"The Lull before the Storm": Event Detection from Group Absenteeism

***Abstract*—Event detection in online social media has primarily focused on identifying abnormal spikes, or bursts, in activity. However, disruptive events such as socio-economic disasters, civil unrest, and even power outages, often result in abnormal troughs involving group absenteeism of activity. We present the first study, to our knowledge, that models absenteeism and uses detected absenteeism as a basis for event detection in location based social networks (LBSN) such as Twitter. Our framework addresses the challenges of (i) early detection of absenteeism, (ii) identifying the point of origin, and (iii) identifying groups or communities underlying the absenteeism. Our approach uses the formalism of graph wavelets to represent the spatiotemporal structure and user activity in a LSBN. This formalism affords multiscale analysis, enabling us to detect anomalous behavior at different graph resolutions, which in turn allows identification of event location and anomalous groups underlying the network. We introduce a systematic group anomaly detection method using graph wavelets to detect absenteeism group and burst group simultaneously. The effectiveness of our approach is highlighted with three case studies involving Twitter activity over Latin American countries.**

I. Introduction

Social Microblogs such as Twitter and Weibo are experiencing explosive growth, with billions of users globally sharing their daily status updates online. For example, Twitter has more than 310 million average monthly active users (78% from mobile) as of March 31, 2016, and an estimated increase of 25% per year[1]. Various studies have shown that Twitter is viable as a social "sensor", and holds great promise for detecting and forecasting significant societal events [5], [24]. In recent years, a significant body of research [1], [16], [18], [19], [24], [25], [31], [32], [34] has focused on modeling bursts and increases of user activity in social media.

However, real world events are not only correlated with burst signals, but can also exhibit unusually low levels of activity in social networks. As shown in Figure 1, a protest in the city of Natal, Brazil began at 5:00 PM (local time) at the Museum of the Republic, with people gradually joining the demonstration. On Twitter, there was an uncharacteristic lull in activity or *group absenteeism* behavior from 6:00 PM—8:00 PM on the same day. Another example comes from December 24, 2013, southern Brazil experienced widespread flash floods. According to news sources, more than 50,000 people were forced to flee their homes in Minas Gerais and Espirito Santo, in the southern states of Brazil. Immediately following the floods, Twitter activity in this region dropped by 51%, and reached its lowest point that evening.

[1]http://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/

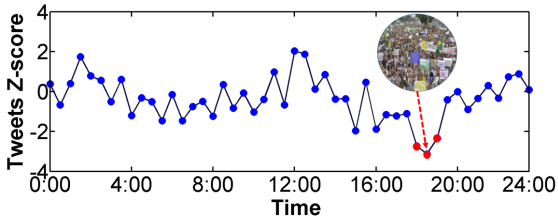


Figure 1: Detected group absenteeism in Natal, Brazil beginning at 6:00 PM on June 17, 2013. This absenteeism event coincides with a large protest that happened in the region.

Investigating this phenomenon of unusually calm behavior online holds enormous potential for understanding localized, disruptive societal events. In this paper we focus on group anomaly detection which is not only able to capture burst events, but also absenteeism events, and introduce this important topic as a key data mining task for social media analytics. An *absenteeism* event in social networks can be defined as an event which is characterized by a significant lull in activity such as a sudden, sharp decrease of Twitter volume within a short period of time (and which may precedes a major burst in re-activity). This paper presents the first study to systematically investigate group anomaly in Location Based Social Networks. To appropriately incorporate absenteeism concepts into our detection approach, we must first address the following questions

- How to define group anomaly in graph? How to reconcile both burst and absenteeism scenarios?
- What scale should we select to model the abnormal groups? Which node should be the central point?
- What is the most efficient approach to select abnormal groups that are spatially and temporally localized?
- How do we model an absenteeism signal for event detection? Even though we have clear examples of real world events which can explain the observed absenteeism, not all absenteeism occurrences can be associated with underlying events. Therefore we must be able to differentiate absenteeism from noisy signals for event detection.

Graph wavelet approach display several outstanding advantages to study the above questions: scalability, localization, low computational complexity, and group compact. In this scenario, the data objects are embedded in a general graph as vertices. By employing wavelet transforms on the graph, we can construct a wavelet function with a graph structure. We propose the graph anomaly index depends on graph structure

and absenteeism score vector, which can define whether a graph is abnormal. When the graph is abnormal, we will calculate its wavelet coefficient, to identify the central node and its coverage cities. In this way, we are able to select abnormal groups at different scales. The group anomaly detection methods are varied and proved to be effective in events detection such as man-made protest and natural disasters.

Our contributions are thus:

- To the best of our knowledge this is the first study to modeling group absenteeism as a basis for event detection. We study different group anomaly, either burst or absenteeism, or first absenteeism then subsequent burst, and prove those anomaly is indicative at detecting events such as man-made protests or natural disasters.
- We incorporate graph wavelets as a mechanism to detect the most anomalous subgraphs at different scales. We demonstrate how this is a powerful appraoch for social media analytics.
- We define the graph anomaly index, which can decide whether a graph is abnormal. Furthermore, we are able to locate the central node and identify the abnormal groups.

The rest of the paper is organized as follows. Section II reviews related work and existing methodologies and Section III formalizes the research problem. In Section IV, we discuss the graph wavelet formalism for group anomaly detection. Section V presents extensive experiments for event detection, and the paper concludes with a summary of the research in Section VI.

II. RELATED WORK

a) Group Anomaly Detection: Anomaly detection in Graphs has been well studied using outlier detection [2]. When considering group concept, two directions has been studied [3]: one is anomalies in unlabeled/plain graphs [20], the other is in attributed graphs. In the plain graph anomaly detection, since the only given information is its structure, various features such as distances, communities [29] have been employed to define graph anomaly. In work of [15], more metrics like vertices, edges, degree, weight, connected components are incorporated into detection framework. In attributed graphs, features regarding nodes behaviors make it possible to have a richer graph representation, which is usually tied with some real-world applications. Such as [35] defines the group based on the term of role, and model the normal groups follow the same pattern with respect to their role mixture rates. Other approaches to group anomaly detection include building generative models of group anomalies [33] where the goal is to automatically infer the groups and detect group anomalies in a social network. Typical to mixture models such methods suffer from high computational complexity due to the size of data and are heavily parameterized. In our work, we consider both graph structure and nodes features, propose a graph wavelet based approach for group anomaly detection, which can guarantee the detected group to be automatically compact, with linear computation complexity and scalability.

b) Event Detection: Event detection based on LSBNs is a research area that has attracted significant attention in the last years. Traditional approaches focus on capturing spatiotemporal burstiness of keywords [18], [19], Kalman filtering to track the geographical trajectories of hot spots of tweets related to earthquakes [24]; detecting topics of interest that are coherent in geographic regions [11], [16], [34]; applying clustering-based approaches search for emerging clusters of documents or terms using predefined similarity metrics that consider factors such as term co-occurrences and social interactions [1], [25], [31], [32]; and using the notion of compactness of a graph [22] to detect events. Several statistical methods have also been used, based on Kulldroff's spatial scan statistic[17], to detect spatial outliers [7] and have been applied to a wide variety of domains including transportation networks, civil unrest forecasting [36], and heterogeneous social media graphs [8].

Our approach to event detection problem is conceptually different from above mentioned studies. It includes a graph-theoretic framework to detect group anomaly and correlate them with future events. Although group absence behavior has been widely studied in the area of organizational behavioral studies [12], [26], it remains unexplored in the area of social network analysis. Resembling closely to group anomaly detection in complex networks, our detection approach is further distinguished by its focus on groups rather than individuals.

c) Graph Wavelet: One of key challenges of our research problem is adapting the detection procedure for both missing and bursty activity groups. For this purpose, we incorporate spectral graph wavelets [14] into our algorithm. This strategy has been quite effectively used in multiscale community mining [30]. Wavelet methods based on spectral graph theory have been applied in a wide array data mining areas such as community detection, anomaly detection [6] and other machine learning tasks [9], [13], [23], [27]. By constructing wavelets over graphs we are able take advantage of local information encoded in graph structure and then cluster and identify nodes which are similar in a scale-dependent fashion.

III. PROBLEM SETTING

In this section, we first introduce mathematical notations. Then we formalize our approach to group anomaly detection. We first describe the accompanying notations in section III-A which will be used throughout the paper. Then we formally present problem statement, provide a brief comparison of our approach to conventional solutions, and review the challenging issues that are relevant to event detection problem.

A. Notations

Given an undirected, weighted graph $\mathbf{G}(V, E; f)$, where $V = \{v_0, v_1, ..., v_{N-1}\}$ represents the set of N cities, E refers to the connections between neighboring cities. W is a matrix of non-negative weights associated with each edge, where $e_{ij} \in E$. The function, $f : V \to \mathbb{R}^N$ maps the vertices of graph $\mathbf{G}$, and $f(n)$ stands for the value on the vertex v_n.

Graph $\mathbf{G}$'s adjacency matrix $\mathbf{A}$ is of size $N \times N$, where each element a_{ij} is represented as:

$$a_{ij} = \begin{cases} w_{ij} & \text{when } e_{ij} \in E \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

Here, $\mathbf{A}$ is symmetric since $a_{ij} = a_{ji}$. Let $d_i = \sum_{v_j \in V} a_{ij}$ be the sum of all edge weights that are incident on v_i, and $\mathbf{D}$ be the diagonal matrix denoted as $\mathbf{D} = diag\{d_1, d_2, \ldots, d_N\}$. A Laplacian matrix $\mathcal{L}$ is defined as $\mathcal{L} = \mathbf{D} - \mathbf{A}$. It is a symmetric matrix and has real eigenvalues λ_i such that $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_{N-1} = \lambda_{max}$. The complete set of $\mathcal{L}$'s normalized eigenvectors [4] χ_i for $i = 0, 1, 2, ..., N-1$ is described as:

$$\mathcal{L}\chi_i = \lambda_i \chi_i \quad (2)$$

The set of eigenvalue and normalized eigenvector pair is denoted as:

$$\sigma(\mathbf{G}) := \{(\lambda_l, \chi_l)\}_{l=0}^{N-1}. \quad (3)$$

$\sigma(\mathbf{G})$ is also called graph spectrum of $\mathbf{G}$.

B. Problem Statement

We focus on the problem of group anomaly detection from online social networks, based on the absenteeism behavior observed in user activity in geographically proximal communities or group of cities. Conventionally, this problem can be described as following: *given a graph and absenteeism score vector,* $\mathbf{G}(V, E; f^t)$ *at time interval* t*, select a subset* $\Sigma \subseteq V$*, such that*

$$\Sigma = \underset{P \subseteq V, P \text{ is compact}}{\arg\min} \sum_{v_k \in P} f(k) \quad (4)$$

However, how to define compactness of the selected subset Σ is an open problem. A general solution to this problem is employing a combinatorial optimization method, by defining a constrained objective function over a network one can identify a subset of vertices which maximize the corresponding function [22]. Therefore, Equation 4 can be modified as:

$$\Sigma = \underset{P \subseteq V}{\arg\min} \sum_{v_k \in P} f(k) + \lambda \mu(P) \quad (5)$$

, where $\mu(P)$ is the compactness penalty function of P (e.g., the sum of distances among all pairs of the vertices in P [22]), and λ is the regularization parameter. However, such methods suffer from the following issues:

1) Definition of the compactness function $\mu(P)$ is subjective.
2) Determination of an appropriate regularizer λ is difficult, as we do not have sufficient training data for this purpose.
3) To solve this objective function is often a NP-hard problem, which makes it unpractical in many real world applications. Sometimes, even the approximate solutions are of high computation complexity, if there are any.

In contrast, our approach proposes a novel group anomaly algorithm in social networks using spectral graph wavelet theory. The graph wavelets focus on the intrinsic geometric structure of the graph by transforming each vertex $v_i \in V$, and mining the topological information of both local and global centered vertices to support multiscale analysis. In addition, the graph wavelet approach identifies anomaly groups which are automatically compact, and provides a fair and low computational method in terms of complexity for identifying abnormal group behavior in a broad application scenarios.

IV. ALGORITHMS

In this section, we first introduce Graph Fourier Transform concept, explain eigenvector and eigenvalue meanings, then based on that define anomaly index of graph. The next is to describe graph wavelet's features such as reconstruction and localization. In section IV-D, we propose the group anomaly detection algorithm via graph wavelet.

A. Graph Fourier Transform

Given a signal f defined on graph $\mathbf{G}$, its Graph Fourier Transform is considered as the projection of f on the complete set of $\{\chi_l\}_{l=0}^{N-1}$, and is written as [14]:

$$\hat{f}(l) = < \chi_l, f > = \sum_{i=1}^{N} \chi_l^*(i) f(i) \quad (6)$$

Since $\{\chi_l\}_{l=0}^{N-1}$ is complete, therefore, f can be recovered by its Graph Fourier Transform coefficients $\hat{f}(l)$ as [14]:

$$f(n) = \sum_{l=0}^{N-1} \hat{f}(l) \chi_l(n) \quad (7)$$

$\hat{f}(l)$ is the coefficient of component χ_l.

1) eigenvector χ_l: As an analog with classical signal processing, eigenvector χ_l is also called frequency of $\mathbf{G}$ by some researchers. In the later part of this paper, χ_l will be called eigenvector or frequency, alternatively. However, unlike the traditional frequency concept in classical signal processing fields, the frequencies of $\mathbf{G}$ is a set of discreet vectors with length of $|V|$. Interestingly, like the classical signal Fourier Transform, Parseval relation still holds; i.e. [28],

$$||\hat{f}||_2^2 = ||f||_2^2 \quad (8)$$

Equation 8 means that energy in vertex domain and frequency domain is equal for any graph signal f. Without loss of generality, we assume $||f||_2 = 1$, if there is no explicit notations.

2) eigenvalue λ_l: According to the definition of eigenvalue λ_l in Equation 2, the following equation holds:

$$\chi_l^T \lambda_l \chi_l = \chi_l^T \mathcal{L} \chi_l = \sum_{e_{mn} \in E} w_{mn} [\chi_l(m) - \chi_l(n)]^2 \quad (9)$$

Since χ_l is normalized, and $||\chi_l||_2 = 1$, then,

$$\chi_l^T \lambda_l \chi_l = \lambda_l = \sum_{e_{mn} \in E} w_{mn} [\chi_l(m) - \chi_l(n)]^2 \quad (10)$$

From equation 10, we can see that λ_l summarizes all the eigenvector deviations on any directly connected vertices v_m

and v_n in $\mathbf{G}$. Since each term in the summation of the right-hand side is non-negative, the eigenvectors associated with smaller eigenvalues are smoother; i.e., the component differences between neighboring vertices are small [28]. As the eigenvalue increases, larger differences in neighboring components of the graph Laplacian eigenvectors is present. Hence, for larger λ_l, its corresponding eigenvector, $\chi_l(n)$, has larger deviation among connected vertices. According to the definition of Laplacian matrix $\mathcal{L}$, it is easy to verify that $\lambda_0 = 0$ since $\mathcal{L} \cdot \vec{\mathbf{1}} = 0 \cdot \vec{\mathbf{1}}$, where $\vec{\mathbf{1}} = \{1, 1, 1, ..., 1\}$, and $\chi_o(n) = \frac{\vec{\mathbf{1}}}{\sqrt{N}}$. Thus, $\chi_o(n) = \frac{\vec{\mathbf{1}}}{\sqrt{N}}$, means $\chi_o(n)$ is constant on each vertex, and there is no deviation among any two vertices in $\chi_0(n)$. For this reason, $\chi_0(n)$ is considered as the least abnormal component of $\mathbf{G}$. Similarly, $\chi_{N-1}(n)$ is considered the most abnormal component of $\mathbf{G}$.

Fig 2 shows an undirected graph $\mathbf{G_1}$, and each edge's weight is 1. Fig 3(a) shows $\mathbf{G_1}$'s six eigenvectors distributions along each vertex. We can see, χ_0 is constant on very vertex, and has the smallest deviations along each edge. χ_5 has the largest deviations, and the difference of χ_5 along each edge is larger than any other eigenvector on average.

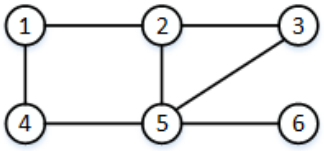


Figure 2: Graph $\mathbf{G_1}$, all edges's weight are 1.

B. Global Anomaly Index

To quantify the anomaly of a vector f defined on a graph $\mathbf{G}$, it's necessary to incorporate the intrinsic structures of $\mathbf{G}$ and f. As discussed above, $\hat{f}(l)$ represents the coefficient of frequency χ_l, and $\hat{f}^2(l)$ is considered as the energy of frequency χ_l. In addition, according to equation 10, λ_l represents the deviation of frequency χ_l along all the connected vertex. Therefore, in this paper, we define the anomaly index of χ_l in f as:

$$\gamma_f(l; \mathbf{G}) = \lambda_l \hat{f}^2(l) = \lambda_l < f, \chi_l >^2 \tag{11}$$

$\gamma_f(l; \mathbf{G})$ depends on two parts, frequency χ_l's deviation sum λ_l, and its energy $\hat{f}^2(l)$. If the energy $\hat{f}^2(l)$ is small, even λ_l is large, the anomaly index of χ_l might be small. Obviously, $\gamma_f(0; \mathbf{G})$ is always 0 since $\lambda_0 = 0$. Further, we use the maximal value of $\gamma_f(l; \mathbf{G})$ to represent the global anomaly of f on $\mathbf{G}$:

$$\gamma_f(\mathbf{G}) = \max_{0 \le l \le N-1} \gamma_f(l; \mathbf{G}). \tag{12}$$

Roughly speaking, $\gamma_f(l; \mathbf{G})$ means the anomaly extension of χ_l in f defined on $\mathbf{G}$, in stead of meaning anomaly extension of vertex v_l. For brevity, $\gamma_f(l; \mathbf{G})$ and $\gamma_f(\mathbf{G})$ are shortened as $\gamma_f(l)$ and γ_f, respectively, when $\mathbf{G}$ is known.

Figure 3(b) plots the anomaly index $\gamma_f(l)$ of f_1 on graph $\mathbf{G_1}$, where $f_1 = [2, 3, 4, 3, 2, 1]$. The six markers on the dashed blue are the six eigenvalues of $\mathbf{G}$. The yellow line is $|\hat{f}(l)|$,

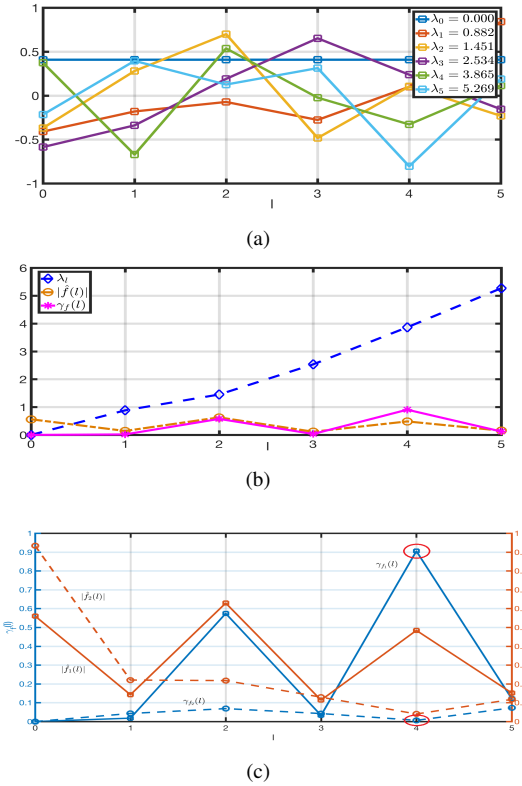
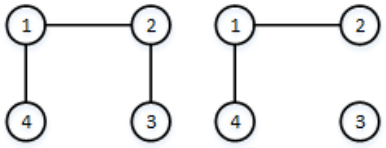


Figure 3: (a): Eigenvector distribution along each vertex in graphs $\mathbf{G_1}$. (b): anomaly index $\gamma_f(l)$ of $f_1 = [2, 3, 4, 3, 2, 1]$ on graph $\mathbf{G_1}$. (c): anomaly index $\gamma_f(l)$ of $f_1 = [2, 3, 4, 3, 2, 1]$ and $f_2 = [2, 2, -3, 4, 3, 1]$ on graph $\mathbf{G_1}$, where $\gamma_{f_1} = 0.905$, and $\gamma_{f_1} = 0.073$, labelled in red oval.

and the pink line is the anomaly index, $\gamma_f(l)$ for frequency χ_l. Because $\gamma_f(l)$ depends on both λ_l and its power $\hat{f}^2(l)$, in the yellow line, even though χ_0 has the strongest power, while its deviation $\lambda_0 = 0$, thus $\gamma_f(0) = 0$. On the other hand, χ_5 has the largest deviation; but its power $|\hat{f}(5)|^2$ is small, which makes $\gamma_f(5)$ is also small. Considering χ_4 has a high deviation (eigenvalue) and a strong power of frequency, thus χ_4 has the largest anomaly index index. To compare the influence of different f on anomaly index, we show an example in Fig 3(c). Set $f_1 = [2, 3, 4, 3, 2, 1]$ and $f_2 = [2, 2, -3, 4, 3, 1]$, we plot their anomaly index γ_f and energy $|\hat{f}(l)|$ respectively. The light blue curves stands for anomaly index, and yellow stands for $|\hat{f}(l)|$. The solid line stands for f_1, and dashed line stands for f_2. As we can see, for high frequency χ_l, f_1 has a larger power than f_2, and hence a higher anomaly index than f_2, where $\gamma_{f_1} = 0.905$ and $\gamma_{f_2} = 0.073$. This is consistent with that f_1 has larger deviations than f_2.

As we discussed before, anomaly index depends on graph structure and f. Shown in 3(c), different f might have very different anomaly index because the power of χ_l distribution is different. Similarly, even same signal f on two different graphs might have very different anomaly indices. Fig 4 shows two graphs with the same $f = [1, 2, 5, 2]$. Fig 5 illustrates the



(a) $\mathbf{G_2}$ (b) $\mathbf{G_3}$

Figure 4: $f = [1, 2, 5, 2]$ on two graphs $\mathbf{G_2}$ and $\mathbf{G_3}$.

anomaly index of f on $\mathbf{G_2}$ and $\mathbf{G_3}$, where $\gamma_f(\mathbf{G_2}) = 0.073$ and $\gamma_f(\mathbf{G_3}) = 0.235$. (This is because in $\mathbf{G_3}$ there is not edge connecting v_2 and v_3, the difference between $f(2)$ and $f(3)$ is not considered as anomaly.)

Remarks: In this subsection, we introduce the anomaly index $\gamma_f(l; \mathbf{G})$ to measure the anomaly of χ_l in f defined on $\mathbf{G}$ by combing the spectrum structure of $\mathbf{G}$ and f. $\gamma_f(l; \mathbf{G})$ depends on two parts: (1) the eigenvalue which reflects the deviations of χ_l; (2) the $|\hat{f}(l)|^2$ which represents the power of χ_l in f. $\gamma_f(l; \mathbf{G})$ reflects the anomaly index of χ_l and not about the vertex v_l. We use the maximal value of $\gamma_f(l; \mathbf{G})$ to define the anomaly index of f, and it denotes the global anomaly index of f on $\mathbf{G}$.

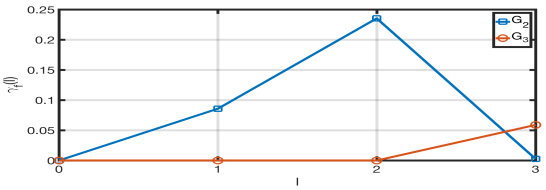


Figure 5: Anomaly index of $\mathbf{G_2}$ and $\mathbf{G_3}$.

C. Graph Wavelets

Classic wavelet is called mathematical microscope because of its capability of showing signal anomaly with different scales. In the case of complex networks, graph wavelets render the graph with good localization properties both in frequency and vertex (i.e. spatial) domains. Their scaling property allows us to zoom in/out of the underlying structure of the graph.

Recall from Equation 6, the anomaly pattern $\hat{f}(l)$ presents the anomaly components of f from the whole graph prospective. However, information concerning the vertex-location can not be identified from the Fourier transform. To address this issue, Hammond et al. [14] proposed constructing wavelet transforms functions over the vertices using weighted graphs, described in the following steps:

1) Define a continuous generating kernel functions $g(x)$ on $\mathbb{R}^+$;
2) Then, select a central vertex $a \in V$ and scale s, set the frequency coefficients as $g(s\lambda_l)\chi_l^*(a)$ for each frequency component χ_l;
3) Finally, sum up all those frequency components χ_l.

In this way, the graph wavelet at central vertex a is constructed as:

$$\psi_{s,a}(n) = \sum_{l=0}^{N-1} g(s\lambda_l)\chi_l^*(a)\chi_l(n) \tag{13}$$

After setting up the graph wavelet, the wavelet coefficients for f can be defined as

$$W_f(s, a) =< \psi_{s,a}, f >= \sum_{l=0}^{N-1} g(s\lambda_l)\hat{f}(a)\chi_l(n) \tag{14}$$

Similar to classical wavelets, graph wavelets provide following three properties, which are presented in detail in [14].

1) **Reconstruction.** When the kernel function $g(x)$ satisfies the admissibility condition and $g(0) = 0$, $f(n)$ can be reconstructed by the wavelet coefficients.
2) **Discretization and Wavelet Frames.** For practical applications, scale s of graph wavelet $\psi_{s,a}$ should be sampled with a finite number of scales. Given a real valued function $h(x)$, satisfying

$$\hat{h}(\omega) = \sqrt{\int_{\omega}^{\infty} \frac{|\hat{g}(\omega')|^2}{\omega'} d\omega'} \tag{15}$$

, where $\hat{g}$ and $\hat{h}$ are the classical Fourier transform of $g(x)$ and $h(x)$, the scaling function $\phi_a(n)$ can be generated as:

$$\phi_a(n) = \sum_{l=0}^{N-1} h(\lambda_l)\chi_l^*(a)\chi_l(n) \tag{16}$$

Accordingly, the scaling coefficients are defined as

$$S_f(a) =< \phi_a, f > \tag{17}$$

Using scale set $\Theta := \{s_j\}_{j=1}^{J}$, the discretized graph wavelet set $\{\psi_{s_j,a}\}_{j=1}^{J}{}_{a=0}^{N-1}$, and scaling function set $\{\phi_a\}_{a=0}^{N-1}$ constitute a frame [14]. According to frame theory [10], $f \in \mathbb{R}^N$ can be reconstructed from those $NJ + J$ wavelet and scaling coefficients as

$$f(n) = \sum_{a=v_0}^{v_{N-1}} [\sum_{j=1}^{J} W_f(s_j, a)\psi_{s,a}(n) + S_f(a)\phi_a(n)]. \tag{18}$$

For brevity, we assume that

$$\phi_a(n) = \psi_{s_0,a}(n) \tag{19}$$

, and

$$S_f(a) = W_f(s_0, a). \tag{20}$$

Therefore, equation 18 can be written as

$$f(n) = \sum_{a=v_0}^{v_{N-1}} \sum_{j=0}^{J} W_f(s_j, a)\psi_{s,a}(n). \tag{21}$$

In the later part of this paper, we do not differentiate scaling coefficient and wavelet coefficient, and call them both wavelet coefficient. A detailed algorithm and treatment concerning the choice of Θ can be found in [14].

3) **Localization in vertex domains.** Given a central vertex v_a and its graph wavelet $\psi_{s,a}(n)$, suppose the kernel function g is $K+1$ times continuously differentiable, let v_n be an vertex of $\mathbf{G}$ with $d_G(n,a) > K$, then there exists constants D and β, such that

$$\frac{|\psi_{s,a}(n)|}{||\psi_{s,a}||} \leq D\beta \tag{22}$$

for all $s < \beta$. $d_G(n,a)$ is the shortest path distance, which is the minimum number of edges in any path that connect vertices v_n and v_a [14]. Equation 22 shows for any vertex v_n that is far away from center vertex v_a ($d_G(n,a) > K$), $\frac{|\psi_{s,a}(n)|}{||\psi_{s,a}||}$ is upper bounded by $D\beta$. In other words, for vertex v_n which is far away form vertex v_a, its wavelet value is linearly attenuated by scale s. When the scale s is small, their wavelet value of marginal vertices will be vanished quickly. The marginal vertices are those which satisfy equation 22. All the other vertices are called kernel vertices, denoted by $\mathcal{K}(s,a)$. Obviously, $\forall v_n \in \mathcal{K}(s,a)$, $d_G(n,a) \leq K$. Thus $\mathcal{K}(s,a)$ is automatically compact. Figure 6 shows two graph wavelets centered on the same vertex a, but with two different scales, $\psi_{s_1,a}$ and $\psi_{s_2,a}$, where $s_1 < s_2$. The length of the vertical bar on each vertex denotes its graph wavelet value. The highlighted areas denote the kernel vertices($d_G(n,a) \leq 1$), and the others are marginal vertices. We can see that the wavelet values on marginal vertices in figure 6(a) are smaller than those in figure 6(b). Figure 6(c) is f's distribution along each vertex, and figure 6(d) shows the wavelet coefficients with center node a for different scales, which indicates $W_f(s_2,a)$ has the largest value, and $W_f(s_3,a)$ has the smallest.

D. Group Anomaly Detection via graph wavelet

According to Equation 22, when s is small, the weights of the marginal vertices are severely attenuated. Essentially, $W_f(s,a)$ is equivalent to the sum of f with large weights on kernel vertices, and small weights on marginal vertices. When f is of uniformly large negative/positive value on kernel vertices, then $W_f(s,a)$ will be a large negative/positive value with scale s.

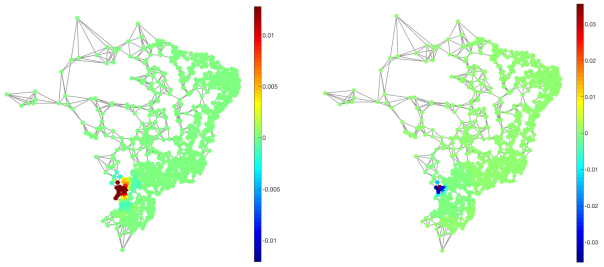
The localization property of graph wavelet makes it appropriate for group anomaly detection since it automatically identifies the kernel vertices from marginal vertices. These kernel vertices form a compact subset since each one of them is close to the same center vertex a, which avoids the compactness constrain condition in equation 5, thus making its computational complexity greatly reduced. We propose our group anomaly detection algorithm based on graph wavelet, illustrated in Algorithm 1. It iterates $NJ+J$ times, and each iteration selects a vertex as the center node, and computes wavelet coefficient $W_f(s_j,a)$ with $J+1$ scales. When $W_f(s_j,a)$ is larger than some pre-set threshold ω_{th}, it considers the corresponding kernel vertices, $\mathcal{K}(a)$, as anomaly burst group. Similarly, when $W_f(s_j,a)$ is smaller than $-\omega_{th}$, it considers $\mathcal{K}(a)$ as anomaly absenteeism group. The computational complexity is $O(J|V|^2)$.

Algorithm 1 Group Anomaly Detection Using Graph Wavelet

1: **Input:** graph and absenteeism score vector $\mathbf{G}(V,E;f^l)$ at time interval l, wavelet threshold ω_{th}.
2: **Output:** anomaly burst group set $\mathcal{I}^{bur}$ and absenteeism group set $\mathcal{I}^{abs}$.
3: compute graph spectral $\sigma(\mathbf{G})$;
4: set graph wavelets $\psi_{s,a}(n)$ and scales set $\{s_j\}_{j=0}^{J}$ for all $a \in V$;
5: **for all** center node $a \in V$ and $s_j \in \{s_j\}_{j=0}^{J}$ **do**
6: compute $W_f(s_j,a)$;
7: **if** $W_f(s_j,a) \geq \omega_{th}$ **then**
8: add group $\mathcal{K}(s_j,a)$ to $\mathcal{I}^{bur}$
9: **end if**
10: **if** $W_f(s_j,a) \leq -1 * \omega_{th}$ **then**
11: add group $\mathcal{K}(s_j,a)$ to $\mathcal{I}^{abs}$
12: **end if**
13: **end for**
14: **return** anomaly burst group $\mathcal{I}^{bur}$ and absenteeism group set $\mathcal{I}^{abs}$.

Remarks:

1) Graph wavelets form a frame, the function f can be reconstructed by their coefficients. As long as the scale level J is high enough, f can be well decomposed into the frame basis. Thus, using graph wavelets to exploit structure of functions defined on graphs is much more reasonable.
2) Graph wavelet transforms selected kernel vertices, $\mathcal{K}(s,a)$, that are close to the central vertex a, and attenuate the impact of other marginal vertices that are far away from a. The abnormal group selected by graph wavelet approach is automatically compact, and circumvent high computational complexity, which makes is easily adaptable to wide variety of application scenarios.
3) Graph wavelet ia able to identify abnormal burst group and absenteeism group simultaneously without extra computation cost.

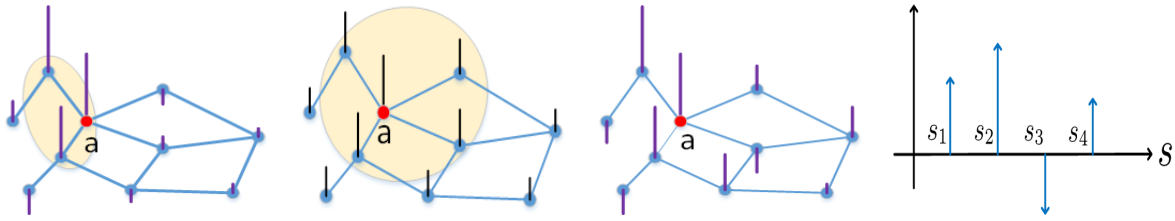


(a) wavelet $\psi_{s_1,a}$ (b) wavelet $\psi_{s_2,a}$

Figure 7: Graph Wavelets with center city v_{83}.

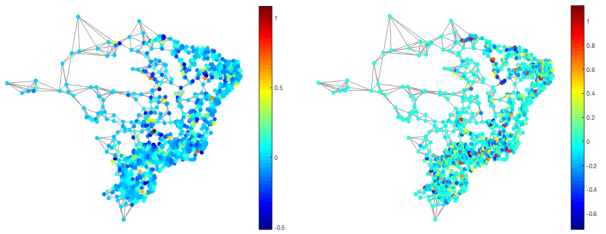
V. EXPERIMENTAL RESULTS

This section discusses the applications of our approach for detecting group anomaly. We begin by briefly describing the dataset we used for our experiments in Section V-A. Then we discuss in Section V-B the implementation details of how we assemble the graph $\mathbf{G}$ and construct the graph wavelets $\psi_{s,a}$.



(a) wavelet $\psi_{s_1,a}$ (b) wavelet $\psi_{s_2,a}$ (c) $f(n)$ vs vertices (d) $W_f(s,a)$ vs scale s

Figure 6: Using graph wavelets for abnormal group identification.



(a) wavelet $W_f(s_1,a)$ (b) wavelet $W_f(s_2,a)$

Figure 8: Graph Wavelet coefficient $W_f(s_1,a)$ and $W_f(s_2,a)$.

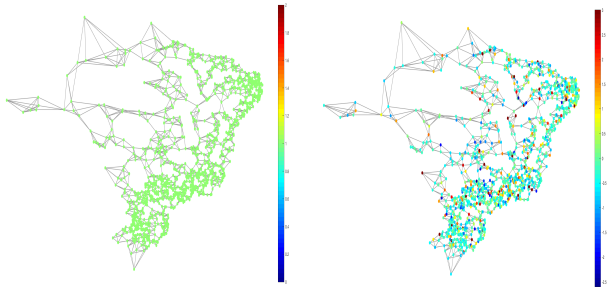
After that, we present group anomaly detection performance in protest events detection. In Section V-D, we describe three case studies that illustrate how the graph wavelet is able to capture absenteeism events like disaster scenarios.

A. Data Collection and Preprocessing

The study described in this paper uses tweets in Latin America that were collected over 2 years (Jan 2012 to June 2014). We query Datasift's streaming API to collect these tweets that also have meta-information including geotag bounding boxes (structured geographical coordinates), Twitter Places (structured data), user profile location (unstructured, unverified strings), and 'mentions information' about locations present in the body of the tweet. Typically, we found the number of tweets with readily available geo-coordinates is too low for conducting meaningful experiments. To circumvent this, we use the geo-enrichment algorithm described in [21]. This algorithm uses a gazetteer-based approach to look-up location names and geo-coordinates. To identify location-specific tweets, we configure the geocoding tool to first consider the tweet's text for mentions of place names and geographical landmarks (e.g., say, Plaza de la Independencia (Quito, Ecuador)). In cases when no geographical location was found in the Tweets text, it then proceeds to process the geographical coordinates and the self-reported location string in user's profile metadata. Using the geocoding tool, we were able to extract tweets corresponding to $598,300$ unique cities from Latin America.

B. Experimental Setup

a) Graph Setup and Z_{score} Time Series: Each city's, v_i's location is represented by its geographical coordinate pair lat_i and lon_i. Instead of using the real physical distance, we define the distance of any two cities v_i and v_j as $d_{ij} = \sqrt{(lat_i - lat_j)^2 + (lon_i - lon_j)^2}$. We setup graph G as a k neighbors graph, which means each city is only connected to its k-nearest-neighbors. In this paper, we set $k = 5$, and all the edges' weights in G are 1. Figure 9(a) shows Brazil' 5 nearest-neighbor Graph with 1276 cities.



(a) (b)

Figure 9: (a) Brazil 5 nearest-neighbor Graph: 1276 cities an all edge weights are 1. (b) Brazil Z_{score} on July 31th, 2013, the color bar shows the scale of Z_{score}.

b) Time Series Z_{score}: Considering the Twitter count X varies vastly among cities, instead of using X itself, we use the normalized value Z_{score}, which is defined as:

$$Z_{score} = \frac{X - \mu}{\sigma} \tag{23}$$

Where, μ is the mean value of the previous 30 days Twitter counts, and σ is the corresponding standard deviation. As shown in Figure 9(b), different node color shows cities' different Z_{score} value.

c) kernel function $g(x)$ and scaling function $h(x)$: Our choice for the wavelet generating kernel function, $g(x)$ and scaling function $h(x)$ is motivated by our goal to achieve scale-dependent localization. We set $g(x)$ and $h(x)$ as:

$$g(x) = \begin{cases} x & \text{for } x < 1 \\ s(x) & \text{for } 1 \le x \le 2 \\ 2x^{-1} & \text{for } x > 2 \end{cases} \tag{24}$$

, where $s(x) = -5 + 11x - 6x^2 + x^3$.

$$h_x = 1.385 * exp(-(\frac{20x}{0.6\lambda_{max}})^4) \tag{25}$$

The scale set $\{s_j\}_{j=1}^{J}$ is selected to be equally logarithmically spaced between the minimum and maximum scales s_1 and s_J, which are defined in [14]. We set $J = 6$ in the experiments.

d) anomaly index $\gamma_f(G)$ and ω_{th}: We claim that the event frequency η is linear to $\gamma_f(\mathbf{G})$, and the linear equation is

$$\eta = k_0 * \gamma_f(\mathbf{G}) + k_1 \tag{26}$$

We use historical data for training of k_0 and k_1 by least square error criterion. Once we know k_0 and k_1, given a new $\gamma'_f(\mathbf{G})$, the events number is estimated as $m = \lceil \eta' \rceil$. After that the threshold ω_{th} is set as the m largest $W_f(s_j, a)$, for all $a \in V$, $0 \leq j \leq J$.

C. Performance

We did experiment on three major protest countries: Brazil, Mexico and Venezuela, from Jan 2013 to Dec 2014. Taken the Gold Standard Report (GSR) as grand truth, we run the graph wavelet approach. For each day, decide whether there is any anomaly, if there is, identify the group of abnormal cities, then compare with GSR, see if the selected cities have protest events on that day, how many of them are matched with grand truth, how many of them do not have protest. We use recall, precision and F-measure to evaluate the performance. To evaluate our graph wavelet approach, we also compare with some intuitive approaches, like frequency based random assign, we call it baseline model, and Z-score based selection methods. The baseline model is built according to the historical protest records of each city, based on frequency, predict the future protests occurrence. The Z-score approach is, selecting the group of cities whose Z-score across some threshold, say $|Zscore| > 3$.

We compare three models performance over two years period test, and show the overall results on Table I. Generally speaking, graph wavelet has better precision, recall, and F-measure scores than baseline across countries. The mean F-measure for graph wavelet detection across models and countries is greater than that for other predictions. Interestingly, we find that the graph wavelet work at different efficiency levels depending on each countries. Compared with Figure 10, Figure 11 and Figure 12, we found graph wavelet model has a much higher recall in Venezuela than in Brazil, while has a inferior quality of event detection in Mexico than in Venezuela.

Table I: The performance of graph wavelet vs. baseline and Zscore.

Country	Method	Precision	Recall	F-measure
Brazil	Baseline	0.052	0.104	0.060
	Zscore	0.117	0.307	0.159
	Graph wavelet	0.404	0.262	0.292
Mexico	Baseline	0.074	0.124	0.090
	Zscore	0.221	0.147	0.168
	Graph wavelet	0.397	0.384	0.408
Venezuela	Baseline	0.078	0.053	0.059
	Zscore	0.197	0.197	0.189
	Graph wavelet	0.292	0.554	0.355

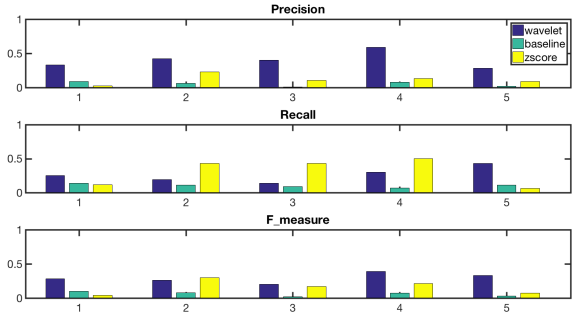


Figure 10: Brazil protest detection performance.

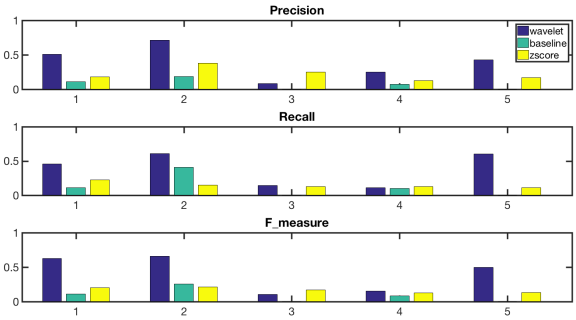


Figure 11: Mexico protest detection performance.

D. Case Studies

Case study 1: Iquique Earthquake, Chile. On April 1, 2014 around 8:46 PM (local time) a large earthquake struck off the coast of Chile, northwest of the port city of Iquique. We show the distribution of absenteeism scores and normalized wavelet coefficient values of the graph wavelets from the beginning of this event and over a 24 hour period. As shown in Figure 13(a) we observe an absenteeism behavior, where the scores are dominated by very low (blue spectrum) of Z-score values (indicating high absenteeism). Likewise in Figure 13(b), we witness low coefficients values for the northern regions of Chile, where the impact of the earthquake was most significant. As the news of earthquake spread throughout the next day, user activity on social media increased. This bursty behavior is seen on April 2nd, around 11:00 AM. We can observe from Figure 13(c) that our z-scores have increased (red spectrum) significantly and the coefficient value distribution (see figure 13(d)) of graph wavelets, for northern regions of Chile is now in red spectrum. From the graph wavelet distributions in Figs. 13(b) 13(d), we can see that the kernel area of the absenteeism/burst wavelets cover most large negative/positive values. In this way, the wavelet identifies the abnormal negative/positive groups in absent/burst time intervals, respectively. Furthermore, a high correlation score of 0.726 was calculated for the wavelets from absenteeism and bursty periods of this episode. As a result, we note that there is a strong connection between the burst in activity and the previously observed absenteeism, signaling an event was detected.

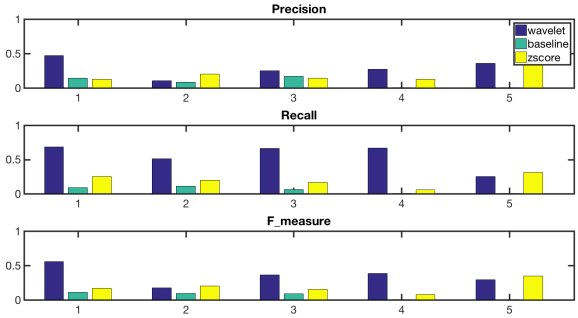
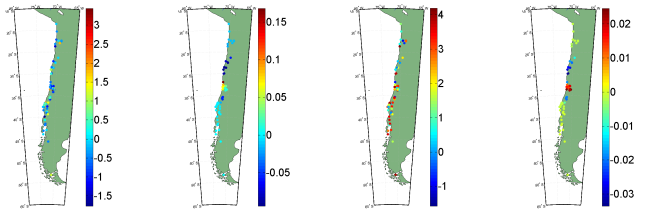


Figure 12: Venezuela protest detection performance.



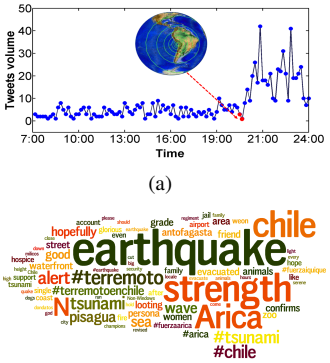
(a) (b) (c) (d)

Figure 13: Iquique Earthquake, Chile. Above plots show differences in distributions of absenteeism score and wavelet coefficients calculated at 8:45 PM, April 1, 2014 (a-b) involving group absenteeism and later when burst in activity is captured at 11:00 AM, April 2, 2014 (c-d), respectively.

From the graph wavelets generated in absenteeism time period 13(b), we found the central node to be the city of 'Iquique'. We study the timeseries (Fig. 14(a)) of Twitter activity for Iquique and word clouds (see Fig.. 14(b)) generated from their content, to see how events unfolded during the course of the earthquake. Strong absenteeism is observed from 8:45 PM to 9:20 PM. We also checked user mobility through geotagged tweets from city of Iquique, on April 1, 2014 and found that the user mobility fraction had increased by 15.4%.

Case Study 2: Massive power outage in Venezuela. A massive power outage in Venezuela plunged several major cities including the capital city, Caracas in to darkness around 7:40 PM (local time) on December 2, 2013. News media reported [2], that the power outage lasted for 10-15 minutes, and the people of Caracas soon went out in the streets to protest. This action at the beginning of the episode coincides with the absenteeism period detected by our algorithm. The scatter plots showing distribution of absenteeism scores and wavelet coefficients (Figs. 15(a), 15(b)) indicate that most of the low values are less than 0. Shortly after the absenteeism, we detected a huge burst in activity around 8:45 PM, signaled by the increased z-scores (low absenteeism) and coefficient values (Figs. 15(c), 15(d)). A correlation score of 0.617 was calculated on comparing the graph wavelets from both

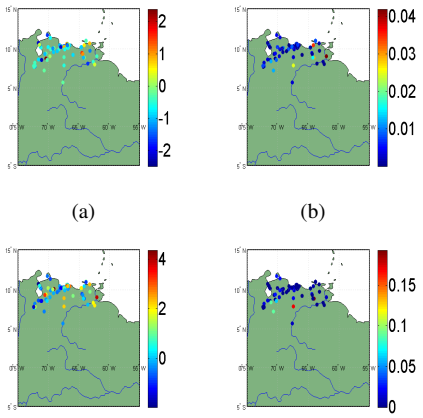
[2]http://www.usatoday.com/story/news/world/2013/12/02/power-failure-caracas-venezuela/3823327/



(a)

(b)

Figure 14: Iquique earthquake, Chile, April 1, 2014. (a) Tweet time series for Iquique on April 1, 2014. (b) Word cloud of tweets which mention 'Iquique'.



(a) (b)

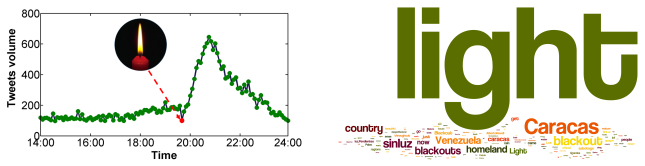
(c) (d)

Figure 15: Power Outage in Venezuela. Above plots show differences in distributions of absenteeism score and wavelet coefficients calculated at 7:40 PM, December 2, 2013 (a-b) involving group absenteeism and later when burst in activity is captured at 8:45 PM in the same day (c-d), respectively.

absentee and burst period.

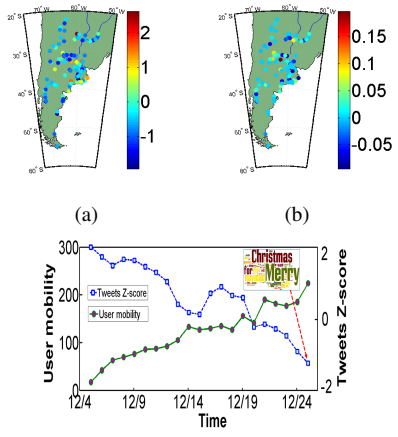
The absenteeism related graph wavelets indicated that the city of Caracas was the central node. Taking a close look at the twitter volume and tweets from Caracas and surrounding cities, we observed a sharp decline in user activity right around 7:40 PM and then a huge spike at starting at 8:45 PM. The word clouds of tweet content show a very similar story. The most dominant words are 'light' and 'blackout', even the Spanish phrase 'sin luz' which means 'no light' became a trending hashtag #sinluz on Twitter.

Case Study 3: Christmas Day. As mentioned earlier, an absenteeism behavior may not always lead to a spike in activity. In this case, our model detected strong absenteeism in social media activity for major holidays such as Christmas day, however it was not followed by a bursty period in Twitter activity. One explanation of this behavior is that people tend



(a) (b)

Figure 16: Caracas, Venezuela power outage, December 2, 2013. (a) Time series of tweets volume. (b) Word cloud of tweets mentioning 'Caracas'.



(a) (b)

(c)

Figure 17: The Christmas Day in Argentina: Above plots shows distributions of (a) absenteeism score and (b) wavelet coefficients calculated on December 25, 2013. (c) Time series comparing absenteeism score and user mobility corresponding to tweets between December 5, 2013 and December 25, 2013.

to travel back to visit family during the holidays. This is supported by low values of z-scores or high absenteeism in Figure 17(a) and wavelet coefficients in Figure 17(b) with respect to Argentina tweets on December 25, 2013. Hence, no subsequent burst period was detected for this event.

Interestingly as the Christmas day approached we observed (see Figure 17(c)) that user mobility gradually increases and z-score decreases signaling greater absenteeism. We used Pearson's correlation coefficient to measure the two time series and found a correlation score of -0.94.

E. Why Absenteeism Group Detection?

Previous research has demonstrated the importance of burst detection in Twitter. In our study, we argue that group absenteeism can also be vital for detecting disruptive societal events. Modeling absenteeism is crucial, because it can serve as a surrogate signal for event detection. For example, in the case of the Iquique earthquake, where our algorithm detected an absenteeism behavior on Twitter followed by a spike in user activity. Unlike traditional event detection methods which identify real time events only after they have occurred i.e., once the burst signal has been identified; the absenteeism signal can be observed much earlier, and it renders a foresight or view into the future events. Our approach thus offers a significant advantage over current strategies that focus solely on modeling spike or burst related patterns for event detection.

VI. Conclusion

Existing approaches for event detection suffer from an inherent latency in their detection process. It is because they use the bursty signals from abnormal activity on social networks, but miss the absenteeism signal that precedes these bursts. Our approach bridges this shortcoming by successfully modeling this *lull-ness*. We have presented a systematic and unified framework for detecting, identifying event's location and distinguishing anomalous groups in Twitter. From the three case studies, we have shown that the initial phase in the evolution of an disruptive, event is characterized by group absenteeism behavior. This behavior is further underlined by an increase in user mobility. As in the case of "Christmas Day" event we observed absenteeism from Argentina Twitter users in days leading to December 25th was characterized by increased mobility (inferred from geolocated tweets). We defined an absenteeism score over the groups of cities that form our Twitter network and used it construct wavelet transforms, that not only to detect the anomalous subgraphs at different scales, but also to find the geographical focal point of the anomaly.

References

[1] C. C. Aggarwal and K. Subbian, "Event detection in social streams." in *Proc. SDM'12*, vol. 12, 2012, pp. 624–635.

[2] L. Akoglu, M. McGlohon, and C. Faloutsos, "Anomaly detection in large graphs," in *In CMU-CS-09-173 Technical Report*. Citeseer, 2009.

[3] L. Akoglu, H. Tong, and D. Koutra, "Graph based anomaly detection and description: a survey," *Data Mining and Knowledge Discovery*, vol. 29, no. 3, pp. 626–688, 2015.

[4] R. B. Bapat, *Graphs and matrices*, 2010.

[5] U. Bügel and A. Zielinski, "Multilingual analysis of twitter news in support of mass emergency events," *International Journal of Information Systems for Crisis Response and Management*, vol. 5, no. 1, pp. 77–85, 2013.

[6] S. Calderara, U. Heinemann, A. Prati, R. Cucchiara, and N. Tishby, "Detecting anomalies in people's trajectories using spectral graph analysis," *Computer Vision and Image Understanding*, vol. 115, no. 8, pp. 1099–1111, 2011.

[7] D. Chen, C.-T. Lu, Y. Kou, and F. Chen, "On detecting spatial outliers," *Geoinformatica*, vol. 12, no. 4, pp. 455–475, 2008.

[8] F. Chen and D. B. Neill, "Non-parametric scan statistics for event detection and forecasting in heterogeneous social media graphs," in *Proc. KDD'14*. ACM, 2014, pp. 1166–1175.

[9] D. I Shuman, B. Ricaud, and P. Vandergheynst, "Vertex-frequency analysis on graphs," *Applied and Computational Harmonic Analysis*, 2013.

[10] I. Daubechies *et al.*, *Ten lectures on wavelets*, 1992, vol. 61.

[11] J. Eisenstein, B. O'Connor, N. A. Smith, and E. P. Xing, "A latent variable model for geographic lexical variation," in *Proc. EMNLP'10*, 2010, pp. 1277–1287.

[12] A. P. Gaudine and A. M. Saks, "Effects of an absenteeism feedback intervention on employee absence behavior†," *Journal of Organizational Behavior*, vol. 22, no. 1, pp. 15–29, 2001.

[13] S. Ghosh-Dastidar and H. Adeli, "Wavelet-clustering-neural network model for freeway incident detection," *Computer-Aided Civil and Infrastructure Engineering*, vol. 18, no. 5, pp. 325–338, 2003.

[14] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 129–150, 2011.

[15] K. Henderson, T. Eliassi-Rad, C. Faloutsos, L. Akoglu, L. Li, K. Maruhashi, B. A. Prakash, and H. Tong, "Metric forensics: a multi-level approach for mining volatile graphs," in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2010, pp. 163–172.

[16] L. Hong, A. Ahmed, S. Gurumurthy, A. J. Smola, and K. Tsioutsiouliklis, "Discovering geographical topics in the twitter stream," in *Proc. WWW'12*, 2012, pp. 769–778.

[17] M. Kulldorff, "A spatial scan statistic," *Communications in Statistics-Theory and methods*, vol. 26, no. 6, pp. 1481–1496, 1997.

[18] T. Lappas, B. Arai, M. Platakis, D. Kotsakos, and D. Gunopulos, "On burstiness-aware search for document sequences," in *Proc. KDD'09*, 2009, pp. 477–486.

[19] T. Lappas, M. R. Vieira, D. Gunopulos, and V. J. Tsotras, "On the spatiotemporal burstiness of terms," *Proceedings of the VLDB Endowment*, vol. 5, no. 9, pp. 836–847, 2012.

[20] C. C. Noble and D. J. Cook, "Graph-based anomaly detection," in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2003, pp. 631–636.

[21] N. Ramakrishnan, P. Butler, S. Muthiah, N. Self, R. Khandpur, P. Saraf, W. Wang, J. Cadena, A. Vullikanti, G. Korkmaz *et al.*, "'beating the news' with embers: forecasting civil unrest using open source indicators," in *Proc. KDD'14*, 2014, pp. 1799–1808.

[22] P. Rozenshtein, A. Anagnostopoulos, A. Gionis, and N. Tatti, "Event detection in activity networks," in *Proc. KDD'14*, 2014, pp. 1176–1185.

[23] R. Rustamov and L. Guibas, "Wavelets on graphs via deep learning," in *Proc. NIPS'13*, 2013, pp. 998–1006.

[24] T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake shakes twitter users: real-time event detection by social sensors," in *Proc. WWW'10*, 2010, pp. 851–860.

[25] H. Sayyadi, M. Hurst, and A. Maykov, "Event detection and tracking in social streams." in *ICWSM'09*, 2009.

[26] B. C. Seamonds, "Stress factors and their effect on absenteeism in a corporate employee group." *Journal of Occupational and Environmental Medicine*, vol. 24, no. 5, pp. 393–hyhen, 1982.

[27] G. Sheikholeslami, S. Chatterjee, and A. Zhang, "Wavecluster: a wavelet-based clustering approach for spatial data in very large databases," *The VLDB Journal*, vol. 8, no. 3-4, pp. 289–304, 2000.

[28] D. I. Shuman, B. Ricaud, and P. Vandergheynst, "Vertex-frequency analysis on graphs," *Applied and Computational Harmonic Analysis*, 2015.

[29] J. Sun, H. Qu, D. Chakrabarti, and C. Faloutsos, "Neighborhood formation and anomaly detection in bipartite graphs," in *Data Mining, Fifth IEEE International Conference on*. IEEE, 2005, pp. 8–pp.

[30] N. Tremblay and P. Borgnat, "Graph wavelets for multiscale community mining," 2014.

[31] K. Watanabe, M. Ochi, M. Okabe, and R. Onai, "Jasmine: a real-time local-event detection system based on geolocation information propagated to microblogs," in *Proc. GIS'08*, 2011, pp. 2541–2544.

[32] J. Weng and B.-S. Lee, "Event detection in twitter." *ICWSM'11*, pp. 401–408, 2011.

[33] L. Xiong, B. Póczos, J. G. Schneider, A. Connolly, and J. VanderPlas, "Hierarchical probabilistic models for group anomaly detection," in *Proc. AISTATS'11*, 2011, pp. 789–797.

[34] Z. Yin, L. Cao, J. Han, C. Zhai, and T. Huang, "Geographical topic discovery and comparison," in *Proc. WWW'11*, 2011, pp. 247–256.

[35] R. Yu, X. He, and Y. Liu, "Glad: group anomaly detection in social media analysis," in *Proc. KDD'14*. ACM, 2014, pp. 372–381.

[36] L. Zhao, F. Chen, J. Dai, T. Hua, C.-T. Lu, and N. Ramakrishnan, "Unsupervised spatial event detection in targeted domains with applications to civil unrest modeling," *PloS one*, vol. 9, no. 10, p. e110206, 2014.