# Discovering Temporal Retweeting Patterns for Social Media Marketing Campaigns

Guannan Liu*, Yanjie Fu†, Tong Xu‡, Hui Xiong† and Guoqing Chen*
*School of Economics and Management, Tsinghua University, Beijing, China
{liugn.10, chengq}@sem.tsinghua.edu.cn
†Rutgers Business School, Rutgers University, USA
{yanjie.fu, hxiong}@rutgers.edu
‡School of Computer Science and Technology, University of Science and Technology of China
tongxu@mail.ustc.edu.cn

***Abstract*—Social media has become one of the most popular marketing channels for many companies, which aims at maximizing their influence by various marketing campaigns conducted from their official accounts on social networks. However, most of these marketing accounts merely focus on the contents of their tweets. Less effort has been made on understanding tweeting time, which is a major contributing factor in terms of attracting customers' attention and maximizing the influence of a social marketing campaign. To that end, in this paper, we provide a focused study of temporal retweeting patterns and their influence on social media marketing campaigns. Specifically, we investigate the users' retweeting patterns by modeling their retweeting behaviors as a generative process, which considers temporal, social, and topical factors. Moreover, we validate the predictive power of the model on the dataset collected from *Sina Weibo*, the most popular microblog platform in China. By discovering the temporal retweeting patterns, we analyze the temporal popular topics and recommend tweets to users in a time-aware manner. Finally, experimental results show that the proposed algorithm outperforms other baseline methods. This model is applicable for companies to conduct their marketing campaigns at the right time on social media.**

## I. Introduction

One basic task of marketing is to get the brands' voices heard among the potential customers. This also works for marketing in social media, one of the most trending marketing channels, and it has become an impactful way to connect the brands with their customers. Nowadays more and more companies have set up official accounts on social media such as Twitter and Facebook, in order to promote new products, send out coupons and sharing some interesting comments to build up their brand image among their followers. The sharing and retweeting mechanisms in social media enable information to propagate extremely fast from one user to another, enhancing the word-of-mouth effect [1]. As a matter of fact, every time the official account gets retweeted, it does not only influence the users who actually retweet, but also the followers who might see the tweet along the propagation trace. The number of retweets generally reflect the marketing influence they exert in social media. Therefore, the official accounts regularly post tweets, aiming at getting more retweets among the large amount of users.

In order to maximize the influence of their posted tweets from the official accounts, companies should understand how their tweets can be more attractive to the audience. The *viral feature* and *viral characteristics* of a product can decide how far it can propagate in social networks [2], and many efforts to optimize the tweets have been made along the line. Zarrela [3] has found that the tweets with URLs are more likely to get retweeted, he also analyzed the words and phrases that are highly retweeted and the influence of number of characters on retweeting. Suh et al. [4] proved the *viral features* such as URLs and hashtags can help gain more retweets in an analytic way. However, these efforts optimize the tweets mainly from content level.

In real marketing practice, consumers' purchase time on particular categories of products such as clothes usually follows some nature and patterns, thus, merchants usually capture these patterns to launch some seasonal promotions at appropriate time. This is similar to propagate tweets in social media platforms, because users' behaviors in using social media may also follow some innate nature and temporal patterns. For example, some users only sign in their Twitter at their leisure time for entertainment; while others may browse their micrblogs to seek for some news and information during working hours.

Some statistical analysis regarding the tweeting time has been reported in [7], [8], which is obtained by analyzing the retweets from a global view. However, the statistics do not apply for all situations. This is due to the fact that users behave distinctively in using social media, and meanwhile different topics can be prevailing at different time periods. Therefore, a more precise model is needed to capture each users' retweeting behaviors in regard with the posting time of tweets.

In this paper, we aim to help the official accounts in social media to discover the temporal retweeting patterns of the users. Since the users' preferences are all manifested in what they have shared and posted, we start by analyzing the retweeting history of each user. We formalize users' retweeting behaviors as (**W**hen, **W**ho, **W**hat), i.e., **3W** retweeting patterns, in order to capture the collaborative effects of the posting accounts, posting time, and the contents of the tweets in influencing users' retweeting behaviors. We propose a generative process to model each users' retweeting history, in which the *shared retweeting context* is introduced to capture the mutual influence of the factors, meanwhile the words of the retweeted tweets are generated in a typical topic modeling process.

By discovering the retweeting patterns of users, we can

1550-4786/14 $31.00 © 2014 IEEE
DOI 10.1109/ICDM.2014.48

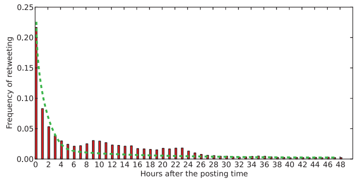
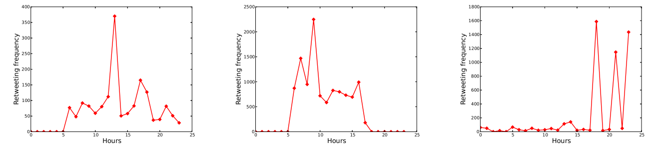


Fig. 1. Retweeting frequency after the posting time.



(a) McDonald's (b) Starbucks (c) iTunesMovies

Fig. 2. Average retweets at different time.

predict the popular topics with regard to different time periods for each account. Moreover, we can further predict users' retweeting decisions in different time and recommend the tweets with large probability to be retweeted to the users.

The remainder of this paper is organized as follows. In Section II, we analyze the retweets in regard with time and propose assumptions for modeling. We introduce the model and parameter estimation in Section III. Section IV presents the experimental results on a real-world dataset. We discuss the related work in Section V and conclude the paper in Section VI.

## II. Time-aware Retweeting Analysis

The official accounts of some big name companies are quite active on social media. They post tweets in hope of getting exposed to the large audience in social media. To analyze the retweets of the official accounts, we crawled the recent 3,000 tweets and their retweets from some selected official Twitter accounts, such as McDonald's, Starbucks, itunesMovies, etc.

We firstly observe the time gaps between the posting time and the retweeting time. We round up the continuous time gaps to hours and calculate the frequency for the retweets that happen in each hour after the tweets are posted. Figure 1 displays the retweeting frequency in a range of 2 days (48 hours). It is obvious that most retweets happen immediately after the tweet is posted, and the retweets become extremely rare after one day (24 hours). The overall distribution of retweeting frequency follows long tail distribution as the dashed line reveals. Thus, we have the following assumption.

***Assumption 1:*** **The posting time of the tweets is significant in influencing the number of retweets.**

We also observe the average number of retweets at different tweeting time for the official Twitter accounts, where the tweeting time is divided into 24 time slots. As we see from Figure 2, the peak in the McDonald's curve occurs at around 12:00-13:00 p.m., which corresponds to the lunch hours. While the tweets of Starbucks are mostly retweeted in the morning (7:00-9:00 a.m.) and afternoon (4:00 p.m.). The tweets posted by iTunesMovies, which are mostly introductions and trailers of some movies, are greatly retweeted in the evening. We see from the curves in Figure 2 that the number of retweets received by the accounts fluctuate with time, and the fluctuation patterns vary over different accounts. This may be due to the fact that each account usually posts tweets on particular topics, and the topics may be prevalent in particular time periods. Thus, we have another assumption as follows,

***Assumption 2:*** **Each account has its own optimal time to post tweets on specific topics in order to get most retweets.**

## III. User Retweet Model

### A. General Ideas

From the above analysis, we assume that the posting time of tweets can influence the numbers of retweets, since users behave periodically in retweeting different topics posted by different accounts. In order to understand how temporal factors can influence the propagation of marketing campaigns in microblogs, we start by studying the retweeting behaviors of each user. We assume that the following factors can influence users' retweeting decisions.

a) **W**hen to sign in? Users have their own habits in using microblogs, and the time for the user to sign in can directly decide what the users would see and like at that moment, and finally influence their retweeting decisions. Thus, we define this factor as *Temporal Context* to capture users' preferences towards time.

b) **W**hose tweets? Social relations in microblogs can also decide how the user react to the tweet [9]. Users generally show different attitudes towards the authors in terms of their trustworthiness and authoritativeness. We define the authors of tweets as *Social Context* to represent users' preferences towards social relationships.

c) **W**hat topics? As pointed out in [10] that users' retweeting can happen when the topics of tweets fit in users' topical preferences. Thus, the topics of the tweet can significantly influence users' decisions in retweeting.

As a matter of fact, the above factors can jointly influence users' retweeting decisions. Users have preferences towards *Temporal Context* in using social media, which decides the usage scenarios. Under the specific scenarios, users show their interests in particular authors. While the authors are usually expert in several topics, users are influenced by the accounts in topic level [11]. In sum, the *Temporal Context* and *Social Context*, along with the topics of the tweets, collaboratively influence users' retweeting decisions. Hence, we formally define the collaborative retweeting patterns as (**W**hen, **W**ho, **W**hat), i.e., **3W** patterns. Therefore, we aim at discovering the retweeting patterns from users' retweeting history in this study, and the problem can be defined as follows.

**Problem Definition:** Suppose that we have users' retweeting history $\mathcal{T}$ in microblogs, each trace is represented with a tuple $(u, t_r, f_d, d, t_p)$, where $u$ retweets $d$ at time $t_r$, which is posted by $f_d$ at time $t_p$. We aim to model all the users' retweeting history to discover the **3W** retweeting patterns. By capturing the retweeting behavioral patterns of the users, we can predict the temporal popular topics and design appropriate viral marketing campaigns and push them to the potential customers at right time.

### B. Modeling Shared Retweeting Context

Each retweeting trace is composed of particular retweeting time, author and content, which are correlated with each other. In order to discover the common retweeting patterns that underlies these specific **3W** tuples, we introduce a latent variable, i.e., *shared retweeting context*, $c$. Each category of retweeting context can be viewed as a mixture of time, author and topics, representing some common retweeting behavioral patterns among all the users, hence the specific retweeting tuples can be categorized into one *shared retweeting context*. This is similar to the *common context* discussed in [12], where the particular context logs of mobile device is generated from a latent common context.

The *shared retweeting context* bridges all the factors that can influence users' retweeting decisions. In modeling $u$'s retweets, the *shared retweeting context* $c$ is firstly sampled from the multinomial distribution of user' preferences towards the retweeting contexts, i.e., $c \sim p(c|u)$. Therefore, users' preferences towards time, author, and topics (**3W**) are simultaneously captured by the *shared retweeting context*, and each particular **3W** factor is generated from the *shared retweeting context*.

To simplify the analysis, we assume that these specific retweeting tuples, including time, author, and topics, are conditional independent given the *shared retweeting context*. Specifically, the generative process for the each user $u$'s retweeted tweet $d \in \mathcal{R}(u)$ is as follows:

(1) Generate a shared retweeting context $c \sim p(c|u)$;
(2) Generate a retweeted author $f \sim p(f|c)$;
(3) Generate the retweeting timeslot $t \sim p(t|c)$;
(4) For each word $w$ in $\mathbf{w}_d$:
  a) Generate a topic $z \sim p(z|c)$;
  b) Generate a word $w \sim p(w|z)$.

### C. Parameter Estimation

Given all the retweeting traces in the collection, the log likelihood is as follows,

$$\begin{aligned} L(\Psi; D) &= \log p(D|\Psi) \\ &= \sum_{u} \sum_{d \in \mathcal{R}(u)} \log \sum_{c} p(u)p(c|u)p(t|c)p(f|c)p(\mathbf{w}_d|c) \end{aligned} \quad (1)$$

where $p(\mathbf{w}_d|c)$ can be calculated as follows,

$$p(\mathbf{w}_d|c) = \prod_{w \in \mathbf{w}_d} p(w|c)^{n(w,d)} \quad (2)$$

$$p(w|c) = \sum_{z} p(w|z)p(z|c) \quad (3)$$

The parameters in the model are $p(c|u)$, $p(t|c), p(f|c), p(z|c), p(w|z)$, which are denoted by $\Psi$. In order to estimate these parameters, we use Expectation Maximization (EM) algorithm.

In the **E-step**, we have two latent variables in the model, thus we need to calculate the joint expecation $p(z, c|u, t, f, \mathbf{w}_d)$. However, as we have assumed $z$ is conditional independent with $t$ and $f$ given the shared retweeting context $c$, the joint expectation of the two latent variables $z$ and $c$ can be written as,

$$\begin{aligned} p(z, c|u, t, f, \mathbf{w}_d) &= p(z|u, t, f, \mathbf{w}_d, c)p(c|u, t, f, \mathbf{w}_d) \\ &= p(z|\mathbf{w}_d, c)p(c|u, t, f, \mathbf{w}_d) \end{aligned} \quad (4)$$

Therefore we can update $p(c|u, t, f_d, \mathbf{w}_d)$ and $p(z|\mathbf{w}_d, c)$ separately, according to Bayes rule as follows,

$$p(c|u, t, f_d, \mathbf{w}_d) = \frac{p(c|u)p(f_d|c)p(t|c)p(\mathbf{w}_d|c)}{\sum_{c'} p(c'|u)p(f_d|c')p(t|c')p(\mathbf{w}_d|c')} \quad (5)$$

$$p(z|w, c) = \frac{p(z|c)p(w|z)}{\sum_{z'} p(z'|c)p(w|z')} \quad (6)$$

In the **M-step**, we find $\Psi$ that maximize the log-likelihood as follows,

$$p(c|u) = \frac{\sum_{d \in R(u)} p(c|u, t, f_d, \mathbf{w}_d)}{\sum_{d \in R(u)} \sum_{c'} p(c'|u, t, f_d, \mathbf{w}_d)} \quad (7)$$

$$p(f|c) = \frac{\sum_{d \in R(u)} p(c|u, t, f_d, \mathbf{w}_d)}{\sum_{d \in R(u)} \sum_{f'} p(c|u, t, f', \mathbf{w}_d)} \quad (8)$$

$$p(t|c) = \frac{\sum_{d \in R(u)} p(c|u, t, f_d, \mathbf{w}_d)}{\sum_{d \in R(u)} \sum_{t'} p(c|u, t', f_d, \mathbf{w}_d)} \quad (9)$$

$$p(z|c) = \frac{\sum_{d \in R(u)} \sum_{w} n(w, d)p(c|u, t, f_d)p(z|w, c)}{\sum_{z'} \sum_{d \in R(u)} \sum_{w} n(w, d)p(c|u, t, f_d)p(z|w, c)} \quad (10)$$

$$p(w|z) = \frac{\sum_{u} \sum_{d \in R(u)} n(w, d)p(c|u, t, f_d)p(z|w, c)}{\sum_{w'} \sum_{u} \sum_{d \in R(u)} n(w', d)p(c|u, t, f_d)p(z|w', c)} \quad (11)$$

By iterating between E-step and M-step until the log likelihood get converged, we can obtain a good estimation of the parameters.

## IV. Experiments

### A. Experiment Setup

The dataset used in the experiments is from a public data set crawled from *Sina Weibo* [13], one of the most popular microblogs in China. It originally contains 300,000 tweets and 23,755,810 retweets. We randomly sampled a collection of popular tweets with large retweeting numbers from the dataset, and we also obtain the users frequently retweet the tweets from the sampled tweets (i.e., the times of retweeting surpasses 5). After preprocessing, we get a dataset with 10,046 users, 21,272 tweets, and the number of retweets is 89,578.

We split the dataset according to a pre-defined cutoff point of the tweets' tweeting time, and then fit the model with the training set and test the model on the test set [1].

### B. Temporal Popular Topics

With the inferred parameters from the model, we can analyze popular topics for different time slots. Specifically, we calculate the topic distribution for each time slot, i.e., $p(z|t)$. By Bayes rule, $p(z|t) \propto p(z, t) = \sum_c p(z, t, c)$, and we also assume that $z$ and $t$ are conditional independent given the latent context $c$. Thus, $p(z|t)$ can be calculated as follows,

$$p(z|t) = \frac{\sum_c \sum_u p(z|c)p(t|c)p(c|u)p(u)}{\sum_{z'} \sum_c \sum_u p(z'|c)p(t|c)p(c|u)p(u)} \quad (12)$$

In this experiment, we discretize the continuous retweeting time into divided timeslots. We assume that users' retweeting behaviors shift in the granularity of hours and also differ in

[1]The sampled dataset and code is available at http://bit.ly/1rwss9h

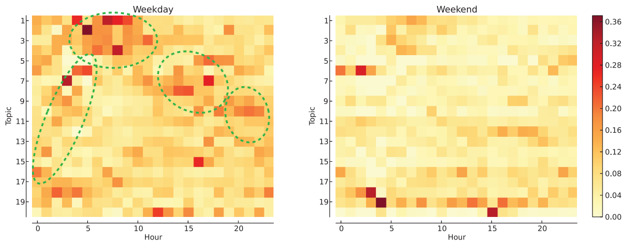



Fig. 3. Temporal topic distribution.

days of week, thus we divide each day into 24 timeslots separately for weekdays and weekends, which makes 48 timeslots in total. We set $c = 50$ and $z = 20$ to learn the model and estimate the topic distribution for different timeslots. Figure 3 presents the topic distribution for each timeslot, with the darkness of colors representing the values of $p(z|t)$, and we also list the representative words for each discovered topics in Table I.

We can see from Figure 3 that the topic distribution shows an obvious temporal pattern on weekdays. As the green circles highlight in Figure 3, popular topics shift over the time axis, and adjacent timeslots have similar prevalent topics. For example, in the early morning 5:00-8:00, topic 1 and 2 are with large probability values. As can be seen in Table I, topic 1 is represented with words concerning life wisdom and religious idioms, while words in topic 2 are mostly about social news and events. We also notice that topics on technology (topic 3) and business (topic 4) are popular in working hours during weekday morning, when most users concentrate on working issues. Moreover, it is easy to observe that the popular topics shift during noon, afternoon, and evening time.

We can see that the temporal patterns of topics on weekends are less obvious than that at weekdays. Users generally have no routine schedules during weekends, and the usage scenarios of social media and retweeting contexts can be much more random, which gives rise to the lack of significant temporal patterns in topics.

From the above analysis, we can discover the temporal patterns for popular retweeted topics, and we see that the proposed User Retweet Model capture the temporal dynamics of retweeted topics, especially in routine weekdays.

### C. Time Aware Tweets Recommendation

Considering the temporal effect in influencing users' retweeting behaviors, we can design a time-aware tweets recommendation method based on the estimated parameters from the proposed model. Specifically, given a timeslot $t$ and a user $u$, candidate tweets can be ranked according to the probability that $u$ would retweet a tweet $d$ at time $t$, where $d$ is associated with words $w_d$ and also the author $f_d$. Thus, the task is to estimate $p(\mathbf{w}_d, f_d|u,t)$, which can be calculated as follows,

$$p(\mathbf{w}_d, f_d|u,t) = \frac{\sum_c p(u)p(c|u)p(t|c)p(f_d|c)p(\mathbf{w}_d|c)}{\sum_{d'}\sum_c p(u)p(c|u)p(t|c)p(f_{d'}|c)p(\mathbf{w}_{d'}|c)} \tag{13}$$

In real world situation, given an active time slot $t$ for a user $u$ in using microblog, $u$ would choose from all the tweets $d \in \mathcal{S}_t$ that have been seen before finally retweets, where $\mathcal{S}_t$ denotes the tweets that are posted by $u$'s followees and viewed by $u$ at the timeslot. However, we cannot know exactly what the users have seen before they retweet, i.e., we cannot obtain the tweets collection $\mathcal{S}_t$ directly from the data. Jiang et al. [9] introduced *online session* to simulate the time when the users are online and what they have seen in this session.

*1) Online session:* To simulate the online sessions for users, we employ the real retweeting time $t_r$ as a cut off point, since user $u$ is certain to be active during the retweeting time. Therefore, it is reasonable to assume that users may see their followees' tweets which are near the real retweeting time (e.g., within half an hour), and these tweets compose a pseudo online session for the user. We only retain the sessions with the number of tweets exceeding 15.

Thus, we recommend tweets for users based on the pseudo online sessions. For each user $u$ in the test set, we generate the online sessions for each retweet in time sequence, and we calculate $p(\mathbf{w}_d, f_d|u,t)$ with the parameters inferred from the training set.

*2) Baseline Methods:* We compare the proposed time aware recommendation method (*URM*) with the following baseline methods.

*Retweeted Times*: Retweeted times reflect the popularity of tweets. Therefore, we extract the retweeted times of each tweet before the time of users' online sessions, and rank the candidate tweets in this session in descending order of retweeted times.

*Content-based LDA*: According to content-based recommendation [17], items that are similar to users' retweeting history are recommended. We firstly model the retweeted tweets with LDA [18], with each tweet represented with a topic distribution. We then calculate the relevance score between the users' retweeting history and the candidate tweet $d$ as follows,

$$D(u,d) = \sum_{d' \in \mathcal{R}(u)} D_{KL}(d'||d) \tag{14}$$

Here $D_{KL}$ calculates the KL-divergence between the topic distributions of the previously retweeted $d'$ by $u$ and the candidate tweet $d$.

*Temporal-based LDA*: As an extension of *Content-based LDA*, this method incorporates the temporal information in calculating relevance score between the user and the tweet. Given the user $u$'s online session at time $t$, the relevance score is calculated as follows,

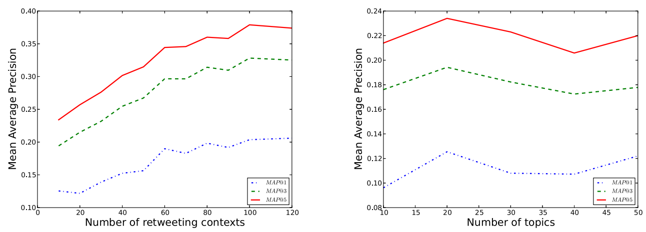
$$D(u^t,d) = \frac{\sum_{d' \in \mathcal{R}(u)^t} D_{KL}(d'||d)}{|\mathcal{R}(u)^t|} \tag{15}$$

where $\mathcal{R}(u)^t$ denotes the collection of tweets that have been retweeted by $u$ at time slot $t$ in history.

*RankSVM*: RankSVM can be used to rank the candidate tweets with various features integrated. We adopt the features extracted from tweet content, authors, and relationships [15]. We also extract other viral features including the number of URLs, the number of hashtags, the number of '@' symbols, etc. To facilitate the learning of RankSVM model, we build online sessions for the training set, with the true retweeting tweets labeled as positive cases while others viewed in the session but not retweeted as negative.

TABLE I. EXAMPLE TEMPORAL TOPICS FOR WEEKDAYS

| Morning Topics | | | | Noon/Afternoon Topics | | | Evening Topics | | | Night Topics | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 | Topic 7 | Topic 8 | Topic 10 | Topic 11 | Topic 12 | Topic 15 | Topic 16 |
| life | Nobel | mobile phone | Internet | Jingdong | campaign | travel | Eason | music | trailer | computer | constellation |
| Buddha | writer | Meizu | data | price | points | trip | classical | live | Depp | password | scorpio |
| trouble | literature | android | economy | Suning | credit card | food | rolling | stage | master | night | leo |
| all beings | news flash | os | bank | shopping site | pattern | domestic | songs | super junior | director | wisdom | libra |
| Bodhisattvas | tourists | processor | nation | customers | brand | Xiamen | adel | mv | artist | knowledge | sagittarius |
| merits | users | smart | mobile | offline | exchange | Tibet | emotion | album | poster | sleep | fortunes |
| intelligence | committee | price | platform | product | business | guides | voice | solo | movie | tricks | champion |
| children | report | dual-processor | marketing | e-business | merchandise | global | piano | TVXQ | animation | after meal | ranklist |
| beneficent | road | ipad | start-up | advantage | credits | restaurant | concert | EXO | Noah's Ark | breakfast | luck at love |




(a) Number of retweeting contexts (b) Number of Topics

Fig. 4. Mean Average Precision of different parameters settings.

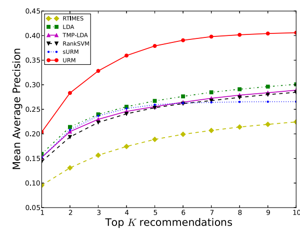



Fig. 5. $MAP@K$ for Top-$K$ recommendations.

*sURM*: We simplify the proposed model by removing the *Temporal Context*, i.e., only the author and the topics are modeled. The simplified model is similar to the model proposed in [16], which only exploits the social influence between users to do recommendation. This model can be used to validate the influence of temporal factors in influencing users' retweeting behaviors.

*3) Evaluation Metrics:* We use Mean Average Precision $MAP@K$ to evaluate the effectiveness of the time aware tweets recommendation method. $MAP@K$ represents the mean value for $AP@K$ over all test sessions, i.e., $MAP@K = \frac{\sum_{s\in\mathcal{S}} AP_s@K}{|\mathcal{S}|}$, where $\mathcal{S}$ is the collection of test sessions. $AP_s@K$ denotes the average precision at top $K$ recommendation for the online session $s$.

$$AP_s@K = \frac{\sum_{r=1}^{K} P(r) \times ret(r)}{|\mathcal{S}(u)|} \tag{16}$$

Here $P(r)$ denotes the precision at cut-off $r$ in the rank list, $ret(r)$ is an indicator function if $u$ actually retweet the tweet at rank $r$, and $|\mathcal{S}(u)|$ is $u$'s retweet counts in the session.

*4) Parameter Settings:* Since we introduce two latent variables in the model, we investigate the recommendation performances with different parameter settings in this section.

**Number of Shared Retweeting Contexts**. We train the model with different number of retweeting contexts, and evaluate the recommendation performance with respect to these numbers. We fix the number of topics to be $z = 10$, and train the model with the number of $c$ ranging from 10 to 120. Figure 4(a) demonstrates the $MAP$ at top 1, 3 and 5 recommendations with different number of retweeting contexts. We see from Figure 4(a) that the $MAP$ of recommendation increases as the number of retweeting contexts increases, and the performance becomes stable after the number reaches 80, thus we choose $c = 80$ in the experiments.

**Number of Topics** The number of latent topics can also influence the model performance, thus we fixed the number of retweeting contexts $c = 20$, and experiment with $z$ ranging from 10 to 50. The results are shown in Figure 4(b). It shows obviously that when the number of topics reaches 20, the performance is best, and the performance does not improve much when $z$ continues to increase. Therefore, we choose $z = 20$ in the experiments.

*5) Recommendation Results:* The Mean Average Precision at Top-$K$ recommendation for the proposed method and the baseline methods is presented in Figure 5. We can see the baseline methods all perform similarly, while the proposed method outperforms the other baseline methods, with the MAP at top 5 recommendation ($MAP@5$) increased by $41.86\%$. It is notable that the model significantly outperforms the model without *temporal context* (*sURM*), which proves the effectiveness of *temporal context* in recommending tweets for users and it also verifies the assumption that users' retweeting behaviors are time dependent.

## V. RELATED WORK

In this section we discussed the work that is related to this study.

*Viral Marketing*. Viral marketing has become a focal point of study recently, especially in the context of social media. Research in marketing fields has explored the effects of viral characteristics in influencing diffusion. Aral et al. [2] found that incorporating viral features in product design can generate large influence and social contagions in new product diffusion. Szabo et al. [19] proposed a linear model to predict the popularity of online contents, however they did not consider the semantics of popularity. There also exist studies on designing viral features to maximize its influence. Barbieri et al. [6] proposed a probabilistic model to capture the viral features' influence on item adoption. Dabeer et al. [8] considered timing effect in increasing the effectiveness of information diffusion in microblogs. Peng et al. [10] modeled retweeting with conditional random fields to predict retweet patterns.

*Topic Modeling*. Recent work on topic models has incorporated richer information to better fit different scenarios. For example, Blei et al. [20] proposed Dynamic Topic Models (DTM) to analyze the evolution of topics. Kawamae [21] introduced trend class to capture the trending topics and proposed Trend Analysis Model (TAM). Yin et al. [22] proposed Latent Periodic Topic Analysis (LPTA) model which exploits the periodicity of terms. Yuan et al. [23] proposed a unified generative model $W^4$ (Who, Where, When and What) that captured user, location, time, and activity jointly. Some other work also exploited social influence in modeling documents generated in social network. Liu et al. [11] assumed that tweets of users are generated from both social influence and users' topic preferences.

*Tweets Recommendation*. Recently much work has been done to improve tweets recommendation accuracy by exploiting the contextual information such as spatio-temporal, social context etc. Chen et al. [14] assumed that users' future rewteeting behaviors are similar to their past retweeting statuses, thus they applied collaborative ranking with tweets' and authors' features. Ye et al. [16] proposed a generative model to capture social influence in item adoption, and used it to recommend items for users and groups. Tang et al. [24] assumed that users can be influenced by both their local friends and users with high global reputations, thus they exploited both local and global social contexts for recommendation.

## VI. Conclusions

In this paper, we studied the effect of tweeting timing in social media marketing. Specifically, by analyzing the retweets across different accounts, we first revealed that the time, authors, and content of tweets could collaboratively influence users retweeting behaviors. Then, we formalized the problem as a task of mining the **3W** (When, Who, and What) patterns from the retweeting records. Along the line, we developed a generative model to identify the retweeting patterns. In this model, the *shared retweeting context* is introduced to capture the correlations among various influencing factors. With the identified retweeting patterns, we analyzed the temporal popular topics. Moreover, we designed a time-aware recommendation method, which is able to recommend tweets during users' online sessions.

It is worth noting that the proposed User Retweet Model can be easily extended to model more influencing factors. In the future work, we can incorporate more retweeting contexts such as users' geographical information, devices, and some other viral features to better capture the retweeting contexts.

## Acknowledgment

The work was partly supported by the MOE Project of Key Research Institute of Humanities and Social Sciences at Universities(12JJD630001), the National Natural Science Foundation of China (71110107027, 71028002), and Tsinghua University Initiative Scientific Research Program (20101081741).

## References

[1] M. Trusov, R. E. Bucklin, and K. Pauwels, "Effects of word-of-mouth versus traditional marketing: findings from an internet social networking site," *Journal of marketing*, vol. 73, no. 5, pp. 90–102, 2009.

[2] S. Aral and D. Walker, "Creating social contagion through viral product design: A randomized trial of peer influence in networks," *Management Science*, vol. 57, no. 9, pp. 1623–1639, 2011.

[3] D. Zarrella, "All about retweets," http://danzarrella.com/all-about-retweets/.

[4] B. Suh, L. Hong, P. Pirolli, and E. H. Chi, "Want to be retweeted? large scale analytics on factors impacting retweet in twitter network," in *Social computing (socialcom), 2010 IEEE second international conference on*. IEEE, 2010, pp. 177–184.

[5] M. Miah, G. Das, V. Hristidis, and H. Mannila, "Determining attributes to maximize visibility of objects," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 21, no. 7, pp. 959–973, 2009.

[6] N. Barbieri and F. Bonchi, "Influence maximization with viral product design," in *SDM*, 2014.

[7] D. Adams, "The best time of day to tweet (to get the most rts)," http://www.bitrebels.com/social/the-best-time-of-day-to-tweet-to-get-the-most-rts/, Jun. 2012.

[8] O. Dabeer, P. Mehendale, A. Karnik, and A. Saroop, "Timing tweets to increase effectiveness of information campaigns." in *ICWSM*, 2011.

[9] M. Jiang, P. Cui, R. Liu, Q. Yang, F. Wang, W. Zhu, and S. Yang, "Social contextual recommendation," in *CIKM*. ACM, 2012, pp. 45–54.

[10] H.-K. Peng, J. Zhu, D. Piao, R. Yan, and Y. Zhang, "Retweet modeling using conditional random fields," in *ICDMW*. IEEE, 2011, pp. 336–343.

[11] L. Liu, J. Tang, J. Han, M. Jiang, and S. Yang, "Mining topic-level influence in heterogeneous networks," in *CIKM*. ACM, 2010, pp. 199–208.

[12] H. Zhu, E. Chen, K. Yu, H. Cao, H. Xiong, and J. Tian, "Mining personal context-aware preferences for mobile users." in *ICDM*, vol. 12, 2012, pp. 1212–1217.

[13] J. Zhang, B. Liu, J. Tang, T. Chen, and J. Li, "Social influence locality for modeling retweeting behaviors," in *IJCAI*. AAAI Press, 2013, pp. 2761–2767.

[14] K. Chen, T. Chen, G. Zheng, O. Jin, E. Yao, and Y. Yu, "Collaborative personalized tweet recommendation," in *SIGIR*. ACM, 2012, pp. 661–670.

[15] Y. Pan, F. Cong, K. Chen, and Y. Yu, "Diffusion-aware personalized social update recommendation," in *RecSys*. ACM, 2013, pp. 69–76.

[16] M. Ye, X. Liu, and W.-C. Lee, "Exploring social influence for recommendation: a generative model approach," in *SIGIR*. ACM, 2012, pp. 671–680.

[17] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 17, no. 6, pp. 734–749, 2005.

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

[19] G. Szabo and B. A. Huberman, "Predicting the popularity of online content," *Communications of the ACM*, vol. 53, no. 8, pp. 80–88, 2010.

[20] D. M. Blei and J. D. Lafferty, "Dynamic topic models," in *ICML*. ACM, 2006, pp. 113–120.

[21] N. Kawamae, "Trend analysis model: trend consists of temporal words, topics, and timestamps," in *WSDM*. ACM, 2011, pp. 317–326.

[22] Z. Yin, L. Cao, J. Han, C. X. Zhai, and T. Huang, "Lpta: A probabilistic model for latent periodic topic analysis," in *ICDM*. IEEE, 2011, pp. 904–913.

[23] Q. Yuan, G. Cong, Z. Ma, A. Sun, and N. M. Thalmann, "Who, where, when and what: discover spatio-temporal topics for twitter users," in *KDD*. ACM, 2013, pp. 605–613.

[24] J. Tang, X. Hu, H. Gao, and H. Liu, "Exploiting local and global social context for recommendation," in *IJCAI*. AAAI Press, 2013, pp. 2712–2718.