An Efficient Approach to Event Detection and Forecasting in Dynamic Multivariate Social Media Networks

Minglai Shao[†], Jianxin Li[†], Feng Chen[‡], Hongyi Huang[†], Shuai Zhang[†], Xunxun Chen[†]
[†]School of Computer Science and Engineering, Beihang University
[‡]Department of Computer Science, State University of New York at Albany
{shaoml, lijx, huanghy, zhangs}@act.buaa.edu.cn, fchen5@albany.edu, xx-chen@139.com

ABSTRACT

Anomalous subgraph detection has been successfully applied to event detection in social media. However, the subgraph detection problembecomes challenging when the social media network incorporates abundant attributes, which leads to a multivariate network. The multivariate characteristic makes most existing methods incapable to tackle this problem effectively and efficiently, as it involves joint feature selection and subgraph detection that has not been well addressed in the current literature, especially, in the dynamic multivariate networks in which attributes evolve over time.

This paper presents a generic framework, namely dynamic multivariate evolving anomalous subgraphs scanning (DMGraphScan), to addressthis problem in dynamic multivariate social media networks. We generalize traditional nonparametric statistics, and propose a new class of scan statistic functions for measuring the joint significance of evolving subgraphs and subsets of attributes to indicate the ongoing or forthcoming event in dynamic multivariate networks. We reformulate each scan statistic function as a sequence of subproblems with provable guarantees, and then propose an efficient approximation algorithm for tackling each subproblem. This algorithm resorts to the Lagrangian relaxation and a dynamic programming based on tree-shaped priors. As a case study, we conduct extensive experiments to demonstrate the performance of our proposed approach on two real-world applications (flu outbreak detection, haze detection) in different domains.

Keywords

dynamic multivariate networks; social media; evolving subgraphs detection; feature selection; nonparametric statistics; approximation algorithm.

1. INTRODUCTION

Over the recent years, the surge of social media, such as Twitter, Weibo and Facebook, has significantly advanced the way that people acquire and share daily events. Besides, many governments, enterprises and social media microblog accounts are registered to accelerate the spread of particular events [1, 6]. Social media presents several advantages for event detection [7]. First, owing to the real-time nature of online social services, the public awareness of real world happenings could be raised in a much quicker fashion than with the traditional media. Second, due to the large amount of users posting content online, more complete pictures of the real world events with descriptions from different angles are offered with fast and large-scale coverage [11, 15].

This paper focuses on the problem of domain-specific event detection and forecasting, such as disease outbreaks and air pollution events in social media. Naturally, social media are structured as dynamic multivariate networks with: 1) vertices, such as users or locations; 2) relationships, such as spatial neighborhood and followers; 3) attributes, such as frequencies of domain-specific keywords, which evolve over time. Based on the dynamic multivariate networks, events can be represented as evolving anomalous subgraphs (e.g., connected subsets of vertices with abnormally high frequencies of domain-specific keywords), and the problem of event detection and forecasting is formulated as the detection of the most anomalous evolving subgraphs in dynamic multivariate social media networks. These subgraphs can be used for indicating the ongoing or forthcoming events.

Majority of existing methods to the problem of event detection based on social media resort to different functions

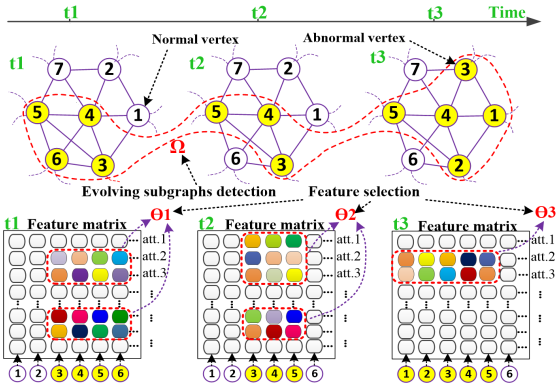


Figure 1: The proposed work focuses on the search of evolving subgraphs (Ω) and the subsets of features ($\Theta = \{\Theta 1, \Theta 2, \Theta 3\}$) that are jointly the most anomalous. Each subset of Θ corresponds to a snapshot of Ω in a separate time slice.

WWW 2017, April 3–7, 2017, Perth, Australia.
ACM 978-1-4503-4913-0/17/04.
http://dx.doi.org/10.1145/3038912.3052588

which aggregate multivariate attributes, and assume that the relevant attributes are known beforehand and these predefined attributes are mostly signal attributes. Then these methods focus on the search of subgraphs whose attributes are the most anomalous on the whole. Specifically, Kulldorff first calculates a separate log-likelihood ratio score for each feature and then aggregate these scores in to a single score by adding [9]. Burkom presents a simple, univariate aggregation of the multiple attributes for each vertex, and then converts the problem into uni-variate subgraph detection problem [3]. A brunch-and-bound method is proposed in [10] to search space-time regions where the aggregated counts of predefined terms are abnormally higher compared with the counts outside the regions. A two-stage empirical calibration process is proposed in [5] to convert multiple attributes of each vertex into a single empirical p-value. The empirical p-value estimate the probability that a randomly selected sample would have observed attributes as extreme as the current attributes of this vertex, under the null hypothesis that no events of interest are occurring. The proposed empirical calibration process is basically a feature extraction process. The reduction of multiple attributes to a single feature (empirical p-value) may lead to potential loss of valuable information relevant to events.

Nevertheless, these assumptions mentioned above are inadequate for event detection and forecasting in social media owing to the dynamic of attributes caused by events. Different events usually have different contexts, and their correlated attributes (e.g., frequencies of keywords) are unpredictable. As the result, the dynamic detection of attributes that are correlated with ongoing or forthcoming events becomes critical and challenging.

In general, it is necessary to trace a lot of keywords, but often only a few and unknown keywords will be relevant to a specific event in different separate time slices. Unfortunately, the majority of noise attributes will potentially dominate the aggregation of all attributes. This paper provides an alternative optimization framework in which the target is to optimize a score function of "anomalousness" over all subgraphs and attributes in dynamic multivariate social media networks. This optimization task contains a serious computational challenge: an exhaustive search over all evolving subgraphs and the corresponding attributes is computationally infeasible, scaling exponentially with the number of subgraphs and attributes. To the best of our knowledge, very limited work has been conducted to address this computational challenge.

Our main contributions are summarized as follows:

- **Formulation of the DMGraphScan framework.** A general framework, named as dynamic multivariate evolving anomalous subgraphs scanning (DMGraphScan), is proposed for tackling the domain-specific event detection and forecasting problem on dynamic multivariate social media networks. The events are decomposed into multidimensional subsets of vertices and attributes, and their signal strengths are characterized as nonparametric scan statistics that are free of distribution assumptions.

- **Design of an approximation algorithm for dynamic multivariate evolving anomalous subgraphs scanning**. We first efficiently reformulate the DMGraphScan problem as a sequence of subproblems with provable guarantees. Then, an approximation algorithm is proposed for solving the reformulated problems of nonparametric scan statistics in dynamic multivariate networks. This algorithm resorts to the Lagrangian relaxation and a dynamic programming based on tree-shaped priors, and can efficiently find an approximation solution for every subproblem.

- **Comprehensive experiments to validate the effectiveness and efficiency of the proposed framework.** Extensive experiments are conducted to evaluate the DMGaphScan in flu outbreak detection and haze detection. The results demonstrate that dGraphScan outperforms representative techniques in both performance and quality.

The rest of this paper is organized as follows. Section 2 presents preliminaries. Section 3 performs the proposed DMGraphScan framework. Section 4 presents an efficient approximation algorithm to the DMGraphScan with theoretical analysis. Experiments are presented in Section 5, and conclusion and future work are presented in the last Section.

2. PRELIMINARIES

This section presents several definitions, including dynamic multivariate network, p-value, evolving subgraphs and nonparametric scan statistics.

Definition 1 (Dynamic Multivariate Network). *A dynamic multivariate network* $\mathbf{G} = \{\mathbb{V}, \mathbb{E}, \mathfrak{F}\}$ *is an undirected connected graph, where* $\mathbb{V} = \{v_1, ..., v_N\}$ *is the set of vertices,* $\mathbb{E} \subseteq \mathbb{V} \times \mathbb{V}$ *refers to the set of edges (relations), and* $\mathfrak{F} = \{f^1, ..., f^T\}$ *is a family of feature vectors of the kind* $f^t : \mathbb{V} \to \mathbb{R}^D$ *which associates each vertex* $v \in \mathbb{V}$ *with a d-dimensional feature vector* $(f^t(v))$ *for each vertex v at time slice t, D refers to the total number of features, and* f^t *corresponds to a discrete time slice t.*

Through the paper, we consider T time slices Twitter as a case study, in which each vertex v refers to a Twitter user, and its d-th feature value (at time slice t) $f^t_d(v)$ refers to the frequency of a specific keyword in the tweets that are posed by this user at time slice t. For $f^t_d(v)$, we measure the significance of observing this feature value at time slice t as its statistical p-value, denoted as $p^t_d(v)$, according to its empirical distribution. The p-value $(p^t_d(v))$ is computed as the fraction of the historical observations of this feature in which an equal or higher value is observed [4, 5, 13]. Besides, we employ two-stage empirical p-value as the p-value of v at time slice t, denoted as $p^t(v)$. The nice theoretical property of two-stage empirical p-value has been discussed in [5]. Intuitively, the p-value is a measure of anomalousness within the range between 0 and 1: *the smaller the p-value of a feature value, the higher the degree of anomalousness of this feature value.* We prepare to define nonparametric statistics for evaluating the significance of p-values, which will be used to define the score functions used for measuring the degree of anomalousness of a subset of vertices and features.

Definition 2 (Evolving Subgraphs). *Given a dynamic multivariate network* $\mathbf{G}$*, the evolving subgraphs* $\Omega = \{G^1, ..., G^T\}$ *is sequence of subgraphs (each one in a separate time slice) of* $\mathbf{G}$ *that satisfies:*

- *Every subgraph is connected within its time slice;*

• *Two contiguous subgraphs share at least one vertex, e. g.,* $V^t \bigcap V^{t+1} \neq \varnothing$, $\forall t \in [0, T-1]$, *where* $V^t \in \mathbb{V}$ *is the set of vertices of* G^t *which denotes the projection of the evolving subgraphs at time slice* t.

Definition 3 (Nonparametric Statistics). *Given a set of p-values* S, *nonparametric statistics (also called aggregation functions of p-values) refer to a class of scoring functions* $\mathcal{F}(S)$ *that measure the joint significance of multiple p-values in* S *and have the general form:*

$$\mathcal{F}(S) = \phi(\alpha, N_\alpha(S), N(S)) \tag{1}$$

where α *is a predefined significance level of p-values;* $N_\alpha(S)$ *refers to the number of p-values in* S *that are less than or equal to* α*; and the function* $\phi(\alpha, N_\alpha, N)$ *satisfies two intuitive properties:*
• ϕ *is monotonically* ***increasing*** *with respect to (w.r.t.)* N_α*;*
• ϕ *is monotonically* ***decreasing*** *w.r.t.* α *and* N.

For the purpose of illustration, we explore to use the Berk-Jones (BJ) statistic [2], as this statistic has been shown effective in a number of real-world applications [5, 12]. It is defined as:

$$\phi_{\text{BJ}}(\alpha, N_\alpha(S), N(\Omega)) = N(S) \times \text{KL}(N_\alpha(S)/N(S), \alpha), \tag{2}$$

where $\text{KL}(\cdot)$ is the Kullback-Leibler divergence between the observed and expected proportions of p-values less than α. KL divergence is defined as:

$$\text{KL}(x, y) = x \log(x/y + (1-x) \log((1-x)/(1-y))). \tag{3}$$

The BJ statistic, which uses the KL divergence, can be interpreted as the log-likelihood ratio statistic for testing whether the empirical p-values follow a uniform or piecewise constant distribution. Berk and Jones demonstrated that this statistic fulfills several optimality properties and has greater power than any weighted Kolmogorov statistic [2].

3. DYNAMIC MULTIVARIATE EVOLVING ANOMALOUS SUBGRAPHS SCANNING

This section first generalizes nonparametric statistics and presents a new class of evolving multivariate subgraphs scan statistic functions for anomalous subgraphs detection and feature selection in dynamic multivariate networks, and then reformulate each function as a sequence of subproblems with provable guarantees.

3.1 Problem Formulation

Given a dynamic multivariate network $\mathbf{G} = \{\mathbb{V}, \mathbb{E}, \mathfrak{F}\}$, to find which evolving multivariate subgraphs are the most anomalous in $\mathbf{G}$, the general form of the nonparametric scan statistic in dynamic multivariate networks is defined as:

$$F(\Omega, \Theta) = \max_{\alpha \leq \alpha_{max}} \phi(\alpha, \psi(\Omega, \Theta, \alpha), N(\Omega) \cdot N(\Theta)) \\ s.t.\ \delta(\Omega) \leq B, N(\Omega) \leq K \tag{4}$$

where $\delta(\Omega) = \sum_{t=1}^{T-1} (|V^t| + |V^{t+1}| - 2|V^t \bigcap V^{t+1}|)$ refers to the total count of change of graph Ω in adjacent time slices; B is the upper bound of the change; Θ refers to subsets of attributes from T time slices (each subset in a separate time slice); $N(\Omega)$ and $N(\Theta)$ refer to the sizes of Ω and Θ, respectively; K is the upper bound of the number of vertices of Ω; $\psi(\Omega, \Theta, \alpha) = \sum_{v \in \Omega, d \in \Theta, t \in \{1,...,T\}} I(p_d^t(v) \leq \alpha)$ refers to the number of p-values among those related to Ω and Θ that are less than or equal to α, the predefined significance value; and the function ϕ is defined in Definition 3. The function $I(\cdot) = 1$ if its input is true; otherwise, $I(\cdot) = 0$. In this paper, we consider the evolving multivariate subgraphs scan statistic function $F_{\text{BJ}}(\Omega, \Theta)$ based on the BJ statistic (Equation (2)) as a case study. Our proposed techniques will be applicable to other multivariate subgraph scan statistic functions as well. The $F_{\text{BJ}}(\Omega, \Theta)$ is shown as:

$$F_{\text{BJ}}(\Omega, \Theta) = \max_{\alpha \leq \alpha_{max}} \phi_{\text{BJ}}(\alpha, \psi(\Omega, \Theta, \alpha), N(\Omega) \cdot N(\Theta)) \\ = \max_{\alpha \leq \alpha_{max}} N(\Omega) \cdot N(\Theta) \times \text{KL}(\frac{\psi(\Omega, \Theta, \alpha)}{N(\Omega) \cdot N(\Theta)}, \alpha) \cdot \\ s.\ t.\ \delta(\Omega) \leq B, N(\Omega) \leq K \tag{5}$$

Based on the nonparametric scan statistics, the detection of the most anomalous evolving subgraphs and features can be formalized as the following optimization problem:

$$\max_{\Omega, \Theta} \max_{\alpha \leq \alpha_{max}} \phi(\alpha, \psi(\Omega, \Theta, \alpha), N(\Omega) \cdot N(\Theta)) \\ s.\ t.\ \delta(\Omega) \leq B, N(\Omega) \leq K \tag{6}$$

which is equivalent to the problem:

$$\max_{\alpha \in \text{U}(\mathbb{V}^*, \alpha_{max})} \max_{\Omega, \Theta} \phi(\alpha, \psi(\Omega, \Theta, \alpha), N(\Omega) \cdot N(\Theta)) \\ s.\ t.\ \delta(\Omega) \leq B, N(\Omega) \leq K \tag{7}$$

where $\text{U}(\mathbb{V}^*, \alpha_{max})$ refers to the union of $\{\alpha_{max}\}$ and the set of distinct p-values no more than α in $\mathbb{V}^*$, $\mathbb{V}^* = \{(v, d) | v \in \mathbb{V}, d \in \{1, ..., D\}, t \in \{1, ..., T\}\}$ denotes the total number of combinations of vertices and features of the whole dynamic multivariate network.

3.2 Problem Reformulation

To analyze the nonparametric scan statistics problem in dynamic multivariate networks is very difficult as it involves a nonlinear objective function, and can not be reduced from the known NP-hard problems that often involve linear objective functions. What's more, the completion of the subgraphs detection and the feature selection concurrently makes the this task harder. Owing to the hardness of analyzing the aforementioned problem, we propose reformulating the nonparametric scan statistics problem, in dynamic multivariate networks, as a sequence of subproblems with provable guarantees. The reformulation is shown in Theorem 1.

Theorem 1 (Problem Reformulation). *Denote* $\bar{\psi}(\Omega, \Theta, \alpha) \equiv N(\Omega) \cdot N(\Theta) - \psi(\Omega, \Theta, \alpha)$, $(\Omega, \Theta, \alpha) = \{(v, d) | v \in \Omega, d \in \Theta, t \in \{1, ..., T\}\}$. *The Problem (7) is equivalent to the following problem:*

$$(a^*, \Omega^*, \Theta^*) = \operatorname*{argmax}_{\alpha \in \mathbf{C1}} \operatorname*{argmax}_{\psi \in \mathbf{C2}} \phi(\alpha, \psi(\Omega, \Theta, \alpha), N(\Omega) \cdot N(\Theta)) \\ s.\ t.\ \delta(\Omega) \leq B, N(\Omega) \leq K \tag{8}$$

where $\mathbf{C1} = \text{U}(\mathbb{V}^*, \alpha_{max})$, $\mathbf{C2} = \{\psi^0, ..., \psi^{\sum_{t=1}^{T} (N \cdot D)}\}$. *Each* $\psi^M \in \mathbf{C2}$ *refers to the solution to the following* M-budget *evolving subgraphs detection and feature selection problem:*

$$(\alpha, \Omega, \Theta)^M = \operatorname*{argmax}_{\Omega,\ \Theta} \psi(\Omega, \Theta, \alpha) \\ s.\ t.\ \delta(\Omega) \leq B, N(\Omega) \leq K, \bar{\psi}(\Omega, \Theta, \alpha) \leq M \tag{9}$$

Proof. Let $(\Omega, \Theta, \alpha)^- \equiv \{(v, d) |\ p_d^t(v) \leq \alpha, v \in \Omega, d \in \Theta, t \in \{1, ..., T\}\}$, $(\Omega, \Theta, \alpha)^+ \equiv \{(v, d) |\ p_d^t(v) > \alpha, v \in \Omega, d \in$

$\Theta, t \in \{1, ..., T\}\}$. Each feasible (Ω, Θ, α) can be decomposed into the subset of normal combination of vertices and features ($(\Omega, \Theta, \alpha)^+$) and the subset of abnormal combination of vertices and features ($(\Omega, \Theta, \alpha)^-$) satisfying the conditions: $\bar{\psi}(\Omega, \Theta, \alpha) \leq M$ and $(\Omega, \Theta, \alpha) = (\Omega, \Theta, \alpha)^+ \bigcup (\Omega, \Theta, \alpha)^-$. Suppose the tuple $(\alpha^*, \Omega^*, \Theta^*)$ is the optimal solution to the Problem (5), and $\bar{\psi}(\Omega^*, \Theta^*, \alpha^*) = m$, where $0 \leq m \leq N \times T \times D$. Then, it can be readily derived that $(\Omega, \Theta, \alpha^*)^m = (\Omega^*, \Theta^*, \alpha^*)$. Based on properties in Definition 3, there does not exist other $(\Omega, \Theta, \alpha)^M$, where $\alpha \neq \alpha^*$ or $M \neq m$, such that $\phi(\alpha, \psi((\Omega, \Theta, \alpha)^M), N((\Omega, \Theta, \alpha)^M)) > \phi(\alpha^*, \psi((\Omega, \Theta, \alpha^*)^m), N((\Omega, \Theta, \alpha^*)^m))$. Otherwise, this is in contradiction to the fact that $(a^*, \Omega^*, \Theta^*)$ is the optimal solution to Problem (8). □

4. APPROXIMATION ALGORITHMS

Owing to the hardness of conducting the subproblems mentioned in Theorem 1 under the constraints, we focus on finding an approximation solution instead of the exact solution for each subproblem (9) through the work: find a $\Omega^{'}$, a $\Theta^{'}$ and:

$$\psi(\Omega^{'}, \Theta^{'}, \alpha) \geqslant \mathcal{C} \cdot \max_{\Omega,\ \Theta} \psi(\Omega, \Theta, \alpha) \ s.t.\ \delta(\Omega) \leq B, N(\Omega) \leq K, \quad (10)$$

where $\mathcal{C} > 0$ is constant. And we provide a multiplicative guarantee for this approximation. Moreover, *Lagrangian relaxation, tree-shaped priors and dynamic programming* are mainly used for obtaining the best solution for Problem (10). The details of them are shown below.

4.1 Lagrangian Relaxation

The Lagrangian relation which guarantees $N(\Omega) \leq K$ is described as:

$$\begin{cases} \max\limits_{\Omega,\ \Theta} \psi(\Omega, \Theta, \alpha) \\ \max\limits_{\Omega} N_\alpha(\Omega) - \lambda N(\Omega) \end{cases} s.t.\ \delta(\Omega) \leq B, N(\Omega) \leq K, \quad (11)$$

where $N_\alpha(\Omega)$ is the number of vertices that maximize $\psi(\Omega, \Theta, \alpha)$, the parameter λ controls the trade-off between the approximation result and the number of vertices in Ω. Furthermore, the Problem (11) can be rewritten as:

$$\begin{cases} \max\limits_{\Omega,\ \Theta} \psi(\Omega, \Theta, \alpha) \\ \max \sum\limits_{v \in \Omega} (I(p^t(v) \leq \alpha) - \lambda) \end{cases} s.t.\ \delta(\Omega) \leq B, N(\Omega) \leq K. \quad (12)$$

What's more, for the sake of obtaining different number of vertices of Ω, we can obtain $0 \leq \lambda \leq \frac{N_\alpha(\Omega)}{N(\Omega)}$. For each $\lambda \in [0, \frac{N_\alpha(\Omega)}{N(\Omega)}]$, we also can seek out a corresponding $\lambda^{'}$ between 0 and α which makes Problem (12) equivalent to:

$$\begin{cases} \max\limits_{\Omega,\ \Theta} \psi(\Omega, \Theta, \alpha) \\ \max \sum\limits_{v \in \Omega} (I(p^t(v) + \lambda^{'} \leq \alpha)) \end{cases} s.t.\ \delta(\Omega) \leq B, N(\Omega) \leq K. \quad (13)$$

4.2 Tree-shaped Priors

To obtain efficient solutions for (11), we propose approximating all snapshot graphs (each one in a separate time slice) of $\mathbf{G}$ as the same tree Γ originating from the same given root vertex $\tau \in \mathbb{V}$ and the search of the best connected evolving subgraphs Ω and the anomalous subsets of features Θ for the nonparametric scan statistics problem is approximated as the search of the best sub-trees and the concurrent anomalous features in all Γ_τ (each one in a separate time slice). In order to obtain Γ, we first label abnormal vertices whose p-values are no more than α and normal vertices whose p-values are more than α as 1 and 0, respectively. If $p^t(v) \leq \alpha$, denote $l^t(v) = 1$; otherwise, $l^t(v) = 0$, where $l^t(v)$ is the label of vertex v at time slice t. Then we denote $L(v) = l^1(v) \vee l^2(v) \vee ... \vee l^T(v)$ as the label of vertex v. Specifically, if $L(v) = 0$, the vertex v is normal in all time.

Several heuristic approaches have been proposed to obtain the tree Γ based on vertex labels mentioned above, such as (1) *Breadth-first search tree (BFS-T)*, (2) *Random spanning tree (Random-T)*, (3) *Steiner tree (Steiner-T)*, (4) *Geodesic shortest path tree (Geodesic-SPT)*. The tree-shaped priors have been successfully applied to event detection based on graphs [8, 14, 16]. Based on the tree-shaped priors and Theorem 1, DMGraphScan is presented in Algorithm 1.

In the paper, ***Steiner-T*** is selected owing to its outstanding comprehensive performance [8, 14, 16]. Intuitively, a tree is good if abnormal vertices are interconnected with the least number of normal vertices. If we denote each abnormal vertex as a terminal vertex, and each normal vertex as a steiner vertex, this tree can be identified by generating the steiner tree of the input graph. The Steiner-T heuristic computes the steiner tree for each $\alpha \in \mathrm{U}(\mathbb{V}^*, \alpha_{max})$, computes the best sub-tree for each (9), and then returns the best solution.

Algorithm 1 DMGraphScan

Input: Dynamic multivariate network $\mathbf{G}$, $\mathbf{R} = 5$, $\alpha_{max} = 0.15$.
Output: The evolving anomalous subgraphs Ω^*, the subsets of features Θ^*.
for $\mathbf{r} \in \{1, ..., \mathbf{R}\}$ **do**;
  Select seed vertex τ from $\{v | v \in \mathbb{V}, p^t(v) \leq \alpha_{max}\}$;
  Approximate the graphs as the tree Γ_τ;
  for $\alpha \in \mathrm{U}(\mathbb{V}^*, \alpha_{max})$ **do**
    for $M = 0, ..., (N(\mathbb{V}^*) - \psi(\Omega, \Theta, \alpha))$ **do**
      $(\Omega, \Theta, \alpha)^M \leftarrow$ Algorithm 2 $(K, \alpha, c, \gamma, B)$;
    end for
  end for
  $(\Omega, \Theta, \alpha)^{\mathbf{r}} = \underset{(\Omega, \Theta, \alpha)^M}{\operatorname{argmax}} \phi(\alpha, \psi((\Omega, \Theta, \alpha)^M), N((\Omega, \Theta, \alpha)^M))$;
end for
Calculate $\mathbf{r}^* = \operatorname{argmax}_{\mathbf{r}} \Phi(\alpha, \psi((\Omega, \Theta, \alpha)^{\mathbf{r}}), N((\Omega, \Theta, \alpha)^{\mathbf{r}}))$;
Return $(\Omega, \Theta, \alpha)^{\mathbf{r}^*}$

4.3 Dynamic Programming

In order to get the most desired solution to the Problem (11), a dynamic programming (DP) algorithm (as shown in Figure 2) is designed when the dynamic network $\mathbf{G}$ is input in the form of trees Γ_τ with the root vertex τ. We first introduce a few notations:

- Γ_v: sub-trees of $\mathbf{G}$ with the same root vertex v, each sub-tree corresponding to a snapshot of $\mathbf{G}$ in a separate time slice .
- $\Omega_v = \{\Omega_{v^1}, \Omega_{v^2}, ..., \Omega_{v^B}\}$: candidate solutions of subgraphs to Γ_v, thereinto, $\Omega_{v^b} \in \Omega_v$ corresponds to $\delta(\Omega_v) = b$, where $b \in \{0, 1, ..., B\}$. Moreover, Ω_{v^b} owns the maximum $\psi(\Omega, \Theta, \alpha)$ under the constraint $\delta(\Omega_v) = b$.
- $\Theta_v = \{\Theta_{v^1}, \Theta_{v^2}, ..., \Theta_{v^B}\}$: the corresponding subsets of features to $\Omega_v = \{\Omega_{v^1}, \Omega_{v^2}, ..., \Omega_{v^B}\}$, respectively.

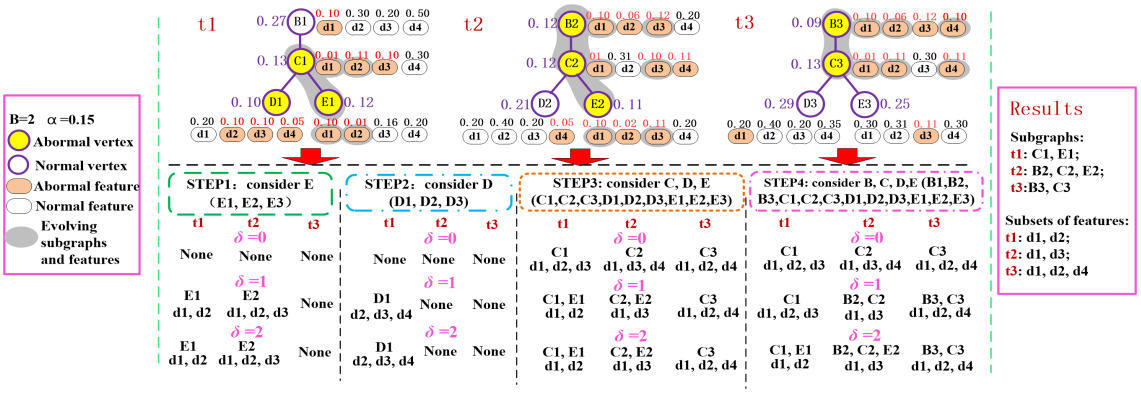


Figure 2: An illustration of dynamic programming for finding evolving anomalous subgraphs Ω and the subsets of features Θ concurrently in a dynamic multivariate network. In each iteration from leaf vertices to root vertices, we select one Ω_{v^b} and one Θ_{v^b} such that $\psi(\Omega, \Theta, \alpha)$ is maximized under the constraint $\delta(\Omega_v) = b$ from all kinds of combination of candidate solutions to v and the candidate solutions to every child of v.

- $p^{t\prime}(v)$: Updated p-value of vertex v at time slice t by $p^{t\prime}(v) = p^t(v) + \lambda^{'}$.
- $\pi^t(v)$: a value that indicates whether $p^{t\prime}(v) \leq 0$. If $p^{t\prime}(v) \leq \alpha$, set $\pi^t(v) = 1$; otherwise, set $\pi^t(v) = 0$.

The procedure of dynamic programming is shown below.

Leaf vertex. For a leaf vertex v, let $\delta(v) = \sum_{t=1}^{T-1} |\pi^t(v) - \pi^{t+1}(v)| = 0, 1, ..., B$, respectively. $\delta(v)$ which is equal to $\delta(\Omega_v)$ denotes the change of vertex v in T time slices when v is the leaf vertex. Then, by maximizing $\psi(\Omega, \Theta, \alpha)$ under different change constrains, we can obtain the corresponding candidate solutions: Ω_{v^1}, Ω_{v^2},..., Ω_{v^B} and Θ_{v^1}, Θ_{v^2},..., Θ_{v^B}.

Non-leaf vertex. Let $v_{child} = \{v_{c1}, v_{c2}, ..., v_{cw}\}$ refer to the w child vertices of the vertex v. Each child $v_{ci} \in v_{child}$ owns $\Omega_{v_{ci}}$ and v its own candidate solutions including the subgraphs and the subsets of features, where $i \in \{1, 2, ..., w\}$. Finding each of the candidate solutions of vertices to Γ_v can be reduced to an approximation 0-1 multiple-choice knapsack combinatorial optimization problem from the vertex v and v_{child}. The problem is to select a Ω_{v^b} and a Θ_{v^b} such that $\psi(\Omega, \Theta, \alpha)$ is maximized under the constraint $\delta(\Omega_v) = b$ from all kinds of combination of candidate solutions to v and the candidate solutions to v_{child} (e. g., $b = 1$, the available combinations of $\{\delta(v), \delta(\Omega_{v_{c1}}), ..., \Omega_{\delta(v_{cw})}\}$ include: $\{1, 0, ..., 0\}$, $\{0, 1, ..., 0\}$,..., $\{0, 0, ..., 1\}$). Then all candidate solutions Ω_v and Θ_v of Γ_v can be found. From the leaf vertices to root vertex τ, we can obtain all candidate solutions Ω_τ and Θ_τ of Γ_τ. Finally, we get the solution:

$$(\Omega^{'}, \Theta^{'}) = \underset{\Omega_{\tau^b}, \Theta_{\tau^b}}{\operatorname{argmax}} \psi(\Omega_{\tau^b}, \Theta_{\tau^b}, \alpha) \;\; s.t.\; b \in \{0, 1, ..., B\}. \quad (14)$$

Theorem 2. *The dynamic programming algorithm is guaranteed to obtain a local maximum solution to Problem (9). and the dynamic programming algorithm has the time complexity $O(NTB(B+D))$, where N is the number of vertices at single time slice, T is the total time slices, B is the upper bound of changes of Ω in T time slices, and D is the number of features of each vertex.*

Proof. 1) As we update the B candidate solutions including the vertices in Ω and the features in Θ in each iteration, we find an optimum solution that maximizes $\psi(\Omega, \Theta, \alpha)$ in each iteration under the different change times ($b \in \{1, ..., B\}$) of Ω. Suppose (Ω^*, Θ^*) is the optimum solution for the Problem (9) under the change constraint b, if an vertex v is added or delated, the number of changes will not be b and the subset of features Θ may change; if a feature is added or delated, the subset of features is changed, the Ω is changed as well, as a result, the number of changes will not be b. These are in contradiction to the fact that (Ω^*, Θ^*) is not the optimal solution to Problem (9) under the change limit b. 2) About the time complexity of the dynamic programming algorithm, for every vertex v in the whole dynamic multivariate network, we need to find B kinds of solutions from v and its child vertices that each of them owns B candidate solutions. And, for each solution under the constraint $b \in \{1, ..., B\}$, we select the same subset of features from $(B+1)$ solutions where each contains D features. Moreover, every vertex spans T time slices. Hence, the global time complexity is $O(NTB(B+D))$. □

4.4 Approximation Solutions

Since we only gives indirect control over the cardinality constraint of the number of the evolving subgraphs, namely $N(\Omega) \leq K$, based on the Lagrangian relaxation in Section 4.1, we then perform Algorithm 2 ,named as ApproAlg, over $\lambda^{'}$ to find a suitable value in this section. The results are shown in Theorem 3.

Theorem 3. *Let Ω denotes the evolving subgraphs that potentially spans T time slices. Moreover, let η, $\gamma > 0$. Then Algorithm 2 returns a solution of evolving subgraphs satisfying:*

$$N_\alpha(\Omega^{'}) \geq (\frac{\alpha K\eta - \gamma\eta + \gamma}{\alpha K\eta - \alpha K}) \max_{\Omega} N_\alpha(\Omega). \quad (15)$$

Proof. Let Ω_K be the solution with $N_\alpha(\Omega_K) = \max_{\Omega} N_\alpha(\Omega)$. Let Ω_l and Ω_r be the solutions corresponding to λ_l and λ_r, respectively. We maintain two invariants $K_r \geq \eta \cdot K$ and $K_l < K$ in the iterations. The invariants also hold before the first iteration in Algorithm 2 due to our initial choices for λ_l and λ_r. From the dynamic programming, we can get a $\Omega^{'}$ satisfying $N_\alpha(\Omega^{'}) - \lambda N(\Omega^{'}) = \max_{\Omega} N_\alpha(\Omega) - \lambda N(\Omega)$

and:

$$\begin{aligned} N_\alpha(\Omega_r) - \lambda_r N(\Omega_r) &\geq N_\alpha(\Omega_K) - \lambda_r N(\Omega_K) \\ \lambda_r &\leq (1/(K - K\eta))N_\alpha(\Omega_K) \end{aligned}. \quad (16)$$

At the end of iterations, we have $\lambda_l^{'} - \lambda_r^{'} \leq \varepsilon$, which can be approximated as $\lambda_l - \lambda_r \leq \frac{\varepsilon \cdot N_\alpha(\Omega_K)}{\alpha \cdot K}$. Then we get:

$$\lambda_l \leq (\frac{\varepsilon K(1-\eta) + \alpha}{\alpha K(1-\eta)})N_\alpha(\Omega_K). \quad (17)$$

Employing the dynamic programming, we also obtain:

$$\begin{aligned} N_\alpha(\Omega_l) - \lambda_l N(\Omega_l) &\geq N_\alpha(\Omega_K) - \lambda_l N(\Omega_K) \\ N_\alpha(\Omega_l) &\geq N_\alpha(\Omega_K) + \lambda_l(N(\Omega_l) - N(\Omega_K)). \\ &\geq N_\alpha(\Omega_K) + \lambda_l(-K) \end{aligned} \quad (18)$$

Combine (17) with (18):

$$\begin{aligned} N_\alpha(\Omega_l) &\geq N_\alpha(\Omega_K) + (\frac{\varepsilon K(1-\eta) + \alpha}{\alpha K(1-\eta)})N_\alpha(\Omega_K)(-K) \\ &= N_\alpha(\Omega_K)(\frac{\alpha K\eta - \gamma\eta + \gamma}{\alpha K\eta - \alpha K}) \end{aligned}. \quad (19)$$

To sum up, (15) can be obtained. □

Algorithm 2 ApproAlg

Input: K, α, η, γ, B.
Output: Optimal evolving subgraphs to Problem (10).
if there is a $\Omega^{'}$ with $N(\Omega^{'}) \leq K$ and $\delta(\Omega^{'}) \leq B$:
  return $\Omega^{'}$;
$\lambda_l^{'} \leftarrow \alpha$, $\lambda_r^{'} \leftarrow 0$, $\varepsilon \leftarrow \frac{\gamma}{K}$;
while $\lambda_l^{'} - \lambda_r^{'} > \varepsilon$ **do**
  $\lambda_m^{'} = (\lambda_l^{'} - \lambda_r^{'})/2$, $\Omega^{'} \leftarrow \mathrm{DP}(p, \lambda_m^{'}, \alpha, B, \tau, \Gamma)$;
  if $N(\Omega^{'}) \geq K$ and $N(\Omega^{'}) \leq \eta \cdot K$ **then return** $\Omega^{'}$
  if $N(\Omega^{'}) > \eta \cdot K$ **then** $\lambda_r^{'} \leftarrow \lambda_m^{'}$ **else** $\lambda_l^{'} \leftarrow \lambda_m^{'}$.
end while
Return $\Omega^{'} \leftarrow \mathrm{DP}(p, \lambda_l^{'}, \alpha, B, \tau, \Gamma)$

5. EXPERIMENTS

This section evaluates the effectiveness and efficiency of the proposed DMGraphScan framework based on two real-world datasets. Compared with other proposed techniques, DMGraphScan outperforms in both subgraph detection and feature selection.

Datasets: We consider the detection and forecasting of haze and flu outbreak events as two case study scenarios.

1) Flu outbreak dataset. We randomly collected ten percent of all the raw Twitter data from Jan 1, 2011 to May 1, 2015 (totally 226 weeks) in the United States. From this dataset, we selected 0.16 million tweets such that each tweet contains at least two terms from a set of 72 terms relevant to flu outbreaks collected from domain experts, which are posted by 39,565 users. According to co-mentions in tweets and following relations, we construct a connected user-user network with 49,204 edges. Each user is geocoded with a province from location in profiles. For each day d and user u, we calculated the corresponding empirical p-value for each keyword. In total, we have 226 snapshot graphs, corresponding to the 226 weeks. Golden Standard Reports (GSR) of 2,260 official flu outbreak records (ILI $\geq$ 2000) were collected from official website (`http://www.cdc.gov/flu/weekly/.`)

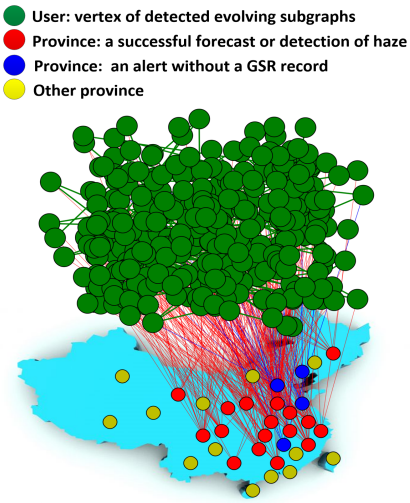
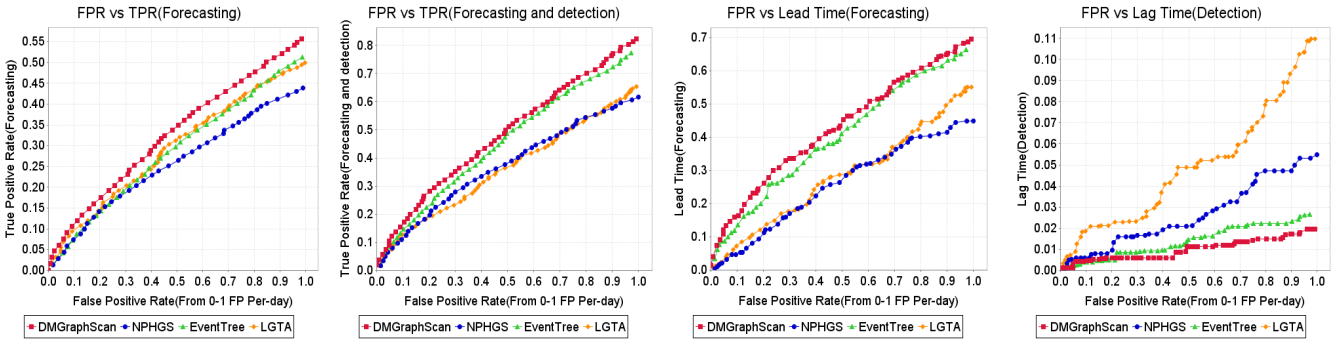


Figure 3: An haze event from Dec. 24, 2014 in China. We transform the anomalous subgraphs to alerts of states (provinces of China). The green vertices are the users of detected evolving subgraphs. The red and blue lines indicate the affiliation between users and the provinces. Within the 7 day window before and after that day, a red vertex refers to a successful forecast or detection; a blue vertex indicates an alert without a GSR record, a yellow vertex refers to a province that there is not an alert.

that is maintained by Centers for Disease Control and Prevention (CDC). CDC publishes the weekly influenza-like illness (ILI) activity level for each state in the United States based on the proportional level of outpatient visits to health care providers for ILI (influenza-like-illness). An example of a CDC flu outbreak event is: (STATE ="Virginia", COUNTRY = "United States", WEEK = "01-06-2013 to 01-12-2013").*For the haze dataset, the time unit is "day", but for the flu outbreak dataset, the time unit is "week", because CDC reports flu outbreaks on a week interval.*

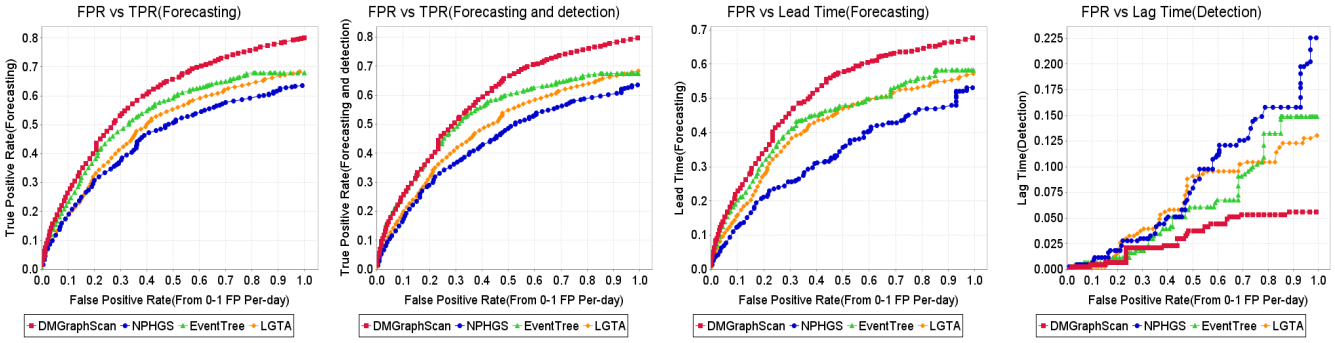
2) Haze dataset. We randomly collected 10 percent of the whole Weibo data from Apr 11, 2014 to Jan 11, 2015, including 1,433,937,815 tweets in total. After removing tweets that contain less than two terms from a dictionary of 68 terms about haze outbreaks collected from domain experts, we obtained 0.35 million tweets that were posted by 49,644 users. According to co-mentions in tweets and following relations, we construct a connected user-user network with 149,408 edges. Each user was geocoded with a province from location in profiles. For each day d and user u, we calculated the corresponding empirical p-value for each keyword using the strategy proposed in [5]. In total, we have 276 snapshot graphs, corresponding to the 276 days. Gold Standard Reports (GSR) of 9,384 official haze outbreak records (level $\geq$ 3) were collected from official websites (MEP), and an example of the GSR record is (Province = "Beijing", COUNTRY = "China", DAY = "11-04-2014").

Comparison Methods: We compared our proposed approach, named as DMGraphScan, with three existing representative baseline methods, including Non-Parametric Heterogeneous Graph Scan (NPHGS) [5], EventTree [14], and Latent Geographical Topic Analysis (LGTA) [17]. We strictly



(a) FPR vs TPR(Fore.) (b) FPR vs TPR(Fore. & Dete.) (c) FPR vs Lead Time(Fore.) (d) FPR vs Lag Time(Dete.)

Figure 4: The comparison between DMGraphScan and baseline methods based on haze dataset.



(a) FPR vs TPR(Fore.) (b) FPR vs TPR(Fore. & Dete.) (c) FPR vs Lead Time(Fore.) (d) FPR vs Lag Time(Dete.)

Figure 5: The comparison between DMGraphScan and baseline methods based on the flu outbreak dataset.

followed strategies recommended by authors in their papers to tune the related model parameters. Specifically, for EventTree, the set of λ values $\{0.1, 0.2, \cdots, 1.0, 50, 100, \cdots, 1500\}$ is tested.

Our Proposed DMGraphScan Algorithm: In this paper, our proposed algorithm is denoted as DMGraphScan. We employ 10-fold cross validation to identify the best combination of all the related parameters. Specifically, the parameter α_{max} is denoted as 0.15.

Performance Metrics: This paper focuses on the evaluation of both event detection and forecasting for different methods. The related metrics include:

- 1) False positive rate (FPR);
- 2) True positive rate (TPR) for forecasting;
- 3) True positive rate for both detection and forecasting;
- 4) Average lead time for forecasting;
- 5) Average lag time for detection.

For each method, the reported alerts are structured as tuples of (date, location), where "location" is defined at the province level. For each GSR event, we decide whether the method:

- 1) Had an alert in the province within 7 days before the event, which means to be "predicted";
- 2) Had an alert in the province within 7 days after the event, which means to be "detected";
- 3) Had no alert in the province within 7 days before or after the event, which is "undetected".

Transformation of Anomalous Subgraphs to Alerts: For each time slice, each approach will output a detected user subgraph with an anomalousness score (the value of the objective function is maximized). A set of places are retrieved from the geocoded places of the users within this subgraph, within which each place leads to an alert with the place name, time slice, and an anomalousness score. As shown in Figure 3, a haze event detection from Dec. 24, 2014 in China, the green vertices consist of the subgraph, the yellow, red and blue vertices are the transformed provinces of China. What's more, a red vertex refers to a successful forecast or detection, a blue vertex indicates an alert without a GSR record. The main reason of the few deviations is that there are few negative posts. Such as active users may discuss the hot events happened in their adjacent areas.

The Results of Event Detection and Forecasting

The results of the comparison between the proposed DMGraphScan approach and three baseline methods are shown in Figure 4 and Figure 5. And an example of the results of DMGraphScan is shown in Figure 6. The figures 4 and 5 show that the comparison at various false positive rates (FPR) for the target of detection and forecasting haze and flu outbreak events. The results indicate that DMGraphScan obtained much higher forecasting TPR, and much higher forecasting and detection TPR than all the baseline methods, and there is a trend that when the FPR increases, the margin between the TPR of DMGraphScan and those of baseline methods consistently increases for both forecasting and detection. Specifically, based on flu dataset, the margin for forecasting is more than 10% shown in Figure 5(a), and the margin for detection and forecasting is more than 10% shown in Figure 5(b). In addition, DMGraphScan ob-

Table 1: Comparison between the keywords selected by methods DMGraphScan and LGTA on the haze and flu outbreak datasets.

Events		DMGraphScan			LGTA				
Haze	Event1	污染(pollution)	预警(warning)	空气(air)	浑浊(turbid)	朦胧(obscure)	黄色(yellow)	中度(middle level)	污染(pollution)
		黄色(yellow)	严重(serious)	雾霾(haze)	能见度(visibility)	健康(health)	污染源(pollution sources)	大雾(fog)	灰霾(haze)
		口罩(mask)							
	Event2	健康(health)	质量(quality)	雾霾(haze)	大雾(fog)	鞭炮(firecracker)	减排(emission reduction)	健康(health)	净化器(purifier)
		污染(pollution)	重度(serious)	空气(air)	雾霾(haze)	肺癌(lung cancer)	环保(environmental protection)	PM10	污染(pollution)
	Event3	灰霾(haze)	健康(health)	呼吸(breathe)	肺癌(lung cancer)	智力(intelligence)	超标(exceed standard)	天空(sky)	健康(health)
		严重(serious)	环境(environment)	肺癌(lung cancer)	灰霾(haze)	能见度(visibility)	环保(environmental protection)	污染(pollution)	感冒(influenza)
Flu	Event1	flu	fever	cold	flu	cold	virus	cough	stomach
		ache	stomach	headache	sleep	runny	sneeze	fever	head
	Event2	flu	cough	ache	flu	cold	cough	fever	stomach
		headche	infection	asthma	virus	sneeze	heart	tired	head
	Event3	flu	head	cold	flu	cold	stomach	runny	head
		sleep			tired	medicine	cough	heart	hurt

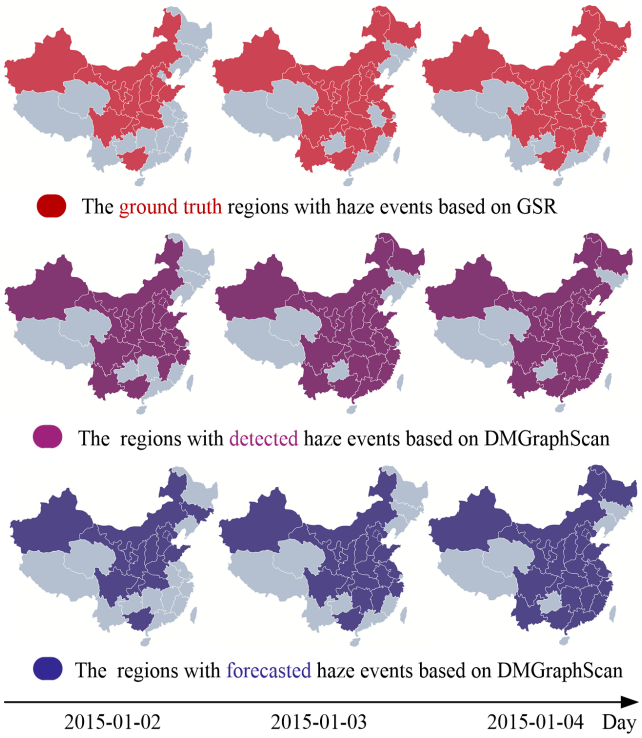


Figure 6: An illustration of the comparison between forecasted, detected alert results based on DMGraphScan and the ground truth of haze events from 2015-01-02 to 2015-01-04 in China. The first line, the second line and third line refer to the ground truth, the forecasted results and the detected results of haze events, respectively. Where each separate region prefers to a state (province of China). On a certain day, an alert of the province within the 7 days window before and after that day.

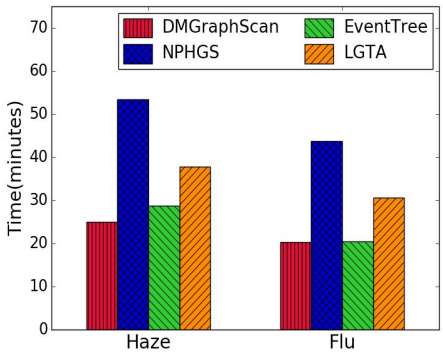


Figure 7: Average run time of each method based on haze and flu outbreak datasets.

tained longer Lead Time, and shorter Lag Time than all the baseline methods at various FPR.

Among these baseline methods, only the method of LGTA is designed for feature selection and subgraph detection concurrently. Nevertheless, this method performs worse than the baseline method EventTree that only conducte subgraph detection but perform the second best on all the metrics. Although the method LGTA has considered feature selection and the subgraph detection concurrently, their strategies do not perform well on the quality of features and subgraph that are identified.

The average runtime of each of methods, including DMGraphScan and the three baseline methods, is presented in Figure 7 based on both the flu outbreak and haze datasets. The results show that DMGraphScan is faster than the other methods on the two real-world datasets.

The Results of Feature Selection

The results of feature selection are shown in Table 1. Table 1 presents the features (keywords) that are selected by DMGraphScan and the 10 highest probable keywords by LGTA for each of the six example GSR events, respectively. First, the results indicate that the number of keywords selected by DMGraphScan is much less than that selected by LGTA in each of the six event examples, and vary in different events. We find that DMGraphScan is capable to select

a number of features, which is different from most existing approaches where a fixed number of features need to be predefined, the LGTA method included. Second, the keywords selected by both DMGraphScan and LGTA methods overlap for a small subset, which could potentially represent the set of core keywords that are related to these events. However, the keywords detected by both methods are still significantly different for all the events. As DMGraphScan performs much better than LGTA in both the datasets in any measurement methods as discussed in the above subsection, we conclude that DMGraphScan is able to identify a small number of signal keywords that are more effective than those detected by LGTA for event detection and forecasting.

The quality of the keywords identified by DMGraphScan is illustrated using Event 1 for haze event detection. The date of Event 1 is Oct. 9, 2014, and there is a corresponding news article that was published in the same day that reported the haze event: "中新网北京10月9日电(BEIJING, Oct. 9, 2014 (Chinanews).), 受不利气象条件影响(Affected by the adverse weather conditions,), 北京民众再次饱受雾霾之苦(Beijing's residents suffered from the haze again.), 全城空气质量维持在严重污染级别(The air quality of the whole city remained at the level of serious pollution.)。北京于9日先后升级发布霾橙色、空气重污染橙色预警(Orange alert for haze and orange alert for serious air pollution were issued successively today.)。预计本次雾霾过程将持续至11日((This haze was expected to last until Otc. 11, 2014.)。...", As shown in this news article, five of the seven selected keywords were mentioned, including "预警(warning)", "空气(air)", "严重(serious)", "污染(pollution)", "雾霾(haze)", where the first three keywords were not identified by LGTA.

6. CONCLUSION AND FUTURE WORK

A generic approach, named as DMGraphScan, is proposed for solving the problem of dynamic multivariate anomalous subgraph detection in this paper. DMGraphScan performs significantly better than several state-of-the-art approaches on the real-world haze and flu outbreak datasets. For the future work, we plan to extend DMGraphScan to detect evolving anomalous subgraphs in dynamic multivariate and heterogeneous networks, where the vertices or edges may have different types and evolve over time.

7. ACKNOWLEDGMENTS

The corresponding author is Jianxin Li. This work is supported by NSFC program (No.61472022, 61421003), China 973 program (No. 2014CB340300),SKLSDE-2016ZX-11 and partly by the Beijing Advanced Innovation Center for Big Data and Brain Computing.

8. REFERENCES

[1] L. Akoglu, H. Tong, and D. Koutra. Graph based anomaly detection and description: a survey. *Data Mining and Knowledge Discovery*, 29(3):626–688, 2015.

[2] R. H. Berk and D. H. Jones. Goodness-of-fit test statistics that dominate the kolmogorov statistics. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 47(1):47–59, 1979.

[3] H. S. Burkom. Biosurveillance applying scan statistics with multiple, disparate data sources. *Journal of Urban Health*, 80(1):i57–i65, 2003.

[4] F. Chen and D. B. Neill. Non-parametric scan statistics for disease outbreak detection on twitter. *Online journal of public health informatics*, 6(1):e155, 2014a.

[5] F. Chen and D. B. Neill. Non-parametric scan statistics for event detection and forecasting in heterogeneous social media graphs. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1166–1175, 2014b.

[6] F. Chen and D. B. Neill. Human rights event detection from heterogeneous social media graphs. *Big Data*, 3(1):34–40, 2015.

[7] X. Dong, D. Mavroeidis, F. Calabrese, and P. Frossard. Multiscale event detection in social media. *Data Mining and Knowledge Discovery*, 29(5):1374–1405, 2015.

[8] A. Gionis, M. Mathioudakis, and A. Ukkonen. Bump hunting in the dark: Local discrepancy maximization on graphs. *IEEE Transactions on Knowledge and Data Engineering*, 2016.

[9] M. Kulldorff, F. Mostashari, L. Duczmal, W. Katherine Yih, K. Kleinman, and R. Platt. Multivariate scan statistics for disease surveillance. *Statistics in medicine*, 26(8):1824–1833, 2007.

[10] T. Lappas, M. R. Vieira, D. Gunopulos, and V. J. Tsotras. On the spatiotemporal burstiness of terms. *Proceedings of the VLDB Endowment*, 5(9):836–847, 2012.

[11] J. Li, J. Wen, Z. Tai, R. Zhang, and W. Yu. Bursty event detection from microblog: a distributed and incremental approach. *Concurrency and Computation: Practice and Experience*, 2015.

[12] E. McFowland, S. Speakman, and D. B. Neill. Fast generalized subset scan for anomalous pattern detection. *Journal of Machine Learning Research*, 14(1):1533–1561, 2013.

[13] M. Mongiovi, P. Bogdanov, R. Ranca, E. E. Papalexakis, C. Faloutsos, and A. K. Singh. Netspot: Spotting significant anomalous regions on dynamic networks. In *Proceedings of the 2013 SIAM International Conference on Data Mining*, pages 28–36, 2013.

[14] P. Rozenshtein, A. Anagnostopoulos, A. Gionis, and N. Tatti. Event detection in activity networks. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1176–1185, 2014.

[15] T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World Wide Web*, pages 851–860, 2010.

[16] N. Wu, F. Chen, J. Li, B. Zhou, and N. Ramakrishnan. Efficient nonparametric subgraph detection using tree shaped priors. In *AAAI*, pages 1352–1358, 2016.

[17] Z. Yin, L. Cao, J. Han, C. Zhai, and T. Huang. Geographical topic discovery and comparison. In *Proceedings of the 20th international conference on World Wide Web*, pages 247–256, 2011.