Discovering Contexts and Contextual Outliers Using Random Walks in Graphs

Xiang Wang  Ian Davidson
Department of Computer Science
University of California, Davis
xiang@ucdavis.edu; davidson@cs.ucdavis.edu

Abstract—The identifying of contextual outliers allows the discovery of anomalous behavior that other forms of outlier detection cannot find. What may appear to be normal behavior with respect to the entire data set can be shown to be anomalous by subsetting the data according to specific spatial or temporal context. However, in many real-world applications, we may not have sufficient a priori contextual information to discover these contextual outliers. This paper addresses the problem by proposing a probabilistic approach based on random walks, which can simultaneously explore meaningful contexts and score contextual outliers therein. Our approach has several advantages including producing outlier scores which can be interpreted as stationary expectations and their calculation in closed form in polynomial time. In addition, we show that point outlier detection using the stationary distribution is a special case of our approach. It allows us to find both global and contextual outliers simultaneously and to create a meaningful ranked list consisting of both types of outliers. This is a major departure from existing work where an algorithm typically identifies one type of outlier. The effectiveness of our method is justified by empirical results on real data sets, with comparison to related work.

I. INTRODUCTION

Outlier detection, also called anomaly detection, is an important but understudied branch of the data mining research: only recently did the first data mining survey on this topic become available [1]. Most of the existing approaches identify outliers from a global point of view, where each data instance is examined as deviating from normality as defined by the entire data set. This type of outlier detection is called global outlier detection [1]. However, sometimes an instance may not be an outlier when compared to the rest of the data set but maybe an outlier in the context of a subset of the data. This type of outlier detection is called contextual outlier detection, where the subset with respect to which the outlier is examined is called the context. For example, in a population demographic data set, a six-foot person may not be anomalous, but in the context of individuals aged under ten years of age would be an outlier.

As compared to global outlier detection, contextual outlier detection is even more understudied [1]. A major challenge of contextual outlier detection is identifying the contexts which then allow the identification of outliers. A data instance may appear anomalous in one context but not in others. Therefore, the meaningfulness of the context essentially decides the interestingness of the contextual outliers. In order to define the proper contexts, existing contextual outlier detection techniques require the user to a priori specify the contextual information, or the contextual attributes, in the data set. Typical contextual attributes used by previous work include partition labels [2], spatial and/or temporal information [3]–[5], adjacency in graphs [6], and profiles [7].

Unfortunately, the a priori contextual information is not always available in practice. And even if we have well-defined contexts within the data set, it is nontrivial to find contextual outliers therein. Since defining contexts and detecting contextual outliers are mutually dependent, a logical extension of existing contextual outlier detection work is to fold identifying the contexts into the outlier detection question itself by asking under which natural contexts do outliers occur. Our work explores this more elaborate question.

We propose a graph based random walk model where we can formally define contexts and contextual outliers therein. Modeling the problem using graph model does not limit our work to graph based data such as social networks. Our work is applicable so long as a transition (probability) matrix can be generated from the data set. For example, we can build a random walk graph by representing each data instance as a node and by converting the similarity between data instances into a transition probability between the two nodes. Based on our random walk graph model, we develop an algorithm using eigendecomposition, which can automatically and simultaneously find different contexts in the data set and rank all the data instances by a probabilistically principled outlier score with respect to their contexts. Our contributions are:

1) To the best of our knowledge, this is the first work to find contextual outliers without a priori contextual information by automatically discovering the contexts.
2) We provide a flexible method of finding the contexts and the contextual outliers, which is applicable to both graphical and vector data.
3) We create an easily interpretable contextual outlier score for any node as being the difference in the chance that a random walk on the entire graph visits the node from respective contexts. This allows us to meaningfully rank both global and contextual outliers.
4) We propose an efficient polynomial time algorithm based on eigendecomposition that can automatically and simultaneously find contexts as well as contextual outliers that are interpretable from a probabilistic perspective.

II. BACKGROUND AND PRELIMINARIES

In this section, we introduce the notion of a random walk graph, which is essentially a homogeneous Markov chain as characterized by a transition matrix. We show the well-known result that the principal eigenvector of the transition matrix gives the stationary distribution of the nodes being visited in the graph under a global random walk. We then survey the previous work [8] that uses the principal transition matrix gives the stationary distribution of the well-known result that the principal eigenvector of the walk graph, which is essentially a homogeneous Markov chain. Specifically, let \( W \) be the transition matrix from node \( i \) to node \( j \). We can construct a random walk graph \( G = (V, E) \) from \( A \) as follows. Each node in \( G \) represents a data instance in \( D \) and the directed edge from node \( i \) to node \( j \) means that the transition from node \( i \) to \( j \) happens with the probability as specified by a transition matrix \( W \). Let \( D \) be a diagonal matrix where

\[
d_{ij} = \begin{cases} 
\sum_{i=1}^{n} a_{ij} & \text{if } i = j \\
0 & \text{if } i \neq j 
\end{cases}
\]

Then the transition matrix \( W \) is defined as \( W = AD^{-1} \). The entries in \( W \) represent the probability of the transition from one node in \( G \) to another. Formally:

\[
w_{ij} = p(X^{t+1} = i | X^t = j), \quad \forall i, j, 1 \leq i, j \leq n,
\]

where \( X^t \in \{1, \ldots, n\} \) is the state random variable of the Markov chain. Note that we assume the Markov chain is time-homogeneous, i.e. \( w_{ij} = p(X^{t+1} = i | X^t = j) \) remains the same for any time \( t \geq 0 \).

Consider the eigendecomposition of the transition matrix \( W \): \( Wu_i = \lambda_i u_i \), \( \forall i, 1 \leq i \leq n \), where \( u_i \) is the \( i \)th eigenvector associated with eigenvalue \( \lambda_i \). According to the Perron-Frobenius Theorem, the following two properties hold:

**Property 1.** If we sort the eigenvalues in descending order, we have \( 1 \geq \lambda_1 \geq \cdots \geq \lambda_n \geq -1 \).

**Property 2.** Given a transition matrix \( W \) with all its entries being strictly positive, there exists an eigenvector \( u \) associated with the largest eigenvalue \( 1 \), whose entries satisfy \( u(i) > 0, \forall i, 1 \leq i \leq n \), where \( u(i) \) is the \( i \)th entry of \( u \); and \( \sum_{i=1}^{n} u(i) = 1 \). We call \( u \) the (normalized) principal eigenvector of \( W \).

The stationary distribution is a time-invariant measure that characterizes the behavior of a Markov random walk. Specifically, given a \( n \)-state time-homogeneous Markov random walk, as defined by the \( n \times n \) transition matrix \( W \), we can define its stationary distribution as follows:

**Definition 1** (Stationary Distribution). Let \( \pi = (\pi_1, \ldots, \pi_n)^T \), where \( \pi_i = p(X^t = i) \) is the probability of node \( i \) being visited by a random walk (as defined by \( W \)) at time \( t \). If at any time \( t \geq 0 \), \( \pi \) satisfies:

\[
\pi_i = \sum_{j=1}^{n} \pi_j w_{ij}, \quad \forall i, 1 \leq i \leq n,
\]

then \( \pi \) is called the stationary distribution of the random walk.

It is well-known that the stationary distribution of a given random walk can be derived from the (normalized) principal eigenvector of the transition matrix. Formally, given a strictly positive transition matrix \( W \), from Property 2 and Definition 1, we have

**Property 3.** Given a strictly positive transition matrix \( W \), the stationary distribution of the random walk is equal to the (normalized) principal eigenvector of the transition matrix \( W \): \( \pi = u \).

Intuitively, given a (global) random walk in graph \( G \), the less likely a node is visited by the random walk, the more likely it is a (global) outlier. Therefore previous work [8] used the stationary distribution as the global outlier score. Formally:

**Definition 2** (Global Outlier Score). Given a random walk graph \( G \) and its transition matrix \( W \), \( \pi_i \) is the global outlier score for node \( i \), \( \forall i, 1 \leq i \leq n \).

The smaller the score is, the more likely node \( i \) is a (global) outlier. Informally, the global outlier score of a node is the chance that a random walk in the graph will visit that node when the random walk has reached equilibrium.

III. CONTEXTUAL RANDOM WALKS AND CONTEXTUAL OUTLIERS

In the previous section we discussed finding global outliers using the global random walks, where an outlier is a node that is unlikely to be visited regardless of where the random walk starts. Though useful, this approach cannot identify contextual outliers since no contextual information is present or used. We now discuss our approach which can identify contextual outliers using the non-principal eigenvectors of a transition matrix and interpret them as the stationary expectation of contextual random walks.
In our model, each non-principal eigenvector of the transition matrix uniquely defines a 2-labeling/2-coloring of the graph. Intuitively, given a 2-coloring of the graph, each subgraph can be considered as a context. Let \( S^+ \) be one subgraph and \( S^- \) the other, we can then determine the chance of a node being visited given the random walk starts from \( S^+ \) and \( S^- \), respectively. Without loss of generality, if a node in \( S^+ \) is more likely to be visited by the random walk starting from \( S^+ \) than from \( S^- \), then it can be considered as a contextual inlier w.r.t. \( S^+ \). On the other hand, there will be some unusual nodes whose chance of being visited by the random walk starting from either \( S^+ \) or \( S^- \) is about the same, i.e., these nodes don’t belong strongly to either \( S^+ \) or \( S^- \). We call these nodes contextual outliers. We assign contextual outlier scores to every node in the graph so that the contextual outliers can be discovered.

A. Contextual Random Walk and Stationary Expectation

We first introduce the definitions and properties of the contextual random walk and the stationary expectation. Assume \( G \) is a random walk graph associated with a strictly positive transition matrix \( W \). First we define:

**Definition 3 (Contexts and Contextual Random Walk).** Let \((S^+, S^-)\) be a 2-coloring of \( G \), where \( S^+ \) is the index set of nodes labeled as + while \( S^- \) is the index set of nodes labeled as –. \( S^+ \) and \( S^- \) satisfy \( S^+ \neq \emptyset \), \( S^- \neq \emptyset \), \( S^+ \cup S^- = \{1, \ldots, n\} \). We call \((S^+, S^-)\) a pair of contexts of the graph \( G \). A random walk in \( G \) with the existence of contexts is then called a contextual random walk.

Now we consider the following indicator random variable:

\[
Y^i_t = \begin{cases} 
1 & X^t = i, X^0 \in S^+ \\
-1 & X^t = i, X^0 \in S^- \\
0 & \text{otherwise}
\end{cases}
\]  

(4)

where \( 1 \leq i \leq n \) and \( t = 0, 1, 2, \ldots \). To put it into words, if node \( i \) is visited by a random walk at time \( t \), and if the random walk started from the context \( S^+ \), we set \( Y^i_t \) to 1; if the random walk started from \( S^- \), we set \( Y^i_t \) to −1; if node \( i \) is not visited at time \( t \), then we set \( Y^i_t \) to 0. We calculate the mathematical expectation of \( Y^i_t \) as follows:

\[
E(Y^i_t) = p(X^t = i, X^0 \in S^+) - p(X^t = i, X^0 \in S^-).
\]  

(5)

Consequently, if \( E(Y^i_t) \) is (relatively) close to 1 (or −1), it indicates that node \( i \) is more likely to be visited by the random walk starting from \( S^+ \) (or \( S^- \)). However, if the expectation is close to 0, it means that node \( i \) is (almost) equally likely to be visited by a random walk starting from either context, which effectively makes the node anomalous as compared to the other nodes which are more strongly "aligned" with \( S^+ \) or \( S^- \).

Though \( E(Y^i_t) \) is informative for identifying contextual outliers, it cannot be used as a contextual outlier score directly as it has the problem of being time-dependent, i.e., it is not a constant and always changes as \( t \) increases. Therefore we introduce a time-invariant measure which can better help characterizing the structure of the random walk graph and identifying contextual outliers. The time-invariant measure is, similar to the stationary distribution of a global random walk, the stationary expectation of a contextual random walk. First we define the stationary expectation of a contextual random walk:

**Definition 4 (Stationary Expectation).** Given the random walk graph \( G \) and its transition matrix \( W \), we say the expectation of \( Y^i_t \), which is \( \mu_i \), is stationary if for all \( t \) the following condition holds:

\[
\mu_i = c \sum_{j=1}^{n} \mu_j w_{ij}, \quad \forall i, 1 \leq i \leq n.
\]  

(6)

where \( c \) is a time-independent constant. We shall refer to \( \mu = (\mu_1, \ldots, \mu_n)^T \) as the stationary expectation of the contextual random walk in \( G \) w.r.t. \( S^+ \) and \( S^- \).

Now the question becomes how we can find a stationary expectation \( \mu \) given the transition matrix \( W \). We will show that if \( W \) is strictly positive, then each of its non-principal eigenvectors uniquely determines a pair of contexts and the corresponding stationary expectation. Specifically, following Property 1, let \( v \) be an eigenvector of \( W \) associated with the eigenvalue \( \lambda < 1 \). We call \( v \) a non-principal eigenvector of \( W \) and the following lemma holds:

**Lemma 1.** Given a non-principal eigenvector \( v \) of a strictly positive transition matrix \( W \), we have \( \sum_{i=1}^{n} v(i) = 0 \), where \( v(i) \) is the \( i \)th entry of \( v \).

**Proof:** The proof is trivial and omitted due to page limit. Please refer to textbooks on spectral analysis, say [10].

With Lemma 1, we can use \( v \) to define a 2-coloring of \( G \), which gives us a pair of contexts:

\[
S^+ = \{i : v(i) > 0\}, \quad S^- = \{i : v(i) < 0\}.
\]  

(7)

Now consider the contextual random walk in \( G \) w.r.t. \((S^+, S^-)\), we have the following theorem:

**Theorem 1 (The Stationary Expectation of a Contextual Random Walk).** If we set \( \mu = (\mu_1, \ldots, \mu_n)^T \) to be

\[
\mu_i = \frac{v(i)}{\sum_{j=1}^{n} |v(j)|}, \quad \forall i, 1 \leq i \leq n,
\]  

(8)

where \( v \) is a non-principal eigenvector of \( W \) associated with the eigenvalue \( \lambda \), then Eq.(6) will hold. Hence \( \mu \) as defined in Eq.(8) is a stationary expectation of the contextual random walk.

**Proof:** Omitted due to page limit. Please contact the authors for the extended version.
Theorem 1 shows that each non-principal eigenvector uniquely determines a 2-coloring of the graph, \((S^+, S^-)\), and its stationary expectation, \(\mu\).

### B. Contextual Outlier and Stationary Expectation

With Theorem 1, we can now define the contextual outlier score using the stationary expectation.

**Definition 5** (Contextual Outlier Score). Given a random walk graph associated with the transition matrix \(W\), the contextual outlier score of node \(i\) is \(|\mu_i|\), where \(\mu_i\) is the stationary expectation as defined in Eq.(8), w.r.t. the contexts \((S^+, S^-)\) as defined in Eq.(7).

According to our definition, the contextual outlier score of any node is always between 0 and 1. A large score means that the node is highly expected to be visited by a random walk starting from one of the two contexts, and is thus a contextual *inlier*; a small score means that the node is equally likely to be visited by the random walk starting from either context, and is thus a contextual *outlier*.

Our contextual outlier score is time-invariant, and is solely determined by the structure of the random walk graph. Note that since the transition matrix \(W\) has \(n - 1\) non-principal eigenvectors, thus we can potentially have \(n - 1\) pairs of contexts, and we can compute for every node in the graph a contextual outlier score w.r.t. each pair of contexts.

An important advantage of our contextual outlier score is that it covers the global outlier score (based on the stationary distribution) as a special case. Formally, we have the following corollary:

**Corollary 1.** The stationary distribution \(\pi\) is a special case of stationary expectation, where \(\lambda = 1\) and \(S^+ = \{1, \ldots, n\}\), \(S^- = \emptyset\).

Corollary 1 says that we can re-interpret the global outlier score within our framework and compare it directly to our contextual outlier score. Consequently, we can produce a unified ranked list containing both global outliers and contextual outliers, ordered by their anomalousness.

### IV. Algorithm

In this section, we discuss the implementation of our contextual outlier score in practice. We propose a hierarchical algorithm which iteratively partitions the data set until the size of the subgraph is smaller than a user-specified threshold \(\alpha\). Both global and contextual outliers are detected and ranked during each iteration. The outline of our algorithm is shown in Algorithm 1.

The input of our algorithm is a graph \(G\) and its associated transition matrix \(W\). The transition matrix \(W\) is generated by normalizing a given similarity matrix \(A\), where \(a_{ij}\) is the similarity between the \(i^{th}\) and \(j^{th}\) data instances. The choice of similarity function is application-dependent. In our experiments we show that promising results are obtained using the Euclidean distance as well as the inner product.

While we hierarchically partition the graph \(G\) into smaller subgraphs, we use a queue \(Q\) to store the subgraphs to be partitioned. A user-specified threshold \(\alpha\) is used to decide when we stop to further partition a subgraph, since as the subgraph becomes smaller, it’s less likely to have meaningful contexts within itself.

The output of our algorithm is a ranked outlier list \(L\), whose entries are tuples in the form of

\[
\text{instance, context, score},
\]

where `instance` is the index of the data instance; `context` is the context with respect to which that data instance is examined; `score` is the outlier score of that data instance. Note that one instance may appear more than once in \(L\) because it has different outlier scores with respect to each pair of contexts.

#### Algorithm 1: Hierarchical contextual outlier detection

**Input:** Random walk graph \(G\) with transition matrix \(W\), queue \(Q\), threshold \(\alpha\);

**Output:** A sorted list \(L\), consisting of tuples as defined in Eq.(9);

1. \(Q \leftarrow \emptyset, L \leftarrow \emptyset; Q\).enqueue \((G, W)\);
2. repeat
   3. \((G, W) \leftarrow Q\).dequeue();
   4. if \(|G| > \alpha\) then
      5. Compute the (normalized) principal eigenvector of \(W\), which is \(u\);
      6. foreach \(i \in G\) do
         7. Add \(\{i, G, u(i)\}\) to \(L\); /* global outliers */
      end
      9. Compute the Fiedler eigenvector of \(W\), which is \(v\);
      10. \(S^+ \leftarrow \{i : v(i) > 0\}\), \(S^- \leftarrow \{i : v(i) < 0\}\);
      11. foreach \(i \in S^+\) do
           12. Add \(\{i, S^+, v(i) / \sum_{j=1}^{n} v(j)\}\) to \(L\);
                /* contextual outliers in \(S^+\) */
      end
      14. foreach \(i \in S^-\) do
           15. Add \(\{i, S^-, v(i) / \sum_{j=1}^{n} v(j)\}\) to \(L\);
                /* contextual outliers in \(S^-\) */
      end
   end
   17. Generate the transition matrices for \(S^+\) and \(S^-\), respectively;
   18. \(Q\).enqueue \((S^+, W^+);\)
   19. \(Q\).enqueue \((S^-, W^-);\)
3. until \(Q\) is empty;
to different contexts.

Our algorithm involves computing the first and second largest eigenvalues and eigenvectors of an \( n \times n \) matrix, where \( n \) is the number of data instances. Therefore its complexity is dominated by the complexity of eigendecomposition. Note that if the transition matrix is generated from the \( k \)-nearest-neighbor graph, then it will be very sparse when \( k \) is small, which leads to much faster eigendecomposition.

V. EMPirical Study

A. Methodology

The evaluation of contextual outlier detection itself is an open problem because there is no commonly accepted ground truth for contextual outliers. In this work, we generated ground truth using the class labels of real-world data sets. Specifically, given a data set with class labels, we first convert it into a random walk graph. Then we discover two contexts within it using the 2-coloring indicated by the Fiedler eigenvector. Note that this partition is equivalent to a normalized Min-Cut of the graph. Next we label each context by the label of the majority class in that context. The minority class in each context is then labeled as true contextual outliers. This can be interpreted as the contextual outliers being the instances most likely to contain class label noise. The ground truth contexts are given by the class labels and contextual outliers are those instances most likely not to be of this class. We apply our method to rank contextual outliers in each context, respectively, and compare our answer to the ground truth. Hence we essentially turn the outlier detection problem into a retrieval problem, where we can compute the precision, recall and f-score of our method. Note that due to the limited availability of ground truth (labels), we did not let the threshold \( \alpha \) to decide when the iterative partition should end; the number of partitions performed was decided by the number of classes in the data.

We also implemented a baseline method for comparison. It uses the same contexts defined by the Fiedler vector. But instead of computing the contextual outlier score, it ranks global outliers within each context, separately, using the stationary distribution based method as described in [8]. Note that we chose this method instead of popular outlier detection techniques say LOF [11] because this method adopted the same random walk graph model as the one used by us, and it ranks outlier in a probabilistic framework, thus is ready for direct comparison with our algorithm.

B. Results and Analysis

The first data set we used was Iris from the UCI Archive. It has 3 classes: setosa, versicolor and virginica. Each class has 50 instances and each instance has 4 attributes. For visualization purposes, we projected the data set onto a 2-dimensional space, using the standard PCA technique. Then we converted the data set into a transition matrix using Euclidean distance.

When we applied our method to discover contexts, we noticed that the setosa class can be perfectly partitioned from the remaining 2 classes, which means that there is no contextual outlier in these contexts (setosa vs. the remaining two). Thus we removed setosa and continued to partition the rest of data (Fig. 1(a)). As a result, the first context contained 54 instances, among which 43 were versicolor and 11 virginica. Thus the first context was labeled as versicolor and had 11 true contextual outliers. Similarly, the other context was labeled as virginica and had 7 true contextual outliers (Fig. 1(b)).

We scored contextual outliers using our method (Contextual Outlier Detection, COD) as well as the baseline method (Baseline), respectively. Both methods reported the top-10 contextual outliers from each context. We can clearly see in Fig. 1(c) that our method effectively identified most of the contextual outliers, while the baseline method tended to report data points that are far away from the majority of the entire data set, but ignored the true contextual outliers (Fig. 1(d)). In fact, as shown in Fig. 2, our method consistently outperformed the baseline method in terms of precision, recall and f-score, in both contexts. More importantly, our method had high precision when only reporting a small number of outliers, which is favorable in practice. Recall that our contextual outlier score covers the global outlier score as a special case and thus makes it possible to measure the interestingness of global and contextual outliers in a unified framework.

To further justify the effectiveness and advantage of our method, we used the 20 Newsgroups data set, which consists of articles from 20 different topics, 1000 of each. Each time we extracted two classes from the data set and constructed the random walk graph using inner product distance. We tried random combinations of classes and com-
Table I

RESULTS ON 20 NEWSGROUPS DATA (COD VS. BASELINE)

<table>
<thead>
<tr>
<th>Selected classes</th>
<th>Outlier ratio</th>
<th>Precision</th>
<th>Recall</th>
<th>F-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>crypt &amp; electr.</td>
<td>.0440</td>
<td>.5463</td>
<td>.2720</td>
<td>.2445</td>
</tr>
<tr>
<td>crypt &amp; med</td>
<td>.0155</td>
<td>.5202</td>
<td>.2759</td>
<td>.2814</td>
</tr>
<tr>
<td>crypt &amp; space</td>
<td>.0110</td>
<td>.3412</td>
<td>.2239</td>
<td>.1600</td>
</tr>
<tr>
<td>electr. &amp; med</td>
<td>.2745</td>
<td>.5085</td>
<td>.4395</td>
<td>.2547</td>
</tr>
<tr>
<td>electr. &amp; space</td>
<td>.0910</td>
<td>.5319</td>
<td>.2601</td>
<td>.2659</td>
</tr>
<tr>
<td>med &amp; space</td>
<td>.0215</td>
<td>.4035</td>
<td>.2440</td>
<td>.1945</td>
</tr>
</tbody>
</table>

![Figure 2](image.png)

Figure 2. Precision, recall, and f-score on trimmed iris data, against the number of outliers reported.

We computed the average precision, recall and f-score. Specifically, we chose four science related topics, namely sci.crypt, sci.electrons, sci.med, and sci.space. We enumerated all the 6 possible combinations between them. We report the average precision, recall, and f-score at the point where the number of reported outliers is equal to the number of true outliers, as shown in Table I. Outlier ratio is the ratio of true contextual outliers in the selected data set. Again, the results showed that our method (left columns) performed twice as good as the baseline method (right columns) in terms of precision, recall, and f-score.

VI. CONCLUSION

Contextual outlier detection typically requires a context to be specified a priori. In this work, we explore automatic and unsupervised identification of contexts and the contextual outliers therein. Our approach is applicable to graph data as well as vector data if this data can be converted to a graph where the edge weights correspond to the similarity between points.

We identify contexts as a 2-coloring of a random walk graph. We introduce the notion of stationary expectation, which is a generalization of the stationary distribution, as our contextual outlier score. For a given node its contextual outlier score characterizes the difference in the chance of a random walk performed in the entire graph (not just the subgraph) visiting the node given the walk starts from either context. Our contextual outlier score is time-invariant and is solely determined by the structure of the random walk graph. Note that when we identify contexts, we do not modify the graph structure in any way such as by removing edges. Therefore, our approach is not the same as performing a multi-way cut on the graph and then applying global outlier detection separately in each individual subgraph.

Our algorithm produces a ranked list of tuples of the form {instance, context, score}. Note that an instance may appear multiple times in this list, but w.r.t. different contexts and different contextual outlier scores. We validated the effectiveness of our method by empirical results on real-world data. Our method consistently outperformed the baseline method on different data sets.

ACKNOWLEDGMENT

The authors thank the NSF and ONR for the support of this research via grant IIS-0801528 and N00014-09-1-0712.

REFERENCES