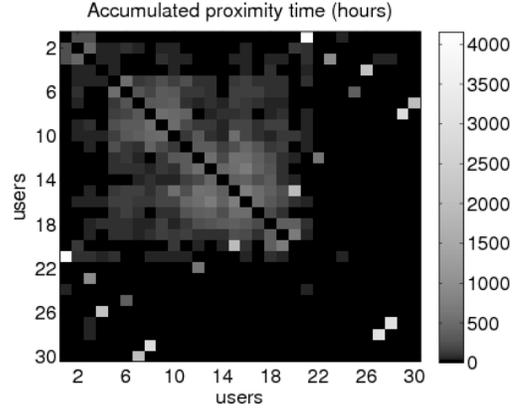


Fig. 4. Pairwise accumulated proximity time over 9 months.

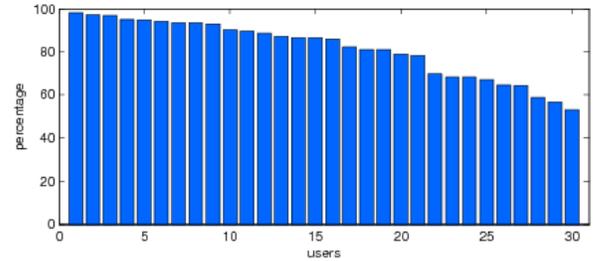


Fig. 5. Percentage of days with Bluetooth data for the set of people.

We show some basic statistics about our data in Figures 4 and 5. Based on the Bluetooth sensor, we can estimate (a lower bound of) the accumulated proximity time between people during the recording period (see Figure 4). As can be seen, for some pairs of users this value can be as high as 15 hours/day on average, while for others there is essentially little or no interaction. Note that some data may be missing due to technical problems or simply human habits related to not carrying their phones with them at all times, turning their phones off, or devices running out of battery [23]. Figure 5 shows the percentage of days with recorded Bluetooth data for the sensed community (sorted in descending order). The lowest number correspond to a person whose amount of days without data is above 50%. There are about two thirds of people who have more than 80% of days with data.

C. Robustness of Bluetooth as proximity sensor

As discussed, Bluetooth proximity data logs are quite noisy since Bluetooth technology was not primarily designed for proximity sensing. It often happens that a Bluetooth device does not detect all nearby devices in a scan. We present in this section a basic analysis of robustness of Bluetooth proximity sensor in a real condition.

Algorithm 1 Collapsed Gibbs Sampling for parameter estimation

```

1: input:  $(\mathbf{r}, \mathbf{c}) = (\mathbf{r}_s, \mathbf{c}_s)_{s=1..S}$ 
2: output: group assignment  $\mathbf{g}$ , grouping mode assignment  $\mathbf{z}$ , and model parameters,  $\theta, \phi, \nu$ .
3: initialization:
   Randomly assign group  $g_{zu}$  for each user  $u$  in each grouping mode  $z$ .
   Randomly assign grouping mode  $z_s$  for each slice  $s$ .
   Compute the count  $n_{zc}, m_{zg}, q_{zgh1}, q_{zgh2}$  according to Eq. (2)
4: while not converged do
5:   for each slice  $s$  do
6:     Decrement the counts involving  $z_s$ :  $n_{zc}, q_{zgh1}, q_{zgh2}$ 
7:     Sample the grouping mode assignment  $z_s$  according to  $P(z_s = z | \mathbf{z}_{-s}, \mathbf{g}, \cdot) \propto \frac{\alpha + n_{zc_s}}{\sum_c (\alpha + n_{zc})} \prod_{g,h} \frac{B(\gamma + \mathbf{q}_{zgh})}{B(\gamma + \mathbf{q}_{zgh})}$ 
8:     Updating the counts:  $n_{zc}, q_{zgh1}, q_{zgh2}$ 
9:   end for
10:  for each grouping mode  $z$  do
11:    for each user  $u$  do
12:      Decrement the counts involving  $g_{zu}$ :  $m_{zg}, q_{zgh1}, q_{zgh2}$ 
13:      Sample the group assignment  $g_{zu}$  according to  $P(g_{zu} = g | \mathbf{g}_{-zu}, \mathbf{z}, \cdot) \propto (\beta + m_{zg}) \prod_h \frac{B(\gamma_{gh} + \mathbf{q}_{zgh})}{B(\gamma_{gh} + \mathbf{q}_{zgh})}$ 
14:      Updating the counts:  $m_{zg}, q_{zgh1}, q_{zgh2}$ 
15:    end for
16:  end for
17: end while
18: Compute  $\nu, \phi, \theta$  according to Eq. (4)

```

We start by considering a subset of the data consisting of the weekly meetings of a group of 10 members for whom we know the exact meeting schedule over the recording period. Based on this, we would like to estimate the rate at which the phone of each person successfully detects other participants. To this end, we divide each group meeting into time slices of short duration, and draw links between people within each time slice. The ground truth for each group meeting is simply a fully connected graph using the people present at the meeting. We consider both directed and undirected graphs for the evaluation:

- An asymmetric link from user u to user v corresponds to the fact that u observed v during the slice.
- A symmetric link between u and v corresponds to the fact that u observed v or v observed u .

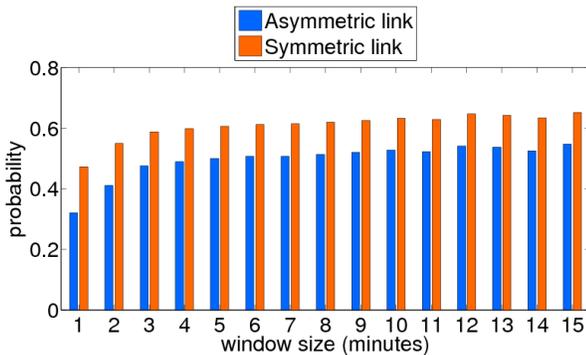


Fig. 6. Proximity detection rate of Bluetooth sensor for group meeting data.

Figure 6 reports the rate of link detection as a function of time slice duration. As can be seen, the duration of the slice is crucial as increasing the “observation” period also increases the rate of link detection. The plot also suggests to consider a slice duration of at least 5 minutes in order to obtain near optimal link detection rate with Bluetooth sensor.

Looking at the result for asymmetric link, we found that the Bluetooth sensor has a proximity detection rate of 0.5 at 10 minutes time slice. The rate can be improved by considering Bluetooth data from two users, this corresponds to the case of symmetric link where the proximity detection rate are roughly 25% better than the case of asymmetric link. Slices of 10 minutes are therefore a conservatively good choice.

VII. GROUPING MODE DISCOVERY RESULTS

In this section, we present results of our model in three series of experiments. First, we study which grouping modes can be captured from the data. Then we evaluate the predictive performance of our model on unseen data and study the influence of training data size, number of grouping modes, and number of groups on the results. Finally, we study how well our model recovers synthetic events in the presence of noise and compare the results with the group-topic model.

In our experiments, we set the slice duration to 10 minutes, resulting in $S = 39312$ non-overlapping time slices. The symmetric proximity link was used, and thus we only consider the interactions r_{su} where $u < v$ as observation. We set $\alpha = 1$ and $\beta = 1$ for the parameters of the Dirichlet distributions. On the relationship side, the Beta distribution prior $P(\nu; \gamma_{gh})$ was set so that we favor high intra-group interaction and low inter-group interaction. Also, since our data are quite sparse, we use a special “inactive group” for people who have very low interactions with others. For instance, family members of co-workers are not likely to interact with other participants of this study on working hours. To this end, we set $\gamma_{gg} = (200, 1)$ for $g = 1..(G - 1)$ and set $(1, 200)$ for all other γ . Finally, since the observed interactions between users are symmetric, the interaction between discovered groups is also symmetric.

A. Discovered grouping modes

First, we highlight some typical examples of grouping modes from our data. The number of grouping modes was

set to $Z = 50$, and we set the number of groups was set to $G = 9$, which means that in each mode, there are up to 8 active groups and the last group is always considered to be the default “inactive” group.

The discovered grouping modes can be visualized with the distribution of interaction between groups (ν_{zgh}) and the distribution of temporal context over the weekly calendar (ϕ_z). Figure 7 shows a discovered popular mode of group assignments: *by working place* (mode $z = 8$). On Figure 7(left), we show (1) the four discovered active groups involving 20 users (all other people were assigned to the inactive group and have very low interaction probability); (2) the number of users assigned to the group; and (3) the interaction probability between these groups. On Figure 7(right), we show the multinomial distribution ϕ_z over temporal context (lighter color is higher probability). As can be seen, we have high probability for working days and working time (very low probability outside office hours and weekends). The intra-group interaction probability of the 3 first groups is quite strong given that the detection rate of the Bluetooth sensor is 0.6. The relatively low interaction probability of the 5 users in group 4 reflects the fact that they do not physically work at the same part of the building, but can occasionally have interactions.

Figure 8 shows another grouping mode discovered by our model, where there is a big group of 10 users (mode $z = 48$). This mode captures some important events of the group considered in the discussion in Section VI-C, where every member is supposed to participate in the events. The corresponding temporal context distribution shows that these events are most likely to happen on Fridays. This correctly reflects that in reality, Friday is the most active day of the group where there are a morning group meeting, group lunch, and a second afternoon group session involving most members. Other interactions between members can also be detected on Mondays and Wednesdays.

Figure 9 reflects important events of the organization that was captured in grouping mode $z = 37$. These events include a weekly all-invited meeting followed by coffee break on Tuesday afternoons and an annual dinner between most members of the organization.

Finally, we show an example of a mode with many small groups in Figure 10, where the temporal context clearly indicates that these are family-like interactions (being together in the morning and at night), and in reality, these small groups are composed of family members or neighbors.

Inspired by the result in Figure 6 where the proximity detection rate might be adequate above 5 minutes time slots (for group meeting data), we also perform some experiments with this shorter time slice setting. Similar grouping modes were found such as working place interactions and group meetings (not shown here for space reasons). However, the model does not discover any big group of more than 11 people. An explanation is that the period of 5 minutes might be too short to capture interaction between people in big events (e.g. organization annual party), resulting in very sparse interaction

graphs from which the model is unable to infer a large group. This result also suggests that the choice of slice duration also depends on the application, for which we might be interested in extended time scales such as months or longer periods.

B. Predictive performance

In this section, we evaluate the predictive performance of our model on unseen data. This is a very important task in context-aware mobile application. The last month of our data set are used as test data set, and we learn the model with different training sets, varying from 2 (last) months to 8 months of data from all users. The learned models are evaluated by estimating the geometric mean of the probability that the model predicts the real interaction correctly given a time slice, called average *predictive probability*. Formally, we compute the term:

$$\left(\prod_{s=1}^{S^{test}} \prod_{u=1}^U \prod_{v=u+1}^U P(r_{suv}^{test} | c_s^{test}, \phi, \theta, \nu) \right)^{\frac{1}{S^{test}U(U-1)/2}} \quad (5)$$

where S^{test} is the number of slices in test data, and r_{suv}^{test} and c_s^{test} denote the observation of slice s in test data. As a baseline model, we compute Bernoulli distributions for each pair of users for each time slot and compute similar terms. This baseline model has no hidden variable and considers pairwise interaction of pair of users separately rather than analyzing group interaction as in our model.

Figure 11(a) shows the predictive probability as a function of training set size. Note that the predictive probabilities are very high as the interaction data is sparse (i.e. it is easy to predict no interaction for pairs of people who rarely meet). One can think of a more sophisticated measure where only some pairs of people are considered. However, we focus here on the relative performance and the measure in Eq. (5) is sufficient for this purpose. As can be seen, our model outperforms the baseline Bernoulli model. Note also that 2 months of data produce good performance for our method. Furthermore, we obtain an optimum performance with 4 months of data, which suggests that using “too old” data might not help improving predictive performance.

We continue the analysis by studying the influence of number of modes and number of groups on the results. Figures 11(b) and 11(c) show the predictive probability obtained by varying the number of groups G and the number of modes Z , respectively. In general, having more groups and more modes slightly helps the model fit the data better. Note that a too large number for these hyper-parameters may lead to over fitting as well as high algorithmic complexity. In practice, these hyper-parameters can be chosen based on the trade off between complexity and performance, e.g. $Z = 50, G = 9$ as set in the previous section.

C. Simulating group events

In order to evaluate how well the model learns from the input data, we designed a final experiment where we introduce a number of “strong” artificial events in the data to see whether



Fig. 7. Mode $z = 8$: Grouping by working place.

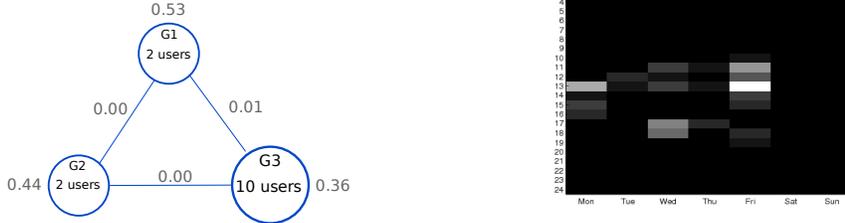


Fig. 8. Mode $z = 48$: Group meeting



Fig. 9. Mode $z = 37$: Big events of the organization.

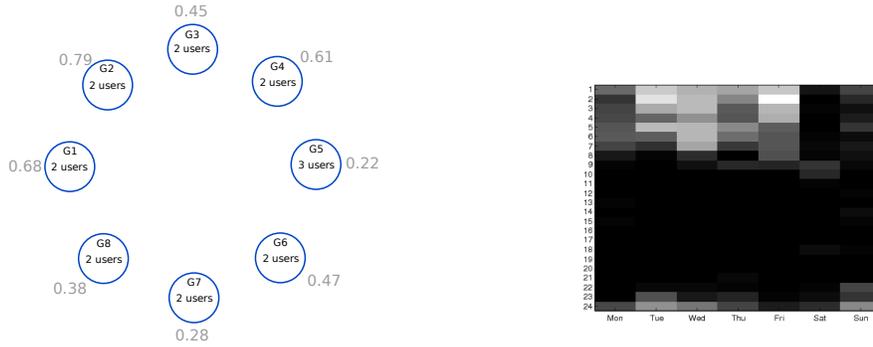


Fig. 10. Mode $z = 41$: Being with family member.

the model can recover them all. In addition to the real data, we select a number of users for each event and simulate their interactions during the artificial events with some noise (i.e. some random links between participants can be missing during the artificial event). We consider three simulated events of various group sizes as listed in Table III.

Ideally, we expect that the model will assign slices from the same event to the same grouping mode, and that slices from different events will be assigned to different grouping modes. Let τ be a matrix of size $K \times Z$ where $K = 3$ is the

TABLE II
SYNTHETIC EVENTS, INCLUDING NUMBER OF PARTICIPANTS AND SCHEDULE.

| description | #part. | time |
|--------------------|--------|--------------------|
| Small Meeting (SM) | 5 | Thursday 10am-12am |
| Group Meeting (GM) | 9 | Wednesday 2pm-3pm |
| Big Meeting (BM) | 20 | Monday 4pm-5pm |

number of simulated events, and τ_{kz} stands for the probability of assigning slices to mode z in event k . Assume that for each event, we have a multinomial distribution over grouping

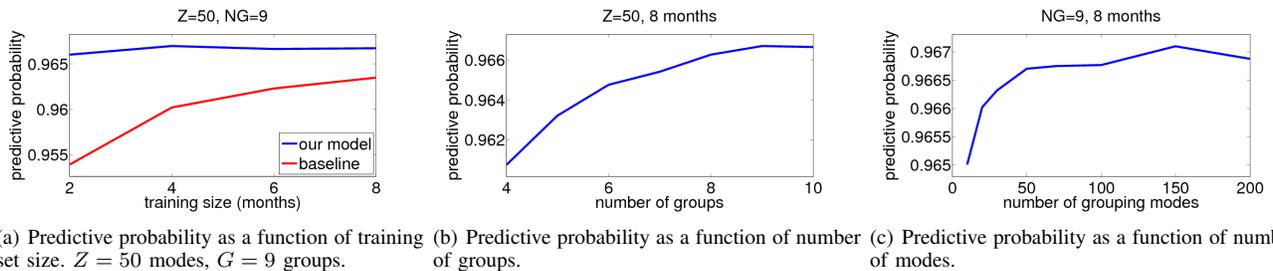


Fig. 11. Predictive performance.

TABLE III
RESULTS WITH SIMULATED EVENTS (LOWER IS BETTER).

| noise | events | This paper | | GT [27] | |
|-------|--------|------------|------|---------|------|
| | | PI | OI | PI | OI |
| 0% | SM | 1.00 | | 2.20 | |
| | GM | 1.00 | 0.00 | 1.00 | 0.00 |
| | BM | 1.00 | | 1.00 | |
| 20% | SM | 1.02 | | 1.03 | |
| | GM | 1.00 | 0.00 | 1.00 | 0.00 |
| | BM | 1.00 | | 1.00 | |
| 40% | SM | 1.02 | | 1.25 | |
| | GM | 1.00 | 0.00 | 1.03 | 1.99 |
| | BM | 1.00 | | 1.00 | |
| 60% | SM | 1.58 | | 22.68 | |
| | GM | 1.00 | 0.00 | 1.67 | 0.81 |
| | BM | 1.00 | | 6.74 | |

modes. Then, the perplexity index (PI) of this distribution can be computed by:

$$PI_k = \exp \left\{ - \sum_{z=1}^Z \tau_{kz} \log(\tau_{kz}) \right\} \quad (6)$$

This measure is appropriate as the smaller the value of PI_k , the better the model captures the event k . However, if we assigned all the slices to the same mode, then $PI_k = 1$ for all k . Hence, we also consider a second criterion, called overlapping index (OI):

$$OI = \sum_{z=1}^Z \sum_{k=1}^K \sum_{k'=1, k' \neq k}^K \tau_{kz} \tau_{k'z} \quad (7)$$

The main idea behind this measure is that if two events have many slices assigned to the same grouping mode then OI will be high. On the contrary, if the events are never overlapped then we obtain the optimal value, $OI = 0$.

We compare our model with the group-topic model [27], which has been re-interpreted so as to use exactly the same input we use for our model, and thus aims at the same goal of finding grouping modes. Table III shows the results of experiments with additional simulated data with different levels of noise, defined by the percentage of missing links in the events. In general, large events are easier to capture than small events. Our model successfully captures the simulated events and robustly reaches optimal scores ($PI = 1$ and $OI = 0$) in most events, even with high level of noise. Besides, the group-topic model recovers the events in the low noise condition, but fails to recover events with high noise.

VIII. CONCLUSION

In this paper, we presented a novel probabilistic approach for real-life social context discovery, which finds emergent grouping modes from longitudinal proximity data. Our framework is validated with large-scale Bluetooth proximity data recorded continuously over 9 months of real-life over a set of 30 people who have real social connections. We conduct an extensive evaluation of our model studying: which social events can be captured; what predictive performance on unseen data can be obtained; and how well the model recovers simulated events. Our model represents an improvement over an existing model on the way interactions are modeled, and show advantages in terms of recovering small-group events. This is in practice important as most of daily-life face-to-face interactions involve small groups of people.

Our framework can be further extended to analyze additional contexts such as the location on which the group event takes place, and to learn from not only Bluetooth data but also other interaction data source such as phone calls, WIFI proximity, GPS, etc. This is part of future work. While our approach is generic, it has some hyperparameters (i.e. the number of groups G , the number of grouping modes Z , and the slide duration) that depend on the application. Finding these hyperparameters automatically from the data is also an interesting direction to explore.

ACKNOWLEDGMENTS

This work was funded by Nokia Research Center Lausanne (NRC) through the LS-CONTEXT project.

REFERENCES

- [1] E. M. Airoldi, D. M. Blei, S. E. Fienberg, and E. P. Xing. Mixed membership stochastic blockmodels. *Journal of Machine Learning Research*, 9:1981–2014, 2008.
- [2] E. M. Airoldi, D. M. Blei, S. E. Fienberg, E. P. Xing, and T. Jaakkola. Mixed membership stochastic block models for relational data with application to protein-protein interactions. In *Proc. International Biometrics Society Annual Meeting*, 2006.
- [3] N. Banerjee, S. Agarwal, V. Bahl, R. Chandra, A. Wolman, and M. D. Corner. Virtual Compass: relative positioning to sense mobile social interactions. In *Proc. Pervasive Computing*, pages 1–21, Helsinki, Finland, May 2010.
- [4] A. Clauset and N. Eagle. Persistence and periodicity in a dynamic proximity network. In *DIMACS Workshop on Computational Methods for Dynamic Interaction Networks*, 2007.
- [5] N. Eagle, A. S. Pentland, and D. Lazer. Mobile phone data for inferring social network structure. In *Social Computing, Behavioral Modeling, and Prediction*, pages 79–88. Springer US, 2008.

- [6] N. Eagle, J. A. Quinn, and A. Clauset. Methodologies for continuous cellular tower data analysis. In *Proc. Pervasive Computing*, pages 342–353, Berlin, Heidelberg, 2009. Springer-Verlag.
- [7] N. Eagle and A. (Sandy) Pentland. Reality mining: sensing complex social systems. *Personal Ubiquitous Comput.*, 10(4):255–268, 2006.
- [8] W. Fu, L. Song, and E. P. Xing. Dynamic mixed membership block-model for evolving networks. In *Proc. International Conference on Machine Learning*, pages 329–336, New York, NY, USA, 2009. ACM.
- [9] W. R. Gilks. *Markov Chain Monte Carlo In Practice*. Chapman and Hall/CRC, 1999.
- [10] J. Gips and A. Pentland. Mapping human networks. In *Proc. Pervasive Computing and Communications*, pages 159–168. IEEE Computer Society, 2006.
- [11] T. L. Griffiths and M. Steyvers. Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101(Suppl. 1):5228–5235, April 2004.
- [12] 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, Feb 2008.
- [13] P. W. Holland and S. Leinhardt. Local structure in social networks. *Sociological Methodology*, page 145, 1975.
- [14] N. K. and S. T. A. B. Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association*, 96:1077–1087, September 2001.
- [15] N. Kiukkonen, J. Blom, O. Dousse, D. Gatica-Perez, and J. Laurila. Towards rich mobile phone datasets: Lausanne data collection campaign. In *Proc. ICPS*, Berlin, 2010.
- [16] S. Mardenfeld, D. Boston, S. Juan Pan, Q. Jones, A. Iammitchi, and B. Cristian. Gdc: Group discovery using co-location traces. In *IEEE Symposium on Social Computing Applications (SCA 2010)*, 2010.
- [17] A. McGovern, L. Friedland, M. Hay, B. Gallagher, A. Fast, J. Neville, and D. Jensen. Exploiting relational structure to understand publication patterns in high-energy physics. *SIGKDD Explor. Newsl.*, 5(2):165–172, 2003.
- [18] E. Minkov and W. W. Cohen. An email and meeting assistant using graph walks. In *Proc Conference on Email and Anti-Spam*, 2006.
- [19] T. Nicolai, N. Behrens, and H. Kenn. Exploring social context with the wireless rope. In *Proc. Workshop MONET: LNCS 4277*, 2006.
- [20] K. Nowicki and T. A. B. Snijders. Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association*, 96(455):1077–1087, September 2001.
- [21] D. O. Olguin, B. N. Waber, T. Kim, A. Mohan, K. Ara, and A. Pentland. Sensible organizations: Technology and methodology for automatically measuring organizational behavior. *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, pages 43–55, 2009.
- [22] E. O’neill, V. Kostakos, T. Kindberg, A. Schiek, A. Penn, D. Fraser, and T. Jones. Instrumenting the city: Developing methods for observing and understanding the digital cityscape. In *Proc. Ubiquitous Computing*, pages 315–332. Springer, 2006.
- [23] S. N. Patel, J. A. Kientz, G. R. Hayes, S. Bhat, and G. D. Abowd. Farther than you may think: An empirical investigation of the proximity of users to their mobile phones. In *Proc. Ubiquitous Computing, P. Dourish*, pages 123–140. Springer, 2006.
- [24] M. Raento, A. Oulasvirta, R. Petit, and H. Toivonen. Contextphone: A prototyping platform for context-aware mobile applications. *IEEE Pervasive Computing*, 4(2):51–59, 2005.
- [25] F. S. Sampson. *A Novitiate in a Period of Change: An Experimental and Case Study of Social Relationships*. PhD thesis, Cornell University, 1968.
- [26] M. Terry, E. D. Mynatt, K. Ryall, and D. Leigh. Social net: using patterns of physical proximity over time to infer shared interests. In *Proc. Human Factors in Computing Systems*, pages 816–817, New York, NY, USA, 2002. ACM.
- [27] X. Wang, N. Mohanty, and A. McCallum. Group and topic discovery from relations and their attributes. In *Proc. Advances in Neural Information Processing Systems*, pages 1449–1456, 2006.
- [28] S. Wasserman and K. Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1994.
- [29] D. Wyatt, T. Choudhury, and H. Kautz. Capturing spontaneous conversation and social dynamics: A privacy sensitive data collection effort. In *Proc. of ICASSP*, 2007.

APPENDIX A: MATHEMATICAL DERIVATION OF EQ. 1

Begin with the joint probability of observation (\mathbf{c}, \mathbf{r}) , mode assignments \mathbf{z} and group assignments \mathbf{g} :

$$\begin{aligned} P(\mathbf{c}, \mathbf{r}, \mathbf{z}, \mathbf{g}; \alpha, \beta, \gamma) &= P(\mathbf{z}) \int_{\phi} (P(\mathbf{c}|\mathbf{z}, \phi) P(\phi|\alpha)) \partial\phi \\ &\int_{\theta} (P(\mathbf{g}|\mathbf{z}, \theta) P(\theta|\beta)) \partial\theta \int_{\nu} (P(\mathbf{r}|\mathbf{z}, \mathbf{g}; \nu) P(\nu|\gamma)) \partial\nu \\ &= \frac{1}{Z^S} \prod_z \int_{\phi_z} \left(\prod_c \phi_{zc}^{n_{zc}} \frac{\prod_c \phi_{zc}^{\alpha-1}}{B(\alpha)} \right) \partial\phi_z \\ &\prod_z \int_{\theta_z} \left(\prod_g \theta_{zg}^{m_{zg}} \frac{\prod_g \theta_{zg}^{\beta-1}}{B(\beta)} \right) \partial\theta_z \\ &\prod_{zgh} \int_{\nu_{zgh}} \nu_{zgh}^{q_{zgh}} (1 - \nu_{zgh})^{q_{zgh} - 1} \frac{\nu_{zgh}^{\gamma h - 1} (1 - \nu_{zgh})^{\gamma g h - 1}}{B(\gamma_{gh})} \partial\nu_{zgh} \end{aligned}$$

After grouping similar terms, we have

$$\begin{aligned} P(\mathbf{c}, \mathbf{r}, \mathbf{z}, \mathbf{g}; \alpha, \beta, \gamma) &= \frac{1}{Z^S} \prod_z \int_{\phi_z} \left(\prod_c \frac{\phi_{zc}^{n_{zc} + \alpha - 1}}{B(\alpha)} \right) \partial\phi_z \\ &\prod_z \int_{\theta_z} \left(\prod_g \frac{\theta_{zg}^{m_{zg} + \beta - 1}}{B(\beta)} \right) \partial\theta_z \\ &\prod_{zgh} \int_{\nu_{zgh}} \frac{\nu_{zgh}^{q_{zgh} + \gamma_{gh} - 1} (1 - \nu_{zgh})^{q_{zgh} + \gamma_{gh} - 1}}{B(\gamma_{gh})} \partial\nu_{zgh} \end{aligned}$$

where the equations inside the integrations have the same form as the prior distribution (Dirichlet and Beta), and can be simplified based on the fact the integration over a prior distribution is always 1. Finally, the joint distribution is greatly simplified:

$$\begin{aligned} P(\mathbf{c}, \mathbf{r}, \mathbf{z}, \mathbf{g}; \alpha, \beta, \gamma) &= \frac{1}{Z^S} \prod_z \frac{B(\alpha + \mathbf{n}_z)}{B(\alpha)} \prod_z \frac{B(\beta + \mathbf{m}_z)}{B(\beta)} \prod_{z,g,h} \frac{B(\gamma_{gh} + \mathbf{q}_{zgh})}{B(\gamma_{gh})} \end{aligned}$$

APPENDIX B: DERIVATION OF EQ. 3

The conditional pdf can be computed efficiently based on the fact that they are proportional to the joint pdf. To perform sampling, we do not need the absolute value of conditional distribution, and any invariant term (wrt. sampling variable) can be eliminated.

$$\begin{aligned} P(z_s | \mathbf{z}_{-s}, \mathbf{g}, \mathbf{c}, \mathbf{r}) &\propto \frac{P(\mathbf{c}, \mathbf{r}, \mathbf{z}, \mathbf{g})}{P(\mathbf{z}_{-s}, \mathbf{g}, \mathbf{c}, \mathbf{r})} \\ &\propto \frac{\frac{1}{Z^S} \prod_z \frac{B(\alpha + \mathbf{n}_z)}{B(\alpha)} \prod_z \frac{B(\beta + \mathbf{m}_z)}{B(\beta)} \prod_{z,g,h} \frac{B(\gamma_{gh} + \mathbf{q}_{zgh})}{B(\gamma_{gh})}}{\frac{1}{Z^{S-1}} \prod_z \frac{B(\alpha + \mathbf{n}_{-s}^z)}{B(\alpha)} \prod_z \frac{B(\beta + \mathbf{m}_{-s}^z)}{B(\beta)} \prod_{z,g,h} \frac{B(\gamma_{gh} + \mathbf{q}_{zgh}^s)}{B(\gamma_{gh})}} \\ &\propto \frac{B(\alpha + \mathbf{n}_{z_s}) B(\beta + \mathbf{m}_{z_s}) \prod_{g,h} B(\gamma_{gh} + \mathbf{q}_{z_s g h})}{B(\alpha + \mathbf{n}_{-s}^z) B(\beta + \mathbf{m}_{-s}^z) \prod_{g,h} B(\gamma_{gh} + \mathbf{q}_{-s g h})} \end{aligned}$$

Note that \mathbf{n}_{z_s} differs \mathbf{n}_{-s}^z by only $n_{z_s c_s}$ then we can show that $\frac{B(\alpha + \mathbf{n}_{z_s})}{B(\alpha + \mathbf{n}_{-s}^z)} = \frac{\alpha + n_{z_s c_s}}{\sum_c (\alpha + n_{z_s c})}$. Furthermore, \mathbf{m}_z does not depend on z_s , then $B(\beta + \mathbf{m}_{z_s}) = B(\beta + \mathbf{m}_{-s}^z)$, we get:

$$\begin{aligned} P(z_s = z | \mathbf{z}_{-s}, \mathbf{g}, \cdot) &\propto \frac{\alpha + n_{z_s c_s}}{\sum_c (\alpha + n_{z_s c})} \prod_{g,h} \frac{B(\gamma + \mathbf{q}_{z g h})}{B(\gamma + \mathbf{q}_{-s g h})} \end{aligned}$$

By analogy, we estimate the second sampling equation:

$$\begin{aligned} P(g_{zu} | \mathbf{g}_{-zu}, \mathbf{z}, \cdot) &\propto \frac{\frac{1}{Z^S} \prod_z \frac{B(\alpha + \mathbf{n}_z)}{B(\alpha)} \prod_z \frac{B(\beta + \mathbf{m}_z)}{B(\beta)} \prod_{z,g,h} \frac{B(\gamma_{gh} + \mathbf{q}_{zgh})}{B(\gamma_{gh})}}{\frac{1}{Z^S} \prod_z \frac{B(\alpha + \mathbf{n}_{-zu}^z)}{B(\alpha)} \prod_z \frac{B(\beta + \mathbf{m}_{-zu}^z)}{B(\beta)} \prod_{z,g,h} \frac{B(\gamma_{gh} + \mathbf{q}_{zgh}^u)}{B(\gamma_{gh})}} \\ &\propto \frac{B(\beta + \mathbf{m}_z) \prod_h B(\gamma_{g_z u h} + \mathbf{q}_{z g_z u h})}{B(\beta + \mathbf{m}_{-zu}^z) \prod_h B(\gamma_{g_{-zu} h} + \mathbf{q}_{-z g_{-zu} h})} \\ &\propto (\beta + m_{zu} - 1) \prod_h \frac{B(\gamma_{g_z u h} + \mathbf{q}_{z g_z u h})}{B(\gamma_{g_{-zu} h} + \mathbf{q}_{-z g_{-zu} h})} \end{aligned}$$