2-2017

Discovering burst patterns of burst topic in Twitter

Guozhong DONG
Harbin Engineering University

Wu YANG
Harbin Engineering University

Feida ZHU
Singapore Management University, fdzhu@smu.edu.sg

Wei WANG
Harbin Engineering University

Published in Computers and Electrical Engineering, 2016 February, Pages 1-9
http://doi.org/10.1016/j.compeleceng.2016.06.012

Discovering burst patterns of burst topic in twitter

Guozhong Dong[a], Wu Yang[a,*], Feida Zhu[b], Wei Wang[a]

[a] *Information Security Research Center, Harbin Engineering University, Harbin, 150001, China*
[b] *Singapore Management University, Singapore*

ARTICLE INFO

Article history:
Received 24 February 2016
Revised 25 June 2016
Accepted 27 June 2016
Available online xxx

Keywords:
Burst topic
Burst pattern
Hierarchical clustering
Frequent sub-graph mining

ABSTRACT

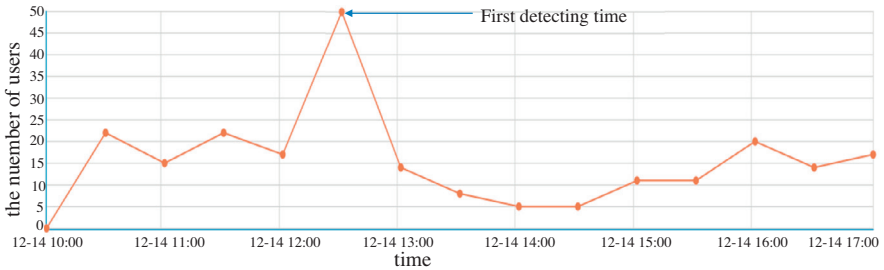
Twitter has become one of largest social networks for users to broadcast burst topics. There have been many studies on how to detect burst topics. However, mining burst patterns in burst topics has not been solved by the existing works. In this paper, we investigate the problem of mining burst patterns of burst topic in Twitter. A burst topic user graph model is proposed, which can represent the topology structure of burst topic propagation across a large number of Twitter users. Based on the model, hierarchical clustering is applied to cluster burst topics and reveal burst patterns from the macro perspective. Frequent sub-graph mining is used to discover the information flow patterns of burst topic from the micro perspective. Experimental results show that several interesting burst patterns are discovered, which can reveal different burst topic clusters and frequent information flows of burst topic.

1. Introduction

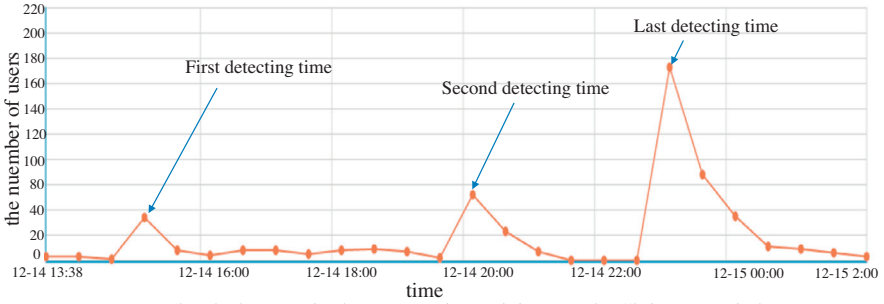
With the development of web 2.0, social media services, such as Twitter, emerge and quickly become popular. Different from traditional news media, Twitter allows users to broadcast short textual messages and express opinions using web-based or mobile-based platforms. When breaking news or events occur, people can post tweets about breaking news and share with friends. Due to large number of people participating in conversation and discussion, some tweets may become hot messages and the source of burst topics. Fig. 1 illustrates the user engagement time series of two burst topics detected by CLEar(Clairaudient Ear) system,[1] in which the arrow denotes each detecting time of burst topic. For example, Fig. 1(a) shows the burst topic that prominent Chinese human rights lawyer Pu Zhiqiang was set to stand trial in Beijing court. The topic was caused after Twitter user (BBCNewsAsia) post a tweet about the event. As shown in Fig. 1(a), the topic had one burst and was detected once by CLEar system in its lifecycle. Fig. 1(b) shows the burst topic about promotion activity over the Christmas period, which had more than one burst in its lifecycle. The different burst patterns raise a question of immense practical value: Can we leverage burst topics detected by CLEar system to discover burst patterns of burst topic in Twitter?

Unfortunately, mining burst patterns in burst topic has not been solved by the existing works. Most prior research works [1-14] focus on detecting burst topics in social media, instead of mining burst patterns in our work. Shen et al. [14] analyze the burst pattern of burst keyword, which can influence the accuracy of burst topics detection. In conclusion, burst pattern of burst topic is an important factor in the studies on burst topic.

* Corresponding author.
E-mail addresses: dongguozhong@hrbeu.edu.cn (G. Dong), yangwu@hrbeu.edu.cn (W. Yang), fdzhu@smu.edu.sg (F. Zhu), w_wei@hrbeu.edu.cn (W. Wang).
[1] http://research.pinnacle.smu.edu.sg/clear/.



(a) The burst topic about Chinese human rights lawyer Pu Zhiqiang



(b) The burst topic about promotion activity over the Christmas period

Fig. 1. Example of user engagement time series of burst topic.

In this paper, we investigate the problem of mining burst patterns of burst topic in Twitter. To solve this problem, we discover burst patterns from both macro and micro perspectives by leveraging a burst topic user graph model. To summarize, the contributions of our work are listed as follows:

(1) We propose a burst topic user graph model which can represent the topology structure of burst topic propagation across a large number of Twitter users. In the burst topic user graph, nodes represent the burst topic users and edges represent the follower/followee relationship between users.
(2) Hierarchical clustering is applied to cluster burst topics and reveal burst patterns from the macro perspective. Combined with extracted 12 topic features, four distinct clusters of burst topic are discovered, which correspond to different burst patterns.
(3) Frequent sub-graph mining is used to discover information flow patterns of burst topic from the micro perspective. Based on the frequent sub-graph mining, several information flow patterns are extracted, which can be applied in several potential applications.

The rest of the paper is organized as follows. Section 2 reviews the related work. Section 3 presents the definition and construction of burst topic user graph. Macro and micro burst pattern mining are presented in Section 4. Section 5 describes the experimental results and findings. Finally, we conclude our work in Section 6.

2. Related work

The study of burst topic in social media and data mining techniques [1–19] have been studied in the last decade. As there are numerous research works focusing on it, here we introduce the ones most related to our work.

Prasadet et al. [1] propose a framework to detect emerging topics through the use of dictionary learning. They determine novel documents in the stream and subsequently identify cluster structure among the novel documents. Agarwal et al. [2] model emerging events detection problem as discovering dense clusters in highly dynamic graphs and exploit short-cycle graph property to find dense clusters efficiently in microblog streams. Alvanaki et al. [3] present the “en Blogue” system for emergent topic detection. En Blogue keeps track of sudden changes in tag correlations and presents tag pairs as emergent topics. Mathioudakiset et al. [4] identify burst keywords and group burst keywords into topics based on their co-occurrences. Cataldi et al. [5] formalize the keyword life cycle leveraging a novel aging theory intended to mine burst keywords and detect burst topics through keyword-based topic graph. Nguyen et al. [6] introduce a novel concept of sentiment burst and employ a stochastic model for detecting bursts in text streams. Takahashi et al. [7] apply a recently proposed change-point detection technique based on Sequentially Discounting Normalized Maximum Likelihood (SDNML) coding to detect abnormal messages and detect the emergence of a new topic from the anomaly measured through the model. Cui et al. [8] study some event-related properties of hashtags, including temporal trends, authorships and pattern of texts. Based on event-related properties of hashtags, they examine the popular hashtags to discover breaking events. Li et al. [9] propose “Twevent” system to detect events in twitter stream which can distinguish the realistic events from the noisy ones. Lee

et al. [10] apply density-based clustering on evolving post network to identify the events. Wang, Liu et al. [11] propose a system called SEA to detect events and conduct panoramic analysis on Weibo events from various aspects. Xie et al. [12–13] present a real-time system to provide burst event detection, popularity prediction, and event summarization. Shen et al. [14] analyze different burst patterns and propose real-time burst topics detection oriented Chinese microblog stream. The method detect burst entities and cluster them to burst topics without requiring Chinese segmentation, which can obtain related messages and users at the same time. In general, while the earlier studies mainly aim at proposing new model and designing systems to detect burst topics. Our work focus on revealing the higher level knowledge of burst topic, that is, discovering burst patterns of burst topic from both macro and micro perspectives.

3. Burst topic user graph model

3.1. Definition of burst topic user graph model

Once a user post a tweet related to the burst topic in Twitter, the tweet can spread to the user's followers, and then followers who are interested in the burst topic may post or retweet the message. In order to represent the topology structure of burst topic propagation across a large number of Twitter users, in the burst topic user graph model, nodes represent the burst topic users and edges represent the follower/followee relationship between users. The burst topic user graph model based on the follower/followee relationship can be formally defined as follows.

Definition 1. Burst Topic User Graph. A burst topic user graph $G_k =< V_k, E_k, l >$ is directed and labeled. In detail, $V_k = \{v_{k_0}, v_{k_1}, \cdots, v_{k_n}, \cdots\}$ is the set of vertices representing Twitter users over the burst topic k. Each Twitter user v_{kn} can be formalized as three tuples: $v_{k_n} = (userId, topicTime, label)$, where *userId* is the user ID of v_{kn}, *topicTime* is the earliest post time of user v_{kn} over burst topic k, *label* is the label of v_{kn}. E_k represents the sets of edges among Twitter users, in which a directed edge $(v_{ki}, v_{kj}) \in E_k$ means that v_{ki} is the follower of v_{kj}. l is the label function that assigns label to users.

First, it is important to consider time information in the burst topic user graph model. Without considering time information, the edges in topic user graph model can't represent the direction of information flow. Secondly, we distinguish the importance of different nodes by labeling nodes in the burst topic user graph with some labels ($L = \{l_1, l_2, \cdots\}$). The label function is introduced using the follower number as the labeling criterion. In this paper, we use four different labels ($L = \{a, b, c, d\}$) to indicate that the follower number of a Twitter user is larger than *10k*, between *1k* and *10k*, between *100* and *1k*, and less than *100*.

3.2. Construction of burst topic user graph model

In this section, we introduce how to collect burst topics and construct burst topic user graph model. In order to detect burst topics in real-time, Xie et al. [12,13] proposed a two-stage integrated solution TopicSketch to detect burst topics as early as possible. In this paper, we collected burst topic dataset from CLEar system which is based on TopicSketch framework. As the relationship among Twitter users may be dynamic and the limit of Twitter platform, it is difficult to obtain the complete diffusion process of burst topic. We formulate the burst topic user graph in Algorithm 1 to get the optimal approximation of the burst topic diffusion path.

Given a tweet list *TL* of burst topic k, the algorithm first sort the tweet in descending order by post time and init burst topic user set V_k. Afterwards, for each tweet in *TL*, we update burst topic user set and label topic users based on labeling function(lines 2–9). Finally, for each topic user in V_k, burst topic user graph are generated based on user relationship and topic time of topic user(lines 10–17).

4. Burst patterns mining

In this section, we attempt to mine burst patterns of burst topic from macro level and micro level perspectives.

4.1. Macro burst pattern mining

Burst topic detection in social media has been a hotspot and difficult research point. It is obviously that the burst topic features vary much from different burst topics. It is important to cluster burst topics based on burst topic features, which can reveal the macro level burst patterns of burst topic. We first present the features used for macro burst pattern mining.

To describe burst topic comprehensively, we extract the features of burst topic from three aspects: the perspective of tweet, the perspective of user and burst features. The extracted features are listed in Table 1.

To characterize the size and type of tweet involved in burst topics, we extract *3* features including the number of tweets (*Tweet Number*), the ratio of retweets to *Tweet Number* (*Retweet Ratio*), as well as the ratio of reply tweets to *Tweet Number* (*Reply Ratio*).

To characterize the size and type of user involved in burst topics, we extract 6 features including the number of users (*User Number*), the number of users who have more than 10k followers (*Big User*). Meanwhile, to measure the user interest in burst topics, we calculate the number of users who post more than one tweet involved in burst topics, and name it as

Algorithm 1 Burst topic user graph construction algorithm.

Input: tweet list for a given burst topic *TL*
Output: burst topic user graph in the pajek [15] format
(1) Sort *TL* using the post time from late to earlier and init user set V_k of burst topic k
(2) **for** each tweet $m \in TL$ **do**
(3) init author of tweet m as v_{kn}
(4) **if** $v_{kn} \notin V_k$ **then**
(5) add v_{kn} into V_k and label v_{kn} using the labeling function l
(6) **else**
(7) update the topic time of v_{kn}
(8) **end if**
(9) **end for**
(10) **for** each topic user $v_{kn} \in V_k$ **do**
(11) get the followee set F_{vkn}
(12) **for** each followee $u_f \in F_{vkn}$ **do**
(13) **if** $uf \in V_k$ **and** $v_{kn}.topicTime > u_f.topicTime$
(14) record the user relationship v_{kn} u_f in the pajek format
(15) **end if**
(16) **end for**
(17) **end for**

Table 1
The features of burst topic.

Features of burst topic	Description
Tweet Number	The number of tweets
Retweet Ratio	The ratio of retweets to *Tweet Number*
Reply Ratio	The ratio of reply tweets to *Tweet Number*
User Number	The number of users
Overlap User	The number of users who post more than one tweet involved in burst topic
Overlap User Ratio	The ratio of *Overlap User* to *User Number*
Big User	The number of users who have more than *10k* followers
Big User Ratio	The ratio of *Big User* to *User Number*
Verified User Ratio	The ratio of verified users to *User Number*
Burst Number	The number of burst
Burst Interval	Average interval of each burst's detecting time
Burst Time Span	The time span between first burst and last burst

Overlap User. Furthermore, we calculate the ratio of *Overlap User* to *User Number* (*Overlap User Ratio*), the ratio of *Big User* to *User Number* (*Big User Ratio*) as well as the ratio of verified users to *User Number* (*Verified User Ratio*).

To characterize the burst features of burst topic, we introduce another 3 features. We define the number of burst in burst topics as *Burst Number*. For burst topic with more than one burst, we define the time span between first burst and last burst as *Burst Time Span*, the average interval of each burst's detecting time as *Burst Interval*.

In total, we have extracted 12 features to describe burst topic, as shown in Table 1. Based on the burst topic features, macro burst pattern can be defined as follows:

Definition 2. Macro burst pattern. Given a burst topic set *BT* and a distance measure based on burst topic features, the macro burst pattern is defined as finding all clusters in *BT*.

In this paper, hierarchical clustering algorithm is adopted to cluster burst topics and mine macro burst pattern, where the Euclidean distance is used as the distance measure. By hierarchical clustering, the correlations between burst topic clusters can be detected.

4.2. Micro burst pattern mining

In this section, frequent pattern mining techniques are used to discover micro burst pattern of burst topic. We first give the definitions of sub-graph, minimum support, and micro burst pattern. Afterwards, we present the micro burst pattern mining algorithm.

Definition 3. Sub-graph. Given two graphs $G =< V, E, l >$ and $G_s =< V_s, E_s, l_s >$, a sub-graph isomorphism of *Gs* to *G* is an injective function *f*: $V_s \rightarrow V$, *Gs* is called a sub-graph of *G*, if

(a) $\forall v \in V_s, l_s(v) = l(f(v))$ and
(b) $\forall (u, v) \in E_s, (f(u)f(v)) \in E, l_s(u, v) = l(f(u), f(v))$

Algorithm 2 Frequent sub-graph mining algorithm.

Input: burst topic user graph G, minimum support threshold τ
Output: all sub-graphs Gs of G such that $\sup_G(Gs) > \tau$
(1) init sub-graph set of G as S, $S \leftarrow \phi$
(2) calculate the set of all frequent edges of G, FE
(3) **for** each $e \in FE$ **do**
(4) $S \leftarrow S \cup subgragh\_extension(e, G, \tau, FE)$
(5) remove e from G and FE
(6) **end for**
(7) **return** S

Algorithm 3 Sub-graph extension algorithm.

Input: A sub-graph Gs of topic user graph G, minimum support threshold $\tau$and the frequent edges FE
Output: all frequent sub-graphs of G such that extend Gs
(1) init frequent sub-graphs set that extend Gs as FS_{Gs}, $FS_{Gs} \leftarrow Gs$
(2) init candidate frequent sub-graph set as CS_{Gs}, $CS_{Gs} \leftarrow \phi$
(3) **for** each e in FE and user u of Gs **do**
(4) **if** e can be used to extend u **then**
(5) Let ex_e be the extension of Gs with e
(6) **if** ex_e is not already generated **then**
(7) $CS_{Gs} \leftarrow CS_{Gs} \cup ex_e$
(8) **end if**
(9) **end if**
(10) **end for**
(11) **for** each $c \in CS_{Gs}$ **do**
(12) **if** $\sup_G(c) > \tau$ **then**
(13) $FS_{Gs} \leftarrow FS_{Gs} \cup subgraph\_extension(e, G, \tau, FE)$
(14) **end if**
(15) **end for**
(16) **return** FS_{Gs}

Definition 4. The minimum support. Let $f_1, \ldots, f_n$ be the set of isomorphisms of a sub-graph $G_s =< V_s, E_s, l_s >$ in a graph G. Also let $F(V) = \{f_1(v), \cdots, f_n(v)\}$ be the set that contains the nodes in G whose functions $f_1, \cdots, f_n$ map a node $v \in Vs$. The minimum support of Gs in G, is defined as $\sup_G(Gs) = \min\{t | t = |F(V) \forall v \in V_s\}$

Definition 5. Micro burst pattern. Given a burst topic user graph G and a minimum support threshold τ, the micro burst pattern in burst topic user graph G is defined as finding all sub-graph G_s in G such that $\sup_G(Gs) > \tau$.

Based on the graph mining and frequent sub-graph mining algorithms [16–18], we propose a frequent sub-graph algorithm which is able to discover frequent pattern in burst topic user graph. The procedure of mining process is presented in Algorithm 2.

Frequent sub-graph mining algorithm starts by identifying set S that contains all frequent edges (i.e., with support greater or equal to τ) in the burst topic user graph (Lines 1–2). Based on the anti-monotone property, only these edges may participate in frequent sub-graphs. For each frequent edge, sub-graph extension algorithm (Algorithm 3) is executed (Lines 3–6).

This algorithm takes as input a sub-graph Gs and tries to extend it with the frequent edges of FE (Lines 3–10). All applicable extensions that have not been previously considered are stored in CS_{Gs}. To exclude already generated extensions, we adopt the DFS(Depth First Search) code canonical form as in Gspan [16]. Then, sub-graph extension algorithm (Lines 11–15) eliminates the members of CS_{Gs} that do not satisfy the support threshold τ. Finally, sub-graph extension algorithm is recursively executed (Line 13) to further extend the frequent sub-graphs.

5. Experimental results and findings

In this section, we first describe the dataset used in burst pattern mining, and then present the mining results and findings.

5.1. Dataset

We collected burst topic dataset from CLEar system developed by Pinnacle lab.[2] The system can detect and summarize burst topics in Singapore Twitter stream as soon as they emerge in real-time, which is convenient for us to collect burst topic features, tweet data and users data involved in burst topics. The collected burst topic dataset covered the period from November 1 to November 30 in 2015. Furthermore, in order to conduct burst topic user graph presented in Section 3.2, the

[2] http://pinnacle.smu.edu.sg/.

Table 2
The main field description of collected data.

Field	Description
tweet_id	The id of the tweet
is_geo	When true, indicates that the tweet is a geo tweet
is_retweet	When true, indicates that the tweet is a retweet
is_reply	When true, indicates that the tweet is a reply tweet
follower_count	The number of followers
user_id	The id of the user
is_verified	When true, indicates that the user is a verified account.
burst_number	The number of burst that system detected
detect_time	The detecting time of each burst

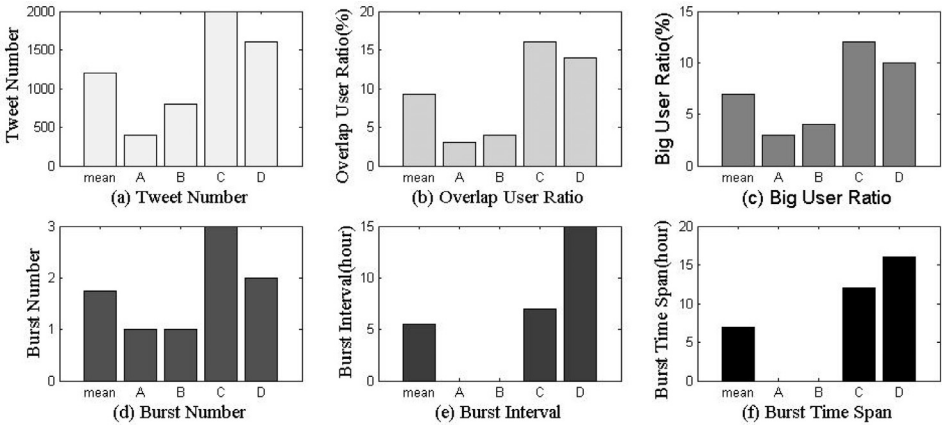


Fig. 2. The characteristics of the four Clusters based on selected features.

follower/followee relationships of burst topic users were also collected. For each burst topic, the main field description of collected data is shown in Table 2.

5.2. Results of macro burst pattern mining

In Section 4.1, we presented the definition of macro burst pattern and the features used for macro burst pattern mining. In this section, we perform hierarchical clustering on burst topics until the number of burst topic clusters reaches the maximum value and stops to increase. Specifically, we ignored clusters that consist of less than 10 burst topics (clusters that do not represent significant burst patterns) and as a result four significant clusters are obtained. For brevity, we name the four clusters as Cluster A, B, C, and D. As the clustering results cannot be easily interpreted, six features with coefficients larger than 0.5 are selected to analyze the clusters. The characteristics of the four Clusters based on selected features are shown in Fig. 2.

As shown in Fig. 2, the four clusters have very distinct patterns regarding to the selected features. Burst topics in Cluster A and B have the shortest *Burst Interval, Burst Time Span* and only one burst in their lifecycle. Specifically, burst topics in Cluster A have smaller *Tweet Number, Overlap User Ratio*, and *Big User Ratio* than Cluster B, which indicates that the scale of burst topics in Cluster A is smaller than Cluster B. In order to distinguish between Cluster A and Cluster B, we present the user engagement time series of four burst topic clusters, which is shown in Fig. 3.

As shown in Fig. 3, the user engagement of burst topics in Cluster A decline rapidly after reaching their peak, which are defined as non-persistent single burst topics in this paper. Burst topics in Cluster B decrease slowly after reaching their peak, which are defined as persistent single burst topics. Unlike Cluster A and B, burst topics in Cluster C and D have more than one burst in their lifecycle and the value of other features are larger than the mean. Besides, *Burst Interval* and *Burst Time Span* of burst topics in Cluster C are shorter than Cluster D. Based on the *Burst Interval* and *Burst Time Span* of each burst in burst topic, burst topics in Cluster C are defined as persistent multiple burst topics and burst topics in Cluster D are defined as non-persistent multiple burst topics.

5.3. Results of micro burst pattern mining

Based on the micro burst pattern mining algorithm presented in Section 4.2, we present the information flow patterns of burst topic in this section. According to the results of macro burst pattern mining presented in Section 5.2, single burst topics and multiple burst topics have distinct pattern. We discover the frequent patterns of *FP_AB* (consisting of Cluster

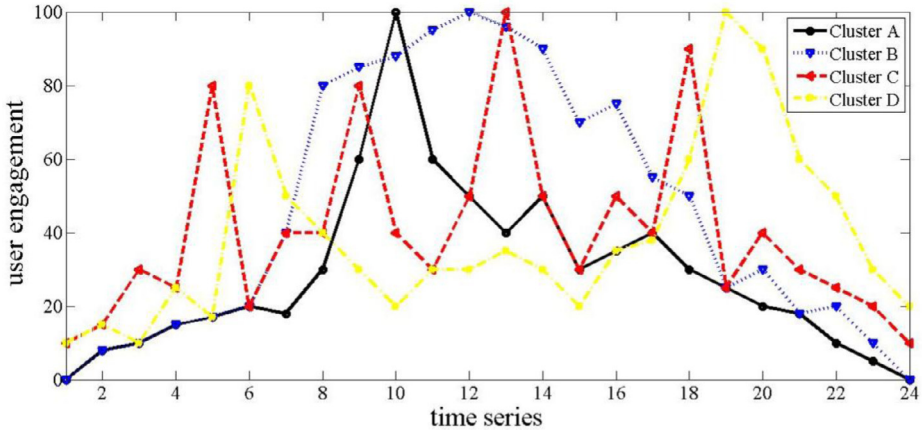
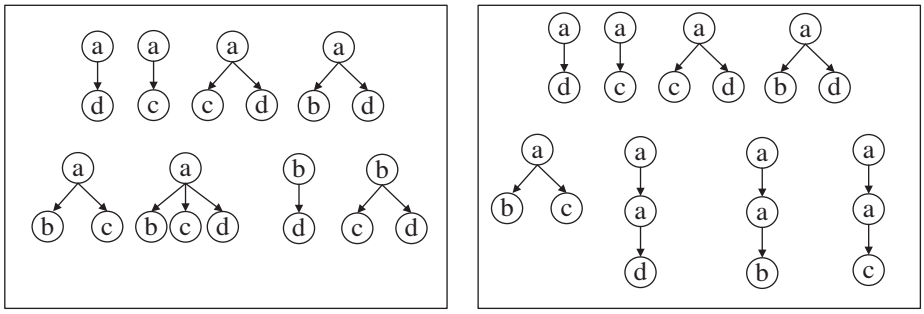


Fig. 3. The user engagement time series of four burst topic clusters.



(a) FP_AB (b) FP_CD

Fig. 4. Top 8 frequent burst patterns of single burst topics and multiple burst topics.

A and Cluster B) and *FP_CD* (consisting of Cluster C and Cluster D), respectively. As shown in Fig. 4, we select the top *8* frequent patterns, which are ranked by their support values.

By comparing the frequent patterns that represent information flow in burst topic, it can be observed that burst topics tend to propagate in certain paths.

(1) Information flow in burst topics tends to propagate from influential users to normal users. It indicates that the participation of influential users is the main reason of topic diffusion. For instance, information flows of most frequent patterns start with users labeled as *a* and end with other labels.
(2) Information flow in single burst topics tends to be wider rather than deeper. It indicates that the probability of information spreading into deeper level is low in single burst topics. For instance, the depth of all frequent patterns in Fig. 4(a) are two.
(3) Information flow in multiple burst topics tends to propagate between influential users and be deeper than single burst topics. On one hand, information propagation between influential users to can get more retweets and cause burst again. On the other hand, the probability of information spreading into deeper increase when more influential users are interested in this burst topic.

6. Conclusions

In this paper, we proposed the problem of mining burst patterns of burst topic in Twitter. In order to represent the topology structure of burst topic propagation across a large number of Twitter users, a burst topic user graph model is proposed. On one hand, hierarchical clustering is applied to cluster burst topics and reveal burst patterns from the macro perspective. On the other hand, frequent sub-graph mining is used to discover the information flow patterns of burst topic from the micro perspective. Experimental results show that several interesting burst patterns are discovered, which can reveal different burst topic clusters and frequent information flows of burst topic.

In future work, we will propose more qualitative and quantitative index to mine burst patterns, trying to discover micro burst patterns of all topic clusters. We will also apply our mining results on relevant burst topic researches.

Acknowledgements

This work is supported by the International Exchange Program of Harbin Engineering University for Innovation-oriented Talents Cultivation, China Scholarship Council, the Fundamental Research Funds for the Central Universities (no. HEUCF100605), the National High Technology Research and Development Program of China (no.2012AA012802) and the National Natural Science Foundation of China (no.61170242,no. 61572459); the National Research Foundation, Prime Minister's Office, Singapore under its International Research Centres in Singapore Funding Initiative and Pinnacle lab for analytics at Singapore Management University.

References

[1] Kasiviswanathan SP, Melville P, Banerjee A, Sindhwani V. Emerging topic detection using dictionary learning. In: Proceedings of the 20th ACM international conference on Information and knowledge management. ACM; 2011, October. p. 745–54.
[2] Agarwal MK, Ramamritham K, Bhide M. Real time discovery of dense clusters in highly dynamic graphs: identifying real world events in highly dynamic environments. Proc VLDB Endowment 2012;5(10):980–91.
[3] Alvanaki F, Sebastian M, Ramamritham K, Weikum G. EnBlogue: emergent topic detection in web 2.0 streams. In: Proceedings of the 2011 ACM SIGMOD international conference on management of data. ACM; 2011, June. p. 1271–4.
[4] Mathioudakis M, Koudas N. Twittermonitor: trend detection over the twitter stream. In: Proceedings of the 2010 ACM SIGMOD international conference on management of data. ACM; 2010, June. p. 1155–8.
[5] Cataldi M, Caro Di, Schifanella C. Emerging topic detection on twitter based on temporal and social terms evaluation. In: Proceedings of the tenth international workshop on multimedia data mining. ACM; 2010, July. p. 1–10.
[6] Nguyen T, Phung D, Adams B, Venkatesh S. Event extraction using behaviors of sentiment signals and burst structure in social media. Knowl Inf Syst 2013;37(2):279–304.
[7] Takahashi T, Tomioka R, Yamanishi K. Discovering emerging topics in social streams via link anomaly detection. In: 2011 IEEE 11th international conference on data mining. IEEE; 2011, December. p. 1230–5.
[8] Cui A, Zhang M, Liu Y, Ma S, Zhang K. Discover breaking events with popular hashtags in twitter. In: Proceedings of the 21st ACM international conference on Information and knowledge management. ACM; 2012, October. p. 1794–8.
[9] Li C, Sun A, Datta A. Twevent: segment-based event detection from tweets. In: Proceedings of the 21st ACM international conference on Information and knowledge management. ACM; 2012, October. p. 155–64.
[10] Lee P, Lakshmanan LV, Milios E. Keysee: Supporting keyword search on evolving events in social streams. In: Proceedings of the 19th ACM SIGKDD international conference on knowledge discovery and data mining. ACM; 2013, August. p. 1478–81.
[11] Wang Y, Liu H, Lin H, Wu J, Wu Z, Cao J. SEA: a system for event analysis on chinese tweets. In: Proceedings of the 19th ACM SIGKDD international conference on knowledge discovery and data mining. ACM; 2013, August. p. 1498–501.
[12] Xie W, Zhu F, Jiang J, Lim EP, Wang K. Topicsketch: Real-time bursty topic detection from twitter. In: 2013 IEEE 13th international conference on data mining. IEEE; 2013, December. p. 837–46.
[13] Xie R, Zhu F, Ma H, Xie W, Lin C. CLEar: a real-time online observatory for bursty and viral events. Proc VLDB Endowment 2014;7(13):1637–40.
[14] Shen G, Yang W, Wang W, Yu M. Burst topic detection oriented large-scale mi-croblogs streams. J Comput Res Dev 2015;52(2):512–21 (in Chinese).
[15] Batagelj V, Mrvar A. Pajek-program for large network analysis. Connections 1998;21(2):47–57.
[16] Yan X, Han J. GSPAN: Graph-based substructure pattern mining. In: 2002 IEEE international conference on data mining. IEEE; 2002. p. 721–4.
[17] Thomas LT, Valluri SR, Karlapalem K. Margin: Maximal frequent subgraph mining. In: 6th international conference on data mining. IEEE; 2006, December. p. 1097–101.
[18] Ranu S, Singh AK. Graphsig: A scalable approach to mining significant subgraphs in large graph databases. In: IEEE 25th international conference on data engineering. IEEE; 2009, March. p. 844–55.
[19] Li Y, Lu H, Wang Y, Zhang L, Yang S, Serikawa S. Robust color image segmentation method based on weighting Fuzzy C-Means Clustering. In: 2012 IEEE/SICE international symposium on system integration. IEEE; 2012, December. p. 133–7.

Guozhong Dong is currently a PhD. candidate in the Department of Computer Science and Technology and Information Security Research Center, Harbin Engineering University. He received his B.E. degree in 2011 from Harbin Engineering University. His main research interests include data mining, social computing and information security.

Wu Yang received his PhD. degree in computer system architecture from Harbin Institute of Technology in 2005. He is currently a professor and doctoral supervisor of Harbin Engineering University. His main research interests include data mining, information security and wireless sensor network.

Feida Zhu is an assistant professor in the School of Information Systems at the Singapore Management University (SMU). He hold a PhD degree in Computer Science from the University of Illinois at Urbana–Champaign and obtained his B.S. degree in Computer Science from Fudan University. His current research interests include graph pattern mining, information/social network analysis and business intelligence.

Wei Wang received his PhD. degree in computer system architecture from Harbin Institute of Technology in 2005. He is currently an associate professor of Harbin Engineering University. His main research interests include data mining and information security.