Absenteeism Detection in Social Media

Fang Jin[*] Feng Chen[†] Rupinder Paul Khandpur[*]
Chang-Tien Lu[*] Naren Ramakrishnan[*]

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 involve abnormal troughs or lack of activity, leading to absenteeism. We present the first study, to our knowledge, that models absenteeism and uses detected absenteeism instances as a basis for event detection in location-based social networks such as Twitter. The proposed framework addresses the challenges of (i) early detection of absenteeism, (ii) identifying the locus of the absenteeism, and (iii) identifying groups or communities underlying the absenteeism. Our approach uses the formalism of graph wavelets to represent the spatiotemporal structure of user activity in a location-based social network. This formalism facilitates multiscale analysis, enabling us to detect anomalous behavior at different graph resolutions, which in turn allows the identification of event locations and underlying groups. The effectiveness of our approach is evaluated using Twitter activity related to civil unrest events in Latin America.

1 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 September 30, 2016 Twitter had more than 317 million average monthly active users (78% of whom were using mobile devices)[1]. Various studies have shown that Twitter is a viable 'social sensor', and thus holds great promise for detecting and forecasting significant societal events [16]. In recent years, a significant body of research [1, 10, 11, 16, 17, 23, 24] has focused on modeling bursts and increases of user activity on 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 phenomenon 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 to 8:00 PM that day.

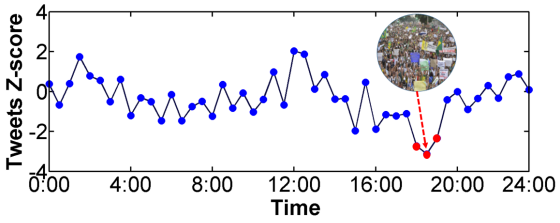


Figure 1: Detected group absenteeism in Natal, Brazil beginning at 6:00 PM on June 17, 2013. This absenteeism event coincides with a large protest that happened in the region.

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 absenteeism as a key phenomenon of interest and develop novel group anomaly detection algorithms for this purpose. An absenteeism event in a social network 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 anomalies in location-based social networks, and has the added advantage of accommodating both absenteeism and bursts. To appropriately incorporate absenteeism concepts into our detection approach, we must first address the following questions:

- How can we define/adapt anomaly detection algorithms to capture not just bursty situations but also those that involve absenteeism?
- At what scale should we model the absenteeism activity and how can we isolate the locus of interest?
- What is the most efficient way to select abnormal

[*]Virginia Tech.
[†]SUNY Albany.
[1]http://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/

groups that are spatially and temporally localized?

- How do we model an absenteeism signal for event detection?

A graph wavelet approach offers several outstanding advantages to study the above questions, including scalability, localization, low computational complexity, and compactness in defining groups. In this formalism, 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 the 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. Such group anomaly detection methods are varied and proven to be effective in detecting events such as protest marches.

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, both bursts and absenteeism, we demonstrate that these anomalies are indicative of key disruptive events such as protests.
- 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 the graph wavelet to locate the central node and identify the abnormal groups.

2 Related work

Group Anomaly Detection. Anomaly detection in graphs has been well studied using outlier detection methods [2]. When considering group concepts, two directions have been explored [3], namely anomalies in unlabeled/plain graphs [12] and those in attributed graphs. In plain graph anomaly detection, since the only information provided is its structure, features such as distances and communities [21] have been employed to define graph anomalies. In one interesting study [9], 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 [25] define groups based on the nature of the role, and model normal groups that follow the same pattern with respect to their role mixture rates.

Event Detection. Traditional approaches focus on capturing the spatiotemporal burstiness of keywords [11]; Kalman filtering to track the geographical trajectories of hot spots of tweets related to earthquakes [16]; detecting topics of interest that are coherent within specific geographic regions [6, 10, 24]; 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, 17, 23]; and using the notion of compactness of a graph [14] to detect events.

Graph Wavelets. One of the key challenges facing our 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 [8] into our algorithm. This strategy has previously been found to be quite effective for multiscale community mining [22]. Wavelet methods based on spectral graph theory have been applied to a wide array of data mining tasks such as community detection, anomaly detection [4], and other machine learning problems [5, 7, 15, 18, 20].

3 PROBLEM SETTING

3.1 Notation We are 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 and 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, $f : V \to \mathbb{R}^N$ operates on 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:

$$a_{ij} = \begin{cases} w_{ij} & \text{when } e_{ij} \in E \\ 0 & \text{otherwise} \end{cases} \tag{3.1}$$

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 χ_i for $i = 0, 1, 2, ..., N-1$ is described as:

$$\mathcal{L}\chi_i = \lambda_i \chi_i \tag{3.2}$$

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

$$\sigma(\mathbf{G}) := \{(\lambda_l, \chi_l)\}_{l=0}^{N-1}. \tag{3.3}$$

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

3.2 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* $\mathbf{G}(V, E; f^t)$*, where* f^t *represents absenteeism score vector at time interval* t*, select a subset* $P \subseteq V$*, such that*

$$P = \underset{P \subseteq V, P \text{ is compact}}{\arg\min} \sum_{v_k \in P} f(k) \tag{3.4}$$

Defining compactness of the selected subset P is, of course, the key issue here. A general solution to this problem involves 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 [14]. Therefore, Equation 3.4 can be modified as:

$$P = \underset{P \subseteq V}{\arg\min} \sum_{v_k \in P} f(k) + \lambda \mu(P), \tag{3.5}$$

where $\mu(P)$ is the compactness penalty function of P (e.g., the sum of distances among all pairs of the vertices in P [14]), 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 [14], which makes it impractical 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.

4 ALGORITHMS

4.1 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 [8]:

$$\hat{f}(l) = < \chi_l, f > = \sum_{i=1}^{N} \chi_l^*(i) f(i) \tag{4.6}$$

Since $\{\chi_l\}_{l=0}^{N-1}$ is complete, f can be recovered by its graph Fourier transform coefficients $\hat{f}(l)$ as [8]:

$$f(n) = \sum_{l=0}^{N-1} \hat{f}(l) \chi_l(n) \tag{4.7}$$

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

4.1.1 Eigenvector χ_l. As an analog with classical signal processing, the eigenvector χ_l is also referred to as the frequency of $\mathbf{G}$ by some researchers. In the latter part of this paper, χ_l will be referred to as the eigenvector or frequency, alternatively. However, unlike the traditional frequency concept in classical signal processing fields, the frequency of $\mathbf{G}$ is a set of discreet vectors with length of $|V|$. Interestingly, like the classical signal Fourier transform, the Parseval relation [19] still holds, i.e.,

$$||\hat{f}||_2^2 = ||f||_2^2 \tag{4.8}$$

Equation 4.8 means that the energy in the vertex domain and frequency domain is equal for any graph signal f. Without loss of generality, we assume $||f||_2 = 1$.

4.1.2 Eigenvalue λ_l. According to the definition of eigenvalue λ_l in Equation 3.2, the following equation holds:

$$\chi_l^T \lambda_l \chi_l = \chi_l^T \mathcal{L} \chi_l = \sum_{e_{mn} \in E} w_{mn} [\chi_l(m) - \chi_l(n)]^2 \tag{4.9}$$

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

$$\chi_l^T \lambda_l \chi_l = \lambda_l = \sum_{e_{mn} \in E} w_{mn} [\chi_l(m) - \chi_l(n)]^2 \tag{4.10}$$

From equation 4.10, we can see that λ_l summarizes all the eigenvector deviations on any directly connected

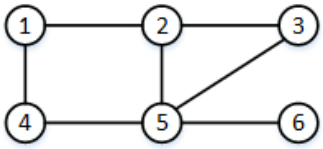


Figure 2: Example graph $\mathbf{G_1}$ where all edges' weights are 1.

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 [19]. As the eigenvalue increases, larger differences in neighboring components of the graph Laplacian eigenvectors are 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 that $\chi_o(n)$ is constant on each vertex, and that 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 $\mathbf{G}$.

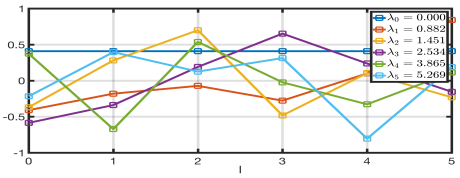
Figure 2 shows an undirected graph $\mathbf{G_1}$ where each edge's weight is 1. Figure 3(a) shows $\mathbf{G_1}$'s six eigenvectors distributions along each vertex. We can see that χ_0 is constant on very vertex, and has the smallest deviations along each edge. χ_5 has the largest deviations, and the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

4.2 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 4.10, 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:

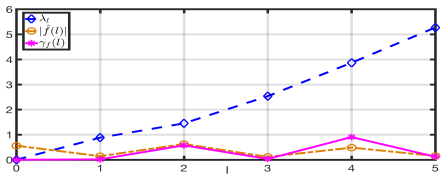
$$\gamma_f(l; \mathbf{G}) = \lambda_l \hat{f}^2(l) = \lambda_l < f, \chi_l >^2 \tag{4.11}$$

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

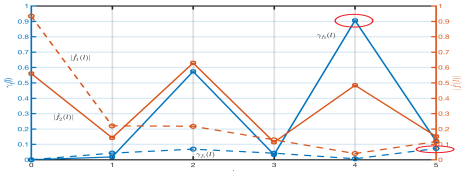
$$\gamma_f(\mathbf{G}) = \max_{0 \le l \le N-1} \gamma_f(l; \mathbf{G}). \tag{4.12}$$



(a)



(b)

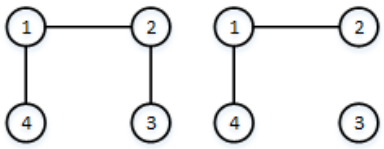


(c)

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

Here, $\gamma_f(l; \mathbf{G})$ refers to the anomaly extension of χ_l in f defined on $\mathbf{G}$, instead of implying the 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_2 on graph $\mathbf{G_1}$, where $f_2 = [2, 2, -3, 4, 3, 1]$. 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, with $\gamma_f(l)$ 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 that χ_4 has a high deviation (eigenvalue) and a strong power of frequency, it has the largest anomaly index. To compare the influence of different f on anomaly index, we show an example in Figure 3(c). Setting $f_1 = [2, 3, 4, 3, 2, 1]$ and $f_2 = [2, 2, -3, 4, 3, 1]$, we plot their anomaly index γ_f and energy $|\hat{f}(l)|$ respectively. The blue curves stand for anomaly indices and the orange curves stand for $|\hat{f}(l)|$. The dashed line stands for f_1, and the solid line stands for f_2. As we can see, for high frequency χ_l, f_2 has a larger power than f_1, and hence a higher anomaly



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

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

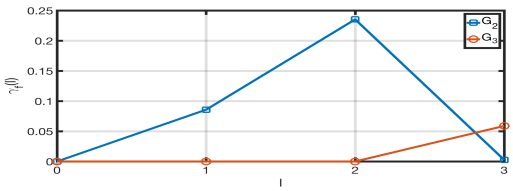


Figure 5: Anomaly indices of $\mathbf{G_2}$ and $\mathbf{G_3}$.

index than f_1, where $\gamma_{f_1} = 0.073$ and $\gamma_{f_2} = 0.905$. This is consistent with that f_2 has larger deviations than f_1.

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 $\mathbf{G_3}$ because there is no edge connecting v_2 and v_3, the difference between $f(2)$ and $f(3)$ is not considered as an anomaly.)

Remarks: In this subsection, we have introduced 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 $\mathbf{G}$.

4.3 Graph Wavelets Classic wavelet formalisms have been referred to as mathematical microscopes because of their capability to depict signal anomalies at 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 that, from Equation 4.6, the anomaly pattern $\hat{f}(l)$ represents the anomaly components of f from the whole graph perspective. However, information concerning the vertex-location cannot be identified from the Fourier transform. To address this issue, Hammond et al. [8] 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:

$$\psi_{s,a}(n) = \sum_{l=0}^{N-1} g(s\lambda_l)\chi_l^*(a)\chi_l(n) \tag{4.13}$$

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

$$W_f(s, a) =< \psi_{s,a}, f >= \sum_{l=0}^{N-1} g(s\lambda_l)\hat{f}(a)\chi_l(n) \tag{4.14}$$

Similar to classical wavelets, graph wavelets obey following three properties, which are presented in detail in [8].

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, the 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

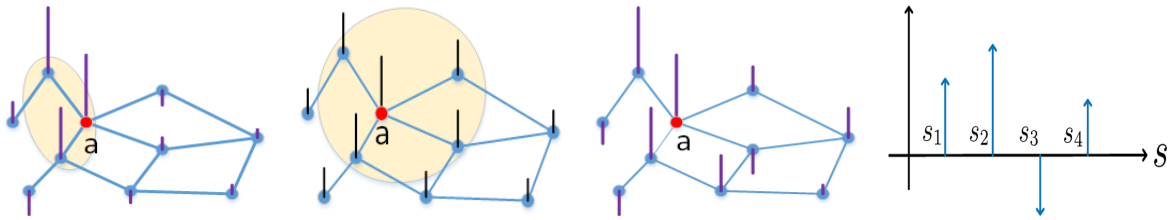
$$\hat{h}(\omega) = \sqrt{\int_{\omega}^{\infty} \frac{|\hat{g}(\omega')|^2}{\omega'} d\omega'}, \tag{4.15}$$

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:

$$\phi_a(n) = \sum_{l=0}^{N-1} h(\lambda_l)\chi_l^*(a)\chi_l(n) \tag{4.16}$$

Accordingly, the scaling coefficients are defined as

$$S_f(a) =< \phi_a, f > \tag{4.17}$$



(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.

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 [8]. $f \in \mathbb{R}^N$ can be reconstructed from those $NJ + J$ wavelet and scaling coefficients as

(4.18)
$$f(n) = \sum_{a=v_0}^{v_{N-1}} [\sum_{j=1}^{J} W_f(s_j, a)\psi_{s,a}(n) + S_f(a)\phi_a(n)].$$

For brevity, we assume that

$$\phi_a(n) = \psi_{s_0,a}(n), \tag{4.19}$$

$$S_f(a) = W_f(s_0, a). \tag{4.20}$$

Therefore, equation 4.18 can be written as

$$f(n) = \sum_{a=v_0}^{v_{N-1}} \sum_{j=0}^{J} W_f(s_j, a)\psi_{s,a}(n). \tag{4.21}$$

In the later part of this paper, we do not differentiate between scaling coefficient and wavelet coefficient. A detailed algorithm and treatment concerning the choice of Θ can be found in [8].

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 exist constants D and β, such that

$$\frac{|\psi_{s,a}(n)|}{||\psi_{s,a}||} \leq D\beta \tag{4.22}$$

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 [8]. Equation 4.22 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 4.22. 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 that $W_f(s_2, a)$ has the largest value, and $W_f(s_3, a)$ with the smallest.

Algorithm 1 Group Anomaly Detection using Graph Wavelets

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 $\mathcal{I}^{abs}$.

4.4 Group Anomaly Detection via Graph Wavelets According to Equation 4.22, 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 wavelets makes them appropriate for group anomaly detection since they automatically identify 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 constraint condition in Equation 3.5, thus reducing its computational complexity greatly. We propose our group anomaly detection algorithm based on graph wavelets in Algorithm 1. It iterates $NJ + J$ times, where each iteration selects a vertex as the center node, and computes the 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 an abnormal burst group. Similarly, when $W_f(s_j, a)$ is smaller than $-\omega_{th}$, it considers $\mathcal{K}(a)$ as an abnormal absenteeism group. Using the fast algorithm for computing graph wavelets [8], the total computational complexity in Algorithm 1 would be $O(J|V|^2)$.

Remarks:

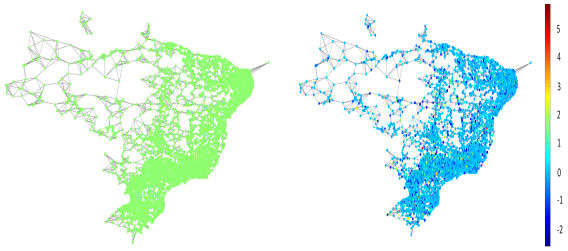
1. Graph wavelets form a frame where the function f can be reconstructed by their coefficients. As long as the scale level J is high enough, f can be well decomposed into the frame basis. Thus, using graph wavelets to exploit the structure of functions defined on graphs is much more reasonable.

2. Graph wavelets transform 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 wavelets are able to identify abnormal burst groups and absenteeism groups simultaneously without extra computation cost.

5 EXPERIMENTAL RESULTS

5.1 Data Collection and Preprocessing The study described in this paper uses tweets geolocated to Latin America and collected over a period of two years (Jan 2013 to Dec 2014). We query Datasift's streaming API to collect tweets that also have meta-information including geographical coordinates, Twitter places, user profile location, and 'mentions information' about locations present in the body of the tweet. In cases when no geographical location was found in the tweet text, we proceed to process the geographical coordinates and the self-reported location string in user's profile metadata.



(a) (b)

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

5.2 Experimental Setup

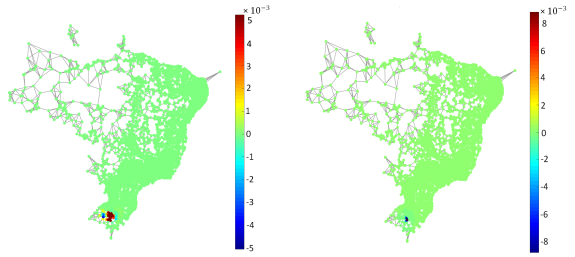
Graph Setup. Each city v_i's location is represented by its geographical coordinate pair lat_i and lon_i. Instead of using the real physical distance, we define the distance of any two cities v_i and v_j as $d_{ij} = \sqrt{(lat_i - lat_j)^2 + (lon_i - lon_j)^2}$. We setup graph $\mathbf{G}$ as a k neighbors 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 $\mathbf{G}$ are 1. Figure 7(a) shows Brazil's 5 nearest-neighbor graph with 5321 cities.

Absenteeism Score. Considering that the tweet volume X varies vastly among cities, instead of using X itself, we use the normalized value of z-score as absenteeism score, which is defined as:

$$\text{z-score} = \frac{X - \mu}{\sigma} \tag{5.23}$$

where μ is the mean value of the previous 30 day tweets volume and σ is the corresponding standard deviation. As shown in Figure 7(b), different node colors denote different z-score values.

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 [8], which behaves as a monic power near the origin, and has power law decay



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

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

for large x. $g(x)$ and $h(x)$ are set as:

$$g(x) = \begin{cases} x & \text{for } x < 1 \\ s(x) & \text{for } 1 \leq x \leq 2 \\ 2x^{-1} & \text{for } x > 2 \end{cases} \tag{5.24}$$

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

$$h_x = 1.385\, exp(-(\frac{20x}{0.6\lambda_{max}})^4) \tag{5.25}$$

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 [8]. We set $J = 6$ in the experiment. Figure 8 shows two different scaled wavelets on Brazil's 5-nearest-neighbor graph. Comparing Figure 8(a) with Figure 8(b), we can see that, when scale increases, more cities (with deeper color) are selected. We also try another kernel function, i.e. the Mexican hat function, and find that as long as the kernel function monotonicity is the same, the differences in wavelet coefficients are negligible.

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

$$\eta = k_0 * \gamma_f(\mathbf{G}) + k_1 \tag{5.26}$$

We use historical data to train k_0 and k_1 by a least squares approach. Once we know k_0 and k_1, given a new $\gamma'_f(\mathbf{G})$, the event number is estimated as $m = \lceil \eta' \rceil$. Subsequently the threshold ω_{th} is set as the m_{th} largest $W_f(s_j, a)$, for all $a \in V$, $0 \leq j \leq J$.

5.3 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) [13] as representing ground truth, we applied our new graph wavelet approach as follows. For each day, we determine whether there are any anomalies detected. If there are, for each anomaly, we identify the group of anomalous cities and compare this set with the 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. (Note that there are many causes of absenteeism besides civil unrest but here we are attempting to determine if our detected events can serve as a signal for such protest events.) We use recall, precision, and the 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 and thus the model's predictions of the future occurrence of protests were based on frequency. The z-score approach entails selecting the group of cities whose z-score crosses the threshold with |z-score| > 3.

Table 1: The performance of graph wavelet vs. a baseline method and a vanilla Z-score approach.

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

We compared the performance of these three models over the two year test period; the overall results are shown in Table 1. 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. The false positives from this study are likely to be useful themselves as they could be indicators of other types of events, e.g., natural disasters, holidays, blackouts, and other situations.

6 Discussion

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. 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. Our approach addresses this shortcoming by successfully modeling the 'lull before the storm'. 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. In the future, we will investigate the impact of k when setting up the k-nearest-neighbors graph, and the scale level J as well. We also plan to extend the absenteeism detection approach to other social media platforms.

Acknowledgment

Supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior National Business Center (DoI/NBC) contract number D12PC000337, the US Government is authorized to reproduce and distribute reprints of this work for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/NBC, or the US Government.

References

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

[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*, 29(3):626–688, 2015.

[4] S. Calderara, U. Heinemann, A. Prati, R. Cucchiara, and N. Tishby. Detecting anomalies in peoples trajectories using spectral graph analysis. *Computer Vision and Image Understanding*, 115(8):1099–1111, 2011.

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

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

[7] S. Ghosh-Dastidar and H. Adeli. Wavelet-clustering-neural network model for freeway incident detection. *Computer-Aided Civil and Infrastructure Engineering*, 18(5):325–338, 2003.

[8] D. K. Hammond, P. Vandergheynst, and R. Gribonval. Wavelets on graphs via spectral graph theory. *Applied and Computational Harmonic Analysis*, 30(2):129–150, 2011.

[9] 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*, pages 163–172. ACM, 2010.

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

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

[12] C. C. Noble and D. J. Cook. Graph-based anomaly detection. In *Proc. KDD'03*, pages 631–636. ACM, 2003.

[13] N. Ramakrishnan, P. Butler, et al. 'beating the news' with embers: forecasting civil unrest using open source indicators. In *Proc. KDD'14*, pages 1799–1808, 2014.

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

[15] R. Rustamov and L. Guibas. Wavelets on graphs via deep learning. In *Proc. NIPS'13*, pages 998–1006, 2013.

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

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

[18] G. Sheikholeslami, S. Chatterjee, and A. Zhang. Wavecluster: a wavelet-based clustering approach for spatial data in very large databases. *The VLDB Journal*, 8(3-4):289–304, 2000.

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

[20] A. Silva, X.-H. Dang, P. Basu, A. K. Singh, and A. Swami. Graph wavelets via sparse cuts. In *Proc. KDD'16*. ACM, 2016.

[21] 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*, pages 8–pp. IEEE, 2005.

[22] N. Tremblay and P. Borgnat. Graph wavelets for multiscale community mining. *Signal Processing, IEEE Transactions on*, 62(20):5227–5239, 2014.

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

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

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