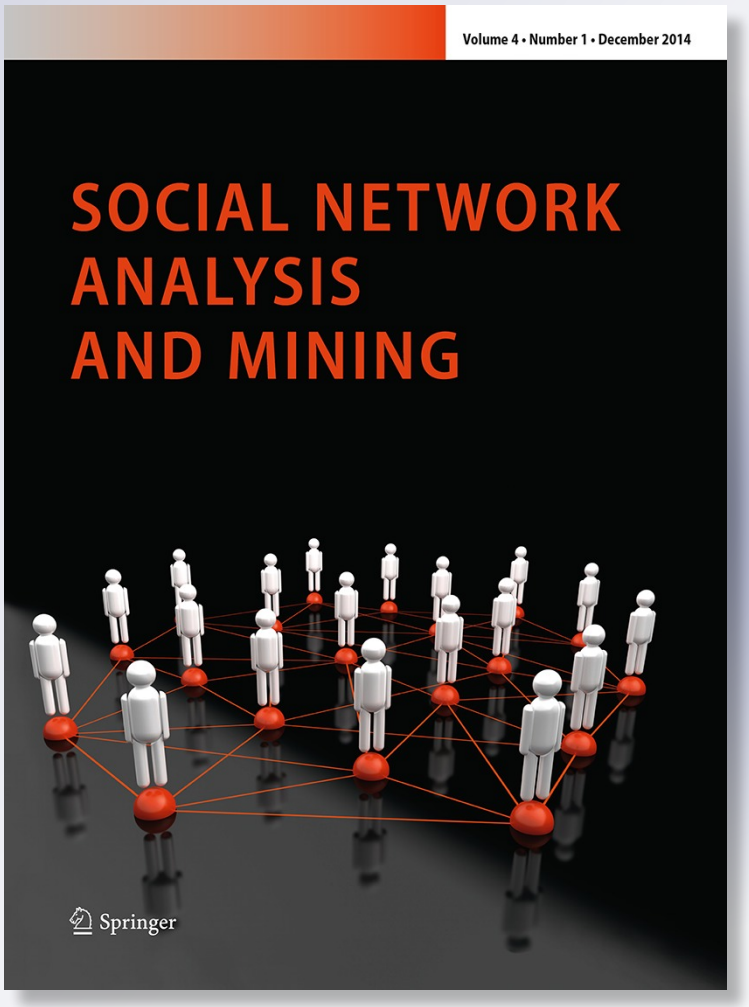
Early detection of persistent topics in social networks

Shota Saito, Ryota Tomioka & Kenji Yamanishi

Social Network Analysis and Mining

ISSN 1869-5450
Volume 5
Number 1

Soc. Netw. Anal. Min. (2015) 5:1-15
DOI 10.1007/s13278-015-0257-1



Springer

Soc. Netw. Anal. Min. (2015)5:19
DOI 10.1007/s13278-015-0257-1

ORIGINAL ARTICLE

Early detection of persistent topics in social networks

Shota Saito[1] · **Ryota Tomioka**[2] · **Kenji Yamanishi**[1,3]

Received: 10 December 2014 / Revised: 6 May 2015 / Accepted: 13 May 2015

Abstract In social networking services (SNSs), persistent topics are extremely rare and valuable. In this paper, we propose an algorithm for the detection of persistent topics in SNSs based on Topic Graph. A topic graph is a subgraph of the ordinary social network graph that consists of the users who shared a certain topic up to some time point. Based on the assumption that the time evolutions of the topic graphs associated with persistent and non-persistent topics are different, we propose to detect persistent topics by performing anomaly detection on the feature values extracted from the time evolution of the topic graph. For anomaly detection, we use principal component analysis to capture the subspace spanned by normal (non-persistent) topics. We demonstrate our technique on a real dataset we gathered from Twitter and show that it performs significantly better than a baseline method based on power-law curve fitting, the linear influence model, ridge regression, and Support Vector Machine.

Keywords Social networks · Information diffusion · Anomaly detection · Principal component analysis · Complex networks · Topic graph

✉ Shota Saito
ssaito@sat.t.u-tokyo.ac.jp

[1] Graduate School of Information Science and Technology, The University of Tokyo, Tokyo, Japan

[2] Toyota Technological Institute at Chicago, Chicago, Illinois, USA

[3] CREST, JST, Tokyo, Japan

1 Introduction

Various human behaviors are highly influenced by social networks (Christakis and Fowler 2008; Watts and Strogatz 1998). In particular, online social networking services (SNSs), such as Facebook and Twitter, are increasing their roles in our daily life (Purcell et al. 2010). Hence, SNSs have been studied from many perspectives (Bakshy et al. 2012; Boyd and Ellison 2007; Cha et al. 2010; Lerman and Ghosh 2010).

Previous studies in data mining have addressed the issue of detecting emerging topics or trends from social network streams(Allan et al. 1998a, b; Cataldi et al. 2010; Kleinberg 2002; Takahashi et al. 2011). Here the main concern is the speed or earliness of the detection. However, we may argue that the value of a topic that bursts for a day or two and then fades out is questionable. On the other hand, a topic that receives continuous interest may be considered as a valuable topic. For example, if we can detect such a topic before it becomes obvious, we can start an action before everyone else.

In this paper, we aim to detect topics that receive continuous attention, which we call persistent topics, as early as possible. This is a challenging task, because it appears that by definition, a long period of observation would be necessary to decide if a topic is persistent or not.

To this end, we leverage the rich graphical structure among the people who shared a topic and build a model that predicts whether a topic becomes persistent in the future. Here persistency of a topic is measured by a quantity we call amplification factor; see Fig. 1. Note that although the amplification factor can only be calculated after a relatively long period of time (say 50 days), our model allows us to make quantitative prediction from a short (say 10 days), but richer, sequence of observation.

Published online: 29 May 2015

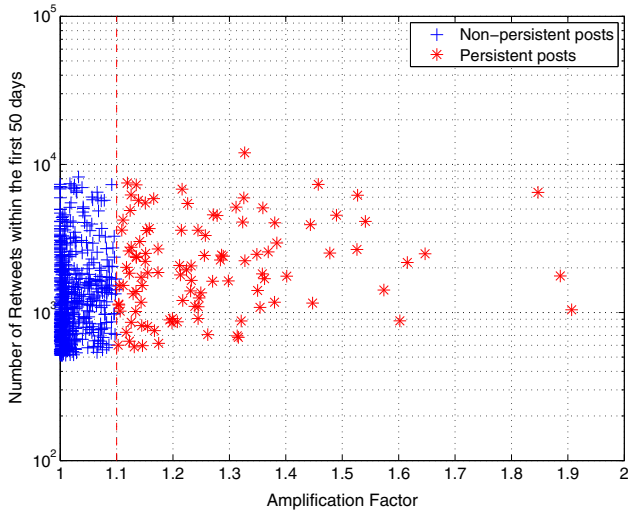


Fig. 1 The number of retweets within the first 50 days n_{50} against the amplification factor n_{50}/n_{10}, i.e., the number of retweets within the first 50 days n_{50} divided by that within the first 10 days n_{10}. The "persistent" posts having amplification factor larger than $\theta_{\text{amp}} = 1.1$ are marked by *red asterisks*. Note that we can only get this picture after 50 days and our goal is to detect persistent posts as early as possible. We also remark that the choice of threshold $\theta_{\text{amp}} = 1.1$ (shown as the *vertical dashed line*) is not essential; see Sects. 3.3 and 5.3.3 for more discussion

More precisely, we collect data from Twitter and aim to predict whether the amplification factor exceeds a predetermined threshold based on the connectivity of the users who shared (retweeted) a topic. The connectivity of the users who shared a topic is modeled as a time-varying directed graph, which we call a topic graph. Here a topic graph is a subgraph of the ordinary follower/followee graph that consists of the users who shared a certain topic up to some time point (see Fig. 3). Since the number of nodes of a topic graph equals the number of people who shared the topic up to a time point, our goal is to show that the proposed method, which looks at the global graphical structure, performs better than simply extrapolating the number of people who shared the topic.

Assuming that the temporal profiles of the topic graphs associated with persistent and non-persistent topics are different (see Fig. 4), we formulate the problem as an anomaly detection problem over the time series of feature vectors that we extract from the topic graph. We apply a principal component analysis (PCA)-based anomaly detection method (Lakhina et al. 2004) on various features we derive from the time series of topic graphs.

Our experiments show that our approach is significantly better than a power-law curve fitting approach and the linear influence model (Yang and Leskovec 2010), and supervised methods, ridge regression and Support Vector Machine. The power-law curve fitting is related to our definition of ground truth based on the amplification factor; see Sect. 3.3. The linear influence model proposed by Yang and Leskovec (2010) predicts the number of retweets as a superposition of non-negative influence functions. Ridge regression and Support Vector Machine are supervised methods, and the experiments on these methods are conducted based upon the same assumption as proposed method, that is the temporal profiles of topic graph associated with persistent and non-persistent topics are different.

Our contribution can be summarized as follows:

- Instead of conventionally studied bursty or trending topics, we focus on topics that receive continuous attention over long period of time, which we call persistent topics.
- The proposed method considers the global temporal/graphical structure of the topic graph. Thus it is agnostic to the language or content of the posts.
- We empirically validate our assumption that the time evolution of graphical structure of the people who share a persistent topic (topic graph, see Fig. 3) is different from a non-persistent topic by the experiment on proposed method and supervised methods.

The preliminary version of this article appeared in the proceedings of ASONAM 2014 (Saito et al. 2014). This extended and revised version uses two supervised methods, ridge regression and Support Vector Machine based upon the topic graph assumption, as comparison methods. As we have a ground truth on the label of either persistent or non-persistent topic, we could try supervised method while the methods in the preliminary version have not been repeated here. We show that our proposed method outperforms these two supervised methods as well as the power-law curve fitting and LIM. Given these supervised methods work although the performance is lower than the proposed method, the results of these two supervised methods support the topic graph assumption. Moreover, we give more discussion for the experiments, namely for the normal and anomalous subspace taken by PCA. Furthermore, the present version contains more illustrative examples and figures, and more details of our proposed method.

The remainder of this paper is organized as follows. In Sect. 2, we give an overview of related work. In Sect. 3, we explain the data we use in our paper and give a criterion to distinguish between a persistent topic and a non-persistent topic. We present our method in Sect. 4. In Sect. 5, we empirically compare our technique to the existing ones and also discuss the effect of feature combination. The concluding remarks are given in Sect. 6.

2 Related work

In this section, we review earlier studies that are related to our paper.

Detecting topics in sequential data is studied in the area of topic detection and tracking (TDT) (Allan et al. 1998a). Allan et al. (1998b) analyzed new topics from news sentences. Kleinberg (2002) studied the bursty structure in the time series of intervals when some pieces of information arrive. Emerging events or topics were studied intensively in the context of SNS based on the textural context (Cataldi et al. 2010; Phuvipadawat and Murata 2010; Sakaki et al. 2010) and also based on the graphical structure induced by the users' mentioning behavior (Takahashi et al. 2011). In our view, the above studies are mainly focused on the emergence or bursts of topics and not on persistent topics.

Information diffusion in SNSs is another well-developed research area. Trusov et al. (2009) modeled diffusion in SNSs and applied to viral marketing.

The works of Asur et al. (2011) and Wang and Huberman (2011) are very similar to our work in terms of their motivation. However, their work is more explanatory than predictive; in particular, they do not aim to predict if a post is going to be persistent or not.

Cha et al. (2010) compares three different measures of influence in Twitter: number of followers, number of retweets, and number of mentions. They revealed that an influencer in Twitter is not always the most followed user. Bakshy et al. (2012) found that social networks' weak ties are important in information spreading in Facebook. Further studies modeled information diffusion process in Twitter. Kwak et al. (2010) modeled the diffusion process of the retweets, which they called retweet tree. Bakshy et al. (2011) made a model for the diffusion of URLs, which they call information cascade. The idea of information cascade is similar to our idea of a topic graph, because both approaches consider networks associated with a topic. However, Bakshy et al. simplified their networks to tree structures while our present work does not. Furthermore, they do not analyze the time evolution of their graph structure.

Analysis of non-stationary time series with principal component analysis (PCA) is known in meteorology (Preisendorfer and Mobley 1988) and the analysis of electroencephalography (Donchin and Heffley 1978); however, it seems to be rather rare in data mining. Note that due to the non-stationarity of the dynamics of topic graphs, popular techniques that assume stationarity, such as, auto-regressive modeling, cannot be applied here.

3 Data

This section provides an overview of the data we collected from Twitter. In Sect. 3.1, we provide a brief overview of Twitter and in Sect. 3.2 we explain how our dataset is constructed.

3.1 Twitter

To study persistent topics in social networking sites, we analyze data from the microblogging service Twitter, which is an extremely popular social networking service, consisting of over 100 million users. This service has a directed social network, where each user can choose to subscribe certain other users if they wish to follow. Twitter users can post a message about any topic within 140 characters called a tweet.

Twitter provides a function called retweet so that users can share any tweet by other users with their followers. Retweet is a key feature that spreads a topic over Twitter.

3.2 Collecting data

We collected data using Twitter API and from a third party service Favstar. Twitter API enables us to crawl and collect data efficiently. We also used Favstar to get a full list of users who retweeted a particular post.

We randomly selected topics that are retweeted by over 500 users that have passed at least 50 days after the original post from the trending topics listed by Favstar and toptweets offered by Twitter officially.

Next, for each topic, we obtained a list of people who retweeted this topic from Favstar and using Twitter API, extracted their followers and the time they retweeted it. In this way, we obtain link information with time stamps, which enables us to define the topic graph. For some of the retweets that we cannot get the exact time stamps, we assumed that the intervals between retweets are regular and used a linear interpolation. Furthermore, we removed a few users whose retweet time was missing due to their privacy setting.

Using this procedure for data collection, we obtained 698 topics retweeted by about 1.6 million users over the period of July 2010–May 2012.

Although follower/followee relationship on Twitter is a directed relation, we ignored the directions in this work, because we are only interested in the topological features of an information diffusion process.

In addition, although users can retweet any post regardless of whether they follow the sender of the post or not, we ignore this case since Twitter API is unable to track this, and it is also a rather rare event.

3.3 Amplification factor

In order to define whether a topic is persistent, i.e., going on for a long period of time, we look at a quantity we call amplification factor. To be more concrete, we have examined 698 posts from Twitter that have been retweeted (shared) at least 500 times. Fig. 1 plots the number of retweets (shares) within the first 50 days n_{50} against the amplification factor n_{50}/n_{10}, i.e., the number of retweets within the first 50 days divided by that within the first 10 days. We can clearly see that there is a dense concentration of posts around amplification factor almost one (shown as blue crosses), that is, although they received many retweets in the first ten days, after that they are not retweeted anymore. On the other hand, there are certain fractions of posts on the right side of the plot that have large amplification factors (shown as red asterisks), that is, although these posts do not necessarily receive much attention in the beginning, they grow steadily in the number of retweets. Note that the total number of retweets shown on the vertical axes does not discriminate these two sets of points well; although they receive roughly the same number of retweets, they have quite different temporal profiles. We remark that we define persistent or not without any human judgement. Nevertheless, our definition of persistency involves a parameter (threshold) which may seem arbitrary. However, the choice of threshold $\theta_{\text{amp}} = 1.1$ (shown as the vertical dashed line in Fig. 1) is not essential; see Sect. 5.3.3 for more discussion.

Examples of persistent tweets in Fig. 2 illustrate that the persistent topics are not like ongoing issues, which can be outdated after certain time. Persistent topics are more like long-run social marketing campaign, emerging new trends or time-invariant valuable aphorism, which can be valuable pieces of information even after a long time. An example, in this case, is a social ad campaign ran by Dropbox, Inc. The persistency of the post can be considered as an indication of the success of the campaign. Another example of persistent topics we found is like an aphorism related to information technology and innovation, which are emerging topics posted in Japanese by an engineer. The tweet is about an insight when the author of the tweets worked as an intern at Apple. Persistency of this post can be considered as a sign of an emergence of an opinion leader. We note that persistent topics are not only in English, but also in other languages. Also we would like our method to be agnostic to the language of the post.

The motivation of using amplification factor is from a power-law model

$$\Delta_t = \beta t^{-\alpha}, \tag{1}$$

where Δ_t is the number of retweets that a post receives in the tth time interval, and α and β are positive parameters (see Fig. 6 for an illustrative example). Integrating the above model, we have

$$n_t \propto \begin{cases} t^{1-\alpha} & (\text{if} \quad 0<\alpha<1), \\ \log(t) & (\text{if} \quad \alpha = 1), \\ 1 - t^{-\alpha+1} & (\text{if} \quad \alpha > 1), \end{cases}$$

where the interval of integration is taken as $[1, t]$ for $\alpha \geq 1$. Then the amplification factor is written as follows:

$$\frac{n_{t_2}}{n_{t_1}} = \begin{cases} \left(\frac{t_2}{t_1}\right)^{1-\alpha} & (\text{if} \quad 0<\alpha<1), \\ \frac{\log(t_2)}{\log(t_1)} & (\text{if} \quad \alpha = 1), \\ \frac{1 - t_2^{-\alpha+1}}{1 - t_1^{-\alpha+1}} & (\text{if} \quad \alpha > 1). \end{cases} \tag{2}$$

From the above expressions, we can see that if $\alpha > 1$, the cumulative number of retweets will asymptotically reach a constant, and the amplification factor will be close to one for sufficiently large t_1 and t_2. On the other hand, if $0<\alpha<1$, the cumulative number of retweets will continuously grow, in fact, the amplification factor becomes strictly larger than one for any $t_2 > t_1$.

The power-law model (1) suggests a simple method for detecting a persistent post, that is, we estimate the exponent α from a short, say 10 days of, observation and then use the

Fig. 2 Examples of persistent tweets in English and Japanese. Persistent posts are typically not related to any ongoing issues, which can be outdated after several time elapsed from the post. They look more like what can be valuable even after a long time elapsed, such as long-run marketing campaign, emerging new trends or time-invariantly valuable aphorism. These persistent posts can be challenging to detect from their textual contents. We would also like our method to be agnostic to the language of the post

value of the estimated α as the criterion for persistency (small α corresponds to high amplification factor); see Sect. 5.1.3 for details.

The motivation of using amplification factor is also from Pareto's law, which says that a significant portion of something occurs in the first 20 % of time (Newman 2005).

4 Proposed method

In this section, our proposed algorithm is presented. Our algorithm consists of three components: a topic graph, feature extraction from a topic graph, and anomaly detection. These three components are described in the following three subsections.

4.1 Topic graph

The topic graph concerning a certain topic is a subgraph of the original social network graph that consists of nodes that correspond to the user who posted the original post, which we call the topic origin, and other users who retweeted the post. The edges of the topic graph correspond to the friendship relation of the underlying social network, which we assume to be symmetric and stationary.

More precisely, let $G = (V, E)$ be the original social network graph, where $V = \{v_1, \ldots, v_n\}$ is the set of nodes and $(v_i, v_j) \in E$ if and only if there is an edge between node v_i and v_j. Let S be the set of topics and T be the set of time points. The topic graph $G^{(s,t)}$ concerning a certain topic $s \in S$ at time $t \in T$ can be written as $G^{(s,t)} = \left(V^{(s,t)}, E^{(s,t)}\right)$, where $V^{(s,t)}$ is the set of nodes

$$V^{(s,t)} = \left\{ v_0^{(s,t)}, v_1^{(s,t)}, \ldots, v_{n_t^{(s)}}^{(s,t)} \right\},$$

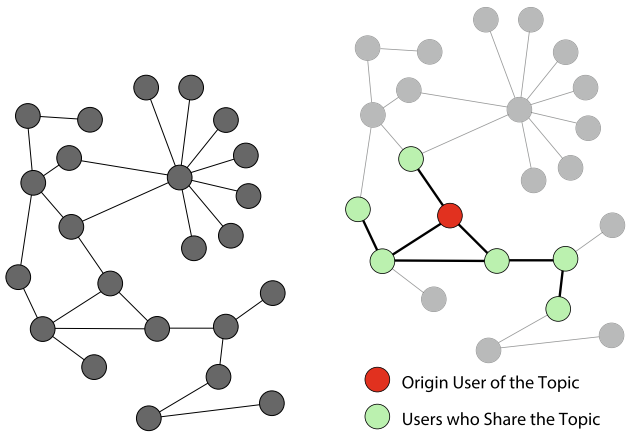
where $v_0^{(s,t)}$ is the *topic origin*, or the node that corresponds to the user who posted the original post s, and the nodes $v_1^{(s,t)}, \ldots, v_{n_t^{(s)}}^{(s,t)}$ correspond to the $n_t^{(s)}$ users who retweeted the post up to time t. $E^{(s,t)}$ is the subset of edges such that $(v_i, v_j) \in E^{(s,t)}$ if and only if $v_i, v_j \in V^{(s,t)}$ and $(v_i, v_j) \in E$.

An example of a topic graph is illustrated in Fig. 3b.

We also define the adjacency matrix and the degree matrix of the topic graph $\left(V^{(s,t)}, E^{(s,t)}\right)$ as follows. Adjacency matrix $A^{(s,t)}$ is a $\left(n_t^{(s)} + 1\right) \times \left(n_t^{(s)} + 1\right)$ matrix and is defined as

$$A_{ij}^{(s,t)} = A_{ji}^{(s,t)} = \begin{cases} 1 & \left(\text{if} \left(v_i^{(s,t)}, v_j^{(s,t)}\right) \in E^{(s,t)}\right), \\ 0 & (\text{otherwise}). \end{cases} \tag{3}$$

The *degree matrix* is defined as



(a) a relationship networks on SNS

(b) a topic graph, we propose in this paper

Fig. 3 Previous work focuses on the networks composed of friendships; see panel (**a**). To study the time evolution of networks associated with a certain topic, we focus on the notion of topic graph. A topic graph is a subgraph of the ordinary social network graph that consists of the users who shared a certain topic; see panel (**b**). Details are described in Sect. 4.1

$$D_{ij}^{(s,t)} = \begin{cases} d\left(v_i^{(s,t)}\right) & (\text{if i} = \text{j}), \\ 0 & (\text{otherwise}), \end{cases} \tag{4}$$

where $d\left(v_i^{(s,t)}\right)$ is the *degree* of node $v_i^{(s,t)}$, given by

$$d\left(v_i^{(s,t)}\right) = \sum_j A_{ij}.$$

4.2 Features of topic graph

We assume that the temporal profiles of the topic graphs (shown in Fig. 4) associated with persistent and non-persistent topics are different. To track the temporal profile of topic graph, we extract the various features from the topic graph on each time and compose a vector to represent temporal profiles of one sequence of topic graphs made of one topic.

In this subsection, we describe five feature values we use to characterize the topic graph, namely, the number of times the topic is shared, the number of communities in the topic graph, the maximum distance from the initial user, and the two eigenvalues (second smallest and the largest) of the graph Laplacian. These features include both local characters and global characters of the topic graph.

Concatenating the five feature values for m time points, a topic is represented by the $5m$ dimensional feature vector

$$\begin{aligned} \mathbf{y}^{(s)} = \Big(& n_{t_1}^{(s)}, \ldots, n_{t_m}^{(s)}, N_{\text{com}}^{(s,t_1)}, \ldots, N_{\text{com}}^{(s,t_m)}, \\ & l^{(s,t_1)}, \ldots, l^{(s,t_m)}, \lambda_2^{(s,t_1)}, \ldots, \lambda_2^{(s,t_m)}, \\ & \lambda_{\max}^{(s,t_1)}, \ldots, \lambda_{\max}^{(s,t_m)} \Big)^{\top}, \end{aligned} \tag{5}$$

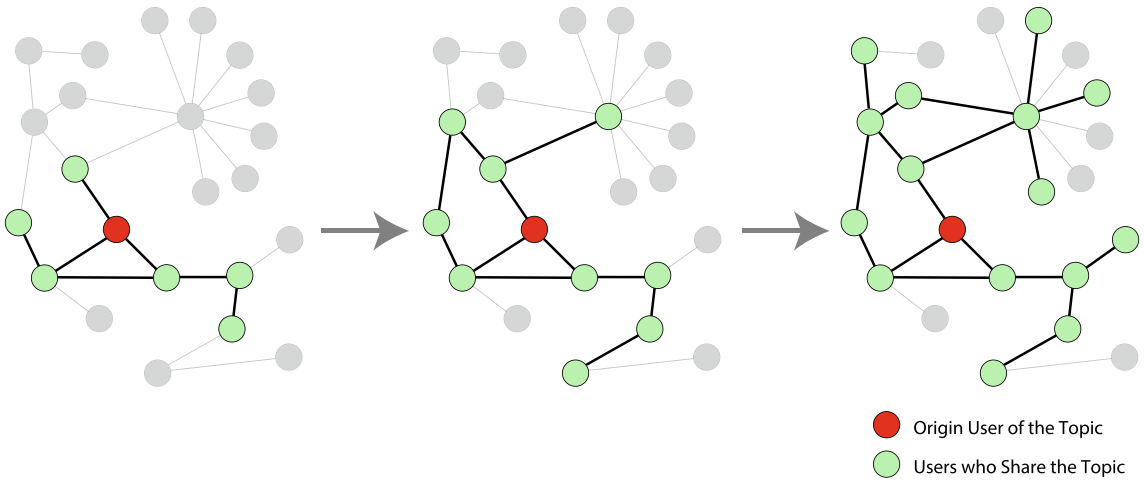


Fig. 4 The illustration of time evolution of topic graph. For the topic graph shown in Fig. 3b, we hypothesize that the temporal profiles of the topic graphs associated with persistent and non-persistent topics are different

where $t_1, \ldots, t_m$ are regularly sampled time points. When the length of the observed period is 10 days and the interval is 30 minutes, m is 240. See below for the precise definition of the feature values. We also denote by the $N_S \times 5m$ matrix Y the matrix obtained by concatenating the feature vectors for all topics S along rows, where $N_S = |S|$ is the number of topics we use for training.

4.2.1 Number of nodes $n_t^{(s)}$

As in the previous section, $n_t^{(s)}$ denotes the number of people who shared a topic $s \in S$ by time $t \in T$. In other words, $n_t^{(s)}$ is the number of nodes in topic graph $G^{(s,t)}$.

4.2.2 Number of communities $N_{\text{com}}^{(s,t)}$

The distribution of edges of a naturally occurring network is not only globally, but also locally inhomogeneous, with high concentrations of edges within special groups of vertices, and low concentrations between these groups (Fortunato 2010). This feature of real networks is called community structure. We compute the number of communities $N_{\text{com}}^{(s,t)}$ using Fast Modularity algorithm (Newman 2004), which costs $O((|E| + |V|)|V|)$ for each graph.

4.2.3 Maximum distance from the topic origin $l^{(s,t)}$

The maximum distance $l^{(s,t)}$ from the topic origin $v_0^{(s,t)}$ is defined as

$$l^{(s,t)} = \max_{v_j^{(s,t)}} d_{G^{(s,t)}}\left(v_j^{(s,t)}, v_0^{(s,t)}\right),$$

where $d_{G^{(s,t)}}(u, v)$ is the distance along the shortest path between node u and node v on topic graph $G^{(s,t)}$. The maximum distance tells how far at most a certain topic is distributed. The maximum distance can be calculated by solving the well-known single-source shortest path problem, which we solve using the Dijkstra method (Cormen 2001; Dijkstra 1959), which requires $O(|E| + |V| \log |V|)$ for each graph.

4.2.4 Eigenvalues of graph Laplacian

Let a graph Laplacian $L^{(s,t)}$ be

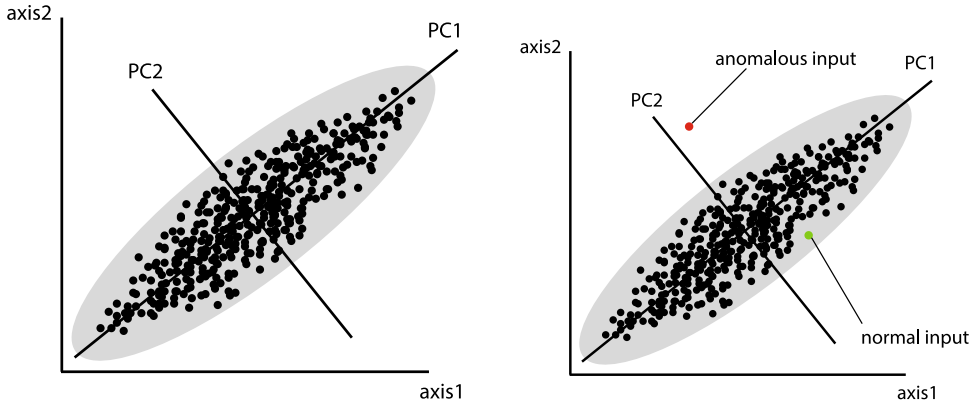
$$L^{(s,t)} = D^{(s,t)} - A^{(s,t)}, \tag{6}$$

where $A^{(s,t)}$ is the adjacency matrix (3) and $D^{(s,t)}$ is the degree matrix (4) of topic graph $G^{(s,t)}$. Note that the smallest eigenvalue of graph laplacian is always 0. To compute eigenvalues, the QR method, which is classical, costs $O(|V|^3)$ for each graph.

Since we want same dimension vector for all topics, we use only the second smallest eigenvalue $\lambda_2^{(s,t)}$ and the largest eigenvalue $\lambda_{\max}^{(s,t)}$ rather than use all eigenvalues. The second smallest eigenvalue $\lambda_2^{(s,t)}$ represents a density of graph, and the largest eigenvalue $\lambda_{\max}^{(s,t)}$ is used in calculating the number of communities (Kim and Motter 2007; Newman 2006). We refer to the literature by Von Luxburg (2007) for more details of graph Laplacian.

4.3 Anomaly detection via principle component analysis

This section presents an anomaly detection method based on the principle component analysis (PCA) (Pearson 1901; Bishop 2007) proposed by Lakhina et al. (2004). The basic



(a) The illustration of PCA **(b)** The illustration of anomaly detection via PCA

Fig. 5 The illustration of PCA and anomaly detection via PCA. PCA converts possibly correlated basis to uncorrelated basis. Basis is chosen to maximize the variance, see (**a**, **b**) illustrates how anomaly detection via PCA works. If the data shown in green are given, these should be the normal data, as the projection onto PC2, which takes less variance of the original data, is small. On the other hand, data shown in red can be regarded as an anomalous input since the projection onto PC2 is large. More detail is given in Sect. 4.3

idea of this method is to use PCA to define normal subspace and abnormal subspace; the anomaly score is given by the projected variance on the anomaly subspace.

Figure 5 provides intuitive illustration how anomaly detection via PCA works. As shown in Fig. 5a, PCA converts possibly correlated basis, axis 1 and axis 2, to uncorrelated one, PC1 and PC2. As PC1 extracts variance of the data more than PC2, it is reasonable to assume that PC1 is a normal subspace and PC2 is an anomalous subspace. In this setting, if we observe two more data shown in Fig. 5a, data drawn in green seem to be normal input as the projection of the data onto the anomalous subspace PC2 is small, whereas data drown in red can be recognized as an anomalous input from the observation that projection of these data onto the anomalous subspace PC2 is large.

Based on this intuitive idea, anomaly detection via PCA is formalized as follows. Define Y as $N_{com} \times 5m$ matrix of non-persistent topics, and is given by

$$Y = \left(\mathbf{y}^{(s_1)}, \mathbf{y}^{(s_2)}, \cdots, \mathbf{y}^{(s_{N_{non}})}\right)^\top, \tag{7}$$

where N_{non} is a number of non-persistent topics. A first principal component $v_1 \in \mathbb{R}^{5m}$ is taken to maximize the variance of Y's projection onto the component. Hence, Y is given by

$$\mathbf{v}_1 = \underset{\|\mathbf{v}\|=1}{\operatorname{argman}} \|Y\mathbf{v}\|. \tag{8}$$

Let $v_j \in \mathbb{R}^{5m}$ be further principal components. The kth component is taken to extract the maximum variance of the space, that is the original space Y subtracted by the first $k-1$ principal components:

$$\mathbf{v}_k = \arg\max_{\|\mathbf{v}\|=1} \left\| \left(Y - \sum_{i=1}^{k-1} Y\mathbf{v}_i\mathbf{v}_i^\top \right)\mathbf{v} \right\| \tag{9}$$

Let C_k be the fraction of variance explained up to the kth principal component as follows:

$$C_k = \frac{\sum_{j=1}^{k} \sigma_j^2}{\sum_{j=1}^{n} \sigma_j^2}. \tag{10}$$

Let

$$P = (\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k),$$

where the number of components k is chosen such that $C_k > \delta$.

We define the decomposition of a feature vector $\mathbf{y}$ into the normal part $\hat{\mathbf{y}}$ and the abnormal part $\tilde{\mathbf{y}}$ as follows:

$$\mathbf{y} = \hat{\mathbf{y}} + \tilde{\mathbf{y}}, \tag{11}$$

where

$$\hat{\mathbf{y}} = PP^\top\mathbf{y}, \tilde{\mathbf{y}} = (I - PP^\top)\mathbf{y}$$

A useful statistic for detecting the abnormal part $\tilde{\mathbf{y}}$ is the squared prediction error (SPE):

$$\mathrm{SPE} \equiv \|\tilde{\mathbf{y}}\|^2 = \|(\mathrm{I} - \mathrm{PP}^\top)\mathbf{y}\|^2,$$

and we consider a topic to be anomalous, or persistent, if

$$\mathrm{SPE} > \delta_{\mathrm{PCA}}. \tag{12}$$

5 Experiments

In this section, we present the result of applying our proposed method to Twitter dataset we described in Sect. 3, and compare our method to a baseline method based on the power-law curve fitting, and the linear influence model (Yang and Leskovec 2010), and two supervised methods, that are ridge regression and Support

Vector Machine. In addition, we discuss the effect of feature combination, qualitative difference between persistent and non-persistent topics, sensitivity to the definition of the threshold θ_{amp}, contribution of axes taken by PCA to persistent and non-persistent topics, and overfitting of supervised methods.

5.1 Experimental setup

5.1.1 Objective

The goal in this experiment is to predict whether a topic is persistent or not (defined by the amplification factor n_{50}/n_{10} being greater or less than 1.1), only looking at the data up to $T = 1$, 3, 5, or 10 days. Since we have 589 non-persistent topics and 109 persistent topics (see Figs. 1 and 2), we randomly left out 109 non-persistent topics for testing and used the remaining $N_S = 480$ non-persistent topics for training. For two supervised methods, we randomly left out 55 persistent topics and 294 non-persistent topics for the test data, and the remaining 54 persistent topics and 295 non-persistent topics are used for the training data. We used the area under the receiver operator curve (AUC) as the performance criterion. To compute AUC, we compute the area under the receiver operating characteristic curve, which is drawn by plotting true-positive rate against the false-positive rate at various thresholds. We remark that when the AUC of a classifier is 1.0, it performs the best, and AUC is 0.5 with the random guess setting. Note that we make sure that the number of persistent topics and non-persistent topics for testing is the same for proposed method, power-law fitting, and LIM. The random split was repeated 200 times and the AUC scores were averaged. The performance of all methods we compare depends on the sampling interval τ_s; we report their performance for $\tau_s = 1$, 3, 6, and 12 h for the proposed approach and power-law curve fitting, and $\tau_s = 12$ and 24 h for the linear influence model; note that the number of time points $m = 24T/\tau_s$ in Eq. (5).

5.1.2 Proposed approach.

We used $\delta = 0.9999$ for determining the number of PCA components in Eq. (10). AUC was computed by changing the threshold parameter δ_{PCA} in Eq. (12).

5.1.3 Power-law curve fitting.

For comparison, we employed power-law curve fitting to the difference sequence $\Delta_t^{(s)}$ of the number of users who shared a certain topic.

We assume that the number of retweets $\Delta_t^{(s)}$ that a post receives in the tth time interval follows the power-law model (1). In order to estimate the coefficients α and β, we used the linear least squares method

$$\left(\hat{\alpha}^{(s)}, \hat{\beta}^{(s)}\right) = \underset{\alpha,\beta}{\operatorname{argmin}} \sum_{t=1}^{m} \left(\log \Delta_t^{(s)} + \alpha \log t - \log \beta\right)^2. \tag{13}$$

See Fig. 6 for an illustration. We replaced zero entries in $\Delta_t^{(s)}$ by $\Delta_t^{(s)} = 10^{-2}$ to avoid the log diverging to infinity.

Since low value of the exponent α indicates high persistency, we used the value $-\hat{\alpha}^{(s)}$ as the anomaly score and computed the AUC.

Note again that the notion of power law is behind the criterion for persistent topics (see Sect. 3.3). We also remark that power-law method can be considered as an unsupervised method.

5.1.4 Linear influence model (LIM)

The linear influence model (LIM) proposed by Yang and Leskovec (2010) can be used to predict the number of future retweets as a superposition of positive influence functions that corresponds to the users who retweeted the original post in the past.

More specifically, we learn the influence functions for K_{lim} key users from N_S training topics. This is a non-negative least squares problem and can be minimized by the MATLAB function `lsqnonneg`. Once we have the influence functions, we can make a prediction into the future. Let $I_u(t)$ be the influence function of user u. Then the cumulative number $\hat{n}_t^{(s)}$ of retweets at time t can be predicted as follows:

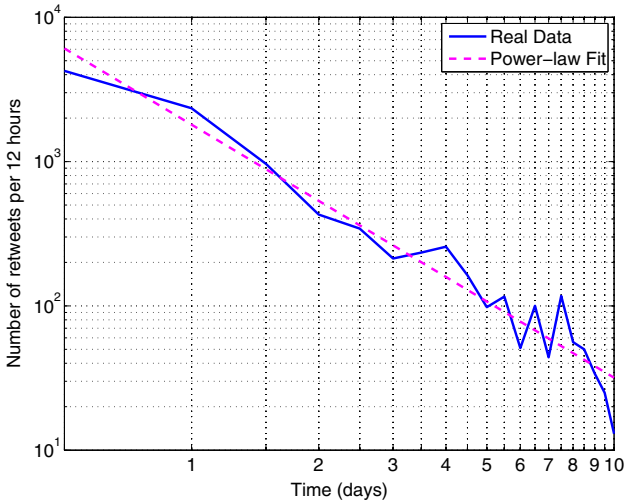


Fig. 6 An illustration of the power-law curve fitting. See Sect. 3.3 for the model and Sect. 5.1 for the estimation procedure

$$\hat{n}_t^{(s)} = \sum_{u:t_u \leq T} I_u(t - t_u),$$

where t_u is the time that user u retweeted topic s ($t_u = \infty$ if she/he did not retweet).

We use the predicted amplification factor $\hat{n}_{50}^{(s)}/\hat{n}_{10}^{(s)}$ as the anomaly score and computed the AUC.

We used the following parameters for LIM: the number of key users $K_{\text{lim}} = 22$, which was the number of users who retweeted at least 100 posts out of the 698 topics we analyzed, the length of the influence function $L_{\text{lim}} = 24 \cdot 50/\tau_s$ (i.e., 50 days), the length of training data $T_{\text{lim}} = 24 \cdot 100/\tau_s$ (i.e., 100 days). The training data and test data are split in the same way as other two methods. The numbers L_{lim} and T_{lim} depend on the sampling interval τ_s, and, in this case, we used $\tau_s = 12$ and 24 h. We did not compute LIM for τ_s smaller than 12 h, because it was extremely time consuming and we did not expect the performance for smaller τ_s to be better than the larger two settings. Note that the LIM always observes training data up to 100 days, whereas the proposed method only uses the training data up to $T(\leq 10)$ days.

5.1.5 Ridge Regression

Ridge regression is like least squares method, but favors more sparseness in coefficients. Given a training example (x_i, y_i), where $x_i \in \mathbb{R}^p$, ridge regression learns a linear function

$$f(\mathbf{x}^*) = \mathbf{w}^\top \mathbf{x}$$

that predicts the output y^*, and favors $\mathbf{w}$ to be shrunk towards zero at the same time. This problem can be formulated as

$$w = \underset{\mathbf{w}}{\text{argmin}} \, \|\mathbf{y} - X\mathbf{w}\| + \lambda\|\mathbf{w}\|, \tag{14}$$

where $\mathbf{y} = (y_1, y_2, \ldots, y_n)^\top$ and $X = (\mathbf{x}_1^\top, \mathbf{x}_2^\top, \ldots, \mathbf{x}_n^\top)$. The first term of the right-hand side of Eq. (14) corresponds to training error, and the second term is called sparseness term, gaining sparseness of $\mathbf{w}$. Taking gradient of Eq. (14) yields a solution

$$\mathbf{w} = (X^\top X + \lambda E_p)^{-1} X^\top \mathbf{y}, \tag{15}$$

where E_p is a $p \times p$ identity matrix. Note that when $\lambda = 0$, Eq. (14) is the same as the optimization formulation of least square.

In the experiment, we compose X with both non-persistent and persistent topics, and y is either 1 or -1, the label on either non-persistent or persistent topic, respectively. In order to track the sensitivity to the coefficient of sparseness term $\lambda = 10^l$, we applied our method for $l = -3, -2, \ldots, 3$, and use the best one.

5.1.6 Support vector machine

Support vector machine (SVM) chooses the decision hyperplane to be the one maximizing the margin, and also allows some of the training points to be misclassified with penalty (Boser et al. 1992; Vapnik 1998). In the following section, we restrict SVM to binary classification, since our problem has only two labels, non-persistent or persistent. With a sparsity term, this problem can be written as

$$f = \underset{f \in \mathcal{H}}{\text{argmin}} \frac{1}{n} \sum (1 - y_i f(\mathbf{x}_i))_+ + \lambda \|f\|_{\mathcal{H}}, \tag{16}$$

where $\mathcal{H}$ is a hypothesis space of functions, and $(a)_+ = \max(a, 0)$. Given the fact that f can be written down as

$$f(\cdot) = \sum_i^n c_i K(\cdot, \mathbf{x}_i),$$

where K is a kernel function, by introducing slackness variables ξ_i, we can rewrite Eq. (16) to

$$\begin{aligned} &\underset{c \in \mathbb{R}^n \xi \in \mathbb{R}^n}{\text{argmin}} \frac{1}{n} \sum (1 - y_i f(\mathbf{x}_i))_+ + \lambda c^\top K c, \\ &\text{s.t. } \xi_i \geq 1 - \sum_{j=1}^n c_j K(\mathbf{x_i}, \mathbf{x_j}) \\ &\quad \xi_i \geq 0. \end{aligned} \tag{17}$$

This problem can be readily solved if we solve the dual problem of Eq. (17).

We used two kernel function, one is linear kernel,

$$K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i \cdot \mathbf{x}_j \tag{18}$$

and the other is Gaussian kernel,

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\sigma \|\mathbf{x}_i - \mathbf{x}_j\|^2), \tag{19}$$

with $\sigma = 0.25, 0.5$, and 0.75. Like the setting in ridge regression, we compose X with both non-persistent and persistent topics, and y is either 1 or -1, labeling the topic. We move the coefficient of sparseness term $\lambda = 10^l$ as $l = -3, -2, \ldots, 3$, and use the best one.

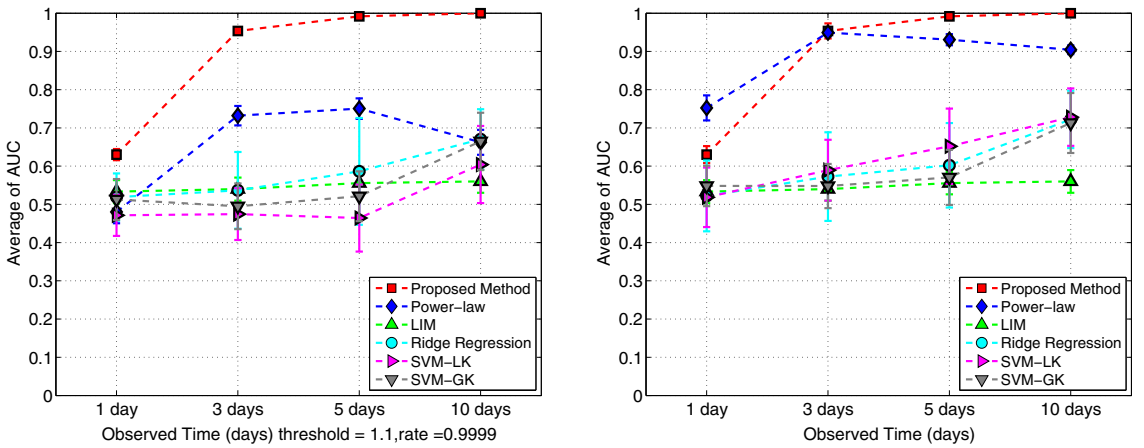
5.2 Results

The results are summarized in Table 1 and Fig. 7. We can see that the proposed method can indeed detect persistent topics in less than 10 days with high AUC, and it clearly outperforms the other two methods that do not take the graph structure into account and the two supervised methods that assume a form of decision hyperplane, for observation time $T \geq 3$ days.

Table 1 The average AUC for the proposed topic graph-based anomaly detection, power-law curve fitting, and linear influence model (Yang and Leskovec 2010)

	1 day	3 days	5 days	10 days
Proposed (interval: 1 h)	0.6302	**0.9536**	**0.9919**	**0.9998**
Proposed (interval: 3 h)	0.5538	0.7587	0.9257	0.9922
Proposed(interval: 6 h)	0.5169	0.6730	0.8655	0.9703
Proposed(interval: 12 h)	0.5119	0.5830	0.8374	0.9533
Power-law (interval: 1 h)	0.4801	0.7320	0.7504	0.6622
Power-law (interval: 3 h)	0.6449	0.7570	0.8477	0.8348
Power-law (interval: 6 h)	0.7345	0.8553	0.8561	0.8966
Power-law (interval: 12 h)	**0.7523**	0.9493	0.9307	0.9043
LIM (interval: 12 h)	0.4979	0.4942	0.4998	0.4945
LIM (interval: 24 h)	0.5333	0.5339	0.5556	0.5600
Ridge regression (interval: 1 h)	0.51903	0.53661	0.58631	0.71641
Ridge regression (interval: 3 h	0.50571	0.54014	0.58859	0.67954
Ridge regression (interval: 6 h)	0.51769	0.54364	0.59133	0.6849
Ridge regression (interval: 12 h)	0.52217	0.57263	0.60222	0.72208
SVM with linear kernel (interval: 1 h)	0.51575	0.53477	0.52748	0.6662
SVM with linear kernel (interval: 3 h)	0.50274	0.53701	0.56991	0.67345
SVM with linear kernel (interval: 6 h)	0.51538	0.58931	0.62697	0.70667
SVM with linear kernel (interval: 12 h)	0.51841	0.58545	0.65189	0.72812
SVM with Gaussian kernel $\sigma = 0.25$ (interval: 1 h)	0.5127	0.47374	0.50294	0.62227
SVM with Gaussian kernel $\sigma = 0.25$ (interval: 3 h)	0.53104	0.51862	0.5426	0.63474
SVM with Gaussian kernel $\sigma = 0.25$ (interval: 6 h)	0.54815	0.54797	0.5429	0.62953
SVM with Gaussian kernel $\sigma = 0.25$ (interval: 12 h)	0.5038	0.52217	0.55919	0.6623
SVM with Gaussian kernel $\sigma = 0.5$ (interval: 1 h)	0.51116	0.48368	0.51544	0.65601
SVM with Gaussian kernel $\sigma = 0.5$ (interval: 3 h)	0.51384	0.50961	0.53642	0.6562
SVM with Gaussian kernel $\sigma = 0.5$ (interval: 6 h)	0.5288	0.5289	0.53622	0.65982
SVM with Gaussian kernel $\sigma = 0.5$ (interval: 12 h)	0.50917	0.52145	0.56665	0.7029
SVM with Gaussian kernel $\sigma = 0.75$ (interval: 1 h)	0.50997	0.49501	0.52125	0.66494
SVM with Gaussian kernel $\sigma = 0.75$ (interval: 3 h)	0.5086	0.50297	0.53689	0.65379
SVM with Gaussian kernel $\sigma = 0.75$ (interval: 6 h)	0.53328	0.51829	0.55637	0.67225
SVM with Gaussian kernel $\sigma = 0.75$ (interval: 12 h)	0.48013	0.53493	0.5706	0.71281

See Sect. 5.1 for the details



(a) The AUC for sampling interval $\tau_s = 1$ hour.

(b) The AUC for the best sampling interval τ_s for each method.

Fig. 7 The performances for the detection of persistent topics measured in AUC of the proposed method (*square*), power-law (*diamond*), LIM (*upward-triangle*), ridge regression (*circle*), SVM with linear kernel (*right-pointing triangle*), and SVM with Gaussian kernel are (*downward-triangle*) shown. For SVM with Gaussian kernel, we chose σ with the best performance. The *error bars* show the standard deviation of 200 random splitting around the mean

Power-law curve fitting performed well in some case (e.g., $T = 3$) but it was more sensitive to the choice of the sampling interval τ_s than the proposed method. In fact, when the sampling interval is too short, the power-law model does not fit well to the sequence $\Delta_t^{(s)}$, because $\Delta_t^{(s)} = 0$ for many ts in such case.

LIM takes the influence of key users into account. However, it did not perform even as good as the simple power-law method. Note that in the original paper, the maximum length of an influence function was a day, whereas we are using the length of 50 days. Thus it might be fair to say that this is not the best setting to use the method.

We can observe that the result of the proposed method is better than that of the two supervised methods. Note that these supervised methods are based upon the same assumption that the temporal profiles of the topic graphs associated with persistent and non-persistent topics are different. The reason of this performance is supposedly because our proposed method does not assume any forms of a decision hyperplane whereas two supervised methods assume kernel functions or linear decision hyperplane. From the perspective of boundary, the PCA model implies that it forms a decision boundary as a surface of the normal subspace composed by principal components and their variance. Moreover, the result of the proposed method shows that the hyperplane as a surface of the normal subspace PCA learns works as a decision boundary. This suggests that there may be a boundary, which classifies well-persistent and non-persistent ones, and therefore we may be able to improve the performance of supervised methods, namely for the selection of proper kernel functions for SVM.

In terms of the run time, the power-law curve fitting was the fastest; it also requires no training. The proposed method was faster than the LIM, ridge regression, and SVM. Note that the proposed method mainly consumes run time in computing the topic graph-based features, whereas the LIM spends time in solving the non-negative least squares problem, ridge regression and SVM computes various sparseness coefficients λ, and SVM has to solve the more costly optimization problem, that is the dual problem of Eq. (17). Therefore, there is room for improving the efficiency of the proposed method by only computing the features that contribute sufficiently to the performance.

5.3 Discussion

5.3.1 Effect of feature combination

We study the effect of feature combination in this subsection. We ask if one of the five features in Eq. (5) could do the job of all of them. The results are presented in Fig. 8

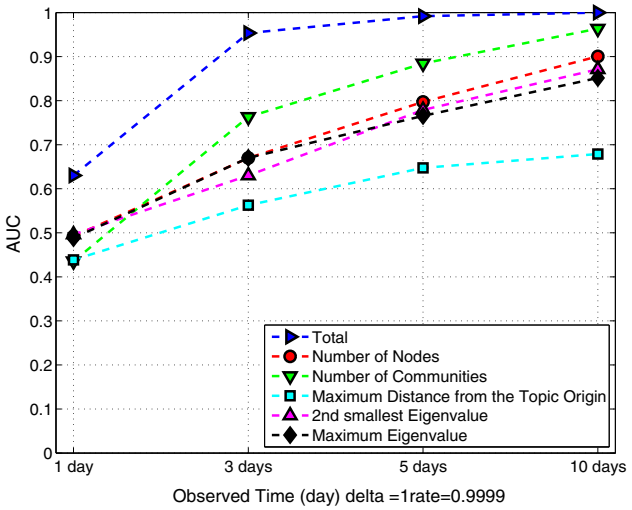


Fig. 8 Average AUC using only one of the features. Sampling interval $\tau_s = 1$ h

The blue curve in Fig. 8 shows the performance of the proposed method using all the features. The other five curves show the performance of the proposed method using only one of the features.

We can see that while there are clearly some features, like the number of communities, that come close to the feature combination in Eq. (5), other features, like the maximum distance, that performs poorly on its own. Using all the features seems to boost and stabilize the performance.

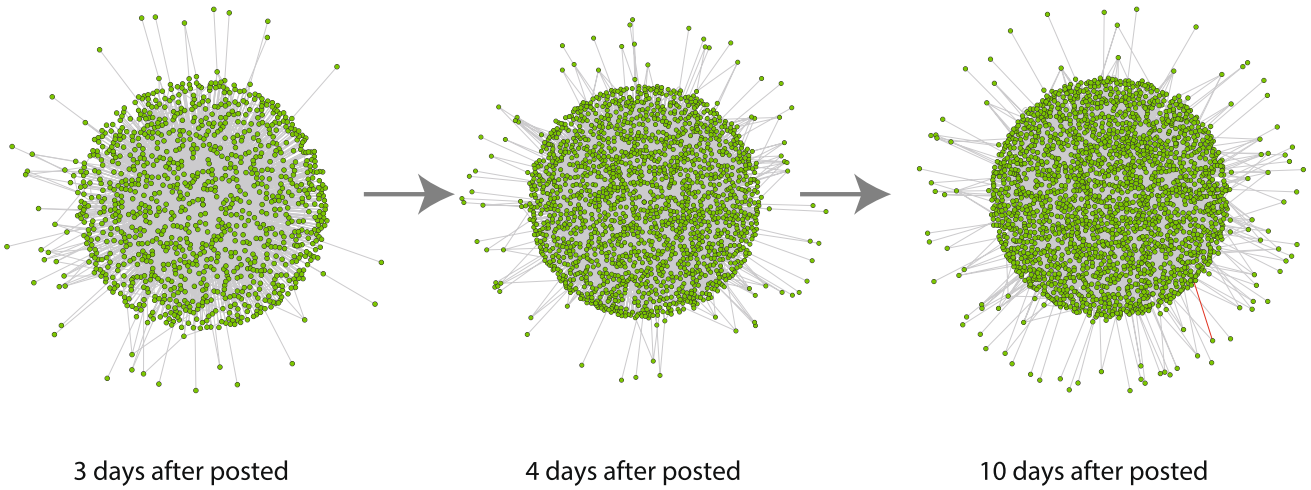
The above results support our strategy to incorporate as many features of the topic graph as possible. In addition, it seems that we need not worry having redundant features, because PCA can detect correlation in the features.

5.3.2 Difference between persistent and non-persistent posts

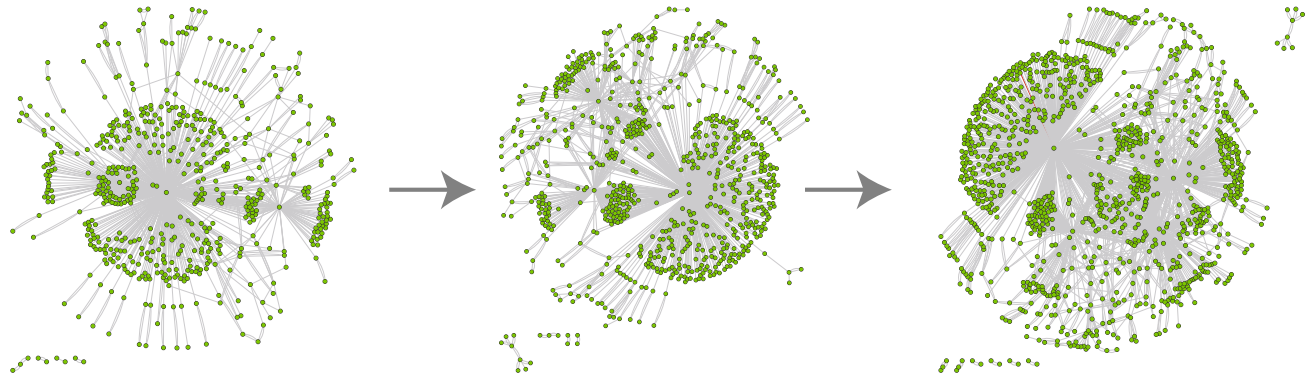
Fig. 9 compares the dynamics of topic graphs of a typical persistent post and a non-persistent post. We can see that the topic graph of a persistent topic is more tightly connected and grows denser and denser as time progresses, which means that the number of communities does not grow too big. On the other hand, the topic graph of a non-persistent topic is relatively loosely connected and consists of several communities. In addition, it relies more on hub users who are connecting different communities.

5.3.3 Sensitivity to the Definition of the Ground Truth

Fig. 10 shows the same plot as in Fig. 7a but a different definition of the amplification factor based on the threshold $\theta_{\text{amp}} = 1.25$ for the proposed method, power law, and LIM. From the plot, we can see that the general trend is



(a) Persistent post.



(b) Non-persistent post.

Fig. 9 Comparison of topic graph dynamics of a persistent post and a non-persistent post found by the proposed method

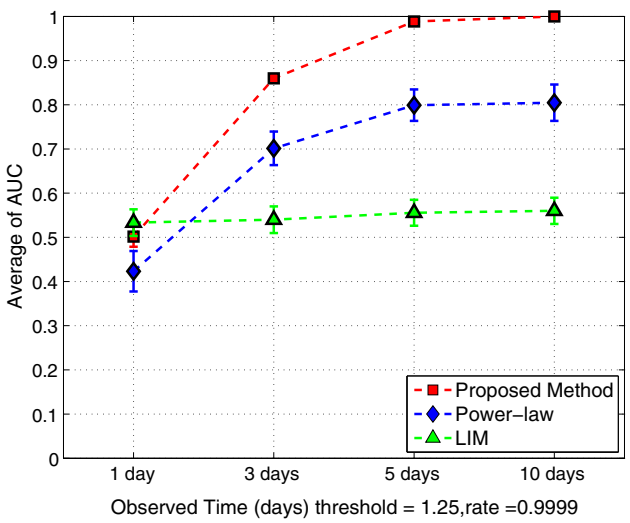


Fig. 10 The AUC for sampling interval $\tau_s = 1$ h when $\theta_{\text{amp}} = 1.25$ is used as the definition of the ground truth; see also Fig. 1 and Sect. 3.3

unchanged. Therefore, the proposed method does not rely on a particular choice of threshold parameter θ_{amp}. It would be an interesting work in the future to extend the current approach to predict the amplification factor in a regression setting.

5.3.4 *Anomalous subspace*

We also study the contribution of axes converted by PCA to the persistent and non-persistent topics. The contribution of principal component k to the topic s can be written as

$$c_k^{(s)} = \frac{\|u_k * \mathbf{y}^{(s)}\|}{\|\mathbf{y}^{(s)}\|}. \tag{20}$$

Figure 11 compares the contribution of principal components to a non-persistent topic with the ones to a persistent topic. The red dashed line shows where $C_k > \delta = 0.9999$, meaning that left-hand side of the red line is normal

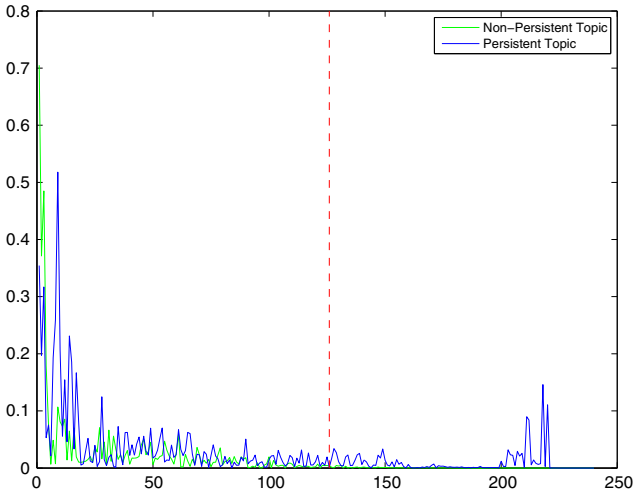
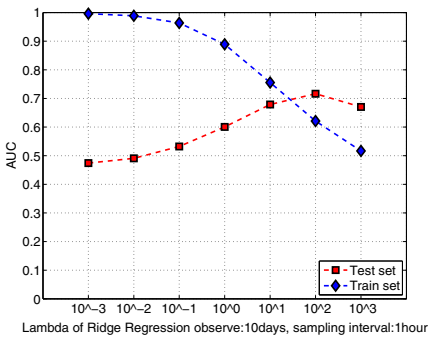


Fig. 11 The comparison of the principal components' contribution to non-persistent and persistent topics. Contribution of kth principal component is defined as $\|u_k * \mathbf{y}^{(s)}\|/\|\mathbf{y}^{(s)}\|$. The *red dashed line* shows where $C_k > \delta = 0.9999$, meaning that left-hand side of the line is normal and right-hand side is anomalous. We can see that there are large contribution of principal components in anomalous subspace (*right-hand side*) for a persistent topic, while there is little contribution in anomalous subspace for non-persistent one. This result implies that PCA composes normal subspace by non-persistent topics, which can distinguish non-persistent and persistent

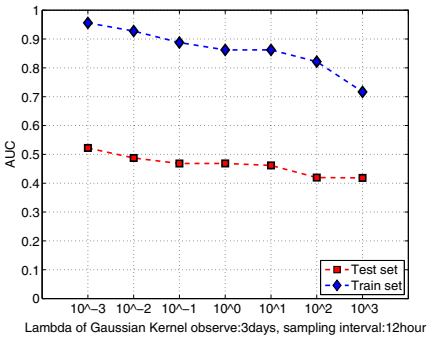
subspace and right-hand side of the red line is anomalous subspace. We can recognize that the persistent topic has large values of contribution to the anomalous subspace, while non-persistent does not. This result supports our assumption that persistent topics have large projection onto the anomalous subspace composed by subtracting principal components of non-persistent ones from the original space.

5.3.5 *Overfitting for supervised methods*

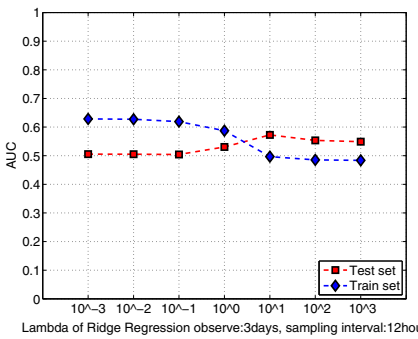
Figure 12a shows the result of ridge regression with observation $T = 10$ days and sampling interval $\tau = 1$ h. We can recognize that AUC for test data improves as the coefficient of the sparseness term λ gains, while AUC for training data falls from 1. This result implies that in ridge regression learning of the labels of the dynamics of topic graph occurs as the overfitting of the classifier is suppressed by gaining λ. Moreover, we can say that this result supports the assumption on the proposed method that dynamics of a topic graph of a non-persistent post is different from that of the persistent one, since these supervised methods are based on the assumption similar to that of our method. However, we should point out that for Gaussian kernel with observing $T = 3$ days and sampling interval $\tau = 1$ h shown in Fig. 12b, AUC for test data keeps around 0.5 although AUC for training data decreases from 1 as λ gains. Furthermore, for ridge regression with observing $T = 3$ days and sampling interval $\tau = 12$ shown in Fig. 12c, AUC with all λ is around 0.5 even for training data, meaning that the hyperplane learnt by these methods fails to classify even for training data as same accuracy as random guess. These results suggest that some of supervised methods may work poorly if we have few data. One reason may be the number of persistent topics is so small compared to the number of non-persistent topics, we could easily overfit to the type of persistent topics we have in our



(a) Ridge regression with observing 10 days and sampling interval 1 hour.



(b) Support Vector Machine with Gaussian kernel $sigma = 0.25$



(c) Ridge regression with observing 3 days and sampling interval 12 hour.

Fig. 12 The Averac and training data. **a** We observe that AUC for test data improves as the coefficient of the sparseness term λ gains and the AUC for training data decrease. This result implies that learning occurs as the overfitting of the classifier is suppressed by gaining λ. **b** Although hyperplane learnt by SVM suppresses overfitting given AUC for SVM with Gaussian kernel with $\sigma = 0.25$ to training data decreases from 1 as we gain the sparseness coefficient λ, the AUC to test data does not gain. **c** The AUC for ridge regression with observing 3 days 12 h is around 0.5 for test data and training data, meaning that hyperplane learnt by ridge regression does not classify well. The results **b** and **c** imply that a supervised method may poorly work if we have few data

training set and would not be suitable to detect unseen type of persistent topics.

6 Conclusion

In this paper, we have proposed a method for early detection of *persistent* topics, which are topics that keep receiving people's attention for a long period of time. The proposed method is based on the notion of topic graph. A topic graph is a dynamically growing subgraph of the social network and consists of nodes that correspond to users who participated in the topic by sharing it with their friends or followers. We have proposed to extract the time series of five network theoretic features from the topic graph. The feature space is further expanded by considering the whole (non-stationary) sequence of the feature values. An anomaly detection method based on PCA is applied to these expanded features to detect persistent topics.

We have applied the proposed method to Twitter data we collected and have shown that the proposed method can reliably detect persistent topics that goes on for 50 days within the first 5 days from posted. We have also compared our algorithm to a simple baseline method based on the power-law curve fitting, the linear influence model (Yang and Leskovec , 2010), ridge regression, and SVM, and have shown that the proposed method performs consistently better than the other four methods.

There are several future directions. Although we have shown that the five graph-based features of topological structure of topic graph were useful in characterizing a growth pattern of the non-persistent topics, they are by no means exhaustively nor systematically chosen. It would be fruitful to combine our framework with the existing graph mining (Inokuchi and Kashima 2003) and spectral methods (Ide and Kashima 2004; Hirose et al. 2009). It would also be highly valuable to consider the trade-off between the computational cost of including some feature and the performance gain that we obtain by incorporating that feature. In addition, it is worth to try to reveal what real-world phenomena are the keys to make topics persistent or non-persistent. Our method has found the difference in time-sequential change of topological structure of topic graph. In particular, as shown in Sect. 5.3.2 and Fig. 9, we have found that topic graphs of persistent topics are tightly connected than those of the non-persistent topics. Thus, we can say that in this paper, we have shown that these topological phenomena in topic graph make topics persistent. However, we should say that we still do not know what real-world phenomena make time sequence of topic graphs have these topological features of persistent topic. Another future direction would be to try to predict not only using topological information but also using external sources such as other features in Twitter like favorite and other websites than Twitter. Finally, as we mentioned in Sect. 5.3.5, with more data it would also be interesting to try a supervised method.

Acknowledgments This work was partially supported by MEXT KAKENHI 23240019 and JST–CREST. This work was supported by MEXT KAKENHI 23240019 and JST–CREST.

Conflict of interest The authors declare that they have no conflict of interest.

References

Allan J, Carbonell J, Doddington G, Yamron J, Yang Y (1998) Topic detection and tracking pilot study: Final report. Evaluation 1998:194–218

Allan J, Papka R, Lavrenko V (1998b) On-line new event detection and tracking. In: Proceedings of SIGIR, pp 37–45

Asur S, Huberman B, Szabó G, Wang C (2011) Trends in social media: Persistence and decay. In: Proceedings of ICSWM

Bakshy E, Hofman JM, Mason WA, Watts DJ (2011) Everyone's an influencer: quantifying influence on twitter. In: Proceedings of WSDM, pp 65–74

Bakshy E, Rosenn I, Marlow C, Adamic L (2012) The role of social networks in information diffusion. In: Proceedings of WWW, pp 519–528

Bishop CM (2007) Pattern recognition and machine learning. Springer

Boser BE, Guyon IM, Vapnik V (1992) A training algorithm for optimal margin classifiers. In: Proceedings of ACM, COLT, pp 144–152

Boyd D, Ellison N (2007) Social network sites: definition, history, and scholarship. J Comput Mediat Commun 13(1–2):210–230

Cataldi M, Torino U, Caro L, Schifanella C (2010) Emerging topic detection on twitter based on temporal and social terms evaluation. In: Proceedings of MDMKDD, pp 1–10

Cha M, Haddadi H, Benevenuto F, Gummadi K (2010) Measuring user influence in twitter: The million follower fallacy. In: Proceedings of ICWSM, pp 10–17

Christakis N, Fowler J (2008) The Collective Dynamics of Smoking in a Large Social Network. N Eng J Med 358(21):2249–2258

Cormen T (2001) Introduction to algorithms. The MIT press

Dijkstra E (1959) A note on two problems in connexion with graphs. Numerische mathematik 1(1):269–271

Donchin E, Heffley E (1978) Multivariate analysis of event-related potential data: a tutorial review. U.S. Gov, Printing Office

Fortunato S (2010) Community detection in graphs. Phys Rep 486(3–5):75–174

Hirose S, Yamanishi K, Nakata T, Fujimaki R (2009) Network anomaly detection based on eigen equation compression. In: Proceedings of KDD

Ide T, Kashima H (2004) Eigenspace-based anomaly detection in computer systems. In: Proceedings of KDD, pp 440–449

Inokuchi A, Kashima H (2003) Mining significant pairs of patterns from graph structures with class labels. In: Proceedings of ICDM

Kim D, Motter A (2007) Ensemble averageability in network spectra. Phys Rev Lett 98(24):248701

Kleinberg J (2002) Bursty and hierarchical structure in streams. In: Proceedings of KDD, pp 91–101

Kwak H, Lee C, Park H, Moon S (2010) What is twitter, a social network or a news media? In: Proceedings of WWW, pp 591–600

Lakhina A, Crovella M, Diot C (2004) Diagnosing network-wide traffic anomalies. In: Proceedings of SIGCOMM, pp 219–230

Lerman K, Ghosh R (2010) Information contagion: An empirical study of the spread of news on digg and twitter social networks. In: Proceedings of ICWSM

Newman M (2004) Fast algorithm for detecting community structure in networks. Physics Review E 69:066–133

Newman M (2005) Power laws, Pareto distributions and Zipf's law. Contemp Phys 46(5):323–351

Newman M (2006) Modularity and community structure in networks. Proc Natl Acad of Sci USA 103(23):8577

Pearson K (1901) On lines and planes of closest fit to systems of points in space. Phil Mag 2(11):559–572

Phuvipadawat S, Murata T (2010) Breaking news detection and tracking in twitter. In: Proceedings of WICACM, vol 3, pp 120–123

Preisendorfer R, Mobley C (1988) Principal component analysis in meteorology and oceanography. Elsevier, Developments in atmospheric science

Purcell K, Rainie L, Mitchell A, Rosenstiel T, Olmstead K (2010) Understanding the participatory news consumer. Pew Internet and American Life Project 1

Saito S, Tomioka R, Yamanishi K (2014) Early detection of persistent topics in social networks. In: Proceedings of ASONAM, pp 417–424

Sakaki T, Okazaki M, Matsuo Y (2010) Earthquake shakes twitter users: real-time event detection by social sensors. In: Proceedings of WWW, pp 851–860

Takahashi T, Tomioka R, Yamanishi K (2011) Discovering emerging topics in social streams via link anomaly detection. In: Proceedings of ICDM, pp 1230–1235

Trusov M, Bucklin R, Pauwels K (2009) Effects of word-of-mouth versus traditional marketing: Findings from an internet social networking site. J Mark 73(5):90–102

Vapnik V (1998) Statistical learning theory, vol 2. Wiley, New York

Von Luxburg U (2007) A tutorial on spectral clustering. Statistics and computing 17(4):395–416

Wang C, Huberman B (2011) Long trend dynamics in social media. CoRR abs/1109.1852

Watts D, Strogatz S (1998) Collective dynamics of 'small-world' networks. Nature 393(6684):440–442

Yang J, Leskovec J (2010) Modeling information diffusion in implicit networks. In: Proceedings of ICDM, pp 599–608