Temporal Latent Space Modeling
for Community Prediction

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

[1] Faculty of Computer Science, University of New Brunswick,
Fredericton, Canada
{hfani,wdu}@unb.ca
[2] Laboratory for Systems, Software and Semantics (LS3), Ryerson University,
Toronto, Canada
{hossein.fani,bagheri}@ryerson.ca

Abstract. We propose a temporal latent space model for user community *prediction* in social networks, whose goal is to predict *future* emerging user communities based on past history of users' topics of interest. Our model assumes that each user lies within an unobserved latent space, and similar users in the latent space representation are more likely to be members of the same user community. The model allows each user to adjust its location in the latent space as her topics of interest evolve over time. Empirically, we demonstrate that our model, when evaluated on a Twitter dataset, outperforms existing approaches under two application scenarios, namely news recommendation and user prediction on a host of metrics such as mrr, ndcg as well as precision and f-measure.

1 Introduction

Social networks have been an effective medium for communication and social interaction. Predicting users' behaviour, interactions, and influence are of interest due to their wide range of applications such as personalized recommendations and marketing campaigns. Community-level analytics provide the means to understand social network dynamics at a higher collective level. In order to support community-level models, various community detection methods have been proposed, which employ information such as users' social connections and content engagement to identify communities. The objective of these models is to identify past and current user communities; however, little work has been done on *community prediction*, that is, to determine how the community structure of a social network will look like in a future yet-to-be-observed time interval.

In this paper, we focus on an instance of this problem, namely *content-based* (*topical*) *future community prediction*. Specifically, given a sequence of users' contributions towards a set of topics from time interval 1 to T, the goal is to predict topical user communities in a future interval T + 1. To perform topical future community prediction, we construct graph snapshots G_t for each time interval t in which users are linked based on pairwise topical similarity

at time interval t. Given the sequence of graph snapshots $[G_{1:T}]$ from time 1 to T, we propose temporal latent space modeling to predict inter-user topical similarities in G_{T+1} after which a community detection method yields future user communities.

Latent space modeling [26] has been successfully employed for link prediction in graphs where given the observed links in the graph, the location of each node in a latent space is learned such that the closer two nodes are in that space, the higher the probability of a link between them would be. In other words, similarity in latent space translates into links in graph space. Latent space modeling preserves homophily [23] where links between nodes are considered clues for similarity, and so, densely connected groups of nodes imply communities.

Different approaches based on matrix factorization and deep neural networks [1,31–33] have been proposed to learn the latent space model of users in the social network. For instance, Akbari et al. [1] propose to learn a single multi-modal latent space representation from users' social views including network structure as well as contents, inter-user interactions (e.g., reply or retweet), and prior knowledge (if any). However, such studies are concerned with static graphs, where the latent representation of the users are fixed. They overlook the fact that latent space representations need to evolve over time and, hence, fall short when identifying user communities of the *future*.

Temporal tensor-factorization approaches [11], or temporal latent space models [34] go beyond static networks and assume that the network is dynamic and changes over time. Such models endeavor to learn low-rank latent space representations based on the intuition that nodes can move in latent space over time. While suitable for predicting links in a social network, dynamic link prediction models are inherently deficient when the communities need to take users' content similarity into account, i.e., identify *content-based* user communities in the future. Temporal content-based user community prediction is of interest due to the following reasons: (i) there are many users on a social network that have similar interests but are not explicitly linked to each other, (ii) an explicit social link does not necessarily indicate user interest similarity but could be owing to sociological processes such as conformity, sociability or other factors such as friendship and kinship [10,28], (iii) there are some cases where the network structure is not accessible [4] or misleading, e.g., when links are fraudulent due to link-farming [21], and (iv) empirical research has shown that link evolution happens at a much lower pace compared to content changes [24], particularly because links are often not removed when they become effectively 'dead'.

Temporal content-based user community detection methods currently exist that incorporate temporal aspects of users' content and stress that users of the same community would ideally show similar interest patterns for similar topics over time [12,17,18]. However, users' temporal content is only used for pairwise user similarity calculation to build content-based user communities as opposed to user community *prediction*. As a result, they have limited applicability for identifying user communities of the future. Regression techniques such as autoregressive integrated moving average (arima) and support vector regression (svr)

that leverage temporal information to predict users' future interests have shown promising results and can be utilized to identify user communities in the future based on users' pairwise content similarity [3]. However, they require building predictive models on a per user basis and, hence, are practically prohibitive.

In this work, we propose temporal latent space modeling to predict content-based user communities in the future. First, contrary to *non*-temporal methods [31–33], our approach incorporates temporality. Second, in contrast to Zhu et al.'s method [34] and the likes [11,14,27,35] that focus on dynamics of social network structure, our approach employs dynamics of social content. Third, although we use temporal information to predict future users' topical interests similar to regression methods, we train only one model for all users and, thus, significantly reduce computational cost compared to regression techniques. Last, unlike Fani et al. [12] and Hu et al. [18] who employ users' temporal and topical interests for performing pairwise user similarities in order to identify user communities up until 'now', our work in this paper employs such information for predicting user communities 'in the future', which is a step forward compared to the state of the art. We perform experiments on a Twitter corpus and compare our work for user community prediction with several state-of-the-art baselines in the context of news recommendation and user prediction. The results show that our method, which incorporates the temporal evolution of users' topics of interest within latent space, exhibits a stronger predictive power compared to the baselines.

The key contributions of this paper are as follows:

1. We propose a temporal latent space model for user community prediction in social networks that (i) allows users to change their latent representations as their topics of interest evolve over time, and (ii) users who are similar not only in their contribution towards the same set of topics of interest, but also have similar temporal behaviour, remain close in d-dimensional latent space.
2. We illustrate how temporal latent space modeling can be effectively employed to predict future emerging content-based user communities given the past history of users' topics of interest.
3. We perform experimentation on a Twitter dataset to demonstrate the superiority of our proposed model compared to the state of the art methods under two application scenarios, namely, news recommendation and user prediction on a host of metrics such as mrr, ndcg as well as precision and f-measure.

The rest of the paper is organized as follows: In Sect. 2 we describe the related work. Sections 3 and 4 are dedicated to the problem definition and the details of our proposed approach. Section 5 presents our experimental work after which Sect. 6 concludes the paper.

2   Related Work

In this paper, we assume that an existing state of the art technique such as those proposed in [5,30,36] can be employed for extracting and modeling users' topics of interest. Therefore, we will not be engaged with the process of identifying

topics and will only focus on determining content-based user communities in a future time interval based on the temporal interest of users from the past up until now towards those topics. Given this focus, the related works to this paper are largely centered around two areas: (1) temporal user community detection; and, (2) temporal latent space modeling.

2.1 Temporal User Community Detection

There is a rich line of research on user community detection; ranging from link-based community detection methods [16,19,31], which rely only on network structure, to content (topic)-based approaches, which mainly focus on information content generated by the users [4,22]. More recently, several effective approaches have been proposed which integrate both the network structure (links) and content to improve community detection performance [1,7]. All these works assume that the user's topics of interest remain stable across time. However, very few consider the notion of temporality in users' topics of interest [12,17,18], particularly in online social networks such as Twitter.

From among the work that consider temporality, Hu et al. [18] have proposed a probabilistic generative model jointly over text, time and links, namely community level diffusion (cold), to simultaneously identify both user communities and topics in order to uncover inter-community influence dynamics. The generative process can be summarized in three steps. First, per community topic distribution is sampled according to Dirichlet. Then for each community-topic pair, temporal distribution (timestamp) of a topic for a community is also sampled according to Dirichlet. Finally, a user chooses a community based on her community membership distribution and selects a topic from the community according to the community's topic distribution to generate a post (e.g., a tweet). The time of the post is from the temporal distribution of that topic for that community. Contrary to the unified generative model, Fani et al. [12] have proposed a neural embedding approach to identify temporally like-minded user communities given the user generated textual content. They model the users' temporal contribution towards topics of interest by introducing the notion of regions of like-mindedness between users. These regions cover users who share not only similar topical interests but also similar temporal behaviour. By considering the identified set of regions of like-mindedness as a context, a neural network is trained such that the probability of a user in a region is maximized given other users in the same region. The final weights of the neural networks form the low-dimensional vector representation of each user that incorporates both topics of interest and their temporal nature. Finally, a graph partitioning technique is applied on a weighted user graph in which the similarity of two users is based on the cosine similarity of their respective vectors to identify like-minded user communities. While both of these works sketch different architectures to incorporate temporality, they have shown performance improvement in modelling content-based user communities, particularly in time-sensitive applications such as item recommendation. However, in both work, users' temporal content is only

used to build user communities up until now and have limited applicability for identifying user communities of the future.

2.2 Temporal Latent Space Modeling

Temporal latent space modeling aims at learning the evolution of a temporal graph node and/or an edge over time into a *dynamic* low dense d-dimensional vector in latent space which is allowed to modify its position in the latent space over time. This can be employed for different underlying applications such as link prediction and node classification [13]. Various works such as Temporal tensor-factorization [11,34], and neural embeddings [14,27,35] have been proposed in this respect. For instance, Singer et al. [27] extend the prior neural-based embedding approaches on static graphs, e.g., node2vec [15], to temporal graphs. They propose a semi-supervised algorithm, namely tNodeEmbed, that learns to combine a node's historical temporal embeddings into a final embedding such that it can optimize for a given underlying task, e.g., link prediction. tNodeEmbed initializes node embeddings for time interval t using a static node embedding method. Since initialization for each time interval happens in isolation and independently, the coordinates of the embeddeding space are not guaranteed to align. tNodeEmbed learns a rotation matrix to align coordinates of node embeddings within the time intervals given the fact that a node's temporal behaviour between two consecutive time intervals is gradual and does not follow a bursty pattern (temporal smoothness). Each node embedding at time t, after alignment, is fed to a recurrent neural network (rrn) with long short term memory (lstm) to output the final temporal embedding of the node by optimizing for the specific task. Zhu et al. [34], however, propose temporal latent space modeling for the task of dynamic link prediction via non-negative matrix factorization followed by a scalable inference algorithm based on block coordinate gradient descent (bcgd) to obtain embeddings in linear time.

While suitable for predicting links in a social network structure, proposed temporal latent space models are inherently deficient when the communities need to take users' content similarity into account, i.e., identify *content-based* user communities in the future. To the best of our knowledge, no approach investigates the application of temporal latent space modeling for temporal content-based community detection, which is the main objective of this paper.

3   Problem Definition

Given a set of topics $\mathcal{Z}$ from a social network, such as Twitter, within T time intervals extracted by a topic detection method (e.g., lda) and a set of users $\mathcal{U}$, we represent the topic preference of user $u \in \mathcal{U}$ towards topic set $\mathcal{Z}$ at time interval $t : 1 \leq t \leq \mathrm{T}$ as a vector $\mathbf{x}_{ut} = [x_{ut,1:|\mathcal{Z}|}]$, namely *topic preference vector*, where $x_{ut,z} \in \mathcal{R}^{[0,1]}$ indicates the preference by user u for topic z at time interval t. We let temporal graph $G_t = (\mathcal{U}, \mathcal{E}_t, s)$ represent the content-based similarity between the users of the social network whose nodes are users in $\mathcal{U}$ and $\mathcal{E}_t$ is

the set of weighted undirected edges whose weights are based on a similarity function s, which is defined as the cosine similarity of topic preference vectors of the users at time interval t, i.e., $\forall u, v \in \mathcal{U} : s(u, v : t) = \frac{\mathbf{x}_{ut} \cdot \mathbf{x}_{vt}}{|\mathbf{x}_{ut}||\mathbf{x}_{vt}|}$. Given $[G_{1:T}]$, we aim to accurately predict a set of induced subgraphs in G_{T+1} to form content-based user communities at time interval $T+1$.

4   Proposed Approach

Our approach consists of three subsequent phases: users' topic preference detection, temporal latent space inference, and community prediction. In the following, we lay out the details of each step.

4.1   Topic Preference Detection

To instantiate the topic preference vector, we find (i) a set of topics $\mathcal{Z}$ that have been observed within T time intervals, and (ii) u's degree of interest at time interval t towards each topic $z \in \mathcal{Z}$, i.e., $x_{ut,z}$ in topic preference vector $\mathbf{x}_{ut} = [x_{ut,1} \,..\, x_{ut,z} \,..\, x_{ut,|\mathcal{Z}|}]$. We derive the set of topics from the collection of users' posts using lda [5]. To this end, we view all tweets authored by each user u at time interval t as a single document d_{ut}. Given the document corpus $\mathcal{D} = \{d_{ut} | \forall u \in \mathcal{U}, 1 \leq t \leq T\}$ and the number of topics $|\mathcal{Z}|$, lda distills $\mathcal{D}$ into two probability distributions: (i) distribution of words in each topic (ϕ_z); and (ii) distribution of each topic z in each document ($\theta_{d_{ut,z}}$) showing u's degree of interest toward z at time t. Formally, $x_{ut,z} = \theta_{d_{ut,z}}$. Once users' topic preference vectors have been identified at time interval t, we are able to calculate the similarity function s for all pairs of users and build the temporal graph G_t.

4.2   Temporal Latent Space Inference

Within time period T, the stream of graphs $[G_1 \,..G_t.. \, G_T]$ could be considered as a dynamic graph $\mathscr{G}$ which is evolving over time. We map each user u up until time interval $t \leq T$ to a low-rank d-dimensional latent space, denoted by $\mathbf{y}_{ut}$, while imposing the following assumptions: (i) users change their latent representations over time, (ii) two users that are close to each other in $\mathscr{G}$ remain close in latent space, (iii) two users who are close in latent space share similar topics of interest with each other.

Formally, given a dynamic network $\mathscr{G}$, we find a d-dimensional latent space representation for $\forall u \in \mathcal{U}$ for time interval $1 \leq t \leq T$ that minimizes the *quadratic loss* with temporal regularization:

$$\arg \min \Big[\sum_{t=1}^{T} \sum_{u,v \in \mathcal{U}} \overbrace{|s(u,v:t) - \mathbf{y}_{ut}\mathbf{y}_{vt}^{\top}|_F^2}^{\text{quadratic loss}} + \lambda \sum_{t=1}^{T} \sum_{u \in \mathcal{U}} \overbrace{\big(1 - \mathbf{y}_{ut}\mathbf{y}_{u(t-1)}^{\top}\big)}^{\text{temporal smoothness}} \Big] \quad (1)$$

$$\forall u \in \mathcal{U}; \mathbf{y}_{ut} \geq 0, \mathbf{y}_{ut}\mathbf{y}_{ut}^{\top} = 1$$

where in the quadratic loss component, $s(u, v : t)$ is the similarity score for a pair of users u and v in G_t, $\mathbf{y}_{ut}$ is the d-dimensional latent representation for u up until time interval t, λ is a regularization parameter, and the temporal smoothness component $(1 - \mathbf{y}_{ut}\mathbf{y}_{u(t-1)}^{\top})$ penalizes user u for a sudden change in its location in latent space. Our model maps each user to a point in a unit hypersphere rather than simplex, because sphere modeling gives a clearer boundary between similar users and dissimilar users when mapping all user pairs into latent space. It is worth noting that in our proposed model, a user's position in latent space up until time interval t depends on preceding movement of the user in the latent space since the first time interval $1 \leq t' < t$ via observation of $[G_1 .. G_t]$. This is contrary to static models that obtain latent representation based solely on G_t.

Optimizing Eq. 1 is expensive in terms of space and time complexity as it requires all graphs in $\mathscr{G}$ to jointly update all temporal latent representations for users in all time intervals. To optimize Eq. 1, we use the *local* block coordinate gradient descent (bcgd) algorithm [34], in which inference happens sequentially. Specifically, we optimize users' latent representation locally by minimizing the following objective function at each time interval t:

$$\arg\min \sum_{u,v \in \mathcal{U}} (s(u, v : t) - \mathbf{y}_{ut}\mathbf{y}_{vt}^{\top})^2 + \sum_{u \in \mathcal{U}} (1 - \mathbf{y}_{ut}\mathbf{y}_{u(t-1)}^{\top}) \qquad (2)$$

The local bcgd algorithm infers users' latent representation from a single graph snapshot G_t and prior initialization from $\mathbf{y}_{u(t-1)}$. The algorithm iteratively updates $\mathbf{y}_{ut}$ until it converges and then moves to the computation of temporal latent space in the next time interval $t + 1$. This local sequential update schema greatly reduces the computational cost in practice. We refer readers to [34] for in-depth analysis of the local bcgd algorithm.

4.3   User Community Detection in the Future

Our goal is to predict those user communities whose members share similar temporal expositions toward similar topics of interest in the future graph G_{T+1}. To do so, we first estimate the future graph G_{T+1}. Based on our model, the topical similarity between two users depends only on their latent representations. In other words, the more two latent representations for a pair of users are close, the more similar the users are in terms of topics of interest. As a result, given $\forall u, v \in \mathcal{U} : \mathbf{y}_{u(T+1)}$ and $\mathbf{y}_{v(T+1)}$, we are able to predict future graph $G_{T+1} = (\mathcal{U}, \mathcal{E}_{T+1}, s)$ assuming $s(u, v : T+1) = \mathbf{y}_{u(T+1)}\mathbf{y}_{v(T+1)}^{\top}$. However, $\mathbf{y}_{u(T+1)}$ and $\mathbf{y}_{v(T+1)}$ are not available and have to be approximated based on temporal latent representations up until time interval T, i.e., $\mathbf{y}_{u(T+1)} = \eta(f(\mathbf{y}_{u1}, .. \mathbf{y}_{ut}, .. \mathbf{y}_{uT}))$ where η is a link function and f is a temporal function.

In order to ensure that our proposed model considers all user information from time interval 1 to T when learning the representation for $\mathbf{y}_{u(T+1)}$, we define the user representation at each time interval to be a summarized representation of that user's activities in all previous time intervals. In other words, $\mathbf{y}_{u(T+1)}$ will encode information for user u in time intervals 1 to T. Similarly, the user

representation for u for the time period 1 to $T-1$ is captured in $\mathbf{y}_{uT}$. As such, a user's latent position at time interval T depends on her latent representation at time interval $T-1$ (which already captures user information up to $T-2$), notationally, $\mathbf{y}_{u(T+1)} = \eta(f(\mathbf{y}_{uT}))$. Without loss of generality, we choose f as the identity function and η as the identity link function. Hence, user's latent representation at time T becomes the proxy for her latent representation at time $T+1$, as suggested by Zhu et al. [34], i.e.,

$$\mathbf{y}_{u(T+1)} \simeq \mathbf{y}_{uT} \tag{3}$$

$$s(u, v : T+1) = \mathbf{y}_{uT}\mathbf{y}_{vT}^{\top} \tag{4}$$

where $\mathbf{y}_{uT}$ encapsulates the latent representations of all snapshots from 1 to $T-1$.

Now, given G_{T+1}, we employ a graph partitioning heuristic to extract clusters of users that form our final user communities in the future. We leverage the Louvain method [6] as it is a linear heuristic for the problem of graph partitioning based on modularity optimization. Louvain can be applied to weighted graphs, does not require *a priori* knowledge about the number of communities, and is computationally efficient on large graphs [25]. The application of Louvain on G_{T+1} produces a set of induced subgraphs such as $G_{T+1}[\mathcal{C}]$ whose vertex set $\mathcal{C} \subset \mathcal{U}$ and edge set consists of all of the edges in $\mathcal{E}_{T+1}$ that have both endpoints in $\mathcal{C}$. Subgraphs with $|\mathcal{C}| \geq 2$ form instances of user communities.

5   Evaluation

5.1   Dataset and Experimental Setup

We adopted a Twitter dataset consisting of 2,948,742 tweets authored by 135,731 users in Nov. and Dec. 2010. The two month time period is sampled on a daily basis, i.e., $T+1 = 61$. The settings in each step of our method are as follows:

Topic Preference Detection. We applied lda using Mallet api after removing stopwords. The number of topics used for reporting results in this paper has been set to $|\mathcal{Z}| = 50$ noting that other topic sizes did not change the findings of this paper. We created $\forall u \in \mathcal{U} : \mathbf{x}_{ut} = [x_{ut,1} \,.. \, x_{ut,z} \,.. \, x_{ut,50}]$ for $t = 1$ up to day 60 as our observation to build $\mathscr{G} = [G_1 \,.. \, G_t \,.. \, G_{T=60}]$ in order to predict $G_{T+1=61}$ at the future day 61.

Model Training. We adopt sequential (local) version of block coordinate gradient descent proposed by Zhu et al. [34]. By setting the temporal smoothness (regularization) parameter $\lambda = 0.01$, we performed experiments on increasing number of dimensions $d \in \{10, 20, ..., 100\}$ for learning temporal latent representation of users in 1,000 iterations.

User Community Detection in Future. We apply Louvain with resolution parameter 1.0 using Pajek[1] to identify subgraphs.

---

[1] mrvar.fdv.uni-lj.si/pajek/.

5.2 Baselines

We compare our work against four categories of baselines:

Community Prediction Baseline. To the best of our knowledge, the most related baseline to our work is a temporal content-based latent space model proposed by **Appel et al.** [2] where *shared* matrix factorization has been used to embed social network dynamics and temporal content in a shared feature space followed by a traditional clustering technique, such as k-means, to identify user communities.

Temporal Community Detection Baselines. Hu et al. [18] and **Fani et al.** [12] are two temporal content-based community detection baselines. The former is a generative process for predefined number of topics and communities. This method is a mixture model in which all users are members of all communities with a probability distribution. We only consider the community with the highest probability as the user's community. The latter is based on temporal user embeddings. This method learns a mapping from the user space to a low-rank latent space that incorporates both topics of interest and their temporality.

Non-temporal Community Detection Baselines. Ye et al. [33] and **Louvain** [6] are *non*-temporal link-based community detection methods from two extremes of neural-based non-negative matrix factorization (nmf) and modularity optimization, respectively. To select the best setting for each method, we performed experiments on increasing number of communities $C = \{5, 10, 20, 30\}$ for Appel et al., Hu et al., and Ye et al., and varying embedding dimensions $d = \{100, 200 .. 500\}$ and $d = \{5, 10, 20, 30\}$ for Fani et al. and Appel et al., respectively. The final communities which are based on the users' temporal content until day 60 are used to predict communities in day 61.

Collaborative Filtering Baselines. Temporal collaborative filtering methods are able to predict users' topics of interest in future and, hence, can be used for the task of content-based community prediction, among which we choose the strongest methods, namely, **timesvd++** [20] and **rrn** [29] as our baselines. We performed grid search over the bin size in $\{1, 2, 4 .. 64\}$ and factor size in $\{10, 20, 40, 80\}$ to select the best settings.

5.3 Evaluation Methodology

Contrary to small real social networks or synthetic ones, gold standard communities are often not available for real world applications [8]. As such, well-defined quality measures such as rand index or normalized mutual information (nmi) that require comparison to the gold standard cannot be used. On the other hand and in the absence of a golden standard, quality functions such as *modularity* are not helpful either since they are based on the explicit links between users. In our approach and the baselines, the links between the users are inferred through a learning process and are not explicit. For instance, a near perfect method may result in low modularity because graph edges are sparse and do not form densely
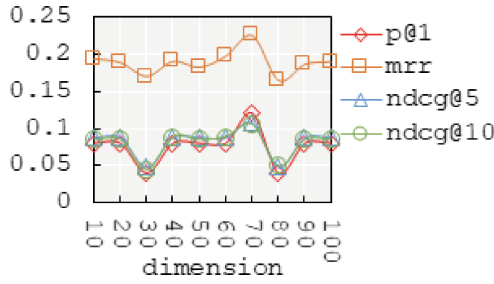
Fig. 1. The impact of dimension size on our method.

connected user sets. Conversely, a weak method may connect topically dissimilar users together forming communities of users that do not share similar interests but have a high modularity. So, the communities that achieve high structural quality in an inferred similarity graph are not necessarily optimal.

Alternatively, the performance of community detection methods can be measured through observations made at the application level. In these *extrinsic* evaluation strategies, a user community detection method is considered better *iff* its output communities improve an underlying application. We deploy two applications, namely news recommendation, and user prediction. By using these applications, we explore whether our proposed method is able to provide stronger performance compared to the state of the art.

To this end, we first build a gold standard dataset for these applications by collecting news articles to which a user has explicitly linked in her tweets (or retweets). We postulate that users post news articles since they are interested in the topics of the news articles. We build the gold standard from a set of news articles whose urls have been posted by user u at time $\mathrm{T}+1$. We see each entry as a triple $(u, a, \mathrm{T}+1)$ consisting of the news article a, user u, and the time interval $\mathrm{T}+1$ to form our gold standard.

5.4   Results

We compare the quality of the communities predicted by our method against the baselines in the context of news recommendation and user prediction.

News Recommendation. To evaluate communities of the future in the context of the news recommendation, we recommend news articles in two steps:

1. For each community $\mathcal{C}$, we recommend news articles in a ranked list based on the similarity of the article a and the community's overall topic preference vector at time $\mathrm{T}+1$. The overall topic preference vector for a community is the sum over all users' topic preference vector belonging to the community, i.e., $\sum_{u \in \mathcal{C}} \mathbf{x}_{u(\mathrm{T}+1)}$.
2. We recommend news article a to user $u \in \mathcal{C}$ based on the same ranked list as her community's list. A true community is one whose members are interested

in the same topics of interest in the future. As a result, at time $T + 1$, a news article is about the same topics of interest as the community's overall interests *iff* all the members post about the same or similar news articles.

Table 1. Comparison with baselines. Asterisk (*) indicates statistically significant improvement over other baselines using paired t-test at $p < 0.05$.

Method	News recommendation			User prediction		
	mrr	ndcg5	ndcg10	precision	recall	f-measure
Community prediction						
Our approach	**0.225***	**0.108***	**0.105***	**0.012***	0.035	**0.015***
Appel et al. [PKDD'18]	0.176	0.056	0.055	0.007	0.094	0.0105
Temporal community detection						
Hu et al. [SIGMOD'15]	0.173	0.056	0.049	0.007	0.136	0.013
Fani et al. [CIKM'17]	0.065	0.040	0.040	0.007	0.136	0.013
Non-temporal link-based community detection						
Ye et al. [CIKM'18]	0.139	0.056	0.055	0.008	0.208	0.014
Louvain [JSTAT'08]	0.108	0.048	0.055	0.004	0.129	0.007
Collaborative filtering						
rrn [WSDM'17]	0.173	0.073	0.08	0.004	**0.740***	0.008
timesvd++ [KDD'08]	0.141	0.058	0.064	0.003	0.657	0.005

We evaluate the recommended list of news articles using standard retrieval metrics such as mrr, ndcg5, and ndcg10. Foremost, we analyze the effect of dimension d on our inference algorithm. We vary d from 10 to 100 and report the performance in Fig. 1. As seen, the overall trend indicates that the recommendation performance in terms of all ranking metrics increases with the number of dimensions up to an extremum at $d = 70$. Next, we compare our proposed method at its best setting ($d = 70$) against the baselines at their best settings in Table 1. As shown, our proposed method outperforms other baselines in terms of all ranking metrics in the context of news recommendation. We attribute the accuracy of our proposed approach to the fact that it directly models and leverages the impact of users' pairwise similarity over their topics of interest within the time dimension, i.e., sequence of similarity graphs, which has been overlooked in all of the other baselines. For instance, Hu et al. is neither a predictive model nor aware of temporal similarity among users, and Ye et al. and Louvain do not take temporal information into account at all. However, it is worth noting that due to capturing sequences of inter-user similarities indirectly through collaborative filtering, rrn was able to become the runner-up in terms of ndcg5 and ndcg10.

User Prediction. The other application with which we evaluate our approach is the user prediction application. Here, given the user communities of the future,
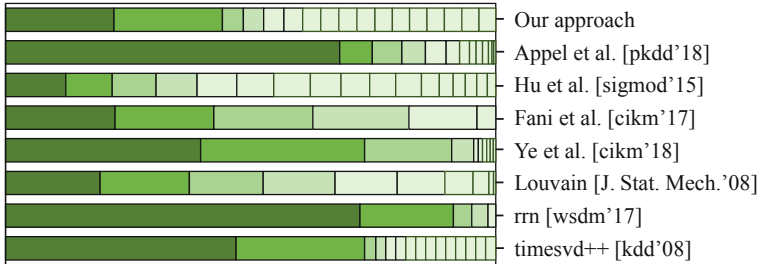
Fig. 2. User distribution in communities. Our method leads to a higher number of communities with a proportional distribution of users in the communities while the baseline methods have a higher skewness. Disproportionate distribution of users in communities can lead to poor application-level performance.

the goal is to predict which users will post news article a at time $T + 1$. To do so, we consider members of the *closest* community to a news article in terms of topics of interest at time $T + 1$ to be the potential posters. We use precision, recall, and f-measure to report user prediction performance. We further compare our method at its best which happens to be at $d = 80$, against the baselines at their best setting in Table 1. In terms of precision, our proposed method was able to outperform other baselines. In terms of *recall*; however, some of the baselines could achieve higher performance and our method is not as strong. The reason for such high recall for some baselines is the fact that the they cluster users into very few, yet large user communities, as seen in Fig. 2. For instance, rrn was able to excel in recall due to its low number of communities. In an extreme, if a method only identifies one community that includes all of the users, recall would be 1. As such the lower the number of the communities is, the higher the recall would be. However, this comes at the cost of precision. Overall, the f-measure metric points to higher quality communities identified based on our proposed work. This reinforces the fact that when users' pairwise similarity with respect to the topics of interest over time are explicitly embedded in a sequence of graphs, it will lead to higher quality user communities in the future. Further, Fig. 2 shows that unlike some of the baselines where the majority of the users are placed in only a few communities and the other communities only have a few members (leading to higher recall but poor performance on precision), our approach could proportionally distribute users across different communities and hence show superior performance over precision and f-measure.

6   Conclusion and Future Work

Our work is among the first to explore the idea of *predicting* topical user communities on social networks. We learn to represent users within a latent space that preserves users' topical similarities over time. Our experiments show that our approach is able to predict communities of like-minded users with respect

to topics of interest in future yet-to-be-observed time interval and outperform the state of the art. The area that we would like to work on in our future work pertains to the fact that our approach penalizes significant and sudden changes in the position of a user's representations in latent space. In other words, our approach favors smooth transition of user representations across different time intervals. However, there may be cases where sudden change in the position of the user representation in latent space may be warranted such as in reaction to bursty topics. As future work, we plan to generalize our approach to support for such cases based on intuitions from Deng et al. [9].

References

1. Akbari, M., Chua, T.: Leveraging behavioral factorization and prior knowledge for community discovery and profiling. In: Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, WSDM 2017, Cambridge, United Kingdom, 6–10 February 2017, pp. 71–79 (2017)
2. Appel, A.P., Cunha, R.L.F., Aggarwal, C.C., Terakado, M.M.: Temporally evolving community detection and prediction in content-centric networks. In: Berlingerio, M., Bonchi, F., Gärtner, T., Hurley, N., Ifrim, G. (eds.) ECML PKDD 2018. LNCS (LNAI), vol. 11052, pp. 3–18. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-10928-8_1
3. Arabzadeh, N., Fani, H., Zarrinkalam, F., Navivala, A., Bagheri, E.: Causal dependencies for future interest prediction on Twitter. In: Proceedings of the 27th ACM International Conference on Information and Knowledge Management, CIKM 2018, Torino, Italy, 22–26 October 2018, pp. 1511–1514 (2018)
4. Barbieri, N., Bonchi, F., Manco, G.: Efficient methods for influence-based network-oblivious community detection. ACM TIST **8**(2), 32:1–32:31 (2017)
5. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet allocation. J. Mach. Learn. Res. **3**, 993–1022 (2003)
6. Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. J. Stat. Mech: Theory Exp. **2008**(10), P10008 (2008)
7. Cao, J., Wang, H., Jin, D., Dang, J.: Combination of links and node contents for community discovery using a graph regularization approach. Future Gener. Comput. Syst. **91**, 361–370 (2019)
8. Chakraborty, T., Cui, Z., Park, N.: Metadata vs. ground-truth: a myth behind the evolution of community detection methods. In: Companion of the The Web Conference 2018 on The Web Conference 2018, WWW 2018, Lyon, France, 23–27 April 2018, pp. 45–46 (2018)
9. Deng, D., Shahabi, C., Demiryurek, U., Zhu, L., Yu, R., Liu, Y.: Latent space model for road networks to predict time-varying traffic. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016, pp. 1525–1534 (2016)
10. Diehl, C.P., Namata, G., Getoor, L.: Relationship identification for social network discovery. In: Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence, Vancouver, British Columbia, Canada, 22–26 July 2007, pp. 546–552 (2007)
11. Dunlavy, D.M., Kolda, T.G., Acar, E.: Temporal link prediction using matrix and tensor factorizations. TKDD **5**(2), 10:1–10:27 (2011)

12. Fani, H., Bagheri, E., Du, W.: Temporally like-minded user community identification through neural embeddings. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM 2017, Singapore, 06–10 November 2017, pp. 577–586 (2017)

13. Milani Fard, A., Bagheri, E., Wang, K.: Relationship prediction in dynamic heterogeneous information networks. In: Azzopardi, L., Stein, B., Fuhr, N., Mayr, P., Hauff, C., Hiemstra, D. (eds.) ECIR 2019. LNCS, vol. 11437, pp. 19–34. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-15712-8_2

14. Goyal, P., Chhetri, S.R., Canedo, A.: dyngraph2vec: capturing network dynamics using dynamic graph representation learning. Knowl.-Based Syst. (2019)

15. Grover, A., Leskovec, J.: node2vec: scalable feature learning for networks. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016, pp. 855–864 (2016)

16. He, D., Liu, D., Jin, D., Zhang, W.: A stochastic model for detecting heterogeneous link communities in complex networks. In: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, Austin, Texas, USA, 25–30 January 2015, pp. 130–136 (2015)

17. Hu, Z., Yao, J., Cui, B.: User group oriented temporal dynamics exploration. In: Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, Québec City, Québec, Canada, 27–31 July 2014, pp. 66–72 (2014)

18. Hu, Z., Yao, J., Cui, B., Xing, E.P.: Community level diffusion extraction. In: Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, Melbourne, Victoria, Australia, 31 May–4 June 2015, pp. 1555–1569 (2015)

19. Jin, D., Chen, Z., He, D., Zhang, W.: Modeling with node degree preservation can accurately find communities. In: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, Austin, Texas, USA, 25–30 January 2015, pp. 160–167 (2015)

20. Koren, Y.: Collaborative filtering with temporal dynamics. Commun. ACM **53**(4), 89–97 (2010)

21. Labatut, V., Dugué, N., Perez, A.: Identifying the community roles of social capitalists in the Twitter network. In: 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2014, Beijing, China, 17–20 August 2014, pp. 371–374 (2014)

22. Li, C., Cheung, W.K., Ye, Y., Zhang, X., Chu, D., Li, X.: The author-topic-community model for author interest profiling and community discovery. Knowl. Inf. Syst. **44**(2), 359–383 (2015)

23. McPherson, M., Smith-Lovin, L., Cook, J.M.: Birds of a feather: homophily in social networks. Ann. Rev. Sociol. **27**(1), 415–444 (2001)

24. Myers, S.A., Leskovec, J.: The bursty dynamics of the Twitter information network. In: Proceedings of the 23rd International Conference on World Wide Web, WWW 2014, pp. 913–924. ACM, New York (2014)

25. Rotta, R., Noack, A.: Multilevel local search algorithms for modularity clustering. ACM J. Exp. Algorithmics **16** (2011)

26. Sarkar, P., Moore, A.W.: Dynamic social network analysis using latent space models. In: Advances in Neural Information Processing Systems 18 [Neural Information Processing Systems, NIPS 2005, Vancouver, British Columbia, Canada, 5–8 December 2005], pp. 1145–1152 (2005)

27. Singer, U., Guy, I., Radinsky, K.: Node embedding over temporal graphs. In: Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, 10–16 August 2019, pp. 4605–4612 (2019)

28. Snijders, T.A., Lomi, A.: Beyond homophily: Incorporating actor variables in statistical network models. Netw. Sci. **7**(1), 1–19 (2019)
29. Wu, C., Ahmed, A., Beutel, A., Smola, A.J., Jing, H.: Recurrent recommender networks. In: Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, WSDM 2017, Cambridge, United Kingdom, 6–10 February 2017, pp. 495–503 (2017)
30. Yan, X., Guo, J., Lan, Y., Cheng, X.: A biterm topic model for short texts. In: 22nd International World Wide Web Conference, WWW 2013, Rio de Janeiro, Brazil, 13–17 May 2013, pp. 1445–1456 (2013)
31. Yang, L., Cao, X., He, D., Wang, C., Wang, X., Zhang, W.: Modularity based community detection with deep learning. In: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI 2016, New York, NY, USA, 9–15 July 2016, pp. 2252–2258 (2016)
32. Yang, L., Cao, X., Jin, D., Wang, X., Meng, D.: A unified semi-supervised community detection framework using latent space graph regularization. IEEE Trans. Cybern. **45**(11), 2585–2598 (2015)
33. Ye, F., Chen, C., Zheng, Z.: Deep autoencoder-like nonnegative matrix factorization for community detection. In: Proceedings of the 27th ACM International Conference on Information and Knowledge Management, CIKM 2018, Torino, Italy, 22–26 October 2018, pp. 1393–1402 (2018)
34. Zhu, L., Guo, D., Yin, J., Steeg, G.V., Galstyan, A.: Scalable temporal latent space inference for link prediction in dynamic social networks. IEEE Trans. Knowl. Data Eng. **28**(10), 2765–2777 (2016)
35. Zuo, Y., Liu, G., Lin, H., Guo, J., Hu, X., Wu, J.: Embedding temporal network via neighborhood formation. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018, London, UK, 19–23 August 2018, pp. 2857–2866 (2018)
36. Zuo, Y., Zhao, J., Xu, K.: Word network topic model: a simple but general solution for short and imbalanced texts. Knowl. Inf. Syst. **48**(2), 379–398 (2015). https://doi.org/10.1007/s10115-015-0882-z