Uncovering Specific-Shape Graph Anomalies in Attributed Graphs

Nannan Wu,[*] Wenjun Wang,[*] Feng Chen,[†] Jianxin Li,[‡] Bo Li,[‡] Jinpeng Huai[‡]
[*]College of Intelligence and Computing, Tianjin University, Tianjin 300072, China
[†]Dept. of Computer Science, University at Albany, SUNY, Albany, NY 12203
[‡]Dept. of Computer Science & Engineering, Beihang University, Beijing 100191, China
{wunannan, lijx, libo}@act.buaa.edu.cn wwj@pku.org.cn fchen5@albany.edu huaijp@buaa.edu.cn

Abstract

As networks are ubiquitous in the modern era, point anomalies have been changed to graph anomalies in terms of anomaly shapes. However, the *specific-shape* priors about anomalous subgraphs of interest are seldom considered by the traditional approaches when detecting the subgraphs in attributed graphs (e.g., computer networks, Bitcoin networks, and etc.). This paper proposes a nonlinear approach to *specific-shape* graph anomaly detection. The nonlinear approach focuses on optimizing a broad class of nonlinear cost functions via *specific-shape* constraints in attributed graphs. Our approach can be used to many different graph anomaly settings. The traditional approaches can only support linear cost functions (e.g., an aggregation function for the summation of node weights). However, our approach can employ more powerful nonlinear cost functions, and enjoys a rigorous theoretical guarantee on the near-optimal solution with the geometrical convergence rate.

Introduction

In numerous network applications, the computer network data of interest consist of *"star-shape"* attacking subgraphs (Wu et al. 2017), and the political blog data of interest consist of *"core-periphery"* graph shape anomalies (Zhang, Martin, and Newman 2015). Anomalies appear in the real network applications in the form of *specific-shapes* rather than point shapes. In the cyber attack detection applications in Figure 1, the deep node (i.e., computer) color represents the higher "transfer rate". Given the *"star-shape"* anomaly query in Figure 1, we formulate the following specific-shape constrained minimization problem to uncover the *specific shape* attack subgraph anomalies:

$$\min_{\mathbf{x}\in\mathbb{R}^n} \varphi(\mathbf{x}) \qquad s.t. \quad supp(\mathbf{x}) \in \mathcal{M}(\mathbb{Q}). \tag{1}$$

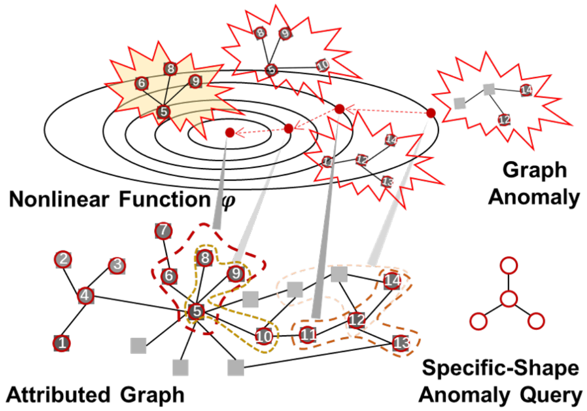


Figure 1: The main idea of uncovering *specific-shape* graph anomalies by the powerful nonlinear function within the continuous space (Akoglu, Tong, and Koutra 2015). In the attributed graph, abnormal vertices come with red circles.

where $\varphi : \mathbb{R}^n \mapsto \mathbb{R}$ is a powerful nonlinear cost function. An attributed graph $\mathbb{G} = (V, E, \mathbf{W})$ is comprised of a set of n vertices $V = [v] = \{1, \cdots, n\}$, an edge set $E \subseteq V \times V$, and an attribute matrix $\mathbf{W} \in \mathbb{R}^{n\times p}$ (p, the number of attributes). Given the *specific-shape* anomalous graph prior $\mathbb{Q}$, let $V_{\mathbb{Q}}$ denote the vertex set of $\mathbb{Q}$, and $\mathcal{M}(\mathbb{Q}) := \{V_{\mathbb{C}} \mid \mathbb{C} \subseteq \mathbb{G}, \mathbb{C} \cong \mathbb{Q}\}$ be the family set of "*specific-shape*" vertex sets, whose graphs $\{\mathbb{C}\}$ are subgraph of the attributed graph $\mathbb{G}$, and are isomorphic to the *specific-shape* anomalous graph prior $\mathbb{Q}$. The basic intuition for Problem (1) is that the cost function φ is minimized at the constrained $\mathbf{x}$ whose nonzero entry index set $supp(\mathbf{x}) := \{v \mid \mathbf{x}_v \neq 0, v \in V\}$ corresponds to the *specific-shape* vertex set $V_{\mathbb{C}}$ from $\mathcal{M}(\mathbb{Q})$.

To motivate this problem, in an attributed computer network $\mathbb{G}$, we represent the attribute matrix $\mathbf{W}$ as "*transfer rate*" attributes. Let $\mathbf{W}_v := \mathbf{W}^T_{v,:}$ be the p-dimensional attribute vector for the vertex $v \in V$ in the attributed graph $\mathbb{G}$, and $\mathbf{W}(i) := \mathbf{W}_{:,i}$ be the n-dimensional attribute vector for the attribute $i \in [p]$. Let $\mathbf{W}(0)$ denote as the vector $\mathbf{w}$. Problem (1) for identifying cyber attacking scenarios can be formulated as the least square problem: $\min_{\mathbf{x}\in\mathbb{R}^n} ||\mathbf{w} - \mathbf{x}||_2^2$, subject to $supp(\mathbf{x}) \in \mathcal{M}(\mathbb{Q})$ (Chen and Zhou 2016).

Related work. Point anomalies (i.e., outliers) can be considered as a binary 0/1 classification problem (Akoglu, Tong, and Koutra 2015). These approaches just assign what is called an outlierness score to each node. The *specific-shape* anomaly is not considered in these approaches. A number of approaches are proposed specifically for the graph anomaly detection problem. They can be directly classified as vector-based and graph-based approaches from the

Table 1: Comparison between our work and several representative works in graph anomaly detection. The nonzero entry index set of vector $\mathbf{x}$ implies a graph anomaly in the attributed graph, and $s = |V_{\mathbb{Q}}|$.

Related Work	Target Solution	Anomaly Shape	Nonlinear Cost Function
(Hegde, Indyk, and Schmidt 2015)	s-sparse signal $\|\|\mathbf{x}\|\|_0 \leq s$	Connected subgraph anomaly	$\surd$
(Chen and Zhou 2016)	s-sparse signal $\|\|\mathbf{x}\|\|_0 \leq s$	Connected subgraph anomaly	$\surd$
(Gupta et al. 2014)	Subgraph $\mathbb{C}$	*Specific-shape* graph anomaly	$\times$
(Yang et al. 2016)	Subgraph $\mathbb{C}$	*Specific-shape* graph anomaly	$\times$
(Wu et al. 2017)	s-sparse signal $\|\|\mathbf{x}\|\|_0 \leq s$	Tree-shape graph anomaly	$\surd$
Our work	s-sparse signal $\|\|\mathbf{x}\|\|_0 \leq s$	*Specific-shape* graph anomaly	$\surd$

target solution (e.g., in Table 1). In the vector-based approaches, a graph anomaly is considered as a sparse signal, and the true signal is recovered by the nonlinear cost functions (e.g., the norm-2 function). Graph-structured matching pursuit (Chen and Zhou 2016) and Graph-CoSaMP (Hegde, Indyk, and Schmidt 2015) approaches are proposed to connected subgraph anomaly detection by general nonlinear functions. The graph tree projection pursuit approach (Wu et al. 2017) can employ the nonlinear cost functions, such as Kulldorff (Kulldorff 1997) and Expectation-based Poisson (EBP) graph scan statistics (Neill 2009b), to the tree shape graph anomaly detection. The *specific-shape* prior can be used to the powerful nonlinear cost function for anomaly detection in attributed graphs with network structures and vertex attributes. In the graph-based approaches, a graph anomaly is considered as a *matching subgraph* in attributed graphs, and the index is built on vertices and edges. Most traditional approaches (Cordella et al. 2004; Huan, Wang, and Prins 2003; Gupta et al. 2014; Yang et al. 2016; 2014; Zou, Chen, and Lu 2007) can employ just linear cost functions for the index structure. Those approaches can not handle the nonlinear anomalies exhibited in attributed graphs.

In Table 1, our work can employ more powerful nonlinear cost functions than the graph-based approaches (Gupta et al. 2014; Yang et al. 2016). The other vector-based approaches (Hegde, Indyk, and Schmidt 2015; Chen and Zhou 2016) consider just the connected subgraph anomaly without the *specific shape* anomaly prior. Although the work (Wu et al. 2017) considers the tree-shape graph anomaly, the other complex *specific shapes* can not be handled in this approach.

Our work aims to develop a nonlinear approach to *specific shape* sparsity anomaly structure via attributed graphs. Our approach has three main features: (a) *Generality*: the approach can encompass several previously studied subgraph matching methods and projection-pursuit oracles (e.g., model-projection algorithms). (b) *Theoretical basis*: the approach achieves the near-optimal solution at a constant error bound with the geometrically convergence rate. (c) *Computational efficiency*: we present a near-linear time algorithm for Problem (1). The main contributions of our work are summarized as follows:

- **Develop an efficient nonlinear approach to graph anomalies.** Given the *specific shape* anomaly query graph $\mathbb{Q}$, a new and efficient approach, namely, Query-based **ma**tching **p**ursuit (Query-map), is developed to optimize a nonlinear function over the *specific shape* graph-structured sparsity model $\mathcal{M}(\mathbb{Q})$.
- **Theoretical basis.** The proposed approach, Query-map, achieves the near-optimal solution at a constant error bound with the geometrically linear convergence rate. The cost function φ satisfies a weaker condition than the popular strong condition such as Restricted Strong Convexity/Smoothness (RSC/RSS).
- **Comprehensive experiments to verify the proposed approach on the real datasets.** The powerful graph scan statistic nonlinear functions are employed in our approach, Query-map, for the task of *specific shape* graph anomaly detection. The extensive experiments on the real datasets show that Query-map performs competitively with a variety of representative methods for the graph anomaly detection.

This work aims to study the graph anomaly with a *specific shape* prior. For the tree shape prior, (Wu et al. 2017) work can be employed. For the connected subgraph prior, (Chen and Zhou 2016), and (Hegde, Indyk, and Schmidt 2015) works can be employed. Without any graph anomaly priors, our method can not guarantee the near-optimal solution.

Query-based *ma*tching *p*ursuit algorithm (Query-map)

The *specific-shape* graph anomaly in attributed graphs can be considered as a ***sparse signal*** (Chen and Zhou 2016). In referring to matching pursuit techniques in compressive sensing works, the main idea in this paper is summarized as: a) from the gradient, getting the most *specific-shape* anomalous vertex set $A := \arg\max_{A \in \mathcal{M}(\mathbb{Q})} \parallel \nabla_A \varphi \parallel_1$ (e.g., $2||\mathbf{x} - \mathbf{w}||_1$, i.e., aggregation function) where $(\nabla_A \varphi)_v = (\nabla \varphi)_v$ for $v \in A$, and $(\nabla_A \varphi)_v = 0$ otherwise; b) from the previous solution $\mathbf{x}$, with $\Omega := A \cup supp(\mathbf{x})$, getting the vector $\mathbf{b}$ minimizing φ over Ω; c) from $\mathcal{M}(\mathbb{Q})$, pruning $\mathbf{b}$ into the new better solution $\mathbf{x}$; and the three procedures are repeated until the halting condition holds. The final solution $\mathbf{x}$ implies the underlying *specific-shape* graph anomaly.

The proposed approach to Problem (1) is illustrated in Algorithm 1. Query-map is an iterative selection method for approximately optimizing the nonlinear cost function φ in Problem (1). The method generates a sequence of intermediate *specific shape* s-sparse vectors $\mathbf{x}^0, \mathbf{x}^1, \ldots$ from a start approximation $\mathbf{x}^0 = \mathbf{0}$ (i.e., $\mathbf{0} \in \mathbb{R}^n$). At the i-th iteration, Step 4, $\mathbf{g} = \nabla\varphi(\mathbf{x}^i)$, computes the gradient of φ at the current solution $\mathbf{x}^i$. Step 5, $A = \mathcal{P}(\mathbf{g})$, selects the best *specific shape* vertex set A of the restricted gradient $\mathbf{g}_A$ leading to the best approximation to $\mathbf{g}$. The *specific shape* projection

Algorithm 1: Query-map

```
1  Pick the specific shape anomaly query $\mathbb{Q}$, attributed
     graph $\mathbb{G}$;
2  Set $i = 0, \mathbf{x}^i = \mathbf{0}$;
3  repeat
4      $\mathbf{g} = \nabla\varphi(\mathbf{x}^i)$;
5      $A = \mathcal{P}(\mathbf{g})$;        ▷ Specific Shape Projection Oracle
6      $\Omega = A \cup supp(\mathbf{x}^i)$;
7      $\mathbf{b} = \arg\min_{\mathbf{x}\in\mathbb{R}^n} \varphi(\mathbf{x}) \quad s.t. \ \mathbf{x}_{\Omega^c} = 0$;
8      $B = \mathcal{P}(\mathbf{b})$;        ▷ Specific Shape Projection
         Oracle
9      $\mathbf{x}^{i+1} = \mathbf{b}_B$;
10     $i = i + 1$;
11 until halting condition holds;
12 return the specific shape vertex set $B$;
```

oracle, $\mathcal{P}(\mathbf{g})$, projects the gradient vector $\mathbf{g}$ onto the nearest point vector $\mathbf{g}_A$ in the *specific shape* graph-structured sparsity model $\mathcal{M}(\mathbb{Q})$ in Problem (2).

$$\mathcal{P}(\mathbf{g}) = \arg \min_{A\in\mathcal{M}(\mathbb{Q})} \parallel \mathbf{g} - \mathbf{g}_A \parallel_1 \tag{2}$$

We define in detail the *specific shape* graph-structured sparsity model $\mathcal{M}(\mathbb{Q})$. A graph $\mathbb{C}$ is a subgraph of $\mathbb{G}$, denoted as $\mathbb{C} \subseteq \mathbb{G}$, if $V_\mathbb{C} \subseteq V_\mathbb{G}$, $E_\mathbb{C} \subseteq E_\mathbb{G}$ and $\forall(u, v) \in E_\mathbb{C}$, $u, v \in V_\mathbb{C}$. A graph $\mathbb{C}$ is isomorphic to a query graph $\mathbb{Q}$, denoted as $\mathbb{C} \cong \mathbb{Q}$, if there is a bijection $\psi : V_\mathbb{C} \to V_\mathbb{Q}$ such that, for every pair of vertices $u, v \in V_\mathbb{C}$, $(u, v) \in E_\mathbb{C}$ if and only if $(\psi(u), \psi(v)) \in E_\mathbb{Q}$. Therefore, $\mathcal{M}(\mathbb{Q}) := \{V_\mathbb{C} \mid \mathbb{C} \subseteq \mathbb{G}, \mathbb{C} \cong \mathbb{Q}\}$ represents the *specific shape* vertex sets of interest in the attributed graph $\mathbb{G}$, whose corresponding subgraphs are isomorphic to the *specific shape* anomaly query graph $\mathbb{Q}$.

Obviously, Problem (2) is equal to the *subgraph matching* problem of $\arg\max_{A\in\mathcal{M}(\mathbb{Q})} \parallel \mathbf{g}_A \parallel_1$, where $\mathbf{g}$ implies a node-weighted graph. The target is to construct a maximum weight *specific shape* subgraph in the node-weighted graph. The *subgraph matching* problem is well studied in the graph matching field (Yang et al. 2016; 2014). In Steps 5 and 8, our approach can encompass the previously studied subgraph mathcing methods or projection projection-pursuit oracles. The main work of our paper is that we proposed a generic approach to optimize a broad class of nonlinear cost function via *specific shape* constraints in attributed graphs.

In Step 6, the set, $\Omega = A \cup supp(\mathbf{x}^i)$, is chosen as the support. Pursuing the minimization of a nonlinear cost function in Ω will be most effective. In Step 7, we find a vector $\mathbf{b}$ with this support, and $\mathbf{b}$ minimizes the nonlinear cost function. In Step 8, the *specific shape* projection oracle, $\mathcal{P}(\mathbf{b})$, projects the vector $\mathbf{b}$ onto the desired *specific shape* vertex set B. In Step 9, prune $\mathbf{b}$ into the new better solution $\mathbf{x}^{i+1}$ by the set B. The above steps are repeated until the halting criterion is satisfied. The natural halting criterion is $\mathbf{x}^{i+1} = \mathbf{x}^i$. In practice there are two popular options to define the halting criterion: (1) the difference between the current φ and the previous one is less than a threshold $|\varphi(\mathbf{x}^i) - \varphi(\mathbf{x}^{i+1})| < \epsilon$; and (2) the difference between the current $\mathbf{x}$ and the previous one is less than a threshold $||\mathbf{x}^i - \mathbf{x}^{i+1}|| < \epsilon$ (e.g., $\epsilon = 0.01$).

Our approach connects to the existing works. In the special case where the *specific shape* projection oracles are the head and tail approximations for connected subgraph sparsity signals, Query-map reduces to Graph-MP (Chen and Zhou 2016) for the connected subgraph anomaly detection. Specifically, the support set selection in Step 6 is tuned by the gradient descent parameter η for $\Omega = supp(\mathbf{x}^i - \eta\nabla_A\varphi(\mathbf{x}^i))$, and the *specific shape* projection oracles consider just the tree-shape graph anomaly. Query-map reduces to Graph-TPP (Wu et al. 2017) for the tree-shape subgraph anomaly detection.

Theoretical Analysis

In this section, for our proposed algorithm Query-map, we analyze its theoretical properties on the two aspects: 1) Studying the convergence rate; and 2) time complexity of Query-map.

Before obtaining the theoretical properties, we require the following key technical condition under which the convergence of Query-map is guaranteed. Without loss of generality, we assume the cardinality of the vertex set of *specific shape* query graph to $s = |V_\mathbb{Q}|$.

Definition 1. *Weak Restricted Strong Convexity* (WRSC) condition for the objective function φ (Wu et al. 2017; Chen and Zhou 2016; Yuan, Li, and Zhang 2014). As $\mathcal{M}(\mathbb{Q})$ is the vertex sets of attributed graphs, if $\forall S \in \mathcal{M}(\mathbb{Q})$ with cardinality $|S| \leq 4s$ and $\forall \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$ with $supp(\mathbf{y}) \cup supp(\mathbf{z}) \subseteq S$, the following inequality holds for some $\xi > 0$ and $0 < \delta_{4s} < 1$. The objective function φ has the technical condition $(\xi, \delta_{4s}, \mathcal{M}(\mathbb{Q}))$-*WRSC*.

$$||\mathbf{y} - \mathbf{z} - \xi\nabla_S\varphi(\mathbf{y}) + \xi\nabla_S\varphi(\mathbf{z})|| \leq \delta_{4s}||\mathbf{y} - \mathbf{z}|| \tag{3}$$

where $\nabla_S\varphi(\mathbf{y})$ is a restriction of $\nabla\varphi(\mathbf{y})$ in S: we have $(\nabla_S\varphi(\mathbf{y}))_v = (\nabla\varphi(\mathbf{y}))_v$ for $v \in S$, and $(\nabla_S\varphi(\mathbf{y}))_v = 0$ otherwise. The WRSC condition is derived from the Restricted Strong Convexity/Smoothness (RSC/RSS) conditions (Yuan, Li, and Zhang 2014). The RSC condition is first characterized for the functions have quadratic bounds on the derivative of the objective function. Then a function must have the WRSC condition if the function has the RSC or RSS conditions (Yuan, Li, and Zhang 2014).

Now we analyze the convergence property of Query-map. We make a simple observation about Query-map that the sequence solution $\{\mathbf{x}^i\}$ defined by Query-map is eventually periodic on a vertex set S, due to the fact that the limit of φ in the support vertex set of size $4s$ is exactly achieved after a finite number of iterations.

Theorem 1. ***Query-map Convergence Rate.*** *Consider a nonlinear cost function $\varphi : \mathbb{R}^n \mapsto \mathbb{R}$ that satisfies the condition $(\xi, \delta_{4s}, \mathcal{M}(\mathbb{Q}))$-WRSC. Given the prior query graph $\mathbb{Q}$, the attributed graph $\mathbb{G}$, let $\mathbf{x}^*$ be the* ***true*** *specific shape vector: $\forall\mathbf{x} \in \mathbb{R}^n$, $\varphi(\mathbf{x}) \geq \varphi(\mathbf{x}^*)$ subject to $supp(\mathbf{x}), supp(\mathbf{x}^*) \in \mathcal{M}(\mathbb{Q})$, the sequence solution $\{\mathbf{x}^i\}$ defined by Query-map holds*

$$||\mathbf{x}^{i+1} - \mathbf{x}^*||_2 \leq \alpha||\mathbf{x}^i - \mathbf{x}^*||_2 + \beta||\nabla_I\varphi(\mathbf{x}^*)||_2 \tag{4}$$

where the parameters are specified as $\alpha = 4\sqrt{\frac{\delta_{4s}}{1-\delta_{4s}}}$, $\beta = \frac{\xi}{1-\delta_{4s}}[\frac{4}{1-2\delta_{4s}} + \frac{2(1-2\delta_{4s})}{\sqrt{\delta_{4s}-\delta_{4s}^2}} + 2]$, *and the vertex set* $I = \arg\max_{S\in\mathcal{M}(\mathbb{Q})} ||\nabla_S\varphi(\mathbf{x}^*)||_2$. *We ensure* $\alpha < 1$, *when selecting* $\delta_{4s} < \frac{1}{17}$.

We analyze the constant $||\nabla_I\varphi(\mathbf{x}^*)||_2$, and the parameters, α β, impact on the convergence to the near-optimal $\mathbf{x}^*$ for the sequence solution $\{\mathbf{x}^i\}$. The following result is an immediate corollary achieved by Theorem 1.

Corollary 1.1. ***Estimation Error Level.*** *For the specific shape anomaly detection (i.e., isomorphism to* $\mathbb{Q}$*), the vector* $\mathbf{x}$ *is calibrated to* $\mathbf{x} \in [0,1]^n$ *where* v *corresponds to the desired anomaly node in* $\mathbb{G}$ *if* $\mathbf{x}_v \neq 0$*, and* v *is not anomaly node otherwise. For the* i*-th solution* $\mathbf{x}^i$*, and the true specific shape vector* $\mathbf{x}^*$*, we have*

$$\| \mathbf{x}^i - \mathbf{x}^* \|_2 \geq \frac{\beta}{1-\alpha} \| \nabla_I\varphi(\mathbf{x}^*) \|_2 \tag{5}$$

$$\| \nabla_I\varphi(\mathbf{x}^*) \|_2 \leq \frac{1-\alpha}{\beta}\sqrt{s} \tag{6}$$

Before reaching the estimation error level (5), Query-map geometrically converges to the near-optimal $\mathbf{x}^*$. The estimation error of Query-map is determined by the multipliers of $||\nabla_I\varphi(\mathbf{x}^*)||_2$. Especially, when $||\nabla_I\varphi(\mathbf{x}^*)||_2 = 0$, Query-map guarantees that the true $\mathbf{x}^*$ is exactly recovered within finite iterations. The shrinkage rate α relates to the parameter δ, where $\delta_s \leq \delta_{2s} \leq \delta_{4s}$ (Tropp and Needell 2008). The smaller δ is attained, and the faster convergence rate of Query-map is achieved. In this work, as $\sqrt{s}(1-\alpha)/\beta$ in the upper bound (6) is a small constant, we consider the case where $||\nabla_I\varphi(\mathbf{x}^*)||$ is a sufficiently small constant, such that $||\mathbf{x}^{i+1} - \mathbf{x}^*|| \leq ||\mathbf{x}^i - \mathbf{x}^*||$. Thus the speed of convergence of our algorithm Query-map mainly depends on the shrinkage rate α for $||\nabla_I\varphi(\mathbf{x}^*)||$ is a sufficiently small constant.

Theorem 2. ***Query-map Time Complexity.*** *The absolute upper bound to the estimation error* $||\mathbf{x}^i - \mathbf{x}^*||$ *is* $\sqrt{2s}$ *for* $\mathbf{x}^i, \mathbf{x}^* \in [0,1]^n$*. Within an approximate estimation error, we obtain a near-optimal solution* $\hat{\mathbf{x}} \in [0,1]^n$*, such that* $||\mathbf{x}^* - \hat{\mathbf{x}}||_2 \leq \big(||\mathbf{x}^*||_2 + \beta/(1-\alpha)\big)||\nabla_I\varphi(\mathbf{x}^*)||_2$*, subject to,* $supp(\hat{\mathbf{x}}) \in \mathcal{M}(\mathbb{Q})$*. The time complexity of Query-map is*

$$O\big(T\log(1 \,/\, ||\nabla_I\varphi(\mathbf{x}^*)||_2)\big) \tag{7}$$

where for each iteration, the time T *consists of two parts: 1) One execution of the subproblem in Line 5 and 8; and 2) one execution of the subproblem in Line 7.*

The estimation error, $\big(||\mathbf{x}^*||_2 + \beta/(1-\alpha)\big)||\nabla_I\varphi(\mathbf{x}^*)||_2\big)$, is a tighter bound for the estimation error level (5). The total time complexity relates to which well-studied oracle is employed to the projection pursuit for *specific shape* anomaly query. Our approach in Algorithm 1 reaches a near-linear time complexity with linear projection pursuit oracles. We observe that $\mathcal{M}(\mathbb{Q})$ is not a convex set. The optimization methods, such as Frank-Wolfe, can not be directly applied to the specific shape anomalous subgraph discovery based problem as studied in this paper before. We first proposed an approach to optimize nonlinear functions in *specific shape* anomaly queries with the better theoretical basis.

Application to Well-known Objective Function

The nonlinear cost function φ in Problem (1) can be applied to least square function, *Kulldorff's original Poisson scan statistic* (KULL) and *Expectation-based Poisson statistic* (EBP) (Neill 2009b). Least square is one of the most popular models in regression analysis that finds "*specific-shape*" sets of best fit for the attributed graphs (Chen and Zhou 2016). In this model, given the attribute vector $\mathbf{w} \in \mathbb{R}^n$, the least square function is optimized via *specific-shape* constraints in attributed graphs.

$$\min_{\mathbf{x}\in\mathbb{R}^n} \| \mathbf{w} - \mathbf{x} \|_2^2 \qquad s.t. \quad supp(\mathbf{x}) \in \mathcal{M}(\mathbb{Q})$$

It is well-known that the least square function is strongly convex. Its Bregman divergence ($\triangle\varphi := \varphi(\mathbf{x}) - \varphi(\mathbf{y}) - < \nabla\varphi(\mathbf{y}), \mathbf{x} - \mathbf{y} >$) is $\| \mathbf{x} - \mathbf{y} \|_2^2$. We obtain that the least square function satisfies the condition $(\xi, \delta_{4s}, \mathcal{M}(\mathbb{Q}))$-WRSC that $\delta_{4s} = 1 - 2\xi$ for $\xi < 1$ (Yuan and Liu 2014). For ensuring that Query-map converges to the near-optimal solution, the parameter ξ ranges between $8/17$ and 1.

In this work, there are the other two well-known graph scan statistics functions: KULL and EBP that are widely employed in pattern detection in graphs. For the statistics, there are just two attributes, ***observed count*** and ***expected count*** (Neill 2009a). Let $\mathbf{c} = \mathbf{W}_{:,0} \in \mathbb{R}^n$ denote the "observed count" attribute values. Similarly, let $\mathbf{d} = \mathbf{W}_{:,1} \in \mathbb{R}^n$ denote the "expected count" attribute values. The log forms of EBP and KULL are examined in the following functions.

$$\begin{aligned} \varphi_{EBP}(\mathbf{x}) &:= -\mathbf{x}^T\mathbf{c}\log(\mathbf{x}^T\mathbf{c} \,/\, \mathbf{x}^T\mathbf{d}) - \mathbf{x}^T\mathbf{d} + \mathbf{x}^T\mathbf{c} \\ \varphi_{KULL}(\mathbf{x}) &:= -\mathbf{x}^T\mathbf{c}\log(\mathbf{x}^T\mathbf{c} \,/\, \mathbf{x}^T\mathbf{d}) - \\ &\qquad (1-\mathbf{x})^T\mathbf{c}\log\big((1-\mathbf{x})^T\mathbf{c} \,/\, (1-\mathbf{x})^T\mathbf{d}\big) \end{aligned}$$

The forms of EBP and KULL are derived from the statistics $F(V_{\mathbb{C}})$ in (Neill 2009a) that is formulated as the detection problem: $\min_{\mathbb{C}\subseteq\mathbb{G}} -F(V_{\mathbb{C}})$ $s.t.$ $V_{\mathbb{C}} \in \mathcal{M}(\mathbb{Q})$.

Experiments

Our experiments consist of two parts: (i) uncovering high quality *specific shape* attacking anomalies in the real-world $*$edu.cn network dataset by our method; and (ii) demonstrating the efficiency of our method on different applications.

Uncovering High Quality Specific Shape Attacking Anomaly

Dataset. *Real-World* $*$**edu.cn** *Network Dataset.* An Internet security company[1] provided us with the total 3,978,073 web sites browsing logs from May 31, 2014 to May 13, 2015. The real-world traffic network of 131,107 nodes and 358,386 edges, is built from the browsing logs (i.e., the edge (IP site A, IP site B) denotes that A visited B). For a day t and a node v in this network, we denote the number of logs within v on that day t as the *observed value* $\mathbf{c}_v$, and the average number of logs within v before t as the *expected value* $\mathbf{d}_v$.

[1] An Internet security company in China with more than 0.6 billion users.

Our method. Query-map employs the graph scan statistic φ_{EBP} as the objective function to detect *specific shape* attack subgraphs in real networks.

Result. In Figure 2, given the query graphs Q1, Q2 and Q3, choose a random day (i.e., March 13, 2015) from the period of Jan 1, 2015 and May 13, 2015. The *specific shape* attacking anomalies are presented in Figure 2. The right subfigure of Figure 2 denotes the *star-chain-shape* anomaly queries. The middle subfigure of Figure 2 presents an attacking sub-network. The left subfigure of Figure 2 shows the *specific shape* cyber attacking cases. All of the detected sites with our method are identified as true attacking sources by the experts from the Internet security company. (1) *Specific shape* anomaly Q1. The user *x.x.223.66* attacking the server *xkb.hlu.edu.cn* is categorized to "FckEditor", "Upload Webshell" and "Common Vulnerability" attacks. This user attacking the servers *tw.hlu.edu.cn*, *zsb2.hlu.edu.cn* and *kydown.hlu.edu.cn* with the same approach is categorized to "Upload Webshell" attacks. (2)*Specific shape* anomaly Q2. The server *www.hlu.edu.cn* is attacked from the users *x.x.78. 34* by "SQL Inject", *x.x.3.69* and *x.x.103.149* by "Dedecms" and "Common Vulnerability". The user *x.x.103.149* also attacked the servers *xkb.hlu.edu.cn* and *tw.hlu.edu.cn* with "Dedecms" and "Common Vulnerability" attacks. (3)*Specific shape* anomaly Q3. The user *x.x.223.66* attacked the server *www.hlu.edu.cn* with "FckEditor", "Upload Webshell" and "Common Vulnerability" approaches. These attacks caused this server to be a *bot machine*. This infected server attacked the server *www.cq51edu.cn* with "Common Vulnerability". The user *x.x.32.220* attacked the server *www. cq51edu.cn* with "Upload Webshell" and "Common Vulnerability" approaches. The user *x.x.21.210* attacked the server *www.cq51edu.cn* with the "nginx parse" approach. From the results, we can observe that the *specific shape* anomalies are exactly detected from the network.

Efficiency Comparison of Query-map and Baseline

Datasets: 1) Water Pollution Dataset. By the chemical contaminant plumes are distributed at 4 nodes within different areas (Chen and Zhou 2016), we collected the real-world water pollution network of 12,527 nodes and 14,831 edges. The network is constructed by the K-Nearest Neighbor (KNN) algorithm. "The spreads of contaminant plumes were simulated using the water network simulator EPANET for 8 hours" (Chen and Zhou 2016). For the *observed count* attribute, the value at each node v is collected from the corresponding sensor per hour, $\mathbf{c}_v \leftarrow 1$ if it is polluted and $\mathbf{c}_v \leftarrow 0$ otherwise. For testing the robustness of methods to noises, we randomly flipped K percent sensor binary values, where $K \in \{2, 4, 6, 8, 10\}$. The noise ratio $\mathbf{d}_v \leftarrow K\%$ is considered as the *expected count* attribute (Chen and Zhou 2016). **2) *Respiratory Emergency Department* (ED) Dataset.** In a grid network of 10,000 nodes and 14,850 edges, for each node v, we collected the T day period of respiratory ED visit data $\mathbf{W}_v \in \mathbb{R}^T$ (e.g., $T = 28$, and the time $t = 0$ denotes the current day). The outbreak linearly increases in cases over the outbreak duration (Neill 2009a). During non-outbreak period, the number of patients visiting ED in v is $\mathbf{W}_v^t \leftarrow$ Poisson(μ) for $t = 0, \cdots, T$ where $\mu \in \{1, \cdots, 34\}$ denotes the expected number in v on that days. During outbreak period, we randomly select the outbreak duration of U from $\{1, \cdots, 7\}$, normalizes the weight $w_v \propto \sum_t \mathbf{W}_v^t$ so that the total weight is equal to 1 in infected nodes, and set the outbreak severity Δ (e.g., Δ=800) (Neill 2009a). On each day $t \in \{0, \cdots, U\}$, we inject cases into each infected node v i.e., $\mathbf{W}_v^t \leftarrow \mathbf{W}_v^t + \text{Poisson}((T-t)w_v\Delta)$ for $t = 0, \cdots, U$ (*medium-size* outbreaks injected for 10 percent nodes (Neill 2009a)). For testing the robustness of methods, we flipped values of $K \in \{2, 4, 6, 8, 10\}$ percent nodes randomly, i.e., no inject outbreak cases if the nodes are infected, inject outbreak cases otherwise. Let $\mathbf{c}_v = \mathbf{W}_v^0$ denote the *observed count* of infected cases, and $\mathbf{d}_v = \frac{1}{T}\sum_{t=1}^{T} \mathbf{W}_v^t$ denote the *expected count*. The datasets are summarized in Table 2.

Comparison Methods. We compared our method "Query-map" with the two state-of-the-art baselines: *Top-k* (Gupta et al. 2014) and *Fast-k* (Yang et al. 2016). The baselines are designed specifically for *specific shape* anomaly discovery in attributed graphs. The baselines aim to obtain top k subgraphs with the maximal sum of its node scores, which are isomorphic to the query graph. The parameters k and D are tuned completely based on the author recommendation in the original papers for $k = 10$, $D = 2$ to *Top-k* (Gupta et al. 2014) and $k = 20$, $d = 2$ to *Fast-k* (Yang et al. 2016). Let $w(v) \leftarrow \mathbf{c}_v$ for each vertex $v \in V_{\mathbb{G}}$.

Performance Metrics. 1) *Precision*. We compute the precision of the target subgraph (i.e., the ratio of the number of correct anomalous nodes and the number of nodes). The recall metric is ignored for the fixed size of returned target subgraph. 2) *Function Score* and *Running Time*. The optimization power of our method is examined in the scores of graph scan statistics. We compare our method to the baselines on running times.

Results: *Precision for target subgraphs detection.* Figure 3 illustrates the precisions obtained by our methods Query-map (EBP and KULL) corresponding to *red bars* and *cyan bars*, and the competitive methods Fast-K and Top-K corresponding to *green bars* and *blue bars*. For the 2% noise level, Figure 3 (a) and (c) illustrate that our methods Query-map (KULL) for Water Pollution dataset and Query-map (EBP) for Emergency dataset outperformed all the baselines on the precision. For the Water Pollution dataset in Figure 3 (a), our methods and baselines perform similar results. Our methods recovered at least 99.5% *true target subgraphs* on the anomaly queries Q2 and Q3. With the query graph expanding in Figure 3(c), the precisions of baselines decrease quickly, however the precisions of our method Query-map (EBP) even increase to 1.0. Especially for the query graph Q3, we can observe that the precisions of our methods are greater than at least 0.67 for the best baseline. For the 10% noise level, the most results in Figure 3(b) and (d) indicate that our methods outperform the baselines. In Figure 3(b), the precisions of baselines decrease with the query graph size. However, even at the 10% noise level, the precisions of our methods still increase to 1.0. For the Emergency dataset in Figure 3(d), although our Query-map (KULL) is little better than the best baselines, our Query-map (EBP) achieves at least 0.27 improvement in the precision larger than the best

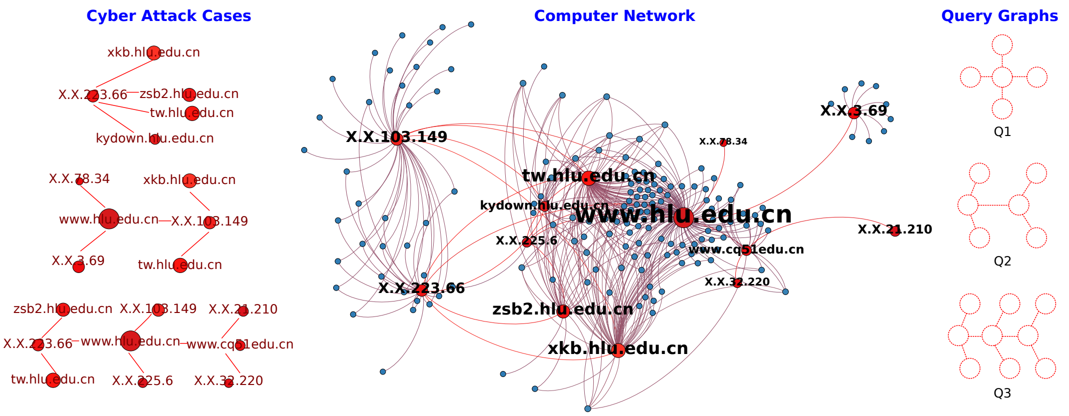
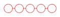
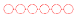
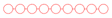
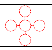
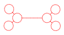
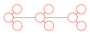


Figure 2: The cyber attack cases on March 13, 2015 are detected by our method Query-map (EBP) in the $*edu.cn$ real-world network dataset (red solid nodes denote the attacking or attacked sites, and blue solid nodes denote normal nodes).

Table 2: Summary of datasets (WP: Water Pollution; ED: Emergency Department).

Attributed Networks					**Interesting Anomaly Queries**		
Data set	**# of nodes**	**# of edges**	**Attribute c_v** (i.e., ***observed value***) **for each node v**	**Attribute d_v** (i.e., ***expected value***) **for each node v**	**Q1 query graph**	**Q2 query graph**	**Q3 query graph**
WP	12,527	14,831	Sensor value (0 or 1)	Noise level			
ED	10,000	14,850	Number of patient visits	Average number of visits			

baseline. Although Fast-K and Top-K performed not bad in Figure 3(a), these heuristic algorithms do not have any theoretical guarantees on identifying *specific shape* anomaly subgraphs. From the results in Figure 3, we can observe that our methods not only have a significant theoretical property, but also always have a high precision.

In Table 2, the graph anomaly shapes Q1, Q2 and Q3 in Water Pollution dataset are different from graph anomaly shapes in Emergency Department dataset for the pollution area showing strip shapes and the outbreak area showing star shapes (Wu et al. 2017). Query-map achieved the precisions for "strip" anomaly shapes in Water Pollution data and "star" anomaly shapes in Emergency Department data illustrated in Figure 3. The Top-K is a path index-based method (Gupta et al. 2014), and the Fast-K is a heuristic method for assembling star components. The baselines tend to perform better on Water Pollution data with "strip" shape graph anomalies, however, perform not better on Emergency Department data with "star" shape graph anomaly.

Evolving curves of graph scan statistics. Figure 4(a-b) report the graph scan statistic (EBP) scores of detected subgraphs for the $\{0, \cdots, 9\}$ iterations. The results in Figure 4(a-b) illustrate that our method Query-map (EBP) converges in less than 10 iterations. Especially in Figure 4(a) for the Water Pollution dataset, Query-map (EBP) converges in less than 3 iterations. The empirical results show the fast convergence trends of our methods Query-map.

Scalability analysis of running time. Table 3 reports the comparison between our methods Query-map and the competitive baseline methods on the running time. In Table 3, the running times were collected from the computer with Intel Xeon E3-1220 (e.g., 4 CPU, 3.1 GHz) and 24GB RAM. The results in Table 3 show that our proposed method Query-map ran faster than all the baseline methods in most settings, except for the *specific shape* anomaly graph Q1. Even though the baseline method Fast-K ran the fastest over the query graph Q1, this method can not detect the target subgraph with high qualities in Figure 3(b-d). In Emergency data, for the Q3, our methods returned results within 97 seconds, however, Top-K did not output any results within 7 hours. The method can performs on the edu.cn data with 131,107 nodes. Results in Table 3 also imply that the running time of our methods is insensitive to the noise level, which is consistent with the time complexity of Query-map as discussed in Theorem 2.

Conclusion

This paper presents an efficient algorithm to optimize nonlinear functions subject to *specific shape* anomaly. Our approach is guaranteed to the near-optimal solution under the estimation error upper bound of $\big(||\mathbf{x}^*||_2 + \beta/(1-\alpha)\big)||\nabla_I \varphi(\mathbf{x}^*)||_2\big)$. A wide variety of attributes can be employed to the *specific shape* anomaly discovery in graphs. We will extend Query-map on the best-effort match to the *specific shape* graph anomaly.

Appendix

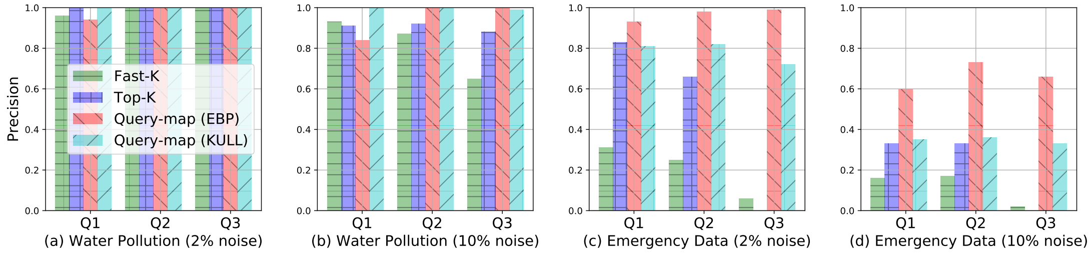
A1. Proof of Theorem 1

We first introduce Lemma 3 to bound residues between the solution $\mathbf{x}^i$ and the optimal $\mathbf{x}^*$.

Lemma 3. *Let* $\mathbf{r}^i = \mathbf{x}^i - \mathbf{x}^*$. *Given* $A \in \mathcal{M}(\mathbb{Q})$*, we have*

Table 3: Efficiency Measure: Comparison on the running times of our methods and the baselines.

Run Time (second)	Water Pollution Data						Emergency Data			
	2 % noise			10 % noise			2 % noise		10 % noise	
	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q1	Q2
Query-map (EBP)	7.99	13.26	22.57	8.18	14.25	21.92	0.95	0.98	0.73	0.74
Query-map (KULL)	7.90	**12.81**	**22.22**	7.10	**12.57**	**19.21**	0.81	**0.82**	**0.35**	**0.36**
Fast-K	**2.37**	34.03	86.13	**2.27**	46.83	190.54	**0.47**	73.41	0.47	72.32
Top-K	33.27	389.69	1,343.22	56.62	660.79	2,313.82	315.99	6,433.98	389.73	6,640.92



(a) Water Pollution (2% noise)
(b) Water Pollution (10% noise)
(c) Emergency Data (2% noise)
(d) Emergency Data (10% noise)

Figure 3: Effective Validation: Precision comparisons of our methods Query-map (EBP) and Query-map (KULL), and the baseline methods Fast-K and Top-K.

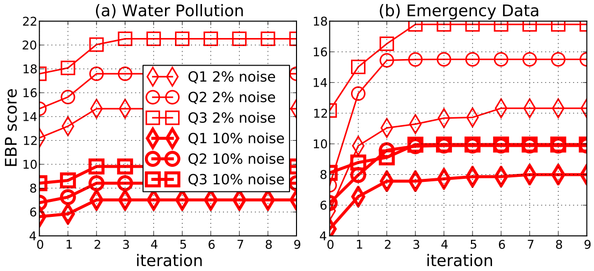


Figure 4: Efficiency Measure: Evolving curves of graph scan statistic scores (EBP) for our method (Query-map) in different iterations. These scores are evaluated on these two datasets with 2% and 10% noises and the query graphs.

the following inequality.

$$||\mathbf{r}^i_{A^c}|| \leq 2\sqrt{\delta_{4s}-\delta_{4s}^2}||\mathbf{r}^i|| + \left(\frac{2\xi}{1-2\delta_{4s}} + \frac{(1-2\delta_{4s})\xi}{\sqrt{\delta_{4s}-\delta_{4s}^2}}\right)||\nabla_I\varphi(\mathbf{x}^*)|| \quad (8)$$

where $I = \arg\max_{S\in\mathcal{M}(\mathbb{Q})} ||\nabla_S\varphi(\mathbf{x}^*)||_2$.

Now, we present the proof of Theorem 1.

Proof. Similarly, we denote $\mathbf{r}^{i+1} = \mathbf{x}^{i+1} - \mathbf{x}^*$. We have $||(\mathbf{x}^{i+1}-\mathbf{b}) + (\mathbf{b}-\mathbf{x}^*)||_2 \leq ||\mathbf{x}^{i+1}-\mathbf{b}||_2 + ||\mathbf{b}-\mathbf{x}^*||_2$ by the triangular inequality property. In Algorithm 1, by Line 8, we have $||\mathbf{x}^{i+1}-\mathbf{b}||_2 = ||\mathbf{b}_B-\mathbf{b}||_2$. According to Problem (2), $\mathbf{b}_B$ is restricted to the largest elements of $\mathbf{b}$ by the projection oracle $\mathcal{P}$, and thus $||\mathbf{b}_B-\mathbf{b}||_2 \leq ||\mathbf{x}^*-\mathbf{b}||_2$. The upper bound of $||\mathbf{r}^{i+1}||_2$ is $2||\mathbf{x}^*-\mathbf{b}||_2$, for the deduction of $||\mathbf{r}^{i+1}||_2 = ||\mathbf{x}^{i+1}-\mathbf{x}^*||_2 \leq ||\mathbf{x}^{i+1}-\mathbf{b}||_2 + ||\mathbf{x}^*-\mathbf{b}||_2 = ||\mathbf{b}_B-\mathbf{b}||_2 + ||\mathbf{x}^*-\mathbf{b}||_2 \leq 2||\mathbf{x}^*-\mathbf{b}||_2$.

Compute the upper bound of $||(\mathbf{x}^*-\mathbf{b})_\Omega||_2^2$

$$\begin{aligned}
&=< \mathbf{b}-\mathbf{x}^*, (\mathbf{b}-\mathbf{x}^*)_\Omega > \\
&=< \mathbf{b}-\mathbf{x}^* - \xi\nabla_\Omega\varphi(\mathbf{b}) + \xi\nabla_\Omega\varphi(\mathbf{x}^*), (\mathbf{b}-\mathbf{x}^*)_\Omega > - \\
&\quad < \xi\nabla_\Omega\varphi(\mathbf{x}^*), (\mathbf{b}-\mathbf{x}^*)_\Omega > \\
&\leq ||\mathbf{b}-\mathbf{x}^* - \xi\nabla_\Omega\varphi(\mathbf{b}) + \xi\nabla_\Omega\varphi(\mathbf{x}^*)||_2||(\mathbf{b}-\mathbf{x}^*)_\Omega||_2 + \\
&\quad \xi||\nabla_\Omega\varphi(\mathbf{x}^*)||_2||(\mathbf{b}-\mathbf{x}^*)_\Omega||_2 \\
&\leq \delta_{4s}||\mathbf{b}-\mathbf{x}^*||_2||(\mathbf{b}-\mathbf{x}^*)_\Omega||_2 + \xi||\nabla_\Omega\varphi(\mathbf{x}^*)||_2||(\mathbf{b}-\mathbf{x}^*)_\Omega||_2
\end{aligned}$$

where the second equality is derived from $\nabla_\Omega\varphi(\mathbf{b}) = 0$ for the function φ is minimized at $\mathbf{b}$ over the set Ω in Line 7 of Algorithm 1. The last inequality is derived from the condition $(\xi, \delta_{4s}, \mathcal{M}(\mathbb{Q}))$-WRSC. Thus the upper bound of $||(\mathbf{x}^*-\mathbf{b})_\Omega||_2$ is $\delta_{4s}||\mathbf{b}-\mathbf{x}^*||_2 + \xi||\nabla_\Omega\varphi(\mathbf{x}^*)||_2$.

For the sets Ω and A, we denote the complement sets are $\Omega^c = V_\mathbb{G} \setminus \Omega$ and $A^c = V_\mathbb{G} \setminus A$. We have $||\mathbf{x}^*-\mathbf{b}||_2 \leq ||(\mathbf{x}^*-\mathbf{b})_\Omega||_2 + ||(\mathbf{x}^*-\mathbf{b})_{\Omega^c}||_2$. By the upper bound of $||(\mathbf{x}^*-\mathbf{b})_\Omega||_2$, we have $||(\mathbf{x}^*-\mathbf{b})_\Omega||_2 \leq \delta_{4s}||\mathbf{x}^*-\mathbf{b}||_2 + \xi||\nabla_\Omega\varphi(\mathbf{x}^*)||_2 + ||(\mathbf{x}^*-\mathbf{b})_{\Omega^c}||_2$. We have $||\mathbf{x}^*-\mathbf{b}||_2 \leq \frac{||(\mathbf{x}^*-\mathbf{b})_{\Omega^c}||_2}{1-\delta_{4s}} + \frac{\xi||\nabla_\Omega\varphi(\mathbf{x}^*)||_2}{1-\delta_{4s}}$. By $supp(\mathbf{b}) \subseteq \Omega$ at Line 7 and $supp(\mathbf{x}^i) \subseteq \Omega$ at Line 6 of Algorithm 1, we have $\frac{||(\mathbf{x}^*-\mathbf{b})_{\Omega^c}||_2}{1-\delta_{4s}} + \frac{\xi||\nabla_\Omega\varphi(\mathbf{x}^*)||_2}{1-\delta_{4s}} = \frac{||\mathbf{x}^*_{\Omega^c}||_2}{1-\delta_{4s}} + \frac{\xi||\nabla_\Omega\varphi(\mathbf{x}^*)||_2}{1-\delta_{4s}} = \frac{||(\mathbf{x}^*-\mathbf{x}^i)_{\Omega^c}||_2}{1-\delta_{4s}} + \frac{\xi||\nabla_\Omega\varphi(\mathbf{x}^*)||_2}{1-\delta_{4s}} = \frac{||\mathbf{r}^i_{\Omega^c}||_2}{1-\delta_{4s}} + \frac{\xi||\nabla_\Omega\varphi(\mathbf{x}^*)||_2}{1-\delta_{4s}}$. As $\Omega^c \subseteq A^c$, we have $\frac{||\mathbf{r}^i_{\Omega^c}||_2}{1-\delta_{4s}} + \frac{\xi||\nabla_\Omega\varphi(\mathbf{x}^*)||_2}{1-\delta_{4s}} \leq \frac{||\mathbf{r}^i_{A^c}||_2}{1-\delta_{4s}} + \frac{\xi||\nabla_I\varphi(\mathbf{x}^*)||_2}{1-\delta_{4s}}$. At last we obtain the upper bound as follows

$$||\mathbf{x}^*-\mathbf{b}||_2 \leq ||\mathbf{r}^i_{A^c}||_2 \,/\, (1-\delta_{4s}) + \xi||\nabla_I\varphi(\mathbf{x}^*)||_2 \,/\, 1-\delta_{4s}$$

Combing $||\mathbf{r}^{i+1}||_2 \leq 2||\mathbf{x}^*-\mathbf{b}||_2$, the above inequality, and Lemma 3, we finish proving this theorem. □

A2. Proof of Theorem 2

Proof. As $\mathbf{x}^i, \mathbf{x}^* \in [0,1]^n$ and $|supp(\mathbf{x}^i)|, |supp(\mathbf{x}^*)| \leq s$, the absolute upper bound to $||\mathbf{x}^i - \mathbf{x}^*||$ is $\sqrt{2s}$. For the small constant $||\nabla_I \varphi(\mathbf{x}^*)||_2$, we present the acceptable upper bound $\big(||\mathbf{x}^*||_2 + \beta/(1-\alpha)\big)||\nabla_I \varphi(\mathbf{x}^*)||_2$.

By Inequality (4), at the i-th iterate of Algorithm 1, we obtain the bound $||\mathbf{x}^* - \mathbf{x}^i||_2 \leq \alpha^i ||\mathbf{x}^*||_2 + \frac{\beta}{1-\alpha}||\nabla_I \varphi(\mathbf{x}^*)||_2$. Let $\alpha^i ||\mathbf{x}^*||_2 + \frac{\beta}{1-\alpha}||\nabla_I \varphi(\mathbf{x}^*)||_2 \leq \big(||\mathbf{x}^*||_2 + \beta/(1-\alpha)\big)||\nabla_I \varphi(\mathbf{x}^*)||_2$. After $\lceil \log(\frac{1}{||\nabla_I \varphi(\mathbf{x}^*)||_2})/\log\frac{1}{\alpha}\rceil$ iterations, the estimate $\hat{\mathbf{x}}$ satisfies $||\mathbf{x}^* - \hat{\mathbf{x}}||_2 \leq \big(||\mathbf{x}^*||_2 + \beta/(1-\alpha)\big)||\nabla_I \varphi(\mathbf{x}^*)||_2$. As the time complexity is T for one iteration, the overall time complexity of Query-map follows the result in Theorem 2. □

Acknowledgments

This work was supported by the National Key R&D Program of China (No. 2018YFC0809800), and partly supported by the NSF grant (No. IIS-1815696). The corresponding author is Wenjun Wang.

References

Akoglu, L.; Tong, H.; and Koutra, D. 2015. Graph based anomaly detection and description: a survey. *Data mining and knowledge discovery* 29(3):626–688.

Chen, F., and Zhou, B. 2016. A generalized matching pursuit approach for graph-structured sparsity. In *IJCAI*, 1389–1395.

Cordella, L. P.; Foggia, P.; Sansone, C.; and Vento, M. 2004. A (sub) graph isomorphism algorithm for matching large graphs. *IEEE TPAMI* 26(10):1367–1372.

Gupta, M.; Gao, J.; Yan, X.; Cam, H.; and Han, J. 2014. Top-k interesting subgraph discovery in information networks. In *ICDE*, 820–831. IEEE.

Hegde, C.; Indyk, P.; and Schmidt, L. 2015. A nearly-linear time framework for graph-structured sparsity. In *ICML*, 928–937.

Huan, J.; Wang, W.; and Prins, J. 2003. Efficient mining of frequent subgraphs in the presence of isomorphism. In *ICDM*, 549–552. IEEE.

Kulldorff, M. 1997. A spatial scan statistic. *Communications in Statistics-Theory and methods* 26(6):1481–1496.

Neill, D. B. 2009a. An empirical comparison of spatial scan statistics for outbreak detection. *International journal of health geographics* 8(1):1.

Neill, D. B. 2009b. Expectation-based scan statistics for monitoring spatial time series data. *International Journal of Forecasting* 25(3):498–517.

Tropp, J. A., and Needell, D. 2008. Cosamp: Iterative signal recovery from incomplete and inaccurate samples. *CoRR* abs/0803.2392.

Wu, N.; Chen, F.; Li, J.; Huai, J.; and Li, B. 2017. Query-driven discovery of anomalous subgraphs in attributed graphs. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*, 3105–3111.

Yang, S.; Wu, Y.; Sun, H.; and Yan, X. 2014. Schemaless and structureless graph querying. *Proceedings of the VLDB Endowment* 7(7):565–576.

Yang, S.; Han, F.; Wu, Y.; and Yan, X. 2016. Fast top-k search in knowledge graphs. In *ICDE*, 990–1001.

Yuan, X.-T., and Liu, Q. 2014. Newton greedy pursuit: A quadratic approximation method for sparsity-constrained optimization. In *CVPR*, 4122–4129.

Yuan, X.; Li, P.; and Zhang, T. 2014. Gradient hard thresholding pursuit for sparsity-constrained optimization. In *ICML*, 127–135.

Zhang, X.; Martin, T.; and Newman, M. E. 2015. Identification of core-periphery structure in networks. *Physical Review E* 91(3):032803.

Zou, L.; Chen, L.; and Lu, Y. 2007. Top-k subgraph matching query in a large graph. In *the ACM first Ph.D. workshop in CIKM*, 139–146. ACM.