A Renewal Theory Approach to Anomaly Detection in Communication Networks

Brian Thompson
Rutgers University
bthom@cs.rutgers.edu

Tina Eliassi-Rad
Rutgers University
eliassi@cs.rutgers.edu

Introduction

Any medium for human interaction can be modeled by a network graph, where nodes represent people or computers, and an edge signifies a relationship between two entities. However, communication networks such as email and phone-call networks are characterized by their highly dynamic nature. For example, the fact that Alice and Bob are friends says nothing about the frequency or regularity of their communication. Analyzing communication patterns across a network should therefore take into account not just the graph structure, but also a wealth of temporal information. In this work, we build a model for representing and understanding activity in a communication network, and propose a novel approach for identifying anomalous behavior of individuals and groups. Experiments on a variety of real-world datasets show the effectiveness and scalability of our approach, as well as a clear and intuitive visual interface.

Model and Approach

We propose that to understand communication, which is inherently a temporal process, one must study it on the level of communication patterns over time between two entities. Only then can edges be placed within the context of their local graph structure or the network as a whole. This enables a more fine-grained analysis that is sensitive to both sudden and gradual changes, and also provides a sound basis for quantifying the degree of anomaly in a subgraph at any scale.

We model communication between two entities as a sequence of time-stamped events, signifying the times at which communication took place. To analyze these sequences, we appeal to the field of renewal theory. A *renewal process* is a continuous-time Markov process where new events occur with inter-arrival times (IATs) sampled from independent and identical distributions (IIDs).
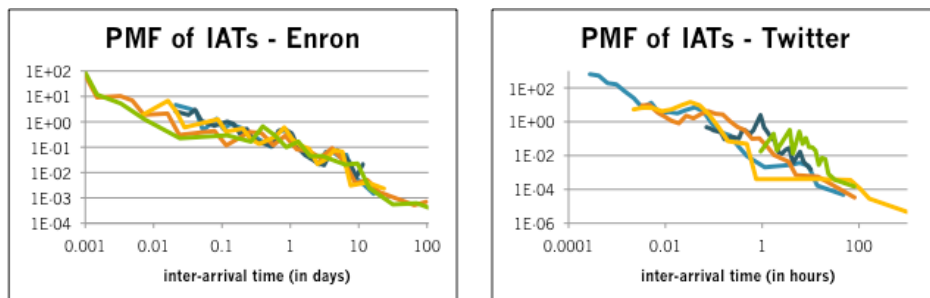

Figure 1. Distribution of IATs for the five most active edges in the Enron and Twitter datasets on a log-log scale.

It has been observed that IATs reflecting human interaction frequently exhibit a power-law distribution. This is evident in the linearity (on a log-log scale) of the IAT distributions for the Enron email and Twitter message datasets (see Figure 1), despite the different time scales. Similar results hold for Bluetooth-connection and IP-traffic datasets. The Bounded Pareto Distribution is a common power-law distribution for modeling real-world data. For each edge, we learn the parameters of the distribution (x_{min}, x_{max}, and shape parameter α) using a Maximum-Likelihood Estimation method, which allows us to maintain and update our model in real-time and constant space.

We now define the *recency* of an edge. Formally, a recency function $Rec: 2^T \times T \rightarrow [0,1]$ assigns a weight to an edge e at time t based on the age of its renewal process – i.e., the time since the last event. Specifically, $Rec(e)$ is 0 at the time a new event occurs, 1 at time x_{max}, and uniform over $[0,1]$ when sampled uniformly in time. Given an IAT distribution, there is a unique recency function that satisfies these criteria.

One problem with describing edges based on the absolute amount of time since the last event is that high-activity edges would always appear to be recent, overshadowing any changes in behavior of other edges. We call this phenomena *time-scale bias*. By defining recency to be uniform over $[0,1]$ for all edges, we effectively eliminate time-scale bias, thus providing an unbiased quantitative comparison of edges across all activity levels.

We could stop now and claim that edges with recency score below a fixed threshold are anomalous. A problem with this approach, however, is that an edge exhibiting normal behavior (according to the model) is guaranteed to be labeled as anomalous a constant fraction of the time, leading to potentially many false positives. Furthermore, we have not considered the relationships between different edges in the graph. In fact, we have not made use of any properties of the graph structure, which ought to be a central theme of anomaly detection in communication networks. In our next step, we consider the collective behavior of a subset of edges.

Divergence and the MCD Algorithm

Consider the weighted graph $G = (V, E)$ representing a communication network at a time t, with $w(e) = Rec(e, t)$. For $E' \subseteq E$ and $p \in [0,1]$, let $X_{E',p} = |\{e \in E': w(e) \leqslant p\}|$ -- i.e., the number of edges in E' with

$Rec(e, t). \leqslant p$. We define the *p-divergence* of E' as follows:

$$Div_p(E') = \frac{1}{P(X \geq X_{E',p})}$$

where $X \sim Bin(|E'|, p)$.

If our IAT distribution model is accurate and edges are independent, a *p*-divergence of *d* means that the probability of at least $X_{E',p}$ edges occurring *p*-recently is *1/d*. A subgraph with high divergence thus indicates significant correlation of edges occurring recently. But how do we choose the correct threshold *p*? Anomalies across different edge sets at different times may only be apparent at different thresholds. We address this challenge by introducing the concept of *max-divergence*: $Div_{max}(E') = \max \{Div_p(E')\}$ over all $p \in [0,1]$.

To take graph structure into account, we look at sets of edges that form connected components in the graph. We define a *maximal p-component* of *G* to be a connected subgraph $C = (V', E')$ for which the following conditions hold: (1) $w(e) \leqslant p$ for all $e \in E'$; and (2) $w(e) > p$ for all $e \notin E'$ incident to V'. Throughout the paper, *component* refers to a maximal *p*-component for some *p*. For convenience, we define the max-divergence of a vertex $v \in V(G)$ as $Div_{max}(v) = Div_{max}(E(v))$ and the divergence of a *p*-component $C \subseteq G$ as $Div(C) = Div_p(E(C))$.

We now present the MCD (Maximal Component Divergence) Algorithm (see Figure 2), which identifies multiple components of high divergence in a graph. (1) Calculate edge weights using the Recency function. (2) Increase the threshold, updating component divergence values as necessary. (3) Output disjoint components with max divergence.

Let $\hat{C}(G)$ denote the component with the highest divergence in *G*. The algorithm builds the MCD Tree of all components, and returns the following set of disjoint components: $C^1 = \hat{C}(G)$, $C^2 = \hat{C}(G - C^1)$, …, $C^k = \hat{C}(G - C^1 - \cdots - C^{k-1})$.
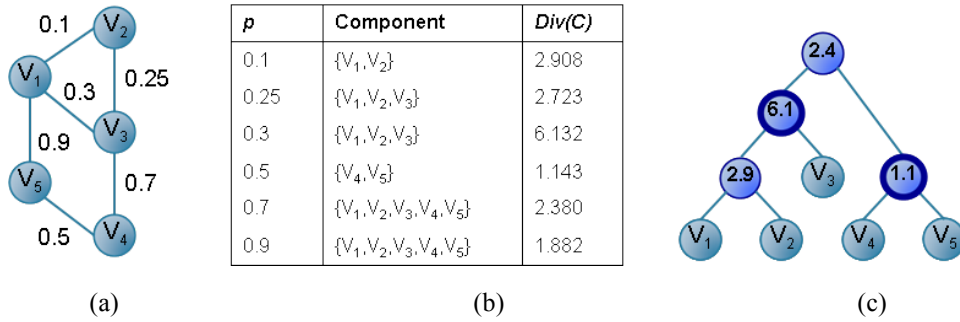
Figure 2. The MCD Algorithm: (a) a recency-weighted graph; (b) components
formed at various thresholds; (3) the MCD Tree with output highlighted

Experimental Results

We test the effectiveness of our approach using four datasets: (1) Enron – a collection of emails sent between Enron employees over the 5 years preceding the Enron scandal; (2) Bluetooth – collected by the Reality Mining Lab at MIT, showing proximity of Bluetooth devices; (3) LBNL – IP traffic over a 1-hour period, with over 9 million timestamps, including hand-labeled scanning activity; and (4) Twitter – messages sent between 250,000 users from 2007-2009.
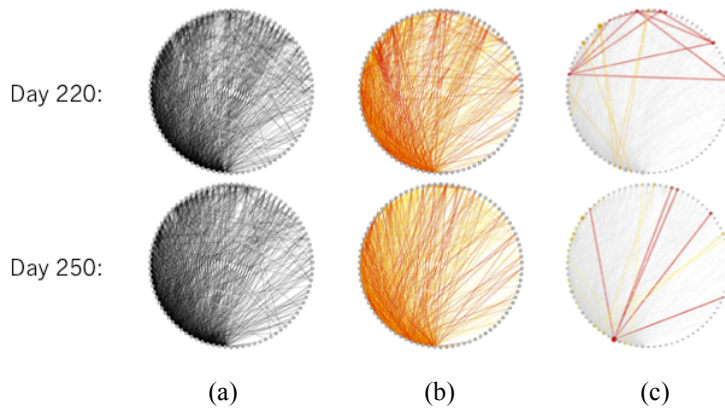

Figure 3. Reality Mining Bluetooth dataset: (a) shows edges sorted by degree,
(b) shows edges colored by recency, and (c) shows edges ranked by the MCD analysis

Figure 3 illustrates how MCD analysis can reveal anomalous activity in a network. This is a key contribution of our work: our algorithm can be run as a stand-alone application, or as a tool to assist IT administrators in identifying nodes with suspicious behavior. In Figure 4a, a simple plot of MCD over time identifies hand-labeled scanning activity, as well as other anomalies overlooked by human analysts. Figure 4b shows the scalability of our algorithm.
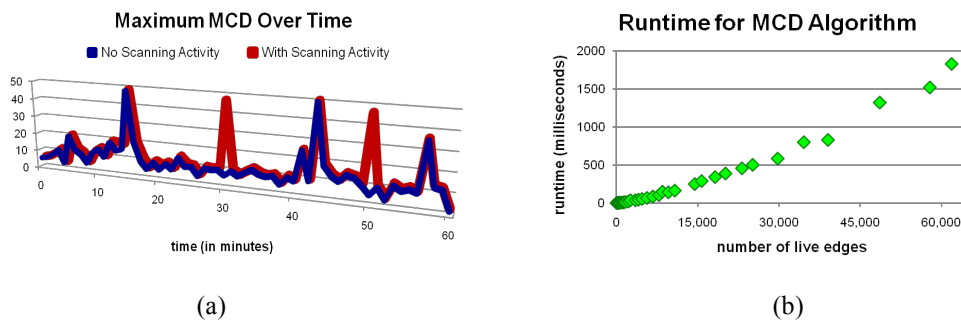

Figure 4. (a) MCD analysis for the LBNL dataset; (b) runtime analysis on the Twitter dataset

Related Work and Conclusions

Time-evolving networks have recently become a hot research topic. [1] provides an empirical study of laws governing weighted time-evolving graphs. [2] gives an overview of change mining. [3] is a recent survey on anomaly detection.

In [4] and [6] algorithms are presented to identify vertices with anomalous neighborhoods, and [5] develops an anomaly detection algorithm to flag times of abnormally high activity. However, these approaches rely on simple graph or vertex statistics or expensive local computations, and are subject to time-scale bias.

Studying inter-arrival times is a novel approach for analyzing communication networks. Our algorithm is streaming and runs in $O(m)$ space and $O(m\ log\ m)$ time, where m is the number of edges in the dataset. MCD analysis provides output that can be easily visualized and used as a tool for monitoring activity in a variety of real-world domains.

References

[1] L. Akoglu. M. McGlohon, and C. Faloutsos. *RTM: Laws and a Recursive Generator for Weighted Time-Evolving Graphs*. ICDM, 2008.

[2] M. Boettcher, F. Hoeppner, and M. Spiliopoulou. *On Exploiting the Power of Time in Data Mining*. SIGKDD Explorations 10(2), pp. 3-11, 2008.

[3] V. Chandola, A. Banerjee, and V. Kumar. *Anomaly Detection: A Survey*. ACM Computing Surveys, July 2009.

[4] K. Henderson, T. Eliassi-Rad, C. Faloutsos, L. Akoglu, L. Li, K. Maruhashi, B.A. Prakash, and H. Tong. *MetricForensics: A Multi-level Approach for Mining Volatile Graphs*. KDD, 2010.

[5] J. Sun, C. Faloutsos, S. Papadimitriou, and P. S. Yu. *GraphScope: Parameter-Free Mining of Large Time-Evolving Graphs*. KDD, 2007.

[6] Y. Park, C. Priebe, D. Marchette, and A. Youssef. *Anomaly Detection Using Scan Statistics on Time Series Hypergraphs*. SDM, 2009.