VOLTIME: Unsupervised Anomaly Detection on Users' Online Activity Volume

Daniel Y. T. Chino * Alceu F. Costa* Agma J. M. Traina* Christos Faloutsos†

Abstract

Is it possible to spot review frauds and spamming on social media and online stores? In this paper we analyze the joint distribution of the inter-arrival times and volume of events such as comments and online reviews and show that it is possible to accurately rank and detect suspicious users such as spammers, bots and fraudsters. We propose VOLTIME, a generative model that fits well the inter-arrival time distribution (IAT) of real users. Thus, VOLTIME automatically spots and ranks suspicious users. Experiments on several real datasets, ranging from `Reddit` comments and phone calls to `Flipkart` product reviews, show that VOLTIME is able to accurately fit the activity volume and IAT of real data. Additionally, we show that VOLTIME ranks suspicious users with a precision higher than 90% for a sensitivity of 70%.

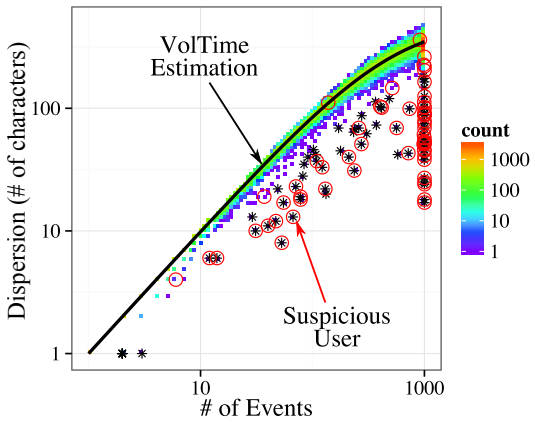
1 Introduction

Suppose that user 'Alice' uploaded 20 reviews to an app-store, all exactly 85 characters long - is this suspicious? How about 'Bob', who uploaded 30 reviews, one every 10 minutes (but of variable length)? Most people would agree that both 'Alice' and 'Bob' are suspicious, and deserve further investigation. The reason for this agreement is that, for such a high count of events, a real human's activity would have higher variety ("dispersion") - that is, higher count of distinct values. This is one of the main insights behind this paper, and we show how to use it, to model real user behavior, and to spot impostors.

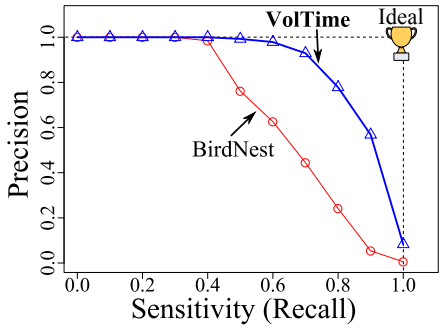
Social media services and online review platforms influence opinions [21, 12, 23, 8] and even purchasing decisions [16]. This has created issues such as spam [10], spreading of rumors [3] and fake reviews [26]. Detecting these issues is important to improve user's experience. Thus, given the activities of a large number of users, can we find the user with the strangest behavior? Specifically, we present two inter-related problems:

1. **Modeling:** How can we model human behavior across different platforms (social media, online stores)? We want to model both the *temporal*-aspect (such as the inter-arrival time of the events of a user), jointly with the *volume* of activity (number of characters in the review, or phone call duration, among other aspects).
2. **Anomaly Detection:** How can we use these models to detect anomalies such as spammers, bots and fraudsters, like 'Alice', and 'Bob' in our earlier example.

To answer these questions, we analyze data from different domains, including comments from a social media service (Reddit), reviews from an online store (Flipkart) and phone calls from a large Asian city. From each platform, we analyze the joint distribution of inter-arrival times (IAT) and volume (comment and review length; phone call duration) of



(a) DISPERSION-PLOT



(b) Bot-Detection using DispersionScore

Figure 1: **VOLTIME detects anomaly successfully**: (a) The DISPERSION-PLOT reveals strange behaviors as outliers. (b) VOLTIME outperforms competitors on accuracy.

*University of Sao Paulo, {chinodyt, alceufc, agma}@icmc.usp.br
†Carnegie Mellon University, SCS and iLab, christos@cs.cmu.edu

communication events.

Our first contribution is the introduction of the *dispersion* metric, which we use to measure the variability of users' behavior. Considering the example of 'Alice', her dispersion would be 1 since all her comments have the same length. However, users with a larger variability in the comment length would have a large dispersion. We also propose a visualization named DISPERSION-PLOT, which illustrates the relationship between the dispersion and number of events for different users. Figure 1(a) shows the DISPERSION-PLOT of `Reddit` users. While typical users form a single cluster, suspicious users (indicated by red circles), clearly deviate from this pattern.

The second contribution of this paper is VOLTIME, a model that generates synthetic inter-arrival times (IATs) and event volumes. An important property of VOLTIME is that it closely matches the typical users' dispersion. In Figure 1(a), the black line, which corresponds to the expected dispersion of our VOLTIME, accurately follows the behavior of typical users. This allow us to use VOLTIME to generate a score that measures users' suspiciousness. That is, users whose dispersion deviate most from VOLTIME's dispersion will have a higher VOLTIME score.

Figure 1(b) shows the VOLTIME performance for the task of detecting bots in Reddit. VOLTIME detected 70% of the bots with a precision of 90%. The main contributions of this paper are summarized as follows:

- **Patterns - Population behavior:** We proposed the dispersion (Equations 5.5 and 5.9) to analyze how the joint distribution of inter-arrival times and volume changes as users produce more events. By analyzing the dispersion across several diverse datasets through the DISPERSION-PLOT, we show that normal users present a similar behavior while bots, fraudsters and spammers clearly deviate from this pattern;
- **VOLTIME - Generative model:** Based on the patterns observed using DISPERSION-PLOT we propose VOLTIME, a generative model that is able to describe the inter-arrival times of communication events across all the studied domains;
- **Usefulness - Anomaly detection:** We used VOLTIME to automatically rank users according to their suspiciousness. VOLTIME was able to detect bots using only time-stamp and event volume data;

For *reproducibility*, our code is open-sourced at `http://chinodyt.github.io/`. The rest of the paper is organized starting with related work, followed by the problem formulation, dataset patterns, the proposed model, anomaly detection, experiments and conclusions.

2 Background and Related Work

Modeling Human Dynamics: The dynamics of human activity is a widely studied topic [18, 28, 11, 24, 14, 19], as it has applications that range from resource management [17] and user clustering [9, 22] to anomaly detection [26]. A well-known model for the timing of human activity is the Poisson-Process [15, 5, 27]. Other works argue that IAT distribution of human activities can be better modeled by heavy-tailed distributions such as power-laws [2]. Recent models for human dynamics include the Self-Feeding Process (SFP) [32], Cascading Non-homogeneous Poisson Process [22] and Rest-Sleep-and-Comment model (RSC) [6]. There are also works that focus on the activity volume (number of characters or call duration) as Truncated Lazy Contractor (TLAC) [31] for call duration. In this paper we propose a model for human activity (VOLTIME) that describe both the timing of activities and the volume of an event, what to the best of our knowledge was not done so far.

Anomaly Detection: There are many works devoted to detect anomalies based on user activity [4, 30, 20, 13, 29, 25]. In [33] the authors proposed a method that consists in creating a scatter-plot of the minute vs. the second for all comment time-stamps of a user. This plot is then used to spot users from Twitter that are suspicious of being bots. In [16] the authors proposed BIRDNEST, which consist of two steps. A model named BIRD (Bayesian Inference for Rating Data), which describes the statistical properties of the timing and ratings in online commerce using a Bayesian model. Based on BIRD, the authors introduced NEST (Normalized Expected Surprise Total), a suspiciousness metric to detect fraudsters. Table 1 compares our proposed method with existing methods.

Table 1: Summary of different models for human dynamics.

	Kleinberg [18]	Poisson [22]	SFP [32]	RSC [6]	TLAC [31]	BIRDNEST [16]	Zhang/Paxson [33]	VOLTIME
Models IAT	✓	✓	✓	✓		✓		✓
Models Volume					✓	✓		✓
Models Both						?		✓
Visualization							✓	✓
Spots anomalies	✓	✓		✓	✓	✓	✓	✓

3 Problem Formulation

In this Section, we outline the problem of modeling the user behavior on online social media. Table 2 gives the list of symbols used throughout the paper.

3.1 How do individual users behave online? On social services, users interact with each other by posting comments

Table 2: Concepts and Symbols

Concepts	**Interpretation**
Activity Volume	Characteristic of the online activity.
Dispersion	Number of non-empty bins.
DISPERSION-PLOT	Visualization tool of population behavior.
DispersionScore	Suspicioness of a user.
Symbols	**Definitions**
n	Number of events of a user.
$\mathcal{T} = \{t_1, t_2, \dots\}$	Multiset of timestamps of a user.
$\mathbf{\Delta} = \{\Delta_1, \Delta_2, \dots\}$	Multiset of inter-arrival times of a user.
$\mathcal{V} = \{v_1, v_2, \dots\}$	Multiset of activity volumes of a user.
$e_i = (\Delta_i, v_i)$	Event at instant t_i.
$\mathcal{E} = \{e_1, e_2, \dots\}$	Multiset of events of a user.
$D(\mathcal{E})$	Dispersion of events $\mathcal{E}$.
$\hat{D}(n)$	Expected dispersion of n events.
τ	DispersionScore.
p_s	Probability of user entering state s.
LL	Log-logistic distribution.
$\theta_{k,s} = \{\alpha_{k,s}, \beta_{k,s}\}$	Log-logistic parameters of attribute k and state s.
Θ	Set of VOLTIME parameters.

on a Reddit forum or making mobile phone calls on a certain timestamp. We are given a user, as shown on Figure 2, with a multiset of activities timestamps $\mathcal{T} = \{t_1, t_2, \dots\}$, where $t_i \leq t_{i+1}$. As the user interacts with a social service, he/she can generate an activity volume v_i at every t_i. The activity volume v is an attribute that describes the amount of the user interaction, for example, v can describe the length (number of characters) of comments/reviews or the duration of a phone call.

For simplicity, we will denote each user interaction with social services as an event e_i represented by the ordered pair (Δ_i, v_i), where $\Delta_i = t_{i+1} - t_i$. A user that interacts n times will generate a multiset of activities events $\mathcal{E} = \{e_1, \dots, e_n\}$. It is important to note that among the infinite possibilities of describing a user, on this paper we will be using the inter-arrival time Δ (IAT) between events. The issue of using timestamps directly is that it is not able to generalize the behavior of users that are more active during

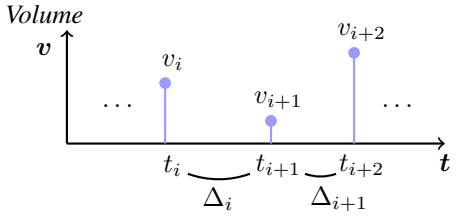


Figure 2: Users can post at any time t_i an activity of volume v. The volume may correspond to the number of characters (e.g. textual comments) or duration (e.g. phone calls).

different times of day. With these considerations in mind, the first problem can be stated as follows:

PROBLEM 1. *(*MODELING STATISTICAL PROPERTIES*) Given a multiset of events $\mathcal{E} = \{e_1, \dots, e_n\}$, where each event $e_i = (\Delta_i, v_i)$, $1 \leq i \leq n$, is a pair of IAT (Δ_i) and activity volume (v_i). What is the joint distribution of the multiset?*

3.2 How can we spot anomalies? A right community on online social services helps the users to have better experience. With that in mind, is it possible to describe the behavior of the community of users? Do the more active users have the same behavior of the less active? These questions bring the main problems of this paper:

PROBLEM 2. *(*SUCCINCT FEATURE EXTRACTION*) Given a multiset of n events (Δ, v), find few features to describe its behavior.*

PROBLEM 3. *(*SPOT SUSPICIOUS USERS*) Given several multisets of events from different users, find a score describing how suspicious a specific user is.*

Our ultimate goal is to solve the Problem 3. To achieve this goal, we first handle with the Problem 1 by understanding and describing how the majority of users behave in terms of the joint distribution of IAT and volume (Section 4). Then in Section 5, we answer Problem 2 by extracting two features from a user's behavior (multiset of events) (see Equation 5.5). In the same Section 5, we answer Problem 3 using Equation 5.6.

4 Modeling Statistical Properties

Is it possible to model the patterns of the users' behavior? In this section we discuss the patterns found on users' online activity on real-world datasets. We also point the implications of our findings and how to model their behavior.

4.1 Datasets Description We analyzed four real-world datasets of users activity events, such as social media posts, e-commerce reviews and mobile phone calls. The datasets are summarized in Table 3 and described in details as follows.

Table 3: Summary of real-world datasets.

Dataset	# of Users	# of Events
Reddit	94,739	35,979,723
Flipkart	158,638	409,679
SWM	113,145	163,873
LAC	1,696,602	280,814,170

Reddit: The Reddit dataset consists of comments posted by users on Reddit. Reddit allows users to submit content, as text posts or URL links. The dataset was originally collected and used in [6]. Of the 94 thousand users, 60 users are known bots inserted by the authors. Since the authors aimed at bot detection, we also checked the dataset for spammers and users that now got their account deleted or banned.
Flipkart: The Flipkart dataset consists of reviews written by users on the Flipkart e-commerce network, which provides a platform for sellers to market products to costumers. Users can write reviews of products using between 100 and 5000 characters.
Software Marketplace (SWM): The SWM dataset contains reviews in an anonymous online software marketplace. For this dataset the timestamp has a granularity of a day, there is no information about the time the review was posted. The dataset was originally collected by [1].
Large Asian City (LAC): The LAC dataset has information of phone calls made on a large anonymous Asian city. For this dataset, it was collected the timestamp of the beginning of a call and its duration.

For the Reddit, Flipkart and SWM datasets the activity volume represents the length of the text comment (number of characters). The IAT is calculated as the difference between timestamps of consecutives events. For the LAC dataset the activity volume represents the duration of a phone call in seconds. Since phone calls have a different nature, the IAT was calculated as the difference of the end of call timestamp and the beginning of the next call.

4.2 Online Activity Event Patterns The focus of this paper is to analyze the behavior of the user's online activity events. As stated in the begining of Section 3, an activity event is the ordered pair of IAT and activity volume. The activity events of a user can be seen by his/her heatmap, a visualization that shows the relationship between the IAT and the activity volume. The frequency of (Δ, v) is shown using a color coding, more frequent events are reddish and less frequent are bluish. The heatmap can show the behavior of a single user or show how the entire population behaves.

Figures 3 and 6 show the heatmap for the population of each dataset. When analyzing the activity volume, we make the following observation:

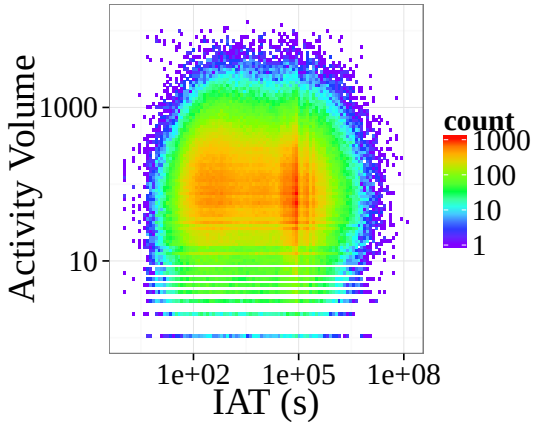


Figure 3: The heatmap of the Reddit dataset showing the behavior of users. Users have two distinct behaviors, an in-session with activities in short bursts and an out-session with a larger IAT.

OBSERVATION 1. *The Activity Volume can be accurately modeled by a mixture of log-logistic distribution.*

The log-logistic (LL) distribution was previously used to model human activity, as phonecall duration [31] and users' activity on social media (number of posts, likes and photos) [7]. The log-logistic PDF is:

$$LL_{PDF}(x;\alpha,\beta) \sim \frac{(\beta/\alpha)(x/\alpha)^{\beta-1}}{(1+(x/\alpha)^{\beta})^2} \tag{4.1}$$

where α is a scale parameter and β is a shape parameter.

It is also possible to notice that there are two modes on the activity events for all datasets (see Figure 6). During the first mode, users appear to be more active, generating events with inter-arrival times between 5 to 10 minutes. On the other hand, during the second mode (around 3 hours), they make a post and rest before generating a new event. We summarize these observations as follows:

OBSERVATION 2. *The events' IAT can be described by a mixture of two log-logistics. The first log-logistic corresponds to short intervals, generated by bursts of activity. The second log-logistic is generated when users are less active or resting.*

4.3 VOLTIME Model How can we generate a model capable of following the Observations 1 and 2? In this section we introduce VOLTIME, a generative model that is capable to describe the interval and volume of human communication in different media. The goal of VOLTIME is to describe two aspects of human communication: (i) the inter-arrival times (IAT) between events and (ii) the volume of each event. VOLTIME is a generative model that creates pairs of synthetic IAT and event volumes that matches statistical properties from real data. With VOLTIME, we can answer the Problem 1.

In order to respect Observation 2, VOLTIME uses a Markov chain to transition between two states: in-session and out-session. Figure 4 shows the state digram for VOLTIME. If VOLTIME is in the in-session state, there is a probability p_{out} to transition to the out-session state and a probability $1 - p_{out}$ to remain in the in-session state. Similarly, if VOLTIME is in the out-session state, the transition probability to the in-session state is p_{in}.

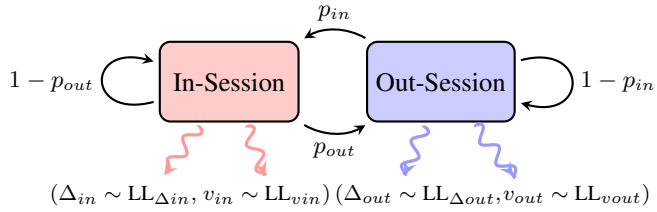


Figure 4: State diagram of VOLTIME. After each state transition VOLTIME generates an IAT Δ and an event volume v sampled from independent log-logistic distributions.

As noted on Observation 1, VOLTIME uses a LL distribution to model volume and IAT. After each state transition, VOLTIME generates an event tuple $e = (\Delta_s, v_s)$ for the current state s (either in-session or out-session). In each state, VOLTIME waits a time interval Δ_i sampled from a LL distribution with parameters $\theta_{\Delta,s}$ and generates an event volume v_i sampled from a LL with parameters $\theta_{\Delta,s}$.

To estimate the parameters of VOLTIME we are given an observed input multiset of IAT and activity volumes. We start by finding the probabilities $P(s_i = in)$ and $P(s_i = out)$ that the i-th event is in the in-session and out-session states, respectively. We assume that the distribution of IAT is mixture of two log-logistics with two components corresponding to the in-session and out-session. This allows us to estimate $P(s_i = in)$ and $P(s_i = out)$ using an expectation-maximization (EM) algorithm.

In order to estimate the log-logistic parameters $\theta_{\Delta,in}$ and $\theta_{v,in}$ that will be used to generate the IAT and event volumes for the in/out-session state, we randomly sample the IAT and volumes from the input sequences while weighting according to the probabilities $P(s_i = in)/P(s_i = out)$. Finally, the sampled IAT and event volumes are used to estimate the log-logistic parameters using the maximum-likelyhood estimation (MLE) method. The complexity of the EM algorithm is linear on the size of the multiset of events. Similarly, the complexity of the MLE algorithm is linear on the number of samples used to estimate the parameters of the log-logistic distributions. Now, let $\text{LL}(X;\theta)$ denote a log-logistic distribution with random variable X and parameters θ.

LEMMA 4.1. (VOLTIME PDF) *The joint probability distribution $f(\Delta, v)$ of the events IAT and volume generated by* VOLTIME *is given by:*

$$f(\Delta, v) = w_{in} \cdot LL(\Delta; \theta_{\Delta,in}) \cdot LL(v; \theta_{v,in}) + w_{out} \cdot LL(\Delta; \theta_{\Delta,out}) \cdot LL(v; \theta_{v,out}) \tag{4.2}$$

where:

$$w_{in} = \frac{p_{in}}{p_{in} + p_{out}}, w_{out} = \frac{p_{out}}{p_{in} + p_{out}} \tag{4.3}$$

5 Spotting Suspicious Activities

How can we spot suspicious users by analyzing their behavior? In Section 4 we proposed VOLTIME to model users' behavior. However, instead of using all 10 parameters of VOLTIME to spot anomalies, we propose a succinct feature extraction, allowing us to visually spot anomalies.

5.1 Population behavior In Section 1 we introduced 'Alice' and 'Bob' who have suspicious behaviors. How could we describe them? For example, if we consider the multiset of activity volume $\{85, 85, \ldots, 85\}$ of the 20 reviews that 'Alice' wrote, a natural feature is the size of the multiset ($n = 20$). What other features can we extract? Entropy? Second moment? We now introduce you the definition of dispersion. The dispersion summarizes how the users behave online and can be used to spot anomalies. Suspicious users will have lower values of dispersion than typical users. And this is our proposed answer to the Problem 2, for each user, with a multiset of events, we extract two features: (a) the number n of events and (b) the dispersion, as defined below:

DEFINITION 1. (DISPERSION) *Given a multiset of n integer numbers $\mathcal{X} = \{x_1, \ldots, x_n\}$. The dispersion D_{1d} of the multiset $\mathcal{X}$ is the count of distinct values ('vocabulary').*

For example, given a multiset of integers $\{1, 2, 1, 5, 5, 2, 5\}$, $n = 7$ and $D_{1d} = 3$. Formally, given x_i an integer in $(1, 2, \ldots, \infty)$, let I_j denote an indicator variable such that $I_j = 1$ if there is at least one i so that $x_i = j$. The dispersion D_{1d} is given by:

$$D_{1d} = \sum_{j=1}^{\infty} I_j \tag{5.4}$$

The same idea can be applied to a multiset of 2-d points.

DEFINITION 2. (DISPERSION 2-D) *Given a multiset $\mathcal{Y}$ of n two-dimensional points (x, y), where both x and y are integers, the dispersion is calculated as follows:*

$$D_{2d} = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} I_{i,j} \tag{5.5}$$

For example, the multiset $\{(1, 1), (1, 3), (1, 1)\}$ has dispersion $D_{2d} = 2$. In our case, the pairs correspond to events

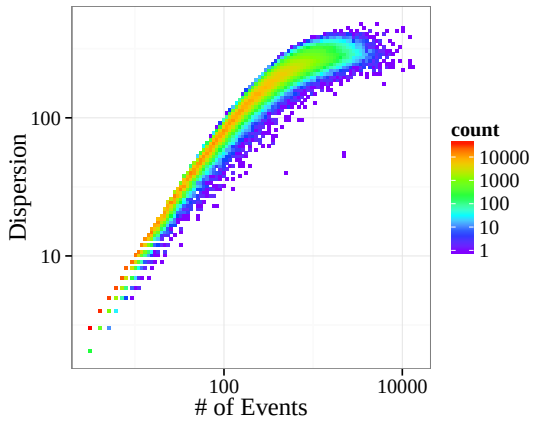


Figure 5: DISPERSION-PLOT shows the relationship between the number of events and dispersion on LAC dataset.

(Δ, v). Since Δ and v have a continuous nature, the vocabulary would be huge and we may lose information. To overcome this, we make them integers using bucketization. We partionate them in log-bins, because we expect skewed distributions in both of them. This concludes our response to the Problem 2: For a given user, with a multiset of (IAT, volume) pairs, we map him/her to a 2-d point: (n, D_{2d}). For the remaining of this paper we will denote the 2-d dispersion as D.

We are ready to tackle Problem 3, namely, *how strange is a given user "X", as compared to a large set of users*. The intuition behind our response, is to map all those users (including user "X"), to such 2-d points, as shown in Figures 1(a) and 5. We propose to name such a plot as a DISPERSION-PLOT, and, since there is heavy over-plotting, we make it a heatmap. We expect to see a clear trend, and specifically, a (non-linear) correlation between dispersion and event-count n; this correlation would of course depend on the joint distribution of (IAT, volume). The upcomming Lemmas 5.1 and 5.2 quantify this correlation, between n and expected dispersion, which gives the black line on Figure 1(a).

Our final answer to the question '*how strange is user "X"?*' is intuitively the distance of the 2-d image of user "X", from the expectation ("black line" in Figure 1(a)).

Formally, we have the following: Let $D(\mathcal{E})$ denote the dispersion (Equation 5.5) of the event multiset $\mathcal{E} = \{e_1, \cdots, e_n\}$. Let $\hat{D}(n)$ denote the expected dispersion from n samples randomly sampled from a joint probability distribution of VOLTIME. The DispersionScore is computed as follows:

$$\tau = |\log \hat{D}(n) - \log D(\mathcal{E})| \tag{5.6}$$

5.2 Expected Dispersion The only missing part is how to estimate the expected dispersion $\hat{D}$, as a function of the sample size n, and given the joint distribution of (IAT, volume). The answer is Equation 5.9, but we need some lemmas first. We start by showing the Expected Dispersion lemma for one dimension:

LEMMA 5.1. (EXPECTED DISPERSION 1D) *Given a multiset of n integers $\mathcal{X} = \{x_1, \ldots, x_n\}$ and P_i the probability of an $x \in \mathcal{X}$ to be equal i. The expected dispersion is:*

$$\hat{D}_{1d}(n) = \sum_{i=1}^{\infty} [1 - (1 - P_i)^n] \tag{5.7}$$

Proof. Let $\mathcal{X}$ and P_i be as described in Lemma 5.2. Let I_i denote an indicator variable such that $I_i = 1$, if there is at least one w where $x_w = i$. The expected value of I_i is:

$$E(I_i) = 1 - (1 - P_i)^n \tag{5.8}$$

Equation 5.7 can be obtained combining Equations 5.4 and 5.8. ■

Lemma 5.1 can be extended for a 2-d multiset.

LEMMA 5.2. (EXPECTED DISPERSION) *Given a multiset of n 2-d points $\mathcal{Y} = \{(x_1, y_1), \ldots, (x_n, y_n)\}$ and $P_{i,j}$ the probability of a $(x, y) \in \mathcal{Y}$ to be equal (i, j). The expected dispersion is:*

$$\hat{D}(n) = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} [1 - [1 - P_{i,j})]^n] \tag{5.9}$$

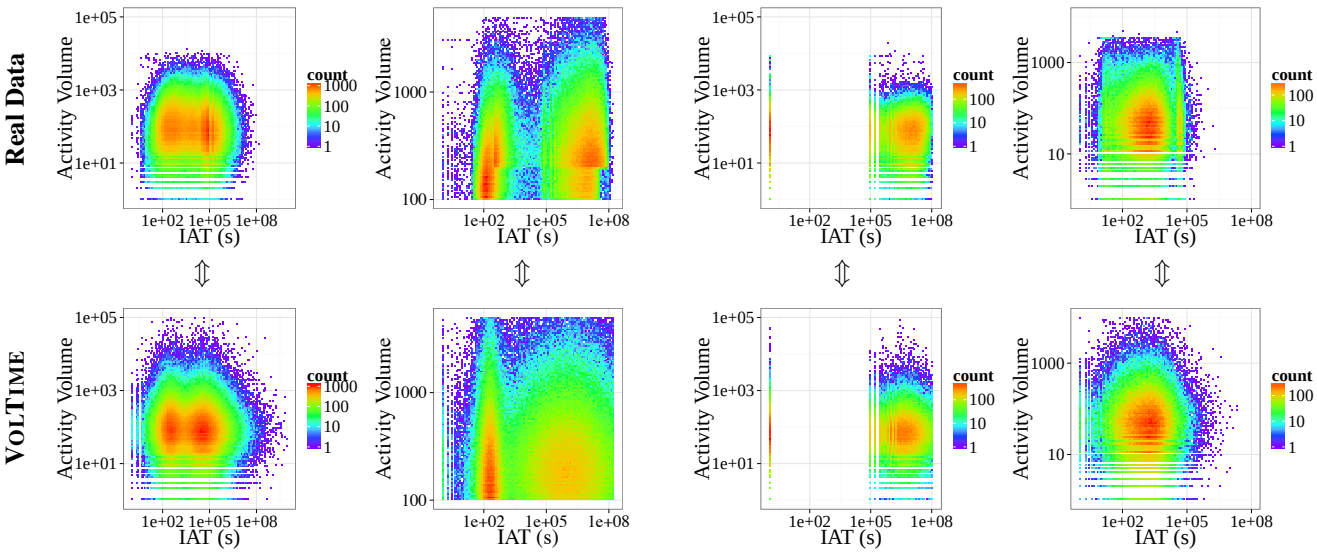
Proof. Easy generalization of Lemma 5.1. ■

Notice that if we have a continuous 2-d distribution, we can always digitize it to an integer-valued 2-d distribution. Formally, for our setting, the joint probability $P_{i,j}$ of an event falling in the (i, j) bin is computed as follows:

$$P_{i,j} = \int_{\Delta'_j}^{\Delta'_{j+1}} \int_{v'_i}^{v'_{i+1}} f(\Delta, v) d\Delta dv \tag{5.10}$$

where $f(\Delta, v)$ is the VOLTIME PDF described by Equation 4.2.

The complexity to calculate the DispersionScore is the complexity to calculate the expected dispersion $\hat{D}$ and the user's dispersion D. Considering that we already have the VOLTIME PDF, the complexity of $\hat{D}$ is $\mathcal{O}(m)$, where m is the total number of discrete bins. The Expected Dispersion can be calculated only once for each number of events n. The complexity to calculate D is $\mathcal{O}(n)$, where n is the user's number of events. Since we only need to count the total number of events and the number of distinct events. The complexity to compute the dispersion for each user is linear to the size of the dataset.



(a) Reddit (b) Flipkart (c) SWM (d) LAC

Figure 6: Heatmap of synthetic data generated by VolTime model for each dataset. On all datasets, VolTime was able to correctly model the In/Out-sessions behavior, showing the capability of correctly modeling the activity events of users.

6 VolTime in Action

In this section we show how well VolTime can fit real data. To the best of our knowledge, there is no work aimed at modeling the joint distribution of IAT and volume. The parameters were estimated using the algorithm described in Section 5 on all datasets. Figure 6 shows the heatmap for the synthetic data generated by VolTime.

For all datasets, VolTime managed to model the in/out-session behavior. We made a modification on the activity volume generation of VolTime due to the Flipkart dataset limitation on the number of characters. We also modified VolTime to only generate IAT with intervals of one day for the SWM dataset, due to its granularity. The correctness of VolTime shows its robustness to different granularities. The LAC dataset has a different behavior than the other datasets. The VolTime model was able to generate the in-session correctly, but did not manage to model the less intense out-session spike. Although VolTime presents this issue, Section 7 shows that VolTime can predict the behavior of the population.

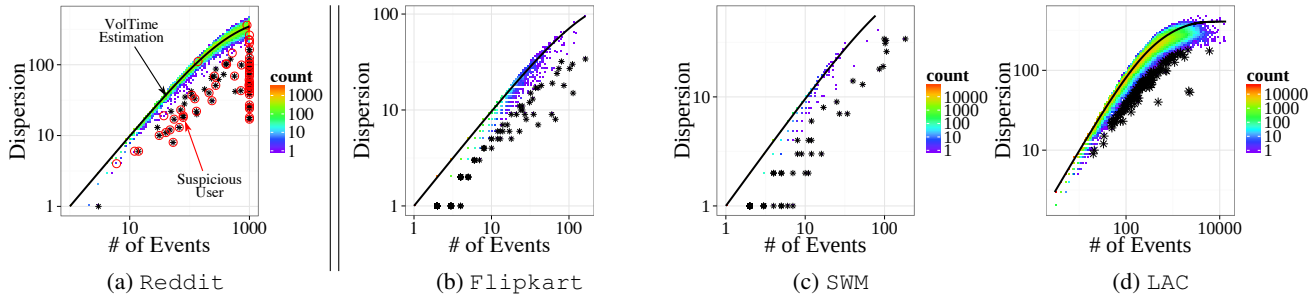
7 Spotting Suspicious Activities with the DispersionScore

In this section we show how well the DispersionScore can spot suspicious users. We used Equation 5.9 to estimate the expected dispersion and calculate the DispersionScore (τ) for a given number n of events. Figure 7 shows the DisperSion-Plot for each dataset. The solid black line is the expected dispersion $\hat{D}(n)$, where n is the number of events. For all datasets, the $\hat{D}(n)$ falls on the behavior of typical users, represented by the green and red areas, showing that VolTime was able to correctly predict the dispersion given the number of activity events. The black stars (*) represent the suspicious users spotted by VolTime and known suspicious users are marked as red circles. Since only the Reddit dataset has a ground truth, experiments on the other datasets discuss the top suspicious users found by our method. The results will be detailed as follows.

Reddit: The result obtained by VolTime on Reddit users are shown in Figure 7(a). More than 80% of the known suspicious users are marked with a black star, showing the correctness of VolTime. We compared VolTime with BIRDNEST [16], but considering the activity volume as its ratings. The activity volume was log-binned to better adapt to BIRDNEST. Figure 8(a) shows the precision vs sensitivity (recall) obtained by VolTime and BIRDNEST. VolTime spotted **80%** of the suspicious users with a precision greater than **85%**, being up to **2.39 times more accurate** than BIRDNEST.

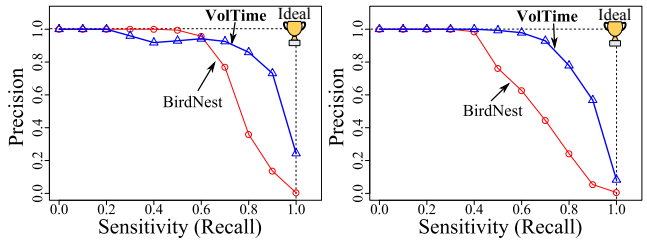
Note that on Figure 7(a), there are some black star users that were not labeled as suspicious on the ground truth. We manually checked these users and spotted suspicious activities: users that only post URL or spammers or had their accounts deleted/banned. The same procedure was done with the BIRDNEST output. Figure 8(b) shows the result considering the new suspicious users. This time, DisVolT spotted **70%** of the suspicious users with a **precision greater than 90%**, while BIRDNEST had a precision of **44%**.

Flipkart and SWM: On both datasets, VolTime spotted spammer users. On Flipkart, the majority of the spam reviews do not add too much information for future buyers, since it has generic adjectives. Usually the top suspicious users post all their reviews in short bursts and in a short



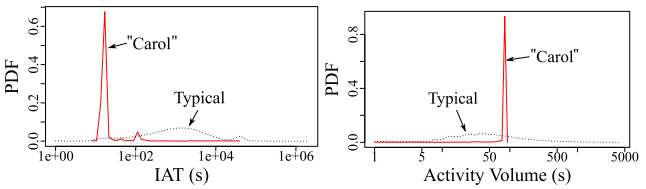
(a) Reddit (b) Flipkart (c) SWM (d) LAC

Figure 7: **DISPERSION-PLOT spots outliers:** DISPERSION-PLOT showing the usefulness of VOLTIME. The solid black line is the expected dispersion. The black stars are the spotted suspicious users ($\tau \geq 1$). (a) The red circles are the confirmed suspicious users.



(a) Ground truth (b) Relabeled ground truth

Figure 8: Precision of VOLTIME spotting suspicious users on the Reddit dataset. VOLTIME in blue is closer to ideal.



(a) "Carol"'s IAT distribution. (b) "Carol"'s activity volume distribution.

Figure 9: **Closer inspection of a suspicious user:** The top suspicious user ("Carol") found by VOLTIME on the LAC dataset. The red line shows the behavior of the suspicious users and the black dotted line is the typical user behavior.

time span. One of the top suspicious user wrote the same 60 reviews on different products in less than 1 hour. VOLTIME was also able to spot users that use a variety of template review texts on different products. One user wrote the same review for different movies of the same actor, just changing the title of the movie. Also, we noted that different users sometimes used the same review text to review different products. All of the top 20 reviewers spotted by VOLTIME had this same behavior.

On SWM, the majority of the top suspicious users just promote some kind of code associated with an app. They usually promote these codes to their own benefit by saying that new users that use their codes will get free points or cash. Usually the top suspicious users posts all their reviews on the same day or in less than a week. Every user from the top 20 users spotted by VOLTIME have similar review texts that promote their codes, always offering promises of free points and cash. We listed below some reviews:

- Flipkart: *"The item quality is very good and its look is very well really appreciate. Highly Recommended item buy again. Fast shipping."*
- SWM: *"Download [redacted] for some free cash!!! Sign up using [redacted] for some points."*

LAC: VOLTIME spotted users with suspicious behavior, like "Carol". The behavior of "Carol" is shown on Figure 9, the solid red line is "Carol"'s behavior and the dotted black line the typical user behavior. "Carol" has over two thousand calls in short bursts to the same person. Notice that the typical user has a smoother distribution of IAT and activity volume. The majority of the top suspicious users found by the VOLTIME also have this same behavior.

8 Conclusions

The contributions of this paper are as follows:

- **Patterns:** We proposed dispersion (Definitions 1 and 2) to quantify the variability of inter-arrival times and volume of events generated by users of different platforms, such as social media services and phone networks.
- **Model:** We introduced VOLTIME, a model for the joint distribution of IAT and volume of events generated by users (Figure 4). We show that our model can accurately fit real data (Figure 6), and, more importantly, match the dispersion metric of human users (Figure 7).
- **Anomaly Detection:** We used VOLTIME to calculate DispersionScore that measures users' suspiciousness (Equation 5.6). Users whose dispersion deviate most from VOLTIME's dispersion will have a higher score. Taking advantage of DispersionScore, we managed to spot **70%** of the suspicious users with a precision higher than **90%** on the Reddit dataset (Figure 8).

Acknowledgement

We are grateful to Flipkart for providing the data in the experiments. This material is based upon work supported by FAPESP, CNPq, CAPES, i-LAB, the National Science Foundation under Grants No. CNS-1314632, IIS-1408924,

by the Army Research Laboratory under Cooperative Agreement Number W911NF-09-2-0053, and by ARO/DARPA under Contract Number W911NF-11-C-0088. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation, or other funding parties. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

References

[1] L. Akoglu, R. Chandy, and C. Faloutsos. Opinion Fraud Detection in Online Reviews by Network Effects. In *ICWSM 2013*, pages 2–11, 2013.

[2] A. Barabási. The origin of bursts and heavy tails in human dynamics. *Nature*, 435(7039):207–211, 2005.

[3] A. Bessi, F. Petroni, M. Del Vicario, F. Zollo, A. Anagnostopoulos, A. Scala, G. Caldarelli, and W. Quattrociocchi. Viral misinformation: The role of homophily and polarization. pages 355–356, 2015.

[4] H. Cheng, P.-N. Tan, C. Potter, and S. Klooster. Detection and characterization of anomalies in multivariate time series. In *SDM*, volume 9, pages 413–424. SIAM, 2009.

[5] J. Cho and H. Garcia-Molina. Estimating frequency of change. *ACM TOIT*, 3(3):256–290, 2003.

[6] A. F. Costa, Y. Yamaguchi, A. J. M. Traina, C. Traina Jr., and C. Faloutsos. RSC: Mining and Modeling Temporal Activity in Social Media. In *KDD*, pages 269–278, 2015.

[7] P. Devineni, D. Koutra, M. Faloutsos, and C. Faloutsos. If walls could talk: Patterns and anomalies in facebook wallposts. In *ASONAM '15*, pages 367–374, 2015.

[8] P. A. Dow, L. A. Adamic, and A. Friggeri. The Anatomy of Large Facebook Cascades. In *ICWSM*, pages 145–154, 2013.

[9] J.-P. Eckmann, E. Moses, and D. Sergi. Entropy of dialogues creates coherent structures in e-mail traffic. *PNAS*, 101(7):14333–14337, 2004.

[10] S. Fakhraei, J. Foulds, M. Shashanka, and L. Getoor. Collective Spammer Detection in Evolving Multi-Relational Social Networks. In *KDD*, pages 1769–1778, 2013.

[11] A. Goyal, F. Bonchi, and L. V. Lakshmanan. Learning influence probabilities in social networks. In *WSDM*, pages 241–250. ACM, 2010.

[12] P. H. C. Guerra, A. Veloso, W. Meira Jr., and V. Almeida. From bias to opinion: a transfer-learning approach to real-time sentiment analysis. In *KDD*, pages 150–158, 2011.

[13] S. Günnemann, N. Günnemann, and C. Faloutsos. Detecting Anomalies in Dynamic Rating Data: a Robust Probabilistic Model for Rating Evolution. In *KDD*, pages 841–850, 2014.

[14] C. A. Hidalgo R. Conditions for the emergence of scaling in the inter-event time of uncorrelated and seasonal systems. *Physica A*, 369(2):877–883, sep 2006.

[15] P. G. Hoel, S. C. Port, and C. J. Stone. *Introduction to Stochastic Processes*. Waveland Pr. Inc., 1986.

[16] B. Hooi, N. Shah, A. Beutel, S. Gunneman, L. Akoglu, M. Kumar, D. Makhija, and C. Faloutsos. Birdnest: Bayesian inference for ratings-fraud detection. In *SDM*, volume 16, pages 495–503. SIAM, 2016.

[17] A. Ihler, J. Hutchins, and P. Smyth. Adaptive event detection with time-varying poisson processes. In *KDD*, pages 207–216, 2006.

[18] J. Kleinberg. Bursty and Hierarchical Structure in Streams. In *KDD*, pages 373–397, Edmonton, Alberta, Canada, 2003.

[19] N. C. Krishnan and D. J. Cook. Activity recognition on streaming sensor data. *PMC*, 10:138–154, 2014.

[20] T. Lappas, M. R. Vieira, D. Gunopulos, and V. J. Tsotras. On the spatiotemporal burstiness of terms. In *VLDB*, pages 836–847, 2012.

[21] J. Leskovec, L. Backstrom, and J. Kleinberg. Meme-tracking and the dynamics of the news cycle. In *KDD*, pages 497–505, 2009.

[22] R. D. Malmgren, J. M. Hofman, L. A. N. Amaral, and D. J. Watts. Characterizing Individual Communication Patterns. In *KDD*, pages 607–616, 2009.

[23] Y. Matsubara, Y. Sakurai, B. A. Prakash, L. Li, and C. Faloutsos. Rise and fall patterns of information diffusion: model and implications. In *KDD*, pages 6–14, 2012.

[24] R. Ottoni, D. L. Casas, J. P. Pesce, W. Meira Jr., C. Wilson, A. Mislove, and V. Almeida. Of Pins and Tweets: Investigating How Users Behave Across Image-and Text-Based Social Networks. In *ICWSM*, pages 386–395, 2014.

[25] J. Pan, Y. Liu, X. Liu, and H. Hu. Discriminating bot accounts based solely on temporal features of microblog behavior. *Physica A*, 2016.

[26] S. Rayana and L. Akoglu. Collective Opinion Spam Detection: Bridging Review Networks and Metadata. In *KDD*, pages 985–994, 2015.

[27] K. C. Sia, J. Cho, and H.-K. Cho. Efficient monitoring algorithm for fast news alerts. *TKDE*, 19(7):950–961, 2007.

[28] C. Tantipathananandh, T. Berger-Wolf, and D. Kempe. A framework for community identification in dynamic social networks. In *KDD*, pages 717–726, 2007.

[29] M. Tsytsarau, T. Palpanas, and M. Castellanos. Dynamics of News Events and Social Media Reaction. In *KDD*, pages 901–910, 2014.

[30] A. Vahdatpour and M. Sarrafzadeh. Unsupervised discovery of abnormal activity occurrences in multi-dimensional time series, with applications in wearable systems. In *SDM*, volume 10, pages 641–652. SIAM, 2010.

[31] P. O. S. Vaz de Melo, L. Akoglu, C. Faloutsos, and A. A. F. Loureiro. Surprising Patterns for the Call Duration Distribution of Mobile Phone Users. In *PKDD*, pages 354–369, 2010.

[32] P. O. S. Vaz de Melo, C. Faloutsos, R. Assunção, R. Alvez, and A. A. F. Loureiro. Universal and Distinct Properties of Communication Dynamics: How to Generate Realistic Inter-event Times. *TKDD*, 9(3):24:1–24:31, 2015.

[33] C. M. Zhang and V. Paxson. Detecting and Analyzing Automated Activity on Twitter. *LNCS*, 6579:102–111, 2011.