

## A Novel Event-centric Trend Detection Algorithm for Online Social Graph Analysis

Ling Wang, Haijing Jiang, Tiehua Zhou\*, Wei Ding and Zhiyuan Chen  
*Department of Computer Science and Technology, School of Information Engineering, Northeast Dianli University, Jilin, China*  
*smile2867ling@163.com, haijing\_103702@126.com, thzhou55@163.com, wxj3429@126.com, tjxczy123@sina.com*

### Abstract

Nowadays, the identification of the most popular and important topics discussed over social networks, is became a vital societal concern. For real-time tracking the hot topics, we proposed a novel event-centric trend detection algorithm, which called *Ec\_TD* algorithm to attempt to add event attributes into the structure of the social networks, then, mining the subgraphs induced by specific attributes which using correlation function measures the correlation of event-changing attributes based on the attribute-extended social network structure. Our experiment shows that *Ec\_TD* algorithm is performed significantly better in real-time event detecting and mining the potential relationships between attributes and vertexes. Moreover, we used true big data to test this algorithm which has substantially reduced respond time, and to prove the feasible of the idea.

**Keywords:** event-centric trend detection, *Ec\_TD* algorithm, structure correlation, social networks

### 1. Introduction

With the popularization and development of the Internet, the era of the great explosion of the information has already come. Social networks provide large-scale information infrastructures for people to discuss and exchange ideas about different topics. Interested on detecting trends of such topics is for many reasons. For one, trends can be used to detect emergent or suspicious behavior in the network. They can also be viewed as a reflection of societal concerns or even as a consensus of collective decision making. Understanding how a community decides that a topic is trend can help us better understands how ad-hoc communities are formed and how decisions are made in such communities. Figure 1 shows extract a fraction of graph which is the nature of social network. The fraction can be seen as three communities. In general, constructing useful trend definitions and providing scalable detection methods for such definitions will contribute towards a better understanding of interactions in the analysis of social graphs.

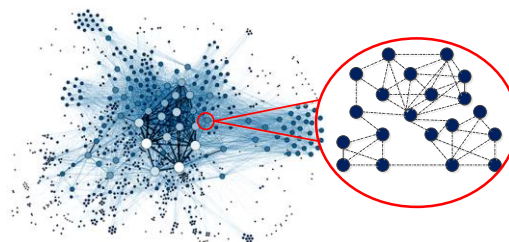


Figure 1. The Nature of Social Networks

---

\* Corresponding Author

Although trends detecting in social networks have been extensively studied, to our knowledge all the published work in this area focusing on the relationships among human mostly, ignores the structural properties of the social network that created these trends. In today's social networks where users are highly influenced by their friends, trend definitions which reach beyond simple heavy-hitters approaches. These solutions integrate the importance of such flow with influence, can be obtain great benefit. The main purpose of this paper is study the relationship between event and human, take the event as an attribute that add to the topology of the human, involved in the operation directly.

Trend detection is as equally popular as it is difficult. Technical, analytical and operational hurdles arise that can derail success when working to uncover and understand trends in a way that meets a business or organization's specific objectives. Trend detection is important to many people and organizations. It makes sense that we would want to know more about what is suddenly increasing in popularity. Knowing what is suddenly increasing in popularity is incredibly important to news organizations, first responders, government entities and more. At a more specific level, knowledge of trending topics tells us about what people are attracted to, what people think is noteworthy or important and what people make the effort to pass on or share. Armed with that knowledge, businesses can begin to make very effective real-time decisions.

The work studies the correlation between vertex attributes and subgraphs, a task we call structural correlation pattern mining. The structural correlation of an attribute set is the probability of a vertex to be member of its induced graph. Attribute structure correlation algorithm is proposed new processing methods in recent years, existing methods is not perfect. And most based on human separately or based on event independent, hybrid their analysis finally. In order to solve such problems as follows: 1) what happened within the community; 2) how the specific events progressed; 3) dynamic analysis of the information; 4) allow the user to see what is happening within group near real-time; 5) foresee what will happen in the group; 6) what the future trend move etc. several actual issues. To the best of our knowledge, this is the first work that takes event as an attribute added to the network topology structure, participate in computing immediately. With the purpose their relationship more intimately, we combine human with event, to introduce event attribute into social network. We introduce a novel event-centric based on attribute motivation definition, emphasize its significance and provide efficient online solution for it. We show that structural trends identify interesting activities in social networks.

Moreover, we propose two methods for computing the structural correlation of attribute sets, and count superposition of event attributes. These techniques are integrated in *Ec\_TD* algorithm, which is described and evaluated in this paper. In particular, we apply *Ec\_TD* to the analysis of real attributed graph: vote network. The results show that *Ec\_TD* is able to extract relevant knowledge regarding how vertex attributes are correlated with subgraph in large attributed graphs.

## 2. Related Works

The hot topic 'event' is a core event or activity, as well as related them directly. Sometimes, an event usually occurred due to some cause or condition, while an event is usually caused due to some conditions, at a specific time and place, involving some objects included a person or thing; also couple with some inevitable result. Top event occurred in real world represent hot topic on SNS platform.

Trends in social network have recently been a focus of interest for many researchers. A bulk of research concentrated on trend from a hot topic. S. Ishikawa *et al.*, [1] propose a classification method that mitigates the variation of posted words related to the same topic. S. B. Hill *et al.*, [2] use the previous data structure improvement, improves on their

heuristic arguments by formalizing the representation with three tunable parameters. B. O'Connor *et al.*, [3] proposed a new retrieval tool 'Twitter Motif', in the navigation form of key terms to present retrieval results, this application to provide convenience for scan, summarizes views *etc.*, demands. M. Mathioudakis [4] constructed 'Twitter Monitor', this system can detect the emerging topic on Weibo flow real time, describe each topic accurately and compound finally. With the popularity of Twitter, although research about event forecast based on twitter data in the beginning, but also have good achievements. L. Guo *et al.*, [5] through analysis of Twitter, Flickr, Delicious and other online social networks, according the user post content to generate the relative significant law. About emerging events, existing techniques for discovering such events from a microblog stream in real time, useful for finding information about known events, a series of studies highlight the importance of event that unraveling in microblog message streams. J. Bollen *et al.*, [6] evaluated user state, based on Google-Profile of Mood States (GPOMS) algorithm; find the calm index that could predict fluctuations of the Dow Jones industrial average index 2-6 days in advance. L. Chen *et al.* [9] through study discussion on twitter users to predict the presidential election. S. Asur *et al.*, [7] to analyze and predict the movie box-office, utilize the movie review of abundant users in a certain period of time. B.P Sharifi *et al.*, [8] introduce a method that a series of post about a topic summarized to a brief summary. B.P Sharifi and Inouye D have carried on the comprehensive description of Weibo abstract. L. Chen *et al.*, [9] with a view to differences from different groups in the generated content and tweeting behavior. In his study can shed light on improving the social media based prediction from the user sampling perspective and more.

Although trend analysis based on temporal and spatial characteristics is important for a better understanding of trends, unlike these studies, we focus structural properties of the network that create trends. S. Nikolov *et al.*, [10] proposes a setting for model specification and selection in supervised learning based on a latent source model. The rise of networks has presented numerous graphs; these graphs not only have topological structures, but also contain events/activities that occurred on their nodes.

Y. Yang *et al.*, [11] using the condensed type average algorithm combining the strategy, the related event which similar to the same topic model as a result of topic detection, then make TD system has the ability to review dependent event. Y. W. Seo *et al.*, [12] introduce a single-pass algorithm that benefit on simple computing and high speed which to detect new event, the post-condition of this algorithm was highly dependent on the processing order of news corpus. Further, the methods in [10-12] are not able to capture the evolution of event, we propose a novel measure for assessing such structural correlations in attribute sets with events. S. Hendrickson *et al.* [13] develop a framework for tracking short, developing scalable algorithms for clustering phrases that travel relatively intact through on-line text. M. K. Agarwal *et al.*, [14] propose two approximation algorithms; first algorithm efficiently computes approximate correlation coefficients of similar signal pairs within a given error bound. The second algorithm efficiently identifies, without any false positives or negatives, all signal pairs with correlations above a given threshold. A. Mueen *et al.*, [15] describe an exact algorithm MOEN, using a novel range across lengths only in linear space to enumerate motifs unlike others motif discovery algorithms. In [16], the authors propose novel algorithms which based on DFT and graph partitioning, to reduce the end-to-end response time of an all-pair correlation query. A. Dong *et al.*, [17] proposed a method to address both problems as quickly crawling relevant content and ranking documents with impoverished. Overall above, community interest was equally important. [18], and its extension [19], address the problem of supporting efficient retrieval of important 2-neighbors of any node, where the importance of a neighbor is related to local and global edge thresholds. C. Cortes *et al.*, [19] introduce a data structure which represented union of subgraphs centered on each nodes, then defined COI what formed by subgraphs.

### 3. Ec\_TD Algorithm for Trend Detection

#### 3.1. Motivation

For instance, consider the shopping rush on Taobao.com in dual 11, which guided by Alibaba Group as Figure 2 shown. Millions of people (at first only in China, spread all over the world later) participated in the event, the concern about that becomes an enormous source of trend detection. All kinds of reports were extensively covered on Weibo, Twitter, the popular micro-blogging service, significantly in advance of traditional media. Personally take an example of youth; his social circle has divided into many categories (e.g., parents, schoolmate, buddy, tradesman and so on). With different ages and complex backgrounds of crowd, there is an interest gap among them. But everyone was given the initial attribute p.

When event A “dual 11 turnover up to 91.2 billion Yuan” occurring, quick response from who always shopping on the internet just like maid and woman that have accounted to 70% of the circle. Then event B “Dual 11’s rate of return reach 63%, and the Alibaba clarification is a rumor” spread explosively, the groups (e.g., parents, middle class, person who unfamiliar with e-commerce) were not meant to pay attention for “dual 11”, since product quality involved, start to concern the continuous events. Now, 90% circle of youth, participated in response of event. When the next correlated event C occur just as “Who is the Alibaba behind the scene?” how the people react that could quantified as a computable problem. And plus a lot of people who was not involved in. We find the people who reacted to the first two events, also concern next close event. Because  $p_{AB}$  and  $p_{ABC}$  correlation is highest that indicated when event C occurring, all of  $p_{AB}$  are associated with C. then will find  $p_{AB}$  was the cause chain of C.

When the event D occurs, according to the event attribute correlation, we could detect trend of the circle of youth. Assume the correlation between C and D was highest, infer the correlation  $p_{ABC}$  and  $p_{ABCD}$  also is high, that is to say, when the D occur which may cause the next peak of attention.

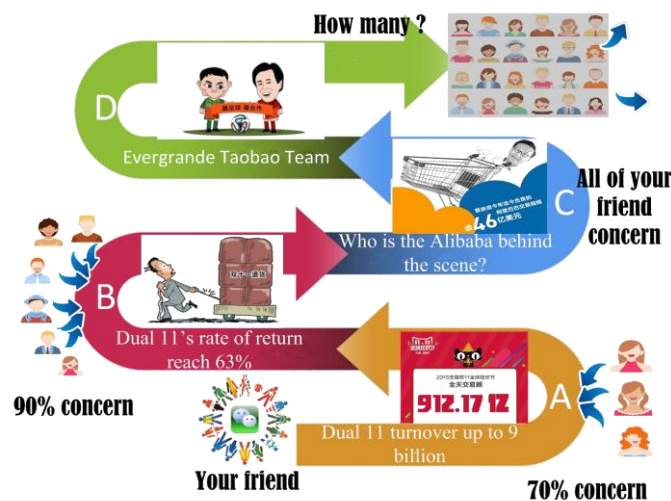


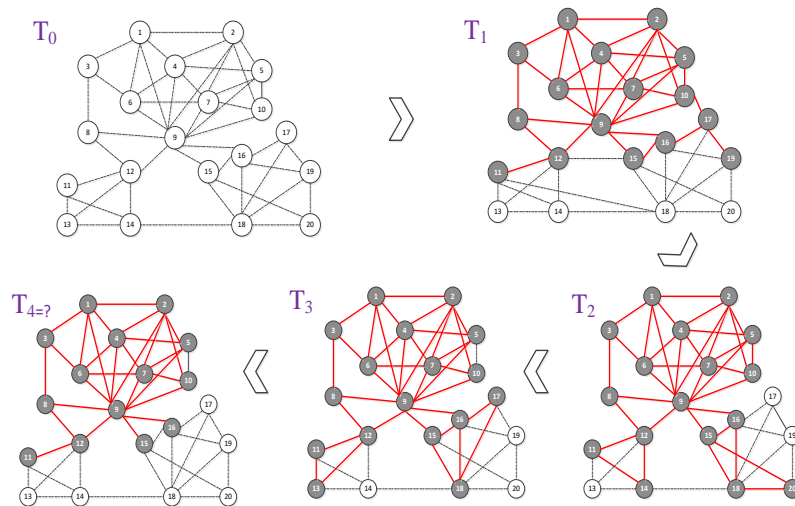
Figure 2. An Example of Motivation Analysis

Every event that is published, results in the attribute update of one or more vertices in the entity graph. The high frequency of event generation, coupled with our need for timely reporting of emerging stories, necessitates that the identification of subgraph structures in the entity graph be highly efficient.

### 3.2. Problem Definition

A social networks with  $n$  individuals and  $m$  link edges can be noted as graph  $G=(V, E)$ , where  $V$  is the set of nodes,  $V=n, n \leq 20, v_i$  is a neighbor of  $v_j$ . At any point nodes of the network can share information on any topic with neighbors. We model each such mention by node  $n_i$  on a specific topic  $X$  as a tuple this section we describe our algorithm with an example. And  $E$  is the set of undirected relationships,  $E \subseteq V * V, E=m$ . A series of topics on behalf of event defined  $a, b \dots$  which  $a, b$  and so on were triggered by the former. We define community as non-overlapping groups that can be assumed have the same attribute  $p$  first at  $T_0$ , then subsequence also define attribute  $B, C, D$  and so on. Attributes  $p, A, B \dots \subset X; A, B, C \dots$  corresponding topics  $a, b, c, \text{ etc.}$ , which from set  $\mathcal{H}$ . Topics in accordance with event happening into each point of time  $T_1, T_2, T_3 \dots T_i$ . (e.g., We divided the process into several stages, respectively defined periods of time  $[T_0, T_1), [T_1, T_2), [T_2, T_3), [T_3, T_4) \dots$  the structure correlation of an attribute is the probability of a vertex to be member of a dense subgraph in its induced graph.

**DEFINITION 1 (STRUCTURAL CORRELATION FUNCTION)** given a changing attribute  $X$ , the structural correlation of  $X, \alpha(X)$ , is given as:  $\alpha(X) = K(X_i)/V$  where  $K(X_i)$  is the set of changing nodes in the time  $T_i$



**Figure 3. Attributes Graph Changed by Real-Time**

1) At the time of  $T_0$ , extraction a part of graph from the circle which divided into three parts naturally, here we define the communities as a list of non-empty node subsets:  $Coms = \{ v_1 \sim v_{10}, v_{11} \sim v_{14}, v_{15} \sim v_{20} \}$ . Generally speaking, the persons in social networks correspond to a set of nodes in graphs as Figure 3 shown.

2) For topics  $a, b, c$  occurring at  $T_1, T_2, T_3 \dots$  when the topic  $a$  spread, about 16 vertices responded, produced a new correlation, formed a new group. Active vertex proportion accounted for 80%. Now, we define the correlation degree between the attribute  $p$  and attribute  $A$  is 0.8.

3) The subsequent topic  $b$  resulted in a new change in the graph at  $T_2$ , which has no action at a time of  $T_1$  originally for topic  $a$ . Several nodes (e.g.,  $v_{14}, v_{18}, v_{20}$ ) does not change before, now has changed. Some vertices continue to change as events occurred in a row. After topic  $c, p_{AB} = 0.7, p_B = 0.15, p_A = 0.1$ .

4) So far, by the end of topic  $d$  at  $T_3$ , every attributes of vertices ultimately as shown in Table 1:

**Table 1. Patterns and Correlation Parameters**

pattern	size	$\alpha$
$(\{p_A\}, \{19\})$	1	0.05
$(\{p_B\}, \{14, 20\})$	2	0.1
$(\{p_C\}, \{13\})$	1	0.05
$(\{p_{AC}\}, \{17\})$	1	0.05
$(\{p_{BC}\}, \{18\})$	1	0.05
$(\{p_{AB}\}, \{1, 2, \dots, 11, 12, 15, 16\})$	14	0.7
$(\{p_{ABC}\}, \{1, 2, \dots, 11, 12, 15, 16\})$	14	0.7

Though the correlation of topics, to calculate the incidence of the attribute. The correlation of the same attribute is high after the occurrence of continuous event, just like the attribute of  $v_1 \sim v_{11}$ ,  $v_{12}$ ,  $v_{15}$ ,  $v_{16}$  were the same. Thus it can be seen no matter whether  $c$  occurs, the people who react to topic  $a$ ,  $b$  will respond to  $d$ .

### 3.3. Ec\_TD Algorithm

This section presents the *Ec\_TD* (Event-centric Trend Detection) algorithm, which adds the event attribute in order to enable trend detection better in large changing graph by time. Traditional detection algorithm Spend a lot of time to traverses the graph for many times, unlike the naive algorithm, *Ec\_TD* does not only focus on the weight update of edges but depend on the structural topology. In order to achieve such goal, in the following paragraph, we describe *Ec\_TD* algorithm in detailed.

In this initialized phase, the graph elements need to get their initial propinquity values. After the propinquity measure ( $\rho$ ) and the hashmap str. ( $\delta$ ) are defined, the procedure starts the voting. Given a series of voter about election, when the queue is not null, for every node add the queue to circulation. Assign the hashmap  $\langle key, value \rangle$ , for  $i$  from string on the basis of '0' to split. The results of the multiple characters save the key. Similarly, split  $i$  on the basis of '1', and take the result to save to value. If the queue concludes the  $key$ , which is always equals to the sum of the current value and forwarding value, recursively. This part of the transformation puts the results of the value into a key.

---

*Ec\_TD* algorithm

---

```

Input:  $G(V,E)$ ,  $\mathcal{H}$  and  $X$ 
Output: str.  $\langle key, value \rangle$ 
1: Initialization  $\rho, \delta$ ;
2:  $T \leftarrow 0$ ,  $queue\ list \leftarrow \rho$ ,  $hashmap \leftarrow \delta$ 
3: start:
4:   visit list;
5:   if list != NULL;
6:   while add list to str
7:   for each i do
8:      $key = str[i].split[0]$ ;
9:      $value = str[i].split[1]$ ;
10:    if str.containskey
11:      str.put( $key, value + str.get(key)$ );
12:    else
13:      str.put( $key, value$ )
14:   return str;
    
```

---

## 4. Experimental Results

This section presents case studies on the structural correlation pattern mining using real datasets. In the attribute graph extracted from large-scale real-world networks like the Wikipedia Vote Network from Stanford university (<http://snap.stanford.edu/data/wiki-Vote.html>). We report the statistics in Table 2, the experiments used Java programming language based on Vulcan 16G memory *Wgraphics* processor. Where  $n_{wcc}$ ,  $e_{wcc}$  is nodes and edges in largest *WCC* respectively, with that  $n_{scc}$ ,  $e_{scc}$  is nodes and edges in largest *SCC*,  $cc_{avg}$  is the average cluster coefficient, and *diam* the network diameter.

**Table 2. Data form Wikipedia Vote Network**

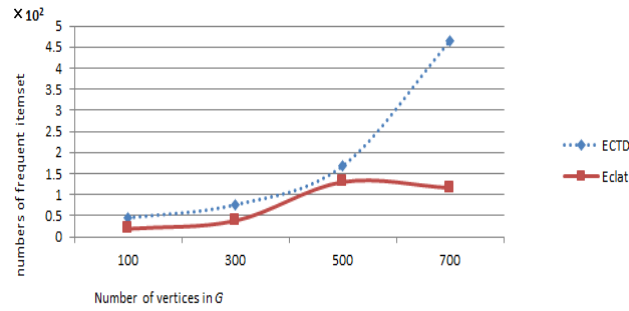
<i>nodes</i>	<i>edges</i>	$n_{wcc}$	$e_{wcc}$	$n_{scc}$	$e_{scc}$	$cc_{avg}$	<i>diam</i>
7115	103689	7066	103663	1300	39456	0.1409	7

Each vertex represents a voter, elections and the candidate are connected after finished a round of voting. The attributes of candidates are terms that appear in the ballot of election voted by them. Using the latest complete dump of *Wikipedia* page edit history (from January 3, 2008) we extracted all administrator elections and vote history data. Nodes in the network represent *Wikipedia* users and a directed edge from node  $i$  to node  $j$  represent that user  $i$  voted on user  $j$ . In this paper, first we take an example of an arbitrary number of information campaigns in a social network. Then we analyze data from *Wikipedia* vote network, a free encyclopedia written collaboratively by volunteers around the world. *Wikipedia* community via a public discussion or a vote decides who to promote to *adminship*. The environment is parallel to the behaviors what user participated in social networks, forward, thumb up, etc.

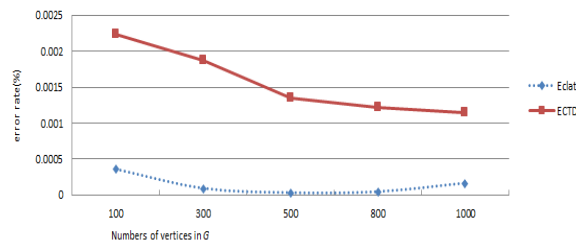
Moreover, we evaluate the performance and study the sensitivity of important input parameters of *Ec\_TD*. This algorithm which applies the strategies for efficient structural correlation pattern mining presented in this section. Since structural correlation pattern mining combines aspects of the frequent itemset mining and the quasi-clique mining problems, we attempt a new measure into structure correlation pattern mining. Further in this paper, we compared our algorithm to *Eclat* algorithm [20].

In this section, we analyzed the performance of our method comparing *Eclat* Algorithm by experiment. As shown in Figure 4, the comparison frequent itemsets at the different number of nodes, Because of advantages in the event-centric graphical structure, *Ec\_TD* algorithm has better performance than *Sclat* algorithm obviously after 500 nodes as seen from the figure; in Figure 5 and Figure 6, Comparing error rate by different number of node or edges, when the data set was small, not seen any advantage yet. But when growth of data sets, our algorithm is much better than *Sclat*, Because we only consider the changing node with events occurring, a total of 132 nodes had been extracted as the core frequent node structure, reduced dataset infrequently expenses when searching. Figure 7 is comparing the query time by different number of node.

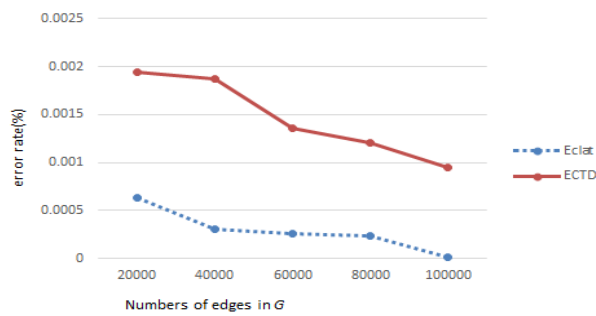
The maximum time is 2.73 seconds that the experiment processing, this demonstrated by result verified our algorithm perform well under above conditions.



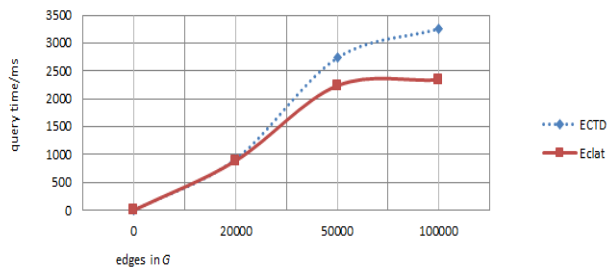
**Figure 4. Frequent Itemsets**



**Figure 5. Error Rate of Different Nodes**



**Figure 6. Error Rate of Different Edges**



**Figure 7. Query Time by Different Nodes**

## 5. Summary

In this paper, we introduced a new method for tracking the event-centric trend over online social graph analysis. Firstly, online motivation analysis could use to find the cause chain of the trigger event. Secondly, structure correlation algorithm of trend detection could predict possible trend of events developing trend. Finally, accomplish deeply



mining the causes of hot events over social platforms, in order to predict the developing direction of the incident.

## Acknowledgments

The Project Sponsored by the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry, by the Education Department Foundation of Jilin Province (No.201698) and by the Science and Technology Plan Projects of Jilin city, China (No.201464059).

## References

- [1] S. Ishikawa, Y. Arakawa, S. Tagashira and A. Fukuda, "Hot topic detection in local areas using Twitter and Wikipedia", Proceeding of International conference on Architecture of Computing Systems, Munich, Germany, (2012) February 28-March 02.
- [2] S. B. Hill, D. K. Agarwal, R. Bell and C. Volinsky, "Building an effective representation for dynamic networks", Journal of Computational and Graphical Statistics, (2012), pp. 584-608.
- [3] B. O'Connor, M. Krieger and D. Ahn, "TweetMotif: Exploratory Search and Topic Summarization for Twitter", Proceeding of the 4th International Conference on Weblogs and Social Media, Washington, USA, (2010) May 23-26.
- [4] M. Mathioudakis and N. Koudas, "TwitterMonitor: trend detection over the twitter stream", Proceedings of the ACM SIGMOD International Conference on Management of Data, Indianapolis, USA, (2010) June 6-10.
- [5] L. Guo, E. Tan, S. Chen, X. Zhang, and Y. Zhao, "Analyzing patterns of user content generation in online social networks", Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, Paris, France, (2009) June 28 - July 01.
- [6] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market", Journal of Computational Science., (2011), pp. 1-8.
- [7] S. Asur, B. A. Huberman, G. Szabo and C. Wang, "Trends in social media: Persistence and decay", Proceeding of 5th International AAAI Conference on Weblogs and Social Media, Barcelona, Spain, (2011) July 17-21.
- [8] B. P. Sharifi, "Automatic microblog classification and summarization", University of Colorado at Colorado Springs, (2010).
- [9] L. Chen, W. Wang and A. P. Sheth, "Are Twitter users equal in predicting elections?", Proceeding of the 4th International Conference on Social Informatics, Lausanne, Switzerland, (2012) December 5-7.
- [10] S. Nikolov, "Trend or no trend: a novel nonparametric method for classifying time series", Twitter Inc, (2012).
- [11] Y. Yang, T. Pierce and J. Carbonell, "A study of retrospective and on-line event detection", Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval, Melbourne, Australia, (1998) August 24-28.
- [12] Y. W. Seo, K. Sycara, "Text clustering for topic detection", Carnegie Mello University, (2004).
- [13] S. Hendrickson, J. Montague, J. Kolb and B. Lehman, "Trend detection in social data". Twitter Inc, (2015).
- [14] M. K. Agarwal, K. Ramamritham and M. Bhide, "Real time discovery of dense clusters in highly dynamic graphs: identifying real world events in highly dynamic environments", Journal of the VLDB Endowment., vol. 5, no. 10, (2012), pp. 980-991.
- [15] A. Mueen and N. Chavoshi, "Enumeration of time series motifs of all lengths", Journal of Knowledge and Information Systems., vol. 45, no. 1, (2015), pp. 105-132.
- [16] A. Mueen, S. Nath and J. Liu, "Fast approximate correlation for massive time-series data", Proceedings of the 2010 ACM SIGMOD International Conference on Management of data, Indianapolis, USA, (2010) June 6-10.
- [17] A. Dong, R. Zhang, P. Kolari, J. Bai, F. Diaz, Y. Chang, Z. Zheng and H. Zha, "Time is of the essence: improving recency ranking using twitter data", Proceedings of the 19th international conference on World wide web, Raleigh, USA, (2010) April 26-30.
- [18] J. Leskovec, L. Backstrom and J. Kleinberg, "Meme-tracking and the dynamics of the news cycle", Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, Paris, France, (2009) June 28-July 01.
- [19] C. Cortes, D. Pregibon and C. Volinsky, "Computational methods for dynamic graphs", Journal of Computational and Graphical Statistics., (2012).
- [20] M. J. Zaki, "Scalable algorithms for association mining", Journal of Knowledge and Data Engineering, vol. 12, no. 3, (2000), pp. 372-390.

## Authors



**Ling Wang** is currently an associate professor in the Department of Computer Science and Technology, Northeast Dianli University in China. She received the Ph.D degree in Computer Science from Chungbuk National University of Korea, in 2013.

Her research interests are mainly in the areas of data mining, graph computing, cloud computing, big data processing, Smart Grid and Energy Internet.



**Haijing Jiang** is currently a master student in the Department of Computer Science and Technology, Northeast Dianli University in China. Her research interests are mainly in the areas of social networks and smart grid.



**Tiehua Zhou** is currently an associate professor in the Department of Computer Science and Technology, Northeast Dianli University in China. He received the Ph.D degree in Computer Science from Chungbuk National University of Korea, in 2016.

His research interests are mainly in the areas of multimedia image processing, data mining, spatial-temporal database and Energy Internet.



**Wei Ding** is currently a master student in the Department of Computer Science and Technology, Northeast Dianli University in China. His research interests are mainly in the areas of graph computing and smart grid.



**Zhiyuan Chen** is currently a master student in the Department of Computer Science and Technology, Northeast Dianli University in China. His research interests are mainly in the areas of data mining, smart grid and Energy Internet.