Information Processing and Management 56 (2019) 939–954

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

Item diversified recommendation based on influence diffusion

Huimin Huang[a], Hong Shen[*,a,b], Zaiqiao Meng[a]

[a] School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China
[b] School of Computer Science, University of Adelaide, Adelaide, Australia

ARTICLE INFO

Keywords:
Item recommendation
Influence diffusion
Social networks

ABSTRACT

Recently, the high popularity of social networks accelerates the development of item recommendation. Integrating the influence diffusion of social networks in recommendation systems is a challenging task since topic distribution over users and items is latent and user topic interest may change over time. In this paper, we propose a dynamic generative model for item recommendation which captures the potential influence logs based on the community-level topic influence diffusion to infer the latent topic distribution over users and items. Our model enables tracking the time-varying distributions of topic interest and topic popularity over communities in social networks. A collapsed Gibbs sampling algorithm is proposed to train the model, and an improved diversification algorithm is proposed to obtain item diversified recommendation list. Extensive experiments are conducted to evaluate the effectiveness and efficiency of our method. The results validate our approach and show the superiority of our method compared with state-of-the-art diversified recommendation methods.

1. Introduction

The time-varying user preferences have put essential importance on the dynamic modeling of users and items in recommender systems. Traditional recommendation algorithms rank recommended items either using historical interactions, i.e., items are recommended to a target user based on the similar items this user has ever selected (item-based collaborative filtering) or other users with similar preferences (user-based collaborative filtering) (Ren, Liang, Li, Wang, & Rijke, 2017). This tends to make the recommendation lists to be monotonous, limited to narrowing popular or over-specified items, and hard to cover all of the user's interests (Wu et al., 2016). Thus, to improve users' satisfaction with the recommendation results, it is reasonable to take both relevance and diversification into consideration and provide diversified recommendation results.

A variety of methods to improve item diversified recommendation, including heuristic algorithms (Adamopoulos & Tuzhilin, 2014; Lee & Lee, 2015; Adomavicius & Kwon, 2012; Gan & Jiang, 2013; Belém, Santos, Almeida, & Gonçalves, 2013; Ziegler, Mcnee, Konstan, & Lausen, 2005) and approximation algorithms (Wu et al., 2016; Qin & Zhu, 2013), have enjoyed impressive success but still have large drawbacks. First, they tend to directly exploit item categories or tags as topics. However, in the real world, items are usually with multiple features and semantic information is latent, hence, a more capable method to extract semantic information is needed. Second, it may be unreliable for these methods to define diversity with reverse order of similarity. Moreover, the existing item diversified recommendation algorithms are static, while users preferences are likely to change over time.

To overcome above limitations, we propose i**T**em **D**iversified **R**ecommendation based on **I**nfluence **D**iffusion **(TDRID)**, which integrates a dynamic probabilistic generative model with a diversification algorithm to extract semantic information and perform

* Corresponding author.
E-mail addresses: huanghm45@gmail.com (H. Huang), hongsh01@gmail.com (H. Shen).

https://doi.org/10.1016/j.ipm.2019.01.006
Received 19 August 2018; Received in revised form 15 January 2019; Accepted 19 January 2019
Available online 15 February 2019
0306-4573/

diversified recommendation effectively. Our motivation is from the observations of marketing and psychology (Kalish, 1985; Chatterjee & Eliashberg, 1990) that (1) Friend relationship influences user making decision on item selection; (2) The process of influence propagation can reveal users' preference; (3) User interest may change over time. Thus, we take an influence-diffusion perspective to learn dynamic user interest and item characteristics, and propose a new diversified recommendation method, jointly applying influence diffusion modeling and probabilistic topic modeling. Besides, we consider to capture user interest by use of community-level extraction because studies on sociology and marketing demonstrate that the behaviors of some individual users are highly volatile (Kalish, 1985), which makes it difficult to accurately mine user preferences and diffusion properties at individual level, whereas community-level model not only avoids volatility of individual users but also alleviates the problem of data sparsity.

The proposed method of item diversified recommendation is based on two premises. First, purchase behaviors are relevant to communities' interests. Fundamentally, community is a group of users that have denser connections among the group than with the rest of the network (Su & Havens, 2015). Users engage in social network as members of communities. Lots of previous studies have found that users in communities share common properties or attributes (Yang, McAuley, & Leskovec, Yang et al.; Li, Hu, Jian, & Liu, 2017; Fang, Cheng, Luo, & Hu, 2016). In other words, communities have their attributes including community interest and community popularity. Assuming each user may behave as the member of different communities in different settings, user interest is interrelated with community interest by the community membership distribution of the user. As is known that purchase behaviors are relevant to user interest, in consequence, purchase behaviors are related with community interest.

Second, community interest is dynamic. This is because in the real world, user interest is varying over time. Besides, a user's interests is likely different from the others' interests toward him/her. For example, assuming user u potentially influences user v making decision on selection of item i, user v's interest is unnecessarily same as user u's. Similar to cases happening between users, assuming community c' potentially influences community c making decision on selection of item i, interest of community c is unnecessarily the same as interest of community c'. Hence, we define the topic popularity of community c' and topic interests of community c to capture this kind of phenomenon between information diffusion source and information diffusion target.

Fig. 1 presents the overview of our TDRID. The input is a friend relationship network along with purchase logs. Our goal is to discover the association between influence diffusion and item diversified recommendation as well as information dynamics in propagation process.

Diversified recommendation based on influence diffusion is challenging. Communities and topic distribution over users and items are hidden, which are three critical factors in the process of influence diffusion. We need to model the correlation between them suitably. Besides, user topic interest may change over time, and we are required to simultaneously consider temporal factors in our latent variable model.

To sum up, our major contributions in this paper are:

(1) We propose a new item diversified recommendation method based on influence diffusion, which mines latent semantic

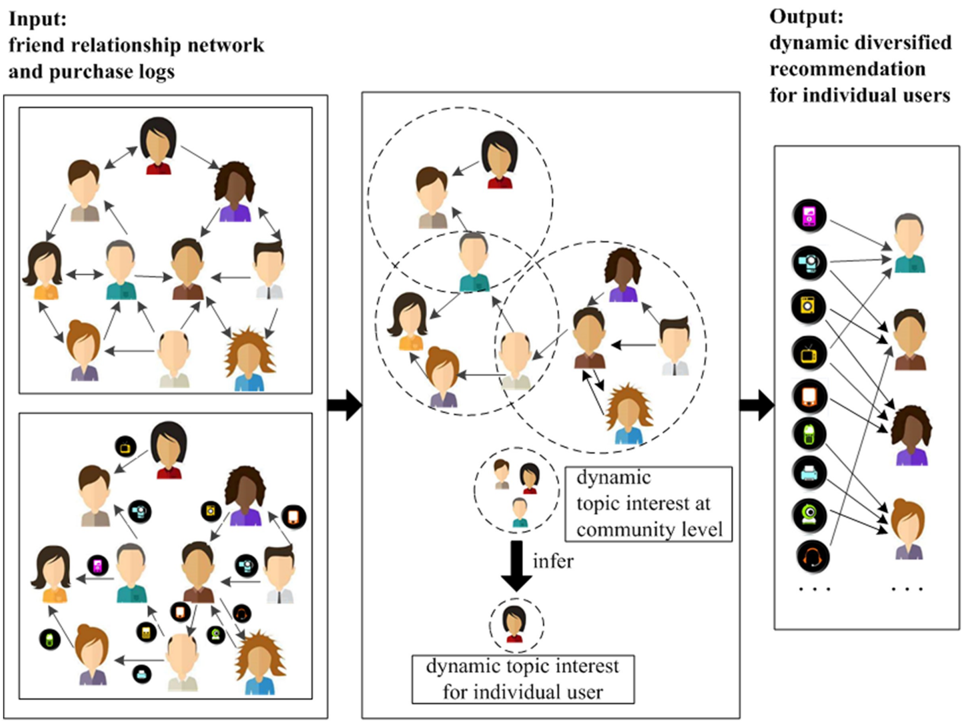

Fig. 1. Overview of our TDRID.

information in the process of influence propagation and returns recommendation lists of items that are recent and relevant to the individual topic interest and cover as many topics as possible. To the best of our knowledge, item diversified recommendation has never been studied from this perspective.

(2) We develop a dynamic latent variable model. It captures dynamic user interest across communities avoiding volatility of individual users and alleviating the data-sparsity problem. Also it addresses the problem of cold-start by conducting probabilistic semantic analysis on the new item added in the system.

(3) We propose an adaptable version of PM-2 diversification algorithm which is time sensitive and applicable to our dynamic model.

(4) We provide a thorough analysis of our item diversification recommendation method, and perform experiments to demonstrate its effectiveness and efficiency compared with state-of-the-art item diversified recommendation algorithms.

The rest of this paper is organized as follows. Section 2 reviews previous work. Section 3 presents the preliminaries including notations, terminologies and our task. Section 4 details the method of item diversified recommendation based on influence diffusion. Section 5 reports our experiment analysis. Section 6 concludes the paper.

2. Previous work

2.1. Item diversified recommendation

Item diversified recommendation algorithms have attracted increasing interests in recent years, but the existing algorithms which can catch semantic information and achieve diversification effectively are scarce. Ziegler et al. (2005) proposes semantic diversification approach which computes similarity between pairs of items through their semantic information feature vectors, and sorts products list in reverse order of similarity to obtain dissimilarity rank, lacking capable method to capture semantic information and perform diversification yet. Belém et al. (2013) exploits explicit categories (tags) as topics and produces a ranking function that assigns scores to candidate tags based on relevance estimates as well as diversification. With an initial ranking of recommended tags based on relevance as input, its diversification algorithm produces diverse ranking with a greedy strategy similar to xQuAD Algorithm (Santos, Macdonald, & Ounis, 2010), but ignores the dynamic of users' interests.

In addition, approximation algorithms of diversified recommendation have been explored. Wu et al. (2016) proposes a recommendation framework that considers both traditional relevance-based scores and the new coverage measure, and further proves that the goals of maximizing relevance and coverage measures are NP-hard. A greedy algorithm is proposed in Qin and Zhu (2013) which integrates a rating function as well as an entropy regularizer, measuring the relevances and the diversifications respectively and reaching a trade-off. This algorithm supplies approximation ratio of $1 - 1/e - \varepsilon$ but has a $O(k(|\Omega| + k)^3)$ time complexity, where $|\Omega|$ denotes the number of rated items of a single user and k is the size of the recommendation list. Although this approximation algorithm has performance guarantee and relatively good execution efficiency, it formulates the objective set function of diversity recommendation as a linear combination of a rating function and an entropy regularizer without considering the dynamic of users' interests.

2.2. Probabilistic topic modeling

Probability topic modeling is aimed at finding the topic structure hidden in the massive documents. Since Latent Dirichlet Allocation (LDA) was first proposed by Blei, Ng, and Jordan (2003), probability topic model has been widely studied in the field of probability semantic analysis. Yan, Guo, Lan, and Cheng (2013) proposes a short text modeling method, biterm topic model (BTM), which directly models the generation of word co-occurrence patterns (i.e. biterms) in the whole corpus. Yin and Wang (2014) applies a Dirichlet multinomial mixture model-based approach for short text clustering. Yang, Kotov, Mohan, and Lu (2015) incorporates demographic information of review authors into topic modeling to accomplish review sentiment classification and user attribute prediction. Besides, the dynamic topic model has been proposed including the online multi-scale dynamic topic model (Iwata, Yamada, Sakurai, & Ueda, 2010), dynamic clustering of streaming short documents (Liang, Yilmaz, & Kanoulas, 2016), dynamic user clustering topic model (Zhao et al., 2016).

With the high popularity of social medias, recently some works propose to integrates content topic discovery and social influence analysis in the same generative process. Representative models include extension of LAD model to handle popular nodes in social networks (Cha & Cho, 2012), social-relational topic model (Guo, Wu, Wang, & Tan, 2015), Followship-LDA (FLDA) model (Bi, Tian, Balmin, Balmin, & Cho, 2014), topic interest generative model for heterogeneous networks (Liu, Tang, Han, Jiang, & Yang, 2010).

In the above probabilistic topic models, the models integrating content topic discovery and social influence analysis are most relevant to our research, but they only focus on extracting influence at individual level, rather than community level, which makes them suffering from high variance result based on volatile individual behaviors.

2.3. Social influence diffusion

Initially motivated by viral marketing, the problem of social influence diffusion emerges which increasingly concerns the propagation of ideas, opinions or rumors etc. through social networks. Barbieri, Bonchi, and Manco (2012) proposes Topic-aware Independent Cascade (TIC) Model, where the user-to-user influence probabilities depend on the topics, and an Expectation-Maximization method is used to learn parameters, but the semantic information in the networks cannot be fully utilized, thereby it is hard to better capture the properties of influence propagation.

Chen, Wei, and Zhang (2012) considers a deadline constraint to reflect the time-critical effect in influence diffusion and develops a new propagation model, independent cascade model with meeting events to capture the delay of propagation in time. It only adds time factor into the traditional influence diffusion model without taking advantage of semantic information, and besides, it only applies to independent cascade diffusion model.

Hu, Yao, Cui, and Xing (2015) models retweet network and posts by a community level diffusion model. Zhang, Lyu, and Zhang (2017) models retweeting process with a hierarchical community-level information diffusion model, which combines semantic analysis and social influence analysis. None of these papers mentioned above consider the simultaneous leaning topic interest and topic-item relevance, nor study item diversified recommendation problem applying social influence diffusion models. To the best of our knowledge, we are the first to study item diversified recommendation through extracting time-varying distributions of topic interest and topic popularity over communities in social networks, jointly modeling influence diffusion and probabilistic topic.

3. Preliminaries

Before we describe our method, we first briefly review the notations and terminologies and introduce our task to be addressed.

3.1. Notations and definitions

Let directed graph $\mathcal{G} = (\mathcal{U}, \mathcal{E})$ be the friend relationship network, where $\mathcal{U}$ denotes the set of nodes and $\mathcal{E}$ denotes the set of directed links. Let $\mathcal{L}$ be the set of *purchase logs* which are represented as triples {(*User, Item, Time*)}, i.e., (u, i, t), indicating that user u selected item i at time t. We assume that each item $i \in \mathcal{I}$ is associated with a mixture of words as its binary features, and let $\boldsymbol{A} \in \{0, 1\}^{M \times F}$ be item attribute matrix with $\boldsymbol{a}_i \in \{0, 1\}^F$ being the attribute vector of i, and M, F being the number of items and attributes respectively. For ease of reference, we list the basic notations in Table 1, and briefly introduce some basic concepts that we use in the remainder.

Definition 1 (*Potential-influence log***).** A potential-influence log $d = (u, v, i)$ is constructed from two purchase logs (u, i, t_p) and (v, i, t_q) where $t_q - t_p \le \Delta$ and $(u, v) \in \mathcal{E}$, meaning that v purchasing item i is potentially influenced by user u. In practice, threshold Δ is set manually (Barbieri et al., 2012). The set of potential-influence logs is denoted as $\mathcal{D}$.

Definition 2 (*Community***).** A community $c \in [1, 2, ..., C]$ consists of nodes which share some common properties or preferences. Each user $v \in \mathcal{U}$ may belong to multiple communities with a multinomial distribution π_v. Accordingly, π_{vc} indicates the probability that user v be the member of community c.

Definition 3 (*Topic***).** A topic $z \in [1, 2...Z]$ is a multinomial distribution over topics of items, denoted as $\boldsymbol{\phi}_i$. Every potential-influence log d is associated with a topic z_i which is sampled for item i according to its topic distribution.

Definition 4 (*Dynamic community-topic relevant***).** Every community c has two dynamic topic-relevant distributions, including dynamic topic interest $\theta_{c,z,t}$ and dynamic topic popularity $\theta'_{c,z,t}$, which represent that c influence the purchase of an item on topic z at time t as the propagation target and the propagation source respectively.

Table 1
Basic notations used in this paper.

Symbol	Description
k	Size of recommendation list for a single user
C, Z, M, F	Number of communities, topics, items and attributes
$\mathcal{I}, \mathcal{U}, \mathcal{E}, \mathcal{D}$	Set of items, users, links, potential-influence logs
U, E, D	Number of users, links, potential-influence logs
$\mathcal{E}_v, \mathcal{D}_v$	The set of links towards v, the set of potential-influence logs with v as a propagation target
E_v, D_v	Size of $\mathcal{E}_v, \mathcal{D}_v$
d	Potential-influence log
$\boldsymbol{\psi}_z$	Multinomial distribution over attribute values in matrix $\boldsymbol{A}$ specific to topic z
$\boldsymbol{\phi}_i$	Multinomial distribution over topics of item i
z_i	The topic of item i
$\boldsymbol{w}_i$	The bag of attribute values of item i
π_v	Multinomial distribution over community specific to user v
c_d, c'_d	Communities associated with user v and u of potential-influence log
s_e, s'_e	Communities associated with v and u of friend relation link
$\boldsymbol{\theta}_{c,t}$	The topic interest of community c at time t
$\boldsymbol{\theta}_{c,t}'$	The topic popularity of community c at time t
$\eta_{c'c}$	Community-level influence strength of community c' and c
I_e	Indicator of the existence of link e
I_d	Indicator of the existence of potential-influence log d
ε	i.e.$(\varepsilon_0, \varepsilon_1)$, Beta priors to $\boldsymbol{\eta}$
$\rho, \boldsymbol{\alpha}_t, \alpha'_t, \beta, \omega$	Dirichlet priors to $\pi_v, \boldsymbol{\theta}_{c,t}, \theta'_{c,t}, \boldsymbol{\psi}, \boldsymbol{\phi}$

Definition 5 (*Community-level influence***).** We assume that each potential-influence log $d = (u, v, i)$ depends on communities c', c of u, v according to the two dynamic community-topic relevant $\theta'_{c',z,t}$, $\boldsymbol{\theta}_{c,z,t}$ and the community-level influence strength $\eta_{c',c}$.

3.2. The diversified recommendation task

We formally define the task of item diversified recommendation based on influence diffusion as follows:

Given result size k, a network $\mathcal{G}$, the set of potential-influence logs $\mathcal{D}$, the item attribute matrix $\boldsymbol{A}$, the diversified recommendation task is to (1) first infer the latent user-topic relevance p(z|v) and topic-item relevance p(i|z) while capturing the community-level influence according to the potential-influence logs; (2) then obtain a final recommended items list of size k which are diverse w.r.t. the latent topics.

According to the task, we need to solve two problems:

(1) designing a model Ξ to infer latent variables:

$$\mathcal{G}, \mathcal{D}, \boldsymbol{A} \xrightarrow{\Xi} p(z|v), p(i|z)$$

(2) running a ranking function $\mathcal{F}$ that satisfies:

$$p(z|v), p(i|z) \xrightarrow{\mathcal{F}} \mathcal{S}$$

where $\mathcal{S}$ is a set of recommendation lists for each user, and $\forall\ \mathcal{S}_v \in \mathcal{S}, |\mathcal{S}_v| = k$.

4. The proposed method

To solve above problems, we propose the **TDRID**, an item diversified recommendation method that integrates a dynamic probabilistic generative model with a diversification algorithm to infer latent variables and perform diversified recommendation effectively. The dynamic probabilistic generative model allows the community-topic relevance to change over time, while the diversification algorithm makes the recommendation list diverse.

4.1. Overview of TDRID

Algorithm 1 presents the overview of our TDRID which composes of two steps.

In Step 1, we utilize the *Dynamic Latent Variable Model* to capture the changes of topic interests and infer the most recent topic interests at community level, i.e., $\boldsymbol{\theta}_{c,t}$, which is assumed to represent the community topic interests properly. Besides, we infer the individual-level topic interests from $\boldsymbol{\theta}_{c,t}$ to be the user-topic relevance $p(z|v, t)$ according to v's community membership distribution π_v. In Step 2, we perform diversification with our improved PM-2 diversification algorithm, which is time sensitive and addresses the drawback of original PM-2. We will discuss the two steps in the following subsections.

4.2. Dynamic latent variable model

4.2.1. Model description

The dynamic latent variable model used in TDRID is a probabilistic generative model of item attributes, user edges and their potential-influence logs inferred from the purchase logs. We take both influence logs and friend relation network into account when modeling influence diffusion for the observation that influence logs are sparser than friend relationship links. Thus, we use friend relation network as a supplementary to interact with potential-influence logs to mimic the process of influence diffusion. There are basically three components in our generative process, i.e., the generation of item attributes, user edges and their potential-influence logs. The probabilistic graphical description of our generative model is presented in Fig. 2, and its generative process is presented in Algorithm 2.

To generate attributes of an item $i \in \mathcal{I}$, we draw a multinomial ϕ_i from a Dirichlet prior ω; then for each attribute w_{ij} in item i: (1) first draw a topic z_i from multinomial ϕ_i; (2) then draw an attribute w_{ij} from multinomial ψ_{z_i}. In the case that new items are added in the system, LDA (Latent Dirichlet Allocation) can still be used to extract item-topic relevance, which makes our model address the problem of cold start.

For generating the user edge, we assume each user may behave as the member of different communities in different settings. Thus, we denote each user to belong to different communities with a multinomial distribution π_v, and use π_{vc} to denote the probability that user v acts as a member of community c. For each link $e(u, v) \in E$, user u acts as member of community s'_e, and user v acts as member of community s_e. A Bernoulli distribution $\eta_{s'_e s_e}$ is used to denote influence strength between community s'_e and community s_e, which determines the probability of the presence of link e.

When generating the potential-influence logs, for each potential-influence log $d = (u, v, i) \in D$, user u acts as member of community c'_d, and user v acts as member of community c_d. We define the topic interest of community c and topic popularity of community c' respectively as $\boldsymbol{\theta}_{c,t}$ and $\theta'_{c',t}$. A Bernoulli distribution $\eta_{c'_d c_d}$ is used to denote influence strength between community c'_d and community c_d. The presence and absence of influence d between community c'_d and community c_d at time t is denoted with a indicator $I_{d,t}$, which is determined by three factors: $\eta_{c'_d c_d}$, $\boldsymbol{\theta}_{c,t}$, $\theta'_{c',t}$. As in Hu et al. (2015), the priors to $\eta_{c'c}$ is set as a $Beta(\varepsilon_0, \varepsilon_1)$, in the

Infer the dynamic user-topic relevance $p(z \mid v, t)$ and topic-item relevance $p(i \mid z)$ through *Dynamic Latent Variable Model*;
Perform diversification to obtain diversified recommendation list $\mathcal{S}_v$ for each user v through improved **PM-2**;

Algorithm 1. Overview of TDRID.

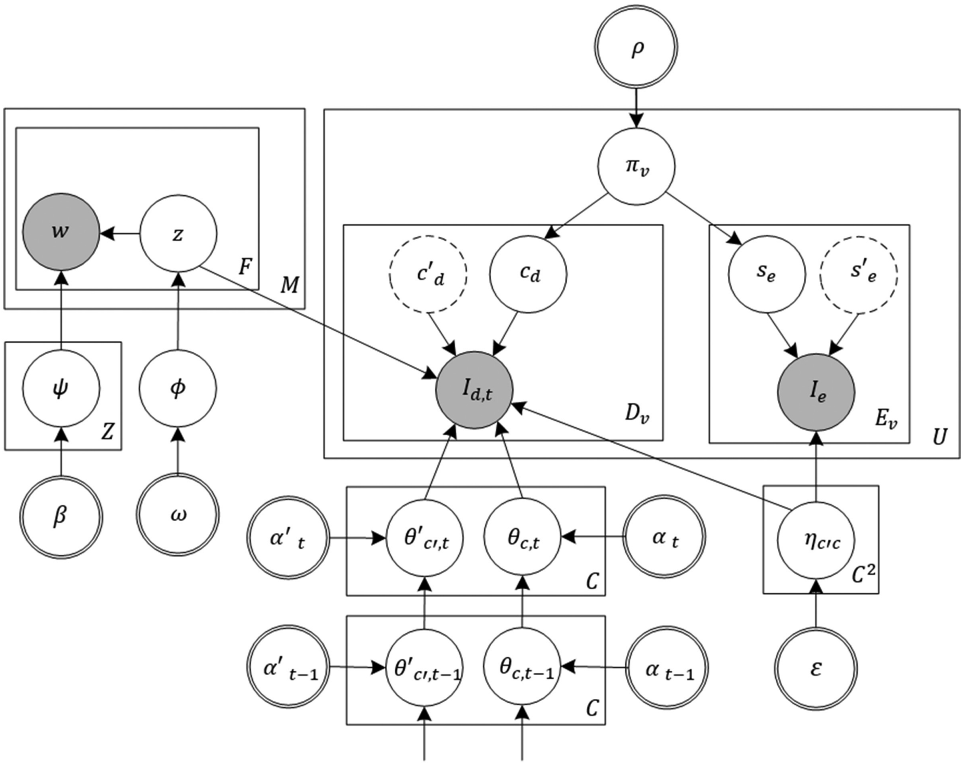

Fig. 2. Graphical representation of our dynamic latent variable model.

hyperparameters of which, negative samples are implicitly modeled, i.e., $\varepsilon_0 = \zeta \ln(N_{neg}/C^2)$ and $\varepsilon_1 = 0.1$, where $N_{neg} = U(U-1)(1+D/E) - D - E$ and ζ is a tunable weight.

To model the dynamic changes of user topic interests at community level, we use the following generative process. Given a potential-influence log $d = (u, v, i) \in D$, suppose we already have the topic interest of target community at the previous time $t-1$, $\theta_{c,t-1}$. The topic interests of target community c at time t, $\theta_{c,t}$ can be drawn from $Dirichlet(\alpha_t \theta_{c,t-1})$ (Liang et al., 2016). Similarly, the topic popularity of source community c' at time t, $\theta'_{c',t}$ can be drawn from $Dirichlet(\alpha'_t \theta'_{c',t-1})$. In this way, we can obtain the most recent user topic interest, which are assumed to represent user interest properly.

4.2.2. Model inference

Following Heinrich (2005), Liang et al. (2016) and Liang, Ren, Yilmaz, and Kanoulas (2017), we use collapsed Gibbs sampling method to estimate the parameters of our model. As item-topic generation is independent of influence diffusion in our model, our Gibbs sampling process is composed of two steps. First, we iteratively sample hidden variables $\boldsymbol{z}$. After the Gibbs Sampling converges, we use those samples to estimate the unknown parameters $\boldsymbol{\phi}, \boldsymbol{\psi}$ and obtain topic distribution z_i of each item i; Second, we iteratively sample hidden variables $\boldsymbol{s}, \boldsymbol{s'}, \boldsymbol{c}, \boldsymbol{c'}$. After the Gibbs Sampling converges, we estimate the unknown parameters π, $\boldsymbol{\theta}, \boldsymbol{\theta'}, \boldsymbol{\eta}$ with the samples. Due to space constraints, we show only the derived Gibbs sampling formulas, omitting the detailed derivation process.

Sample latent topic z_{ij} for each attribute w_{ij} in matrix $\boldsymbol{A}$**:**

$$P(z_{ij} = z|\boldsymbol{z}_{\neg ij}, \boldsymbol{w}) \propto \frac{n_{i,z} + \omega}{\sum_Z (n_{i,z} + \omega)} \cdot \frac{n_{z,w_{ij}} + \beta}{\sum_{M\times F} (n_{z,w_{ij}} + \beta)} \tag{1}$$

where $n_{i,z}$ refers to the number of times that topic z has been observed with an attribute in item i, and $n_{z,w_{ij}}$ denote the number of times that attribute w_{ij} has been observed with topic z.

After the Gibbs Sampling of latent topic converges, we can estimate parameters $\boldsymbol{\phi}, \boldsymbol{\psi}$ as follows.

$$\phi_{iz} = \frac{n_{i,z} + \omega}{\sum_Z (n_{i,z} + \omega)} \tag{2}$$

$$\psi_{zw} = \frac{n_{z,w} + \beta}{\sum_{M\times F} (n_{z,w} + \beta)} \tag{3}$$

Also, we can infer item-topic relevance $P(z|i)$ by counts.

for each topic $z = 1, 2, \ldots Z$ **do**
  Draw $\boldsymbol{\psi}_z \sim Dir(\beta)$;
end for
for each item $i = 1, 2, \ldots M$ **do**
  Draw $z_i \sim Mul(\boldsymbol{\phi}_i)$;
  Draw $w_i \sim Mul(\boldsymbol{\psi}_{z_i})$;
end for
for each community $c = 1, 2, \ldots C$ **do**
  Draw topic interest distribution $\theta_{c,t} \sim Dir(\alpha_t \boldsymbol{\theta}_{c,t-1})$;
  Draw topic popularity distribution $\theta'_{c,t} \sim Dir(\alpha'_t \boldsymbol{\theta}'_{c,t-1})$;
  for each community $c' = 1, 2, \ldots C$ **do**
    Draw the community-level diffusion probability $\eta_{c'c} \sim Beta(\varepsilon_0, \varepsilon_1)$;
  end for
end for
for each user $v = 1, 2, \ldots U$ **do**
  Draw the distribution over communities $\boldsymbol{\pi}_u \sim Dir(\rho)$;
  for each link $e = (u, v) \in \mathcal{E}_v$ **do**
    Draw user u's community $s'_e \sim Mul(\boldsymbol{\pi}_u)$;
    Draw user v's community $s_e \sim Mul(\boldsymbol{\pi}_v)$;
    Draw the existence indicator $I_e \sim Ber(\eta_{s'_e s_e})$;
  end for
  for each potential-influence log $d = (u, v, i) \in \mathcal{D}_v$ **do**
    Draw user u's community $c'_d \sim Mul(\boldsymbol{\pi}_u)$;
    Draw user v's community $c_d \sim Mul(\boldsymbol{\pi}_v)$;
    Draw item i's topic indicator $z_i \sim Mul(\boldsymbol{\phi}_i)$;
    Draw the existence indicator at time t,
    $I_{d,t} \sim Ber(\eta_{c'_d c_d} \theta_{c_d, z_i, t} \theta'_{c'_d, z_i, t})$;
  end for
end for

Algorithm 2. GenerativeProcess.

Sample community indicator s_e, s'_e for each edge $e = (u, v) \in \mathcal{E}$**:**

$$P(s_e = c, s'_e = c' | \mathbf{s}_{\neg e}, \mathbf{s}'_{\neg e}, \cdot) \propto \frac{n_{u,c'} + \rho}{\sum_C (n_{u,c'} + \rho)} \cdot \frac{n_{v,c} + \rho}{\sum_C (n_{v,c} + \rho)} \cdot \frac{n_{c',c} + \varepsilon_1}{n_{c',c} + \varepsilon_0 + \varepsilon_1} \tag{4}$$

where $n_{u,c'}$ indicates the number of times when user u acts as a member of community c' in all links and potential-influence logs, $n_{c',c}$ is the number of links from community c' to community c and potential-influence logs with community c' as source community and community c as target community.

Sample community indicator c_d, c'_d for each potential-influence log $d = (u, v, i) \in \mathcal{D}$**:**

As in Liang et al. (2016), we apply fixed-point iteration to get the optimal α_t at time t and derive updated rule of α_t as follows by maximizing the joint distribution in the fixed-point iteration:

$$\alpha_{t,z} \leftarrow \frac{\alpha_{t,z}(\Psi(n_{c,z,t} + \alpha_{t,z}\theta_{c,z,t-1}) - \Psi(\alpha_{t,z}\theta_{c,z,t-1}))}{\Psi(\sum_Z n_{c,z,t} + \alpha_{t,z}\theta_{c,z,t-1}) - \Psi(\sum_Z \alpha_{t,z}\theta_{c,z,t-1})} \tag{5}$$

where $\Psi(\cdot)$ defined by $\Psi(x) = \frac{\partial \log \Gamma(x)}{\partial x}$ is the digamma function.

For each potential-influence log $d(u, v, i)$, we first draw item i's topic distribution $\boldsymbol{z_i}$ and apply the following Gibbs sampling formula:

$$\begin{aligned} P(c_d = c, c'_d &= c' | \mathbf{c}_{\neg d}, \mathbf{c}'_{\neg d}, \mathbf{z}_{\neg i}, z_i = z, \theta_{c,z,t-1}, \cdot) \\ &\propto \frac{n_{u,c'} + \rho}{\sum_C (n_{u,c'} + \rho)} \cdot \frac{n_{v,c} + \rho}{\sum_C (n_{v,c} + \rho)} \cdot \frac{n_{c',c} + \varepsilon_1}{n_{c',c} + \varepsilon_0 + \varepsilon_1} \\ &\cdot \frac{n_{c,z,t} + \alpha_{t,z}\theta_{c,z,t-1}}{\sum_Z (n_{c,z,t} + \alpha_{t,z}\theta_{c,z,t-1})} \cdot \frac{n_{c',z,t} + \alpha'_{t,z}\theta'_{c',z,t-1}}{\sum_Z (n_{c',z,t} + \alpha'_{t,z}\theta'_{c',z,t-1})} \end{aligned} \tag{6}$$

where $n_{c,z,t}$ denotes the number of potential-influence logs which is relevant to topic z and with community c as propagation target at time t, and $n_{c',z,t}$ denotes the number of potential-influence logs which is relevant to topic z and with community c' as propagation source at time t.

Every potential-influence log d is associated with a topic z_i according to the item i in $d(u, v, i)$. Thus, according to item-topic relevance $P(z|i)$, $n_{c,z,t}$ can be computed as $\Sigma_M n_{c,i,t} \cdot P(z|i)$, where $n_{c,i,t}$ is the number of potential-influence logs which is about item i and with community c as propagation target at time t. Similarly, $n_{c',z,t}$ can be computed as $\sum_M n_{c',i,t} \cdot P(z|i)$, where $n_{c',i,t}$ is the number of potential-influence logs which is about item i and with community c' as propagation source at time t. Thus replacing $n_{c,z,t}$ in Formula (6) with $\Sigma_M n_{c,i,t} \cdot P(z|i)$ and replacing $n_{c',z,t}$ in Formula (6) with $\sum_M n_{c',i,t} \cdot P(z|i)$, we can get Formula (7) as follows.

$$\begin{aligned} P(c_d = c, c'_d &= c' | \mathbf{c}_{\neg d}, \mathbf{c}'_{\neg d}, \mathbf{z}_{\neg i}, z_i = z, \theta_{c,z,t-1}, \cdot) \\ &\propto \frac{n_{u,c'} + \rho}{\sum_C (n_{u,c'} + \rho)} \cdot \frac{n_{v,c} + \rho}{\sum_C (n_{v,c} + \rho)} \cdot \frac{n_{c,c'} + \varepsilon_1}{n_{c,c'} + \varepsilon_0 + \varepsilon_1} \\ &\cdot \frac{\sum_M n_{c,i,t} \cdot P(z|i) + \alpha_t \theta_{c,z,t-1}}{\sum_Z (\sum_M n_{c,i,t} \cdot P(z|i) + \alpha_t \theta_{c,z,t-1})} \\ &\cdot \frac{\sum_M n_{c',i,t} \cdot P(z|i) + \alpha'_t \theta'_{c',z,t-1}}{\sum_Z (\sum_M n_{c',i,t} \cdot P(z|i) + \alpha'_t \theta'_{c',z,t-1})} \end{aligned} \tag{7}$$

After all Gibbs Sampling converges, the unknown parameters can be estimated as follows.

$$\pi_{v,c} = \frac{n_{v,c} + \rho}{\sum_C (n_{v,c} + \rho)} \tag{8}$$

$$\eta_{c',c} = \frac{n_{c',c} + \varepsilon_1}{n_{c',c} + \varepsilon_0 + \varepsilon_1} \tag{9}$$

$$\theta_{c,z,t} = \frac{\sum_M n_{c,i,t} \cdot P(z|i) + \alpha_t \theta_{c,z,t-1}}{\sum_Z (\sum_M n_{c,i,t} \cdot P(z|i) + \alpha_t \theta_{c,z,t-1})} \tag{10}$$

.

$$\theta'_{c',z,t} = \frac{\sum_M n_{c',i,t} \cdot P(z|i) + \alpha'_t \theta'_{c',z,t-1}}{\sum_Z (\sum_M n_{c',i,t} \cdot P(z|i) + \alpha'_t \theta'_{c',z,t-1})} \tag{11}$$

.

4.2.3. Inference topic interests of individual users and topic-item relevance

Now we can map the community-level topic interests into user-level topic interests according to the community membership distribution of the user:

$$P(z|v, t) = \sum_C \pi_{vc} \cdot \theta_{c,z,t}. \tag{12}$$

With item-topic relevance $P(z|i)$ derived from our generative model, we can infer topic-item relevance $P(i|z)$ using Bayesian Formula:

$$P(i|z) = \frac{P(z|i) \cdot P(i)}{\sum_Z (P(z|i) \cdot P(i))} \tag{13}$$

where $P(i)$ denotes the probability of potential-influence logs related with item i among all potential-influence logs.

4.3. Item diversified recommendation

PM-2 (Dang & Croft, 2012) is a probabilistic adaptation of the Sainte–Lague method which assigns seats to the political party with the largest quotient (quotient is based on the votes the political parties receive and the number of seats they have taken) such that the number of seats assigned for each party is proportional to the votes they have received.

Owing to the probabilistic adaptation of the Sainte–Lague method, PM-2 algorithm has been widely adopted in some diversified retrieval methods, which involves probability computation more. However, as far as its application to our diversified recommendation is concerned, PM-2 appears some drawbacks: (1) Due to the usage of standard query likelihood model to estimate the relevance $P(d|z)$ between the document d and the aspect z, PM-2 has a non-trivial complexity. Likewise, it is non-trivial to compute the aspect probability $v_{z|q}$, which is often set to be uniform. (2) The user-topic distribution $P(z|v, t)$ derived from our model is time-sensitive, since our latent topic model is dynamic. Nevertheless, PM-2 is static, that makes it unadaptable to our diversified recommendation.

We adapt PM-2 algorithm by considering some variables relevant to recommendation system, i.e., topic-item distribution, user-topic distribution etc., and adding time element into PM-2 to make it adaptable to our dynamic latent variable model well. Our proposed item diversified recommendation is presented in Algorithm 3.

Input: The number of users U
The number of items M
The number of topics Z
An integer k
Current user-topic distribution $\boldsymbol{\theta}_{u,t}$
Topic-item distribution $P(i|z)$
Output: a diversified recommendation list for each of users.
1: **for** user $v = 1 \ldots U$ **do**
2: $\quad \mathcal{S}_{v,t} \leftarrow \varnothing, \mathcal{R}_{v,t} \leftarrow \mathcal{I}$;
3: **end for**
4: **for** user $v = 1 \ldots U$ **do**
5: $\quad$ **for** k positions in the ranked list $\mathcal{S}_{v,t}$ **do**
6: $\quad\quad$ **for** topic $z = 1 \ldots Z$ **do**
7: $\quad\quad\quad qt[z \mid v, t] = \frac{\theta_{v,z,t}}{2s_{z|v,t}+1}$;
8: $\quad\quad$ **end for**
9: $\quad\quad z^* \leftarrow \arg\max_z qt[z \mid v, t]$;
10: $\quad\quad i^* \leftarrow \arg\max_{i \in R} \lambda \cdot qt[z^*|v, t]P(i|z^*) + \sum_{z \neq z^*}(1-\lambda)qt[z|v, t]P(i|z)$;
11: $\quad\quad \mathcal{S}_{v,t} \leftarrow \mathcal{S}_{v,t} \cup \{i^*\}$;
12: $\quad\quad \mathcal{R}_{v,t} \leftarrow \mathcal{R}_{v,t} \setminus i^*$;
13: $\quad\quad$ **for** topic $z = 1 \ldots Z$ **do**
14: $\quad\quad\quad s_{z|v} \leftarrow s_{z|v} + \frac{P(i^*|z)}{\sum_{z'=1}^{Z} P(i^*|z')}$;
15: $\quad\quad$ **end for**
16: $\quad$ **end for**
17: **end for**

Algorithm 3. Generate Diversified Recommendation Lists.

From line 1 to line 3, Algorithm 3 starts with a ranked list $\mathcal{S}_{v,t}$ with k empty seats and a set of items $\mathcal{R}_{v,t}$ with k items. From line 4 to line 17, for each user, the algorithm fills the ranked list $\mathcal{S}_{v,t}$ by selecting k items, which satisfy the rule that items are relevant to the user's interest over topics and cover topics as many as possible. From line 6 to 8, the algorithm computes the quotient, $qt[z|v, t]$ for each topic z according to Sainte–Lague method (Sainte–Lague method assigning seats to the political party with the largest quotient that is based on the votes the political parties receive and the number of seats they have taken (Dang & Croft, 2012)). For each seat in the item list, the algorithm selects the topic z^* with largest quotient, $qt[z|v, t]$ following SainteLague method, then fills the seat with the item i^*, and thereafter updates $\mathcal{S}_{v,t}$ and $\mathcal{R}_{v,t}$ (Lines 9-12). For each topic, the algorithm updates the "portion" of seats in $\mathcal{S}_{v,t}$ occupied by each of the topics z according to how relevant it is to the selected item i^* (Lines 13-14).

4.4. Complexity analysis

In this section, we analyze the time complexity of our proposed TDRID Algorithm. Firstly, we discuss the time complexity in the generative process. Given parameters C and Z fixed, the time of the generative process is linear to the size of input data and times of iterations, i.e., $O(T_1 \cdot M + T_2 \cdot (E + D))$, where T_1 denotes times of iterations in sampling latent topics and T_2 denotes times of iterations in sampling $\boldsymbol{s}, \boldsymbol{s'}, \boldsymbol{c}, \boldsymbol{c'}$. In each iteration in sampling latent topics, the time complexity is $O(Z \cdot F)$. In each iteration in sampling $\boldsymbol{s}, \boldsymbol{s'}, \boldsymbol{c}, \boldsymbol{c'}$, the time complexity of sampling community indicators associated with each link is $O(C^2)$, and sampling community indicators associated with each potential-influence log runs in $O(C \cdot t_\alpha)$ time, where t_α represents the time expended on computing α_t.

Secondly, it takes the $O(C \cdot U)$ time to infer user-level topic interests from community-level topic interests, and given item-topic relevance, it takes $O(Z \cdot E)$ time to infer topic-item relevance.

Thirdly, we discuss the time complexity in the diversifying process. The time complexity of computing diversified recommendation lists for all users is $O(U \cdot k \cdot Z \cdot M)$, while computing diversified recommendation lists for a single users runs in $O(k \cdot Z \cdot M)$ time.

5. Experimental setup

5.1. Data set

We use two real-world and publicly available datasets: Yelp data set challenge 2013, Yelp data set challenge 2014[1] and Digg

[1] https://www.yelp.com/

Table 2
Feature comparison of different methods.

	Features		
	Social	Topic	Dynamic
MF + MMR			
xTReD			
AIR + MaxSum		•	
ID + xQuAD	•	•	•
TDRID	•	•	•

Table 3
Performance of All Methods on Relevance and Diversification Metrics.

		nDCG	α-nDCG	ERR-IA	S-Recall
Digg dataset	MF + MMR	0.3805	0.2850	0.1723	0.5127
	xTReD	0.4103	0.3004	0.1822	0.5461
	AIR + MaxSum	0.4621	0.3316	0.2143	0.5934
	ID + xQuAD	0.4732	0.3501	0.2215	0.6142
	TDRID	0.5015	0.3716	0.2502	0.6573
Yelp 2013	MF + MMR	0.3960	0.3004	0.1844	0.5436
	xTReD	0.4255	0.3185	0.1902	0.5927
	AIR + MaxSum	0.4803	0.3592	0.2283	0.6414
	ID + xQuAD	0.4902	0.3761	0.2375	0.6542
	TDRID	0.5115	0.3973	0.2652	0.6813
Yelp 2014	MF + MMR	0.4027	0.3240	0.1973	0.5809
	xTReD	0.4358	0.3283	0.2025	0.6336
	AIR + MaxSum	0.4861	0.3708	0.2416	0.6982
	ID + xQuAD	0.5093	0.3914	0.2517	0.7301
	TDRID	0.5369	0.4177	0.2841	0.7517

dataset[2]. Yelp data set challenge 2013 contains 70,816 users, 622,873 links, 15,584 items, Yelp data set challenge 2014 contains 366,715 users, 2,949,285 links, 61,184 items, and Digg dataset contains 30,358 users, 99,846 directed links, 7100 items.

For Yelp dataset challenge 2013 and Yelp dataset challenge 2014, we do not have data directly reflecting purchase time, so we assume the time at which the user reviews the item is the time when the item is purchased. Besides, we disregard the unfrequent behaviors of repeated review for the same item.

5.2. Baselines and parameter settings

We compare our TDRID with several state-of-art baselines. Table 2 lists the features of all methods.

MMR(Maximal Marginal Relevance) (Carbonell & Goldstein, 1998) is a diversification algorithm, and accordingly, MF + MMR is a diversity enhancement algorithm applied to the matrix factorization algorithm. xTReD(explicit Tag Recommendation Diversifier) (Belém et al., 2013) exploits explicit categories(tags) as topics. Although, to some extent, it is of semantic diversification, its tasks exclude topic extraction and dynamic.

AIR + MaxSum and ID + xQuAD are diversification algorithms MaxSum (Gollapudi & Sharma, 2009) and xQuAD (Santos et al., 2010) respectively applied to the topic diffusion model in Barbieri et al. (2012) and our dynamic latent variable model. Therefore, AIR + MaxSum possesses the features of topic extraction and ID + xQuAD possesses the features of social, topic extraction and dynamic.

We train TDRID, the baselines ID + xQuAD by setting λ to vary from 0 to 1.0, setting the number of topics Z to vary from 2 to 16, setting the number of communities C as 50, 100, 150, setting the size of recommendation list as 5, 10, 20 and setting the time interval as week, month, quarter, half year, year. For baselines MF + MMR, xTReD and AIR + MaxSum, we let λ to vary from 0 to 1.0 in the training process, set k as 5, 10, 20.

We use 60% potential influence logs as train set and 20% potential influence logs as validation set, and set 20% potential influence logs and all links in friend relationship graph as test set. We will report only the optimal performance with best parameters for fair comparison.

Note that Dirichlet hyperparameters have low impact on model performance. Therefore, we set hyperparameters {ρ, β, ω, ε} as fixed values following the common strategy in Yin, Cui, Sun, Hu, and Chen (2014), Hu et al. (2015), i.e., $\rho = 50/C$, $\beta = 0.01$, $\omega = 50/Z$, and ε_0, ε_1 are set as Section 4.2.1.

[2] https://www.digg.com/

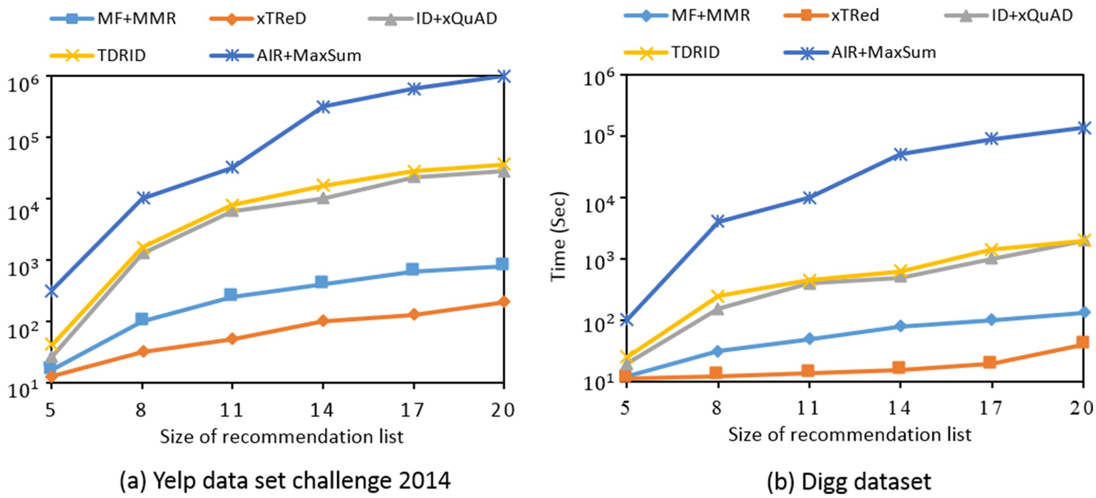

Fig. 3. Comparison on running time.

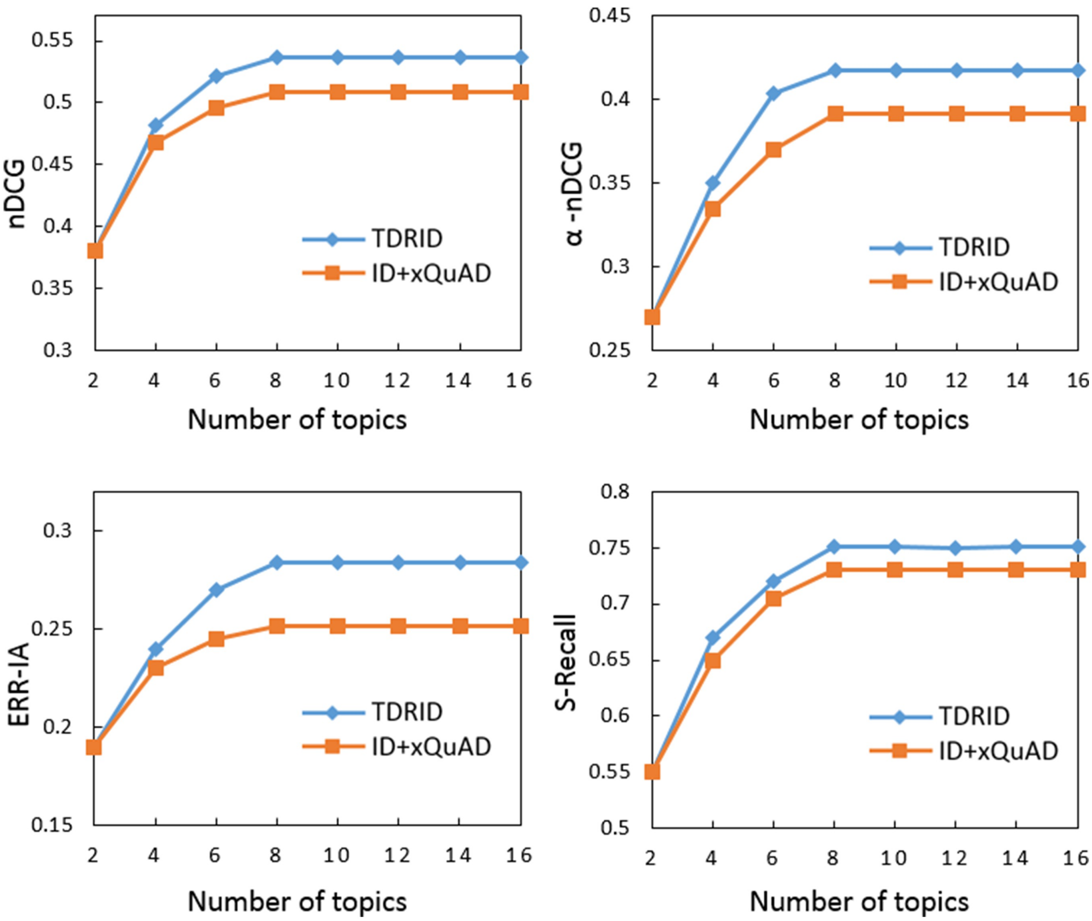

Fig. 4. Comparisons between ID + xQuAD and TDRID on Yelp data set challenge 2014 when varying the number of topics, for nDCG, α-nDCG, ERR-IA, S-Recall, respectively. Figures are not to the same scale.

5.3. Results

For evaluating performance of our method and the baselines, we adopt relevance metric nDCG and diversification metrics α-nDCG, ERR-IA and S-Recall, which have widely used in Belém et al. (2013), Clarke et al. (2008), Zhai, Cohen, and Lafferty (2015),

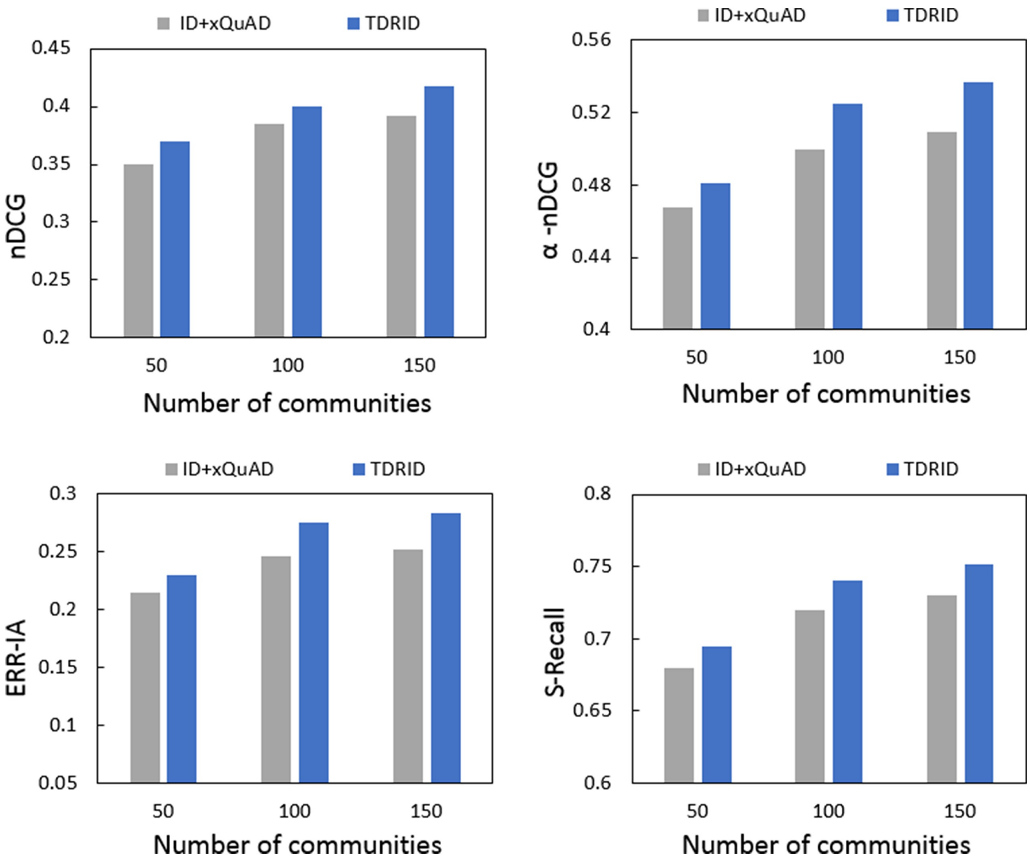

Fig. 5. Comparisons between ID + xQuAD and TDRID on Yelp data set challenge 2014 when varying the number of communities, for nDCG, α-nDCG, ERR-IA, S-Recall, respectively. Figures are not to the same scale.

Agrawal, Gollapudi, Halverson, and Ieong (2009) and Dang and Croft (2012).

In this section, We compare the relevance and diversity performance of TDRID against that of the other methods for three datasets; we compare running time of TDRID against that of the other methods for Yelp data set challenge 2014 and Digg dataset; We then report and analyze the impact of parameters for three datasets.

5.3.1. Performance of TDRID

We evaluate the relevance and diversity performance of our proposed TDRID algorithm by comparing it with the baselines. Table 3 presents the nDCG, α-nDCG, ERR-IA and S-Recall results on three data sets for all methods.

From Table 3, we can observe that for relevance metric nDCG, the performance improvement between methods is presented and the magnitude of improvement is very different. For every data set, the order of performance is TDRID ∼ ID + xQuAD ∼ AIR + MaxSum > xTReD ∼ MF + MMR, where symbol > denotes significantly higher performance and ∼ denotes insignificant improvement. Specifically, for these three datasets, the most significant improvement of relevance performance of TDRID respectively compared with xTEeD and AIR + MaxSum is 23.20% and 14.70% on Yelp data set challenge 2014; the most significant improvement of diversity performance α-nDCG of TDRID respectively compared with xTEeD and AIR + MaxSum is 27.23% and 12.65% on Yelp data set challenge 2014; the most significant improvement of diversity performance ERR-IA of TDRID respectively compared with xTEeD and AIR + MaxSum is 40.30% and 17.60% on Yelp data set challenge 2014; the most significant improvement of diversity performance S-Recall of TDRID respectively compared with xTEeD and AIR + MaxSum is 20.36% and 10.77% on Digg dataset. This illustrates that the integration of latent topics, dynamic and influence diffusion can enhance the performance of relevance and diversity notably. Besides, on the whole, the best experiment result of relevance and diversity generates from Yelp data set challenge 2014 (excluding experiment result w.r.t. S-Recall metric), the largest dataset among these three datasets, that validates the scalability and robustness of our TDRID.

In essence, MF + MMR and xTReD are diversity enhancement algorithms applied to the matrix factorization algorithm, which ignores time factor of user interest and generally computes similarity with the cosine similarity between the item feature vectors. While, AIR + MaxSum is topic-aware, and ID + xQuAD together with TDRID are based on our dynamic latent variable model. This illustrates that the integration of latent topics, dynamic and influence diffusion can enhance the relevance performance.

As is obvious from Table 3, TDRID significantly outperforms other methods on diversification performance with evaluation

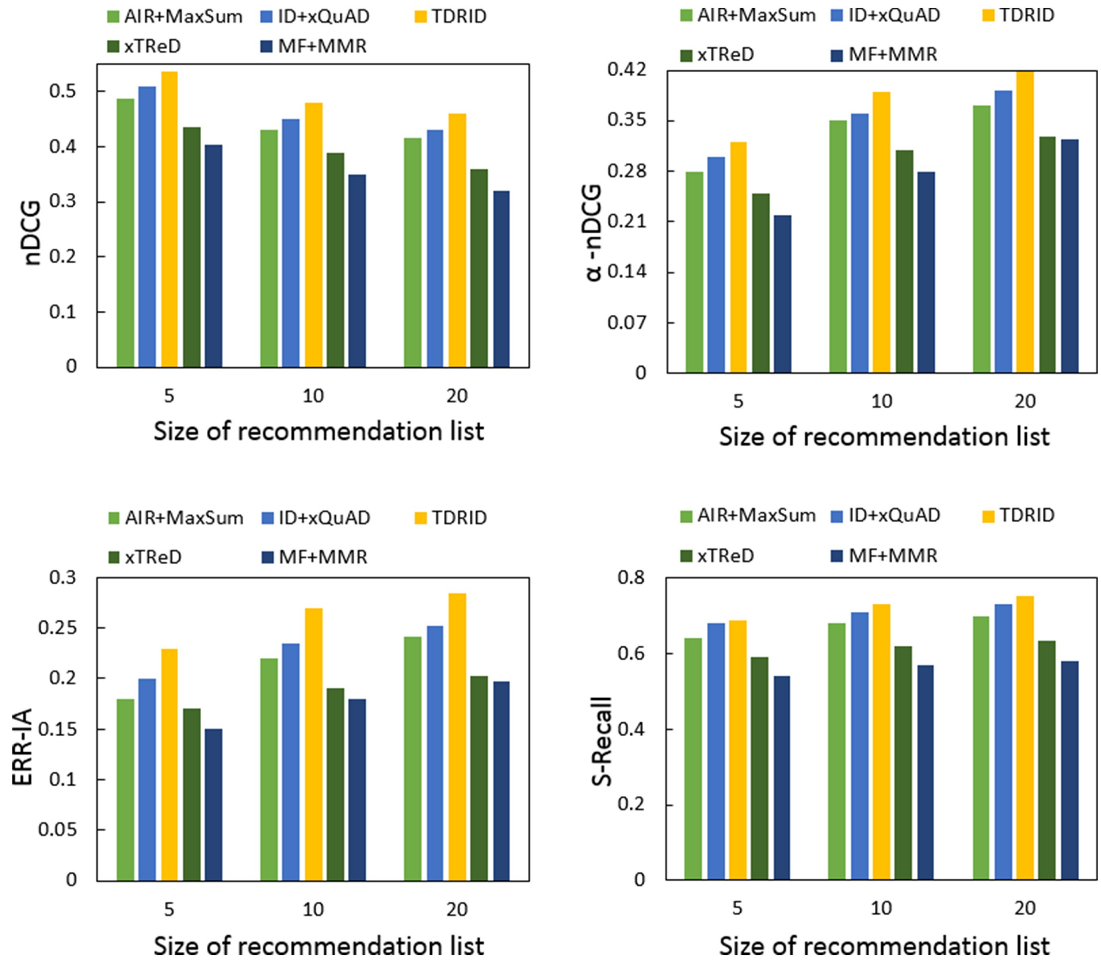

Fig. 6. Comparisons among MF + MMR, xTReD, ID + MaxSum, ID + xQuAD and TDRID on Yelp data set challenge 2014 when varying the size recommendation list, for nDCG, α-nDCG, ERR-IA, S-Recall, respectively. Figures are not to the same scale.

metrics α-nDCG, ERR-IA and S-Recall. MF + MMR is the least effective since it is topic-blind and only provides a linear combination of relevance and diversification. Essentially, xTReD is a diversity enhancement algorithm (nearly similar to xQuAD) applied to a relevance-driven algorithm. While, xQuAD takes into account the topics to penalize redundancy. That is to say, for each item selected, xQuAD downweights each of the topics based on the degree of its relevance to the selected items so that the topics that are less relevant to the selected items will have higher priority in the next round. This exactly explains why xTReD outperforms MF + MMR algorithm in diversification performance. As for xTReD and ID + xQuAD, the difference between them is that the former ignores time information and simply constructs item feature vector with item tags, whereas the latter is based on a dynamic latent topic model, thus better captures user interest. Last, we analyze the diversification performance of AIR + MaxSum, ID + xQuAD and TDRID. Since MaxSum is a diversification algorithm that selects item through a relevance-dissimilarity function, which limits its applicability to topic probability assignment, its diversification performance is lower than ID + xQuAD and TDRID. Our algorithm TDRID evolves from PM-2 which, despite the conceptual difference, can be explained using xQuAD' s framework of reweighting topic still. However, the biggest difference between xQuAD and PM-2 is that the latter uses a more proportionality aware topic weighting function which is based on the Sainte-Lague algorithm. This makes PM-2 more effective than xQuAD.

For Yelp data set challenge 2014 and Digg dataset, we compare running time of our TDRID and the baselines by varying the size of recommendation list from 5 to 20. And the experimental results are respectively shown in Fig. 3(a) and (b). From Fig. 3(a) and (b), we observe that the time efficiency of TDRID outperforms AIR + MaxSum significantly because TDRID is equipped with more efficient topic extraction method than topic diffusion model AIR. We can also see that the curve of TDRID is close to that of ID + xQuAD for the reason that they are with the same latent variable model and the only difference between them is the diversification process. We can see that xTReD and MF + MMR are the best two algorithms w.r.t. running time, because they simply exploit explicit categories (tags) as topics and use an easily tuned ranking function to balance the tradeoff between relevance and diversity.

5.3.2. Impact of parameters

In this part, we analyze the sensitivity of TDRID and baselines to parameters, including *Z, C, k* and length of time interval. Due to

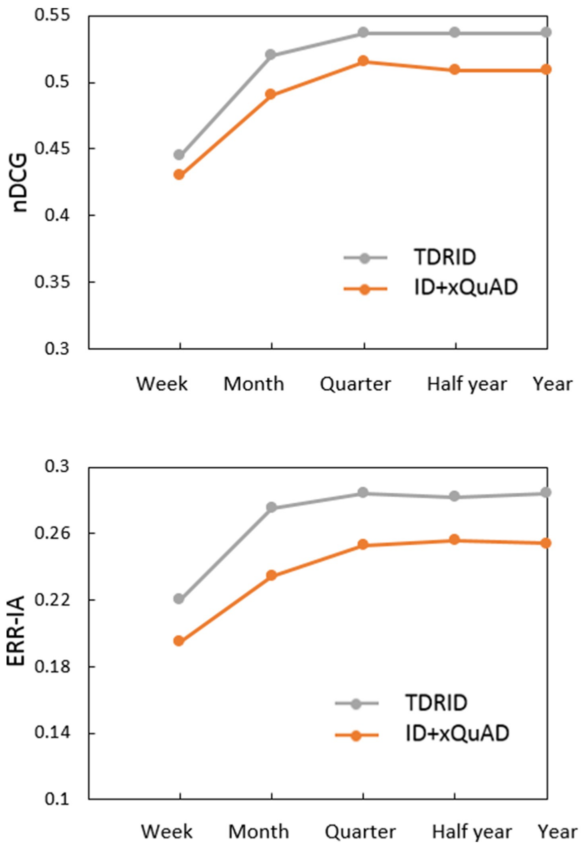

Fig. 7. Comparisons between ID + xQuAD and TDRID on Yelp data set challenge 2014 when varying time interval, for nDCG, α-nDCG, ERR-IA, S-Recall, respectively. Figures are not to the same scale.

space constraints, we show only the impact of parameters on Yelp data set challenge 2014. Similar conclusion is drawn on Yelp data set challenge 2013 and Digg dataset.

First, we compare the sensitiveness of ID + xQuAD and TDRID to number of topics since only these two methods involve tunable latent topics. As illustrated in Fig. 4, when Z is too small, their performance is almost same. When Z varies from 2 to 8, the performance of two methods improves significantly and it is obvious the performance of TDRID outperforms ID + xQuAD. With the number of topics increasing, the performance of two methods reaches a plateau, which exactly testifies the robust and insensitiveness to the number of topics.

Moreover, we analyze the impact of number of communities to these two methods. As demonstrated in Fig. 5, when C increases, their performance improves evidently. This is because with the number of communities increasing, users can be divided into finer-grained communities with more precise influence pattern accordingly, which yields better predicting performance.

Furthermore, the impact of size of recommendation list is analyzed as Fig. 6 shows. It is remarkable that when $k = 5$, the relevance metric nDCG of all methods reaches the highest point but their diversification metrics reach lowest values. It is reasonable when being recommended less items, users usually prefer items which are more relevant, but with the number of recommended items increasing, users' demands for diversity emerge.

We also compare the performance of ID + xQuAD and TDRID in different time interval settings. From Fig. 7, we can observe that when interval increases from week to month, the performance of two methods improves remarkably and it is obvious the performance of TDRID outperforms ID + xQuAD. With the number of interval increasing to quarter, their performance reaches a plateau. The reason is that it is difficult for our dynamic latent variable model to capture user interest in too short interval for the sparsity of data set. With the alleviation of data sparsity problem, the performance of these two methods increases and tends to stable.

6. Conclusions

In this paper, We explore diversified recommendation with influence diffusion. We propose a diversified recommendation method, TDRID, which can return recommendation lists of items that are recent and relevant to the individual topic interests and cover as many topics as possible. We develop a dynamic latent variable model that can capture the changes of topic interest and infer the most recent community-level topic interest, through which we infer the most recent topic interest for each individual user. Moreover, we design a Gibbs sampling algorithm to infer model parameters. We propose a time-sensitive diversification algorithm, which is applicable to perform diversified recommendation over dynamic topics. We evaluate effectiveness and efficiency of our proposals by comprehensive experiments. The experiment results validate our method and show the superiority of our algorithm compared with state-of-the-art diversified recommendation methods. As to future work, it may be interesting to extend item diversified recommendation w.r.t. heterogeneous networks or dynamic evolution of social network.

Acknowledgements

This work is supported by National Key R & D Program of China Project #2017YFB0203201 and Australian Research Council Discovery Project DP150104871. The corresponding author is Hong Shen.

References

Adamopoulos , P., & Tuzhilin , A. (2014). On unexpectedness in recommender systems: Or how to better expect the unexpected. *ACM Transactions on Intelligent Systems and Technology, 5*(54), 1–32.
Adomavicius, G., & Kwon, Y. O. (2012). Improving aggregate recommendation diversity using ranking-based techniques. *IEEE Transactions on Knowledge and Data Engineering, 24*(5), 896–911.
Agrawal, R., Gollapudi, S., Halverson, A., & Ieong, S. (2009). *Diversifying search results. WSDM ACM*5–14.
Barbieri, N., Bonchi, F., & Manco, G. (2012). *Topic-aware social influence propagation models. ICDM IEEE*81–90.
Belém, F., Santos, R., Almeida, J., & Gonçalves, M. (2013). *Topic diversity in tag recommendation. Conference on recommender systems, ACM*141–148.
Bi, B., Tian, Y., Balmin, A., Balmin, A., & Cho, J. (2014). *Scalable topic-specific influence analysis on microblogs. WSDM ACM*513–522.
Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research, 3*(January), 993–1022.
Carbonell, J., & Goldstein, J. (1998). *The use of mmr, diversity-based reranking for reordering documents and producing summaries. SIGIR ACM*335–336.
Cha, Y., & Cho, J. (2012). *Social-network analysis using topic models. SIGIR ACM*565–574.
Chatterjee, R., & Eliashberg, J. (1990). The innovation diffusion process in a heterogeneous population: A micromodeling approach. *Management Science, 36*(9), 1057–1079.
Chen, W., Wei, L., & Zhang, N. (2012). *Time-critical influence maximization in social networks with time-delayed diffusion process. AAAI*592–598.
Clarke, C. L. A., Kolla, M., Cormack, G. V., Vechtomova, O., Ashkan, A., & Mackinnon, I. (2008). *Novelty and diversity in information retrieval evaluation. SIGIR ACM*659–666.
Dang, V., & Croft, W. B. (2012). *Diversity by proportionality: an election-based approach to search result diversification. SIGIR ACM*65–74.
Fang, Y., Cheng, R., Luo, S., & Hu, J. (2016). Effective community search for large attributed graphs. *Proceedings of the VLDB Endowment, 9*(12), 1233–1244.
Gan, M., & Jiang, R. (2013). Constructing a user similarity network to remove adverse influence of popular objects for personalized recommendation. *Expert Systems with Applications, 40*(10), 4044–4053.
Gollapudi, S., & Sharma, A. (2009). An axiomatic approach for result diversification. In www.acm (pp. 381–390).
Guo, W., Wu, S., Wang, L., & Tan, T. (2015). *Social-relational topic model for social networks. CIKM ACM*1731–1734.
Heinrich, G. (2005). *Parameter Estimation for Text Analysis*.
Hu, Z., Yao, J., Cui, B., & Xing, E. (2015). *Community level diffusion extraction. SIGMOD ACM*1555–1569.
Iwata, T., Yamada, T., Sakurai, Y., & Ueda, N. (2010). *Online multiscale dynamic topic models. SIGKDD ACM*663–672.
Kalish, S. (1985). A new product adoption model with price, advertising, and uncertainty. *Management Science, 31*(12), 1569–1585.
Lee, K., & Lee, K. (2015). Escaping your comfort zone: A graph-based recommender system for finding novel recommendations among relevant items. *Expert Systems with Applications, 42*(10), 4851–4858.
Li, J., Hu, X., Jian, L., & Liu, H. (2017). *Toward time-evolving feature selection on dynamic networks. ICDM IEEE*1003–1008.
Liang, S., Ren, Z., Yilmaz, E., & Kanoulas, E. (2017). *Collaborative user clustering for short text streams. AAAI*3504–3510.
Liang, S., Yilmaz, E., & Kanoulas, E. (2016). *Dynamic clustering of streaming short documents. SIGKDD ACM*995–1004.
Liu, L., Tang, J., Han, J., Jiang, M., & Yang, S. (2010). *Mining topic-level influecne in heterogeneous networks. CIKM ACM*199–208.
Qin, L., & Zhu, X. (2013). *Promoting diversity in recommendation by entropy regularizer. IJCAI AAAI*2698–2704.
Ren, Z., Liang, S., Li, P., Wang, S., & Rijke, M. D. (2017). *Social collaborative viewpoint regression with explainable recommendations. WSDM ACM*485–494.
Santos, R. L. T., Macdonald, C., & Ounis, I. (2010). Exploiting query reformulations for web search result diversification. In www.acm (pp. 881–890).
Su, J., & Havens, T. C. (2015). Quadratic program-based modularity maximization for fuzzy community detection in social networks. *IEEE Transactions on Fuzzy Systems, 23*(5), 1356–1371.
Wu, L., Liu, Q., Chen, E., Yuan, N. J., Guo, G., & Xie, X. (2016). Relevance meets coverage:a unified framework to generate diversified recommendations. *ACM Transactions on Intelligent Systems and Technology, 7*(3), 1–30.
Yan, X., Guo, J., Lan, Y., & Cheng, X. (2013). A biterm topic model for short texts. In www.acm (pp. 1445–1456).
Yang, J., McAuley, J., & Leskovec, J. (2013). *Community detection in networks with node attributes. ICDM IEEE*1151–1156.
Yang, Z., Kotov, A., Mohan, A., & Lu, S. (2015). *Parametric and non-parametric user-aware sentiment topic models. SIGIR ACM*413–422.
Yin, H., Cui, B., Sun, Y., Hu, Z., & Chen, L. (2014). Lcars: A spatial item recommender system. *ACM Transactions on Information Systems, 32*(3), 11.
Yin, J., & Wang, J. (2014). *A dirichlet multinomial mixture model-based approach for short text clustering. SIGKDD ACM*233–242.
Zhai, C., Cohen, W. W., & Lafferty, J. (2015). Beyond independent relevance: Methods and evaluation metrics for subtopic retrieval. *ACM SIGIR Forum, 49*(1), 2–9.
Zhang, Y., Lyu, T., & Zhang, Y. (2017). *Hierarchical community-level information diffusion modeling in social networks. SIGIR ACM*753–762.
Zhao, Y., Liang, S., Ren, Z., Ma, J., Yilmaz, E., & Rijke, M. D. (2016). *Explainable user clustering in short text streams. SIGIR ACM*155–164.
Ziegler, C. N., Mcnee, S. M., Konstan, J. A., & Lausen, G. (2005). Improving recommendation lists through topic diversification. In www.acm (pp. 22–32).