Time-Sensitive Topic-Based Communities on Twitter

4 authors, including:

Fattane Zarrinkalam
Ferdowsi University Of Mashhad
17 PUBLICATIONS   **50** CITATIONS

SEE PROFILE

Weichang Du
University of New Brunswick
74 PUBLICATIONS   **352** CITATIONS

SEE PROFILE

Time-Sensitive Topic-Based Communities on Twitter

Hossein Fani[1,3(✉)], Fattane Zarrinkalam[1,2], Ebrahim Bagheri[1],
and Weichang Du[3]

[1] Laboratory for Systems, Software and Semantics (LS[3]),
Ryerson University, Toronto, Canada
hosseinfani@gmail.com
[2] Department of Computer Engineering,
Ferdowsi University of Mashhad, Mashhad, Iran
[3] Faculty of Computer Science, University of New Brunswick,
Fredericton, Canada

Abstract. This paper tackles the problem of detecting temporal content-based user communities from Twitter. Most existing content-based community detection methods consider the users who share similar topical interests to be like-minded and use this as a basis to group the users. However, such approaches overlook the potential temporality of users' interests. In this paper, we propose to identify *time-sensitive topic-based communities* of users who have similar temporal tendency with regards to their topics of interest. The identification of such communities provides the potential for improving the quality of community-level studies, such as personalized recommendations and marketing campaigns that are sensitive to time. To this end, we propose a graph-based framework that utilizes multivariate time series analysis to represent users' temporal behavior towards their topics of interest in order to identify like-minded users. Further, Topic over Time (TOT) topic model that jointly captures keyword co-occurrences and locality of those patterns over time is utilized to discover users' topics of interest. Experimental results on our Twitter dataset demonstrates the effectiveness of our proposed temporal approach in the context of personalized news recommendation and timestamp prediction compared to non-temporal community detection methods.

Keywords: Community detection · Topic modeling · Time series analysis · Graph clustering · Twitter

1 Introduction

Information sharing and communication patterns of users in social network platforms such as Twitter can lead to the formation of communities that consist of like-minded or similarly behaving users. Besides models that utilize network structure for identifying communities, various topic-based approaches have been

investigated in the literature that employ textual content published by the users or jointly with social connections to detect like-minded users [1,13]. However, these approaches consider the users who share similar topics of interest as like-minded users and do not take into account their temporal behaviour towards these topics. For example, some users who are less interested in political issues such as *'Arresting WikiLeaks Chief in the U.K'* might show their inclination only after a few weeks of delay while others might react immediately to the topic. The community of users who contributes to this topic instantaneously should be viewed as a different community from those users who only pay attention to the topic after several weeks. The ability to model these temporal topic-based communities provides the potential for improving the quality of community-level studies, such as personalized recommendations and marketing campaigns [5,22]. For example, in case of news recommendation, it would be unreasonable to recommend a news article about a topic to those users who showed interest in it a few weeks ago and have since moved on. On the other hand would make reasonable sense to recommend the same article to users who are actively pursuing this topic on Twitter at the current point in time.

We propose to identify communities of user who have similar temporal tendency with regards to the active topics on Twitter based on their published tweets. Hence, we support temporality in our community detection process. Further, since tweets are rather short, noisy and do not provide sufficient contextual information for identifying their semantics, we model semantics of tweets by annotating them with unambiguous concepts described in external knowledge bases such as Wikipedia [21].

Our work in this paper makes the following contributions:

– We propose a model to incorporate user's interests towards active topics on Twitter within a temporal framework. We adopt the Topics over Time (TOT) model [19] that jointly captures word co-occurrences and locality of those patterns in time to simultaneously discover topics and model users' temporal inclination towards these topics. User models are utilized for identifying similar users.
– We propose a graph-based framework over multidimensional user time series for discovering time-sensitive topic-based communities in Twitter.
– We demonstrate the performance of our user community detection approach in the context of personalized news recommendation and time stamp prediction in order to compare our work with the state of the art.

This paper is organized as follows: In the next section, we review the related work. The proposed approach is introduced in Sect. 3. Section 4 is dedicated to our evaluations. Finally, in Sect. 5, we conclude the paper.

2   Related Work

Existing user community detection approaches can be broadly classified into two categories [7]: *Topology-based* and *Topic-based* approaches. The *Topology-based*

community detection approaches represent the social network as a graph whose nodes are users and edges indicate explicit user relationships. This approach relies only on the network structure of the social network graph. On the other hand, *Topic-based* approaches mainly utilize textual content of the users' posts in the social network to detect communities. Topology-based view to community detection on social networks may not be able to identify communities of users that share similar conceptual interests because there are many users on a social network that have very similar interests but are not explicitly connected to each other. Further, many of the social connections may not be due to users' interests similarity but can be due to other factors such as friendship and kinship that do not necessarily point to inter-user interest similarity [6]. Since, the goal of our proposed approach is to detect communities formed toward the topics extracted from users' information contents, we review topic-based community detection methods in this section.

Most of these works have proposed a probabilistic model to detect topic-based user communities based on textual content or jointly with social connections [1,16,17,23]. For example, Abdelbary et al. [1] have identified users' topics of interest and extract user communities based on the topics utilizing Gaussian Restricted Boltzmann Machine. Yin et al. [20] have integrated community discovery with topic modeling in a unified generative model to detect communities of users who are coherent in both structural relationships and latent topics. In their framework, a community can be formed around multiple topics and a topic can be shared between multiple communities. Sachan et al. [17] have proposed probabilistic schemes that incorporate users' posts, social connections and interaction types to discover latent user communities in social networks. In their paper, they have considered three types of interactions: conventional tweets, reply tweets and retweets.

Another class of work attempts to transform the topic-based community detection problem into a graph clustering problem. Liu et al. [11] have proposed a clustering algorithm based on topic-distance between users to detect topic-based communities in a social tagging network. In this work, LDA is used to extract hidden topics in tags. Peng et al. [15] have proposed a hierarchical clustering algorithm to detect latent communities from tweets. They have used the predefined categories in SINA Weibo and have calculated the pairwise similarity of users based on their degree of interest in each category. All the above methods do not incorporate temporal aspects of users' interests and undermine the fact that users of like-minded communities would ideally show similar contribution or interest patterns for similar topics throughout time. Hu et al. [10] is one of the few that consider this aspect. These authors propose a unified probabilistic generative model to extract latent communities over temporal topics and analyze topic temporal fluctuation across different communities.

3   Proposed Approach

In this section, we present our approach to identify Twitter user communities, within a specific time period T, who share similar topics of interest at the same

time. We refer to these communities as *time-sensitive topic-based communities*. To this end, we first model temporal behavior of users' interests towards existing active topics on Twitter in time period T by building their *user-topic contribution time series* and then detect latent communities by means of a graph-based clustering algorithm. These steps are described in the following.

User-Topic Contribution Time Series. Assuming there are K active topics $z_1, z_2, ..., z_K$ present on Twitter in time period T, we view T as L consecutive time intervals, and calculate user's contribution to each topic z at each time interval. Thus, user u is represented by temporally ordered vectors denoted as $\mathbf{Y}^u = [\mathbf{y}_t^u]^{1 \leq t \leq L}$, where $\mathbf{y}_t^u$ is a vector of size K, representing u's degree of interest at time interval t to each topic z. Collectively, $\mathbf{Y}^u$ is a K-variate time series for user u, named the *user-topic contribution time series.*

Considering $\mathbb{M}$, the set of tweets posted during time period T as a text corpus, it is possible to extract topics $\mathbb{Z}$ using topic modeling methods. LDA [3] as a topic modelling method assumes that a document d is a mixture of topics and implicitly uses co-occurrence patterns of terms to extract sets of correlated terms to represent topics. However, the dynamics of topics on Twitter can temporarily impact the co-occurrence patterns of terms. For instance, *'Arrest'* and *'Julian Assange'* have appeared much more frequently on Twitter during the December 2010 time period because of the arrests made with regards to the WikiLeaks case in U.K.; however, this association between the two terms are temporal and cannot be generalized. The LDA method does not model time which can confound co-occurrence patterns and result in unclear, suboptimal topic discovery [19]. We need a topic model that explicitly models time jointly with keyword co-occurrence patterns. To address this issue, we exploit Topics over Time (TOT) [19] to identify active topics on Twitter, which simultaneously captures word co-occurrences and locality of those patterns over time and is hence able to discover more event-specific topics.

On the other hand, applying topic modeling methods such as TOT and LDA to extract topics from tweets might suffer from the sparsity problem [12,18]. Because, they are designed for regular documents rather than short, noisy and informal texts like tweets. To obtain better topics from Twitter without modifying the standard topic modeling methods, we annotate each tweet m from our collection of tweets $\mathbb{M}$ with concepts defined in Wikipedia using an existing semantic annotator and employ these concepts instead of the words/terms. For instance, for a tweet such as *'Sweden issues Warrant for Wikileaks exec Julian Assange's arrest* http://bit.ly/9Ho0WM', a semantic annotator such as TagMe [8] is able to identify and extract four Wikipedia concepts, namely *'Sweden'*[1], *'Arrest_warrant'*, *'WikiLeaks'* and *'Julian_Assange'* as basic units of this tweet. We believe that using concepts instead of words leads to the reduction of noisy content within the topic detection process, because each concept implicitly represents a collection of terms which are collectively more meaningful than a single term [8].

---

[1] https://en.wikipedia.org/wiki/Sweden.

Formally, let $\mathbb{C} = \{c_1, c_2, ..., c_N\}$ be a set of N Wikipedia concepts extracted from $\mathbb{M}$, we consider a concept $c \in \mathbb{C}$ as the basic unit of tweets and model a tweet posted by user u at timestamp ts, $\mathbf{m}_{ts}^u \in \mathbb{M}$, as a vector of N nonnegative integers, where the ith number shows the occurrence frequency of the ith concept. Similar to previous works in the literature [9,12], we aggregate the tweets of a user which are published in a given time interval into a single document for use as training data. More specifically, let T be a specified time period divided into L consecutive time intervals. A document $\mathbf{d}_{ts}^u$ is an aggregate over all tweets of user u posted during time interval $1 \leq t \leq L$, i.e., $\mathbf{d}_t^u = \sum_{t \leq ts < t+1} \mathbf{m}_{ts}^u$.

By running TOT [19] over these documents, we can discover topics $\mathbb{Z} = \{\mathbf{z}_1, \mathbf{z}_2, ...\mathbf{z}_K\}$, where a topic $\mathbf{z}$ is a vector of N real numbers in $\mathbb{R}^{[0,1]}$. The ith number shows the participation score of the ith concept in forming the topic. Collectively, $\mathbb{Z} = \{\mathbf{z} \in \mathbb{R}^{[0,1]^n} : ||\mathbf{z}||^1 = 1\}$ is the set of all topics indexed from 1 to K. $|| \cdot ||^1$ is the L^1-norm of $\mathbf{z}$. In TOT, the vector is the Dirichlet distribution of concepts in the topic with the parameter β; notationally, $\phi_{\mathbf{z} \in \mathbb{Z}} \sim Dir(\beta)$.

Finally, *user-topic contribution time series* of each user u, $\mathbf{Y}^u$, is the stream of document-topic contribution $\mathbf{y}_t^u$ over L consecutive time intervals, i.e., $\mathbf{Y}^u = [\mathbf{y}_t^u]^{1 \leq t \leq L} = [\mathbf{y}_1^u, \mathbf{y}_2^u, ..., \mathbf{y}_L^u]$ where $\mathbf{y}_t^u$ is a vector of K real numbers in $\mathbb{R}^{[0,1]}$. The ith number shows the participation score of the ith topic in forming the document $\mathbf{d}_t^u$ of user u for time interval $1 \leq t \leq L$. TOT [5] assumes a Dirichlet distribution of topics in the document d_t^u with the parameter α; notationally, $\theta_{\mathbf{d}_t^u} \sim Dir(\alpha)$.

We believe that the behavior of the user-topic contribution time series can be considered to be a good measure for finding the similarity between two users in that it allows us to find like-minded users based on their *temporally*-correlated contributions in similar topics. In the following, we use user-topic contribution time series of users to discover time-sensitive topic-based user communities.

Time-Sensitive Topic-Based Communities. In order to extract user communities that consist of like-minded users, who have contributed to the same topics with the same temporal behavior and contribution degrees, we first calculate pairwise similarity between two users by computing the 2D variation of cross correlation measure on their corresponding user-topic contribution time series. Formally, the $2D$ cross-correlation measure of two matrices, such as $A_{I \times J}$ and $B_{I \times J}$, denoted by $XC_{(2I-1) \times (2J-1)}$, is calculated as follows:

$$XC[l,m](A,B) = \sum_{i=0}^{I-1} \sum_{j=0}^{J-1} A[i,j] \times B^*[i-l, j-m] \tag{1}$$

where B^* denotes the complex conjugate of B. Intuitively, the 2D cross-correlation measure slides one matrix over the other and sums up the multiplications of the overlapping elements. Positive row index l corresponds to a downward shift of the rows of A over B and negative column index m indicates a leftward shift of the columns. A maximum correlation occurs at $XC[0,0]$ if the two matrix are the same without any shift.

Now, given the fact that user-topic contribution time series of each user, i.e., $\mathbf{Y}^u$, towards the K topics over L time intervals can be represented as a $K \times L$

matrix, the similarity distance between two users can be calculated through the $2D$ cross-correlation of their user-topic contribution time series without a shift in time. The normalized similarity distance of two users u_1 and u_2, denoted as $usd(u_1, u_2)$, is defined as follows:

$$usd(u_1, u_2) = \frac{XC[0,0](\mathbf{Y}^{u_1}, \mathbf{Y}^{u_2})}{\sqrt{(\mathbf{Y}^{u_1} \cdot \mathbf{Y}^{u_1}) \times (\mathbf{Y}^{u_2} \cdot \mathbf{Y}^{u_2})}} \qquad (2)$$

where $\mathbf{Y}^u$ is the user-topic contribution time series for user u.

We are now able to calculate the correlation distance between all user pairs and build a weighted user graph. A user graph, $UG = < \mathbb{V}, \mathbb{E} >$, is a weighted undirected graph where $\mathbb{V}$ is the set of all users and $\mathbb{E} = \{usd(u_i, u_j) | \forall u_i, u_j \in \mathbb{V}, i \neq j\}$. After constructing the user graph, we apply a graph partitioning algorithm namely the Louvain Method (LM) [4] to extract clusters of users that form latent communities. Louvain is a greedy optimization method that initially finds small communities by locally maximizing the modularity and consequently performs the same procedure on the new graph by considering each community extracted in the previous step as a single vertex. We find the Louvain method to be suitable for topic extraction because of its following characteristics: *(i)* this algorithm can be applied to weighted graphs; *(ii)* it does not require *a priori* knowledge of the number of communities, and *(iii)* it is computationally very efficient when applied to large and dense graphs. While modularity maximization is NP-hard, the complexity of Louvain's greedy implementation is $O(n\log n)$, where n is the number of vertices in the graph [4,14]. To calculate the degree of interest of each extracted user community UC at each time interval t to each topic $\mathbf{z} \in \mathbb{Z}$, named *community-topic contribution time series*, we sum over the user-topic contribution time series of all users who are the members of the user community, i.e., $\mathbf{y}_t^{UC} = \sum_{u \in UC} \mathbf{y}_t^u$.

4   Experiments

In this section we show the performance of our proposed approach in the applications of news recommendation and timestamp prediction on a dataset obtained from the Twitter.

4.1   Dataset

In our experiments, we use a Twitter dataset which is publicly available[2] and presented by Abel et al. [2]. It consists of approximately 3M tweets posted by $136\,\mathrm{K}$ users between Nov. 1 and Dec. 31, 2010. We annotated the text of each tweet with Wikipedia concepts using the TAGME RESTful API [8] which resulted in $351\,\mathrm{K}$ concepts.

---

[2] http://wis.ewi.tudelft.nl/websci11/.

4.2   Qualitative Analysis

To make the overall idea of our time-sensitive community detection approach more clear, in Fig. 1, we utilize heat map to visualize the contribution of three sample users in each day for the two months period of our Twitter dataset between Nov. 1 and Dec. 31 (X-axis) to each extracted topic (Y-axis). The number of topics is set to 50. As illustrated in Fig. 1, the users namely @amfumero, @lazarogonzale and @ebarrera highly contribute to similar topics z_1, z_{12} and z_{34}. However, @ebarrera evidently shows a time shift in his contribution to topic z_{12}. While the first two users' contributions to topic z_{12} span from mid November to early December, @ebarrera starts his posts about the topic from mid December to late December. We do believe that the former two users should be the members of a same community different from the community of the latter.
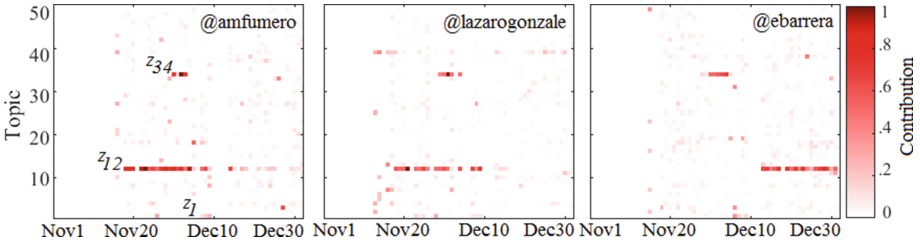

Fig. 1. The heat map of user-topic contribution time series for three Twitter users.

To discover user communities based on the similarity distance between their user-topic contribution time series, we use the implementation of Louvain method available on Pajek[3]. Six of our extracted communities are shown in Fig. 2. Likewise, we use heat map to visualize their community-topic contribution time series in Fig. 3. As seen in this figure, the users of communities UC_1 and UC_2 discuss the same set of topics, z_1, z_{12}, and z_{34}, but in different time intervals with regard to topic z_{12} with a week of delay. Topic-based community detection methods that overlook temporal behavior of users would merge the users of such communities like @amfumero, @lazarogonzale and @ebarrera into a single community. However, our time-sensitive topic-based community detection method incorporates this deviation in users' temporal behavior and therefore group @amfumero and @lazarogonzale into user community UC_1 and @ebarrera to user community UC_2. We believe that this is an important distinguishing feature of our work. To highlight this feature, consider the case of a news recommendation application. It would be unreasonable to recommend a news article on topic z_{12} in December to members of UC_1 who discussed the topic in late November but would make total sense to recommend the same article to members of UC_2 who start to pursue the topic in early December. This is illustrated in Sect. 4.4 by evaluating the performance of a news recommender application. Other communities, UC_3 to UC_6, are formed because there is at least one distinctive topic in their topics of interest. For example, user community UC_3 contributes to topics z_4, z_{34} while members of UC_4 are interested in z_{28}.
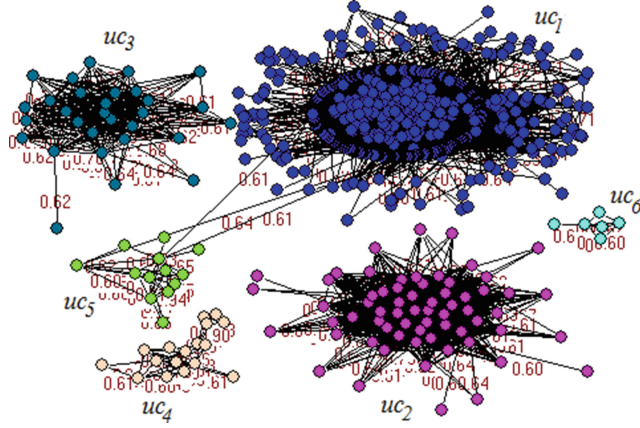
---

[3] http://mrvar.fdv.uni-lj.si/pajek/.

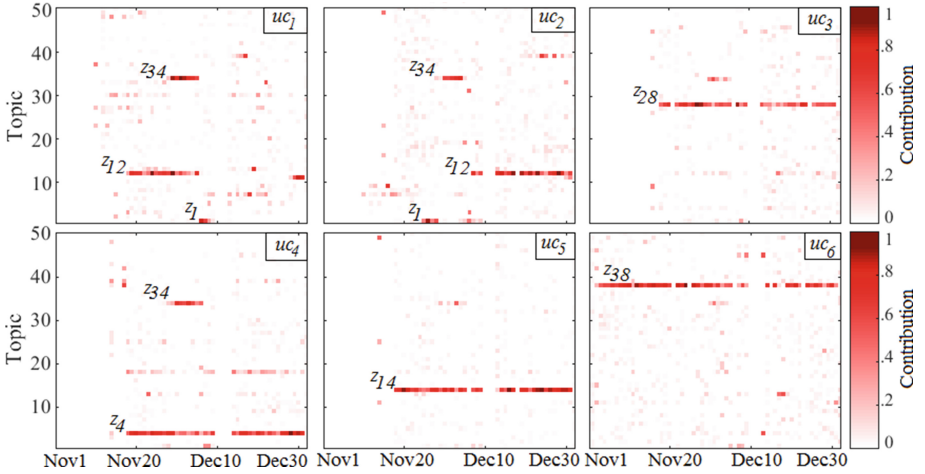**Fig. 2.** The output of the Louvain community detection method.

Fig. 3. The heat maps of community-topic contribution time series for respective user communities in Fig. 2.

4.3   Comparison Methods

We investigated the behavior of our temporal topical community detection method by benchmarking it against the following methods:

(CD-LDA) Non-temporal Community Detection Over LDA Topics. To apply LDA, we aggregated all concepts extracted from tweets of a user in each day as a single document and applied the MALLET implementation of LDA[4] on the collection of such documents to discover the latent topics. To compute

---

[4] http://mallet.cs.umass.edu/topics.php.

the probability distribution of each topic per user, we aggregated the tweets published by a user and performed inference using the LDA model. To discover user communities, we then calculated the pairwise similarity of users based on the cosine similarity of their probability distributions and built a weighted user graph. Finally, the Louvain method was applied for extracting the latent user communities. We set the number of topics to 50.

(CD-TOT) Non-temporal Community Detection Over Topics Over Time. The goal of this method is to evaluate the impact of considering temporal issues on extracting topics. We aggregated the concepts which are mentioned in the tweets of a user in each day as a single document and the date of that day as the timestamp of the document. Then we applied TOT to extract topics. The community detection method and the number of topics that are used are the same as CD-LDA.

(TCD-LDA) Temporal Community Detection Over LDA Topics. In this method, we extract topics and the probability distribution of each topic per user per day, similar to CD-LDA. Then, each user u will be represented by user-topic contribution time series toward the detected LDA topics. To discover communities, we used the proposed method introduced in Sect. 3 under time-sensitive topic-based communities.

(TCD-TOT) Temporal Community Detection Over Topics Over Time. This method is our proposed approach in this paper.

4.4   News Recommendation

The performance of the community detection method can be measured through observations made at the application level such as news recommendation. Therefore, to evaluate our work, we deploy the widely used news recommender application and adopt the same evaluation strategy as mentioned in [2,21]. To this end, we first build ground truth by collecting, for each user, news articles from news agencies' websites to which the user has explicitly linked in her tweets (or retweets). Given this gold standard, the objective is to see whether it would be possible to recommend the right news articles to each Twitter user. The right news articles for a given user would be the ones that they had posted in their timeline. Since our main objective is not to propose a news recommender application, we adopt a simple recommender algorithm as proposed in [2] as follows:

Given K topics extracted in time interval T, we represent each user community UC as a K-tuple vector $\mathbf{I}^{UC}$ over the extracted topics which is calculated by aggregating the community-topic contribution time series of each user community UC over the whole time period T of L consecutive time intervals. Formally,

$$\mathbf{I}^{UC} = \sum_{1 \leq t \leq L} \mathbf{y}_t^{UC} \tag{3}$$

where y_t^{UC} is the K-tuple community-topic contribution vector of user community UC in time interval t. We also represent each news article A as a weighted

vector $\mathbf{I}^A = (i_1^A, i_2^A, ..., i_K^A)$ over the K extracted topics where i_k^A denotes the degree of A's relatedness to topic $\mathbf{z}_k$ and is calculated based on the frequency of the constituent concepts of $\mathbf{z}_k$ in news article A. Given $\mathbf{I}^A$ and $\mathbf{I}^{UC}$, we recommend article A to the users of user community UC based on the cosine similarity of their corresponding vectors.

We use standard information retrieval metrics: Mean Reciprocal Rank (MRR) which is the inverse of the first position that a correct item occurs within the ranked recommendations and Success at rank K (S@K) that shows the probability that at least one correct item occurs within the top-k ranked recommendations. In the following, the five methods are compared to each other in terms of MRR, S@1 and S@10. Figure 4 summarizes the results.

CD-LDA and CD-TOT use the same community detection method over different kinds of topics. CD-LDA uses LDA which extracts topics by only considering co-occurrence between keywords. However, CD-TOT considers time jointly with keywords co-occurrence patterns by applying TOT as its topic detection method. By comparing these two methods it is possible to evaluate the degree of granularity of topics in topical communities for developing news recommendation applications. As illustrated in Fig. 4, CD-TOT outperforms CD-LDA in terms of three metrics. Our observations show that TOT topics which are better localized in time compared to topics of LDA are more effective in recommending related news based on user's interests. By comparing the results of the TCD-LDA and TCD-TOT, this observation is also confirmed.

The difference between the CD-LDA and TCD-LDA is in their community detection method. TCD-LDA produces communities of users who have contributed to the same topics with the same temporal behavior and contribution degrees. However, communities extracted by CD-LDA are based on the similarity of user's topics of interest. As Fig. 4 shows, TCD-LDA outperforms CD-LDA in terms of all metrics. This means that incorporating temporal aspects of users'
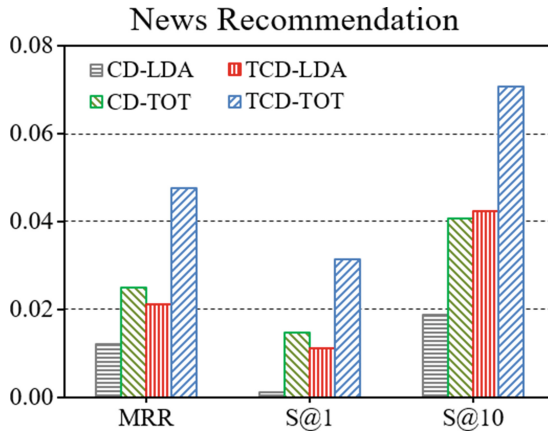

Fig. 4. Comparison between different methods in context of news recommendation.

interests to extract like-minded communities leads to more cohesive communities that consequently results in higher quality news recommendations. This is also confirmed by comparing the results of CD-TOT and TCD-TOT.

Further, when looking at the results in Fig. 4, it can be concluded that our proposed method, i.e. TCD-TOT, that produces temporal topical communities leads to highest quality of recommendations in comparison with the other methods. The reason lies in the fact that our approach incorporate user temporal behavior and contribution degrees of users over TOT topics that takes into account both time and co-occurrence of keywords to discover topics.

4.5   Timestamp Prediction

In this section, adopted from [19], we evaluate our temporal communities in terms of the capability of predicting the timestamp of tweets given their content. To do so, given the extracted user communities and a tweet, we first find the most similar community to the tweet in terms of the degree of their contribution to each topic. Then, within the selected community, we predict the tweet's timestamp by taking the peak time for the tweet's topic within that community. As our test set, we randomly selected 1,000 tweets which were not included when our communities were built (leave-out protocol) and were annotated with at least 5 concepts. The results of comparing TCD-TOT and CD-TOT are illustrated in Fig. 5 in terms of prediction accuracy as a function of the tolerance range in days (e.g. when tolerance range is set to 20 days, we consider a predicted timestamp as accurate if its difference with ground truth timestamp is within 20 days). Based on the results, we can see that applying our proposed temporal community detection method over TOT topics, TCD-TOT, leads to achieving more accurate prediction results compared to the non-temporal community detection method, CD-TOT.
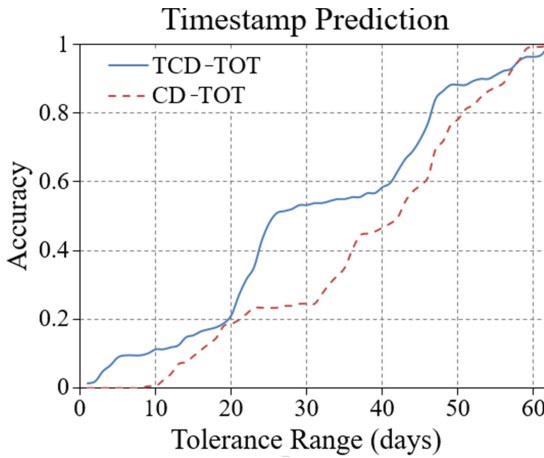

Fig. 5. The results of timestamp prediction in terms of prediction accuracy as a function of the tolerance range in days.

5   Conclusion

In this paper, we have addressed the problem of identifying time-sensitive topic-based user communities on Twitter. We focus on communities which are formed not only based on the different topics of interest to the users, but also based on the temporality of the user contributions. To this end, we have utilized Topic over Time (TOT) that jointly captures term co-occurrences and locality of those patterns over time to discover active topics of Twitter in a given time period based on the tweets published in that time period. Further, we extract user communities through novel multivariate time series modeling of like-minded users. We investigated the performance of the proposed approach in the application of news recommendation and timestamp prediction. The experimental results indicate that the proposed model achieves better performance in comparison with non-temporal community detection methods.

References

1. Abdelbary, H., ElKorany, A., Bahgat, R.: Utilizing deep learning for content-based community detection. In: Science and Information Conference, pp. 777–784 (2014)
2. Abel, F., Gao, Q., Houben, G.-J., Tao, K.: Analyzing user modeling on twitter for personalized news recommendations. In: Konstan, J.A., Conejo, R., Marzo, J.L., Oliver, N. (eds.) UMAP 2011. LNCS, vol. 6787, pp. 1–12. Springer, Heidelberg (2011)
3. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. In: Advances in Neural Information Processing Systems, NIPS 2001, vol. 14, pp. 601–608 (2001)
4. Blondel, V., Guillaume, J., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. J. Stat. Mech. Theor. Exp. **2008**(10), P10008 (2008)
5. Chen, J.M., Tang, Y., Li, J.G., Mao, C.J., Xiao, J.: Community-based scholar recommendation modeling in academic social network sites. In: Huang, Z., Liu, C., He, J., Huang, G. (eds.) WISE Workshops 2013. LNCS, vol. 8182, pp. 325–334. Springer, Heidelberg (2014)
6. Deng, Q., Li, Z., Zhang, X., Xia, J.: Interaction-based social relationship type identification in microblog. In: International Workshop on Behavior and Social Informatics and Computing, pp. 151–164 (2013)
7. Ding, Y.: Community detection: topological vs. topical. J. Infometrics **5**(4), 498–514 (2011)
8. Ferragina, P., Scaiella, U.: Fast and accurate annotation of short texts with wikipedia pages. J. IEEE Softw. **29**(1), 70–75 (2012)
9. Hong, L., Davison, B.: Empirical study of topic modeling in twitter. In: 1st ACM Workshop on Social Media Analytics, pp. 80–88 (2010)
10. Hu, Z., Yao, J., Cui, B.: User group oriented temporal dynamics exploration. In: AAAI 2014, pp. 66–72 (2014)
11. Liu, H., Chen, H., Lin, M., Wu, Y.: Community detection based on topic distance in social tagging networks. TELKOMNIKA Indonesian J. Electr. Eng. **12**(5), 4038–4049 (2014)
12. Mehrotra, R., Sanner, S., Buntine, W., Xie, L.: Improving lda topic models for microblogs via tweet pooling and automatic labeling. In: SIGIR 2013, pp. 889–892. ACM (2013)

13. Natarajan, N., Sen, P., Chaoji, V.: Community detection in content-sharing social networks. In: ASONAM 2013, pp. 82–89 (2013)
14. Noack, A., Rotta, R.: Multilevel local search algorithms for modularity clustering. In: SEA 2009, pp. 257–268 (2009)
15. Peng, D., Lei, X., Huang, T.: Dich: a framework for discovering implicit communities hidden in tweets. World Wide Web **18**(4), 795–818 (2014)
16. Rosen-Zvi, M., Griffiths, T., Steyvers, M., Smyth, P.: The author-topic model for authors and documents. In: UAI 2004, pp. 487–494 (2004)
17. Sachan, M., Contractor, D., Faruquie, T., Subramaniam, L.: Using content and interactions for discovering communities in social networks. In: WWW 2012, pp. 331–340 (2012)
18. Sriram, B., Fuhry, D., Demir, E., Ferhatosmanoglu, H., Demirbas, M.: Short text classification in twitter to improve information filtering. In: SIGIR 10, pp. 841–842 (2010)
19. Wang, X., McCallum, A.: Topics over time: a non-markov continuous-time model of topical trends. In: KDD 2006, pp. 424–433 (2006)
20. Yin, Z., Cao, L., Gu, Q., Han, J.: Latent community topic analysis: integration of community discovery with topic modeling. ACM Trans. Intell. Syst. Technol. **3**(4), 63:1–63:21 (2012)
21. Zarrinkalam, F., Fani, H., Bagheri, E., Kahani, M., Du, W.: Semantics-enabled user interest detection from twitter. In: WI-IAT 2015, pp. 469–476 (2015)
22. Zhao, G., Lee, M., Hsu, W., Chen, W., Hu, H.: Community-based user recommendation in uni-directional social networks. In: CIKM 2015, pp. 189–198 (2013)
23. Zhou, D., Manavoglu, E., Li, J., Giles, C., Zha, H.: Probabilistic models for discovering e-communities. In: WWW 2006, pp. 173–182 (2006)