

Learning Geo-Social User Topical Profiles with Bayesian Hierarchical User Factorization

Haokai Lu, Wei Niu, and James Caverlee

Department of Computer Science and Engineering, Texas A&M University
hlu,wei,caverlee@tamu.edu

ABSTRACT

Understanding user interests and expertise is a vital component toward creating rich user models for information personalization in social media, recommender systems and web search. To capture the pair-wise interactions between geo-location and user's topical profile in social-spatial systems, we propose the modeling of fine-grained and multi-dimensional *user geo-topic profiles*. We then propose a two-layered Bayesian hierarchical user factorization generative framework to overcome user heterogeneity and another enhanced model integrated with user's contextual information to alleviate multi-dimensional sparsity. Through extensive experiments, we find the proposed model leads to a 5~13% improvement in precision and recall over the alternative baselines and an additional 6~11% improvement with the integration of user's contexts.

ACM Reference Format:

Haokai Lu, Wei Niu, and James Caverlee. 2018. Learning Geo-Social User Topical Profiles with Bayesian Hierarchical User Factorization. In *SIGIR '18: The 41st International ACM SIGIR Conference on Research and Development in Information Retrieval*, July 8–12, 2018, Ann Arbor, MI, USA. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3209978.3210044>

1 INTRODUCTION

Uncovering user interests and expertise is a vital component of search and recommendation systems [14, 16, 30, 37, 38]. For example, knowing a user is interested in tennis and proficient in python can be used to augment newsfeed ranking algorithms to surface high-quality content, improve item-based recommenders by leveraging the topical expertise of knowledgeable users [3], and enhance personalized web search [26, 30] and targeted ads [2].

While many existing *user topical profiles* focus on a global view of each user, there are important *geo-social* factors that are critical to consider: (i) First, individual users may be perceived differently in different places; and (ii) Users with similar topical profiles can have very different geo-spatial impact. Without careful consideration of each of these factors, user topical profiles will lead to error-prone recommendation (e.g., recommending content generated by local foodies to people interested in food but living far away), low-quality advertising (e.g., placing online ads of a new product by a tennis equipment brand company in areas little known by local people) and so on.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

SIGIR '18, July 8–12, 2018, Ann Arbor, MI, USA

© 2018 Association for Computing Machinery.

ACM ISBN 978-1-4503-5657-2/18/07...\$15.00

<https://doi.org/10.1145/3209978.3210044>

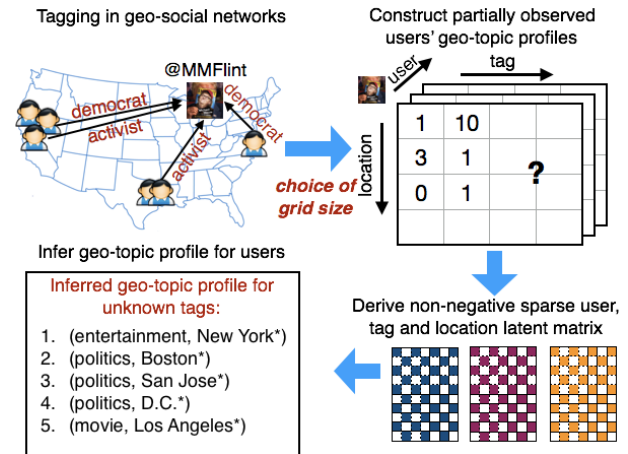


Figure 1: Learning geo-social user topical profiles. * indicates the grid where the cities locate.

To illustrate these two geo-social factors, Figure 2 shows the heat maps of the locations of Twitter users who have labeled Michael Moore (@MMflint) and Roy Blunt (@RoyBlunt) using Twitter Lists. This aggregate crowd-labeling of Twitter users has been used previously to derive user topical profiles [4, 9, 11]. For Factor (i), we see that @MMflint, as a filmmaker, is mainly known in New York and LA, while he has a much broader impact for politics in regions such as San Francisco and D.C. For Factor (ii), @MMflint and @RoyBlunt are both known for politics, but @RoyBlunt is known mainly in Missouri and D.C. while @MMflint is known in a much broader geo-scope.

Hence, these observations motivate the need for a careful study of the impact of social-spatial properties on the creation of high-quality user topical profiles. We propose in this paper to complement traditional user topical profiles with new fine-grained *user geo-topic profiles*. These profiles are designed to capture the variations of user popularity for topics across geo-locations; essentially, geo-topic profile is a multi-dimensional concept to describe and model a user, and is expected to capture the pair-wise interactions involving geo-locations and users' topical profiles (see Figure 1).

Modeling user geo-topic profiles faces two major challenges, however: (i) they are often overdispersed. Unlike the ratings or content studied in many previous works [6, 12, 13], the popularity counts in geo-topic profiles are of great variability in terms of variance which is caused by user heterogeneity (some users are much more popular than others); and (ii) they are often extremely sparse due to multi-dimensionality, with many users often known for very few topics at certain locations.

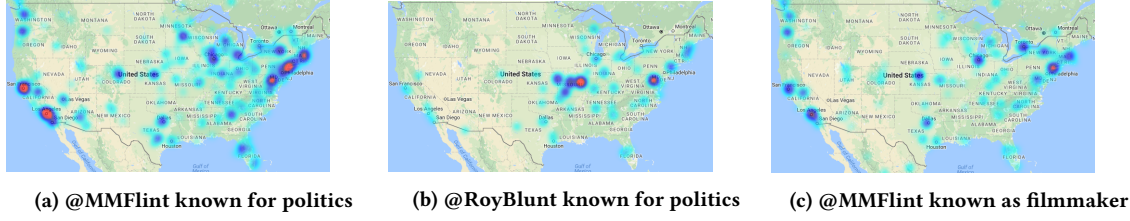


Figure 2: Heatmaps of Twitter users listing (a) @MMflint for politics; (b) @RoyBlunt for politics; (c) @MMflint as a filmmaker.

Given these challenges, we first propose a multi-layered (two-layered in our case) Bayesian hierarchical user factorization which generalizes Poisson gamma belief network [40] from modeling two dimensional non-negative counts to multi-dimensional heterogeneous counts. The extra layer of user factorization learns a more expressive user model than that of single-layered factorization by allowing a larger variance-to-mean ratio on user geo-topic profiles, thus making it better equipped at handling overdispersion and user heterogeneity. To alleviate the sparsity issue, we investigate how user’s contexts, specifically, geo-location and social ties, correlate with one’s geo-topic profile, and then propose to integrate them into the two-layered hierarchical model to learn better representation of user’s geo-topic preference by others. Due to the non-conjugacy of the multi-layered factorization scheme, we exploit a data augmentation scheme for negative binomial (NB) distribution and develop an efficient closed-form Gibbs sampling formula for scalable inference. We summarize the contributions as follows:

- First, we introduce multi-dimensional geo-topic profile to capture the pair-wise interactions between geo-location and user’s topical profile, and formulate the problem of learning user’s geo-topic profile to infer user’s location-sensitive topical profile.
- Second, to overcome overdispersed popularity counts caused by user heterogeneity, we propose a two-layered Bayesian hierarchical user factorization (**bHUF**) generative framework, which can be easily generalized to deep user factorization.
- Third, to alleviate sparsity, we investigate the impact of user’s contexts, specifically, user’s geo-location and social ties, on user’s geo-topic profile, and propose an enhanced model (**bHUF+**) by exploiting user’s contexts. We then develop an efficient closed-form Gibbs sampling scheme for inference using data augmentation schemes for NB distribution.
- Finally, we evaluate bHUF and bHUF+ against several baselines over GPS-tagged Twitter datasets, and observe that bHUF gives about 5%~13% improvement in precision and recall over the best alternative one-layered baseline, and an additional 6%~11% improvement with user’s geo-location and social context.

2 RELATED WORK

Identifying user’s interests and expertise. Finding user interests and expertise is critical for many important applications, including personalized search [24, 26, 30], recommender systems [14, 23, 37, 38], expert mining [9, 11, 33] and targeted advertising [2]. Many of these applications seek to find interests or expertise either explicitly [5, 8, 11, 14] via user relevant features and tags, or uncover latent interest or expertise implicitly [2, 15, 36–38] with

models such as matrix factorization. For example, Hong et al. [15] proposed probabilistic language models to uncover implicit geographical topics and user interests in social media. Cao et al. [5] presented a matrix factorization-based approach to infer users’ topical profiles by exploiting their footprints. To increase user base and engagement, many works [16, 19, 34, 37] create rich user models for better user experience. For example, Zhao et al. [37] proposed a matrix factorization based approach to build topical profiles by distinguishing user behaviors. Jiang et al. [19] presented a probabilistic matrix factorization model to exploit social context to better predict user adoption behaviors on Weibo. Most of these efforts, however, focus on uncovering user’s topical profile or latent interests without explicitly considering geo-social factors. In contrast, we propose modeling user *geo-topic profiles* to explicitly infer location-sensitive topical profiles with the combination of topical tags and locations.

Modeling discrete count data. An emerging line of research [7, 12, 17, 35] has focused on modeling discrete count data with Gamma-Poisson distribution instead of traditional Gaussians for recommender systems [6, 12], topic modeling [13], spatial data analysis [35] and political science [29]. For example, Gopalan et al. [12] proposed Poisson matrix factorization for item recommendation and later developed collaborative topic Poisson factorization [13] by integrating topic modeling for better article recommendation. Yu et al. [35] proposed a weakly-supervised labeled Poisson factor model to predict the number of app openings at different locations from mobile app usage. More recently, Schein et al. [29] proposed Bayesian Poisson tensor factorization for inferring multi-lateral relations between countries based on political events data. Compared with previous works, in our study, we focus on learning geo-social user topical profiles, and propose a multi-layered contextual hierarchical user factorization framework which generalizes Poisson gamma belief network to modeling multi-dimensional overdispersed sparse tensor counts.

Exploiting user’s contextual information. Various contextual signals have been exploited to improve learning user’s latent preferences and behaviors in recommender systems, including text [20, 32], social networks [19, 21, 23], geographical influence [18, 22] and temporal context [7, 34]. For example, Want et al. [32] proposed collaborative topic regression to learn latent user preference by modeling both ratings and content. User’s social relations [19, 23] have also been explored to regulate similar user’s latent preference for better recommendation. Hu et al. [18] considered business geographical location to improve rating prediction. In this work, we explore how user’s geo-location and social ties correlate with

one’s *geo-topic profile*, and then extend the multi-layered user factorization to model user’s contexts so that it can capture the useful information contained within user’s attributes.

3 PROBLEM DEFINITION AND CHALLENGES

We assume there exist a set of users \mathcal{U} in a social-spatial network, a dictionary of tags \mathcal{T} that are used to indicate user’s topical profile and a total of locations \mathcal{L} . We use capital letters U , T and L to represent their sizes, respectively. In a social-spatial network such as Twitter or LinkedIn, a user u is often tagged or followed by people from different locations for her topical expertise t . For instance, Twitter users are often tagged by others in lists with their own selected keywords, with some examples shown in Figure 2. LinkedIn users can also use *skill tags* for their own profiles and be endorsed by others with these tags. With these notations, we first give several preliminary definitions as follows.

User’s Geo-Topic Token. A geo-topic token of user u is defined as a quadruplet $\{u \xleftarrow{t} v, l\}$ indicating that user u is followed/tagged by user v at location l with a tag t . To study user’s popularity of a topic t at a location l , we aggregate all of u ’s geo-topic tokens with respect to t and l , and obtain a count y that indicates the extent to which the user is popular.

User’s Geo-Topic Profile. A user u ’s geo-topic profile in a social-spatial network is defined to be a set of quadruplets \mathcal{P}_u , with each quadruplet p_u representing this user’s popularity y , where $y \in \mathbb{Z}_{\geq 0}$, for topic t at location l , denoted $\{u, t, l, y\}$.

From the definition of user’s geo-topic profile, we can easily obtain user’s topical profile $\{u, t, y\} \in \mathcal{T}_u$ by aggregating quadruplets \mathcal{P}_u with the same topic. Similarly, user’s location profile $\{u, l, y\} \in \mathcal{L}_u$ can be obtained by aggregating \mathcal{P}_u with the same location. Thus, user’s geo-topic profile provides a more generalized and finer approach to profiling users by considering geo-spatial influence of user’s topical profile. Very often, we are only able to observe user’s geo-topic profile for partial tags or locations due to limited visibility or geo-tagging capabilities of users. In light of this, we formally define the following problem to infer user’s geo-topic profiles for unknown topical tags or locations.

Learning User’s Geo-Topic Profile. Given partially observed user geo-topic profile, i.e., a subset of quadruplets $\mathcal{P}_u^o \in \mathcal{P}_u$, our task is to infer the unobserved quadruplets $\mathcal{P}_u - \mathcal{P}_u^o$ by ranking user u ’s popularity for different tags at different locations, i.e., identify top-ranking tag-location combination $\{t, l\}$ according to learned score \hat{y}_{utl} for user u . The overall framework is illustrated in Figure 1.

Basic Approach with Bayesian Poisson Tensor Factorization (BPTF). Bayesian Poisson factorization has shown promising performance for several tasks such as item and content recommendation [12, 13] by modeling count-valued data. By assuming Poisson distribution in modeling counts instead of Gaussian distribution, it is better at modeling the zero observations encountered in sparse data [10] and enjoys scalable learning since only non-zeros are considered during inference. It is extended to BPTF in [29] for modeling high dimensional sparse dyadic tensor data. In our context, to model user u ’s popularity for tag t at location l , denoted y_{utl} in our context, it is generated through Canonical PARAFAC (CP)

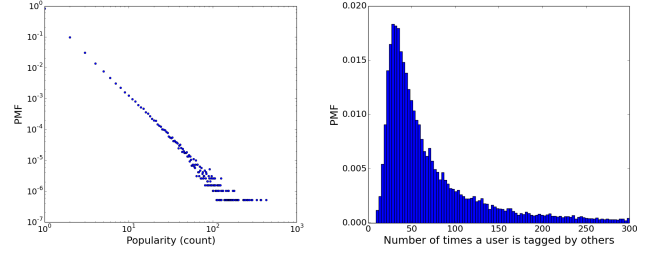


Figure 3: Left: probability mass function of user’s popularity counts with a grid size of $2.5^\circ \times 2.5^\circ$ (latitude by longitude). Right: histogram of the number of geo-topic tokens. Similar distributions have been observed in other sizes of grids.

decomposition with Poisson distribution as follows,

$$y_{utl} \sim \text{Poisson}\left(\sum_{k=1}^K \theta_{uk} \phi_{tk} \psi_{lk}\right) \quad (1)$$

where $\theta_{uk} \in \mathbb{R}_{\geq 0}^K$ is the latent factor of u representing user’s geo-topic preference by others. ϕ_{tk} and $\psi_{lk} \in \mathbb{R}_{\geq 0}^K$ represents tag t ’s and location l ’s latent factor. Each component of these latent factors is drawn from conjugate Gamma distributions with shape parameter α and rate parameter β . Small values of α induces sparsity and better interpretability of inferred latent factors.

As a first pass, we can attack the problem with BPTF by using inferred latent factors θ_{uk} , ϕ_{tk} and ψ_{lk} to compute the expected missing values in a user’s geo-topic profile. However, a preliminary data analysis suggests that users’ geo-topic profiles present the following two key challenges:

Challenge 1: Geo-topic profiles are overdispersed. To illustrate, we first plot the probability mass function of user’s popularity counts in Figure 3. As we can see, the non-zero counts follow the power law distribution with most of counts very low, ranging from 1 to almost a few hundreds with a variance-to-mean ratio about 15.4. The great variability of counts can be explained by the fact that users are often known by others to very different extent, i.e., some users can be very popular while others are lesser known. This heterogeneity from users requires the underlying model to better capture the higher moments (e.g., variance) of the data. However, the modeling capability of Equation 1 can be limited since its variance is constrained to be equal to its mean, thus may cause underfitting when applied on these overdispersed popularity counts.

Challenge 2: Geo-topic profiles are extremely sparse. User’s geo-topic profile, as is observed from our data, is extremely sparse with only a small portion of non-zeros due to its multi-dimensionality. Specifically, many users are often known for very few topics at certain locations. Indeed, our data shows a sparsity level of having 0.047% or 0.018% of non-zero counts at a grid size of $2.5^\circ \times 2.5^\circ$ or $0.5^\circ \times 0.5^\circ$ in terms of latitude by longitude. One common approach to alleviate the issue is to exploit user’s contextual information which commonly exists in social-spatial networks. However, it is not clear whether user’s contexts have positive correlations with one’s geo-topic profile, and if there is, how to model and integrate them to better learn user’s geo-topic profile.

4 BAYESIAN HIERARCHICAL USER FACTORIZATION

To address the challenges, in this section, we first propose a two-layered hierarchical model (bHUF) to overcome geo-topic profile overdispersion caused by user heterogeneity, and then investigate how user’s contexts influence one’s geo-topic profile, and finally present an enhanced model (bHUF+) by integrating user’s contexts to alleviate the sparsity issue.

4.1 Two-layered Hierarchical Model

bHUF is a hierarchically constructed multi-dimensional Bayesian framework which generalizes the recently proposed Poisson Gamma belief network [40] from modeling two dimensional non-negative counts to modeling multi-dimensional tensor counts. Specifically, it generates user latent factor with another layer of factorization by decomposing it into the product of a second-layer user latent factor and a weight-connection matrix. This multi-layered construction is equivalent to modeling the counts with NB distribution which is shown in later section to have a variance-to-mean ratio larger than that of single-layered construction by a multiplicative factor. As a result, it is better equipped at handling overdispersed counts induced by user heterogeneity.

As in Equation 1, we assume user u ’s popularity y_{utl} for tag t at location l is generated through CP decomposition with Poisson distribution. However, in Equation 1, user’s latent factor is only generated through Gamma priors with hyper-parameters. Instead, we let each user’s latent factor be generated through a second layer of factorization, with each first-layer user latent factor conditioned upon a second-layer user latent factor. The overall generative framework, given y_{utl} , is described as follows,

$$\begin{aligned}
 y_{utl} &\sim \text{Poisson}\left(\sum_{k=1}^K \theta_{uk} \phi_{tk} \psi_{lk}\right) \\
 \theta_{uk} &\sim \text{Gamma}(\theta_{uk_2} \Omega, \beta_u) \\
 \phi_{:m} &\sim \text{Dirichlet}(\eta^t, \dots, \eta^t) \\
 \psi_{:m} &\sim \text{Dirichlet}(\eta^l, \dots, \eta^l) \\
 \Omega_{k_2:} &\sim \text{Dirichlet}(\eta_0, \dots, \eta_0)
 \end{aligned} \tag{2}$$

where θ_{uk} is Gamma distributed, with the shape parameter dependent on the second-layer user latent factor θ_{uk_2} . θ_{uk_2} is given a Gamma distribution as a prior. For tag and location latent factor, we impose Dirichlet distributions as their priors, where each column of the tag and location latent factor $\phi_{:m}$ and $\psi_{:m}$ is Dirichlet distributed. The weight matrix $\Omega \in \mathbb{R}^{K_2 \times K}$ connects the first layer user factor θ_{uk} to the second layer user factor θ_{uk_2} , and is expected to capture the correlations between different layers of user factors. Each row of Ω is given Dirichlet distribution as its prior.

Using Dirichlet distribution as the priors for tag, location and weight matrices brings several benefits: (i) it naturally imposes non-negativity constraints [10]; (ii) Since each column of $\phi_{:m}$ and $\psi_{:m}$ is restricted to have L_1 norm, they can be considered as a distribution over the corresponding entities, resulting in better interpretability of inferred tag and location factors; (iii) The L_1 norm of these factors allows the tag and location factor not to directly interact with user’s latent factors. This property is especially desired to generalize our

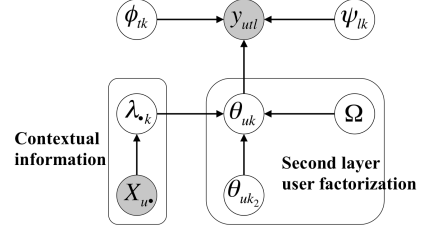


Figure 4: Overall generative framework.

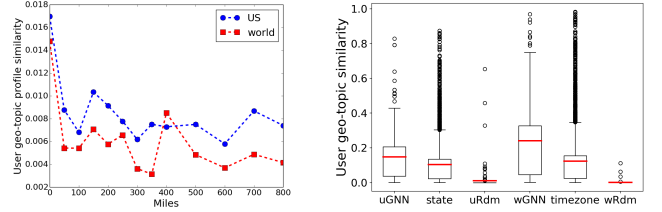


Figure 5: Left: average users’ geo-topic profile similarities with respect to the distance between users. Right: boxplots for US GNN, US state, US random, world GNN, world time-zone, world random.

three-dimensional models to higher dimensions, and as we will show in section 4.3, to maintain the validity of lemma 1.

For both layers of user’s rate parameter β_u and $\beta_u^{(2)}$, we place uninformative gamma priors on them: $\beta_u, \beta_u^{(2)} \sim \text{Gamma}(\epsilon_0, \epsilon_0)$. These user-dependent rate parameters are used to generate user’s latent factor θ_{uk} and θ_{uk_2} to help explain the different degree of overall popularity for each user. Note that we can easily generalize the model to three or more (deep) layers by continuously factorizing the second user’s latent factor θ_{uk_2} .

4.2 Modeling User’s Contextual Information

Next, to alleviate the sparsity issue, we investigate two of the most common user contexts in social-spatial networks — geo-location and social ties by first exploring how these two factors correlate with user’s geo-topic profiles, then followed by a contextualized Bayesian hierarchical model to learn better user representation.

4.2.1 User’s Geo-location Context. To study the influence of user’s geo-location, we need to define how to measure the similarity between users’ geo-topic profiles. To that end, we treat each user’s geo-topic profile \mathcal{P}_u as a distribution of counts indexed by tag and location, and adopt Jensen-Shannon (JS) divergence as the distance measure due to its symmetric and bounded property. Thus, the similarity between two user’s geo-topic profile \mathcal{P}_{u_a} and \mathcal{P}_{u_b} is defined as $1 - \text{JS}(\mathcal{P}_{u_a}, \mathcal{P}_{u_b})$ where $\text{JS}(\mathcal{P}_{u_a}, \mathcal{P}_{u_b}) = \frac{1}{2} \text{KL}(\mathcal{P}_{u_a} || M) + \frac{1}{2} \text{KL}(\mathcal{P}_{u_b} || M)$. M is the average of \mathcal{P}_{u_a} and \mathcal{P}_{u_b} and $\text{KL}(\cdot)$ represents Kullback-Leibler divergence. The similarity ranges from 0 to 1, with a large value indicating similar distributions of two users’ geo-topic profiles.

We first show how geo-topic profile similarity between users changes with respect to the geographical distance in Figure 5. As we

Table 1: Pearson correlation coefficient between user’s profile similarity and their social similarity.

Location	Geo-topic profile	Topical profile	Location profile
US	0.293	0.332	0.142
world	0.320	0.313	0.237

can see, for both users in US and the world, the average similarity is the highest when users are close to each other, specifically, when they are within about 50 miles. It then drastically decreases as the distance increases. This observation aligns with Tobler’s first law [31] of geography which states that “near things are more related than distant things”. In light of this, we next explore how geo-topic similarity between geographically nearest neighbors (GNN) compares against that of two random users. Figure 5 also shows the boxplots of geo-topic similarities for a group of GNN in US/the world (in terms of five neighbors) against a group of random users in US/the world. As we can see, the similarities of GNN for both datasets are statistically much higher than random users, suggesting that GNN are more likely to share similar geo-topic profiles with each other. Furthermore, we also obtain the state for each user in US and timezone for each user in the world through reverse geo-coding. We observe in Figure 5 that users in the same state or timezone have statistically higher similarities than random users, indicating that the state and timezone could also be useful geo-spatial features for finding users with similar geo-topic profiles.

4.2.2 User’s Social Context. Due to the homophily effect [25], socially connected users often share similar interests with each other. Indeed, there have been many works [19, 23] that exploit social ties in the applications such as recommendation and search. Here, we explore how social ties in terms of a user following others may affect geo-topic profiles. To that end, we first define users’ social similarity with their corresponding followers. Let \mathcal{F}_a and \mathcal{F}_b denote the set of users following user u_a and u_b . We define two user’s social similarity as the cosine similarity of \mathcal{F}_a and \mathcal{F}_b : $||\mathcal{F}_a \cap \mathcal{F}_b|| / (||\mathcal{F}_a|| ||\mathcal{F}_b||)$, where each user is treated as a binary vector indexed by each follower $f \in \mathcal{F}$. Thus, a large similarity of two users indicates that they are often followed by the same followers. We then examine the Pearson correlation coefficient between users’ geo-topic profile similarity and their social similarity in Table 1. As we can see, for both users in US and the world, there exists weak positive relationship between user’s geo-topic profile similarity and their social similarity. This indicates that users tend to share similar geo-topic profile, i.e., known for the same tags at the same locations, if they are followed by a similar set of users. Furthermore, we also show Pearson correlation between user’s social similarity and user’s topical and location profile, as previously defined in the Section 3. As we can see, user’s social similarity has a weak positive relationship with user’s topical profile for both users in US and the world. It, however, does not share similar degree of positive relationship with user’s location profile. An intuitive explanation is that users often follow each other for similar topical interest, as the social homophily indicates, and that social ties have more topical influence than geo-spatial influence.

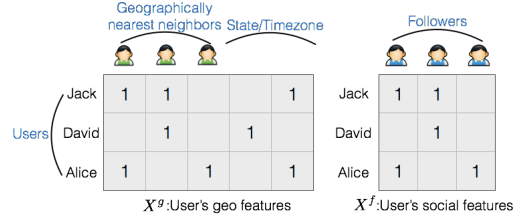


Figure 6: Constructed user’s geo and social feature matrix.

4.2.3 Modeling User’s Contextual Information. Previous analysis shows that user’s geo-location and social information has corresponding influence over user’s geo-topic profile. One immediate question is that how we can integrate these contexts into the model for better representation of users. In this section, we propose an extension to the previous two-layered user factorization model.

Let X denote the discrete matrix of user’s contextual features, where each element is a binary indicating whether a user has a corresponding feature. Specifically, we use $X^g \in \mathbb{Z}^{U \times G}$ and $X^f \in \mathbb{Z}^{U \times F}$ to denote user’s geo-spatial and social features, as illustrated in Figure 6. For user’s geo-spatial features, we empirically select five GNN for each user and their corresponding states (US) or timezone (world), with each column of X^g representing a feature. For social features, each column of X^f represents one of F followers. To integrate X^g and X^f , we propose to generate the first layer of user latent factor θ_{uk} on her corresponding contextual information. Specifically, we linearly add feature’s latent factor to the shape parameter of θ_{uk} , and let the inference process automatically learn feature’s latent factor and its contribution to user’s latent factor. Thus, our contextual user factorization goes as follows,

$$\theta_{uk} \sim \text{Gamma}(\theta_{uk_2} \Omega + \sum_{g=1}^G X_{ug} \lambda_{gk} + \sum_{f=1}^F X_{uf} \lambda_{fk}, \beta_u) \quad (3)$$

where λ_{gk} and λ_{fk} represents the latent factor of geo context g and social context f , and are placed with uninformative Gamma distribution as priors $\lambda_{gk}, \lambda_{fk} \sim \text{Gamma}(\epsilon_0, \epsilon_0)$. All the rest latent factors are kept the same as in Equation 2. Since the first layer user latent factor is also conditioned upon user’s contexts, θ_{uk} is expected to not only reflect user’s preference through geo-topic popularity counts but also user’s contextual information. This formulation is especially useful for “cold-start” users who have very few counts in their geo-topic profiles. Our overall Bayesian contextual hierarchical user factorization (**bHUF+**) is shown in Figure 4.

4.3 Inference

Since our model makes use of Gamma shape parameters for the construction of the multi-layer factorization scheme on tensor counts, it is not tractable to compute the exact posterior of the latent factors due to its non-conjugacy. However, we exploit the recently proposed data augmentation scheme for NB distribution [39], and develop an efficient closed-form Gibbs sampling formula with augmented distribution. Section B in the appendix explains the derived Gibbs sampling scheme in further details.

Table 2: Twitter Datasets.

Dataset	# of users	# of tags	# of labelers	# of taggings
US	6,709	240	148,623	756,771
World	11,103	304	327,313	1,506,355

To see how tensor-based multi-layered user factorization can overcome data overdispersion theoretically, we obtain and compare the variance-to-mean ratio given by one-layered (BPTF) and two-layered (bHUF) user factorization, respectively. Specifically, the following Lemma is reached showing that a two-layered construction of θ_{uk} is more capable of modeling overdispersed counts by allowing a larger variance-to-mean ratio on the data. The proof is given in Section A of the appendix.

Lemma 1. The variance-to-mean ratio of two-layered factorization of θ_{uk} to θ_{uk_2} in Equation 2 is larger than that of one-layered θ_{uk} by a factor of $1 + \frac{1}{(1+\beta_u)\beta_u^{(2)}}$.

5 EXPERIMENTAL EVALUATION

In this section, we conduct several experiments for evaluation. Specifically, we seek to answer the following questions: (i) How well does bHUF perform against alternative baselines? (ii) Do user’s contexts help improve upon bHUF? (iii) How does bHUF and bHUF+ perform in two subproblems — predicting user’s topical and geo profile? (iv) What impact do some important parameters have?

Data. We use a sample of about 12 million geo-tagged Twitter lists collected from 2013 to 2014. Twitter lists [4, 11, 27] are crowd-generated lists where a labeler can choose to place a user if the labeler thinks the user is known for the topic indicated by the list tags. Thus, a user is considered to be widely known for a topic if the user is labeled by many labelers with the same tag. In our experiments, we filter out infrequent tags which appear in less than 50 lists to focus on quality tags. All users and labelers’ geo-coordinates (latitude/longitude) are determined through their geo-tagged tweets by repetitively dividing and selecting the grid which contains the most geo-tagged tweets as in [9]. We randomly sample two datasets, show in Table 2: one is bounded in US in which users and labelers are all in US, and the other is across the whole world. Note that with these datasets, a geo-topic token $\{u \xleftarrow{t} v, l\}$ is formulated by treating a user as u , a labeler as v , the labeler’s geo-coordinates as l and the list tag as t .

Experimental Setup. To evaluate the performance of our proposed bHUF and bHUF+ for inferring and ranking user’s geo-topic profiles, we randomly select 50% of non-zero popularity counts for each user for training and 50% of non-zero popularity for testing. We use the same data split ratio for all other experiments. In our experiments, the latent dimension K is selected through cross-validation from $\{50, 100, 150, 200, 250, 300\}$ to be 200 since the performance of all baselines plateaus around 200. For K_2 , we empirically set it to 50 and examine its impact in later experiments. The hyperparameter ϵ_0 is empirically set to 0.1 to encourage sparse solutions. For Gibbs sampling, we use a burn-in period of 1000 iterations, and then use another 1000 iterations to collect every tenth sample for all models. All latent parameters are randomly initialized.

For evaluation metrics, we adopt precision and recall to measure the ranking performance between predictions and the hold-out ground truth. Specifically, let $GroundTruth(u_i, n)$ denote the set of top n ranked tag-location combination (t, l) according to y out of the entire non-zero counts $GroundTruth(u_i)$ for user u_i in the testing set, and let $Pred(u_i, n)$ denote the set of top n ranked tag-location combination (t, l) according to model’s predictions for user u_i , then, precision $P@n$ is defined as $\frac{1}{N} \sum_{i=1}^N \frac{|Pred(u_i, n) \cap GroundTruth(u_i, n)|}{n}$ and recall $R@n$ is defined as $\frac{1}{N} \sum_{i=1}^N \frac{|Pred(u_i, n) \cap GroundTruth(u_i)|}{|GroundTruth(u_i)|}$. Thus, we can see that $P@n$ represents the fraction of correctly identified tag-location combinations from the top n predictions, and $R@n$ represents the fraction of true non-zero tag-location combinations that are ranked in the top n predictions.

Baselines. We use the following baselines for comparison:

- Pairwise Interaction Tensor Factorization (**PITF**) [28]. This is a pairwise tensor factorization model which captures the pairwise interactions between entities.
- CP Decomposition with gradient-based optimization (**CP-OPT**) [1]. This method is based on minimizing least square between CP and the observations with gradient descent. However, a direct application with CP-OPT does not perform that well due to the power law distribution of users’ geo-topic profiles. Therefore, we first transform the counts y with $\log(y + 1)$ and then apply CP-OPT.
- CP Alternating Poisson Regression (**CP-APR**) [10]. This method is based on maximizing the likelihood of Poisson distribution on the observed users’ geo-topic profiles, and is equivalent to minimizing Kullback-Leibler (KL) divergence.
- Beta-Negative Binomial CP decomposition (**BNBCP**) [17]. This Bayesian Poisson CP decomposition is based on a beta-negative binomial construction with each entity’s latent factor generated by Dirichlet distribution, and is a one-layered factorization model.
- Bayesian Poisson Tensor Factorization (**BPTF**) [29]. This is the state-of-art Bayesian Poisson tensor model with each entity’s latent factor generated through Gamma distribution, and is essentially a one-layered model.
- Bayesian Hierarchical User Factorization (**bHUF**). We use **bHUF-1** and **bHUF-2** to represent one-layered and two-layered model.
- Two-layered bHUF integrated with only user’s geo-location context (**g-bHUF-2**) and only user’s social context (**s-bHUF-2**).

5.1 Comparison with Baselines

We first show the overall performance comparison in Table 3, with all results averaged over 10 runs for Gibbs sampling. The grid size in this experiment is set to $2.5^\circ \times 2.5^\circ$ (equivalent to approximately 175 miles \times 130 miles at latitude 40°) for both datasets. Other choices of grid size are examined in later experiments. As we can see, overall, the proposed bHUF+ gives the best performance among all methods in terms of precision and recall. This confirms the superiority of the two-layered user factorization integrated with user’s geo-location and social contexts in modeling users’ over-dispersed and sparse geo-topic profiles.

Comparison without user’s context. From Table 3, we observe that two-layered user factorization bHUF-2 generally gives better performance than the single-layered Bayesian Poisson models. Specifically, it gives an average improvement of 7.06% and 13.4%

Table 3: Overall performance comparison. ‘†’ and ‘*’ marks statistically significant improvement over the best one-layered baseline and over bHUF-2, respectively. Both are evaluated according to two sample t -test at significance level 0.05.

Method	US					world				
	$P@10$	$P@20$	$R@10$	$R@20$	Imp. (P/R)	$P@10$	$P@20$	$R@10$	$R@20$	Imp. (P/R)
PITF	0.0801	0.0876	0.0582	0.0863	N/A	0.1019	0.1006	0.0832	0.1258	N/A
CP-OPT	0.1065	0.1084	0.1021	0.1372	N/A	0.1270	0.1255	0.1418	0.1814	N/A
CP-APR	0.1333	0.1358	0.1298	0.1674	N/A	0.1465	0.1447	0.1666	0.2159	N/A
BNBCP	0.1419	0.1447	0.1309	0.1691	N/A	0.1604	0.1578	0.1785	0.2218	N/A
BPTF	0.1409	0.1442	0.1300	0.1679	N/A	0.1627	0.1587	0.1788	0.2245	N/A
bHUF-1	0.1420	0.1455	0.1306	0.1709	N/A	0.1534	0.1526	0.1628	0.2157	N/A
bHUF-2	0.1524	0.1554	0.1483	0.1940	7.06%†/13.4%†	0.1703	0.1664	0.1912	0.2462	4.76%†/8.30%†
g-bHUF-2	0.1630	0.1664	0.1590	0.2034	7.01%*/6.03%*	0.1787	0.1742	0.1979	0.2558	4.80%*/3.70%*
s-bHUF-2	0.1599	0.1625	0.1578	0.2017	4.74%*/5.18%*	0.1763	0.1730	0.1964	0.2547	3.74%*/3.08%*
bHUF+	0.1708	0.1691	0.1667	0.2134	10.4%*/10.7%*	0.1834	0.1791	0.2022	0.2636	7.66%*/6.41%*

for precision and recall over the best one-layered model for the US dataset, and an average improvement of 4.76% and 8.30% for precision and recall for the world dataset. Additionally, we also observe that methods modeled with Poisson distribution (including the non-Bayesian model CP-APR) perform significantly better than CP-OPT, and that Bayesian Poisson models (BNBCP, BPTF, bHUF-1) perform better than the non-Bayesian version (CP-APR). This confirms that (i) Two-layered user factorization has more expressive modeling power over the single-layered models; (ii) User’s popularity count values are better modeled with Poisson distribution with its latent parameters regularized by proper priors.

Moreover, by examining the inferred user factors θ_{uk} and θ_{uk_2} , we obtain that the variance for θ_{uk} and θ_{uk_2} is 26.83 and 0.804, indicating that the second-layer user factor has much less variance. This indicates a more uniform distribution of the elements of θ_{uk_2} compared to the first-layer user factor and that using two layers of user factorization can already explain the variance quite well.

Furthermore, to show that bHUF-2 can roughly capture user’s geo-topic profile, we examine the top-ranking tags and locations for user @MMFlint. We first compute the ranking of topical tags for @MMFlint according to the inner product of θ_{uk} and all tag factors, with the top five tags shown as: politics, entertainment, movie, tv and art. We can see that these topical tags can roughly capture @MMFlint’s topical interest and expertise. For the tag politics, we examine the top ranking regions for @MMFlint, with the top five grid center coordinates give as: (40.319°, -74.638°), (41.931°, -71.238°), (37.096°, -122.243°), (38.707°, -76.339°) and (33.872°, -118.843°). Each location is close to New York, Boston, San Jose, Washington D.C. and Los Angeles, respectively, which generally agrees with the heat map in Figure 2. Note that these coordinates are not exactly in these cities due to grid granularity.

Comparison with user’s context. From Table 3, we can observe that bHUF+ enhanced with user’s contexts shows significant improvement over models without them. Specifically, the model integrated with user’s geo-location g-bHUF-2 gives an average improvement of 5.90% for precision and 4.86% for recall over bHUF-2, while s-bHUF-2 gives an average improvement of 4.24% and 4.13% for precision and recall over bHUF-2. This confirms that (i) user’s context, specifically, user’s geo-location and social ties, can be used

to improve predicting user’s geo-topic profile; and (ii) Equation 3 — conditioning the first-layer user factor on a linear combination of the second-layer user factor and her corresponding contexts — can effectively learn the impact of contextual factors. Furthermore, the integration with both contexts gives better performance than either of them, suggesting that user’s geo-location is complementary to the social context in terms of predicting one’s geo-topic profile. These observations confirm that the sparsity issue is able to be mitigated by integrating user’s relevant contextual information.

5.2 Predicting User’s Topical and Geo Profiles

Very often, we are interested in knowing only user’s topic profile — what a user is known for — or user’s geo profile — where this user is known for. Both problems can be considered as lower dimensional subproblems of predicting user’s geo-topic profile by aggregating quadruplets $\{u, t, l, y\} \in \mathcal{P}_u$ with respect to locations/tags. In this section, we show how modeling a more generalized user geo-topic profile can improve on either two subproblems. Specifically, we use user, tag and location latent factor θ_{uk} , ϕ_{tk} and ψ_{lk} inferred by three-dimensional Bayesian Poisson models to obtain user-tag and user-location score by computing the inner product of θ_{uk} and ϕ_{tk} , and of θ_{uk} and ψ_{lk} . To compare with a two-dimensional model, we use BPF [12] for modeling user-tag counts and user-location counts. All models are evaluated with $P@5$ and $R@5$.

Predicting user’s topical profile. We show the performance comparison in Figure 7. As we can see, three dimensional models by considering user’s geo-topic profile generally give much better performance. This indicates that by distinguishing user’s topical profile with geo-space, we can better predict user’s topical profile. Among all methods, bHUF+ gives the best performance overall, while bHUF-2 outperforms the one-layered models, which agrees with the previous observation for predicting user’s geo-topic profile. Furthermore, we can see that models enhanced with user’s contextual information significantly outperforms those without it, confirming the importance of user’s contextual information. Specifically, s-bHUF-2 gives better performance than g-bHUF-2, indicating that social context in terms of user’s followers is more informative than user’s geo-location on learning user’s topical profile, which agrees with the correlation analysis in Table 1.

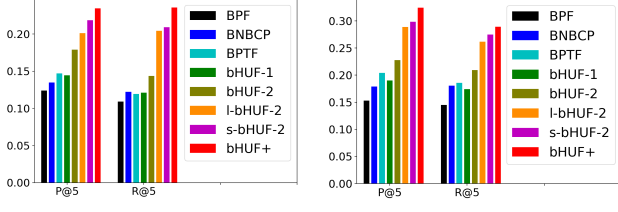


Figure 7: Precision and recall for predicting user’s topical profile. Left: US. Right: world.

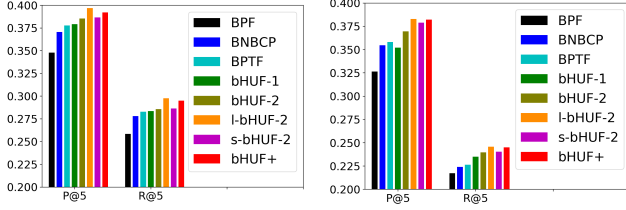


Figure 8: Precision and recall for predicting user’s location profile. Left: US. Right: world.

Predicting user’s geo profile. We also show performance comparison in Figure 8. Similarly with user topical profile prediction, three dimensional methods generally outperform BPF, and bHUF-2 performs better than one-layered models. Different from user topical profile prediction, g-bHUF-2 gives the best performance, while s-bHUF-2 does not show as much improvement over bHUF-2. This indicates that when predicting where a user is known by others, user’s geo-location is more informative than user’s social context.

Together with the previous results, we conclude that user’s geo-location and social context improves on different aspects of user geo-topic profile, and that one should select the appropriate context to improve performance depending on the problem on hand. Note that the improvement among all methods is generally not as significant as that of user topical profile prediction. One possible hypothesis could be that user’s geo profile has less inherent structure to learn from than user’s topical profile, thus making it less likely to be influenced by model improvements.

5.3 Parameter Analysis

Impact of grid size. An important parameter in this study is grid size. A larger grid size, indicating a larger geographical area for each location, results in a smaller total number of locations in our models. Thus, if the grid is large, user geo-topic profile is coarse-grained with respect to locations, while fine-grained when it is small. Thus, it is important to see how the models perform with respect to grid size.

To that purpose, we run our models with respect to four different sizes of grid (latitude * longitude): $0.5^\circ \times 0.5^\circ$, $0.8^\circ \times 0.8^\circ$, $1.2^\circ \times 1.2^\circ$ and $2.5^\circ \times 2.5^\circ$. For each size, we compare precision and recall for BPTF, bHUF-2, bHUF+ in Figure 9. As we can observe, for the US dataset, the precision first increases when the grid size is small,

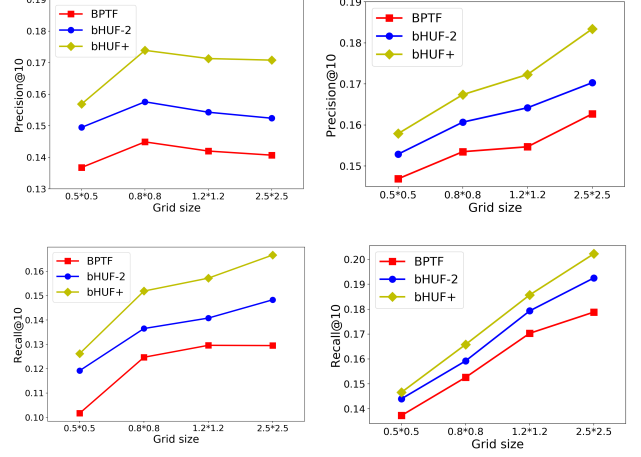


Figure 9: Performance comparison with respect to the size of grid. Left: US. Right: world.

and starts to decrease passing $0.8^\circ \times 0.8^\circ$. This indicates that for US dataset, a grid size of about $0.8^\circ \times 0.8^\circ$ could be the best geo-space partitioning for describing and modeling user’s geo-topic profile. A larger grid may combine geo-topic tokens at two different locations into one grid; while a smaller grid may over-partition a location, thus causing sparsity issue. For precision on the world dataset, we can see that it gives best performance when the grid is $2.5^\circ \times 2.5^\circ$ (possibly larger), and that it always decreases when the grid gets smaller. This indicates that a coarse-grained geo-partitioning is better suited for the world dataset. As for R@10, all performance increases as the grid gets larger. This could be explained by the fact that as grid gets larger, there will be less number of grids with non-zero counts to be retrieved, making the denominator in recall definition smaller. Furthermore, no matter how we choose the grid size, bHUF-2 consistently performs better than BPTF; and that bHUF+, integrated with user’s geo-location and social contexts, also consistently outperforms bHUF-2. This indicates that our proposed models are robust with respect to grid size, and that one could choose an appropriate grid size to her needs without degrading the relative performance.

Impact of K_2 . How does the second-layer user latent dimension size K_2 affect the performance? To that end, we run bHUF-2 for US dataset with different values of K_2 ranging from 10 to 200, and show the boxplots of $P@10$ and $R@10$ in Figure 10. As we can see, as we increase the value of K_2 , both performance increases until it saturates at about 40 for $P@10$ and $R@100$. We obtain similar trends for the world dataset. We use 50 for K_2 for all previous experiments for tradeoff between performance and efficiency. Furthermore, we can see that even when K_2 is as small as 10, it still outperforms the one-layered user models, indicating the superiority of an extra layer of factorization.

6 CONCLUSION

In this paper, we introduced multi-dimensional user geo-topic profile to capture the pair-wise interactions between geo-location and

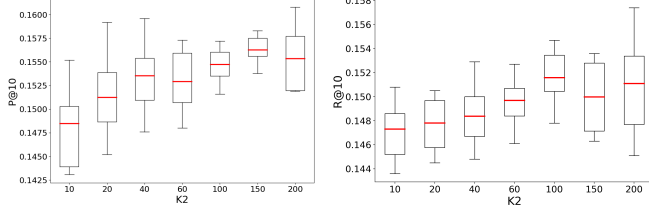


Figure 10: Precision@10 and recall@10 with respect to K_2 .

user topical profile in social-spatial networks. To overcome overdispersion and user heterogeneity in geo-topic profiles, we proposed a two-layered multi-dimensional Bayesian hierarchical user factorization framework. To alleviate the sparsity issue, we studied how user's context — user's geo-location and social ties — correlates with geo-topic profile, and proposed an enhanced model to integrate user's contexts. Through Twitter-based social-spatial datasets, we find bHUF leads to a 5~13% improvement in precision and recall and an additional 6~11% improvement with user's contexts.

ACKNOWLEDGMENTS

This work was supported in part by NSF grant IIS-1149383.

A PROOF OF LEMMA 1

From Equation 4 and Poisson-multinomial equivalence, we have $y_{uk} \sim \text{Pois}(\theta_{uk} \sum_{t=1}^T \sum_{l=1}^L \phi_{tk} \psi_{lk})$. Since $\sum_{t=1}^T \phi_{tk} = 1$ and $\sum_{l=1}^L \psi_{lk} = 1$, we have $y_{uk} \sim \text{Pois}(\theta_{uk})$. We then consider two cases: the first is that θ_{uk} is without further factorization, and the second is that θ_{uk} is further factorized to θ_{uk_2} as in bHUF. For the first case, let $\theta_{uk} \sim \text{Gamma}(\epsilon_0, \beta_u)$. By integrating out θ_{uk} , we have $y_{uk} \sim \text{NB}(\epsilon_0, \gamma_u)$, where $\gamma_u = \frac{1}{1+\beta_u}$. Thus the variance-to-mean ratio of y_{uk} is given by $\frac{1}{1-\gamma_u}$. For case 2, from Equation 5, we have $y_{u \cdot k_2} \sim \text{NB}(\theta_{uk_2} \sum_{k=1}^K \Omega_{k_2k}, \gamma_u)$. Since $\sum_{k=1}^K \Omega_{k_2k} = 1$, we have $y_{u \cdot k_2} \sim \text{NB}(\theta_{uk_2}, \gamma_u)$. Since $\theta_{uk_2} \sim \text{Gamma}(\epsilon_0, \beta_u^{(2)})$, by the law of total expectation and law of total variance, we obtain the expectation $\mathbb{E}(y_{u \cdot k_2}) = \frac{\gamma_u}{1-\gamma_u} \frac{\epsilon_0}{\beta_u^{(2)}}$ and the variance $\text{Var}(y_{u \cdot k_2}) = \frac{\gamma_u}{(1-\gamma_u)^2} \frac{\epsilon_0}{\beta_u^{(2)}} (1 + \frac{\gamma_u}{\beta_u^{(2)}})$. Thus the variance-to-mean ratio of case 2 is $\frac{1}{1-\gamma_u} (1 + \frac{\gamma_u}{\beta_u^{(2)}})$. We then reach the conclusion that the variance-to-mean ratio for case 2 is larger than case 1 by a factor of $1 + \frac{\gamma_u}{\beta_u^{(2)}}$.

B INFERENCE

We first present two lemmas [39] frequently used in the derivation.

Lemma 2. Suppose that y_1, \dots, y_K are independent random variables with $y_k \sim \text{Pois}(\theta_k)$ and $y = \sum_{k=1}^K y_k$. Set $\theta = \sum_{k=1}^K \theta_k$. The Poisson-multinomial equivalence states that $(y_1, \dots, y_K) \sim \text{Mult}(y; \frac{\theta_1}{\theta}, \dots, \frac{\theta_K}{\theta})$ and $y \sim \text{Pois}(\theta)$.

Lemma 3. Suppose $y \sim \text{NB}(r, p)$ and $c \sim \text{CRT}(y, r)$ is a Chinese restaurant table distribution. Then y and c can also be jointly distributed as $y \sim \text{SumLog}(c, p)$ and $c \sim \text{Pois}(-r \ln(1-p))$, where SumLog is defined as $y = \sum_{i=1}^c x_i$ and $x_i \sim \text{Log}(p)$ is logarithmic-distributed random variable.

Then, given a tensor count y_{utl} , we first reparametrize it as $y_{utl} = \sum_{k=1}^K y_{utlk}$, where each y_{utlk} is a latent count and has a Poisson distribution, $y_{utlk} \sim \text{Pois}(\theta_{uk} \phi_{tk} \psi_{lk})$. Then by the Poisson-multinomial equivalence, y_{utlk} can be sampled as

$$\{y_{utlk}\} \sim \text{Mult}(y_{utl}; \frac{\theta_{uk} \phi_{tk} \psi_{lk}}{\sum_{k=1}^K \theta_{uk} \phi_{tk} \psi_{lk}}) \quad (4)$$

With the latent count sampled, we define several collective aggregates that are used in sampling as follows:

$$y_{uk \cdot} = \sum_{t=1}^T \sum_{l=1}^L y_{utlk}, \quad y_{t \cdot k}^t = \sum_{u=1}^U \sum_{l=1}^L y_{utlk}, \quad y_{l \cdot k}^l = \sum_{u=1}^U \sum_{t=1}^T y_{utlk}$$

We sample each latent factor with Gibbs sampling as follows,

Sampling ϕ_{tk} . Due to the Dirichlet and Multinomial conjugacy, each column of $\phi_{\cdot k}$ has a posterior Dirichlet distribution, which can be sampled with

$$\phi_{\cdot k} \sim \text{Dir}(y_{1k}^t + \eta^t, y_{2k}^t + \eta^t, \dots, y_{Kk}^t + \eta^t)$$

Sampling ψ_{lk} . Similarly with ϕ_{tk} , each column of $\psi_{\cdot k}$ also has the posterior Dirichlet distribution with

$$\psi_{\cdot k} \sim \text{Dir}(y_{1k}^l + \eta^l, y_{2k}^l + \eta^l, \dots, y_{Kk}^l + \eta^l)$$

Sampling θ_{uk} . Using Poisson-Gamma conjugacy, the posterior of θ_{uk} is still a Gamma distribution and can be sampled with,

$$\theta_{uk} \sim \text{Gamma}(y_{uk \cdot} + \theta_{uk_2} \Omega + \sum_{g=1}^G X_{ug} \lambda_{gk} + \sum_{f=1}^F X_{uf} \lambda_{fk}, 1 + \beta_u)$$

Since the row of $\phi_{t \cdot}$ and $\psi_{l \cdot}$ are Dirichlet distributed with the L_1 norm 1, the posterior rate parameter of θ_{uk} is simplified to $1 + \beta_u$. Thus, we can see that Dirichlet distributed ϕ_{tk} and ψ_{lk} do not interact with θ_{uk} directly, which provides nice isolation between latent factors of users and other entities (tags and locations).

Sampling Ω . Using Poisson distribution's additivity on $y_{utlk} \sim \text{Pois}(\theta_{uk} \phi_{tk} \psi_{lk})$ for each t and l gives us $y_{uk \cdot} \sim \text{Pois}(\theta_{uk})$. Let $\theta_{uk} = \sum_i \theta_{uki}$, where θ_{uki} are latent parameters defined as

$$\theta_{uki} = \begin{cases} \theta_{k_2} \Omega_{k_2k}, & 1 \leq i \leq K_2 \\ X_{ug} \lambda_{gk}, & K_2 + 1 \leq i \leq K_2 + G \\ X_{uf} \lambda_{fk}, & K_2 + G + 1 \leq i \leq K_2 + G + F \end{cases}$$

Thus, by the Poisson-multinomial equivalence, we can sample corresponding y_{uki} with $\{y_{uki}\} \sim \text{Mult}(y_{uk \cdot}; \theta_{uki})$. For notation convenience, we let $y_{ukk_2} = y_{uki}$ for $1 \leq i \leq K_2$, $y_{ukg} = y_{uki}$ for $K_2 + 1 \leq i \leq K_2 + G$ and $y_{ukf} = y_{uki}$ for the rest of i . Then, for $1 \leq i \leq K_2$, we have $y_{ukk_2} \sim \text{Pois}(\theta_{uki})$ and $\theta_{uki} \sim \text{Gamma}(\theta_{k_2} \Omega_{k_2k}, \beta_u)$ due to infinite divisibility of Gamma distribution. We then integrate out θ_{uki} , and have

$$y_{ukk_2} \sim \text{NB}(\theta_{k_2} \Omega_{k_2k}, \gamma_u) \quad (5)$$

where $\gamma_u = \frac{1}{1+\beta_u}$. Using data augmentation for NB distribution in Lemma 3, we can sample c indexed by u, k and k_2 with,

$$c_{ukk_2} \sim \text{CRT}(y_{ukk_2}, \theta_{k_2} \Omega_{k_2k}) \quad (6)$$

Then, given c_{ukk_2} , each row of $\Omega_{k_2 \cdot}$ has a posterior Dirichlet distribution which can be sampled with,

$$\Omega_{k_2 \cdot} \sim \text{Dir}(\sum_{u=1}^U c_{u1k_2} + \eta_0, \dots, \sum_{u=1}^U c_{uKk_2} + \eta_0)$$

Sampling θ_{uk_2} . With Equation 5 and 6 and using the data augmentation in Lemma 3 again, the posterior θ_{uk_2} has a Gamma distribution, and can be sampled with

$$\theta_{uk_2} \sim \text{Gamma}(\sum_{k=1}^K c_{uk_2} + \epsilon_0, \beta_u^{(2)} - \ln(1 - \gamma_u))$$

Sampling λ_{gk} . For each user who have the geographical feature g , i.e., $X_{ug} = 1$, we have $y_{ukg} \sim \text{Pois}(\theta_{uki})$ for $K_2 + 1 \leq i \leq K_2 + G$, and $\theta_{uki} \sim \text{Gamma}(\lambda_{gk}, \beta_u)$ due to infinite divisibility of Gamma distribution. Integrating out θ_{uki} , we have $y_{ukg} \sim \text{NB}(\lambda_{gk}, \gamma_u)$. By using data augmentation in Lemma 3, we can sample c indexed by u, k and g with $c_{ukg} \sim \text{CRT}(y_{ukg}, \lambda_{gk})$. Given c_{ukg} , λ_{gk} has a posterior Gamma distribution which can be sampled with

$$\lambda_{gk} \sim \text{Gamma}(\sum_{u \in \mathcal{U}(g)} c_{ukg} + \epsilon_0, \epsilon_0 - \sum_{u \in \mathcal{U}(g)} \ln(1 - \gamma_u))$$

where $\mathcal{U}(g)$ represents the set of users who have feature g .

Sampling λ_{fk} . For each user who have follower f , i.e., $X_{uf} = 1$, we have $y_{ukf} \sim \text{Pois}(\theta_{uki})$ and $\theta_{uki} \sim \text{Gamma}(\lambda_{fk}, \beta_u)$ for $K_2 + G + 1 \leq i \leq K_2 + G + F$. Thus, λ_{fk} can be sampled similarly with λ_{gk} as follows,

$$\lambda_{fk} \sim \text{Gamma}(\sum_{u \in \mathcal{U}(f)} c_{ukf} + \epsilon_0, \epsilon_0 - \sum_{u \in \mathcal{U}(f)} \ln(1 - \gamma_u))$$

where $\mathcal{U}(f)$ represents the set of users who have follower f .

Sampling β_u and $\beta_u^{(2)}$. The posterior of β_u and $\beta_u^{(2)}$ are Gamma distributions due to their conjugate Gamma priors, which can be sampled with

$$\begin{aligned} \beta_u &\sim \text{Gamma}(\sum_{k_2=1}^{K_2} \theta_{uk_2} + \sum_{k=1}^K \sum_{g=1}^G X_{gk} \lambda_{gk} + \sum_{f=1}^F X_{fk} \lambda_{fk} \\ &\quad + \epsilon_0, \epsilon_0 + \sum_{k=1}^K \theta_{uk}) \\ \beta_u^{(2)} &\sim \text{Gamma}((K_2 + 1) * \epsilon_0, \epsilon_0 + \sum_{k_2=1}^{K_2} \theta_{uk_2}) \end{aligned}$$

REFERENCES

- [1] Evrim Acar, Daniel M Dunlavy, and Tamara G Kolda. 2011. A scalable optimization approach for fitting canonical tensor decompositions. *Journal of Chemometrics* (2011).
- [2] Amr Ahmed, Yucheng Low, Mohamed Aly, Vanja Josifovski, and Alexander J Smola. 2011. Scalable distributed inference of dynamic user interests for behavioral targeting. In *SIGKDD*.
- [3] Xavier Amatriain, Neal Lathia, Josep M Pujol, Haewoon Kwak, and Nuria Oliver. 2009. The wisdom of the few: a collaborative filtering approach based on expert opinions from the web. In *SIGIR*.
- [4] Parantapa Bhattacharya, Saptarshi Ghosh, Juhi Kulshrestha, Mainack Mondal, Muhammad Bilal Zafar, Niloy Ganguly, and Krishna P Gummadi. 2014. Deep twitter diving: Exploring topical groups in microblogs at scale. In *CSCW*.
- [5] Cheng Cao, Hancheng Ge, Haokai Lu, Xia Hu, and James Caverlee. 2017. What Are You Known For?: Learning User Topical Profiles with Implicit and Explicit Footprints. In *SIGIR*.
- [6] Allison JB Chaney, David M Blei, and Tina Eliassi-Rad. 2015. A probabilistic model for using social networks in personalized item recommendation. In *RecSys*.
- [7] Laurent Charlin, Rajesh Ranganath, James McInerney, and David M Blei. 2015. Dynamic poisson factorization. In *RecSys*.
- [8] Jilin Chen, Rowan Nairn, and Ed Chi. 2011. Speak little and well: recommending conversations in online social streams. In *SIGCHI*.
- [9] Zhiyuan Cheng, James Caverlee, Himanshu Barthwal, and Vandana Bachani. 2014. Who is the barbecue king of texas?: a geo-spatial approach to finding local experts on twitter. In *SIGIR*.
- [10] Eric C Chi and Tamara G Kolda. 2012. On tensors, sparsity, and nonnegative factorizations. *SIAM J. Matrix Anal. Appl.* (2012).
- [11] Saptarshi Ghosh, Naveen Sharma, Fabricio Benevenuto, Niloy Ganguly, and Krishna Gummadi. 2012. Cognos: crowdsourcing search for topic experts in microblogs. In *SIGIR*.
- [12] Prem Gopalan, Jake M. Hofman, and David M. Blei. 2015. Scalable Recommendation with Hierarchical Poisson Factorization. In *UAI*.
- [13] Prem K Gopalan, Laurent Charlin, and David Blei. 2014. Content-based recommendations with poisson factorization. In *NIPS*.
- [14] Ido Guy, Naama Zwerdling, Inbal Ronen, David Carmel, and Erel Uziel. 2010. Social media recommendation based on people and tags. In *SIGIR*.
- [15] Liangjie Hong, Amr Ahmed, Siva Gurumurthy, Alexander J Smola, and Kostas Tsoutsoulouklis. 2012. Discovering geographical topics in the twitter stream. In *WWW*.
- [16] Liangjie Hong, Aziz S Doumith, and Brian D Davison. 2013. Co-factorization machines: modeling user interests and predicting individual decisions in twitter. In *WSDM*.
- [17] Changwei Hu, Piyush Rai, Changyou Chen, Matthew Harding, and Lawrence Carin. 2015. Scalable bayesian non-negative tensor factorization for massive count data. In *ECML-PKDD*.
- [18] Longke Hu, Aixin Sun, and Yong Liu. 2014. Your neighbors affect your ratings: on geographical neighborhood influence to rating prediction. In *SIGIR*.
- [19] Meng Jiang, Peng Cui, Rui Liu, Qiang Yang, Fei Wang, Wenwu Zhu, and Shiqiang Yang. 2012. Social contextual recommendation. In *CIKM*.
- [20] Qi Liu, Enhong Chen, Hui Xiong, Chris HQ Ding, and Jian Chen. 2012. Enhancing collaborative filtering by user interest expansion via personalized ranking. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* (2012).
- [21] Xin Liu and Karl Aberer. 2013. SoCo: a social network aided context-aware recommender system. In *WWW*.
- [22] Haokai Lu and James Caverlee. 2015. Exploiting Geo-Spatial Preference for Personalized Expert Recommendation. In *RecSys*.
- [23] Hao Ma, Dengyong Zhou, Chao Liu, Michael R. Lyu, and Irwin King. 2011. Recommender Systems with Social Regularization. In *WSDM*.
- [24] Anirban Majumder and Nisheeth Shrivastava. 2013. Know your personalization: learning top level personalization in online services. In *WWW*.
- [25] Miller McPherson, Lynn Smith-Lovin, and James M Cook. 2001. Birds of a feather: Homophily in social networks. *Annual review of sociology* (2001).
- [26] Feng Qiu and Junghoo Cho. 2006. Automatic identification of user interest for personalized search. In *WWW*.
- [27] Vineeth Rakesh, Dilpreet Singh, Bhanukiran Vinzamuri, and Chandan K Reddy. 2014. Personalized Recommendation of Twitter Lists using Content and Network Information. In *ICWSM*.
- [28] Steffen Rendle and Lars Schmidt-Thieme. 2010. Pairwise interaction tensor factorization for personalized tag recommendation. In *WSDM*.
- [29] Aaron Schein, John Paisley, David M Blei, and Hanna Wallach. 2015. Bayesian poisson tensor factorization for inferring multilateral relations from sparse dyadic event counts. In *SIGKDD*.
- [30] Ahu Sieg, Bamshad Mobasher, and Robin Burke. 2007. Web search personalization with ontological user profiles. In *CIKM*.
- [31] Waldo R Tobler. 1970. A computer movie simulating urban growth in the Detroit region. *Economic geography* (1970).
- [32] Chong Wang and David M Blei. 2011. Collaborative topic modeling for recommending scientific articles. In *SIGKDD*.
- [33] Jianshu Weng, Ee-Peng Lim, Jing Jiang, and Qi He. 2010. Twitterrank: finding topic-sensitive influential twitterers. In *WSDM*.
- [34] Hongzhi Yin, Bin Cui, Ling Chen, Zhiting Hu, and Zi Huang. 2014. A temporal context-aware model for user behavior modeling in social media systems. In *SIGMOD*.
- [35] Rose Yu, Andrew Gelfand, Suju Rajan, Cyrus Shahabi, and Yan Liu. 2016. Geographic Segmentation via Latent Poisson Factor Model. In *WSDM*.
- [36] Quan Yuan, Gao Cong, Zongyang Ma, Aixin Sun, and Nadia Magnenat Thalmann. 2013. Who, where, when and what: discover spatio-temporal topics for twitter users. In *SIGKDD*.
- [37] Zhe Zhao, Zhiyuan Cheng, Lichan Hong, and Ed H Chi. 2015. Improving user topic interest profiles by behavior factorization. In *WWW*.
- [38] Erheng Zhong, Nathan Liu, Yue Shi, and Suju Rajan. 2015. Building Discriminative User Profiles for Large-scale Content Recommendation. In *SIGKDD*.
- [39] Mingyuan Zhou and Lawrence Carin. 2015. Negative binomial process count and mixture modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2015).
- [40] Mingyuan Zhou, Yulai Cong, and Bo Chen. 2015. The Poisson gamma belief network. In *NIPS*.