Group Anomaly Detection based on Spectral Graph Wavelets

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. The new 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 facilitates multiscale analysis, enabling us to detect anomalous behavior at different graph resolutions, which in turn allows the identification of event location and anomalous groups within the network. We introduce a systematic group anomaly detection method that utilizes graph wavelets to simultaneously detect absenteeism and burst groups. The effectiveness of our approach is then verified using Twitter activity related to protests events in Latin America.

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, as of March 31, 2016 Twitter had more than 310 million average monthly active users (78% of whom were using mobile devices) and were anticipating this to continue to grow by as much as 25% per year[1]. Various studies have shown that Twitter is a viable social "sensor", and thus 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 lead to unusually low levels of activity in social networks. An example of this is shown in Figure 1, where 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 in the area for the two hours from 6:00 PM - 8:00 PM that day. Another example comes from December 24, 2013, when 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
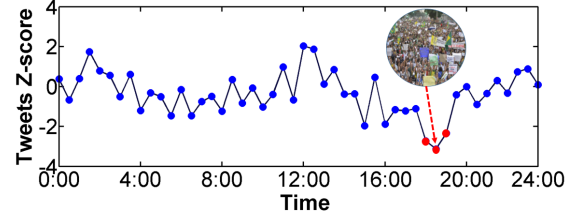

Figure 1: Detected group absenteeism in Natal, Brazil beginning at 6:00 PM on June 17, 2013. This absenteeism event coincides with a large protest that happened in the region.

floods, Twitter activity in this region dropped by 51%, reaching its lowest point later that evening.

Developing a better appreciation of 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, 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 precede a major burst in activity as people react to the event). 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 should group anomaly in graphs be defined? How can both burst and absenteeism scenarios be reconciled?
- At what scale should we model the abnormal groups? Which node should be the central point?
- What is the most efficient way 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 that explain the observed absenteeism, not all absenteeism occurrences will be associated with underlying events so we must be able to differentiate between absenteeism and noisy signals for successful event detection.

A graph wavelet approach offers several outstanding advantages when studying the above questions, including 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 use of a graph anomaly index that depends on graph structure in conjunction with an absenteeism score vector in order to define whether a graph is abnormal. When a graph is deemed to be exhibiting abnormal behavior, we can calculate its wavelet coefficient to identify the central node and its coverage area. This approach will enable us to select abnormal groups at different scales. The group anomaly detection methods are varied and proved to be effective in detecting events such as protest marches and natural disasters.

Our contributions are thus:

- To the best of our knowledge this is the first study to utilize group absenteeism as a basis for event detection. By studying different types of group anomalies, either bursts or absenteeism, we demonstrate that these anomalies are indicative and can be used to detect events such as civil protests or natural disasters.
- We incorporate graph wavelets as a mechanism to detect the most anomalous subgraphs at different scales. We demonstrate the power of this approach for social media analytics.
- We define a graph anomaly index that can be used to determine whether a graph is abnormal. We then apply this new term to locate the central node and identify the abnormal groups.

The rest of this 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 extensive experiments testing our new approach's effectiveness for real-world 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 methods [2]. WWhen considering group concepts, two directions have been studied [3], namely anomalies in unlabeled/plain graphs [20] and attributed graphs. In plain graph anomaly detection, since the only information provided is its structure, features such as distances and communities [29] have been employed to define graph anomalies. In one interesting study [15], additional metrics such as vertices, edges, degree, weight, and connected components are incorporated into the detection framework. In attributed graphs, features regarding node behaviors make it possible to create a richer graphical representation, which is usually connected with one or more real-world applications. Other studies, for example [35] define groups based on the term of the role, and model normal groups that follow the same pattern with respect to their role mixture rates. Alternatively, some researchers working on group anomaly detection have built generative models of group anomalies [33] where the goal is to automatically infer the groups and thus detect group anomalies in a social network. Typically these mixed models suffer from high computational complexity, however, due to the size of the data and are as a result heavily parameterized. In this study, we consider both the graph structure and node features in our proposed graph wavelet based approach for group anomaly detection, as this guarantees that the detected group is 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 recent years. Traditional approaches focus on capturing the 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 within specific geographic regions [11], [16], [34]; applying clustering-based approaches to 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 developed, based on Kulldroff's spatial scan statistic[17], to detect spatial outliers [7] and these have been applied to a wide variety of domains including transportation networks, civil unrest forecasting [36], and heterogeneous social media graphs [8]. The approach we have chosen to address the event detection problem is conceptually different from all the above mentioned studies as it is based on the use of a graph-theoretic framework to detect group anomalies and correlate these to predict future events. Although group absence behavior has been widely studied in the area of organizational behavioral studies [12], [26], it remains largely unexplored in the area of social network analysis. Bearing a close resemblance to group anomaly detection in complex networks, our new detection approach is further distinguished by its focus on groups rather than individuals.

c) Graph Wavelet: One of the key challenges facing this research is the need to adapt a detection procedure to encompass both missing and bursty activity groups. To address this issue, we incorporate spectral graph wavelets [14] into our algorithm. This strategy has previously been found to be quite effective for multiscale community mining [30]. Wavelet methods based on spectral graph theory have been applied to 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 the local information encoded in the graph structure and use this to cluster and identify nodes that are similar to one another in a scale-dependent fashion.

III. PROBLEM SETTING

In this section, we begin by introducing the mathematical notations used in this paper and then go on to formalize our approach to group anomaly detection. After presenting the problem statement, we provide a brief comparison of our approach with conventional solutions, and review the challenging issues that are relevant to this type of 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:

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

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 $0 = \lambda_0 < \lambda_1 \le \lambda_2 \le \ldots \le \lambda_{N-1} = \lambda_{max}$. The complete set of $\mathcal{L}$'s normalized eigenvectors [4] χ_i for $i = 0, 1, 2, ..., N-1$ is described as:

$$\mathcal{L}\chi_i = \lambda_i \chi_i \tag{2}$$

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

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

$\sigma(\mathbf{G})$ 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*

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

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 minimize the corresponding function [22]. Therefore, Equation 4 can be modified as:

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

, 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 [22], 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 for social networks that is based on 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 a multiscale analysis. In addition, the graph wavelet approach identifies anomaly groups that are automatically compact, and provides a fair method at a low computational cost in terms of complexity for identifying abnormal group behavior in broad application scenarios.

IV. ALGORITHMS

In this section, we first introduce Graph Fourier Transform concept, explain eigenvectors and eigenvalue meanings, and then apply them based on our newly defined anomaly index for graphs. We then go on to describe graph wavelet features such as reconstruction and localization. In section IV-D, we propose a new group anomaly detection algorithm based on graph wavelets.

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]:

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

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]:

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

$\hat{f}(l)$ 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],

$$||\hat{f}||_2^2 = ||f||_2^2 \tag{8}$$

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 as the most abnormal component of $\mathbf{G}$.

Figure 2 shows an undirected graph $\mathbf{G_1}$, and each edge's weight is 1. Figure 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.
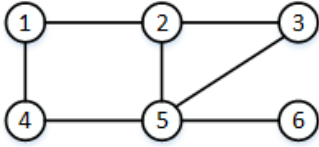

Figure 2: Graph $\mathbf{G_1}$, all edges' weight are 1.

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 vertices. Therefore, in this paper, we define the anomaly index of χ_l in f as:

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

$\gamma_f(l; \mathbf{G})$ 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
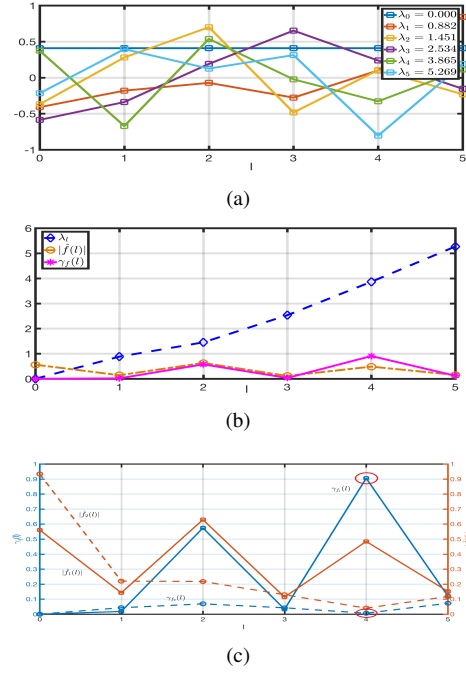

Figure 3: (a): Eigenvector distribution along each vertex in graph $\mathbf{G_1}$. (b): anomaly index $\gamma_f(l)$ of $f_1 = [2, 3, 4, 3, 2, 1]$ on graph $\mathbf{G_1}$. (c): anomaly index $\gamma_f(l)$ of $f_1 = [2, 3, 4, 3, 2, 1]$ and $f_2 = [2, 2, -3, 4, 3, 1]$ on graph $\mathbf{G_1}$, where $\gamma_{f_1}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

maximal value of $\gamma_f(l; \mathbf{G})$ to represent the global anomaly of f on $\mathbf{G}$:

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

Roughly speaking, $\gamma_f(l; \mathbf{G})$ means the anomaly extension of χ_l in f defined on $\mathbf{G}$, instead 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 $|\hat{f}(l)|$, 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 $\hat{f}^2(l)$, for the yellow line, even though χ_0 has the strongest power, 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. To compare the influence of different f on anomaly index, we show an example in Figure 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 curve 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
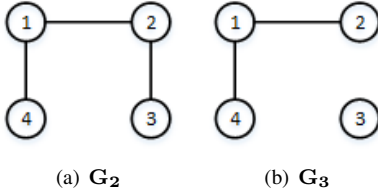
(a) $\mathbf{G_2}$ (b) $\mathbf{G_3}$

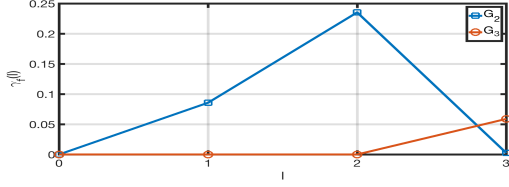Figure 4: $f = [1, 2, 5, 2]$ on two graphs $\mathbf{G_2}$ and $\mathbf{G_3}$.

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

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, the anomaly index depends on graph structure and f. As shown in Figure 3(c), different f might have very different anomaly index because the power of χ_l distribution is different. Similarly, for the same signal f on two different graphs, it might have very different anomaly indices. Figure 4 shows two graphs with the same $f = [1, 2, 5, 2]$. Figure 5 illustrates the 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. We use the maximal value of $\gamma_f(l; \mathbf{G})$ to define the anomaly index of f, which denotes the global anomaly index of f on $\mathbf{G}$.

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 $\hat{f}(l)$ represents 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:

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

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

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

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

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

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:

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

Accordingly, the scaling coefficients are defined as

$$S_f(a) = <\phi_a, f> \quad (17)$$

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

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

For brevity, we assume that

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

and

$$S_f(a) = W_f(s_0, a). \quad (20)$$

Therefore, equation 18 can be written as

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

(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
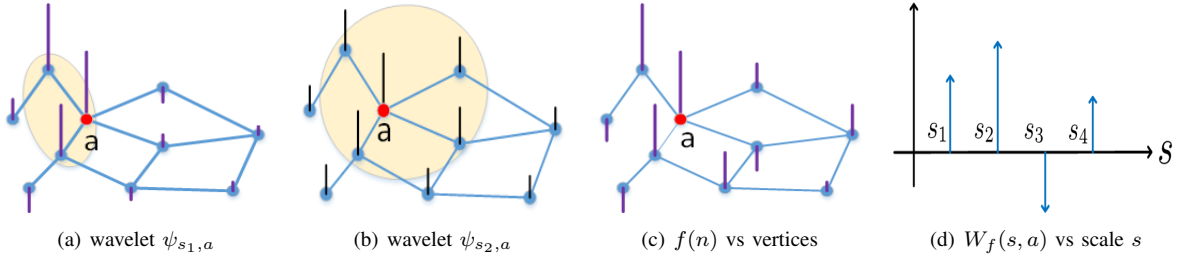
Figure 6: Graph wavelet scale and graph wavelet coefficient.

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 8 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 values 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 abnormal burst group. Similarly, when $W_f(s_j, a)$ is smaller than $-\omega_{th}$, it considers $\mathcal{K}(a)$ as abnormal 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:** abnormal burst group set $\mathcal{I}^{bur}$ and absenteeism group set $\mathcal{I}^{abs}$.
3: compute graph spectrum $\sigma(\mathbf{G})$;
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** abnormal 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 a wide variety of application scenarios.

3) Graph wavelet is able to identify abnormal burst group and absenteeism group simultaneously without extra computation cost.

V. EXPERIMENTAL RESULTS

This section discusses the application of our approach for detecting group anomalies. We begin by briefly describing the dataset used for our experiments in Section V-A and then move on to discuss the implementation details of how the graph **G** is assembled and construct the graph wavelets $\psi_{s,a}$ in Section V-B. The following section presents the group anomaly detection performance for identifying protest events. In Section V-D, we describe three case studies that illustrate how the graph wavelet model is able to capture absenteeism events such as 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: Each city 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 7(a) shows Brazil' 5 nearest-neighbor Graph with 5321 cities.
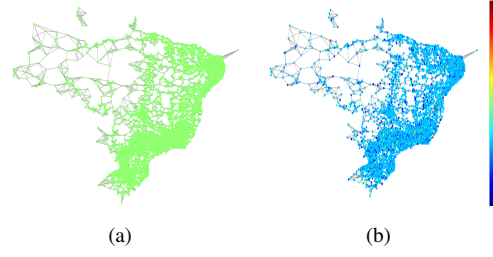

(a)   (b)

Figure 7: (a) Brazil's 5-nearest-neighbor Graph: 5321 cities, all edges' weights are 1. (b) Brazil's Z-score distribution on July 31, 2013. The color bar shows the scale of Z-score.
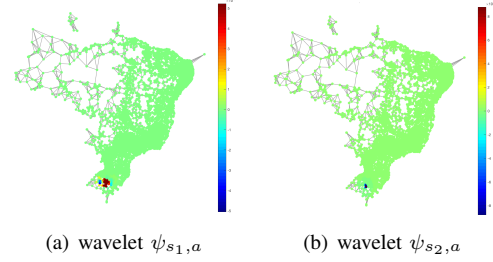

(a) wavelet $\psi_{s_1,a}$ (b) wavelet $\psi_{s_2,a}$

Figure 8: Graph wavelets with center city v_{83}. $s_1 = 1.31$, $s_2 = 0.68$.

b) Absenteeism Score: Considering the Tweets volume X varies vastly among cities, instead of using X itself, we use the normalized value Z-score as absenteeism score, which is defined as:

$$Z - score = \frac{X - \mu}{\sigma} \qquad (23)$$

Where, μ is the mean value of the previous 30 days Tweets volume, and σ is the corresponding standard deviation. As shown in Figure 7(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 follow the kernel function setting in [14], which behaves as a monic power near the origin, and has power law decay for large x. $g(x)$ and $h(x)$ are set as:

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

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

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

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 experiment. Figure 8 shows two different scaled wavelets on Brazil's 5-nearest-neighbor graph. Comparing Figure 9(a) with Figure 9(b), we can see that, when scale increases, more
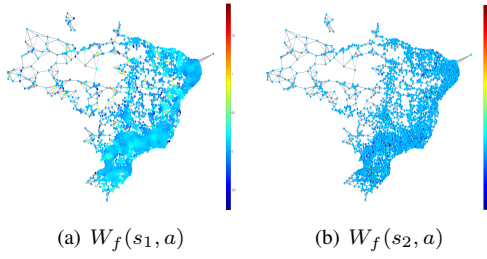
(a) $W_f(s_1, a)$ (b) $W_f(s_2, a)$

Figure 9: Graph wavelet coefficient $W_f(s_1, a)$ and $W_f(s_2, a)$.

Table I: The performance of graph wavelet vs. baseline and Z-score.

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

cities (with deeper color) are selected. Figure 9 shows the corresponding wavelet coefficients. We also try another kernel function, i.e. Mexico hat function, and find that as long as the kernel function monotonicity is the same, the difference of the wavelet coefficient is negligible.

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

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

We use historical data to train 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_{th} largest $W_f(s_j, a)$, for all $a \in V$, $0 \le j \le J$.

C. Performance

The data for this experiment was gathered for three countries experiencing major protest events, namely Brazil, Mexico and Venezuela, from Jan 2013 to Dec 2014. Taking the Gold Standard Report (GSR) [21] as representing ground truth, we applied our new graph wavelet approach as follows. For each day, decide whether there is any anomaly. If there is, identify the group of abnormal cities, then compare this with GSR to determine if the selected cities actually experienced protest events on that day and thus show how many of the model's predictions matched the ground truth and how many did not. We use recall, precision and F-measure to evaluate the model's performance. To evaluate the effectiveness of our new graph wavelet approach, we also compared the results with those obtained using intuitive approaches such as frequency based random assignment, referred to here as the baseline model, and Z-score based selection methods. The baseline model was built according to the historical protest records for each city

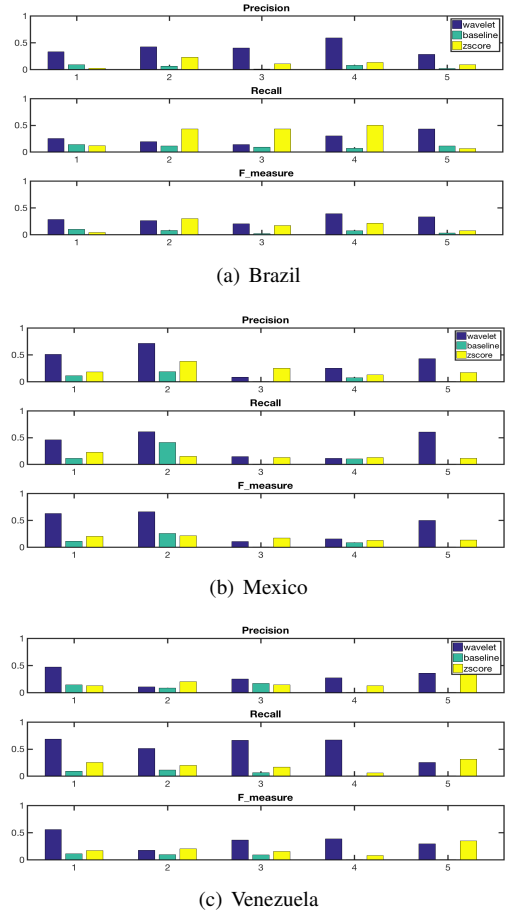(a) Brazil

(b) Mexico

(c) Venezuela

Figure 10: Brazil, Mexico and Venezuela protest detection performance.

and thus the model's predictions of the future occurrence of protests were based on frequency. The Z-score approach is, selecting the group of cities whose Z-score across some threshold, say $|Z - score| > 3$.

We compared the performance of these three models over the two year test period; the overall results are shown in Table I. Generally speaking, the new graph wavelet approach exhibited better precision, recall, and F-measure scores than the baseline model across all three countries. The mean F-measure for the graph wavelet detection across models and countries is greater than that achieved by either of the other prediction models. Interestingly, the graph wavelet approach appears to operate at different efficiency levels for each country. From Figure 10 we can see, the graph wavelet model has a much higher recall in Venezuela than in Brazil, and an inferior quality of event detection in Mexico compared to Venezuela.

D. Case Studies

Case study 1: Iquique Earthquake, Chile. On April 1, 2014 at 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
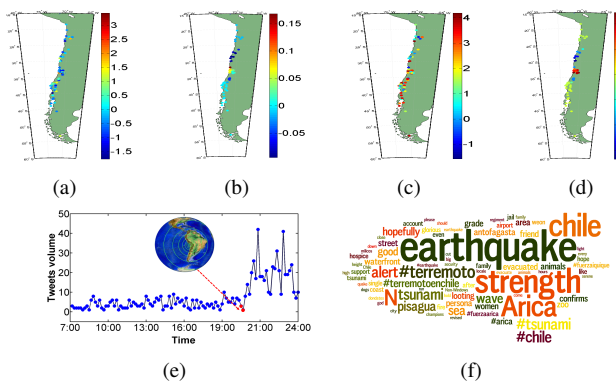
Figure 11: Iquique Earthquake, Chile. (a-d) plots show differences in distributions of absenteeism score and wavelet coefficients calculated at 8:45 PM, April 1, 2014 (a-b) involving group absenteeism and later when burst in activity is captured at 11:00 AM, April 2, 2014 (c-d), respectively; (e) Tweet time series for Iquique on April 1, 2014; (f) Word cloud of Tweets which mention 'Iquique'.
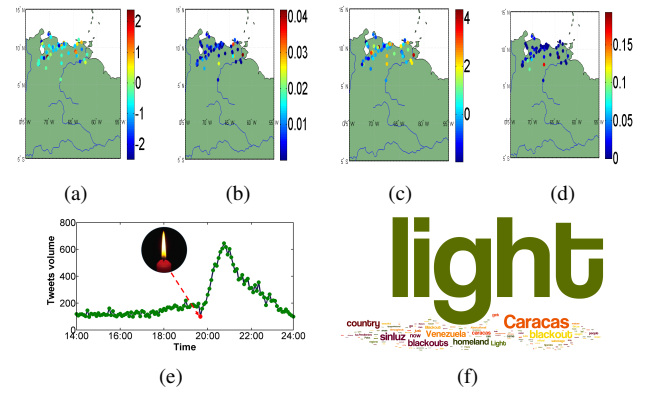

Figure 12: Power Outage in Venezuela. (a-d) plots show differences in distributions of absenteeism score and wavelet coefficients calculated at 7:40 PM, December 2, 2013 (a-b) involving group absenteeism and later when burst in activity is captured at 8:45 PM in the same day (c-d), respectively; (e) Time series of Tweets volume on December 2, 2013; (f) Word cloud of Tweets mentioning 'Caracas'.

normalized wavelet coefficient values of the graph wavelets from the beginning of this event and throughout the subsequent 24 hour period. As shown in Figure 11(a), we can clearly see absenteeism behavior, where the scores are dominated by very low (blue spectrum) Z-score values (indicating high absenteeism). Likewise, Figure 11(b) shows low coefficient 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 clearly visible on April 2nd, at around 11:00 AM. Figure 11(c) shows that the Z-scores increase (red spectrum) significantly and the coefficient value distribution (Figure 11(d)) of the graph wavelets for northern regions of Chile are also in the red spectrum. The graph wavelet distributions in Figures 11(b) 11(d) show that the kernel area of the absenteeism/burst wavelets cover most large negative/positive values. In this way, the wavelets identify 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, indicating a strong connection between the burst in activity and the previously observed absenteeism, signaling an event was detected.

The graph wavelets generated during the absenteeism time period Figure 11(b) have a central node located in the city of 'Iquique'. Looking at the time series (Figure 11(e)) of Twitter activity for Iquique and the associated word clouds (see Figure 11(f)), we can see how events unfolded during the course of the earthquake. Strong absenteeism is observed from 8:45 PM to 9:20 PM. Examining user mobility via their geotagged Tweets from the city of Iquique, on April 1, 2014, 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, into 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 quickly took to 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 the distribution of absenteeism scores and wavelet coefficients (Figures 12(a), 12(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 (Figures 12(c), 12(d)). A correlation score of 0.617 was calculated when comparing the graph wavelets from the absentee and burst periods.

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

Case Study 3: Christmas Day. As noted earlier, absenteeism behavior may not always lead to a spike in activity. For example, our model detected strong absenteeism in social media activity for major holidays such as Christmas Day that was not followed by a bursty period in Twitter activity. This is likely because people tend to travel to visit family during the holidays. This is supported by low values of z-scores or high absenteeism in Figure 13(a) and wavelet coefficients in Figure 13(b) with respect to Argentinian tweets on December

---

[2]http://www.usatoday.com/story/news/world/2013/12/02/power-failure-caracas-venezuela/3823327/
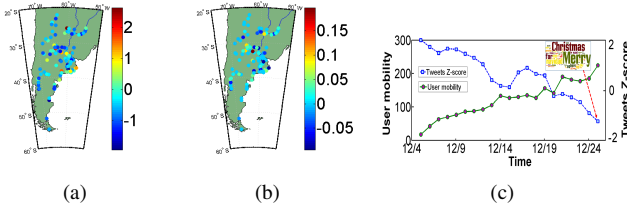
Figure 13: The Christmas Day in Argentina: (a-b) plots show distributions of (a) absenteeism score and (b) wavelet coefficients calculated on December 25, 2013; (c) Time series comparing absenteeism score and user mobility corresponding to Tweets between December 5 - 25, 2013.

25, 2013. Hence, no subsequent burst period was detected for this event. Interestingly as Christmas Day approached. Figure 13(c) shows that user mobility gradually increases and the 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 this 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, our new algorithm detected absenteeism behavior on Twitter that was closely followed by a spike in user activity. Unlike traditional event detection methods, which identify real time events only after they have occurred because the burst signal must first be identified, an absenteeism signal can be observed much earlier, thus providing greater foresight into future events. This means that our proposed approach 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. This is because they are based on the use of bursty signals from abnormal activity on social networks, but miss the absenteeism signal that often precedes these bursts. Our approach addresses this shortcoming by successfully modeling this "lull before the storm". In this study we defined an absenteeism score for groups of cities within the Twitter network and apply it to construct wavelet transforms that not detect anomalous subgraphs (including both burst and absenteeism groups) at different scales, but can also be used to identify the geographical focal point of the anomaly. This localization property of graph wavelets guarantees that the selected groups are compact automatically. The identified abnormal groups have been verified using real-world datasets and proved to be indicative for detecting events such as civil protests or natural disasters.

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 *Proc. KDD'10*. 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 *Proc. KDD'03*. 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," *Signal Processing, IEEE Transactions on*, vol. 62, no. 20, pp. 5227–5239, 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.