PERSONALIZED EXPERT RECOMMENDATION: MODELS AND ALGORITHMS

A Dissertation

by

HAOKAI LU

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	James Caverlee
Committee Members,	Frank Shipman
	Yoonsuck Choe
	Patrick Burkart
Head of Department,	Dilma Da Silva

December 2017

Major Subject: Computer Engineering

ABSTRACT

Many large-scale information sharing systems including social media systems, question-answering sites and rating and reviewing applications have been growing rapidly, allowing millions of human participants to generate and consume information on an unprecedented scale. To manage the sheer growth of information generation, there comes the need to enable personalization of information resources for users — to surface high-quality content and feeds, to provide personally relevant suggestions, and so on. A fundamental task in creating and supporting user-centered personalization systems is to build rich user profile to aid recommendation for better user experience.

Therefore, in this dissertation research, we propose models and algorithms to facilitate the creation of new crowd-powered personalized information sharing systems. Specifically, we first give a principled framework to enable personalization of resources so that information seekers can be matched with customized knowledgeable users based on their previous historical actions and contextual information; We then focus on creating rich user models that allows accurate and comprehensive modeling of user profiles for long tail users, including discovering user's known-for profile, user's opinion bias and user's geo-topic profile. In particular, this dissertation research makes two unique contributions:

First, we introduce the problem of personalized expert recommendation and propose the first principled framework for addressing this problem. To overcome the sparsity issue, we investigate the use of user's contextual information that can be exploited to build robust models of personal expertise, study how spatial preference for personally-valuable expertise varies across regions, across topics and based on different underlying social communities, and integrate these different forms of preferences into a matrix factorization-based personalized expert recommender.

Second, to support the personalized recommendation on experts, we focus on modeling and inferring user profiles in online information sharing systems. In order to tap the knowledge of most majority of users, we provide frameworks and algorithms to accurately and comprehensively create user models by discovering user's known-for profile, user's opinion bias and user's geo-topic profile, with each described shortly as follows:

— We develop a probabilistic model called Bayesian Contextual Poisson Factorization to discover what users are known for by others. Our model considers as input a small fraction of users whose known-for profiles are already known and the vast majority of users for whom we have little (or no) information, learns the implicit relationships between user?s known-for profiles and their contextual signals, and finally predict known-for profiles for those majority of users.

— We explore user's topic-sensitive opinion bias, propose a lightweight semi-supervised system called "BiasWatch" to semi-automatically infer the opinion bias of long-tail users, and demonstrate how user's opinion bias can be exploited to recommend other users with similar opinion in social networks.

— We study how a user's topical profile varies geo-spatially and how we can model a user's geo-spatial known-for profile as the last step in our dissertation for creation of rich user profile. We propose a multi-layered Bayesian hierarchical user factorization to overcome user heterogeneity and an enhanced model to alleviate the sparsity issue by integrating user contexts into the two-layered hierarchical user model for better representation of user's geo-topic preference by others.

DEDICATION

To my parents, Wenye and Emma

ACKNOWLEDGMENTS

I would like to express my thanks and gratitude to all who have helped or supported me in any way during my PhD life.

First and foremost, I am deeply indebted to my advisor, Dr. James Caverlee, for his unlimited help and support throughout my research. Back in 2012 when I lost my interest in the research of my advisor at that time, he took me in and gave me a platform and an opportunity to work with him. When I was in the hard times trying to find dissertation topics, he was always very patient, kept giving me encouragement and guided me through our countless conversations and meetings. I have benefited so much not only from his insightful research suggestions, sharp ideas and clear thinking, but also from his life lessons on how to cope with failures and rejections by seeing the bright side of each and every failure. In that sense, he is not only a great advisor professionally, but also a wonderful mentor, full of cheer and with a big heart. I will always be grateful for his mentorship and support. Without that, this work would not have been possible.

Many thanks to my lab-mates for their support and helpful discussions. Senior lab members, Zhiyuan Cheng, Krishna Kamath, Elham Khabiri, Kyumin Lee, Jeff McGee, welcomed me to the lab and helped me understand how to become a successful researcher. Thanks to Wei, Cheng, Hancheng as fellow colleagues for constructive discussions, exchange of ideas and collaboration on several projects. Thanks to my office mates over the years, Zhao Xing, Zhang Yin, Sindhuja Venkatesh, for bringing joy to the daily life. I would also like to thank many junior graduate students, Parisa Kaghazgaran, Majid Alfifi, Henry Qiu, Haiping Xue, for bringing fresh ideas and new research insights into the lab.

I would like to thank my parents for their love and support. Without them constantly being there for me, I would not have been able to finish my PhD. I also want to thank

my two big sisters for their care and support. Last but not the least, I want to thank my wife, Wenye, for being in this journey with me together no matter what, and for always believing in and enrouraging me whenever I encounter difficulties in my research. This work would be impossible without her love and support.

I would also like to thank all my committee members for their advice and comments in this dissertation research, and Dr. Ruihong Huang for taking the place of Dr Yoonsuck Choe for the dissertation defense and giving additional feedback.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee consisting of Professor James Caverlee, Professor Frank Shipman, Professor Yoonsuck Choe of the Department of Computer Science and Engineering and Professor Patrick Burkart of the Department of Communication.

All other work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by graduate research assistanship funded in part by Air Force Office of Scientific Research (AFOSR) grant FA9550-12-1-0363, National Science Foundation (NSF) grant IIS-1149383 and Google Research Award.

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

1. INTRODUCTION

1.1 Motivation

We live in an age of unprecedented access to information and knowledge enabled by many large-scale online sharing systems – from social media services (e.g., Facebook, Twitter and LinkedIn), to question-answering sites sharing knowledge and expertise (e.g., Quora, Yahoo! Answers and Stack Overflow), to online encyclopedia sharing collective knowledge (e.g., Wikipedia), to rating and reviewing applications publishing crowdsourced reviews and ratings (e.g., Yelp, Foursquare and TripAdvisor).

A key feature of these information sharing systems is their increasing leverage of the power of human crowds: as primary contributors of knowledge and content (e.g., users who answer questions in Quora and share images in Pinterest), as annotators and raters of other resources (e.g., for surfacing high-quality content), as the main channel for information propagation (as in social media and networks), and so on. For example, Facebook — launched in 2004 — has now over 2 billion monthly active users as of June 30, 2017, generating over 695,000 status updates, 79,364 wall posts and 510,040 comments every minute [1]. Quora, a question-answering website dedicated to sharing knowledge and expertise, has over 190 million monthly users since it is launched in 2009 with about 6,000 questions asked daily across 400 hundred topics [2]. Pinterest, an image-based online system, has 150 million monthly active users including 70 million users from the US, with 50 billion pins* and 1 billion boards already generated [3]. Even for the non-profit encyclopedia website Wikipedia, it now has over 5 million English articles as of July 2017 with 671 new articles added every day on average, and a total of about 30 thousand active editors contributing over 3 million edits per month [4]. This evidence showcases the massive

*A pin is an image that has been uploaded or linked from a website; a board is a collection of pins dedicated to a theme such as travel or food.

scale of human crowds powering these online information sharing systems.

Given these large and growing services, there comes the need to enable personalization of information resources for users in these services to surface high-quality content and feeds, to provide personally relevant suggestions, and so on. For example, Facebook and Twitter have adopted large-scale machine learning techniques to personalize timelines, push notifications, and suggest news feed for users [5, 6]. Quora has resorted to supervised machine learning approaches as well to rank answers [7] to show to viewers. Pinterest has also relied upon state-of-the-art machine learnings models to prioritize pins with high relevance scores and show them at the top of user's home feed [8]. These personalization systems are often focused on customizing specific items like timelines and news feed based on user profiles, while not focused on the recommendation of users and experts.

Therefore, in this dissertation, we are focused on user-centered personalization in information sharing systems. User-centered personalization is often different from item-centered personalization due to the unique nature and attributes of users, e.g., user's demographic profile, user's interests and expertise and user's topical opinion, and thus is facing different sets of challenges. In the rest of the section, we will first present some key research challenges in user-centered personalization, and then describe the main contributions of this research, followed by high-level introductions of the rest sections in the dissertation.

1.2 Research Challenges

Previously, we described the massive scale of human crowds present in many information sharing systems, and how personalization is essential to providing user friendly services and experience by creating rich user models. We now identify some of the research challenges associated with these information sharing systems.

- First, a key challenge in large-scale information sharing systems is in effectively

connecting the right person with the right resource. Many of the systems are optimized towards seeking information from popular experts with broad appeal. Consequently, users with specific information needs may not be satisfied with the answers of these general experts. For example, a user may be interested in the expert opinions of nearby local foodies, but less interested in the opinions of globally popular celebrity chefs. Bridging this *personalization gap* is the first key challenge.

- Second, many of these information sharing systems rely on a small pool of active domain experts or prominent (top-ranked) users as the content contributors or sources of expertise. Consequently, these systems are not able to take advantage of the knowledge or expertise of the vast majority of users at large (long tail users). Furthermore, those domain experts or prominent users may become inactive due to retiring or simply loss of interests. Together, these factors may prevent the growth of these systems, with their purpose for information distribution not fully realized. Thus, how to effectively discover *long tail users' profiles* is the second key challenge.

- Third, the geo-spatial variations of these information sharing systems are typically not explored or distinguished. User interest or expertise, however, is inherently constrained within certain geo-scope. For example, a user may be known for some topics only locally, while she is known for other topics more widely across the country. An information sharing system can be improved by uncovering and distinguishing such locality of user's different topical expertise. Thus, how to effectively model *user's geo-spatial profile* is the third key challenge.

1.3 Dissertation Overview

Therefore, in this dissertation research, we propose models and algorithms to facilitate the creation of new crowd-powered personalized information sharing systems. Our vision

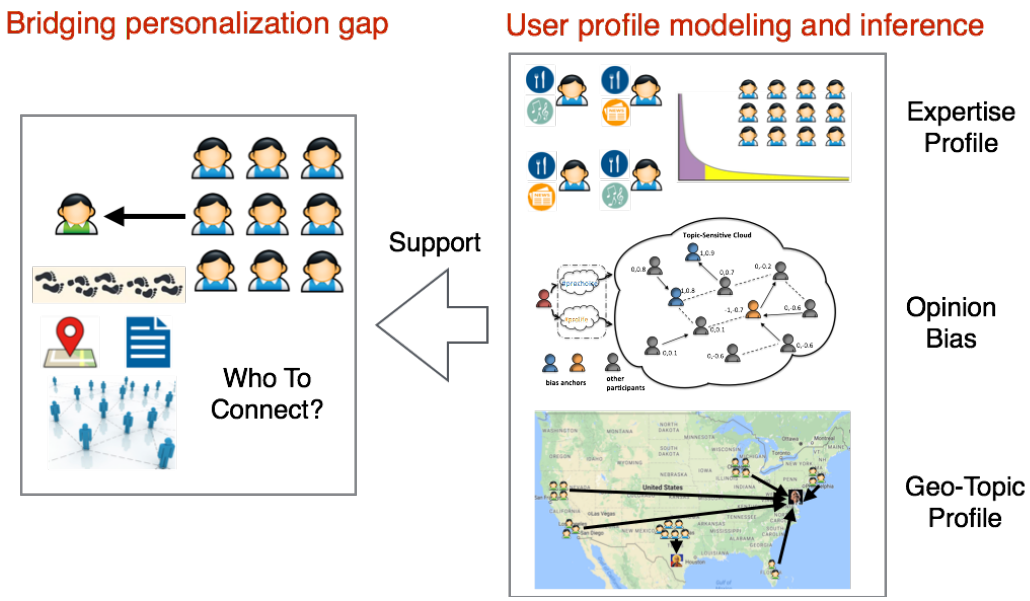


Figure 1.1: Overview of the dissertation.

is of an information sharing system where the majority of users, in addition to a few top domain experts and prominent users, are able to provide high-quality information on their knowledge and contribute to the overall knowledge reservoir. The system should be able to automatically match information seekers to knowledgeable users in a personalized fashion so that their customized information needs can be met. This user-centric information sharing not only is able to tap the knowledge of most majority of users, but also bridge the gap between general information seeking and more personalized needs. Overall, our models and algorithms are able to provide: (i) personalized user recommendation to match information seekers with customized knowledgeable users based on their previous historical actions and contextual information; (ii) accurate modeling of users' profiles for long tail users, where our models can reflect user's salient characteristics such as user's topical interest, expertise and opinion. Figure 1.1 demonstrates a high level framework of the dissertation. Specifically, this dissertation research makes two unique contributions:

- The first contribution of this dissertation research lies in formalizing the problem of personalized expert recommendation and in proposing the first principled framework for addressing this problem. A major challenge for any approach (and indeed, one that has limited prior efforts to typically focus on broad, general experts) is the extreme sparsity of user behaviors that could be leveraged for building personal recommenders. To overcome the sparsity issue, we investigate the use of contextual information that can be exploited to build robust models of personal expertise. Through a fine-grained GPS-tagged social media trace, we study how spatial preference for personally-valuable expertise varies across regions, across topics, and based on different underlying social communities, which we denote region, topic, and social-based locality, respectively. We integrate these different forms of preferences into a matrix factorization-based personalized expert recommender, and conduct extensive experiments. We conclude that our proposed model outperforms several alternative baselines. Specifically, we show that the integration of region, topic and social-based locality gives better performance over matrix factorization, respectively, and that the combination of these influence gives the best result among all, indicating that geo-spatial, topical and social factors are able to complement each other in personalized expert recommendation.

- To support and complement the personalized recommendation on experts, the second contribution of this dissertation research focuses on modeling and inferring user profiles in online information sharing systems and social media. Our previous research establishes a framework of personalized expert recommendation in which users are recommended to connect to experts who are more likely to satisfy their personalized information needs. However, many real-world systems usually rely on a small portion of active domain experts or dominant (top-ranked) users among

all participants, and are not able to take advantage of the knowledge and expertise of the long tail of users. In order to tap the knowledge of most majority of users, we provide frameworks and algorithms to accurately and comprehensively model user profile. Specifically, we have investigated and explored user models to discover user's known-for profile, user's opinion bias, and user's geo-topic profile with each described as follows:

— **Discovering user's known-for profile.** We first develop a context-based model called Bayesian Contextual Poisson Factorization to discover what users are known for by others. Our model considers as input a small fraction of users whose known-for profiles are already known and the vast majority of users for whom we have little (or no) information, learns the implicit relationships between user's known-for profiles and their contextual signals, and finally predict known-for profiles for those majority of users. Our method moves beyond just modeling the content a user generated, and naturally models and integrated additional contextual factors that provide implicit linkages between users for improved known-for profile estimation.

— **Discovering user's opinion bias.** In a different aspect of user profiling, we explore and investigate user's topic-sensitive opinion bias for richer user profiles, and demonstrated how user's opinion bias can be exploited to recommend other users with similar opinion in social media systems. Specifically, we propose a lightweight system called "BiasWatch" for (i) semi-automatically discovering and tracking bias themes associated with opposing sides of a topic; (ii) identifying strong partisans who drive the online discussion; and (iii) inferring the opinion bias of long-tail users. By taking just two hand-picked seeds to characterize the topic-space (e.g., "pro-choice" and "pro-life") as weak labels, we develop an efficient optimization-based opinion bias propagation method over social/information network. These inferred

opinion bias can be used as a different dimension to augment user's profiles.

— **Discovering user's geo-topic profile.** Finally, we have studied how a user's topical profile varies geo-spatially, how we can model a user's geo-topic profile, and the implications for personalized user recommendation in information sharing system. Generally, the quality of user's profile is often dependent on the social-spatial properties of the systems in which they arise. Specifically, our analysis on the impact of geo-location on user's topical profile indicates the pair-wise interactions between geo-locations and user's topical profile. Motivated by the observations, we propose the modeling of fine-grained *user geo-topic profiles* to capture the aforementioned pair-wise interactions. Concretely, we first propose a multi-layered Bayesian hierarchical user factorization which can learn a more expressive user model than single-layered models. We then investigate how user's contextual information, specifically, user's geo-location and social ties, correlates with one's geo-topic profile, and then propose to integrate these contexts into the two-layered hierarchical user model for better representation of user's geo-topic preference by others.

The rest of this dissertation is organized as follows:

- *Section 2:* In this section, we discuss related work on user profiling, more specifically, on user interests, expertise and opinion bias. We also discuss user context in social systems and how it can be used to improve user profiling. Finally, we discuss some related work on modeling user's implicit feedback and more generally, on modeling discrete count data.
- *Section 3:* In this section, we introduce the problem of personalized expert recommendation, and propose a principled matrix factorization based approach to effectively recommend personal experts to users. We evaluate the method against several alternative baselines and study the influence of different localities on the recommendation

performance.

From Section 4 to Section 6, we focus on modeling and inferring user profiles by discovering user's known-for profile, by discovering user's opinion bias, and by discovering user's geo-topic profile.

- *Section 4:* In this section, we introduce the problem of user known-for profile discovery to model and propose a probabilistic generative framework called Bayesian Contextual Poisson Factorization to effectively learn what a user is known for by others. We develop an efficient variational inference to learn latent parameters, evaluate the method against several baselines and study how user's contextual information affects final performance.
- *Section 5:* In this section, we build a systematic framework – BiasWatch – to discover biased themes and estimate user's opinion bias quantitatively under the context of controversial topics in social media. We propose an efficient optimization scheme to propagate opinion bias from prominent users to "regular" users. We then evaluate BiasWatch by comparing against several alternative approaches and study the effect of framework's parameters on classification performance. Finally, we demonstrate how these inferred opinion bias scores can be exploited to recommend similar-minded users.
- *Section 6:* In this section, we introduce a multi-dimensional user profiling concept called user's geo-topic profile, and propose a multi-layered Bayesian hierarchical user factorization to model where users are popular for what topic. We then investigate the impact of user's contexts on user's geo-topic profile and propose an enhanced model to integrate their influence. Finally, we evaluate our proposed models against several baseline methods and study the effect of model parameters on final predicative performance.
- *Section 7:* We conclude with a summary of the contributions of this work, and provide a discussion of two potential future research opportunities.

2. RELATED WORK

In this section, we discuss several lines of work related to our dissertation including user profiling, user context, and its application for different tasks in information systems. We also discuss some related work on modeling user's implicit feedback and more generally, on modeling discrete count data.

2.1 User Profiling

User profiling is an important task, with many efforts building user's topic interests and demographic profile for personalized search [9, 10, 11, 12], recommender systems [13, 14, 15, 16, 17], expert mining [18, 19, 20] and targeted advertising [21]. User profiles are also used to provide recommendations for online activities such as commenting on news stories [15] and mentions in micro-blogging systems [22].

2.1.1 User Interests and Expertise

Finding user interests and expertise is critical for many important applications. Many of these applications seek to find interests or expertise either explicitly [23, 24, 19, 13] via user relevant features and tags, or uncover latent interest or expertise [21, 16, 17] with models such as matrix factorization. Dou et al. [9] built user interest profile by predefined topic categories for personalized webpage re-ranking. Ahmed et al. [21] proposed a statistical framework to extract user's dynamic interest profile for behavioral targeting.

To increase user base and engagement, many works [24, 25, 26, 27, 28, 16] create rich user models for better user experience in social media. For example, Hong et al. [26] focused on modeling retweeting behavior on Twitter using factorization machines on users and tweets. Zhao et al. [16] proposed a matrix factorization based approach to build topical profiles by distinguishing user behaviors. Jiang et al. [27] presented a

probabilistic matrix factorization model to exploit social context to better predict user adoption behaviors on Weibo. Chen et al. [24] built topic profiles of users with bag-of-words model to recommend conversations in Twitter.

Besides user's topic interest profile, other works[29, 30, 31] focus on inferring user's attribute and demographics profile such as gender and education. For example, Mislove et al. [31] proposed community detection based approaches to infer user's college and major. Li et al. [29] presented a weakly-supervised approach to extract user's job and education.

Many previous works [32, 33, 34, 35] have focused on finding general topic experts in many domains (e.g., enterprise corporate, email networks), with a recent emphasis on social media and microblogging sites [36, 37, 38]. Weng et al. [38] proposed a PageRank-based approach to find topic experts by taking advantage of both topical similarity between users and social link structure. Pal and Counts [37] introduced a probabilistic clustering followed by a within-cluster Gaussian ranking framework to find topic authorities using nodal and topical features on Twitter. Ghosh et al. [36] proposed and built a system called Cognos to find topic experts by relying on Twitter Lists (though not with any geo-spatial information, as in this work). Recently, Cheng et al. [39] addressed the problem of identifying local experts on Twitter. Our dissertation work extends on these prior efforts by focusing on *personalized* experts for users.

2.1.2 User Opinion Bias

There has been considerable research effort devoted to exploring political polarization, assessing media bias of major news outlets, and assessing user sentiment towards particular topics.

Political polarization has been a topic of great interest in the past decade and studied in news articles [40], online forums [41] and social media [42, 43, 44, 45, 46, 47, 48]. Adamic and Glance [42] demonstrated the divided community structure in the citation network of

political blogs. Conover et al. [45] and Livne et al. [47] showed that there exists a highly segregated network structure using modularity. Guerra et al. [46] compared polarized and non-polarized networks and proposed a new measure to determine whether a network is polarized given that the network is also modular. Since knowing users' political orientation can be of great importance for understanding the overall political landscape, many approaches have been proposed to classify a user's political identity. Conover et al. [44] and Pennacchiotti and Popescu [48] exploit text and network features for classification. Akoglu [41] proposed to use signed bipartite opinion networks for the classification and ranking of user's political polarity on forum data. Zhou et al. [40] applied semi-supervised learning methods to classify news articles and users' political standing. Cohen et al. [43] employ supervised methods to classify political users into groups with different political activities, and conclude that it is hard to infer "ordinary" users' political orientation. In our work, instead of simply focusing on the classification of user's political orientation, we are interested in developing a flexible tool to explore controversial themes and discover their underlying users' degree of opinion bias on a topic basis. We show that user's opinion bias can be leveraged to improve other applications such as user recommendation.

Apart from user-oriented political orientation, some works have explored media bias. Groseclose et al. [49] proposed a new measure to quantify media bias by comparing the number of citations to think tanks and policy groups to those of Congress members. Gentzkow et al. [50] also proposed a media bias measure which considers the frequency of phrases quoted by Congressional members of Republican and Democratic parties in newspapers. Lin et al. [51] focused on the measure of coverage quantity to compare the extent of bias between blogs and news media. Wong et al. [52] quantified the political leanings of media outlets on Twitter using aggregated retweeting statistics. Our work differs from these in that we target the opinion bias of "regular" users instead of prominent media, and with respect to different controversial topics instead of only political leanings.

There are also prior works which infer user's sentiment toward a topic in social media or online forums. Tan et al. [53] proposed a semi-supervised approach to inferring users' sentiment using social network information. Kim et al. [54] and Gao et al. [55] proposed to use a collaborative filtering like approach to estimate user-level sentiment. Lu et al. [56] proposed to use content and social interactions to discover opinion networks in forum discussions. However, our work has two differences from these and other sentiment-oriented approaches. The first is that many of these works require a significant amount of manually labelled tweets or users as ground truth. In our work, we develop automatic approaches using crowdsourced hashtags as seeds to substantially reduce manual labor. The second is that we focus on intrinsic opinion bias instead of sentiment. Sentiment [57] centers around users' attitude or emotional state, usually reflected by the use of emotional words. However, opinion bias can also be reflected by the news or factual information she chooses to post, which may lack any prominent emotional words.

2.2 Exploiting User Context

Contextual information in social media has been widely used in previous works [58, 59, 60, 61, 62, 63, 28, 64] to learn better user profiles and improve task performance. Various contextual signals have been exploited to improve learning user's interests and expertise, including text [65, 63, 66, 67, 68], social networks [27, 60, 14], geographical footprints [69, 70], behavioral signal [13, 71], temporal context [72, 28], emotions and sentiment [73, 74], and linguistic activity [75].

A seminal work by Singh and Gordon [62] proposed collective matrix factorization to simultaneously factor several matrices encoding contextual information for better prediction of user-movie ratings. Wang and Blei [63] proposed collaborative topic regression to learn latent user preference by modeling both ratings and content. Similarly, Gopalan et al. [58] also modeled both user's ratings and content but with Poisson factorization. Jamal

et al. [59] proposed a context dependent factor model to learn general latent factors of entities in social networks for better recommendation. Guy et al. exploited user's tagging behavior to improve content recommendation [13]. Temporal information has also been used in [28] to learn both user-oriented topics and time-oriented topics. Other contextual information used to learn user preference includes domain-specific communities [64] and social relations [60, 61] which are used to regularize latent factors between socially connected users.

Geographical footprints have also been widely explored in many location-based applications [76, 69, 77, 78, 79, 80, 81, 82, 83]. One of the most popular applications is point-of-interest (POI) recommendation on social networks, where geographical influence is combined with user preference for better performance. For example, Ye et al. [84] explored the spatial clustering phenomenon and proposed a unified POI recommendation framework combining user preference, geographical influence, and social influence. Cheng et al. [76] proposed a multi-center Gaussian model to model user's check-in behavior, which is used as input for a generalized matrix factorization framework. Liu et al. [85] proposed a geographical probabilistic factor analysis framework, which jointly models the effect of geographical distance, user preference, POI popularity and user mobility. Another different application that utilizes geographical footprints is the rating prediction problem in Yelp [86], where Hu et al. observed weak positive correlation between a business's ratings and its neighbor's ratings, and used this observation to improve rating predictions.

Other works have used geographical influence for rating prediction in Yelp [69], activity recommendation with GPS history [83], expert recommendation [80] and event-based group recommendation [82].

2.3 Modeling User's Implicit Feedback

In recommendations when only implicit feedback is available, one-class collaborative filtering approach [87, 88, 82] can be used for learning rank between items. Rendle et al. [88] proposed a pairwise method called bayesian personalized ranking (BPR) framework which models the order of items. Following this, Rendle et at. [89] extended matrix factorization to tensor factorization for tag recommendation. Chen et al. [87] adapted BPR framework for tweet recommendation by incorporating tweet topic factors, social relation factors and other explicit features. Zhang et al. [82] also used pairwise matrix factorization to recommend event-based groups to users with location features, social features and tags. Krohn-Grimberghe et al. [90] extended this framework to model both user's feedback and social relations by optimizing weighted loss functions so that social relation is considered as part of objective function.

More generally, user's implicit feedback is often in the form of binary data, which can also be considered as discrete count more broadly. An emerging line of research [72, 91, 92, 93, 94] has focused on modeling discrete count data with Gamma-Poisson distribution instead of traditional Gaussians for recommender systems [95, 91], topic modeling [58], spatial data analysis [93] and political science [96]. For example, Gopalan et al. [91] proposed a hierarchical Poisson matrix factorization for item recommendation to users based on implicit feedback, and later developed collaborative topic Poisson factorization [58] by integrating topic modeling for better article recommendation with user clicks. Yu et al. [93] proposed a weakly-supervised labeled Poisson factor model to predict the number of app openings at different locations through aggregated spatial data from mobile app usage. Zhou et al. [94] proposed to use beta-negative binomial process as a non-parametric Bayesian prior for an infinite Poisson factor analysis model. More recently, Schein et al. [96] have generalized two dimensional Poisson factorization to

Bayesian Poisson tensor factorization for inferring multilateral relations to analyze international affairs between countries based on political events data. Zhou et al. [97] proposed Poisson gamma belief network, a multi-layer generative probabilistic framework based on Poisson distribution for modeling two-dimensional discrete count data. In our dissertation work, we exploit Bayesian Poisson factorization to model user's implicit feedback and user's popularity counts instead of probabilistic matrix factorization based on Gaussian distribution for better performance.

3. PERSONALIZED EXPERT RECOMMENDATION*

In this section, we begin to address the first challenge in large-scale information sharing system — how to connect the right person with the right resource. To that end, we formalize the problem of *personalized expert recommendation*, and propose a principled matrix factorization based approach to effectively recommend personal experts to users.

3.1 Introduction

Finding and recommending *experts* is a critical component for many important tasks. For example, the quality of movie recommenders can be improved by biasing the underlying models toward the opinions of experts [98]. Making sense of mobile and social information streams such as the Facebook newsfeed and the Twitter stream can be improved by focusing on content contributed by experts. Along these lines, companies like Google and Yelp are actively soliciting *expert reviewers* to improve the coverage and reliability of their services [99]. More generally and in contrast to search engines and question-answer systems, experts can provide ongoing help for evolving and ill-specified needs, as well as personalized access to knowledge and experience that only experts possess.

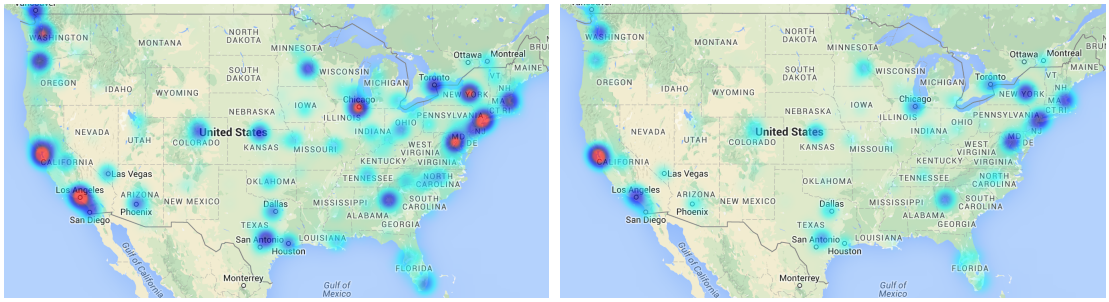
Indeed, there has been considerable effort toward expert finding and recommendation, e.g., [32, 33, 18, 19, 34, 37, 38, 35]. These efforts have typically sought to identify topical experts with broad appeal, e.g., the top Java developer in an enterprise, the best lawyer in Texas. However, there is a research gap in our understanding of both (i) identifying *personal experts*, that is experts who are of significance and importance to me, but perhaps not viewed so more broadly. For example, I may be interested in the expert opinions of nearby

*Reprinted with permission from "Exploiting Geo-Spatial Preference for Personalized Expert Recommendation" by Haokai Lu and James Caverlee, 2015. *Proceedings of the 9th ACM international conference on Recommender Systems*, 67-74, Copyright 2015 by ACM. DOI: http://dx.doi.org/10.1145/2792838.2800189.

local foodies, but less interested in the opinions of globally popular celebrity chefs; and (ii) how spatial preference for personally-valuable expertise varies across topics, across regions, and based on different underlying social communities. For example, technologists in Houston, TX may be more interested in the opinions of experts in nearby Austin and in more distant Silicon Valley, but less so in the opinions of experts from New York. Similarly, the reach of experts may vary by location, so that tech experts from Silicon Valley have a larger footprint than do experts from other regions.

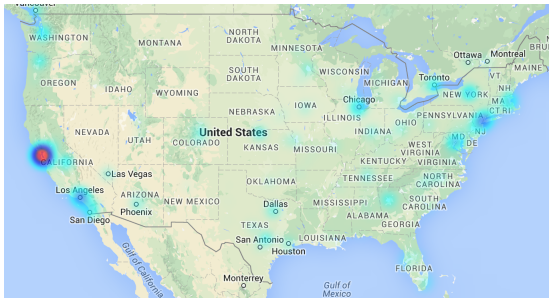
Hence, in this section, we are interested to study the problem of *personalized expert recommendation* by integrating the geo-spatial preferences of users and the variation of these preferences across different regions, topics, and social communities. These geo-spatial preferences are increasingly being revealed through the fine-grained geo-spatial footprints of Instagram, Foursquare, and Twitter, among other mobile location sharing platforms. Concretely, we opportunistically leverage a collection of GPS-tagged Twitter users and their relationships in Twitter lists, a form of crowd-sourced knowledge whereby user A may label user B with a descriptor (like "technology"). In isolation these lists allow a user to organize a personal Twitter stream; in aggregate, the many labels applied to a target user in many lists can provide a crowdsourced expertise profile of the target user. Specifically, we propose and evaluate a matrix factorization-based personalized expert recommender that leverages three key factors:

- *Region-based locality*, reflecting the variation in spatial preference from region to region. For example, Figure 3.1c and Figure 3.1d shows that the preference of users for food experts varies greatly based on the location of the user (in essence, local users prefer local foodies). How can these regional differences be captured and incorporated into a personalized expert recommender?
- *Topic-based locality*, reflecting the variation in spatial preference across different topics.

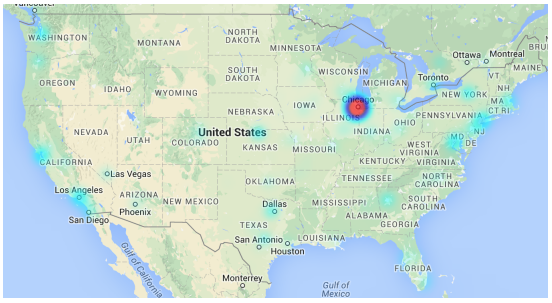


(a) Food experts

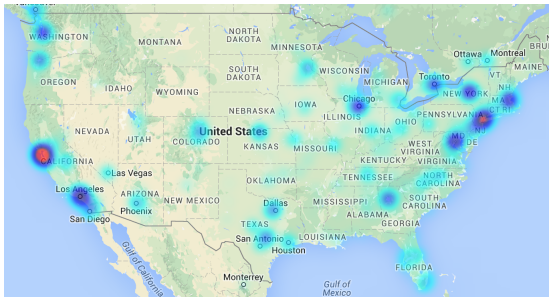
(b) Tech experts



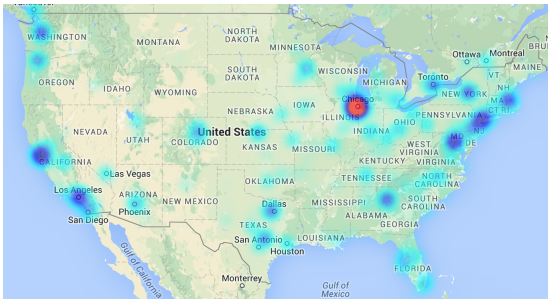
(c) Users for SF-based food experts



(d) Users for Chicago-based food experts



(e) Users for SF-based tech experts



(f) Users for Chicago-based tech experts

Figure 3.1: Spatial distribution of experts (a, b) and for the users who have listed experts (c, d, e, f) based on geo-tagged Twitter lists.

For example, Figure 3.1c and Figure 3.1e demonstrate that spatial preference is much less local for the topic technology than for food. How can this topical variation be integrated into a personalized expert recommender?

- *Social-based locality*, reflecting the social connections between users and experts. For example, are users who are connected in an underlying social network more "similar" in their preferences for experts? Are experts who are more tightly coupled in the underlying social network preferred by the same set of users?

Through extensive experimental validation, we find that each of these factors – region, topic, and social-based locality – improves the quality of personalized expert recommendation. And together, the proposed model achieves around 24% improvement in precision and 21% improvement in recall versus both a collaborative filtering and a baseline matrix factorization based recommender. Furthermore, we also find that the proposed approach can ameliorate the cold start problem when users have few experts on their lists, leading to more than 20% improvement over the baseline in precision and recall.

3.2 Preliminaries

In this section, we introduce the problem of personalized expert recommendation and outline our core approach.

3.2.1 Problem Statement

We assume there exists a set of users $U = \{u_1, u_2, ..., u_N\}$, where N is the total number of users. From this set U, there are a number of recognized experts denoted as $E = \{e_1, e_2..., e_M\}$, where M is the total number of experts. Each user has a preference over some of these experts, expressed as a *personalized expertise list*. For example, Alice may prefer Beth to Candace in the topic of "Java programming", but have no opinion on Doug. We then define the problem of *personalized expert recommendation* as: Given a user u_i, identify the top-n personally relevant experts to u_i. That is, can we further identify experts

Eva and Frank that are of personal interest to Alice?

3.2.2 Recommendation by Matrix Factorization

We tackle *personalized expert recommendation* using latent factor matrix factorization [100]. We assume there is a factor p_i associated with each user u_i and a factor q_j associated with each expert e_j. The model defines a rating score between the list and the expert, denoted as y_{ij}, and factors the score into a latent space through p_i and q_j as follows:

$$y_{ij} = p_i^T q_j + b_j \tag{3.1}$$

Through this factorization, we can think of q_j as the latent properties for expert e_j, p_i as the latent preference of user u_i and b_j as the popularity bias for e_j. However, unlike the standard recommendation task, we do not have a rating score for each expert on the lists. Instead, we only have the implicit feedback for a list, which assumes a user prefers an expert who is already on the list to an expert who is not. Accordingly, the learning objective should be based on the pair-wise ranking between experts. In recommendations when only implicit feedback is available, the one-class collaborative filtering approach [87, 88, 82] can be used for learning a rank order among items. Similar efforts have been targeted at tag recommendation [89], tweet recommendation [87], and event-based groups [90, 82]. Here, we adapt the Bayesian Personalized Ranking (BPR) criterion proposed by Rendle et al. in [88] to our problem.

Formally, for a user u_i, an expert e_k and an expert e_h, suppose u_i puts e_k on the list while not e_h, we denote this pair as $e_k^{u_i} \succeq e_h^{u_i}$, and the likelihood for this preference under BPR can be written as:

$$p(e_k^{u_i} \succeq e_h^{u_i}) = \sigma(y_{ik} - y_{ih}) \text{ where } \sigma(x) = \frac{1}{1 + e^{-x}}$$

Therefore, the likelihood for all users could be written as:

$$p(R|\Theta) = \prod_{e_k^{u_i} \in \mathcal{P}^{u_i}, e_h^{u_i} \in \mathcal{N}^{u_i}, u_i \in U} p(e_k^{u_i} \succeq e_h^{u_i})$$

where R is the set of all preference pairs, Θ is the set of all parameters, $\mathcal{P}^{u_i}$ is the set of experts included on u_i's list and $\mathcal{N}^{u_i}$ is the set of absent experts for u_i. If Θ has a prior density $p(\Theta)$, we can derive a bayesian version of the likelihood, where the prior is used to prevent the overfitting of the parameters as a form of regularization. Thus, the posterior log-likelihood to maximize is

$$p(R|\Theta) = \sum_{e_k^{u_i} \in \mathcal{P}^{u_i}, e_h^{u_i} \in \mathcal{N}^{u_i}, u_i \in U} ln(\sigma(y_{ik} - y_{ih})) - \text{regularization}$$

which can be learned through stochastic gradient descent (SGD) by iterating each of the preference pairs and updating the corresponding parameters.

3.3 Region, Topic, and Social-Based Locality

While promising, the baseline matrix factorization approach ignores the geo-spatial preferences of users and the variation of these preferences across different topics, regions, and social communities (as suggested by Figure 3.1's intuitive support for these notions). Hence, we turn in this section to demonstrating how these factors manifest in real-world Twitter-based data and how each of these factors can be incorporated into a new personalized expert recommendation matrix factorization framework.

3.3.1 Data and Metrics

We begin by highlighting the data used here and two statistical measures – expert entropy and expert spread – to characterize region and topic-based locality. We then turn to the social properties of the dataset to demonstrate social-based locality.

Table 3.1: Geo-tagged Twitter list data.

topic	# of lists	# of experts	# of listings	sparsity(%)
news	35,539	20,295	287,321	0.04
music	17,945	7,896	160,286	0.11
sports	16,018	5,395	139,838	0.16
food	10,476	5,485	96,661	0.17
celebs	9,783	4,090	104,004	0.26
tech	13,046	10,760	125,178	0.26
general	30,000	36,217	289,528	0.03

Data. We use the geo-tagged Twitter lists collected in [18]. In total, there are about 12 million crowd-generated lists and 14 million geo-tagged listings, where a geo-tagged listing indicates a direct link from a list creator to an expert where both of their geo-locations are known. That is, each user $u_i \in U$ is associated with geographical coordinates $coord_{u_i}$. Furthermore, for each list, there exist associated labels that list creators use to indicate the topic of that list. In the following analysis, we selected lists which include the most frequent unigram labels indicating typical topics as follows: news, music, tech, sports, celebs, and food. Additionally, we randomly sampled lists which include any unigram occurring more than 200 times in list labels. We denote this randomly sampled list data as "general". Furthermore, we excluded experts who have only occurred in one list and also excluded lists which includes only one expert. After filtering, we have the geo-tagged Twitter list data statistics shown in Table 6.2. In the following sections, we refer to list creators as *users* and list members as *experts*.

Metrics. We discretize the continental US surface with a 1° by 1° geodesic grid to map the coordinates to discrete regions.* Formally, we have a total number of K grids, which we call regions. We denote K regions as $R = \{r_i | i = 1, 2, ..., K\}$, to which each coordinate

*1° by 1° is approximately 70 miles by 50 miles at latitude 40°. We also tested a finer mesh of 0.1° by 0.1°, which gave quantitatively similar results.

inside the US can be mapped. Furthermore, we assume for an expert e, there are totally n_e users who put e on their lists. Among them, we let U^e be the set of users for expert e, and $U^e_{r_i}$ be the set of users from the region r_i. Thus, the probability of expert e's user from the region r_i can be defined as $p^e_{r_i} = \frac{|U^e_{r_i}|}{\sum_{r_i \in R} |U^e_{r_i}|}$. With these preliminaries, we quantity the geographical characteristics of expertise with:

Expert entropy. The *expert entropy* is defined as

$$H(e) = -\sum_{r_i \in R} p^e_{r_i} log(p^e_{r_i})$$

This measure indicates the degree of randomness in spatial distribution of the users for an expert. It ranges from 0 when all users for the expert are only from one region, to $logK$ when user's distribution is uniform across all regions. Thus, it implicitly reflects the level of an expert's recognizability across the entire country.

Expert spread. While entropy provides insights into the spatial distribution of users, it lacks explicit consideration for the distance between a user and an expert. Hence, we define another measure called *expert spread* as follows:

$$S(e) = Median_{u_i \in U^e}(d(coord_e, coord_{u_i}))$$

where d is the distance between two locations, computed with Haversine function to account for the shape of the earth as follows:

$$d = 2rarcsin(\sqrt{haversin(\phi_2 - \phi_1) + cos(\phi_1)cos(\phi_2)haversin(\psi_2 - \psi_1)})$$

where $haversin(\theta) = sin(\theta/2)^2$ is the Haversine function, r is the redius of the earth, ϕ represents the latitude and ψ the longitude. The expert spread indicates how far a typical labeler is from an expert, thus can be considered as the localness of an expert.

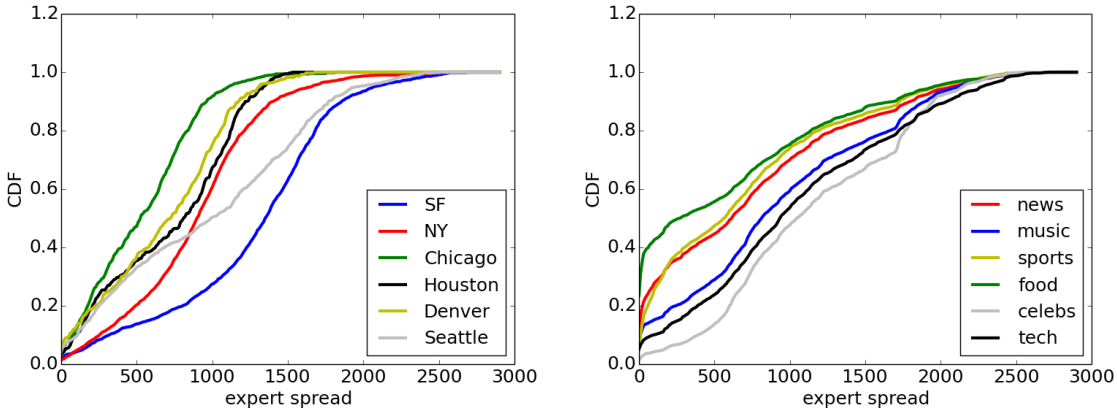
Table 3.2: Average expert entropy for different cities.

topic	SF	NY	Houston	Chicago	Seattle	Denver
news	**2.461**	2.342	2.021	1.950	1.836	1.884
music	**2.514**	2.386	1.946	1.996	2.105	2.162
sports	2.518	**2.703**	1.956	2.281	2.217	2.060
food	1.689	**2.105**	1.315	1.172	1.439	1.327
celebs	**3.274**	2.777	3.013	2.781	2.950	2.842
tech	2.323	**2.400**	2.262	2.249	2.098	1.917
general	**1.954**	1.932	1.645	1.610	1.606	1.585

3.3.2 Region-Based Locality

In Figure 3.1c and 3.1d, we observed that food experts from San Francisco and Chicago are preferred by users nearby. How does this observation manifest according to our statistical measures? To that end, we select experts from the following cities: San Francisco (SF), New York (NY), Chicago, Houston, Denver and Seattle. We first show the average expert entropy for these cities with respect to different topics in Table 3.2. As can be observed from the table: (i) Experts from different geo-locations have different levels of recognizability across the country; and (ii) Generally, experts from SF and NY are popular in more regions than those from other geo-locations, indicating that SF and NY have a greater impact on expertise curation for users on Twitter.

In Figure 3.2a, we examine expert spread for these cities. We can see that generally, experts from different geo-locations have different levels of locality, with experts from Chicago and SF having the smallest and largest expert spread. This indicates that compared to other cities, Chicago has the most local influence on expertise curation while SF reaches the farthest. Combined with the observations from Table 3.2, we conclude that *experts from different regions may have different levels of locality*, i.e., some may reach a wider geographical scope but others may be only locally popular.



(a) (b)

Figure 3.2: CDF of expert spread (a) for different cities and (b) for different topics.

Integrating region-based locality. Since the observed region-based locality reflects collective opinion, how can we integrate it into personalized expert recommendation? That is, if we know the geo-location of a user, can we recommend experts who are popular around the user's geo-location? As can be observed in Figure 3.1, an expert's popularity is not necessarily linear in the distance between user and expert; rather, it is often in the form of "clusters", i.e., experts may be popular in one region but not in other regions. Thus, we introduce the concept of "regional popularity", where we parameterize the popularity of each expert by regions, with the regional popularity to be learned from training.

Concretely, we assume the geographical space is partitioned into K regions. For an expert e_j and a region r_i, we assume there is a popularity parameter s_{ij} associated with r_i and e_j. This parameter is used to capture the degree of popularity that expert e_j receives in region r_i. Thus, if we have a total of M experts, the popularity parameters constitute the matrix S of dimension K by M, which represents regional popularity for all experts. Each column $s_{\cdot j}$ represents the popularity e_j receives in all regions. Then, given a user u_i, the popularity e_j receives at the region where u_i is from is denoted as $s_{c(u_i)j}$, where $c(x)$ is

Table 3.3: Average expert entropy and expert spread (miles) when $CDF = 0.5$ for different topics.

topic	food	news	tech	sports	music	celebs
entropy	1.661	2.048	2.235	2.247	2.267	**2.868**
spread	290	630	950	580	830	**1060**

a function mapping a user to its region. We use $s_{c_i j}$ instead of $s_{c(u_i)j}$ for convenience. By integrating the matrix S to the original matrix factorization, we have:

$$y_{ij} = p_i^T q_j + s_{c_i j} \tag{3.2}$$

We denote Equation 3.2 as the Geo-Enhanced factorization (GEF). Note that GEF is reduced to Equation 3.1 when $K = 1$. The GEF approach has the advantage over the baseline matrix factorization of explicitly capturing and learning expert regional popularity.

3.3.3 Topic-Based Locality

In the previous analysis of expertise, we observed that expert entropy can be impacted by geo-locations (see Table 3.2). Additionally, this table also implies that expert entropy can be impacted by the choice of topic. To further observe the geo-spatial distribution of expertise for different topics, we list the average expert entropy for the six sample topics in Table 3.3. As we can see, celebs has the largest entropy, which indicates that users interested in celebrity are most widely spread across the country; while food has the smallest entropy, indicating that users interested in food experts are most concentrated in certain regions. This is intuitively reasonable since a celebrity is very likely to have a better chance of being known in the whole country than a food expert from a certain location.

In Figure 3.2b, we show the cumulative density function against expert spread for different topics. We can see that, for a fixed spread value, the topic food gives the largest cumulative probability, indicating that users interested in food are closest to the experts; while users interested in celebrity are farthest. We also show the expert spread when the CDF is 0.5 in Table 3.3. We can see that the topics with increasing expert spread are ordered as: food < sports < news < music < tech < celebs, with food having the smallest expert spread of 290 miles, and celebs having the largest expert spread of 1060 miles, which is almost half the distance from the west coast to east coast of the US. Combined with the previous observation on expert entropy, we can conclude that the topic food is the most local among all, with users mostly concentrated in local regions of experts, while the topic celebs is the least local, with users scattered across the country. In another word, users interested in food tend to select food expert nearby, while users interested in celebrity do not have such geographical constraints, and users interested in other topics fall in between.

Overall, we can conclude that *expert's regional popularity can vary by topic*; in other words, users may have different regional preference for experts because of their topic interests.

Integrating topic-based locality. Now that we have observed that topic locality can influence user's preference for experts, it is important that user's interests should be aligned with the interests of the experts to be recommended. Since each Twitter list is labeled with certain keywords, we can aggregate all of the labels for an expert in all lists he appears. As a result, an expert e_j has a description d_{e_j} consisting of the aggregated labels. We then introduce a user latent topic factor t_{u_i}, representing u_i's topical preference, and expert topic factor $\bar{t}_{e_j}$, representing the topical property of e_j. Thus, the inner product $t_{u_i}^T \bar{t}_{e_j}$ indicates an affinity score of user u_i and expert e_j with respect to topic. Here, $\bar{t}_{e_j}$ is treated as known through d_{e_j}, and t_{u_i} is treated as unknown to be learned. The reason to model in this way is that labels for lists often have only one term, e.g., lists with one

term label "food" occupy about 60% percent of total lists with any "food" in its labels. But often, a list is very focused on finer aspects of a topic. For example, a list labeled with "food" may include many "wine" experts, implying that we should also consider expert candidates labeled with "wine". By making t_{u_i} unknown, we are forcing the model to learn topic aspects of a user from those of experts she selected. For convenience, we use t_i instead of t_{u_i} and $\bar{t}_j$ instead of $\bar{t}_{e_j}$ afterwards. Thus, our Topic-Enhanced factorization (TEF) can be written as:

$$y_{ij} = p_i^T q_j + t_i^T \bar{t}_j \tag{3.3}$$

Here, we treat each label as a dimension of $\bar{t}_j$. Through the explicit handling of each user's topic aspects, it is expected that user's interests are aligned with the interests of the experts to be recommended.

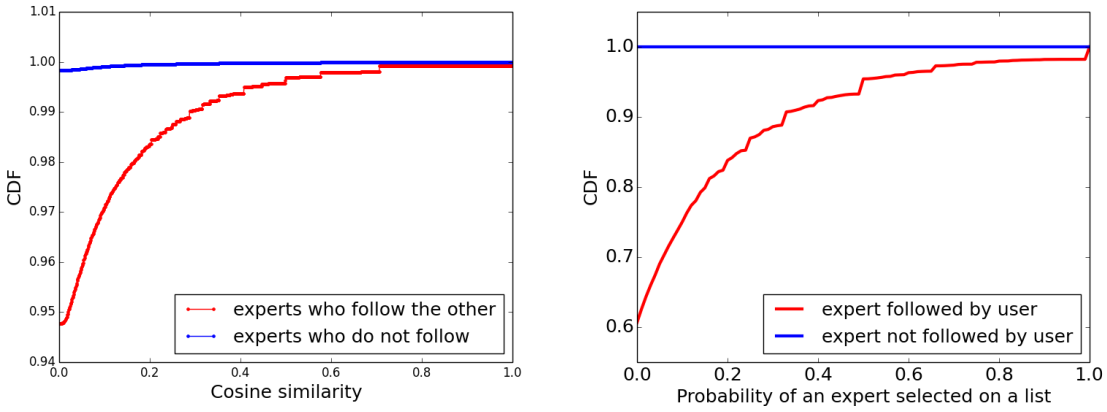
Fusion of region and topic-based locality. Naturally, we can integrate both region and topic-based locality into the model. We adopt a linear model for the integration of Equation 3.2 and Equation 3.3, resulting in our Geo-Topic Enhanced factorization (GTEF):

$$y_{ij} = p_i^T q_j + s_{c_i j} + t_i^T \bar{t}_j \tag{3.4}$$

The intuition is when we know the region of a user and her topic aspects (by looking at the labels of her selected experts), we can recommend an expert both topically and geographically relevant.

3.3.4 Social-Based Locality

In addition to the modeling of region and topic based locality, we are also interested to explore if social connections among users and experts can improve expert recommendation. Our intuition is that (i) people who are connected by social ties have a higher probability to have similar interests; (ii) people who are socially related may have a higher



(a) (b)

Figure 3.3: (a): CDF for similarity between experts; (b) CDF for the probability of an expert on a list.

probability to select who they follow as experts.

As evidence of social-based locality, in Figure 3.3a, we compare the similarity of experts for two cases: (i) when one expert follows the other; (ii) when no tie exists between two experts. Here, similarity of experts is defined as the cosine similarity computed by viewing each expert as a vector of all users, with each element being a value indicating whether the expert is listed by the user or not. Thus, a large similarity of two experts indicates that they often occur on the same list. We can observe that experts who follow the other generally have a larger similarity. We also compare the probability of an expert selected on a list in Figure 3.3b for two cases: (i) when experts are followed by the user; and (ii) when experts are not followed by the user. We can see that the chance for an expert to be listed by a user is boosted significantly when that expert is already followed by the user. Based on these observations, we individually model three kinds of social relationships:

User-user relationship. In this case, the following relationship is from a list creator to another list creator. When one user follows another, we assume that their preference is more similar to each other than those who do not. In terms of modeling, we adopt the

approach of regulating their latent factors as in [101]. Formally, suppose there are user u_i and u_j, assume $\mathcal{F}_i^u$ is the set of users u_i follows, the social regularization incurred by user user relationship can be written as:

$$\sum_{i=1}^{N} \sum_{f \in \mathcal{F}_i^u} w(u_i, u_f) ||p_i - p_f||^2$$

where $w(u_i, u_f)$ represents the similarity between u_i and u_f. Thus, if u_i and u_f is more similar, the latent preference factor p_i and p_f is also closer. Here, we use cosine similarity of users as the weighting scheme. The cosine similarity of users is computed by viewing each user as a vector of all experts, with each element taking a value — 1 if the expert is on the list, 0 if not.

Expert-expert relationship. In this case, the following relationship is from an expert to another expert. Using a similar approach as in the previous case, we regulate their latent factors so that experts following the other have similar latent factors. Formally, assume $\mathcal{F}_i^e$ is the set of experts e_i follows, the social regularization incurred by expert expert relationship can be written as:

$$\sum_{j=1}^{M} \sum_{f \in \mathcal{F}_j^e} w(e_j, e_f) ||q_j - q_f||^2$$

where $w(e_j, e_f)$ represents the similarity between e_j and e_f. Here, we also use cosine similarity of experts as the weighting scheme.

User-expert relationship. In this case, the following relationship is from a user to an expert. Unlike the previous two kinds of social ties, this relationship links two different entities, and so the regularization approach is ill-suited here. Instead, we explicitly model

these relationships with a bias term added to Equation 3.4 as follows:

$$y_{ij} = p_i^T q_j + s_{c_i j} + t_i^T \bar{t}_j + \theta_i b_{ij} \tag{3.5}$$

where b_{ij} takes a boolean value 1 if u_i follows e_j, and 0 if not. θ_i is a weighting parameter to be learned. Thus, by adding a personalized bias term for each user, the model can take advantage of the following ties between user and expert.

3.3.5 Model Training

Combining the social regularization and Equation 3.5, the final objective function to maximize can be written as:

$$\begin{aligned}
&\sum_{e_k^{u_i} \in \mathcal{P}^{u_i}, e_h^{u_i} \in \mathcal{N}^{u_i}, u_i \in U} ln(\frac{1}{1 + e^{-(y_{ik} - y_{ih})}}) \\
&- \frac{\beta_1}{2} \sum_{i=1}^{N} \sum_{f \in \mathcal{F}_i^u} w(u_i, u_f) ||p_i - p_f||^2 \\
&- \frac{\beta_2}{2} \sum_{j=1}^{M} \sum_{f \in \mathcal{F}_j^e} w(e_j, e_f) ||q_j - q_f||^2 - regularization
\end{aligned}$$

where L_2-norm regularization is adopted, with β_1 and β_2 as the corresponding regularization parameters. In summary, the parameter set Θ to be learned through SGD is $\{p_i, q_j, t_i, \theta_i, s_{.j}\}$. For each iteration of SGD, we need to sample a user u_i, and from u_i's list, an expert e_k. Due to the large size of absent experts for each user, we also need to sample the set $\mathcal{N}^{u_i}$. Here, we adopt the strategy of random sampling. Then, for each sampled triplet $< l_i, e_k, e_h >$, we update each parameter value by taking a step along its gradient ascending:

$$\Theta^{t+1} = \Theta^t + \epsilon \frac{\partial L_{ikh}}{\partial \Theta}$$

where L_{ikh} is the posterior log-likelihood for the triplet $< l_i, e_k, e_h >$, and ϵ is the step size.

Update of S. For expert e_k and e_h, the corresponding parameters to update are $s_{.k}$ and $s_{.h}$. If the region of u_i is c_i, then the parameter $s_{.k}$ and $s_{.h}$ can be updated with

$$\frac{\partial L_{ikh}}{\partial s_{jk}} = -I(j = c_i)\hat{e} + \beta s_{c_i k},$$
$$\frac{\partial L_{ikh}}{\partial s_{jh}} = I(j = c_i)\hat{e} + \beta s_{c_i h}$$

where $\hat{e} = \frac{e^{-(y_{ik}-y_{ih})}}{1+e^{-(y_{ik}-y_{ih})}}$, $I(j = c_i)$ is a Kronecker delta function that gives value 1 if and only if $j = c_i$, for $j = 1, ..., K$, and β is a regularization parameter.

Update of t_i and θ_i. Similarly, we have the gradient for t_i and θ_i as follows:

$$\frac{\partial L_{ikh}}{\partial t_i} = -\hat{e}(\bar{t}_k - \bar{t}_h) + \beta t_i,$$
$$\frac{\partial L_{ikh}}{\partial \theta_i} = -\hat{e}(b_{ik} - b_{ih}) + \beta \theta_i$$

Update of p_i, q_k and q_h. Since p and q are socially regularized, we have the following socially regularized gradients:

$$\frac{\partial L_{ikh}}{\partial p_i} = -\hat{e}(p_k - p_h) + \beta p_i + \beta_1 \sum_{f \in \mathcal{F}_i^u} w(u_i, u_f)(p_i - p_f)$$
$$\frac{\partial L_{ikh}}{\partial q_k} = -\hat{e}p_i + \beta q_k + \beta_2 \sum_{f \in \mathcal{F}_i^e} w(e_k, e_f)(q_k - q_f)$$

The gradient for q_h can be obtained similarly as q_k.

3.4 Experimental Evaluation

In this section, we report on experiments to evaluate the proposed Geo-Topic Enhanced Factorization with Social ties (GTEF-S) for personalized expert recommendation. Specif-

ically, we seek answers to the following questions:

- How well does the proposed method perform compared to alternative baselines? Does region, topic and social-based locality give improvement individually, and if they do, do they complement each other?
- How well does it perform in cold-start situation, i.e., for users who have very few experts on their lists?
- Does the number of regions affect performance? If so, how?

3.4.1 Data Preparation and Experimental Setup

For evaluation, we randomly partition experts for a user into 50% for training and 50% for testing. To determine the number of negative experts for each user, we experimented with $\{50, 100, 150, 200, 250\}$ and selected 150 for a tradeoff between accuracy and computational efficiency. For latent factor dimension, we empirically select 20 for all methods. For regularization parameters β, β_1 and β_2, we use cross-validation for tuning and select 0.02, 0.01 and 0.015, respectively. For gradient step, we initialize it with the step size 0.025, and decrease it to its 98% after each pass throughout all triples. This strategy is shown to be effective in reducing the number of iterations for the method to converge [69].

In the modeling of region locality, it is assumed that the continental US has been partitioned into K regions. Instead of using a gridding approach, we resort to k-means clustering to obtain the partitions by clustering the geo-locations of the entire set of users U. We choose a clustering approach based on Euclidean distance because the geo-spatial distribution of users exhibits a clustering effect, as shown in Figure 3.1, and can be satisfactorily captured by k-means clustering. In section 3.4.5, we evaluate the effect of the number of regions K. For other experiments, we select K to be 80.

For evaluation metrics, we adopt Precision@k (Prec@k) and Recall@k (Rec@k). Prec@k represents the percentage of correctly recommended experts out of the top k rec-

ommendations, while Rec@k represents what percentage of experts can emerge in the top k recommendations. Formally, if we define $Test(u)$ as the set of experts selected by user u and $Reco(u)$ as the set of top k recommended experts, we have

$$Prec@k = \frac{1}{N}\sum_{i=1}^{N}\frac{|Test(u_i) \cap Reco(u_i)|}{k}$$

$$Rec@k = \frac{1}{N}\sum_{i=1}^{N}\frac{|Test(u_i) \cap Reco(u_i)|}{|Test(u_i)|}$$

In our experiments, we evaluate k at 5, 10 and 15.

3.4.2 Baselines

We consider the following baselines:

- Expert Popularity (EP). In this baseline, we recommend experts for each user by ranking experts according to the number times each expert is listed by users.
- User-based Collaborative Filtering (UCF). Collaborative filtering method can be used to discover user's implicit preference by aggregating similar users. Formally, let a_{ij} take a boolean value, where $a_{ij} = 1$ represents expert e_j is selected by u_i, while $a_{ij} = 0$ means the opposite. Thus, according to UCF, the prediction score $\bar{c}_{ij}$ of u_i selecting e_j can be obtained by $\bar{c}_{ij} = \frac{\sum_k w_{ik} \cdot c_{kj}}{\sum_k w_{ik}}$, where w_{ik} is computed with cosine similarity. We then rank the candidate experts according to $\bar{c}_{ij}$ and select the top k experts for recommendation. We select the number of neighbors for each user to be 100.
- MF. This is the basic pair-wise latent factor model shown in Equation 3.1 trained by BPR.
- GEF. This model only considers region-based locality manifested through users' geographical footprints, shown in Equation 3.2.
- TEF. This model only considers topic-based locality manifested through experts' labels,

shown in Equation 3.3.

- GTEF. This model is the fusion of GEF and TEF, considering both region and topic-based locality as shown in Equation 3.4.
- Social MF. This model considers three different kinds of social ties. If the model only considers user user relationship, it is denoted as MF-S1; if the model only considers expert expert relationship, it is denoted as MF-S2; and if the model only considers user expert relationship, it is denoted as MF-S3. We denote the model as MF-S if it considers all three kinds of social ties.

3.4.3 Comparison with Baselines

How well does the proposed method compare to alternative approaches? To answer this question, we first show the performance comparison in Figure 3.4, where we report Prec@k and Rec@k for all topics. As we can see, overall, the proposed GTEF-S generally gives the best performance for different k. Specifically, it gives an average improvement of 24.6% over the best of EP, UCF and MF for precision, and 21.3% for recall. GTEF-S generally performs better than either GTEF or MF-S, indicating the superiority of enhanced pair-wise matrix factorization by considering region, topic and social-based locality, and that these three factors are able to complement each other.

Comparison for MF, GEF, TEF and GTEF. By comparing these methods, we can examine if the explicit modeling of region and topic-based locality can provide any improvement. In Figure 3.4, we can see that GEF, TEF and GTEF perform consistently better than MF. Specifically, GEF gives an average improvement of 3.73% for precision and 3.43% for recall over MF for all datasets. This indicates that the introduction of the regional popularity matrix S for modeling expert's regional popularity can help distinguish regionally popular experts if we know the geo-location of the user.

Furthermore, we can see that TEF also performs consistently better than MF, specifi-

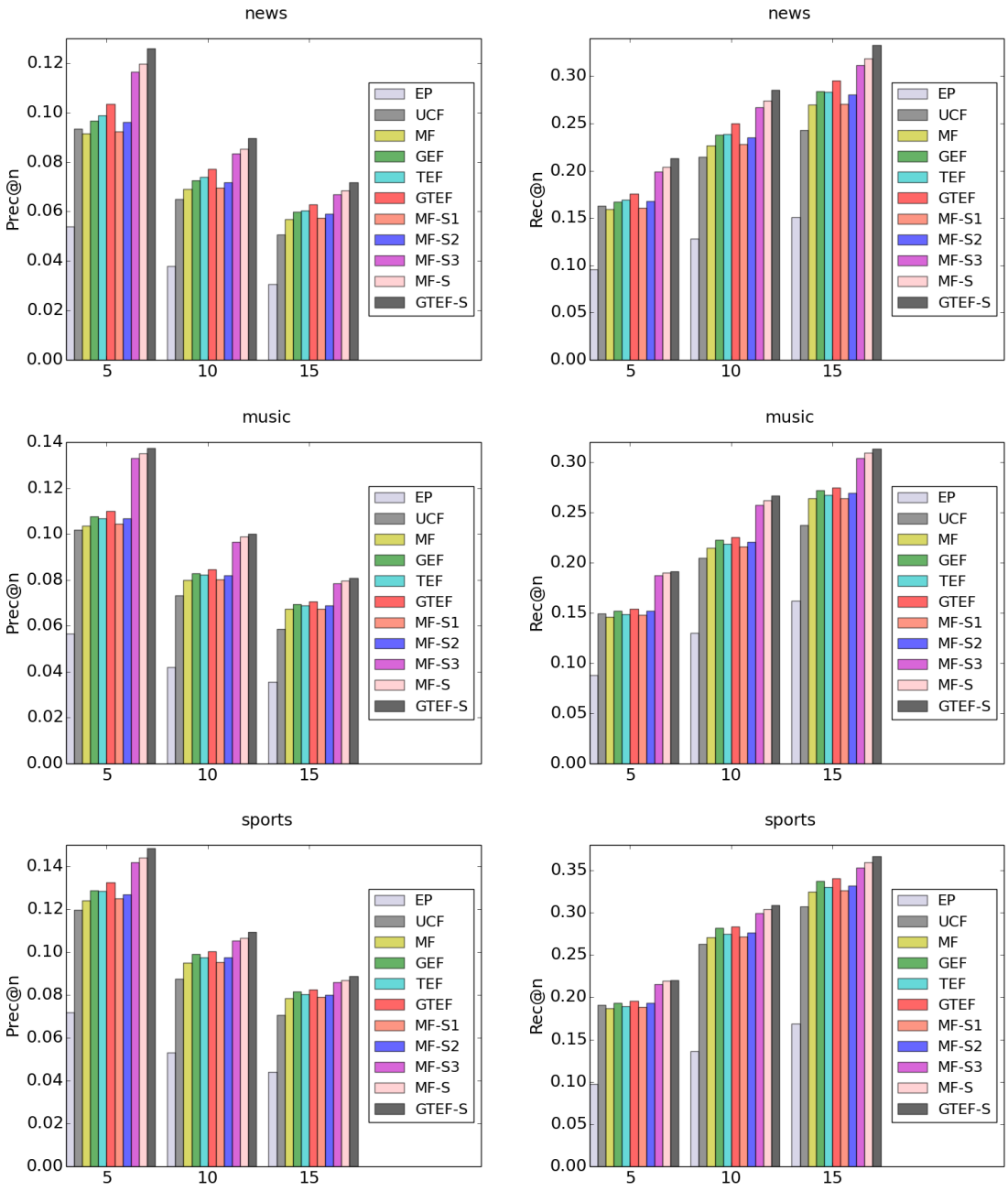


Figure 3.4: Evaluating personalized expert recommendation: Precision and Recall at 5, 10, and 15 for six different topics across 11 approaches.

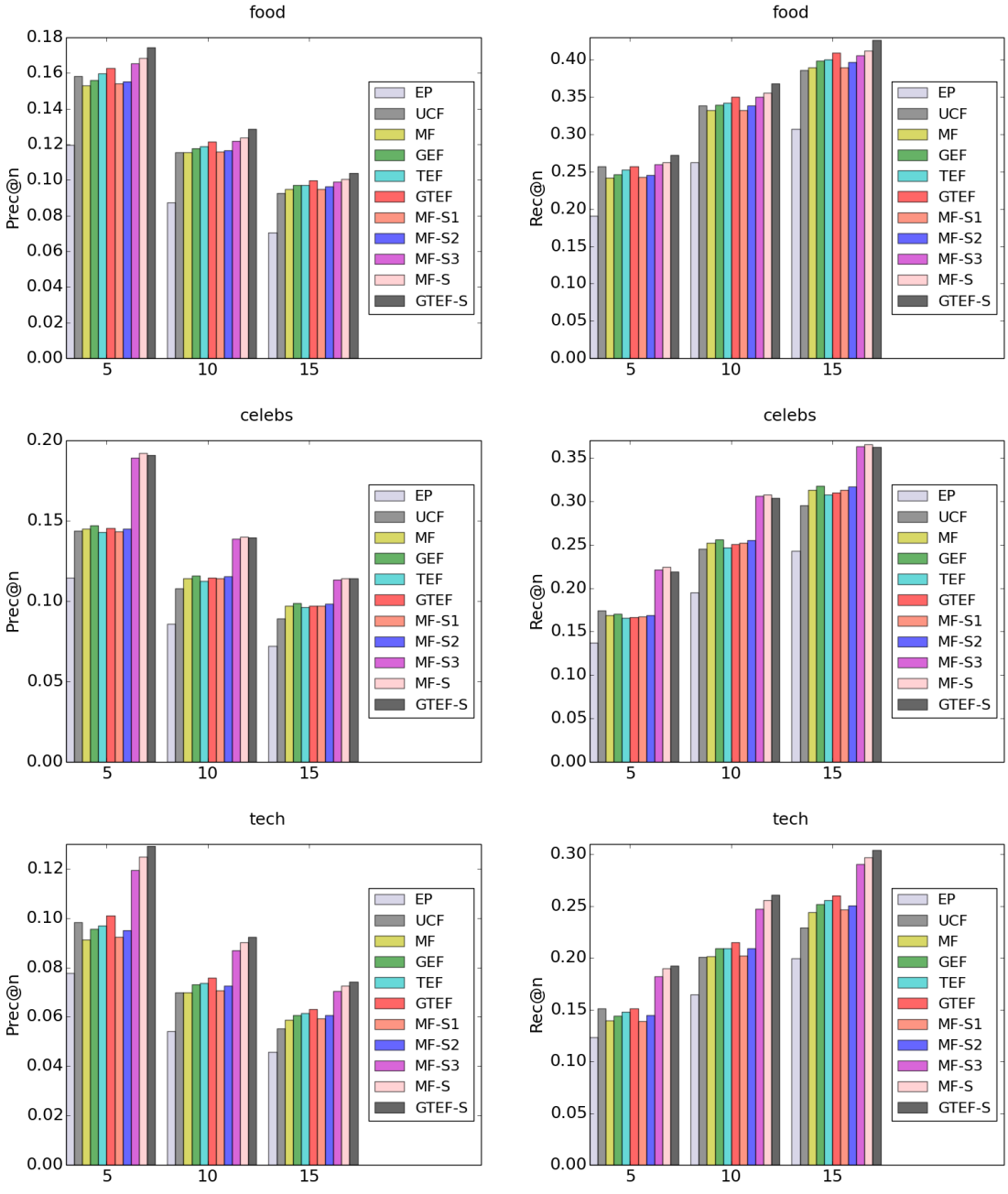


Figure 3.4 Continued.

cally, giving an average improvement of 3.91% for precision and 3.16% for recall. This indicates that modeling user topic factor through the labels of experts can help find experts with similar topic aspect.

Finally, we can see that GTEF gives the best performance among all (an average improvement of 7.35% for precision and 6.28% for recall over MF). These improvements are very close to the additive improvements of both GEF and TEF, thus indicating geographical influence is complementary to topic influence in modeling expert's regional popularity, and that they should be considered together for recommendation. Note that for topic celebs, the improvement of GEF (1.44% for precision) is not as good as those for other topics. This is probably because celebrities are heavily concentrated in the region of Los Angeles, and since its expert entropy is very high, it would not be very useful to model expert regional popularity. Also, TEF performs slightly worse than MF. Upon further examining the labels for experts and user's topic preference factors, we found that the labeling information for experts is scarce and most experts are only labeled with "celebs", without finer topic aspects.

Comparison for MF-S1, MF-S2, MF-S3 and MF-S. By comparing these methods, we can examine if the modeling of social-based locality can help recommend experts. Specifically, we have explored three kinds of relations: user following user (S1), expert following expert (S2) and user following expert (S3). From Figure 3.4, we can see that MF-S1 gives only slightly better performance than MF (0.43% for precision and 0.35% for recall), indicating that social ties between users barely provide additional information for recommending experts. For MF-S2, we see that it provides descent improvement over MF (5.02% for precision and 4.92% for recall), confirming that if an expert is similar to another expert, i.e., if they often co-occur on other lists, it is likely that the other expert can be recommended to this user. For MF-S3, it is shown to be rather effective. On average, it gives an improvement of 16.6% for precision and 14.4% for recall over MF, which confirms that if

a user is already following this expert, it is very likely that this user will include this expert on the list. MF-S, modeling the previous three social relationships together, gives the best improvement of all (21.4% for precision and 18.8% for recall), indicating three kinds of social ties complement each other.

3.4.4 Recommendation for Cold-Start Lists

Previously, we found that the introduction of geographical, topical and social influence in pair-wise MF can improve expert recommendation. In this section, we examine how the proposed methods perform in the cold-start situation. When there is only limited number of experts on a list, there is little positive feedback for training, making it hard to obtain accurate latent factors of users' preferences. In consideration of this, we perform experiments to investigate the recommendation performance of the proposed methods for lists with few experts. Specifically, we select only lists which have fewer than 3 experts on the list to examine the performance. In Figure 3.5, we report the prec@5 and rec@5 for the method MF, GTEF, MF-S and GTEF-S. As we can see, GTEF, MF-S and GTEF-S consistently give better performance than MF for all topics, with GTEF-S showing the best improvement on average (23.7% for precision and 22.3% for recall). This indicates that the knowledge of user's region, topic preference and social relations can help relieve the cold-start problem. Also, MF-S gives better performance than GTEF, indicating that social relation is a stronger signal than user's geo-topic preference. Additionally, note that GTEF-S brings the best improvement over MF-S for news and food. This implies that modeling region and topic-based locality works best when users demonstrate strong regional preference (see Table 3.3).

3.4.5 Effect of Number of Regions

In this section, we study the effect of the number of regions K chosen to cluster the geographical coordinates of users. To that end, we select the number of regions from the set

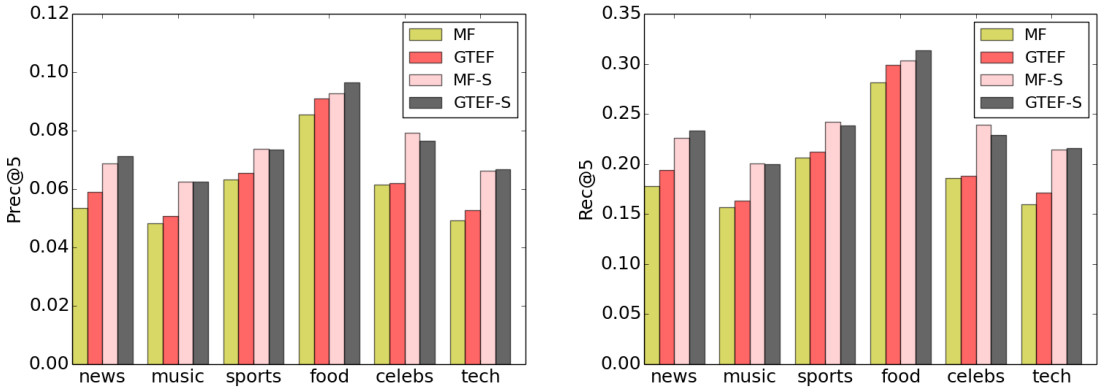


Figure 3.5: Comparing recommenders for cold start lists.

$\{10, 20, 40, 60, 80, 100\}$, and run GTEF with each number for ten random initializations for topic food and tech (we ignore plots for other topics since they show similar trends). In Figure 3.6, we show how prec@5 changes with K. We can see that as the number of regions increases, the precision also generally increases, although the value of K varies for two topics when the performance reaches saturation. Specifically, for tech, the precision reaches almost the best at a smaller K, while for food, the precision gradually increases and reaches the best at a larger K. This can be explained by the observation from the previous analysis about topic locality. Specifically, topics such as food and news are relatively more local, i.e., experts of these topics are listed by local people more often. Also, experts of these topics are usually concentrated in many regions across the country, as shown in Figure 3.1a. As a result, a finer clustering of regions would separate two close regions. For example, it would separate the region of NY and Washington D.C. in Figure 3.1a, so that a user interested in food in NY can be recommended with popular food experts from NY instead of popular food experts from Washington D.C. On the other hand, topics such as tech and celebs are less local, and considering most of these experts are concentrated in fewer regions, it is not necessary to use a finer clustering.

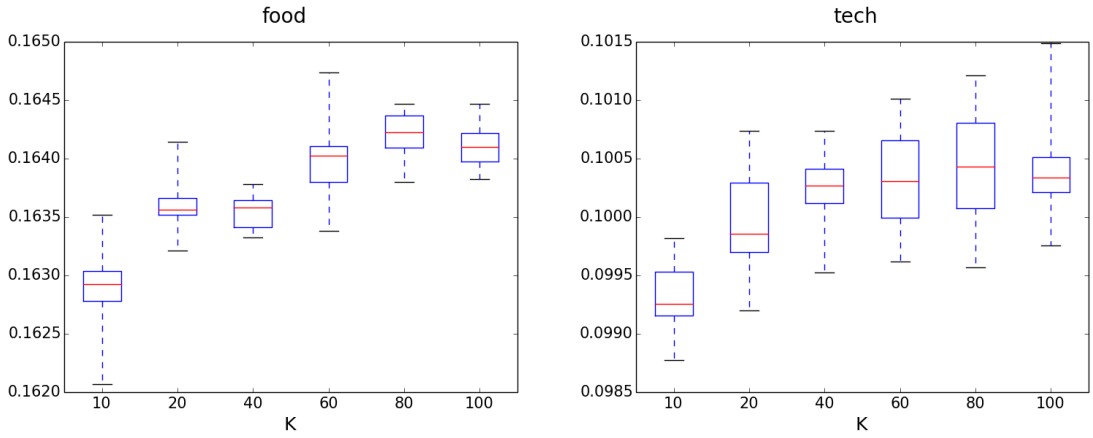


Figure 3.6: Effect of number of regions.

3.5 Conclusion

In this section, we tackled the problem of personalized expert recommendation in GPS-enabled social media. Specifically, we investigated the geo-spatial preferences of users and the variation of these preferences across different regions, topics and social communities. We proposed a matrix factorization-based personalized expert recommender that leverages region, topic and social-based locality. Through experimental evaluation over a Twitter list dataset, we found that the proposed approach achieves more than 20% in precision and recall and can ameliorate the cold start problem compared to several baselines. This confirmed users' geo-spatial preference of expertise and their underlying social communities have great potential for personalized expert recommendation.

4. USER PROFILING: DISCOVERING USER'S KNOWN-FOR PROFILE*

From this section, we will begin to explore and investigate several dimensions of user's profile including user's known-for profile (Section 4), user's opinion bias (Section 5) and user's geo-topic profile (Section 6), and propose models and algorithms to uncover these different aspects of profiles for users. In this section, we start by discovering long-tail user's known-for profile which reflects one's expertise viewed by others.

4.1 Introduction

Discovering what people are *known for* is valuable to many important applications, including recommender systems and question-answering sites. For example, item-based recommenders can be augmented to customize recommendations based on what knowledgeable users prefer, rather than relying on all users [98]. While an individual's *personal interests* are often reflected in the media she consumes and generates, what a user is *known for* is reflected by the views of others and is often not easily discerned. A few high-profile people are easily recognized – for example, a researcher may be interested in basketball, biking, and recommender systems, though mainly known for recommender systems. But there is a long-tail of the vast majority of users for whom we have only limited insight into what they are known for.

And yet accurate identification of a user's known-for profile is challenging. First, the content that a user chooses to post on social media is often noisy and ambiguous. What users are truly known for can be buried among posts about daily routines and personal interests. Second, many users post only infrequently, meaning extreme sparsity for the

*Reprinted with permission from "Discovering What You're Known for: A Contextual Poisson Factorization Approach" by Haokai Lu, James Caverlee and Wei Niu, 2016. *Proceedings of the 10th ACM international conference on Recommender Systems*, 253-260, Copyright 2016 by ACM. DOI: http://dx.doi.org/10.1145/2959100.2959146.

vast majority of all users. Third, while a natural step to overcoming this sparsity is the integration of additional contextual factors (e.g., social links between users), it is not clear how different contextual factors correlate with users' known-for profiles and how we can model these contextual influence and integrate them together.

Hence, in this section, we tackle the problem of *discovering what users are known for* through a probabilistic factorization model called Bayesian Contextual Poisson Factorization (BCPF). Three of the key features of the proposed BCPF model are: (i) It is jointly learned on a small fraction of users whose known-for profiles are already known and the vast majority of users for whom we have little (or no) information. (ii) Moving beyond just modeling the content a user generates, it naturally models and integrates additional contextual factors that provide implicit linkages between users for improved known-for profile estimation. Concretely, we investigate the impact of geo-spatial footprints and social influence as additional contextual signals. (iii) It inherits the strengths of Bayesian Poisson factorization (BPF), a variant of probabilistic matrix factorization recently proposed in [102, 58], which demonstrates scalable inference for sparse data and outperforms traditional matrix factorization [102].

In summary, this section makes the following contributions:

- First, we define the problem of discovering user's known-for profile in social media, and propose a probabilistic method called Bayesian Contextual Poisson Factorization. This model can capture the implicit relationships between user's known-for profile and her content, geo-spatial and social influence. We then develop an efficient approximate variational inference to learn the latent parameters of BCPF.
- Second, we evaluate the proposed method over two Twitter datasets and against several alternative baselines. Overall, we see a significant improvement of 17.5% in precision and 20.9% in recall on average over the next-best method.

- Finally, we study the inferred geo-spatial and social influence latent factors, and observe that the geo-spatial factors are able to capture the underlying distributions of user's known-for profile at different locations. We also show that friends have different social influence on users with respect to different topics, and that our model is able to learn this fine-grained social influence.

4.2 Preliminaries

Problem Formulation. We assume there exist a set of users U in a social network, and a set of tags T used to describe what users are known for with respect to different topics or aspects. The **known-for profile** p_u of a user $u \in U$ is defined to be a subset of tags from T which can best describe what this user is *known-for* by others, instead of what u is personally *interested-in*. To give an example: suppose Alice is known by others as a chef and frequently posts many food recipes, but she also posts news and personal related stories now and then as her personal interest. We associate Alice with tags "chef" and "recipe" instead of "news" to best describe her perceived image by others, and treat "chef" and "recipe" as the *known-for* profile for Alice.

With the above definition, we can define the task of discovering user's known-for profile in social media as follows:

User Known-For Profile Discovery. Given the users U^{old} whose known-for profiles are already known (labeled users), identify the known-for profiles for the rest of the users U^{new} (unlabeled users) based on U^{old}. Equivalently, the task is the same as identifying the top-n tags from T for unlabeled users as their known-for profiles.

However, discovering user's known-for profile is a classic cold-start problem. Since unlabeled users do not have any identified tags in their profile, collaborative filtering techniques fail to infer the profiles of unlabeled users since the corresponding rows of the user-profile matrix are all zeros for unlabeled users. Thus, to overcome the cold-start situation,

we propose to leverage a user's contextual information to extract implicit relationships between their profile and the context for the task as described in the following.

4.3 Known-For Profile Discovery

As a first pass, we can attack the problem of known-for profile discovery with Bayesian Poisson factorization (BPF), a variant of probabilistic matrix factorization recently proposed in [102, 58] to model implicit feedback for recommendation. Due to the assumption of Poisson distribution instead of traditional Gaussian distribution in modeling the discrete data, BPF can capture the long-tailed distribution of user behavior, enjoy scalable inference for sparse data and outperform traditional matrix factorization [102]. In our setting, since a user's known-for profile is inherently of discrete nature, i.e., either a tag exists in the profile or not, it is a natural choice to adopt a Poisson distribution instead of a Gaussian distribution to model user's contextual influence.

Specifically, given the binary occurrence r_{ut} of tag t in the known-for profile for user u, BPF assumes that r_{ut} is generated according to:

$$r_{ut} \sim Poisson(\theta_u^T \theta_t) \tag{4.1}$$

where $\theta_u \in \mathbb{R}^K$ is the latent factor of u representing what u is known for in a latent space, and $\theta_t \in \mathbb{R}^K$ is the latent attribute of t. Each component of these latent parameters is drawn according to a Gamma distribution as follows:

$$\theta_{.k} \sim Gamma(\lambda_a, \lambda_b) \tag{4.2}$$

where λ_a and λ_b are the shape and rate parameter of the Gamma distribution, respectively. Given a binary matrix of known-for profiles, BPF can find each user's latent factors over the tag's latent attributes. These inferred latent factors can be further used to identify other

related tags which may fit the user's known-for profile.

4.3.1 Bayesian Contextual Poisson Factorization

While BPF is able to reveal the latent factors for users whose known-for profiles are already known or partially known, it will fail to infer the known-for profiles of unlabeled users, since these users are "zero rows" in the user-tag matrix. Hence, a natural approach to extend BPF to these missing users is through the integration of additional contextual information that may provide implicit linkages between users.

We refer to this extended framework as **Bayesian Contextual Poisson Factorization** since it inherits the strengths of BPF and is extended to integrate valuable contextual information. Intuitively, BCPF is designed to learn the influence of each contextual feature individually and combine them together for overall representation. Under the assumption of Poisson factorization, since the sum of Poisson random variables is still a Poisson distribution, we can linearly add different contextual influence together without changing the underlying probabilistic assumption.

In the rest of this section, we present how several important contextual factors can be modeled under BCPF – user content, geo-spatial impact, and social influence – and how these factors can be integrated into BCPF. Note that the proposed model is generalizable and other contextual factors could be incorporated as well. Concretely, we model users in a heterogeneous graph $\mathbb{G} = (\mathbb{V}, \mathbb{E})$, consisting of four types of nodes $\mathbb{V} = (U, C, L, P)$. Each user u has a sample of the textual content that she posts online, denoted as $c_u \in C$, and a pair of geographical coordinates $(l_u^{lat}, l_u^{lng}) \in L$ indicating the approximate location of this user's posting activities. The known-for profile of a labeled user from U^{old} is denoted as $p_u \in P$. All of the above relations are directly relevant to users themselves, and can be regarded as user attributes. Furthermore, users may also have social relations with each other. Together, user attributes and social relations constitute all edges $\mathbb{V}$ between

different entities.

Content factor. The first baseline factor is a user's content – that is, we assume that a user's known-for profile can be reflected by the usage of words in posts. Let $\mathbf{R^c}$ be a sparse matrix describing the usage of words from all users, where each element r_{uw} is the count of word w adopted by user u. We then choose to model every count r_{uw} using a Poisson distribution, with the intensity parameter factorizing over u's latent factor θ_u and w's latent factor θ_w as follows:

$$r_{uw} \sim Poisson(\theta_u^T \theta_w) \tag{4.3}$$

where θ_u and θ_w are both K-dimensional non-negative vectors with Gamma priors specified in Equation 4.2.

Factorizing over the word matrix $\mathbf{R^c}$ gives us an unsupervised version of users' latent preference over words. However, for labeled users U^{old}, θ_u should also reflect their known-for profiles. For instance, if both users are labeled with "sports", it is likely that they may share some common words in their sports-related posts. Thus, to capture this intuition, θ_u should also be constrained by users' known-for profiles. Let $\mathbf{R^p}$ be a binary matrix describing known-for profiles for all users, where each element r_{ut} represents if user u is labeled with tag t. Similarly, we model each binary r_{ut} with a Poisson distribution, with its intensity parameter factorizing over θ_u and t's latent factor θ_t as follows:

$$r_{ut} \sim Poisson(\theta_u^T \theta_t) \tag{4.4}$$

where θ_t is a K dimensional non-negative vector parametrized with Gamma priors. By sharing user's latent factor for both factorization 4.3 and 4.4, the model can not only reflect user's latent preference over common use of words, but also ensure that users with similar

profiles have similar θ_u. Note however that a user's posts may be intertwined with other non-revealing texts, and thus can be noisy. To enhance the representation for unlabeled users, we also incorporate each user's geo-spatial footprints and social connections to further refine the model.

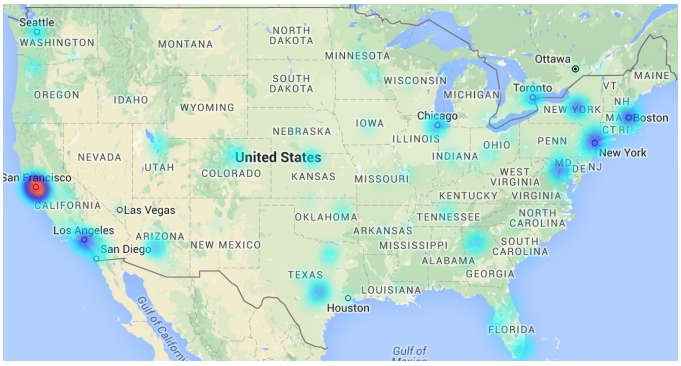
Geo-spatial factor. To demonstrate how geo-spatial location may reflect what a user is known for, we first show heat maps of users in US who are known for "entrepreneur" and "politics" in Figure 4.1 based on a sample of crowd-generated Twitter lists (described more fully in Section 4.4). As we can observe, for both topics, these well-known users are mostly distributed in a few areas, with a majority focused in San Francisco for "entrepreneur" and in Washington D.C. for "politics". This suggests that (i) for a specific tag, users do not distribute evenly across the entire geo-scope and may concentrate in a few areas; (ii) geo-spatial distributions of well-known users may vary by different topics.

Given the above observations, how can we use geo-spatial coordinates to enhance the representation of users? We notice from Figure 4.1 that users known for certain topical tags often appear in clusters, i.e., they concentrate in several discrete regions. In light of this, we introduce a region-dependent variable to represent a region's affinity over tags.

Concretely, we assume that the geographical space is partitioned into $\mathcal{I}$ regions. Each user can then be assigned into the region where she belongs, which gives a matrix $\mathbf{R^{ul}}$. Each region corresponds to a K-dimensional location latent factor, denoted θ_l, indicating the region's affinity over tags. If we consider that each r_{ut} in the matrix R^p is only dependent on user's location, then r_{ut} can be generated through the following Poisson distribution:

$$r_{ut} \sim Poisson(\theta_{l_u}^T \theta_t) \tag{4.5}$$

where l_u represents the region where user u belongs, and the intensity parameter is determined by the inner product of the corresponding location and tag latent factor. Similarly,



(a) entrepreneur

(b) politics

Figure 4.1: Geo-spatial distribution of users who are known for “entrepreneur” and “politics”.

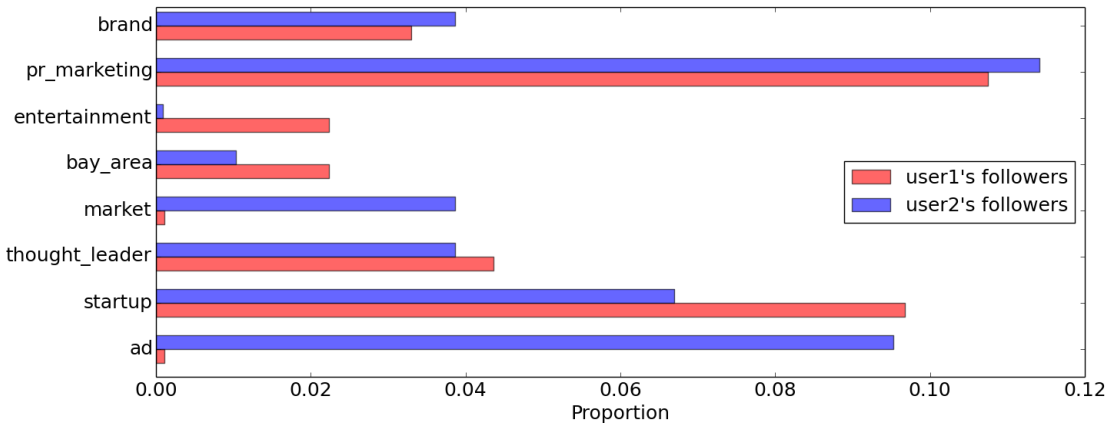


Figure 4.2: Proportion of followers who have the same tag in their known-for profile as the users. user1: red; user2: blue.

each component of θ_l has a conjugate Gamma prior as specified in Equation 4.2. Thus, through the explicit handling of tag's dependence over locations, it is expected that θ_l could indicate which tags are mostly dominant for the corresponding region.

Social influence factor. Homophily in social networks [103] suggests that people tend to connect with others who are similar to themselves. Indeed, there already exists some works [60, 61] exploiting social relations for item recommendations. Here, we would like to explore how social relations may benefit user known-for profile discovery. An immediate problem is that user's social relations are often noisy, i.e., if a friend of a user is known for a tag, it does not necessarily mean that the user is also known for the tag. In Figure 4.2, we randomly select two Twitter users who are known for eight tags, and examine the proportion of followers who have the same tag in their known-for profiles. We can see that even though user1 is known for "ad", very few followers of the user are also known for "ad". On the contrary, a significant portion of followers of user2 are known for "ad". This indicates that users may have different degree of social influence.

Thus, to capture this observation, we introduce a social influence parameter θ_f for each

friend f of all users, indicating the degree of social influence of this friend on any user who follows her. Thus, if we have a total of M friends, the influence parameters constitute a vector of Θ^f with dimension M. Then, given user u's friends $\mathcal{N}_u$, if we consider that each r_{ut} in the matrix $\mathbf{R^p}$ is only dependent on user's friends' known-for profiles, then r_{ut} could be generated through the following Poisson distribution:

$$r_{ut} \sim Poisson(\sum_{f \in \mathcal{N}_u} \theta_f r_{ft}) \tag{4.6}$$

where r_{ft} is the element in the matrix $\mathbf{R^{ft}}$ indicating if friend f has the tag t. The intensity parameter is obtained by aggregating all of the friends' latent influence for the user.

However, since each friend may be labeled with multiple tags, it is very likely that this friend has different levels of influence for different tags, as shown in Figure 4.2. In the figure, many people following user2 have the tag "pr marketing" and "ad". Thus, although user2 is also known for "entertainment", it is obvious that this user is more influential on "pr marketing" and "ad". As a result, if a new user starts to follow her, it is very likely that this user follows her because she is also known for "pr marketing" or "ad".

Hence, to capture this intuition, we assume that for each friend f of all users' friends, there is a latent factor θ_{ft} for each tag t. This parameter is used to capture the degree of influence that f has over t. Thus, if we have a total of N tags, the influence parameters constitute a M by N matrix Θ^{ft}. Thus, given user u's friends $\mathcal{N}_u$, similarly as above, r_{ut} can be generated through the following:

$$r_{ut} \sim Poisson(\sum_{f \in \mathcal{N}_u} \theta_{ft} r_{ft}) \tag{4.7}$$

where the intensity parameter is obtained by aggregating all of the friends' latent influence for this user and the corresponding tag. Consequently, Equation 4.7 is able to model the

above scenario where the new user is more likely to have the tag "pr marketing" or "ad" instead of "entertainment" if she follows user2. Similarly, each nonzero component of Θ^{ft} is put on Gamma priors for inference.

Integrating contextual factors. So far, we can generate independent known-for profiles through user's content, geographical coordinates and social relations. However, it is natural to combine all types of contextual influence to form a better representation for users. Since the sum of Poisson distributions is still a Poisson distribution, we can combine the generative process specified by Equation 4.4, 4.5 and 4.7 together to obtain the overall intensity parameter of the sum Poisson distribution as:

$$\lambda_{ut} = \alpha\theta_u^T\theta_t + \beta\theta_{l_u}^T\theta_t + \gamma \sum_{f \in \mathcal{N}_u} \theta_{ft} r_{ft} \tag{4.8}$$

where α, β and γ are the tradeoff weights for content, geo-spatial and social contributions, respectively. The entire generative process of BCPF is described in Algorithm 1.

4.3.2 Prediction

We have specified a Bayesian probabilistic model over the latent parameters Θ^u, Θ^w, Θ^l, Θ^{ft} and Θ^t, and the observed discrete data matrix $\mathbf{R^p}$, $\mathbf{R^c}$, $\mathbf{R^{ul}}$ and $\mathbf{R^{ft}}$. To predict the known-for profiles for new users, we need to estimate the posterior distributions of the latent parameters $p(\Theta^u, \Theta^w, \Theta^l, \Theta^{ft}, \Theta^t | \mathbf{R^p}, \mathbf{R^c}, \mathbf{R^{ul}}, \mathbf{R^{ft}})$ given the observed data. Once we have the posterior distributions of the latent factors, we can predict the known-for profiles for new users with the expectation of the sum of weighted Poisson distributions. This leads to computing the expectation of the intensity parameter λ_{ut}, which gives:

$$E(r_{ut}) = \alpha E(\theta_u)^T E(\theta_t) + \beta E(\theta_{l_u})^T E(\theta_t) + \gamma \sum_{f \in \mathcal{N}_u} E(\theta_{ft}) r_{ft}$$

Algorithm 1: Generative Process for BCPF

Input: hyper-parameter λ_a^w, λ_b^w, λ_a^u, λ_b^u, λ_a^l, λ_b^l, λ_a^t, λ_b^t, λ_a^{ft}, λ_b^{ft}

for each word w **do**
  for k = 1 to K **do**
    Draw $\theta_{wk} \sim$ Gamma(λ_a^w,λ_b^w)
  end for
end for
for each user u **do**
  for k = 1 to K **do**
    Draw $\theta_{uk} \sim$ Gamma(λ_a^u,λ_b^u)
  end for
end for
for each user u **do**
  for each word w **do**
    Draw $r_{uw} \sim$ Poisson($\theta_u^T \theta_w$)
  end for
end for
for each location l **do**
  for k = 1 to K **do**
    Draw $\theta_{lk} \sim$ Gamma(λ_a^l,λ_b^l)
  end for
end for
for each friend f **do**
  for each tag t of f **do**
    Draw $\theta_{ft} \sim$ Gamma(λ_a^{ft},λ_b^{ft})
  end for
end for
for each tag t **do**
  for k = 1 to K **do**
    Draw $\theta_{tk} \sim$ Gamma(λ_a^t,λ_b^t)
  end for
end for
for each user u **do**
  for each tag t of u **do**
    Draw $r_{ut} \sim$ Poisson(λ_{ut}), where λ_{ut} is determined by Equation 4.8.
  end for
end for

where all expectations are with respect to the posterior distributions. By making use of a new user's contextual information, we can predict the ranking of the tags for this user by their expected occurrence.

4.3.3 Learning with Variational Inference

Since it is not tractable to compute the exact posterior of the latent factors Θ, we propose to use variational methods [104] for approximate inference. Variational inference casts the approximation process as an optimization problem. By defining a freely parameterized family of distributions over latent variables, variational methods seek to fit its parameters so as to minimize the KL-divergence between the defined distribution and the posterior distribution. In our case, variational inference solves the following minimization problem:

$$q^*(\Theta) = argmin_q KL(q(\Theta)||p(\Theta|\mathbf{R^p}, \mathbf{R^c}, \mathbf{R^{ul}}, \mathbf{R^{ft}})) \tag{4.9}$$

where $q^*(\Theta)$ is the optimized variational distribution that is used as the proxy for the exact posterior. To facilitate the inference, we first introduce several auxiliary latent variables [102] for Equation 4.3 and 4.8. Specifically, let's assume $z_{uwk} \sim Poisson(\theta_{uk}\theta_{wk})$, where $\sum_k z_{uwk} = r_{uw}$, and $z^c_{utk} \sim Poisson(\alpha\theta_{uk}\theta_{wk})$, $z^l_{utk} \sim Poisson(\beta\theta_{l_uk}\theta_{tk})$, $z_{utf} \sim Poisson(\gamma\theta_{ft}r_{ft})$, where $\sum_k z^c_{utk} + \sum_k z^l_{utk} + \sum_{f\in\mathcal{N}_u} z_{utf} = r_{ut}$. Since the sum of a set of Poisson random variables is still a Poisson distribution, these auxiliary variables can still preserve the marginal distribution of r_{uw} and r_{ut} when marginalized out. Thus, our latent variables include latent parameters Θ and auxiliary latent variables Z_{uw}, Z_{ut}.

Before solving Equation 4.9, we need to derive the complete conditionals, i.e., the conditional distributions of a latent variable given all other variables, for each of Θ, Z_{uw} and Z_{ut}. Since each of Θ has conjugate Gamma priors, the complete conditionals for Θ are also Gamma distributions. For auxiliary latent variables, we follow the conclusion from [105] that given a vector Z_{ut} of Poisson distributed count, Z_{ut} is distributed as a

multinomial conditioned upon the observed sum r_{ut}. Similarly, Z_{uw} is also a multinomial given r_{uw}. Thus, all complete conditionals can be derived.

Variational inference assumes $q(\Theta, Z_{uw}, Z_{ut})$ is in the same exponential family with the complete conditionals [58]. Thus, we employ the following mean field variational family, where each latent variable is independent with each other and governed by its own variational parameters:

$$
\begin{aligned}
q(\Theta^u, \Theta^w, \Theta^l, \Theta^{ft}, \Theta^t, Z_{uw}, Z_{ut}) = & \prod_{u,k} q(\theta_{uk}|\eta^a_{uk}, \eta^b_{uk}) \\
& \prod_{w,k} q(\theta_{wk}|\eta^a_{wk}, \eta^b_{wk}) \prod_{l,k} q(\theta_{lk}|\eta^a_{lk}, \eta^b_{lk}) \prod_{t,k} q(\theta_{tk}|\eta^a_{tk}, \eta^b_{tk}) \\
& \prod_{f,t} q(\theta_{ft}|\eta^a_{ft}, \eta^b_{ft}) \prod_{u,w,k} q(z_{uwk}|\phi_{uwk}) \prod_{u,t,k} q(z^c_{utk}|\phi^c_{utk}) \\
& \prod_{u,t,k} q(z^l_{utk}|\phi^l_{utk}) \prod_{u,t,f \in \mathcal{N}_u} q(z_{utk}|\phi_{utf})
\end{aligned}
$$

where $\eta^a_{\cdot}$ and $\eta^b_{\cdot}$ are the shape and rate parameter of the variational Gamma distributions; and $\phi_{ut\cdot}$ is the parameter of the variational Multinomial distributions. To obtain optimal values for $\eta^a_{\cdot}$, $\eta^b_{\cdot}$, $\phi_{uw\cdot}$ and $\phi_{ut\cdot}$, we need to solve the minimization in Equation 4.9. Since variational inference requires that the natural parameter of each $q(\cdot)$ is the expectation of the natural parameter of the corresponding complete conditional under $q(\cdot)$ [106], we just need to compute the expectation of the natural parameters of each complete conditional. To give an example, we show how to compute parameter η^a_{tk} and η^b_{tk}. The expectation of the natural parameters on the complete conditional of θ_{tk} gives:

$$
\begin{aligned}
\eta^a_{tk} &= \lambda^a_t + \sum_u r_{ut} \phi^c_{utk} + \sum_u r_{ut} \phi^l_{utk} \\
\eta^b_{tk} &= \lambda^b_t + \alpha \sum_u \frac{\eta^a_{uk}}{\eta^b_{uk}} + \beta \sum_u \frac{\eta^a_{l_u k}}{\eta^b_{l_u k}}
\end{aligned}
\tag{4.10}
$$

where we have applied the fact that the expectation of a Gamma random variable is equal to the ratio of the shape parameter over the rate parameter. Similarly, we can derive the formula to compute other latent parameters. For auxiliary latent variable Z_{ut}, the expectation of its natural parameter of the complete conditional leads to

$$\begin{aligned}
\phi^c_{utk} &\sim \alpha \exp(\psi(\eta^a_{uk}) - \ln(\eta^b_{uk}) + \psi(\eta^a_{tk}) - \ln(\eta^b_{tk})) \\
\phi^l_{utk} &\sim \beta \exp(\psi(\eta^a_{l_u k}) - \ln(\eta^b_{l_u k}) + \psi(\eta^a_{tk}) - \ln(\eta^b_{tk})) \\
\phi_{utf} &\sim \gamma r_{ft} \exp(\psi(\eta^a_{ft}) - \ln(\eta^b_{ft}))
\end{aligned} \tag{4.11}$$

where $\psi(\cdot)$ is the digamma function; and we have applied the fact that $E(\ln(X)) = \psi(\eta^a) - \ln(\eta^b)$ when $X \sim Gamma(\eta^a, \eta^b)$. Similarly, we could obtain the formula for Z_{uw}. We then use coordinate ascent to update each variational parameter by turns to obtain the locally optimal values. The overall update algorithm is shown in Algorithm 2.

4.3.4 Complexity Analysis

One of the benefits using Poisson distribution to model discrete data is that the algorithm only iterates through non-zero occurrence of data matrix [102]. Thus, the computational cost can be dramatically reduced if the data matrix is very sparse, as is the case for our datasets. The complexity of Algorithm 2 is determined by coordinate ascent optimization, which essentially includes two parts: (i) optimizing the variational parameters for auxiliary variables Z_{uw} and Z_{ut}; and (ii) optimizing the variational parameters for Θ.

Suppose the number of non-zeros in $\mathbf{R^p}$ and $\mathbf{R^c}$ is $|\mathbf{R^p}|$ and $|\mathbf{R^c}|$, respectively. Let's also assume the average number of friends for each user is $\mathcal{N}_u$. Then the complexity for the update of $\phi_{uw\cdot}$ is $K|\mathbf{R^c}|$. For the update of $\phi_{ut\cdot}$, since it considers three factors, the complexity of the update is $(2K + \mathcal{N}_u)|\mathbf{R^p}|$. For the update of Θ^{ft}, the complexity must consider the average number of friends $\mathcal{N}_u$ each user has, and is estimated to be $\mathcal{N}_u|\mathbf{R^{ft}}|$, where $|\mathbf{R^{ft}}|$ represents the number of non-zeros in the data matrix $\mathbf{R^{ft}}$. For other parameter

Algorithm 2: Variational inference for BCPF

Input: hyper-parameter λ_a^w, λ_b^w, λ_a^u, λ_b^u, λ_a^l, λ_b^l, λ_a^t, λ_b^t, λ_a^{ft}, λ_b^{ft}, data matrix $\mathbf{R^p}$, $\mathbf{R^c}$, $\mathbf{R^{ul}}$, $\mathbf{R^{ft}}$

repeat

Initialize all variational parameters $\eta_{\cdot}^a$, $\eta_{\cdot}^b$, $\phi_{ut\cdot}$, $\phi_{uw\cdot}$

for each r_{uw}

Update ϕ_{uwk} for each k with

$\phi_{uwk} \sim \exp(\psi(\eta_{uk}^a) - \ln(\eta_{uk}^b) + \psi(\eta_{wk}^a) - \ln(\eta_{wk}^b))$

Normalize $\phi_{uw\cdot}$

for each r_{ut}

Update ϕ_{utk}^c for each k, ϕ_{utk}^l for each k, and

ϕ_{utf} for each $f \in \mathcal{N}_u$ with Equation 4.11

Normalize $\phi_{ut\cdot}$

for each w, update η_{wk}^a and η_{wk}^b for each k with

$\eta_{wk}^a = \lambda_w^a + \sum_u r_{uw}\phi_{uwk}$, $\eta_{wk}^b = \lambda_w^b + \sum_u \frac{\eta_{uk}^a}{\eta_{uk}^b}$

for each u, update η_{uk}^a and η_{uk}^b for each k with

$\eta_{uk}^a = \lambda_u^a + \sum_w r_{uw}\phi_{uwk} + \sum_t r_{ut}\phi_{utk}^c$

$\eta_{uk}^b = \lambda_u^b + \sum_w \frac{\eta_{wk}^a}{\eta_{wk}^b} + \sum_t \frac{\eta_{tk}^a}{\eta_{tk}^b}$

for each l, update η_{lk}^a and η_{lk}^b for each k with

$\eta_{lk}^a = \lambda_l^a + \sum_t \sum_u I(l_u = l) r_{ut}\phi_{utk}^l$

$\eta_{lk}^b = \lambda_l^b + \sum_t \sum_u I(l_u = l) \frac{\eta_{tk}^a}{\eta_{tk}^b}$

for each t, update η_{tk}^a and η_{tk}^b for each k with Equation 4.10

for each f and each t, where $r_{ft} > 0$

Update η_{ft}^a and η_{ft}^b with

$\eta_{ft}^a = \lambda_{ft}^a + \sum_u I(f \in \mathcal{N}_u) r_{ut}\phi_{utf}$

$\eta_{ft}^b = \lambda_{ft}^b + \sum_u I(f \in \mathcal{N}_u) r_{ft}$

until Converge

Table 4.1: Geo-tagged Twitter data.

Loc	# of users	# of tweets	# of following links
US	10,552	317,436	24,676
World	19,776	594,929	30,853
Loc	# of tags	# of records	sparsity
US	1,011	85,994	0.81%
World	1,456	136,625	0.47%

update, their complexities are much lower than the above parameter update. Thus, the total computational complexity for one iteration of coordinate ascent can be estimated to be $K|\mathbf{R^c}| + (2K + \mathcal{N}_u)|\mathbf{R^p}| + \mathcal{N}_u|\mathbf{R^{ft}}|$. Since usually $\mathcal{N}_u \ll K$ and $|\mathbf{R^{ft}}| < |\mathbf{R^p}|$, the complexity is determined by $\mathcal{O}(K \cdot max\{|\mathbf{R^c}|, |\mathbf{R^p}|\})$. Therefore, the learning algorithm is quite efficient for sparse data matrix.

4.4 Experimental Evaluation

In this section, we conduct several experiments to evaluate the proposed BCPF for user known-for profile discovery.

Data. Our data is based on a sample of about 12 million Twitter lists collected from 2013 to 2014. Twitter lists [107, 19] are crowd-generated, for which a labeler can put a user on a tagged list, if the labeler thinks the user is known for the topic indicated by the tags. Thus, if a user is labeled by different labelers with certain tags, for example, "chef" and "recipe", then we consider this user is known for these topics. In our experiments, we use the threshold of three labelers to determine the existence of a tag in user's known-for profile. In addition, we also filter out infrequent tags with fewer than 20 users to focus on quality tags. We randomly sample two datasets (see Table 6.2), one in the US and one across the world. We also crawled about 30 recent tweets for each user and sampled her social relations.

Experimental Setup. For evaluation, we randomly partition all users into 60% for training, 30% for testing, and 10% for cross-validation. For the dimension K of the latent factors, we empirically select 100 for all methods. For the tradeoff weights α, β and γ in Equation 4.8 for BCPF, we set them to 1, 1.2 and 0.2 via cross-validation. To initialize the hyper-parameters of the Gamma priors in BCPF, we follow [58] and set them to 0.3 plus a small random variations on geo-spatial and social latent factors for sparse solutions; for user content and word variational parameters, we adopt LDA [108] and use document topic distribution and topic word distribution to initialize η_{uk} and η_{wk}, respectively. In the modeling of geo-spatial influence, it is assumed that the geographical space is partitioned into discrete regions. Here, we adopt k-means clustering instead of a simple gridding by clustering users' geo-coordinates. We choose a clustering-based approach because the geo-spatial distribution of users exhibits a clustering effect, as shown in Figure 4.1. For evaluation metrics, we adopt Precision@k (Prec@k) and Recall@k (Rec@k), defined as below:

$$Prec@k = \frac{1}{N}\sum_{i=1}^{N}\frac{|pred_k(u_i) \cap p_u|}{k}$$

$$Rec@k = \frac{1}{N}\sum_{i=1}^{N}\frac{|pred_k(u_i) \cap p_u|}{|p_u|}$$

Prec@k measures the percentage of the correctly identified tags over the top k predicted tags. Rec@k represents what percentage of true tags can emerge in top k predicted tags. Both measures are averaged over all testing users. In our experiments, we select k to be 5 and 10.

Baselines. We compare BCPF with the following baselines:

- *k-Nearest Neighbors*. In this baseline, we extract content features from user's sampled tweets, and apply kNN to find n most similar users to the testing user u. We then compute tag's score according to $s_{ut} = \sum_i^n w_{ui} r_{it}$, where w_{ui} is the similarity between

user u and i, and select top k tags for prediction. We use two approaches to compute textual similarity, one with the bag-of-words model (kNN-BoW), another with the topic model (kNN-LDA) [108].

- *One-Vs-Rest multi-label ranking* [109]. We train a One-Vs-Rest multi-label classifier with bag-of-words on user's tweets. Logistic regression is used as the classification method to have a probabilistic output for ranking tags.
- *Wsabie* [110]. Wsabie is an embedding-based model which learns a mapping from a feature space to the joint space with the tags. Here, we train the model with WARP loss on bags-of-words of user's tweets, and use the learned user and tag embeddings for ranking tags.
- *Content-based Poisson Factorization (C-PF)*. This is the method where we only keep the content factor in BCPF. It can be considered a form of CTPF proposed in [58] where we treat user's tweets as documents, and user's known-for profiles as binary ratings. We also ignore the topic offsets since the testing users are in complete cold-start situation.
- *Geo-spatial CPF (GC-PF)*. This is the method in which we keep both content and geo-spatial factor in BCPF.
- *Social influenced CPF (SC-PF)*. This method keeps both content and social influence factor in BCPF.

4.4.1 Effectiveness of BCPF

How well does the proposed BCPF perform compared to alternative baselines? In Table 4.2, we report Prec@k and Rec@k for all methods. Overall, BCPF gives the best results for all metrics and both datasets. Specifically, it gives an average improvement of 17.5% in precision and 20.9% in recall over the best content based method C-PF. This indicates the superiority of BCPF by exploiting geo-spatial and social influence factors other than user's content.

Table 4.2: Comparison of performance with alternative methods. BCPF generally gives the best performance by integrating the contextual influence of textual, geo-spatial and social factors, and these factors are able to complement each other.

Method	US				World			
	Precision		Recall		Precision		Recall	
	@5	@10	@5	@10	@5	@10	@5	@10
kNN-BoW	0.114	0.098	0.120	0.181	0.083	0.072	0.099	0.151
kNN-LDA	0.150	0.127	0.155	0.226	0.120	0.100	0.137	0.200
One-Vs-Rest	0.166	0.140	0.176	0.250	0.136	0.112	0.161	0.227
Wsabie	0.159	0.137	0.169	0.244	0.116	0.101	0.134	0.201
C-PF	0.188	0.160	0.194	0.279	0.147	0.126	0.168	0.246
GC-PF	0.199	0.174	0.212	0.312	0.159	0.138	0.186	0.277
SC-PF	0.209	0.177	0.217	0.309	0.165	0.138	0.189	0.270
BCPF	**0.222**	**0.188**	**0.240**	**0.336**	**0.173**	**0.147**	**0.201**	**0.295**

In content-based methods, neighborhood-based methods generally provide the worst performance of all since only local information can be used in prediction. Here, LDA based kNN performs better than BoW based kNN, since LDA can take advantage of the global topic information in users' posts. Supervised methods, however, give relatively better performance than neighborhood methods, as shown by One-Vs-Rest multi-label ranking and Wsabie. Both methods are trained with multiple labeled tags for each user on the BoW features, where Wsabie obtains embeddings for each user and tag. However, One-Vs-Rest clearly outperforms Wsabie, although Wsabie is more efficient and requires less computation time. C-PF, however, gives the best performance of all content-based methods, indicating the superiority of joint modeling by learning labeled users' profiles on content features and unlabeled users' texts.

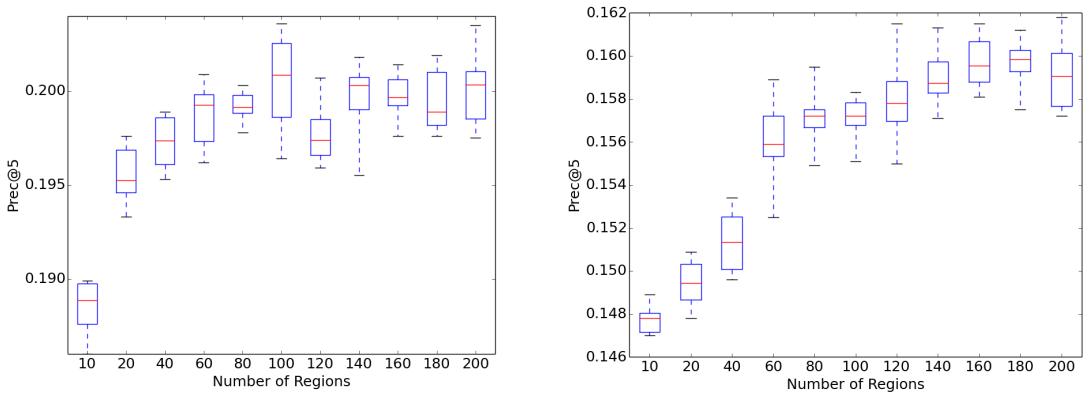
Furthermore, we can see from the table that both GC-PF and SC-PF provides better performance against only content-based C-PF, respectively. Specifically, GC-PF gives an average improvement of 8.07% in precision and 11.1% in recall, respectively. SC-PF gives an average improvement of 10.9% in precision and 11.2% in recall, respectively. This in-

dicates that geo-spatial features and parameterized social influence can both improve the identification of user's known-for profile. Given the overall improvement of the combination of these features, we can also conclude that these factors are able to complement each other.

4.4.2 Geo-spatial Factor Analysis

In the modeling of geo-spatial influence on user's known-for profile, we discretize the US/world area into geographical regions, and treat user's location as the region where she belongs. An important question here is how to select the number of regions $\mathcal{I}$. As we can imagine, if $\mathcal{I}$ is too small, the model may not be able to reflect the geo-spatial distribution of user's known-for profile; if it is too large, regions may be too small, thus leads to sparse observations. To that end, we select $\mathcal{I}$ for k-means clustering from 10 to 200, and run GC-PF with each value for ten random initializations of parameters. In Figure 4.3, we show how precision and recall changes with respect to the number of regions for both datasets. We can see that as $\mathcal{I}$ goes up, precision generally also increases. However, the performance plateaus when it is large. We attribute this to: first, when the number of regions increases, we obtain finer-grained geo-spatial characterization of tags; second, when it is large enough, a finer-grained discretization does not help distinguish tag's local distributions. It may even result in sparse counts in regions and also increases model complexity, thus degrading the final performance. Note that it requires less number of regions, i.e., larger partition of areas, for the US dataset than world to get to the best average performance. The reason is that the world data may be more diverse in terms of the geo-spatial distribution of user known-for profiles, thus requiring finer-grained geo-segments to reach the best performance.

To further examine the inferred region factors, we show in Table 4.3 the top ranking tags associated with selected areas. In particular, we compute the affinity score between a



(a) US

(b) World

Figure 4.3: Box plot for precision@5 with respect to the number of regions.

Table 4.3: Top ranking tags for different areas obtained from latent location factors and tag factors.

Los Angeles	D.C.	SF	Chicago
la	dc	sf	chicago
celebrity	politics	geek	chi
star	progressive	bay area	illinois
tv	baltimore	dev	pr
artist	us	technology	advertising
actor	conservative	mobile	social media
famous people	government	startup	marketing
entertainment	politico	software	marketer

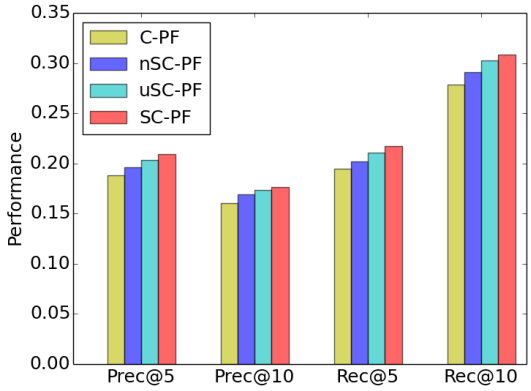
location and a tag by taking the expected inner product of the latent factor of the location and the tag, denoted as $E(\theta_l^T \theta_t)$. We can see from the table that, generally, the top ranking tags can reflect the common knowledge of the characteristics of those selected areas. For example, it is very likely that a user from the Washington D.C. area is well known for her political activities/comments. Thus, these inferred region factors are capable of nudging user's known-for profile toward her region characteristics.

4.4.3 Social Influence Factor Analysis

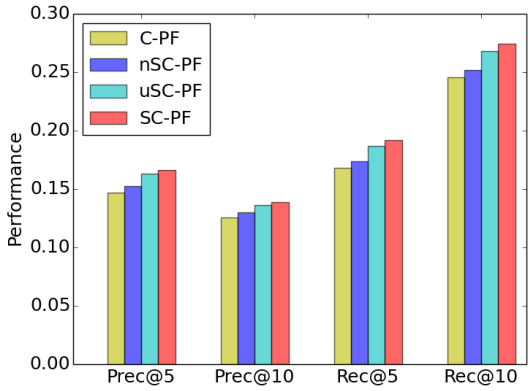
We have seen from the previous experiments that social influence plays an importance role in representing user's known-for profile by looking at her connections. In order to get a further understanding of how social influence varies for different friends and tags, we consider the following methods using social influence: (i) naive SC-PF (nSC-PF), where we replace θ_f with a single parameter θ_{naive} in Equation 4.6 for all friends and tags; (ii) user weighted SC-PF (uSC-PF), which models social influence by a vector of dimension M, with each element indicating the level of a friend's social influence for all tags (see Equation 4.6); and (iii) SC-PF, which models each friend's social influence for every tag of hers by matrix Θ_{ft}. We compare these methods with C-PF to see which method models social influence the best. Note that we also include the content factor as a basic representation for users since not all users have social connections.

As we can see from Figure 4.4, overall, social connection is able to improve the performance for predicting user's known-for profile. Specifically, we can make the following two conclusions: (i) it is generally better to differentiate social influence for different friends, rather than treating all friends with the same influence; (ii) it is generally better to differentiate social influence for different tags of each friend, rather than treating all tags of each friend with the same influence.

Table 4.4 shows the inferred social influence for different users and their top-ranking tags, which further exemplifies the observation that users may have different social influence on their followers with respect to different tags. In particular, user @UnderTheBar shows little influence on his followers' known-for profiles even if he may have tens of thousands of followers. This indicates that users who follow @UnderTheBar are not necessarily also known for "entrepreneur" or "health". On the contrary, @mollyblock has much larger social influence on her followers for her tag "marketer", indicating that users



(a) US



(b) World

Figure 4.4: Comparative performance of the methods modeling social influence. SC-PF performs best by making it dependent on both friends and tags.

Table 4.4: User's inferred social influence θ_f by uSC-PF and social influence θ_{ft} for top ranking tags by SC-PF.

User	θ_f	top-1/θ_{ft}	top-2/θ_{ft}
@UnderTheBar	0.00290	entrepreneur 0.0230	health 0.0226
@skwigg	0.00425	fit 0.0217	nutrition 0.0190
@mollyblock	0.0143	marketer 0.104	pr 0.0226
@bikehugger	0.210	bicycle 0.435	cycling 0.303

following @mollyblock are more likely to also be known for "marketer", even if @mollyblock has fewer followers than @UnderTheBar. One explanation for this observation is that people connect with others for different reasons on social media, and that popularity itself is not necessarily an indicator of one's social influence.

4.5 Conclusion

In this section, we tackled the problem of discovering what users are known for in social media. By integrating user's textual, geo-spatial and social influence, we proposed Bayesian Contextual Poisson Factorization to overcome the noisiness of user's posting activities and social relations. Experimental results showed that BCPF can improve known-for prediction by 17.5% in precision and 20.9% in recall on average. We also showed that user's connections have varying social influence for different topics, confirming our fine-grained modeling of social influence by making it dependent on both users and topics. In our future work, we are interested in exploring the impact of additional contextual signals beyond the textual, geo-spatial and social influence studied here.

5. USER PROFILING: DISCOVERING USER'S OPINION BIAS*

In this section, we move to the second aspect of user profiling in the dissertation: user's opinion bias. Specifically, we will show how we can semi-automatically discover user's opinion bias through a very small amount of human guidance, and how we can exploit user's opinion bias for personalization.

5.1 Introduction

Social media has increasingly become a popular and important platform for "regular" people to express their opinions, without the need to rely on expensive and fundamentally limited conduits like newspapers and broadcast television. These opinions can be expressed on a variety of themes including politically-charged topics like abortion and gun control as well as fun (but heated) rivalries like android vs. iOS and Cowboys vs. 49ers. Our interest in this section is in creating a flexible tool for discovering and tracking the themes of opinion bias around these topics, the strong partisans who drive the online discussion, and the degree of opinion bias of "regular" social media participants, to determine to what degree particular participants support or oppose a topic of interest.

However, assessing topic-sensitive opinion bias is challenging. First, the opinion bias of "regular" users may not be as pronounced as prominent figures, so discerning this bias will require special care. Second, how opinion bias manifests will inevitably vary by topic, so a system should be adaptable to each topic. Third, the themes by which people express their opinions may change over time depending on the circumstances (e.g., gun control debates may take different forms based on the ebb and flow of elections, recent shooting

*Reprinted with permission from "BiasWatch: A Lightweight System for Discovering and Tracking Topic-Sensitive Opinion Bias in Social Media" by Haokai Lu, James Caverlee and Wei Niu, 2015. *Proceedings of the 24th ACM international conference on Information and Knowledge Management*, 213-222, Copyright 2015 by ACM. DOI: http://dx.doi.org/10.1145/2806416.2806573.

incidents, and so forth). As a result, assessing bias should be adaptive to these temporal changes.

Hence in this section, we develop a lightweight system – BiasWatch – for discovering and tracking opinion bias in social media. BiasWatch begins by taking just two hand-picked seeds to characterize the topic-space (e.g., "pro-choice" and "pro-life" for abortion) as *weak labels* to bootstrap the opinion bias framework. Concretely, we leverage these hand-picked seeds to identify other emerging (and often unknown) themes in social media, reflecting changes in discourse as new arguments and issues arise and fade from public view (e.g., an upcoming election, a contentious news story). We propose and evaluate two approaches for expanding the hand-picked seeds in the context of Twitter to identify supporting and opposing hashtags – one based on co-occurrence and one on signed information gain. We use these discovered hashtags to identify strong topic-based partisans (what we dub *anchors*). Based on the social and information networks around these anchors, we propose an efficient opinion-bias propagation method to determine user's opinion bias – based on both content and retweeting similarity – and embed this method in an optimization framework for estimating the topic-sensitive bias of social media participants. In summary, this section makes the following contributions:

- First, we build a systematic framework – BiasWatch – to discover biased themes and estimate user-based opinion bias quantitatively under the context of controversial topics in social media. We propose an efficient optimization scheme – called User-guided Opinion Propagation [UOP] – to propagate opinion bias. By feeding just two opposing hashtags, the system can discover bias-related hashtags, find bias anchors, and assess the degree of bias for "regular" users who tweet about controversial topics.
- Second, we evaluate the estimation of users' opinion bias by comparing the quality of the proposed opinion bias approach versus several alternative approaches over multiple

Twitter datasets. Overall, we see a significant improvement of 20.0% in accuracy and 28.6% in AUC on average over the next-best method.

- Third, we study the effect of different approaches for biased theme discovery to measure the impact of newly discovered biased hashtags as additional supervision. We observe that the newly discovered hashtags are often associated with the underlying community of similar opinion bias, and that they temporally fluctuate due to the impact of new controversial events.
- Finally, we demonstrate how these inferred opinion bias scores can be integrated into user recommendation by giving similar-minded users a higher ranking. We show that the integration can improve the recommendation performance by 26.3% in precision@20 and 13.8% in MAP@20. This result implicitly confirms the principle of homophily in the context of opinion bias, and demonstrates how topic-sensitive opinion bias can enrich user modeling in social media.

5.2 Lightweight Bias Discovery

Problem Statement. We assume there exists a set of users $U = \{u_1, u_2, ..., u_n\}$ sampled from Twitter. Each user has their corresponding tweets $D = \{d_1, d_2, ..., d_n\}$ related to a controversial topic T, where d_i is a collection of tweets by u_i. Since a person's opinion bias represents the intrinsic tendency that she chooses to support or oppose a concept under a controversial context, we choose to quantize the degree of her opinion bias by a numeric score ranging from -1 to 1. Specifically, we assign $B = \{b_1, b_2, ..., b_n\}$ for each user in U, respectively, where $b_i \in [-1, 1]$. When b_i is close to 1, it denotes that user u_i has a strong positive standing toward the topic; when b_i is close to -1, it represents the opposite. Thus, given a controversial topic T, a sampled set of users U and their on-topic tweets D, we identify the following tasks of the system framework: (i) Discovering biased themes that are discussed by opposing sides of users. We denote P as the set of positive themes and N

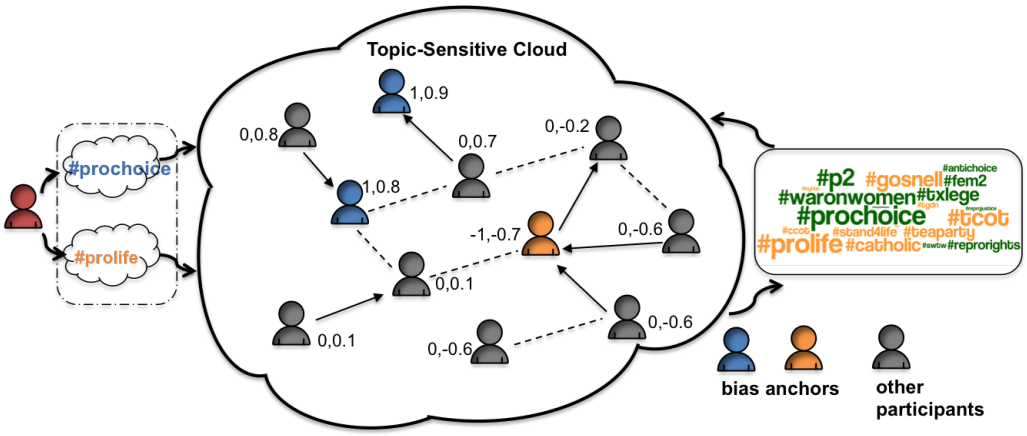


Figure 5.1: Overall BiasWatch Framework

as the set of negative themes; (ii) Finding bias anchors who show strong degree of opinion bias, which we denote as U_{anchor}; (iii) Determining "regular" participants' opinion bias B.

Overall Approach. In order to tackle these tasks, we propose a lightweight framework that propagates opinion bias scores based only on a few hand-picked seeds that characterize the topic-space (e.g., "pro-choice" and "pro-life" for abortion). The BiasWatch framework, illustrated in Figure 6.3, takes as input these hand-picked seeds and then proceeds through the following three key steps:

- **Finding Bias Anchors.** This first step identifies topic-based partisans whose opinion bias is strongly revealed through their choice of hashtags. We develop two automatic approaches to: (i) identify biased themes in the form of hashtags through initial seeds; (ii) expand the pools of bias anchors with these identified biased themes.
- **Propagating Bias.** This second step builds a user similarity network around these expanded anchors and other "regular" participants, and propagates bias along this network. The edges here measure the similarities of two users through content and link features from tweets.
- **Noise-Aware Optimization.** Lastly, we propose to embed the previous two steps into

a noise-aware optimization framework where anchors' opinion bias can be effectively propagated to each "regular" participant throughout the network. A key facet is that this optimization is tolerant of noisy labels on the initial bias anchors, so that initial errors made in identifying bias anchors need not lead to cascading errors in "regular" participants.

5.2.1 Finding Bias Anchors

Our first challenge is to identify strong topic-based partisans (what we dub *anchors*). These anchors serve as the basis for propagating opinion bias throughout the social and information network. One reasonable method for identifying anchors is to manually label a number of users, among whom we hope that there exist a portion of users whose opinion bias is clearly shown. However, there are two disadvantages of this approach: (i) it is potentially expensive and time-consuming; and (ii) because of the random nature of labeling, typically, we have to label many users whose opinion bias is not clear or neutral in order to obtain anchors.

To overcome these difficulties, we propose to exploit crowd-generated hashtags in Twitter. Hashtags are often used to tag tweets with a specific topic for better discoverability or to indicate the community to which the tweets are posted [111]. Some hashtags may be viewed as a rich source of expressing opinions [112], potentially indicating user's opinion bias. For example, some most popular hashtags for "gun control" include #guncontrolnow and #2ndamendment, which reveal strong user bias, with the former often used by supporters of gun control and the latter by opponents. Hashtags of this nature can be essentially considered as *weak labels* of the polarity of tweets with respect to a controversial topic.

Hence, to find bias anchors we first seek to identify a candidate set of hashtags that provide support for the topic and a candidate set of hashtags that express opposition. Since

Table 5.1: Top ten themes at different times for "fracking" discovered by seed expansion; red for pro-fracking; blue for anti-fracking.

Dec 2012	Mar 2013	June 2013	Sept 2013
#shale	#dontfrackny	#shale	#balcombe
#natgas	#shale	#energy	#shale
#oil	#energy	#natgas	#frackoff
#energy	#oil	#oil	#energy
#gas	#gas	#gas	#greatgasgala
#tcot	#natgas	#banfracking	#natgas
#jobs	#frack	#fracked	#banfracking
#frack	#banfracking	#frack	#gas
#marcellus	#nokxl	#tcot	#oil
#naturalgas	#tcot	#dontfrackny	#frack

the hashtags used to express an opinion may change over time as new issues arise and fade from public view (e.g., an upcoming election, a contentious news story) and as new arguments are reflected in online discourse, we leverage these seeds to identify emerging themes in social media via seed expansion. To show a concrete example, Table 5.1 lists top ranking opposing and supporting themes at different times for the topic "fracking" after seed expansion is performed. We can see that new biased themes emerge as controversial events occur. In the following, we consider two approaches for expanding the hand-picked seeds to identify supporting and opposing hashtags:

Seed Expansion via Co-Occurrence. The first approach relies on hashtag co-occurrence statistics to find other hashtags used by users with similar opinion bias. The intuition is that if two hashtags are often used together by different users (whether in the same tweet or different tweets with respect to a topic), these two hashtags are likely to indicate the same opinion bias. Here, hashtag co-occurrence is based on users instead of tweets, considering that a user's opinion bias is not likely to change. Thus, for hashtags which occur in different tweets, as long as they are used by the same user, they are still considered

to occur together.

Let f_i and f_j represent the frequency of the hashtag h_i and h_j related to a topic, respectively. Let f_+ and f_- represent the frequency of the pro-seed h_+ and the anti-seed h_-, respectively. The similarity between two hashtags, denoted as $\sigma(h_i, h_j)$, is computed by *Jaccard coefficient (JC)* as:

$$\sigma(h_i, h_j) = \frac{f_{h_i \cap h_j}}{f_{h_i \cup h_j}}$$

where $f_{h_i \cap h_j}$ represents the co-occurrence frequency of h_i and h_j, and $f_{h_i \cup h_j}$ represents the total occurrence frequency of either h_i or h_j. Thus, $\sigma(h_+, h_i)$ and $\sigma(h_-, h_i)$ represents the similarity of h_i to the pro-seed and anti-seed, respectively. We select top hashtags with the largest similarity to the pro-seed and anti-seed as the candidate set C_+ and C_-. However, since some hashtags, for example, "#guncontrol", are generic and co-occur with both the pro-seed and anti-seed, these hashtags do not indicate any opinion bias. To filter out common hashtags like these, we impose the following constraint on hashtag from C_+ for pro-seed:

$$\sigma(h_+, h_i)/\sigma(h_-, h_i) > \epsilon$$

where a large ϵ reflects more correlation with the pro-seed; similar constraint can also be imposed to filter hashtags in C_-. We then use the resulting m top hashtags for pro-seed as positive theme set P, and m top hashtags for anti-seed as negative theme set N.

Seed Expansion via Signed Information Gain. Although the approach above does find other biased hashtags, there are two disadvantages: (i) it often gives niche hashtags which are only used by a small number of participants; (ii) it often misses event-related short-lasting biased hashtags. In light of these issues, we propose the second approach which relies on weak supervision to select the most distinguishing hashtags for each side. Specifically, we perform the following procedure:

1. Training with pro-seed and anti-seed. First, we aggregate a user's tweets and use a bag-

of-words model to compute TFIDF for each user. Users who have tweeted with at least one hashtag are then selected and used. From these users, we treat users with only pro-seed as positive class c_+, users with only anti-seed as negative class c_-, and the rest for prediction. Finally, an SVM classifier is learned on the training data and used to predict the polarity of users which are left. We now have an expanded set of users who are positive, and an expanded set of users who are negative.

2. Selecting hashtags. From the expanded sets of users, we use *signed information gain (SIG)* proposed by Zheng et al. [113] as the measure to select hashtags for pro-seed and anti-seed, respectively.

$$SIG(h_i, c) = sign(AD - BC) \cdot \sum_{c \in \{c_+, c_-\}} \sum_{h \in \{h_i, \bar{h}_i\}} p(h, c) \cdot log \frac{p(h, c)}{p(h) \cdot p(c)}$$

where A is the number of users with h_i and in class c_+, B is the number of users with h_i and in class c_-, C is the number of users without h_i and in class c_+ and D is the number of users without h_i and in class c_-. Also, $\bar{h}_i$ represents hashtags other than h_i. Here, the probability p is obtained by maximum likelihood estimation. We select m hashtags with the largest SIG as the finalized hashtag set P for pro-seed, and m hashtags with the smallest SIG as the finalized hashtag set N for anti-seed.

As a result, this approach can not only filter out common hashtags used by both sides of users without manually specifying any extra parameters, but also can discover popular yet distinguished biased hashtags.

Bias Anchors. Given the expanded set of hashtags (both supporting and opposing a particular topic), we identify as our strong partisans users who consistently adopt hashtags from only one opinion standpoint, which we denote as U_{anchor}. We assign an initial bias

score $\tilde{b_i}$ to these anchors as follows:

$$\tilde{b_i} = \begin{cases} +1, & \text{if } u_i \in U_P \\ -1, & \text{if } u_i \in U_N \end{cases} \tag{5.1}$$

where U_P and U_N is the set of anchors adopting hashtags from P and N, respectively, and $U_{anchor} = U_P \cup U_N$. These opinion bias anchors serve as the basis for propagating bias throughout the social and information network, which we tackle next.

5.2.2 Bias Propagation Network

After the discovery of bias anchors, how do we determine the opinion bias of those remaining participants? We propose to build a propagation network where two users are only connected if their similarity passes a threshold. In the following, we adopt both content and link features to determine user similarity.

Content-Based Propagation. The assumption of content induced propagation is that if two users have a high textual similarity in their posts, it is likely that they may share similar opinion bias. To compute the content similarity of two users, we aggregate each user's topic-related tweets and treat each user as a document. Thus, content similarity of two users can be computed with document similarity. Here, we adopt cosine similarity of the TFIDF of the two documents with a standard bi-gram model. Tokenization of tweets is done through the tool provided by Owoputi et al. [114] for its robustness. Hashtags and mentions are also included in the model as features. To reduce the size of feature dimensions, we performed stop-word removal and kept only unigrams and bi-grams with occurrence frequency greater than two. The content similarity between u_i and u_j can be

written as:

$$w_{ij}^{content} = \begin{cases} C_{ij}, & \text{if } u_j \in \mathcal{N}^c(u_i) \\ 0, & \text{if } u_j \notin \mathcal{N}^c(u_i) \end{cases} \tag{5.2}$$

where C_{ij} is the cosine similarity of u_i and u_j. To reduce the propagation complexity, we construct a sparse network by only considering k-nearest neighbors $\mathcal{N}^c(u_i)$ for each user u_i. We choose k to be 10 for its efficiency without compromising much accuracy in experiments.

Link-Based Propagation. Retweeting can be considered a form of endorsement for users in Twitter [115, 45]. Based on this observation, it is expected that if a user retweets another user on a topic, both users tend to share similar opinion bias. Thus, we define the link similarity between u_i and u_j as follows:

$$w_{ij}^{link} = \begin{cases} 1, & \text{if } u_j \in \mathcal{N}^l(u_i) \\ 0, & \text{if } u_j \notin \mathcal{N}^l(u_i) \end{cases} \tag{5.3}$$

where u_j is in the neighbors of u_i if u_j retweeted u_i or u_i retweeted u_j. Besides retweeting links, we can also take advantage of the follower/following network information in Twitter. Here, we choose not to use it since following link is not as strong a signal as retweeting. One reason is that a user whose opinion is on one side may choose to follow someone on the opposite side for the purpose of receiving any topic related statuses or refuting their arguments. Furthermore, we notice that the resource of retweeting activities is usually sparse so that the propagation network constituted only from retweeting links is separated into many isolated networks. Thus, a retweeting-based propagation is not enough to be treated by itself but needs to be combined with content-based propagation,

leading to the following fusion of similarity between users:

$$w_{ij} = w_{ij}^{content} + \lambda w_{ij}^{link} \tag{5.4}$$

where λ is a weighting parameter for content and link similarity.

5.2.3 Optimization Framework

Finally, we embed the discovered anchors and bias propagation network into an optimization setting to propagate the opinion-bias score of all users more effectively. Since the initial input to the approach is a set of weak labels (two user-specified opposite hashtags), we call this optimization *User-guided Opinion Propagation [UOP]*. Specifically, by allowing each user's true opinion bias b_i to change as an optimization variable, we force the following conditions: (i) for bias anchors, b_i should be as close to the bias indicated by adopting biased hashtags (Eqn. 5.1); (ii) for other participants, b_i and b_j should be close to the degree indicated by their content and link similarity (Eqn. 5.4). Their opinion bias is initialized randomly between $[-1, 1]$ and can now be iteratively propagated through optimization. Thus, we have the following objective function:

$$\begin{aligned} \min_{b_i \in B} \quad & f = \sum_{u_i \in U_{anchor}} (b_i - \tilde{b_i})^2 + \mu_1 \sum_{i=1}^{n} \sum_{j=i+1}^{n} w_{ij}(b_i - b_j)^2 \\ \text{subject to} \quad & -1 \leq b_i \leq 1 \qquad \forall i \in \{1, ..., n\} \end{aligned} \tag{5.5}$$

where μ_1 is the tradeoff weight for different components. Thus, by solving this optimization, each user's opinion bias is propagated through the network in an optimized fashion. Since the objective function is convex, we can use the standard L-BFGS method with constraints to solve it efficiently. Another advantage of this framework is that other similarity signals such as location, profile demographics, and so on can be easily incorporated into Equation 5.5.

Handling Noisy Bias Anchors. The essence of the above optimization framework is that we propagate users' opinion bias which we know with confidence to other users who we have seldom knowledge of. We can see that anchor's opinion bias $\tilde{b_i}$ is treated as the golden truth and stays unchanged. However, a prominent issue is that users who consistently adopt hashtags from either P or N are not guaranteed to have the corresponding opinion bias. To give an example, one user's tweet reads, "#Gosnell certainly a tragedy, also cautionary tale, but not an argument against abortion rights in the US. Want more Gosnells? Ban abortion." This sarcastic pro-choice user adopted the hashtag #Gosnell, as we can observe from many tweets, is a primary hashtag pro-life users would use. Hence, according to Equation 5.1, this user is falsely identified as a pro-life anchor. Without manual inspection of user's profile and their related tweets, it is very hard to judge whether a bias anchor determined from Equation 5.1 is correctly identified. To relieve this problem, we introduce another variable y_i for $u_i \in U_{anchor}$ as the ideal opinion bias. Intuitively, it should satisfy: (i) y_i should be close to the opinion bias inferred from neighbors; (ii) most y_i should be consistent with $\tilde{b_i}$, with a few of them being noisy. Inspired by the annotation of noisy web images in [116], we propose a modified minimization function as follows:

$$\begin{aligned} \min_{b_i \in B} \quad f = & \sum_{u_i \in U_{anchor}} (b_i - y_i)^2 + \mu_1 \sum_{i=1}^{n} \sum_{j=i+1}^{n} w_{ij} (b_i - b_j)^2 \\ & + \mu_2 \sum_{u_i \in U_{anchor}} |y_i - \tilde{b_i}| \\ \text{subject to} \quad & -1 \leq b_i \leq 1 \quad -1 \leq y_i \leq 1 \quad \forall i \in \{1, ..., n\} \end{aligned} \tag{5.6}$$

where μ_1 and μ_2 are the weighting parameters. We use l-1 norm to constrain the ideal variable y_i to $\tilde{b_i}$ since normally, only a small portion of bias anchors are noisy and l-1 norm could force most anchors to stay as biased. We solve the above minimization through the following steps:

(i) Initialize y_i to $\tilde{b_i}$ and solve Equation 5.5 to obtain b_i.

(ii) Solve the following sub-minimization problem with b_i to obtain y_i:

$$\begin{aligned} \min_{b_i \in B} \quad & f = \sum_{u_i \in U_{anchor}} (b_i - y_i)^2 + \mu_2 |y_i - \tilde{b_i}| \\ \text{subject to} \quad & -1 \leq y_i \leq 1 \qquad \forall i \in \{1, ..., n\} \end{aligned} \tag{5.7}$$

We employ the package L1 General [117] to solve the problem.

(iii) Replace $\tilde{b_i}$ with y_i in Equation 5.5 to get final b_i. We could repeat step (i) and (ii) for several times for further optimization but usually two to three iterations are enough shown by experiments. In this way, errors made in identifying bias anchors can be mitigated, leading to more accurate opinion bias estimation.

5.3 Experimental Evaluation

In this section, we perform several sets of experiments to evaluate the BiasWatch framework for topic-sensitive opinion bias discovery. We investigate the impact of seed expansion, the quality of bias propagation via both content and retweeting links, and compare the performance versus alternative opinion bias approaches. We couple this study with an application of the system on two more controversial datasets.

5.3.1 Data

The datasets that we use are collected with Twitter's streaming API from October 2011 to September 2013. To create topic-related datasets for opinion discovery, we selected three controversial topics: "gun control", "abortion" and "obamacare". We select these topics because they are popular controversial topics discussed by a large number of Twitter users with both opposing sides of opinion expressed in the time period. For each topic, we extracted a base set of tweets (and their corresponding users) containing at least one topic-related keyword: for "gun control": *gun control, gun right, pro gun, anti gun,*

Table 5.2: Datasets

Topic	Users	Tweets	Retweets
gun control	70,387	117,679	60,293
abortion	119,664	173,236	93,690
obamacare	67,937	123,320	70,008
vaccine	27,362	36,822	13,108
fracking	22,231	34,485	14,524

gun free, gun law, gun safety, gun violence; for “abortion": *abortion, prolife, prochoice, anti-abortion, pro-abortion, planned parenthood*; and for “obamacare": *obamacare, #aca*. Additionally, we created another two datasets on the topics “vaccine” and “fracking” for demonstration. We select these two topics for further evaluation because they are relatively recent controversial topics compared to the previous ones, and also their opposing sides may not be fully entrenched in traditional left/right party politics. To extract “vaccine” related tweets, we use the following keywords: *vaccine, vaccination, vaccinate, #vaxfax*; for “fracking”, we use: *fracking, #frack, hydraulic fracturing, shale, horizontal drilling*. We summarize the datasets in Table 6.2.

5.3.2 Gathering Ground Truth

In order to evaluate the framework, we need to know the true opinion of a randomly sampled user set against which we can compare the optimization results. Without direct access to user’s bias and considering the inherent difficulty of knowing a user’s bias degree with respect to a controversial topic, we rely on an external labeling scheme using Amazon Mechanical Turk. Since the bias score obtained from Equation 5.5 is continuous, we discretize the opinion bias of a Twitter user into the following five categories: strong support [+2], some support [+1], neutral or no evidence [0], some opposition [-1], strong opposition [-2].

Thus, we can map the continuous range into the above categories for evaluation. For

Table 5.3: Turker labeling results of HITs

topic	Number of users for each category				
	+2	+1	0	-1	-2
gun control	116	40	60	54	234
abortion	115	54	55	26	254
obamacare	82	26	33	26	337

each topic, we randomly selected 504 Twitter users from the total users in Table 6.2, and assigned eight users in each human intelligence task (HIT), then ask the turkers (human labeler) to select the most appropriate category for these users. For each user, we show her twitter user ID and her topic related tweets for each turker to examine. We also highlight the hyperlinks embedded in the tweets and make them clickable. To ensure good quality of assessment, we follow the suggestions by Marshall and Shipman [118]. For each human intelligence task, we put two additional users in random positions, making a total of ten users in one HIT. Those users' bias are already known through experts, which we refer to as the golden users. If the label given by a turker for any of these golden users is very different from that by experts, we discard the entire answer by this turker for the HIT. Moreover, we ask five turkers to label one user and take the majority vote as the final label for the user. The results are shown in Table 5.3.

Agreement of Opinion Bias Labels. To measure the reliability of the above human labeling tasks, we investigate the inter-rater agreement of the obtained assessment with Fleiss' κ statistic. Specifically, we obtained the 5-category κ statistic of 0.264, 0.393 and 0.418 for "gun control", "abortion" and "obamacare", respectively. These values lie in the interpretation of fair agreement by Landis and Koch [119]. In addition, we also adopt the accuracy of agreement provided by Nowak [120] and adapt it into the following formula

since each user is assessed by more than two turkers:

$$accuracy = \frac{1}{N}\sum_{i=1}^{N}\frac{\text{\# of votes of the majority category for user i}}{\text{\# of votes for user i}}$$

where N is the total number of users to be assessed by turkers for each topic. The accuracy ranges from 0.304 when the majority is obtained by chance to 1 when every user's bias category is agreed by all turkers. The lower bound 0.304 is obtained by calculating the average number of people in a majority out of five when they select one category from five by chance. Hence, an accuracy of 0.6, for example, means that on average, 3 out of 5 turkers agree on a category. The accuracy for "gun control", "abortion" and "obamacare" is 0.646, 0.727 and 0.788, respectively, which means, on average, at least 3 turkers agree on the majority category.

Furthermore, we aggregated the same polarity into one category, namely, category [+1] and [+2] are combined to one category and vice versa. The 3-category κ statistic increases to 0.461, 0.588 and 0.649, correspondingly, while the accuracy increases to 0.811, 0.857 and 0.906. The κ values can now be interpreted as moderate agreement. The accuracy now means at least four out of five people agree on the majority bias polarity on average. This indicates that humans are more capable of discerning the polarity of users' opinion bias than determining the extent of users' bias.

In the following, we choose to use the 3 bias categories as ground truth since it has the most consistent and reliable performance by human labelers. We additionally consider only the support and opposition categories (ignoring the minority of users who are neutral or do not show evidence, namely, category [0]) so we can cast the evaluation as a binary class problem. We adopt the standard classification measures of accuracy and area under the curve (AUC).

5.3.3 Alternative Opinion Bias Estimators

To evaluate our approach of determining users' opinion bias, we consider the following alternative opinion bias estimators:

- **SentiWordNet [SWN]**. This is a simple sentiment detection approach, where we assign a sentiment score to each user's tweets according to SentiWordNet and classify each user's opinion bias with the relative portion of positive and negative tweets. Specifically, for each tweet d_i of user u_i, we classify it positive if $\sum_{w_j \in d_i} pos(w_j) > \sum_{w_j \in d_i} neg(w_j)$, and vice versa. We then classify user u_i as positive if the number of positive tweets is greater than the number of negative tweets, and vice versa.
- **User Clustering with Content [uCC]**. In this baseline, we construct a user graph and perform user clustering with tweets. The nodes of the graph are users and the edges are constructed based on k nearest neighbors with the largest content similarities of tweets. We choose to use cosine similarity of the TFIDF of bi-grams as the similarity measure. To perform graph clustering, we apply normalized cuts for graph partitioning by Shi and Malik [121] due to its simplicity and good performance. This is essentially a max-cut problem explored in [122, 123] to partition newsgroup and online debates into opposite positions, respectively. The purpose of using this unsupervised baseline is to examine whether the selected biased hashtags as a form of weak supervision can provide much improvement.
- **User Clustering with Content and Links [uCCL]**. This baseline is the modified version of uCC in which the edge weight combines both content and link similarity. Specifically, if there exists a retweeting link between two users, we add a constant to its content similarity, i.e., $w = w^{content} + \theta * w^{link}$, where θ is used to balance the weights. The purpose of this baseline is to examine if retweeting links can help distinguish opposing sides of users' opinion bias compared to uCC.

- **Weakly-supervised SVM [wSVM]**. We train an SVM with a bi-gram model of bias anchors' tweets, which is then used to classify the test dataset. The parameters of SVM are determined through 5-fold cross-validation. We denote the trained classifier with bias anchors found through initial seeds (IS) as *wSVM+IS*, and the other two classifiers trained on seed expansions as *wSVM+JC* and *wSVM+SIG*. Here, wSVM+IS is treated as the baseline, and the other two as our improved versions.
- **Local Consistency Global Consistency [LCGC].** This is a semi-supervised method proposed by Zhou et al. [124] and applied in [40] for the classification of the political learning of news articles. This method optimizes the tradeoff between local consistency and global consistency among node labels. Here, we use Equation 5.4 as the affinity between nodes for the method and adapt LCGC into our own version by incorporating seed expansion from SIG, denoted as *LCGC+SIG*.
- **UOP***. This is the framework in which we only consider content based bias propagation without handling noisy bias anchors.
- **UOP**†. This is the framework in which we consider both content and link based bias propagation without handling noisy bias anchors, indicated by Equation 5.5.
- **UOP**. This is the full blown-approach indicated by Equation 5.6.

5.3.4 Biased Theme Discovery

Before experiments, we first select the following pro-seed and anti-seed manually as the input to the system: #guncontrolnow and #2ndamendment for "gun control"; #pro-choice and #prolife for "abortion"; #ilikeobamacare and #defundobamacare for "obamacare". We later show in the experiments the effect of different seed selections.

We then highlight the seed expansion methods – both hashtag co-occurrence (JC) and signed information gain (SIG) – used to identify biased hashtags adopted by users with similar opinion bias. For seed expansion via SIG, we need to determine the value of

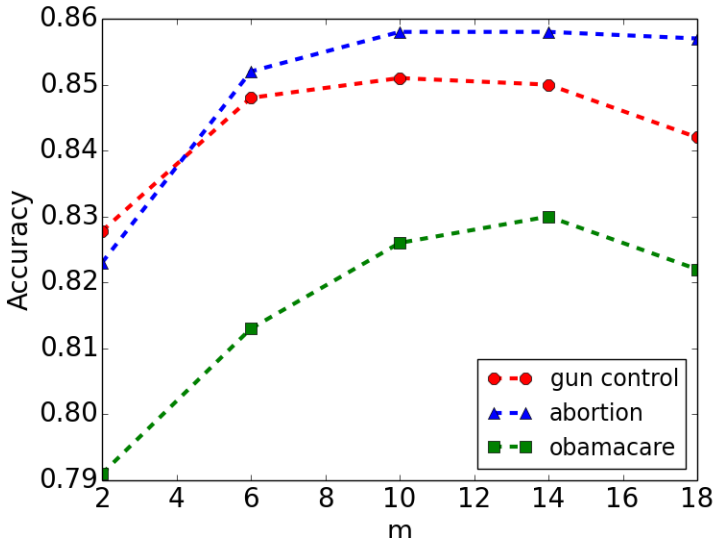


Figure 5.2: Effect of m for seed expansion via SIG.

parameter m. To study the influence of this parameter, we adopt the method UOP* to evaluate performance changes with values from $\{2, 6, 10, 14, 18\}$. Figure 5.2 shows that the accuracy is highest when m is approximately at 10 for "gun control" and "abortion", and is slightly larger for "obamacare". After those values, accuracy levels or even decreases, possibly because the additional discovered hashtags are noisy or do not imply much opinion bias. Thus, in the following experiments, m is fixed at 10 for all topics. For seed expansion via co-occurrence, we choose m to be 10 using the similar approach, and empirically set ϵ to 3. The expanded hashtags, as we observed from the output, can be approximately categorized as:

(i) *Sentiment-oriented.* These hashtags can be easily discerned and used directly by participants to show opinion bias, such as #nowaynra for "gun control" and #dontfundit for "obamacare".

(ii) *Community identification.* These hashtags indicate personal or political identities and are often used in a community, such as #p2, #tcot, #teaparty and #fem2 (for feminists);

(iii) *Thematic.* These hashtags often indicate arguments used by participants to express their opinion. For example, gun control antagonists say #gunrights as a constitutional right protected by #2ndamendment (also #2a); abortion protagonists may emphasize #reprorights or #reprojustice in their arguments;

(iv) *Action-oriented.* Examples include #momsdemandaction, #stand4life, #standwithcruz (stand with Ted Cruz), #swtw (stand with Texas women) and #demandaplan (as in "#demandaplan to end gun violence");

Overall, we can conclude that seed expansion through our proposed approaches is able to find other biased hashtags which are used by people with similar opinion bias. The above categorization also serves as guidance for users to pick initial opposing seeds as input to the system. Now that we have discovered the biased themes related to each side of polarity for different controversial topics, can we leverage those to determine "regular" participants' opinion bias? Specifically, we ask the following questions:

(i) Can these newly discovered biased hashtags help to identify user's opinion bias? If so, how much better can they do?

(ii) Do different pro-seed and anti-seed selections affect performance? If so, can seed expansion help us with the selection?

(iii) Can social ties, in the form of retweeting links, help us determine user's opinion bias?

Effect of Seed Expansion. To evaluate the performance of these expanded hashtags, we use wSVM and UOP* as the base methods since both of them only rely on the information provided by bias anchors and content. For UOP*, the weight parameter μ_1 in Equation 5.5, is empirically determined to be 0.1.

We now compare the performance between the version when opinion bias is propagated only through initial seeds and the version when the seeds are expanded. Since the result of accuracy is similar with AUC, we only show AUC in Figure 5.3. We can see that

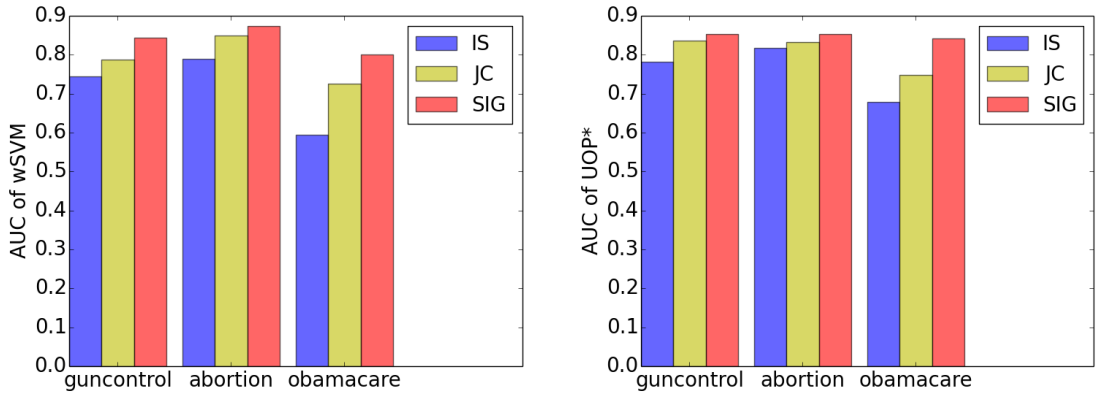


Figure 5.3: Effect of seed expansion.

both seed expansion approaches outperform initial seeds, with seed expansion via SIG giving the best performance. Specifically, AUC from seed expansion via SIG gives a 19.5% and 6.8% improvement over IS and JC for wSVM, respectively, while the improvement is 12.5% and 5.6% for UOP*, respectively. This shows that (i) the newly discovered hashtag set through seed expansion provides additional amount of bias information for users; and (ii) the quality of expanded hashtag set via SIG is generally the best, i.e., it is able to discover higher quality of biased themes.

Effect of Different Seeds. Here, we are interested in studying the influence of different choices of initial seeds to the system. To that end, we first rank hashtags by occurrence frequency and then manually select top five pro-seeds and top five anti-seeds for "gun control" by observing tweets with those hashtags. The top five pro-seeds are: 1. #nowaynra; 2. #guncontrolnow; 3. #demandaplan; 4. #newtown; 5. #whatwillittake. The top five anti-seeds are: 1. #tcot; 2. #2ndamendment; 3. #nj2as; 4. #2a; 5. #gunrights. We represent a pro-seed and anti-seed combination with their corresponding number. For example, "12" represents the combination "#nowaynra #2ndamendment". We then use the method UOP* to evaluate the performance of each combination for two cases: one with

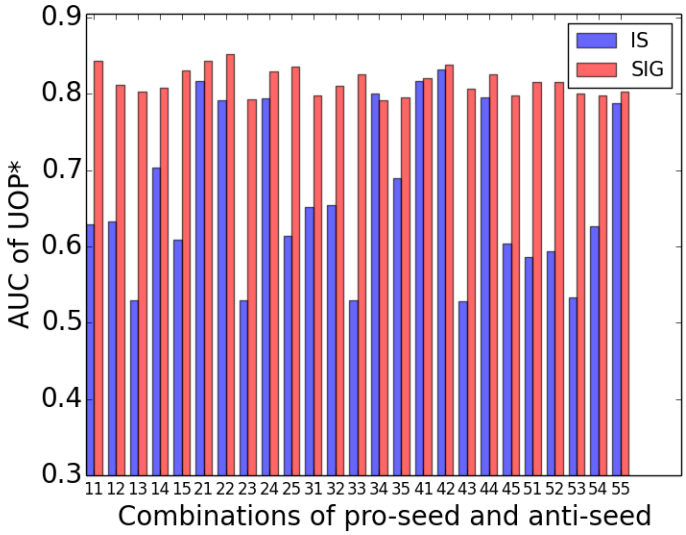


Figure 5.4: Performance for 25 pro-seed and anti-seed combinations.

initial seeds; the other with seed expansion via SIG. The results are shown in Figure 5.4. Overall, we can see that for different choices of input seeds, the performance without seed expansion is very sensitive to the choices, while it gives consistent high accuracy for seed expansion with SIG. We can also observe that #nj2as is not a good anti-seed choice since all combinations with #nj2as performs bad. Also, all combinations with #gunrights give unsatisfactory results except "55". These results indicate that it is difficult to choose the most effective seed combinations, since very often a single pair of seeds can be noisy and do not cover enough bias related themes. However, seed expansion can mitigate this effect by introducing other and often more complete biased hashtags for bias propagation. Hence, even though the initial seeds are not well selected, the final performance does not suffer much due to the benefits of seed expansion. For other topics, similar results are observed, and thus are not reported due to the space limit.

Effect of Social Ties. Here, we evaluate the effect of social ties in the form of retweeting links for different seed expansion approaches. The retweeting links are randomly sampled

for ten times for each fraction since a different sample of retweeting links gives a different propagation network. Here, we adopt UOP† for the evaluation, with the weight given to the link-based propagation λ in Equation 5.4 empirically chosen to be 1. The final results are averaged and plotted in Figure 5.5. We can see that as more retweeting links are added for propagation, the performance increases for all three approaches, which confirms the assumption that users linked via retweeting tend to share similar opinion bias.

Furthermore, among these three approaches, the performance obtained via SIG is always the best for different fractions of retweeting links at all topics. As the fraction gets larger, the improvement generally gets smaller. This indicates that the effect of seed expansion gets reduced when the fraction increases. For seed expansion via JC, the performance is generally better than that of initial seeds when the fraction is small. However, this initial improvement gets reduced or even disappears when retweeting link is abundant, hence, making the extra hashtags obtained via co-occurrence less effective. This is probably because of the noisiness of those extra hashtags. Though beneficial at a small number of retweeting links, they prevent the correct bias from being propagated when more of them are added for optimization. This shows again that the key of seed expansion is that not only more biased hashtags should be discovered, but also they should be of high quality.

5.3.5 Comparison with Baselines

In this section, we compare our proposed BiasWatch framework with the alternative opinion bias estimators. For uCCL, the weight-balancing parameter θ is selected from $\{0.05, 0.1, 0.15, 0.2, 0.25\}$ to be 0.1 for best performance. For LCGC+SIG, we use the same parameter setting as in [40]. Other parameter settings in UOP*, UOP† and UOP are the same as in previous experiments. For μ_2 in Equation 5.6, we empirically set it to 0.2 for handling noisy bias anchors. We choose the best seed expansion approach (via SIG) for all methods. The final results are shown in Table 5.4 and Table 5.5.

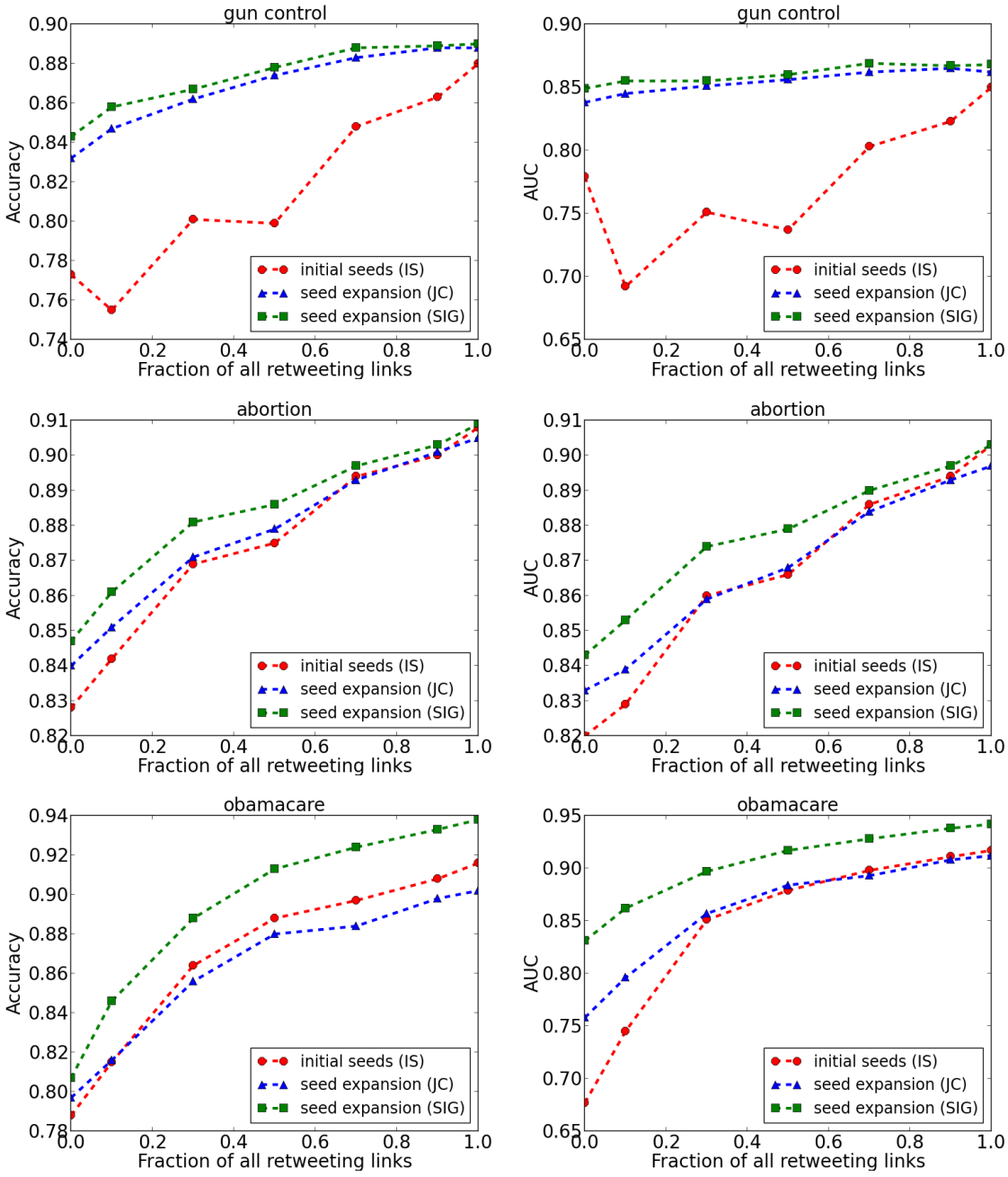


Figure 5.5: Performance for different seed expansion approaches with respect to different fraction of retweeting links for all topics.

Table 5.4: Accuracy Comparison with alternative opinion bias estimators. Boldface: the best result for each topic among all methods. '$*$' marks statistically significant difference against the best of alternative opinion bias estimators (with two sample t-test for $p \leq 0.05$).

Method	gun control	abortion	obamacare	average
SWN	0.560	0.527	0.465	0.517
uCC	0.534	0.537	0.516	0.529
uCCL	0.586	0.530	0.520	0.545
wSVM+IS	0.696	0.825	0.786	0.769
wSVM+SIG	0.860	0.884	0.727	0.824
UOP*	0.851	0.847	0.826	0.841
LCGC+SIG	0.858	0.900	0.811	0.856
UOP†	0.881	0.906	0.894	0.894
UOP	**0.908***	**0.915**	**0.945***	**0.923***

Table 5.5: AUC Comparison with alternative opinion bias estimators. Boldface: the best result for each topic among all methods. '$*$' marks statistically significant difference against the best of alternative opinion bias estimators (with two sample t-test for $p \leq 0.05$).

Method	gun control	abortion	obamacare	average
SWN	0.570	0.531	0.541	0.547
uCC	0.533	0.527	0.522	0.527
uCCL	0.584	0.531	0.546	0.554
wSVM+IS	0.745	0.790	0.594	0.710
wSVM+SIG	0.844	0.874	0.800	0.839
UOP*	0.853	0.843	0.842	0.846
LCGC+SIG	0.857	0.900	0.864	0.874
UOP†	0.861	0.903	0.915	0.893
UOP	**0.883***	**0.910**	**0.945***	**0.913***

Overall, user-guided approaches give much better performance than unsupervised methods, indicating that by just a small amount of human guidance — two opposite seed hashtags, the performance can be boosted significantly. Moreover, UOP gives the best performance, reaching an average accuracy and AUC of 0.923 and 0.913, respectively (an improvement of 20.0% and 28.6% over supervised baseline wSVM+IS). Note that the sentiment based approach SWN gives unsatisfactory results, probably because user's opinion bias is multifaceted and can be reflected by the topical arguments or factual information published by the user. For example, one of the anti-obamacare tweets reads, "Double Down: Obamacare Will Increase Avg. Individual-Market Insurance Premiums By 99% For Men, 62% For Women. #Forbes". Also, we can see that UOP† gives better results than LCGC+SIG, indicating that our framework works better in capturing user's opinion bias. UOP, however, gives better performance than UOP†, confirming that initial bias anchors determined through biased hashtags are noisy, and that UOP is able to correct some wrongly determined bias anchors due to the l-1 norm regularization on ideal bias scores.

Furthermore, we see from the table that UOP† has an improvement of 0.053 and 0.047 for accuracy and AUC over UOP*, respectively. Compared to the corresponding improvement of 0.016 and 0.027 by uCCL over uCC, it is considerably higher. This indicates that retweeting links are more effective in contributing to bias propagation with the help of bias anchors. When some users are correctly "labeled" by discovered biased hashtags, opinion bias can be propagated more effectively through retweeting links.

5.3.6 Multi-Category Classification

In previous experiments, we mainly evaluate our framework as a binary class problem. However, in this way, we lose the finer granularity of the inferred opinion bias score by ignoring users who are neutral or do not show evidence. This category of users can not only be beneficial to our understanding of the general landscape of controversial topics,

but also can be targeted by polarized activists to influence their bias. Thus, in this section, we are interested in a finer level of evaluation by casting it as a three-class problem. Here, instead of combining the category +1 and +2, and the category -1 and -2, we aggregate users in the category of +1, 0 and -1 into one neutral category 0 for a relatively larger pool of users in the middle. We then partition the datasets into 50% for training and 50% for testing, and compare the performance of wSWM+SIG, LCGC+SIG, UOP† and UOP. We are interested to see how these methods perform in the recall defined by $\frac{1}{3}\sum_i \frac{tp_i}{tp_i+fn_i}$. This measure indicates the ability of finding out the correct user category in a three-category setting for different methods.

For wSVM+SIG, we adopt the one-vs-rest multi-class implementation. For LCGC, we modified it to consider three labels. For UOP based methods, we select two optimal thresholds as boundaries: θ_p, used to distinguish category +2 and 0 and θ_n, used to distinguish category -2 and 0. These thresholds are selected by searching the range (-1,1) with a step of 0.05 with the training data. For UOP, θ_n and θ_p are determined as -0.25 and 0.2 for "gun control", -0.35 and 0.35 for "abortion", -0.2 and 0.2 for "obamacare". The final results are shown in Table 5.6. As we can see, UOP gives the best performance of all, indicating that it is the most capable in finding out users who are polarized and neutral. wSVM and LCGC perform worse, probably because of the small amount of training samples for category 0. UOP based methods are able to mitigate this effect by propagating constrained bias score without explicit consideration of neutral users and only with bias anchors. Thus, we can conclude that UOP is effective in representing the degree of user's opinion bias under the controversial topics.

5.3.7 Case Study: Fracking and Vaccines

Now that we have demonstrated the effectiveness of the system in finding general users' opinion bias, we would like to test the framework further in another two datasets:

Table 5.6: Multi-category classification performance.

Method	gun control	abortion	obamacare
wSVM+SIG	0.611	0.592	0.573
LCGC+SIG	0.632	0.657	0.646
UOP†	0.690	0.699	0.673
UOP	0.705	0.701	0.716

"fracking" and "vaccine". Since these two topics are relatively new compared to the above politically-driven well-known topics, it would be interesting to discover what is being posted and debated by opinionated users of both sides. For "fracking", we pick the seeds #dontfrackny and #jobs as the input to the system. We select #jobs as the pro-seed for "fracking" because protagonists tend to emphasize the benefit of creating jobs as one of the arguments for "fracking". For "vaccine", we pick the seeds #health and #autism as the input to the system. Again, we select #autism as the anti-seed because antagonists tend to focus the negative effect of vaccination. Overall, selection of input seeds is intuitive due to the recognizability of some of the crowd-generated tags and thus does not require much human effort.

We first demonstrate the biased themes discovered via SIG. Table 5.1 shows the top-ranking biased hashtags for different time periods. We can see that some themes such as #natgas and #frack remain constantly used by users supporting and opposing fracking, respectively; some other themes, such as #dontfrackny and #balcombe, arise and fade as related controversial events occur. For example, the occurrence of #dontfrackny corresponds with a rally of New Yorkers in the mid February 2013 to urge Governor Cuomo to resist fracking in the state of New York; the occurrence of #balcombe corresponds with a protest against a license to drill near Balcombe in England granted by Environment Agency. These changing themes indicate a strong degree of opinion bias and emerge as a group of users start to use them together, making them a very useful signal to determine

Table 5.7: Top ten themes at different times for "vaccine"; red for pro-vaccine; blue for anti-vaccine.

Feb 2012	June 2012	Nov 2012	Apr 2013
#vaxfax	#health	#health	#vaccineswork
#health	#vaxfax	#vaxfax	#health
#flu	#polio	#flu	#measles
#polio	#hpv	#autism	#mmr
#hpv	#vaccineswork	#news	#hiv
#autism	#pakistan	#polio	#autism
#thrillers	#autism	#hiv	#vaxfax
#measles	#news	#suspense	#polio
#action	#suspense	#thrillers	#flu
#flushot	#action	#hpv	#news

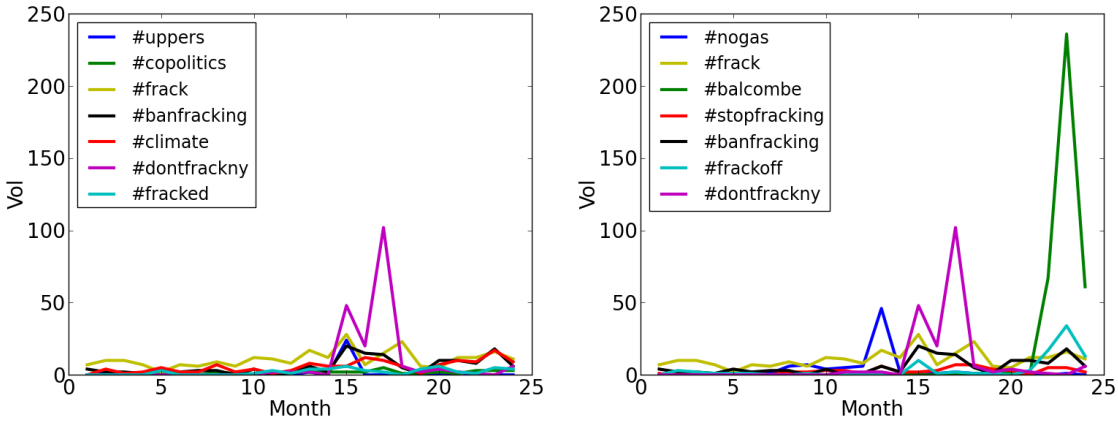


Figure 5.6: Temporal volumes of top anti-fracking themes for seed expansions via co-occurrence (Left) and via SIG (Right).

Table 5.8: Sample opinionated users and their corresponding tweets for "fracking"; positive bias score represents pro-fracking.

users	bias anchor	bias score	tweets
@Energy21	yes	0.91	Four real-life examples of how #shale #energy is creating #jobs and improving the economy: via @FreeEnterprise.
@Duffernutter	yes	-0.92	RT @ABFalecbaldwin: Together we can help @NYGovCuomo see there is no place for #fracking in NY. RT to let him know #DontFrackNY.
@IntellisysUK	no	0.15	RT @JimWoodsUK: Provocative article from @James_BG on New Environmentalism. Worth remembering fracking has reduced US CO2 emissions.
@janet_ewan	no	-0.39	RT @IshtarsGate: A 5.7-magnitude earthquake linked to fracking in Oklahoma in 2011 knocked down 14 homes and injured two people.

and propagate user's opinion bias. Furthermore, the transient property of these changing themes also makes the approach of seed expansion via co-occurrence less effective. Figure 5.6 illustrates the temporal characteristics of discovered anti-fracking themes for different seed expansion approaches. We can see that seed expansion via co-occurrence failed discovering #balcombe and #nogas, as both of these hashtags do not co-occur with #dontfrackny. In contrast, SIG tackles the problem by tapping into the power of content to discover more related biased themes. Table 5.7 shows the top-ranking biased hashtags at different time periods for the dataset "vaccine". To illustrate how the system can uncover strong partisans, Table 5.8 shows two uncovered bias anchors – @Energy21, a pro-fracking account from the Institute for 21st Century Energy, and @Duffernutter, an anti-fracking account associated with TheEnvironmentTV.

5.4 Integrating Opinion Bias into User Recommendation

In this section, we demonstrate the application of integrating opinion bias for user recommendation in social media. The principle of homophily, observed both in political blogosphere [42] and social media [45], states that people tend to associate with others who are like-minded. Thus, when social media services recommend users to follow, it is natural to consider recommending users who have similar opinion bias on shared topic interests. User recommendation can also be considered as a task of link prediction in graphs, and there has already exist many works [125] which specifically address this task, mostly taking advantage of graph structure. Here, our goal is to demonstrate how opinion bias can be utilized for user recommendation as a different dimension.

The task can be formally described as follows: Given a controversial topic T, a sampled set of users U and their corresponding on-topic tweets D, recommend k friends to follow for the target user. Note that we have only tweets to rely on to determine the recommendations. We provide two approaches for the task.

Content-Based approach. In this collaborative filtering approach, users with the highest content similarities to the target user are recommended. Specifically, we aggregated the corresponding tweets for each user and applied vector space model (VSM) with unigrams and bi-grams, with the similarity computed by cosine measure.

Opinion-Weighted (OW) approach. Here, user similarity is considered as a weighted sum of content similarity and opinion similarity. The content similarity is computed in the same way as VSM. The opinion similarity is obtained as follows. First, user's opinion bias score is obtained through UOP. Then, we characterize the opinion similarity of two users as:

$$sim_{opi}(u_i, u_j) = 1 - |f(b_i) - f(b_j)|$$

where $f(x)$ is a normalizing function with the form as $\frac{1}{1+e^{-c \cdot x}}$. The value of parameter

c should be chosen to make b's transition to opposite sign steep enough, so that two bias scores with small difference but opposite signs can still result in large difference after transformation. Also, when b_i and b_j have the same sign, they can have a relatively large opinion similarity. To this end, c is chosen to be 5. The final similarity is computed as the weighted sum (α is the weighting parameter), which we use for ranking users:

$$sim(u_i, u_j) = \alpha sim_{vsm}(u_i, u_j) + (1 - \alpha) sim_{opi}(u_i, u_j)$$

5.4.1 Evaluating User Recommendation

For evaluation, we additionally crawled following links for the dataset "gun control", "abortion" and "obamacare". These following links are used as the ground truth in our experiments. We randomly sampled 500 users who have at least 20 followees for each topic and used the following metrics to evaluate: (i) precision@K: which measures the percent of the correct followees out of the top K recommended users; and (ii) mean average precision@K: which is the average of the precision at the position of each correct followee out of the top K recommended users. This measure considers the positions of the recommended users. K is chosen to be 20.

Figure 5.7 shows the performance comparisons between the two approaches. Here, α is set to 0.5. For both metrics, the OW approach has a better performance than the vanilla VSM for each topic. On average, it gives an improvement of 26.3% in precision@20 and 13.8% in MAP@20. These results indicate that user's opinion similarity boosted the rank of some of the true followees who have similar opinion bias as the target user, while lowering the similarity with users who hold different opinion. Hence, it implicitly confirms the principle of homophily that people tend to make friends who share similar opinions.

In Figure 5.8, we also show the performance at different values of α. As we can see, the best performance is achieved neither at $\alpha = 0$ or 1 for all topics, but at a mixed weight

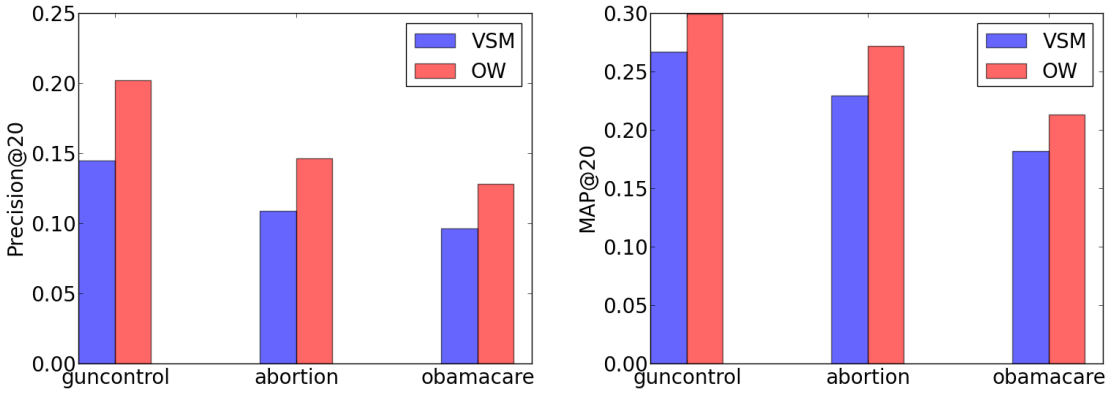


Figure 5.7: Performance comparison for VSM and OW.

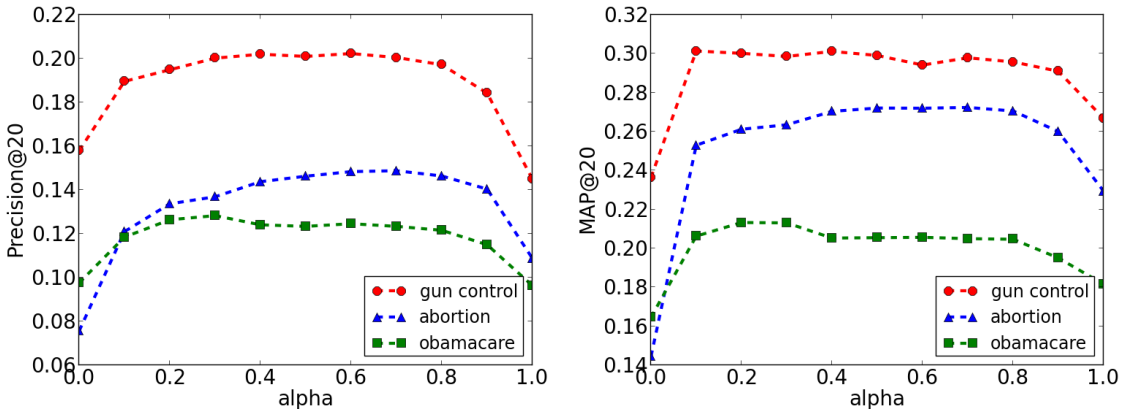


Figure 5.8: Performance at different values of parameter α.

between content and opinion similarity. Even when the weight given to opinion similarity is small, i.e., when $\alpha = 0.9$, the improvement can reach 20.2% for precision@20 and 9.2% for MAP@20 on average. The figure shows that the performance is not very sensitive to the value of α when α is approximately in the range of (0.2, 0.8), indicating it does not require fine tuning to reach a better performance.

5.5 Conclusion and Next Steps

We have seen how the BiasWatch system can lead to an improvement of 20% in accuracy over the next-best alternative for bias estimation, as well as uncover opinion leaders and bias themes that may evolve over time. We also demonstrated how the inferred opinion bias can be integrated into user recommendation and showed that it gives a performance improvement of 26% in precision. While our investigation has focused on textual and relational features, it does not limit us from integrating new signals such as location and profile demographics for better performance in future work. We are also interested in incorporating opinion bias into a social media dashboard, so that participants can be aware of their own opinion dynamics as well as those of others.

6. USER PROFILING: DISCOVERING USER'S GEO-TOPIC PROFILE

In this section, we move to the final aspect of user profiling in the dissertation: user's geo-topic profile, a multi-dimensional concept to describe geo-spatially aware topical profile for users. Specifically, we first give the motivation on why we propose modeling user's geo-topic profile, then present and evaluate a principled framework to uncover what users are known at what location.

6.1 Introduction

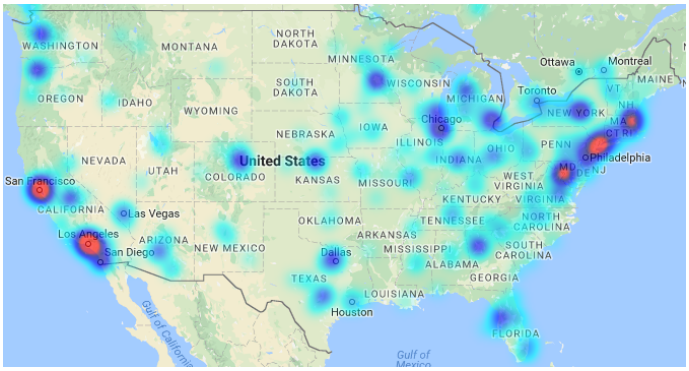
Uncovering user interests and expertise is a vital component of search and recommendation systems on the web and social media [13, 26, 12, 16, 17]. These *user topical profiles* reveal what a user is known for [19] and can be used to augment newsfeed ranking algorithms to surface high-quality content, improve item-based recommenders by leveraging the topical expertise of knowledgeable users [98], and enhance personalized web search [11, 12] and targeted advertising [21].

Of course, the quality of these *user topical profiles* is dependent on the social-spatial properties of the systems in which they arise. For example, several recent efforts [107, 18, 19] have demonstrated how to build profiles from Twitter Lists, where the aggregate crowd-labeling of Twitter users can provide a window into what users are known for. The user topical profiles derived from these Twitter Lists are impacted by the social connections of users on Twitter as well as their geo-locations, meaning that careful consideration of these social-spatial properties is critical. To illustrate, Figure 1 shows the heat maps of the locations of Twitter users who have labeled Michael Moore (@MMflint) and Roy Blunt (@RoyBlunt). We observe that: (i) Generally, a user's topical expertise is geographically bounded, i.e., a user is not uniformly known across the entire geo-space; (ii) @MMFlint and @RoyBlunt, even though both known for politics, are known by users from different

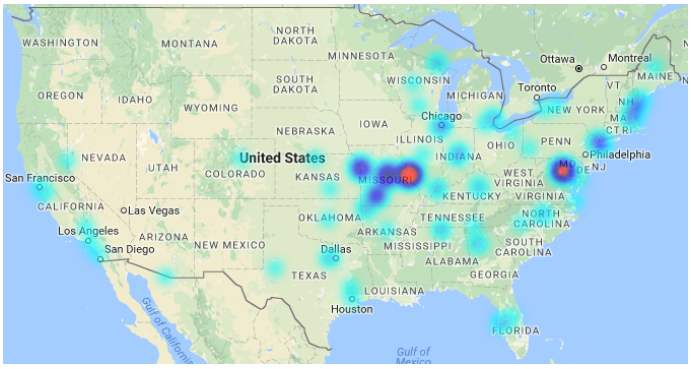
regions, with @RoyBlunt known mainly in Missouri and D.C. while @MMFlint is known in a much broader geo-scope; (iii) @MMFlint, as a filmmaker, is mainly known in New York and LA, while he has a much broader impact for politics in other regions such as San Francisco and D.C. as well. Thus, we can see that geo-location can impact user topical profiles in two aspects: (i) Having the same topical profile, users can have impact in different geo-locations; (ii) One user can be known at different geo-locations with varying popularity for different topical expertise.

Motivated by these observations, we investigate the impact of social-spatial properties on the creation of high-quality user topical profiles. Complementary to traditional user topical profiles, we propose the modeling of fine-grained *user geo-topic profiles*. These geo-topic profiles capture the variations of user popularity for topics across geo-locations; essentially, a 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. Modeling user geo-topic profiles faces two major challenges: (i) they are often overdispersed. Unlike the ratings or content studied in many previous works [95, 91, 58], the popularity counts in geo-topic profiles are heterogeneous with varying scales for different users (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.

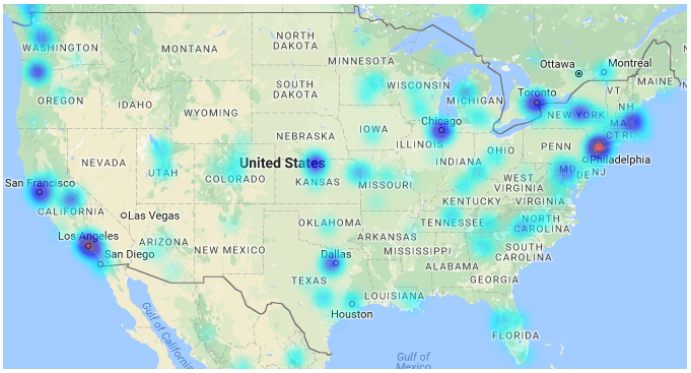
Given these challenges, we first propose a multi-layered (two-layered in our case) Bayesian hierarchical user factorization which extends the recently proposed Poisson gamma belief network [97] from modeling two dimensional non-negative counts to multi-dimensional heterogeneous counts. The extra layer of user factorization learns a more expressive user model than that of single-layered factorization by allowing a larger variance-to-mean ratio from the generative framework, thus making it better equipped at handling overdispersion and user heterogeneity. To alleviate the sparsity issue, we investigate how user's con-



(a) @MMFlint known for politics



(b) @RoyBlunt known for politics



(c) @MMFlint known as filmmaker

Figure 6.1: Spatial distribution of Twitter users who have listed (a) @MMFlint for politics; (b) @RoyBlunt for politics; (c) @MMFlint as a filmmaker.

textual information, specifically, user's geo-location and social ties, correlates with one's geo-topic profile, and then propose to integrate these contexts into the two-layered hierarchical user model for 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 to obtain the posterior distributions of latent factors and parameters. We summarize the contributions as follows:

- First, we introduce a multi-dimensional user profiling concept called user 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 model where users are popular for what topic.
- 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 then propose an enhanced model (**bHUF+**) by generating the first-layer user latent factor based on user's contextual information. We then develop an efficient closed-form Gibbs sampling scheme for inference using a data augmentation scheme 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.

6.2 Preliminaries

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 6.1. Linkedin users can also use *skill tags* for their own profiles and be endorsed by others with these tags. With these notations, we first define a user's geo-topic token in a social-spatial network as follows.

User's Geo-Topic Token. A geo-topic token of user u is defined as a quadruplet $\{u \overset{t}{\leftarrow} v, l\}$ 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. We then give the following definition of user's geo-topic profile.

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, $y \in \mathbb{Z}_{\geq 0}$, 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. Based on the above definition, we formally define the problem studied in this work as below.

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 predict and rank user u's popularity for different tags at different locations, i.e., identify top-ranking $\{u, t, l, \hat{y}\}$ according to learned score $\hat{y}$ from each tag-location combination (t, l) for u.

Bayesian Poisson Tensor Factorization (BPTF). Bayesian Poisson factorization (BPF) has shown promising performance for several tasks [91, 58] on modeling count-valued data. By assuming Poisson distribution in modeling counts instead of Gaussian distribution, it is better at capturing long tailed distribution of sparse discrete data and enjoys scalable learning since only non-zeros are considered during inference. BPF is extended to BPTF in [96] for modeling high dimensional sparse dyadic tensor data. Specifically, 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) decomposition with Poisson distribution as follows,

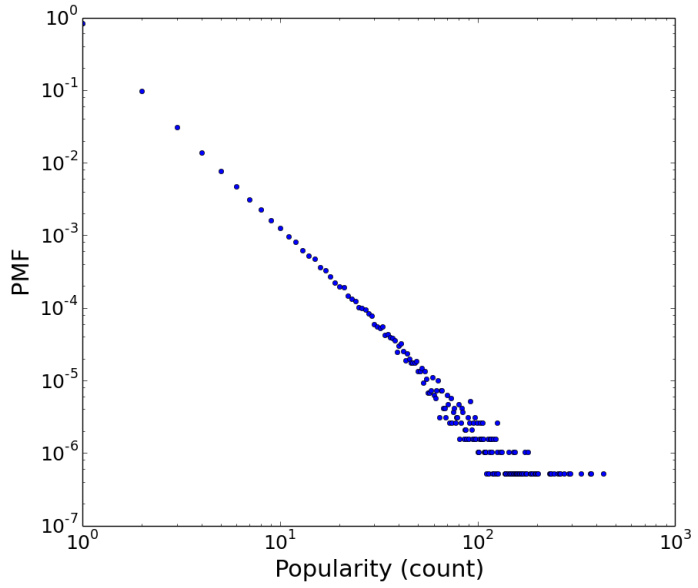
$$y_{utl} \sim Poisson(\sum_{k=1}^{K} \theta_{uk}\phi_{tk}\psi_{lk}) \tag{6.1}$$

where θ_{uk} is the latent factor of u representing user's geo-topic preference by others. ϕ_{tk} and ψ_{lk} represents tag t's and location l's latent factor. Each component of these latent factors are 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.

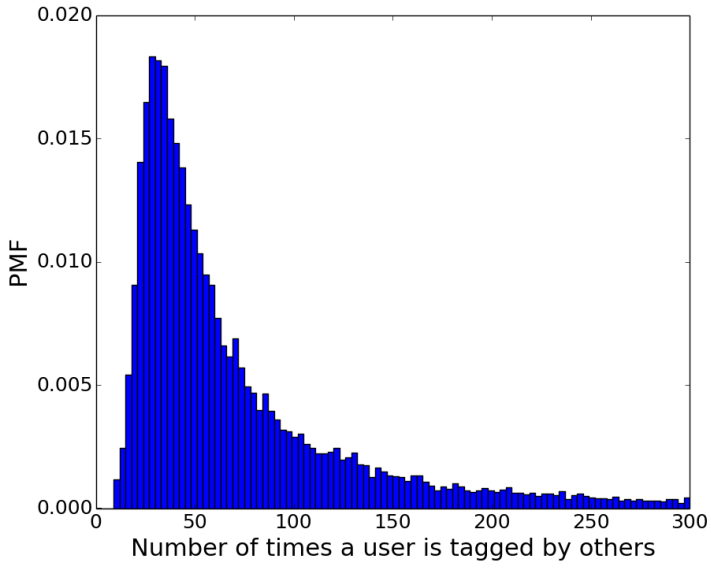
6.3 Bayesian Hierarchical User Factorization

As a first pass, we can attack the problem with BPTF since users' geo-topic profiles constitute a non-negative count tensor. We can then use inferred latent factors θ_{uk}, θ_{tk} and θ_{lk} to compute the expected score to predict the missing values in a user's geo-topic profile. However, BPTF may struggle with the following challenges:

Geo-topic profiles are overdispersed. We first plot the probability mass function of user's



(a) a



(b) b

Figure 6.2: a: probability mass function of user's popularity counts with a grid size of $2.5^\circ \times 2.5^\circ$ in terms of latitude by longitude. b: histogram of the number of geo-topic tokens for a user. Similar distributions have been observed in other sizes of grids.

popularity counts in Figure 6.2a. 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 in the data. Figure 6.2b also shows the histogram of the number of user geo-topic tokens, which exhibits a heavy-tailed distribution. This indicates that users are often known by others to different extent, i.e., some users can be very popular while others are lesser known. This heterogeneity from users requires a model that is capable of capturing this phenomenon. However, the modeling capability of Poisson distribution in Equation 6.1 is limited due to its equal mean and variance, thus may not perform well with these overdispersed popularity counts.

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.

In the following sections, we first propose a two-layered hierarchical model (bHUF) to overcome overdispersion and user heterogeneity, 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.

6.3.1 Two-layered Hierarchical Model

bHUF is a hierarchically constructed multi-dimensional Bayesian framework which extends the recently proposed Poisson Gamma belief network [97] from modeling two

dimensional non-negative counts to modeling sparse and heterogeneous high 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 6.1, 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 6.1, 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 dependent latent factor. The overall generative framework, given y_{utl}, is described as follows,

$$\begin{aligned} y_{utl} &\sim Poisson(\sum_{k=1}^{K} \theta_{uk}\phi_{tk}\psi_{lk}) \\ \theta_{uk} &\sim Gamma(\theta_{uk_2}\Omega, \beta_u) \\ \theta_{uk_2} &\sim Gamma(\epsilon_0, \beta_u^{(2)}) \\ \Omega_{k_2:} &\sim Dirichlet(\eta_0, ..., \eta_0) \end{aligned} \tag{6.2}$$

where θ_{uk} is Gamma distributed according to a second user's latent factor θ_{uk_2}. The second user's latent factor θ_{uk_2} is given a Gamma distribution as a prior.

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 designed to capture the correlations between user's latent factor θ_{uk} of the first layer. Thus, Ω can be treated as the weights between two hidden layers from the perspective of a neural network. Each row of Ω is given Dirichlet distribution as

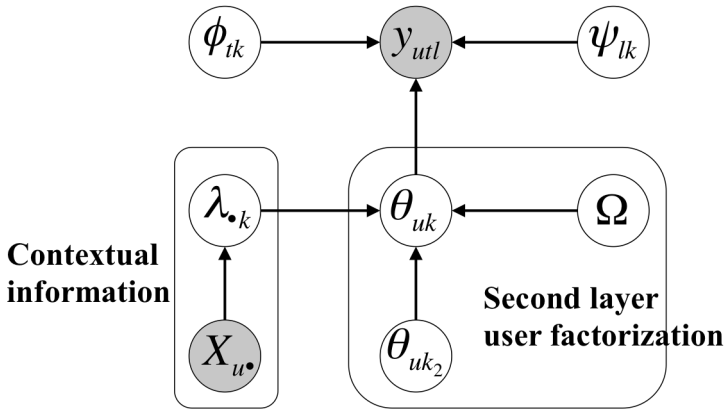


Figure 6.3: Overall generative framework.

its prior so as to push the scale learning to Gamma distributed θ_{uk_2}. As we will later show, it also permits a relatively easy inference process.

For tag and location latent factor, we impose Dirichlet distributions as their priors:

$$\phi_{:m} \sim Dirichlet(\eta^t, ..., \eta^t)$$
$$\psi_{:m} \sim Dirichlet(\eta^l, ..., \eta^l)$$

where each column of the tag and location latent factor $\phi_{:m}$ and $\psi_{:m}$ is Dirichlet distributed. The use of Dirichlet distribution as priors brings several benefits: (i) it naturally imposes non-negativity constraints [126]; (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 can also separate scale from tag and location factors, which makes the tag and location factor not interact with user's latent factor in terms of scale for all latent dimensions. This property is desired to generalize our three-dimensional models to higher dimensions, and as we will show in section 6.3.3, to maintain the validity of lemma 3.

For both layers of user's rate parameter β_u and $\beta_u^{(2)}$, we place uninformative gamma

priors on them: $\beta_u, \beta_u^{(2)} \sim Gamma(\epsilon_0, \epsilon_0)$. These user-dependent rate parameters are used to generate user's latent factor θ_{uk} and θ_{uk_2} to explain the different degree of overall popularity for each user. Thus, β_u and $\beta_u^{(2)}$ shall be learnt to reflect the extent of user's overall popularity at different layers. 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}.

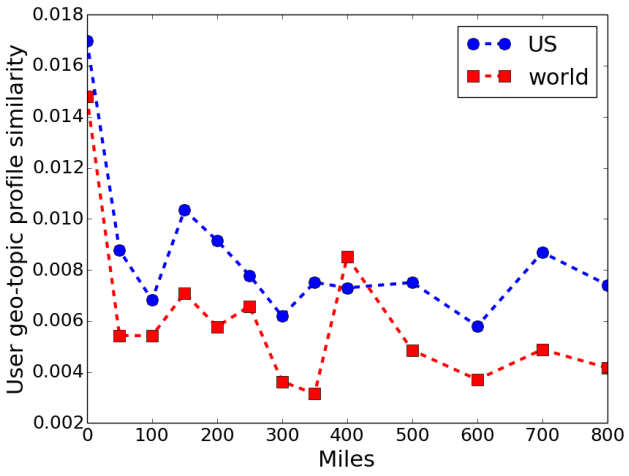
6.3.2 Modeling User's Contextual Information

In this section, we present an enhanced contextual user model based on the previously proposed hierarchical generative framework to alleviate the sparsity issue. Specifically, we investigate two of the most common user contexts in social-spatial networks — geo-location and social ties — among the many user contexts. We first explore how these two factors correlate with user's geo-topic profiles, and then propose a contextualized Bayesian hierarchical user model to learn better representation of users.

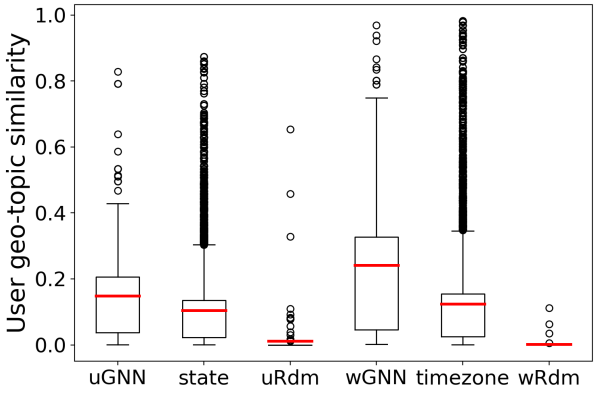
6.3.2.1 User's Geo-location Context

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 Jenson-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 $1 - JS(\mathcal{P}_{u_a}, \mathcal{P}_{u_b})$ where $JS(\mathcal{P}_{u_a}, \mathcal{P}_{u_b}) = \frac{1}{2}KL(\mathcal{P}_{u_a}||M) + \frac{1}{2}KL(\mathcal{P}_{u_b}||M)$. M is the average of $\mathcal{P}_{u_a}$ and $\mathcal{P}_{u_b}$ and $KL(\cdot)$ 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 6.4. As we 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



(a) a



(b) b

Figure 6.4: Left: average users' geo-topic profile similarities with respect to the Haversine distance between users. Right: boxplots for US GNN, US state, US random, world GNN, world timezone, world random.

they are within about 50 miles. It then drastically decreases as the distance increases. This observation aligns with Tobler's first law [127] 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 6.4b 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 6.4b 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.

6.3.2.2 User's Social Context

Due to the homophily effect [103], socially connected users often share similar interests with each other. Indeed, there have been many works [27, 14] 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 the social similarity of two users 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 6.1. As we can see, for both users in US and the world, there exists weak

Table 6.1: Pearson correlation coefficient between user's profile similarity and their social similarity.

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

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 geo profile, as previously defined in the Section 6.2. 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 geo 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.

6.3.2.3 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. 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 at follows,

$$\theta_{uk} \sim Gamma(\theta_{uk_2}\Omega + \sum_{g=1}^{G} X_{ug}\lambda_{gk} + \sum_{f=1}^{F} X_{uf}\lambda_{fk}, \beta_u) \tag{6.3}$$

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 $\lambda_{gk}, \lambda_{fk} \sim Gamma(\epsilon_0, \epsilon_0)$. All the rest latent factors are kept the same as in Equation 6.2. 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 6.3.

6.3.3 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, and develop an efficient closed-form Gibbs sampling formula to learn the latent factors with augmented distribution. We first introduce two lemmas [128] that are frequently used in the derivation.

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

Lemma 2. Suppose $y \sim NB(r,p)$ and $c \sim CRT(y,r)$ is a Chinese restaurant table distribution. Then y and c can also be jointly distributed as $y \sim SumLog(c,p)$ and $c \sim Pois(-rln(1-p))$, where $SumLog$ is defined as $y = \sum_{i=1}^{c} x_i$ and $x_i \sim Log(p)$ is logarithmic-distributed random variable.

Gibbs sampling for bHUF+. 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,

$$y_{utlk} \sim Pois(\theta_{uk}\phi_{tk}\psi_{lk})$$

Then by the Poisson-multinomial equivalence, y_{utlk} can be sampled as

$$\{y_{utlk}\} \sim Mult(y_{utl}; \frac{\theta_{uk}\phi_{tk}\psi_{lk}}{\sum_{k=1}^{K}\theta_{uk}\phi_{tk}\psi_{lk}}) \tag{6.4}$$

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

$$y_{uk\cdot} = \sum_{t=1}^{T}\sum_{l=1}^{L} y_{utlk}$$
$$y^{t}_{tk\cdot} = \sum_{u=1}^{U}\sum_{l=1}^{L} y_{utlk}$$
$$y^{l}_{lk\cdot} = \sum_{u=1}^{U}\sum_{t=1}^{T} y_{utlk}$$

We sample each latent factor and parameter with Gibbs sampling as follows,

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

$$\phi_{:k} \sim Dir(y^{t}_{1k\cdot} + \eta^{t}, y^{t}_{2k\cdot} + \eta^{t}, ..., y^{t}_{Kk\cdot} + \eta^{t})$$

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

distribution with

$$\psi_{:k} \sim Dir(y^l_{1k\cdot} + \eta^l, y^l_{2k\cdot} + \eta^l, ..., y^l_{Kk\cdot} + \eta^l)$$

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

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

Since the row of $\phi_{t:}$ and $\psi_{l:}$ 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. This isolation by scale provides a nice property for the derivation of sampling scheme for user related latent factors.

Sampling Ω. Using Poisson distribution's additivity on $y_{utlk} \sim Pois(\theta_{uk}\phi_{tk}\psi_{lk})$ for each t and l gives us $y_{uk\cdot} \sim Pois(\theta_{uk})$. Note that tag and location latent factor disappear due to the unit L_1 norm of Dirichlet distribution. Let $\theta_{uk} = \sum_i \theta_{uki}$, where θ_{uki} are latent parameters defined as

$$\theta_{uki} = \begin{cases} \theta_{k_2}\Omega_{k_2k}, & 1 \le i \le K_2 \\ X_{ug}\lambda_{gk}, & K_2 + 1 \le i \le K_2 + G \\ X_{uf}\lambda_{fk}, & K_2 + G + 1 \le i \le K_2 + G + F \end{cases} \tag{6.5}$$

Thus, by the Poisson-multinomial equivalence, we can sample corresponding y_{uki} with $\{y_{uki}\} \sim Mult(y_{uk\cdot}; \theta_{uki})$. For notation convenience, we let $y_{ukk_2} = y_{uki}$ for $1 \le i \le K_2$, $y_{ukg} = y_{uki}$ for $K_2 + 1 \le i \le K_2 + G$ and $y_{ukf} = y_{uki}$ for the rest of i. Then, for $1 \le i \le K_2$, we have $y_{ukk_2} \sim Pois(\theta_{uki})$ and $\theta_{uki} \sim Gamma(\theta_{k_2}\Omega_{k_2k}, \beta_u)$ due to infinite divisibility of Gamma distribution. Because $\sum_{t=1}^{T} \phi_{tk} = 1$ and $\sum_{l=1}^{L} \psi_{lk} = 1$ due to unit

L_1 norm, we can integrate out θ_{uki}, and have

$$y_{ukk_2} \sim NB(\theta_{uk_2}\Omega_{k_2k}, \gamma_u) \tag{6.6}$$

where $\gamma_u = \frac{1}{1+\beta_u}$. Using data augmentation for NB distribution in Lemma 2, we can sample c indexed by u, k and k_2 with,

$$c_{ukk_2} \sim CRT(y_{ukk_2}, \theta_{uk_2}\Omega_{k_2k}) \tag{6.7}$$

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

$$\Omega_{k_2:} \sim Dir(\sum_{u=1}^{U} c_{u1k_2} + \eta_0, ..., \sum_{u=1}^{U} c_{uKk_2} + \eta_0) \tag{6.8}$$

Sampling θ_{uk_2}. With Equation 6.6 and 6.7 and using the data augmentation in Lemma 2 again, the posterior θ_{uk_2} has a Gamma distribution, and can be sampled with

$$\theta_{uk_2} \sim Gamma(\sum_{k=1}^{K} c_{ukk_2} + \epsilon_0, \beta_u^{(2)} - ln(1-\gamma_u)) \tag{6.9}$$

Sampling λ_{gk}. For each user who have the geographical feature g, i.e., $X_{ug} = 1$, we have $y_{ukg} \sim Pois(\theta_{uki})$ for $K_2 + 1 \leq i \leq K_2 + G$, and $\theta_{uki} \sim Gamma(\lambda_{gk}, \beta_u)$ due to infinite divisibility of Gamma distribution. Integrating out θ_{uki}, we have

$$y_{ukg} \sim NB(\lambda_{gk}, \gamma_u)$$

By using data augmentation in Lemma 2, we can sample c indexed by u, k and g with,

$$c_{ukg} \sim CRT(y_{ukg}, \lambda_{gk}) \tag{6.10}$$

Given c_{ukg}, λ_{gk} has a posterior Gamma distribution which can be sampled with

$$\lambda_{gk} \sim Gamma(\sum_{u \in \mathcal{U}(g)} c_{ukg} + \epsilon_0, \epsilon_0 - \sum_{u \in \mathcal{U}(g)} ln(1 - \gamma_u)) \tag{6.11}$$

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

Sampling λ_{fk}. For each user who have a follower f, i.e., $X_{uf} = 1$, we have $y_{ukf} \sim Pois(\theta_{uki})$ for $K_2 + G + 1 \leq i \leq K_2 + G + F$, and $\theta_{uki} \sim Gamma(\lambda_{fk}, \beta_u)$. We can sample λ_{fk} similarly with λ_{gk} by integrating out θ_{uki} and use data augmentation for negative binomial distribution in Lemma 2. Specifically, we first sample c_{ukf} as follows,

$$c_{ukf} \sim CRT(y_{ukf}, \lambda_{fk})$$

Then, λ_{fk} can be sampled with

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

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

Sampling β_u. Since the rate parameter of Gamma distribution has conjugate Gamma prior, the posterior of β_u is also a Gamma distribution, which can be sampled with

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

Sampling $\beta_u^{(2)}$. Similarly with β_u, it has a posterior Gamma distribution which can be

sampled with

$$\beta_u^{(2)} \sim Gamma((K_2+1)*\epsilon_0, \epsilon_0 + \sum_{k_2=1}^{K_2} \theta_{uk_2})$$

Sampling CRT distribution. Our inference scheme repeatedly samples from Chinese restaurant table (CRT) distributions, for which we use an equivalent sum of Bernoulli distributions [128] as a substitute. Specifically, given $c \sim CRT(y, r)$, it is equivalent to

$$c \sim \sum_{i=1}^{y} Bernoulli(\frac{r}{r+i-1})$$

As we discover from the inference, the further factorization of θ_{uk} is more capable of modeling overdispersed counts by allowing a larger variance-to-mean ratio. Since all scale learning is pushed to θ_{uk} and θ_{uk_2} due to the Dirichlet distribution of $\phi_{:k}$, $\psi_{:k}$ and $\Omega_{k_2:}$, we can reach the following Lemma, with the proof given in [129].

Lemma 3. The variance-to-mean ratio of two-layered factorization of θ_{uk} to θ_{uk_2} in Equation 6.2 is larger than that of one-layered θ_{uk} by a factor of $1 + \frac{\gamma_u}{\beta_u^{(2)}}$.

Proof: From Equation 6.4 and Poisson-multinomial equivalence, we have

$$y_{uk\cdot} \sim Pois(\theta_{uk} \sum_{t=1}^{T} \sum_{l=1}^{L} \phi_{tk}\psi_{lk})$$

Since $\sum_{t=1}^{T} \phi_{tk} = 1$ and $\sum_{l=1}^{L} \psi_{lk} = 1$, we have $y_{uk\cdot} \sim Pois(\theta_{uk})$. 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 $\theta_{uk} \sim Gamma(\epsilon_0, \beta_u)$. By integrating out θ_{uk}, we have

$$y_{uk\cdot} \sim NB(\epsilon_0, \gamma_u)$$

where $\gamma_u = \frac{1}{1+\beta_u}$. Thus the variance-to-mean ratio of $y_{uk\cdot}$ is given by $\frac{1}{1-\gamma_u}$. For case 2,

from Equation 6.6, we have

$$y_{u\cdot k_2} \sim NB(\theta_{uk_2} \sum_{k=1}^{K} \Omega_{k_2 k}, \gamma_u)$$

Since $\sum_{k=1}^{K} \Omega_{k_2 k} = 1$, we have

$$y_{u\cdot k_2} \sim NB(\theta_{uk_2}, \gamma_u)$$

Since $\theta_{uk_2} \sim Gamma(\epsilon_0, \beta_u^{(2)})$, 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

$$Var(y_{u\cdot k_2}) = \frac{\gamma_u}{(1-\gamma_u)^2} \frac{\epsilon_0}{\beta_u^{(2)}} (1 + \frac{\gamma_u}{\beta_u^{(2)}})$$

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 $1 + \frac{\gamma_u}{\beta_u^{(2)}}$.

6.4 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 [107, 19, 130] 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

Table 6.2: 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

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 [18]. We randomly sample two datasets, show in Table 6.2: 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 $\{u \overset{t}{\leftarrow} v, l\}$ 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 predicting and ranking user's geo-topic profiles, we first randomly select 50% of users for training and 50% of the users for testing. From the testing users, we then randomly remove 50% of non-zero popularity counts from each user for testing, and the rest for training. We use the same data split ratio for all of the other experiments. In our experiments, we set the latent dimension K to a large enough number $K = 200$ so that all models are evaluated with sufficient number of latent factors. For K_2, we empirically set it to 50 and examine its impact in supplementary material [129] due to space limit. 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.

We measure the performance with three different metrics: RMSE, precision, recall. RMSE is defined as the root mean square error on the hold-out non-zero counts. Pre-

cision and recall are used 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 $P@n$ and recall $R@n$ are defined as

$$P@n = \frac{1}{N}\sum_{i=1}^{N}\frac{|Pred(u_i, n) \cap GroundTruth(u_i, n)|}{n}$$

$$R@n = \frac{1}{N}\sum_{i=1}^{N}\frac{|Pred(u_i, n) \cap GroundTruth(u_i)|}{|GroundTruth(u_i)|}$$

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**) [131]. This is a pairwise tensor factorization model which captures the pairwise interactions between entities.
- Beta-Negative Binomial CP decomposition (**BNBCP**) [92]. 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**) [96]. This is the-state-of-art Bayesian Poisson tensor model introduced in Preliminaries, 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

Table 6.3: Overall comparison for the US dataset. '†' marks statistically significant difference over the best one-layered baseline. '$*$' marks statistically significant difference over bHUF-2. Both are evaluated according to two sample t-test at significant level 0.05.

Method	RMSE	$P@10$	$P@20$	$R@10$	$R@20$	Imp. (P/R)
PITF	4.7318	0.1249	0.1176	0.0811	0.1224	-
BNBCP	4.5438	0.1604	0.1578	0.1785	0.2218	-
BPTF	4.4973	0.1627	0.1587	0.1788	0.2245	-
bHUF-1	4.4998	0.1534	0.1526	0.1628	0.2157	-
bHUF-2	4.4869	0.1703	0.1664	0.1912	0.2462	$4.76\%^{\dagger}/8.30\%^{\dagger}$
g-bHUF-2	4.4705	0.1787	0.1742	0.1979	0.2558	$4.80\%^{*}/3.70\%^{*}$
s-bHUF-2	4.4767	0.1763	0.1730	0.1964	0.2547	$3.74\%^{*}/3.08\%^{*}$
bHUF+	**4.4642**	**0.1834**	**0.1791**	**0.2022**	**0.2636**	$7.66\%^{*}/6.41\%^{*}$

Table 6.4: Overall comparison for the world dataset. '†' marks statistically significant difference over the best one-layered baseline. '$*$' marks statistically significant difference over bHUF-2. Both are evaluated according to two sample t-test at significant level 0.05.

Method	RMSE	$P@10$	$P@20$	$R@10$	$R@20$	Imp. (P/R)
PITF	4.7318	0.1249	0.1176	0.0811	0.1224	-
BNBCP	4.5438	0.1604	0.1578	0.1785	0.2218	-
BPTF	4.4973	0.1627	0.1587	0.1788	0.2245	-
bHUF-1	4.4998	0.1534	0.1526	0.1628	0.2157	-
bHUF-2	4.4869	0.1703	0.1664	0.1912	0.2462	$4.76\%^{\dagger}/8.30\%^{\dagger}$
g-bHUF-2	4.4705	0.1787	0.1742	0.1979	0.2558	$4.80\%^{*}/3.70\%^{*}$
s-bHUF-2	4.4767	0.1763	0.1730	0.1964	0.2547	$3.74\%^{*}/3.08\%^{*}$
bHUF+	**4.4642**	**0.1834**	**0.1791**	**0.2022**	**0.2636**	$7.66\%^{*}/6.41\%^{*}$

only user's social context (**s-bHUF-2**).

6.4.1 Comparison with Baselines

We first show overall performance comparison in Table 6.3, with all results averaged over 10 runs for Gibbs sampling. The grid size in this experiment is set to $2.5^{\circ} * 2.5^{\circ}$ (equivalent to approximately 175 miles $*$ 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 RMSE, precision and recall. This indicates the superiority of the two-layered user factorization integrated with user's geo-location and social contexts.

Comparison without user's context. As shown in Table 6.3, two-layered user factorization bUHF-2 generally gives better performance than single-layered model. Specifically, it gives an average improvement of 7.06% and 13.4% 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. This confirms that the two-layered factorization has more expressive modeling power over the single-layered model. 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 and a more uniform distribution compared to the first-layer user factor.

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 6.1. Note that these coordinates are not exactly in these cities due to grid granularity.

Comparison with user's context. From Table 6.3, 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 improve-

ment 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, respectively. 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 6.3 — 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.

6.4.2 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 for each user. 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 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 [91] 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 performance comparison in Figure 6.5. 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

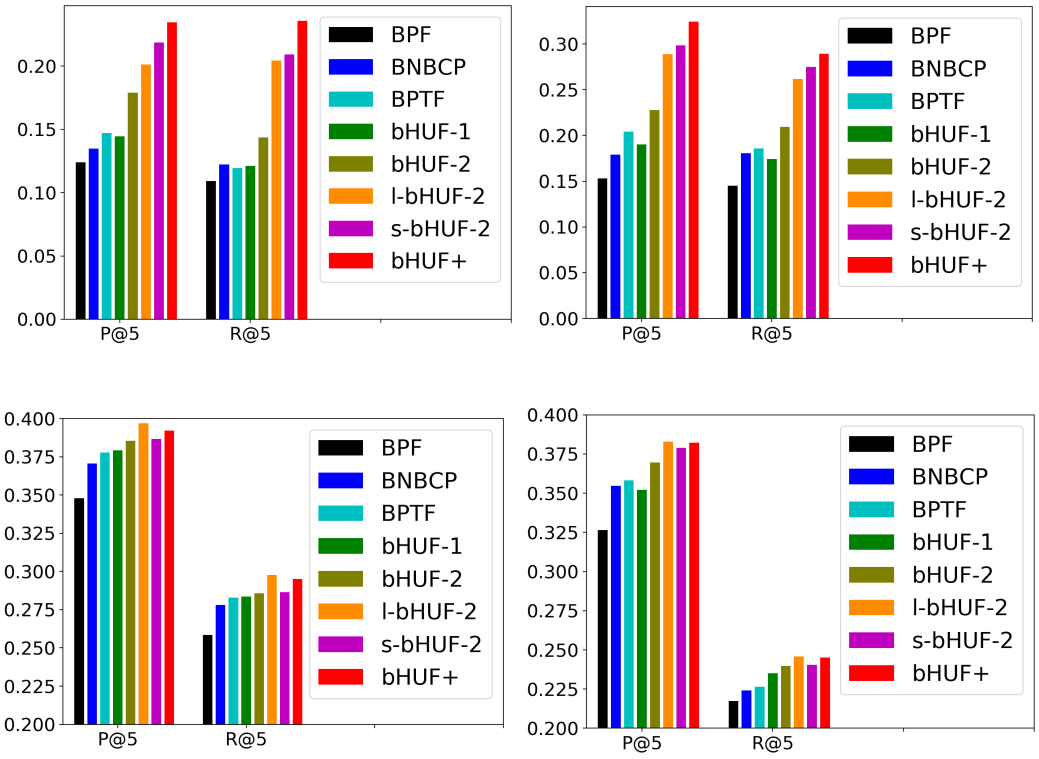


Figure 6.5: Precision and recall for predicting user's topical (top) and geo (down) profile. Left: US. Right: world.

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 6.1.

Predicting user's geo profile. We also show performance comparison in Figure 6.5. Similarly with user topical profile prediction, three dimensional methods generally outperform BPF, and bUHF-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 reason 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.

6.4.3 Parameter Analysis

One important parameter in this study is grid size. A larger size of grid means that the total number of grids is smaller. This indicates a larger geographical area for each location and results in a smaller total number of locations in our models. Thus, if we set the grid large, user geo-topic profile is coarse-grained with respect to locations, while more fine-grained when it is small. Therefore, 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): $0.5^\circ * 0.5^\circ$, $0.8^\circ * 0.8^\circ$, $1.2^\circ * 1.2^\circ$ and $2.5^\circ * 2.5^\circ$. For each size, we compare RMSE, precision and recall for BPTF, bHUF-2, bHUF+. In Figure 6.6, we can see that as the grid size becomes smaller (smaller area for each location), RMSE decreases for both datasets. The reason is that since the total number of user's geo-topic tokens is the same, as we increase the number of grids, the average number of tokens assigned to each grid becomes smaller, resulting in a smaller RMSE. More importantly, we can see that two-layered user factorization performs better than BPTF, and that bHUF+ outperforms bHUF-2 for all grid sizes.

We also show P@10 and R@10 in Figure 6.6. For US dataset, we observe that the precision first increases when the grid size is small, and starts to decrease passing $0.8^\circ * 0.8^\circ$. This indicates that for US dataset, a grid size of about $0.8^\circ * 0.8^\circ$ 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 $2.5^\circ * 2.5^\circ$ (possibly larger), and that it always decreases when the grid gets smaller. This indicates that a coarse-grained geo-partitioning is more 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, similarly with RMSE, 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

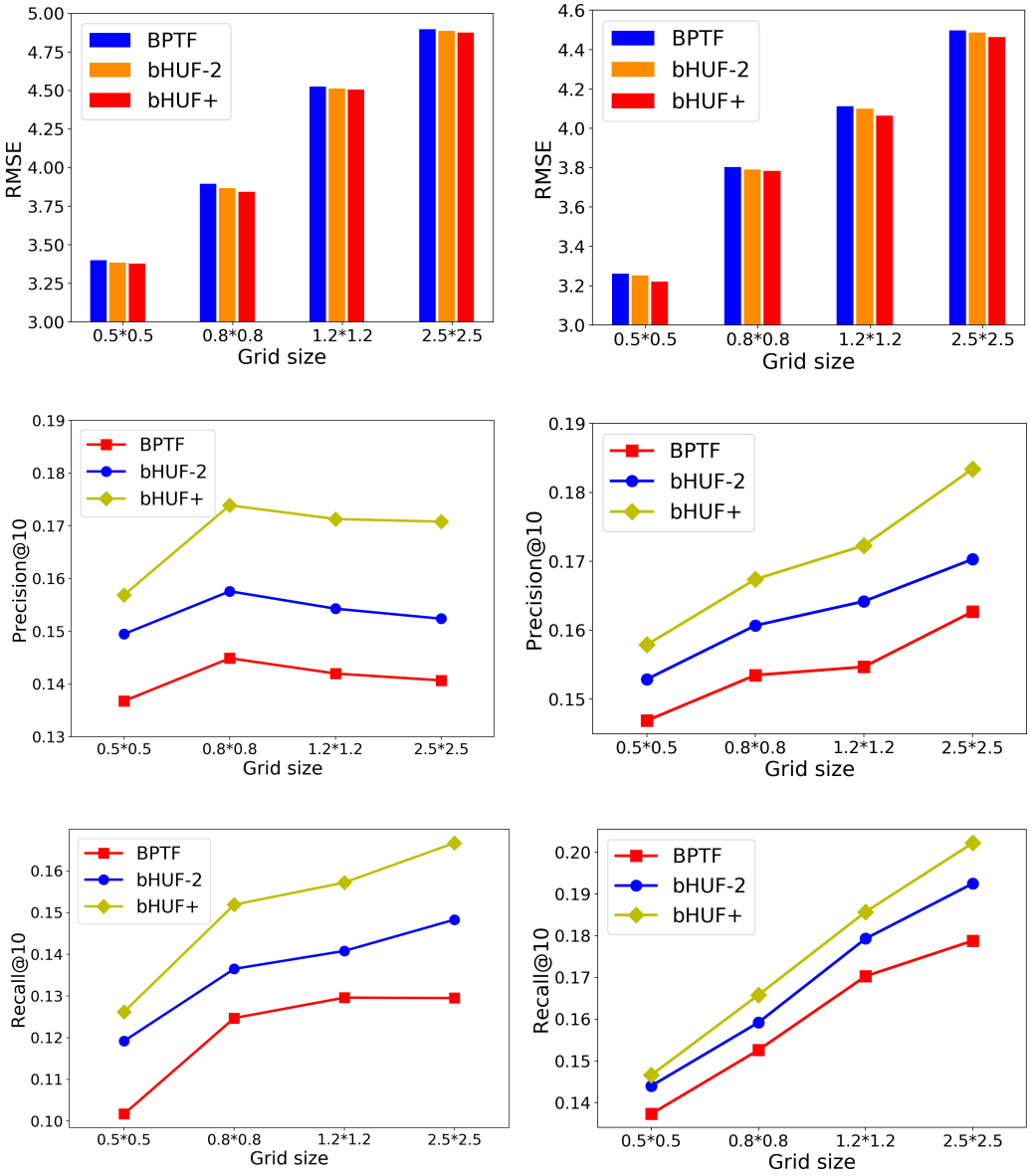


Figure 6.6: Performance comparison with respect to the size of grid. Left: US. Right: world.

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 6.7. 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.

6.5 Conclusion

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

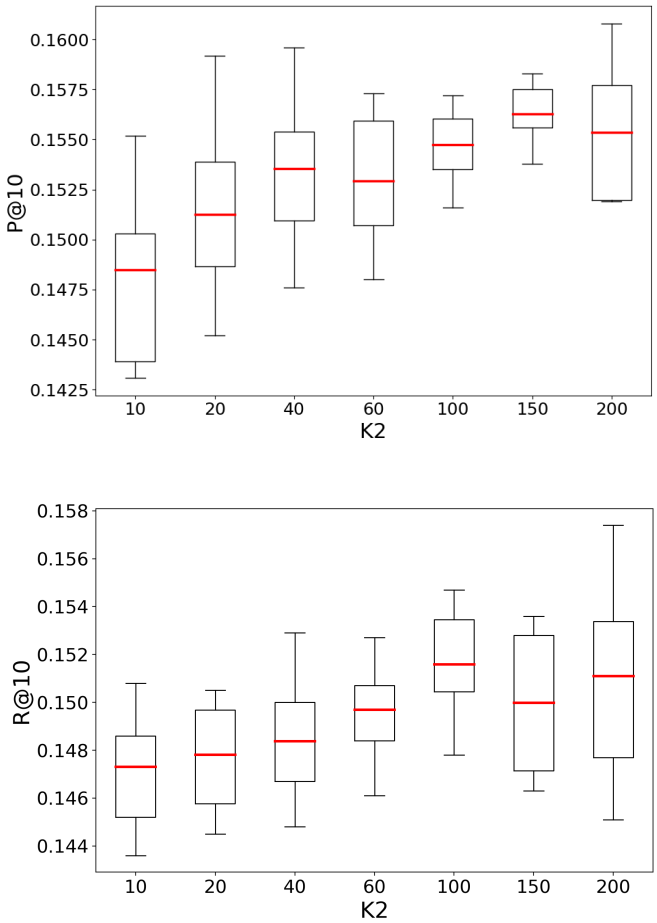


Figure 6.7: Precision@10 and recall@10 with respect to K_2.

7. CONCLUSIONS AND FUTURE WORK

In this section, we present the conclusion of this dissertation and potential future research opportunities.

7.1 Conclusion

In recent years, large-scale information sharing systems – including social media systems, question-answering sites and rating and reviewing applications – have continued to grow and become popular with the number of users who generate, share and consume information. To manage the sheer growth of information generation, there comes the need to enable personalization of information resources for users. A fundamental task in creating personalization systems is to build rich user profiles for better user experience.

Therefore, in this dissertation research, we propose models and algorithms to facilitate the creation of new crowd-powered personalized information sharing systems. Specifically, we first give a principled framework to enable personalization of resources so that information seekers can be matched with customized knowledgeable users based on their previous historical actions; We then focus on creating rich user models that allows accurate modeling of user profiles for long tail users, including discovering user's known-for profile, user's opinion bias and user's geo-topic profile. In particular, this dissertation research made the following contributions:

First, we introduced the problem of personalized expert recommendation and proposed the first principled framework for addressing this problem. To overcome the sparsity issue, we investigated the use of user's contextual information that can be exploited to build robust models of personal expertise. In particular, we studied how spatial preference for personally-valuable expertise varies across regions, across topics, and based on different underlying social communities. We integrated these different forms of preferences into

a matrix factorization-based personalized expert recommender, and demonstrated that the integration of region, topic and social-based locality gives better performance over matrix factorization, and that the combination of these influence gives the best result among all, indicating that geo-spatial, topical and social factors are able to complement each other in personalized expert recommendation.

Second, to tap the knowledge of the majority of users (long tail users) and as a first step to create rich user profiles, we formulated the problem of user known-for profile discovery, and developed a context-based probabilistic model called Bayesian Contextual Poisson Factorization to discover what users are known for by others. Our model considers as input a small fraction of users whose known-for profiles are already known and the vast majority of users for whom we have little (or no) information, learns the implicit relationships between user's known-for profiles and their contextual signals, and finally predicts known-for profiles for those majority of users.

As a second step to create rich user profile, we have explored user's topic-sensitive opinion bias, and demonstrated how user's opinion bias can be exploited to recommend other users with similar opinion in social networks. Specifically, we developed a lightweight semi-supervised system called "BiasWatch" to semi-automatically infer the opinion bias of long-tail users. We proposed an efficient optimization scheme to propagate opinion bias on social and information networks, and evaluated the system by showing the discovered biased themes and their temporal fluctuation, and by comparing against several opinion estimation baselines. This inferred opinion bias can be used as a different dimension to augment user's profiles.

As the last step for user profiling, we have studied how a user's topical profile varies geo-spatially and how we can model a user's geo-spatial known-for profile as the last step in our dissertation for creation of rich user profiles. Specifically, our analysis on the impact of geo-location on user's topical profile indicates the existence of pair-wise inter-

actions between geo-locations and user's topical profile. Motivated by these observations, we proposed the modeling of fine-grained user geo-topic profiles to capture these pair-wise interactions. We then presented a multi-layered Bayesian hierarchical user factorization which can overcome user heterogeneity and learn a more expressive user model. To alleviate the sparsity issue, we investigated how user's contextual information, specifically, user's geo-location and social ties, correlates with one's geo-topic profile, and then proposed to integrate these contexts into the two-layered hierarchical user model for better representation of user's geo-topic preference by others.

7.2 Further Research Opportunities

We identify two future research directions as follows:

- The first is on how we can use inferred user profiles for better resource personalization to users through *interpretable recommendation*. Traditionally, matrix factorization/latent factor models have been at the heart of many popular recommender systems, such as Amazon and Etsy. However, one limitation of the matrix factorization based approach is that the inferred factors can be difficult to interpret and to explain why users prefer some specific resource (items, other users, etc.). One way to improve the interpretability of a personalization system is to utilize user profiles to help explain user's decisions on selected resources. For example, user A interested in cooking selects to follow another user B because B is known as a celebrity cook often posting recipes. This can be interpretable as long as we know user B's known-for profile. As another example, user A selects to become friends with user B because B has a similar opinion on gun control with user A. This is interpretable as long as we know user's opinion profile. These two examples indicate how we can use rich user profiles to improve recommendation performance, not only by providing useful signals about users which can be used by following recommendation machines, but also by providing suggestions to help interpret

user's decisions on resource selections.

- The second is to improve personalization by investigating *temporal changes* of user profiles. Traditionally, user profiles are often considered static without temporal changes. However, a user's taste and preference can often change over time; and a user's profile, e.g., user's interests, opinion and expertise, can also change as users become interested in different items or acquire some new skills. Thus, a research question is to ask how temporal activities of users can affect personalization systems, and how we could develop time-sensitive models to incorporate this influence. Indeed, there have been some works to illustrate and tackle such influence [132, 133, 134, 135]. For example, Lathia et al. [134] have demonstrated how user's rating patterns change over time in Netflix movie datasets. Koren [133] also observed temporal changes in user ratings on movies in Netflix movie datasets and proposed a time-aware latent factor model. More recently, Wu et al. [135] proposed a recurrent recommender system by using Long Short-Term Memory (LSTM) to capture the temporal dynamics of user's rating behavior and achieved state-of-the-art performance. Thus, in this direction, we can ask more specific questions such as how a user's profile changes over time and how we can develop principled time-aware methods to capture these temporal dynamics.

REFERENCES

[1] Facebook, "Facebook key statistics." https://newsroom.fb.com/company-info/, November 2017. Online; accessed 5 November 2017.

[2] Quora, "Quora key statistics." https://www.quora.com/How-many-people-use-Quora-7, Novemeber 2017. Online; accessed 3 November 2017.

[3] S. Aslam, "Pinterest key statistics." https://www.omnicoreagency.com/pinterest-statistics/, Janurary, 2017. Online; accessed 5 November 2017.

[4] Wikipedia, "Wikipedia key statistics." https://stats.wikimedia.org/EN/SummaryEN.htm, September 2017. Online; accessed 5 November 2017.

[5] Facebook, "Evaluating boosted decision trees for billions of users." https://code.facebook.com/posts/975025089299409/evaluating-boosted-decision-trees-for-billions-of-users/, 2017. Online; accessed 5 November 2017.

[6] Twitter, "Using deep learning at scale in twitterâĂŹs timelines." https://blog.twitter.com/engineering/en_us/topics/insights/2017/using-deep-learning-at-scale-in-twitters-timelines.html, 2017. Online; accessed 5 November 2017.

[7] Quora, "A machine learning approach to ranking answers on quora." https://engineering.quora.com/A-Machine-Learning-Approach-to-Ranking-Answers-on-Quora, 2017. Online; accessed 5 November 2017.

[8] Pinterest, "Pinnability: Machine learning in the home feed." https://medium.com/@Pinterest_Engineering/pinnability-machine-learning-in-the-home-feed-64be2074bf60, 2015. Online; accessed 5 November 2017.

[9] Z. Dou, R. Song, and J.-R. Wen, "A large-scale evaluation and analysis of personalized search strategies," in *WWW*, 2007.

[10] A. Majumder and N. Shrivastava, "Know your personalization: learning topic level personalization in online services," in *WWW*, 2013.

[11] F. Qiu and J. Cho, "Automatic identification of user interest for personalized search," in *WWW*, 2006.

[12] A. Sieg, B. Mobasher, and R. Burke, "Web search personalization with ontological user profiles," in *CIKM*, 2007.

[13] I. Guy, N. Zwerdling, I. Ronen, D. Carmel, and E. Uziel, "Social media recommendation based on people and tags," in *SIGIR*, 2010.

[14] H. Ma, D. Zhou, C. Liu, M. R. Lyu, and I. King, "Recommender systems with social regularization," in *WSDM*, 2011.

[15] E. Shmueli, A. Kagian, Y. Koren, and R. Lempel, "Care to comment?: recommendations for commenting on news stories," in *WWW*, 2012.

[16] Z. Zhao, Z. Cheng, L. Hong, and E. H. Chi, "Improving user topic interest profiles by behavior factorization," in *WWW*, 2015.

[17] E. Zhong, N. Liu, Y. Shi, and S. Rajan, "Building discriminative user profiles for large-scale content recommendation," in *SIGKDD*, 2015.

[18] Z. Cheng, J. Caverlee, H. Barthwal, and V. Bachani, "Who is the barbecue king of texas?: A geo-spatial approach to finding local experts on twitter," in *SIGIR*, 2014.

[19] S. Ghosh, N. Sharma, F. Benevenuto, N. Ganguly, and K. Gummadi, "Cognos: crowdsourcing search for topic experts in microblogs," in *SIGIR*, 2012.

[20] J. Weng, E.-P. Lim, J. Jiang, and Q. He, "Twitterrank: finding topic-sensitive influential twitterers," in *WSDM*, 2010.

[21] A. Ahmed, Y. Low, M. Aly, V. Josifovski, and A. J. Smola, "Scalable distributed inference of dynamic user interests for behavioral targeting," in *SIGKDD*, 2011.

[22] B. Wang, C. Wang, J. Bu, C. Chen, W. V. Zhang, D. Cai, and X. He, "Whom to mention: expand the diffusion of tweets by@ recommendation on micro-blogging systems," in *WWW*, 2013.

[23] M. Bastian, M. Hayes, W. Vaughan, S. Shah, P. Skomoroch, H. Kim, S. Uryasev, and C. Lloyd, "Linkedin skills: large-scale topic extraction and inference," in *RecSys*, 2014.

[24] J. Chen, R. Nairn, and E. Chi, "Speak little and well: recommending conversations in online social streams," in *SIGCHI*, 2011.

[25] J. Hannon, M. Bennett, and B. Smyth, "Recommending twitter users to follow using content and collaborative filtering approaches," in *RecSys*, 2010.

[26] L. Hong, A. S. Doumith, and B. D. Davison, "Co-factorization machines: modeling user interests and predicting individual decisions in twitter," in *WSDM*, 2013.

[27] M. Jiang, P. Cui, R. Liu, Q. Yang, F. Wang, W. Zhu, and S. Yang, "Social contextual recommendation," in *CIKM*, 2012.

[28] H. Yin, B. Cui, L. Chen, Z. Hu, and Z. Huang, "A temporal context-aware model for user behavior modeling in social media systems," in *SIGMOD*, 2014.

[29] J. Li, A. Ritter, and E. H. Hovy, "Weakly supervised user profile extraction from twitter.," in *ACL*, 2014.

[30] R. Li, C. Wang, and K. C.-C. Chang, "User profiling in an ego network: co-profiling attributes and relationships," in *WWW*, 2014.

[31] A. Mislove, B. Viswanath, K. P. Gummadi, and P. Druschel, "You are who you know: inferring user profiles in online social networks," in *WSDM*, 2010.

[32] K. Balog, L. Azzopardi, and M. D. Rijke, "Formal models for expert finding in enterprise corpora," in *SIGIR*, 2006.

[33] C. Campbell, P. Maglio, A. Cozzi, and B. Dom, "Expertise identification using email communications," in *CIKM*, 2003.

[34] X. Liu, B. Croft, and M. Koll, "Finding experts in community based question-answering services," in *CIKM*, 2005.

[35] J. Zhang, M. S. Ackerman, and L. Adamic, "Expertise networks in online communities: structure and algorithms," in *WWW*, 2007.

[36] S. Ghosh, N. Sharma, F. Benevenuto, N. Ganguly, and K. Gummadi, "Cognos: Crowdsourcing search for topic experts in microblogs," in *SIGIR*, 2012.

[37] A. Pal and S. Counts, "Identifying topical authorities in microblogs," in *WSDM*, 2011.

[38] J. Weng, E. Lim, J. Jiang, and Q. He, "Twitterrank: Finding topic-sensitive influential twitterers," in *WSDM*, 2010.

[39] Z. Cheng, J. Caverlee, H. Barthwal, and V. Bachani, "Who is the barbecue king of texas?: A geo-spatial approach to finding local experts on twitter," in *SIGIR*, 2014.

[40] X. Zhou, P. Resnick, and Q. Mei, "Classifying the political leaning of news articles and users from user votes.," in *ICWSM*, 2011.

[41] L. Akoglu, "Quantifying political polarity based on bipartite opinion networks," in *ICWSM*, 2014.

[42] L. Adamic and N. Glance, "The political blogosphere and the 2004 u.s. election: Divided they blog," in *LinkKDD*, 2005.

[43] R. Cohen and D. Ruths, "Classifying political orientation on twitter: It's not easy!," in *ICWSM*, 2013.

[44] M. Conover, B. Gonçalves, J. Ratkiewicz, A. Flammini, and F. Menczer, "Predicting the political alignment of twitter users," in *SocialCom*, 2011.

[45] M. Conover, J. Ratkiewicz, M. Francisco, B. Gonçalves, A. Flammini, and F. Menczer, "Political polarization on twitter," in *ICWSM*, 2011.

[46] P. C. Guerra, W. M. Jr., C. Cardie, and R. Kleinberg, "A measure of polarization on social media networks based on community boundaries.," in *ICWSM*, 2013.

[47] A. Livne, M. Simmons, E. Adar, and L. Adamic, "The party is over here: Structure and content in the 2010 election.," in *ICWSM*, 2011.

[48] M. Pennacchiotti and A. Popescu, "Democrats, republicans and starbucks afficionados: User classification in twitter," in *SIGKDD*, 2011.

[49] T. Groseclose and J. Milyo, "A measure of media bias," *The Quarterly Journal of Economics*, 2005.

[50] M. Gentzkow and J. Shapiro, *What Drives Media Slant? Evidence from US Daily Newspapers*. National Bureau of Economic Research Cambridge, Mass., USA, 2006.

[51] Y. Lin, J. Bagrow, and D. Lazer, "Quantifying bias in social and mainstream media," *SIGWEB Newsl.*, 2012.

[52] F. Wong, C. Tan, S. Sen, and M. Chiang, "Quantifying political leaning from tweets and retweets," in *ICWSM*, 2013.

[53] C. Tan, L. Lee, J. Tang, L. Jiang, M. Zhou, and P. Li, "User-level sentiment analysis incorporating social networks," in *SIGKDD*, 2011.

[54] J. Kim, J. Yoo, H. Lim, H. Qiu, Z. Kozareva, and A. Galstyan, "Sentiment prediction using collaborative filtering," in *ICWSM*, 2013.

[55] H. Gao, J. Mahmud, J. Chen, J. Nichols, and M. Zhou, "Modeling user attitude toward controversial topics in online social media," in *ICWSM*, 2014.

[56] Y. Lu, H. Wang, C. Zhai, and D. Roth, "Unsupervised discovery of opposing opinion networks from forum discussions," in *CIKM*, 2012.

[57] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and trends in information retrieval*, 2008.

[58] P. K. Gopalan, L. Charlin, and D. Blei, "Content-based recommendations with poisson factorization," in *NIPS*, 2014.

[59] M. Jamali and L. Lakshmanan, "Heteromf: recommendation in heterogeneous information networks using context dependent factor models," in *WWW*, 2013.

[60] X. Liu and K. Aberer, "Soco: a social network aided context-aware recommender system," in *WWW*, 2013.

[61] H. Ma, H. Yang, M. R. Lyu, and I. King, "Sorec: social recommendation using probabilistic matrix factorization," in *CIKM*, 2008.

[62] A. P. Singh and G. J. Gordon, "Relational learning via collective matrix factorization," in *SIGKDD*, 2008.

[63] C. Wang and D. M. Blei, "Collaborative topic modeling for recommending scientific articles," in *SIGKDD*, 2011.

[64] X. Zhang, J. Cheng, T. Yuan, B. Niu, and H. Lu, "Toprec: domain-specific recommendation through community topic mining in social network," in *WWW*, 2013.

[65] Q. Liu, E. Chen, H. Xiong, C. H. Ding, and J. Chen, "Enhancing collaborative filtering by user interest expansion via personalized ranking," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 2012.

[66] R. Ottoni, D. B. Las Casas, J. P. Pesce, W. Meira Jr, C. Wilson, and A. Mislove, "Of pins and tweets: Investigating how users behave across image-and text-based social networks.," in *ICWSM*, 2014.

[67] M. Qiu, F. Zhu, and J. Jiang, "It is not just what we say, but how we say them: Lda-based behavior-topic model," in *Proceedings of the 2013 SIAM International Conference on Data Mining*, pp. 794–802, SIAM, 2013.

[68] X. Zhang, J. Cheng, T. Yuan, B. Niu, and H. Lu, "Toprec: Domain-specific recommendation through community topic mining in social network," in *WWW*, 2013.

[69] L. Hu, A. Sun, and Y. Liu, "Your neighbors affect your ratings: on geographical neighborhood influence to rating prediction," in *SIGIR*, 2014.

[70] H. Lu, J. Caverlee, and W. Niu, "Discovering what you're known for: A contextual poisson factorization approach," in *RecSys*, 2016.

[71] T. Lappas, K. Punera, and T. Sarlos, "Mining tags using social endorsement networks," in *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pp. 195–204, ACM, 2011.

[72] L. Charlin, R. Ranganath, J. McInerney, and D. M. Blei, "Dynamic poisson factorization," in *RecSys*, 2015.

[73] X. Hu, J. Tang, H. Gao, and H. Liu, "Unsupervised sentiment analysis with emotional signals," in *Proceedings of the 22nd international conference on World Wide Web*, pp. 607–618, ACM, 2013.

[74] X. Hu, L. Tang, J. Tang, and H. Liu, "Exploiting social relations for sentiment analysis in microblogging," in *Proceedings of the sixth ACM international conference on Web search and data mining*, pp. 537–546, ACM, 2013.

[75] Y. Hu, K. Talamadupula, S. Kambhampati, *et al.*, "Dude, srsly?: The surprisingly formal nature of twitter's language.," in *ICWSM*, 2013.

[76] C. Cheng, H. Yang, I. King, and M. R. Lyu, "Fused matrix factorization with geographical and social influence in location-based social networks," in *AAAI*, 2012.

[77] B. Liu, Y. Fu, Z. Yao, and H. Xiong, "Learning geographical preferences for point-of-interest recommendation," in *SIGKDD*, 2013.

[78] D. Lian, C. Zhao, X. Xie, G. Sun, E. Chen, and Y. Rui, "Geomf: joint geographical modeling and matrix factorization for point-of-interest recommendation," in *SIGKDD*, 2014.

[79] Y. Liu, W. Wei, A. Sun, and C. Miao, "Exploiting geographical neighborhood characteristics for location recommendation," in *CIKM*, 2014.

[80] H. Lu and J. Caverlee, "Exploiting geo-spatial preference for personalized expert recommendation," in *RecSys*, 2015.

[81] M. Ye, P. Yin, W.-C. Lee, and D.-L. Lee, "Exploiting geographical influence for collaborative point-of-interest recommendation," in *SIGIR*, 2011.

[82] W. Zhang, J. Wang, and W. Feng, "Combining latent factor model with location features for event-based group recommendation," in *SIGKDD*, 2013.

[83] V. W. Zheng, Y. Zheng, X. Xie, and Q. Yang, "Collaborative location and activity recommendations with gps history data," in *WWW*, 2010.

[84] M. Ye, P. Yin, W. Lee, and D. Lee, "Exploiting geographical influence for collaborative point-of-interest recommendation," in *SIGIR*, 2011.

[85] B. Liu, Y. Fu, Z. Yao, and H. Xiong, "Learning geographical preferences for point-of-interest recommendation," in *SIGKDD*, 2013.

[86] L. Hu, A. Sun, and Y. Liu, "Your neighbors affect your ratings: On geographical neighborhood influence to rating prediction," in *SIGIR*, 2014.

[87] K. Chen, T. Chen, G. Zheng, Q. Jin, E. Yao, and Y. Yu, "Collaborative personalized tweet recommendation," in *SIGIR*, 2012.

[88] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, "Bpr: Bayesian personalized ranking from implicit feedback," in *UAI*, 2009.

[89] S. Rendle and L. Schmidt-Thieme, "Pairwise interaction tensor factorization for personalized tag recommendation," in *WSDM*, 2010.

[90] A. Krohn-Grimberghe, L. Drumond, C. Freudenthaler, and L. Schmidt-Thieme, "Multi-relational matrix factorization using bayesian personalized ranking for social network data," in *WSDM*, 2012.

[91] P. Gopalan, J. M. Hofman, and D. M. Blei, "Scalable recommendation with hierarchical poisson factorization," in *UAI*, 2015.

[92] C. Hu, P. Rai, C. Chen, M. Harding, and L. Carin, "Scalable bayesian non-negative tensor factorization for massive count data," in *ECML-PKDD*, 2015.

[93] R. Yu, A. Gelfand, S. Rajan, C. Shahabi, and Y. Liu, "Geographic segmentation via latent poisson factor model," in *WSDM*, 2016.

[94] M. Zhou, L. Hannah, D. B. Dunson, and L. Carin, "Beta-negative binomial process and poisson factor analysis," in *AISTATS*, 2012.

[95] A. J. Chaney, D. M. Blei, and T. Eliassi-Rad, "A probabilistic model for using social networks in personalized item recommendation," in *RecSys*, 2015.

[96] A. Schein, J. Paisley, D. M. Blei, and H. Wallach, "Bayesian poisson tensor factorization for inferring multilateral relations from sparse dyadic event counts," in *SIGKDD*, 2015.

[97] M. Zhou, Y. Cong, and B. Chen, "The poisson gamma belief network," in *NIPS*, 2015.

[98] X. Amatriain, N. Lathia, J. M. Pujol, H. Kwak, and N. Oliver, "The wisdom of the few: a collaborative filtering approach based on expert opinions from the web," in *SIGIR*, 2009.

[99] Google, "Overview of local guides," April 2015.

[100] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, 2009.

[101] H. Ma, D. Zhou, C. Liu, M. Lyu, and I. King, "Recommender systems with social regularization," in *WSDM*, 2011.

[102] P. Gopalan, J. M. Hofman, and D. M. Blei, "Scalable recommendation with poisson factorization," *arXiv preprint arXiv:1311.1704*, 2013.

[103] M. McPherson, L. Smith-Lovin, and J. M. Cook, "Birds of a feather: Homophily in social networks," *Annual review of sociology*, pp. 415–444, 2001.

[104] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul, "An introduction to variational methods for graphical models," *Machine learning*, 1999.

[105] N. Johnson, A. Kemp, and S. Kotz, *Univariate Discrete Distributions*. John Wiley and Sons, 2005.

[106] M. D. Hoffman, D. M. Blei, C. Wang, and J. Paisley, "Stochastic variational inference," *JMLR*, 2013.

[107] P. Bhattacharya, S. Ghosh, J. Kulshrestha, M. Mondal, M. B. Zafar, N. Ganguly, and K. P. Gummadi, "Deep twitter diving: Exploring topical groups in microblogs at scale," in *CSCW*, 2014.

[108] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *JMLR*, 2003.

[109] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer-Verlag New York, Inc., 2006.

[110] J. Weston, S. Bengio, and N. Usunier, "Wsabie: Scaling up to large vocabulary image annotation," in *IJCAI*, 2011.

[111] L. Yang, T. Sun, M. Zhang, and Q. Mei, "We know what@ you# tag: does the dual role affect hashtag adoption?," in *WWW*, 2012.

[112] X. Wang, F. Wei, X. Liu, M. Zhou, and M. Zhang, "Topic sentiment analysis in twitter: a graph-based hashtag sentiment classification approach," in *CIKM*, 2011.

[113] Z. Zheng, X. Wu, and R. Srihari, "Feature selection for text categorization on imbalanced data," *SIGKDD Explor. Newsl.*, 2004.

[114] O. Owoputi, B. O'Connor, C. Dyer, K. Gimpel, N. Schneider, and N. Smith, "Improved part-of-speech tagging for online conversational text with word clusters," in *NAACL*, 2013.

[115] D. Boyd, S. Golder, and G. Lotan, "Tweet, tweet, retweet: Conversational aspects of retweeting on twitter," in *HICSS*, 2010.

[116] J. Tang, R. Hong, S. Yan, T. Chua, G. Qi, and R. Jain, "Image annotation by knn-sparse graph-based label propagation over noisily tagged web images," *ACM TIST*, 2011.

[117] M. Schmidt, "L1 general," May 2015.

[118] C. Marshall and F. Shipman, "Experiences surveying the crowd: Reflections on methods, participation, and reliability," in *WebSci*, 2013.

[119] J. Landis and G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, 1977.

[120] S. Nowak and S. Rüger, "How reliable are annotations via crowdsourcing: A study about inter-annotator agreement for multi-label image annotation," in *MIR*, 2010.

[121] J. Shi and J. Malik, "Normalized cuts and image segmentation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2000.

[122] R. Agrawal, S. Rajagopalan, R. Srikant, and Y. Xu, "Mining newsgroups using networks arising from social behavior," in *WWW*, 2003.

[123] A. Murakami and R. Raymond, "Support or oppose?: Classifying positions in online debates from reply activities and opinion expressions," in *COLING*, 2010.

[124] D. Zhou, O. Bousquet, T. Lal, J. Weston, and B. SchÃűlkopf, "Learning with local and global consistency," in *NIPS*, 2004.

[125] D. Liben-Nowell and J. Kleinberg, "The link-prediction problem for social networks," *Journal of the American society for information science and technology*, 2007.

[126] E. C. Chi and T. G. Kolda, "On tensors, sparsity, and nonnegative factorizations," *SIAM Journal on Matrix Analysis and Applications*, 2012.

[127] W. R. Tobler, "A computer movie simulating urban growth in the detroit region," *Economic geography*, 1970.

[128] M. Zhou and L. Carin, "Negative binomial process count and mixture modeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015.

[129] "Supplementary material for bayesian hierarchical user factorization," tech. rep., 2017.

[130] V. Rakesh, D. Singh, B. Vinzamuri, and C. K. Reddy, "Personalized recommendation of twitter lists using content and network information," in *ICWSM*, 2014.

[131] S. Rendle and L. Schmidt-Thieme, "Pairwise interaction tensor factorization for personalized tag recommendation," in *WSDM*, 2010.

[132] S. Chang, Y. Zhang, J. Tang, D. Yin, Y. Chang, M. A. Hasegawa-Johnson, and T. S. Huang, "Streaming recommender systems," in *Proceedings of the 26th International Conference on World Wide Web*, pp. 381–389, International World Wide Web Conferences Steering Committee, 2017.

[133] Y. Koren, "Collaborative filtering with temporal dynamics," *Communications of the ACM*, vol. 53, no. 4, pp. 89–97, 2010.

[134] N. Lathia, S. Hailes, L. Capra, and X. Amatriain, "Temporal diversity in recommender systems," in *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, pp. 210–217, ACM, 2010.

[135] C.-Y. Wu, A. Ahmed, A. Beutel, A. J. Smola, and H. Jing, "Recurrent recommender networks," in *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pp. 495–503, ACM, 2017.