

Mass Movements and their Adoption in Social Networks

Fang Jin

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Computer Science and Applications

Naren Ramakrishnan, Chair
Chang-Tien Lu
Chris North
Yang Cao
Feng Chen

May 30, 2016
Arlington, Virginia

Keywords: Information Propagation, Absenteeism, Event Detection, Social Networks
Copyright 2016, Fang Jin

Mass Movements and their Adoption in Social Networks

Fang Jin

(ABSTRACT)

Online social networks have become a staging ground for modern movements, with the Arab Spring being the most prominent example. In an effort to understand and predict those movements, social media is regarded as valuable social sensor to disclose the underlying behavior and pattern. To fully understand the mass movement information propagation pattern in social networks, several problems need to be considered and addressed. Specifically, modeling mass movements that incorporate (i) multiple spaces (ii) dynamic network structure (iii) swift outbreak/slowly evolving transmission (iv) misinformation would be highly propitious in understanding information propagation in social medias.

This dissertation explores four research problems underlying mass movement adoption in social media. First, how do mass movements get mobilized on Twitter, especially in a specific geographic area. Second, how do we detect protest activity in social networks by observing group abnormality in graph? Third, how can we infer the causality of a specific type of protest, say climate related protest? Fourth, how do we distinguish real movements from rumors or misinformation campaigns?

A fundamental objective of this research has been to comprehensively study the mass movement adoption in social networks, it may cross multiple spaces, it may evolve with dynamic network structures, it can be swift outbreaks or long term slowly evolving transmissions, what is more, it may mixed with misinformation campaigns. Each of those issues requires the development of new mathematical models and algorithmic approaches which are explored here. It is my hope that this work will facilitate advancements in information propagation, group abnormality detection and misinformation distinction, and ultimately helps improve the understanding of mass movement and their adoptions in social networks.

Acknowledgments

I would like to acknowledge and express my gratitude to several individuals for their contributions and support through my years as a doctoral student.

My advisor, Naren Ramakrishnan, for giving me the independence to pursue my research interests. I am especially grateful for his advice and wisdom over the past four years.

My doctoral committee - Chang-Tien Lu, Yang Cao, Chris North, Feng Chen - for their valuable comments, guidance and encouragement.

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Methods	2
1.2.1	Geometric Brownian Motion	3
1.2.2	Graph Wavelets	3
1.2.3	Epidemiological Models	4
1.3	Goals of the Dissertation	4
1.4	Organization of the Dissertation	6
2	Modeling Mass Protest Adoption in Social Network Communities using Geometric Brownian Motion	7
2.1	Introduction	7
2.2	Related Work	9
2.3	Formalisms	10
2.3.1	Basics	10
2.3.2	Trust functions and GBM	11
2.3.3	GBM propagation	12
2.3.4	GBM parameter estimation	13
2.4	Bispace Propagation Model	14
2.4.1	GBM with Communities	15
2.4.2	Propagation in Latent Space	17
2.5	Experiments	18

2.5.1	Dataset description	18
2.5.2	GBM Diffusion Model	19
2.5.3	GBM Diffusion with Communities	20
2.5.4	Latent Space Diffusion Model	21
2.6	Evaluation results	22
2.7	Discussion	24
3	Protest Detection from Group Abnormality	25
3.1	Introduction	25
3.2	Related Work	27
3.3	Preliminaries	28
3.3.1	Notations	28
3.3.2	Graph Anomaly	29
3.3.3	Generalized Graph Anomaly	30
3.3.4	Problem Statement	30
3.4	Two-pass Event Detection Model	31
3.4.1	Graph Wavelets	32
3.4.2	Group Anomaly Detection	33
3.4.3	Group Absenteeism Based Event Detection	35
3.5	Experimental Results	36
3.5.1	Data Collection and Preprocessing	37
3.5.2	Experimental Setup	37
3.5.3	Case Studies	38
3.5.4	Why Absenteeism Group Detection?	40
3.6	Conclusion	41
4	Causality Inference to Climate Related Protest	42
4.1	Introduction	42
4.2	Related research	44

4.3	Climate protest classifier	44
4.3.1	Majority assign	44
4.3.2	K Nearest Neighbor	45
4.3.3	Naive Bayes	45
4.3.4	Weighted Support Vector Machine	46
4.3.5	Logistic regression classifier	48
4.3.6	Evaluation	48
4.4	Climate Motivated Protests	49
4.4.1	Frequency Analysis by Country	50
4.4.2	Spatial Distribution of Climate Motivated Protests	51
4.4.3	Temporal Dependency on Climate Events	53
4.5	Climate protests causality	56
4.5.1	Word cloud	56
4.5.2	Analysis of Protest Descriptions	56
4.5.3	Pathways to Climate Motivated Protest	57
4.6	Climate protest pattern	59
4.7	Climate protests in Twitter	60
4.8	Discussion	61
5	Distinguish Real Movement from Rumor on Twitter	62
5.1	Introduction	62
5.2	Related work	63
5.3	Datasets	65
5.3.1	News topics	65
5.3.2	Rumors	66
5.3.3	Preliminary Analysis	67
5.4	Our Approach	68
5.4.1	SIS	68
5.4.2	SEIZ	70

5.4.3	Practical Issues	72
5.4.4	Parameter Identification	74
5.5	Experimental Results	75
5.5.1	Fitting Results	75
5.5.2	Boston Marathon Bombing Analysis	76
5.5.3	Rumor Detection	78
5.6	Conclusion	80
6	Misinformation Propagation on Twitter in the age of Ebola	81
6.1	Introduction	81
6.2	Rumors on Twitter	82
6.3	Epidemiological modeling of rumors	85
7	Concluding Remarks	90

List of Figures

2.1	Mexico teacher protest events from Sep 1 to Sep 7, 2013. The blue pins denote protest cities; the numbers in red denote the sequence of protests as they spread across the country.	8
2.2	An example mentions network. Nodes denote Twitter users, directed edges denote direction of mentions between users, and edges are labeled with mention frequency.	10
2.3	Trust function. A threshold defines the transition between the pre-trust and trusted period.	11
2.4	Bispace propagation model. In the latent space, users infections are explained by a Poisson model, and the red nodes denote the infected users from one time step to another. In the mentions network space, users are infected according to the GBM model. Here, the purple nodes (a, b, c, d, e) denote user infections explained by the GBM model.	14
2.5	Major communities of teacher protest events (Sep 1 to Sep 12, 2013, Mexico).	15
2.6	Key graph properties of communities underlying the Mexican teacher protest events.	16
2.7	Poisson distribution in latent space propagation. (a) shows the raw data outside of the mentions network of teacher protest events on Sep 3, 2013. (b) shows the probability distribution of the number of infections.	16
2.8	Brownian distance vs propagation time for teacher protest events.	19
2.9	GBM and Poisson propagation simulation for Yosoy protests (Mexico) on May 19, 2012.	19
2.10	GBM and Poisson propagation simulation for teacher protests (Mexico) on Sep 1, 2013.	20
2.11	GBM and Poisson propagation simulation for Colombia protests on Dec 4, 2012. .	20

2.12	Normalized mass-protest propagation speed for major communities during teacher protest events.	21
2.13	Total tweets over time from Sep 2 to Sep 11, 2013 (Mexico).	21
2.14	Performance accuracy of the bispace model for the 7 protest scenarios considered here, with and without community structure.	23
2.15	Bispace model simulation results compared against ground truth infected nodes for the Mexican teacher protest events.	24
3.1	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.	26
3.2	Using graph wavelets for abnormal group identification.	34
3.3	Iquique Earthquake, Chile. Above plots show differences in distributions of absenteeism score and wavelet coefficients calculated at 8:45 PM, April 1, 2014 (a-b) involving group absenteeism and later when burst in activity is captured at 11:00 AM, April 2, 2014 (c-d), respectively.	38
3.4	Iquique earthquake, Chile, April 1, 2014. (a) Tweet time series for Iquique on April 1, 2014. (b) Word cloud of tweets which mention ‘Iquique’	38
3.5	Power Outage in Venezuela. Above plots show differences in distributions of absenteeism score and wavelet coefficients calculated at 7:40 PM, December 2, 2013 (a-b) involving group absenteeism and later when burst in activity is captured at 8:45 PM in the same day (c-d), respectively.	39
3.6	Caracas, Venezuela power outage, December 2, 2013. (a) Time series of tweets volume. (b) Word cloud of tweets mentioning ‘Caracas’	40
3.7	The Christmas Day in Argentina: Above plots shows distributions of (a) absenteeism score and (b) wavelet coefficients calculated on December 25, 2013. (c) Time series comparing absenteeism score and user mobility corresponding to tweets between December 5, 2013 and December 25, 2013.	41
4.1	Gold standard report (GSR) format.	43
4.2	Classification methods comparison.	49
4.3	Blue bar shows all the GSR protest events, yellow bar shows climate related protest events, green area shows the climate protest percentage over all the Latin American countries, from July 2012 to March 2015.	50
4.4	(a) Climate related protests events numbers; (b) Climate related protests percentage in Latin American countries, from July 2012 to March 2015. . .	51

4.5	Climate protest events and population (million) of each country. The two series have a Pearson correlation coefficient 0.64.	51
4.6	climate protest and non-climate protest time series from 2011 to March 2015.	52
4.7	Climate and non-climate protests from July 2012 to March, 2015. Red circle represents climate related protest events, and blue circle represents non-climate related protests.	52
4.8	(a) Mexico climate disasters and climate protests. The blue time series shows the climate related protest events, and light red vertical lines show two storm diasters in Mexico, storm Manuel in September 17, 2013 and hurricane Odile in September 15, 2014 respectively. (b)Brazil climate disasters and climate protests. The blue time series shows the climate related protest events, and yellow vertical lines show three drought diasters in Brazil, drought in Feb 2012, Heat wave in Feb 2014, and drought in Oct 2014, respectively. (c) Venezuela climate disasters and climate protests. The blue time series shows the climate related protest events, and rose vertical line show local area flood diaster, and yellow vertical lines drought disasters.	53
4.9	The map shows population density of all Mexico's 32 states. The track shows Tropical Storm Manuel of 2013 and Hurricane Odile of 2014, points in different color represent the wind speeds.	54
4.10	Word cloud of Mexico storm Manuel, Sept 13, 2013.	55
4.11	Word cloud of all the climate related protests, from GSR descriptions.	56
4.12	Climate protest causality diagram. Left bar shows ten countries' climate protest numbers, and right bar shows nine climate event categories which cause climate protests.	57
4.13	Climate protest causality diagram	58
4.14	Climate protest clustering results	59
4.15	Climate protest events in Mexico, Sept 2013 and Brazil, May 2012. Different flag represents different climate disasters. The adjacent world cloud shows Twitter discussion as per that event.	60
5.1	Tweet volume.	66
5.2	Followers/followees distributions. Followers: people who follow the person; Followees: people who are followed by the person.	67
5.3	Retweet cascade for the Amuay Explosion news and Castro rumor. Each node is a user id, and each edge connects the retweet user to the original user. . .	68

5.4	SIS model framework	69
5.5	SEIZ model framework	69
5.6	Numerical implementation work-flow.	70
5.7	Best fit modeling for Boston news.	72
5.8	Best fit modeling for Pope news.	72
5.9	Best fit modeling for Amuay news.	73
5.10	Best fit modeling for Michelle news.	73
5.11	Best fit modeling for Obama news.	74
5.12	Best fit modeling for Doomsday rumor.	74
5.13	Best fit modeling for Castro rumor.	75
5.14	Best fit modeling for Riot news.	76
5.15	SEIZ compartment time-course results.	77
5.16	Ultimate R_{SI} values for eight Twitter datasets.	79
5.17	Dynamic R_{SI} values for Castro rumor and Boston bombing real news.	80
6.1	Word clouds constructed from Ebola-related tweets.	82
6.2	Top 10 Ebola-related rumors (by volume; from 09/28/2014 to 10/18/2014). . .	83
6.3	Distribution of top-10 rumors obtained from geolocated tweets. Data from 10/08/2014 is used for this plot. Icon size is proportional to the logarithm of the tweet volume.	83
6.4	How rumors cluster: (a) 09/29/2014. (b) 10/06/2014. Rumors are color coded consistently across the two frames.	84
6.5	Ebola-related patent rumor propagation over time.	85
6.6	Ebola-related Dallas news propagation over time.	85
6.7	The SEIZ compartmental model. The various states denote: (S) Susceptible. (I) Infected. (E) Exposed. (Z) Skeptic.	86
6.8	Model fits of SEIZ to different rumors: (from left to right) white, zombies, airborne, and patent. (top row) Fitting results. (bottom row) time-course profiles of different compartments.	87
6.9	Ultimate response ratios for 3 news stories (left) and 10 rumors related to Ebola.	88
6.10	Dynamic R_{SI} values for Zombie rumor and Dallas real news.	89

List of Tables

2.1	Mass protests studied in this chapter.	17
2.2	GBM simulation results for teacher protest events on Sep 2, 2013.	22
4.1	Classification methods comparison.	49
5.1	Twitter datasets studied in this chapter.	65
5.2	Parameter definitions in SEIZ model[5]	71
5.3	Fitting error of SIS and SEIZ models	75
5.4	Ratios of SEIZ model for Boston dataset.	78

Chapter 1

Introduction

Social microblogs such as Twitter and Weibo are experiencing explosive growth with billions of users around the globe sharing their daily status updates online. For example, Twitter has more than 255 million average monthly active users (78% from mobile) per month as of March 31, 2014, and an estimated growth of 25% per year. In the technology era, online social networks have become a staging ground for modern movements with the Arab Spring being the most prominent example. Interestingly, the role of social networks is not limited to helping organize the activities of disruptive elements. Many key government and news agencies have also begun to embrace Twitter and other social platforms to disseminate information. Without doubt, the analysis of social media networks has become a crucial and irreplaceable task in understanding the social movements.

Social network analysis is the process of gathering data from stakeholder conversations on digital media, and processing into structured insights. These lead to more information-driven decisions, which include but are not limited to understanding social sentiment, discovering topics, identifying ongoing events, and predicting future trends. Social media as a carrier of information, despite its various forms (Facebook, Twitter, Weibo, etc.) shares some common properties in information propagation, that can be approached using the methods of mathematical modeling and data mining.

1.1 Motivation

This dissertation explores four research problems underlying mass movement adoption in social media. In contrast to popular memes, they constitute modelling protest mobilization, detect graph group abnormality pattern, infer protest causality, and distinguish real movement from rumors. (i) First, how do mass movements get mobilized on Twitter, especially in a specific geographic area, (ii) Second, how do we detect protest activity in social networks by observing group abnormality in graph? (iii) Third, how can we infer the causality

of climate related protests? and (iv) Fourth, how do we distinguish real movements from rumors or misinformation campaigns?

Modeling mobilizations: It is well known that network structure plays a key role in information propagation. Several interesting questions arise in this space. Which node is the key player who exerts influence over others? How do we realistically simulate information propagation process within a network? How do specific memes get adopted in the network? When do they translate into mass movements?

Group abnormality in graph: Group abnormality not only depends on each user's activity, but also closely associates with the graph structure. In recent year, a significant body of research on group abnormality has been focused on two aspects: (1) modeling users behaviors to define the group abnormality, but fail to pay attention to the underlying network structure; (2) define the group in local scale with distance-based restrictions such as distance, radius, or even nodes numbers, but fail to consider in the global perspective, as nodes with far distance could be highly associated. We pay attention to the global level group abnormality, without setting any restriction to the group definition, consider both the users' behavior and the underlying graph structure. Investigating this phenomenon of broad group abnormality behavior online holds enormous potential for understanding large-scale, disruptive societal events, such as mass movements.

Climate related protest causality: The occurrence of either a shift in climate, extreme weather, or environmental catastrophe is not sufficient to guarantee that civil unrest is likely to follow. In general the causal mechanisms leading to civil unrest are very complex, and there is no easy way to determine a linear pathway to protest. What is climate related protest evolution pattern, thus how does the climate disasters lead to armed protests? What is the coherent correlations among the climate protests?

Misinformation campaigns propagation: As millions of users post various messages every second, every one of them is a potential information source, resulting in multiple propagation paths, mixed messages, innuendos, falsehoods, and rumors. How do we track the spread of rumors and misinformation campaigns and can we distinguish them from 'regular' or normal propagation patterns? Can we distinguish real movement from rumors or misinformation campaigns?

1.2 Methods

Here we present an overview of the methods used in this dissertation. They will serve as the foundation to the key new information diffusion models proposed in Chapters 2–5.

1.2.1 Geometric Brownian Motion

Brownian motion is the random motion of particles suspended in a fluid (a liquid or a gas) resulting from their collision with the quick atoms or molecules in the gas or liquid. This term can also refer to the mathematical model used to describe such random movements, which is often called a particle theory [59].

Geometric Brownian motion is a continuous-time stochastic process in which the logarithm of the randomly varying quantity follows a Brownian motion (also called a Wiener process) with drift. It is an important example of stochastic processes satisfying a stochastic differential equation (SDE); in particular, it is used in mathematical finance to model stock prices (such as the price of a stock over time), subject to random noise.

A stochastic process S_t is said to follow a geometric Brownian motion if it satisfies the following stochastic differential equation:

$$dS_t = \mu S_t dt + \delta S_t dW_t$$

we call W_t as a Wiener process (Brownian motion) and μ the drift, δ the volatility.

Consider a Brownian motion trajectory that satisfies the differential equation, $\mu S_t dt$ controls the ‘trend’ of this trajectory and the term $\delta S_t W_t$ controls the ‘random noise’ effect in the trajectory. The analytical solution of this geometric brownian motion is given by

$$S_t = S_0 \exp\left(\left(\mu - \frac{\delta^2}{2}\right)t + \delta W_t\right)$$

According to the GBM properties, $\ln(S_t^{ij})$ is a Gaussian variable given by:

$$\ln(S_t^{ij}) \sim \mathcal{N}\left(\left(\mu - \frac{\sigma^2}{2}\right)t, \sigma^2 t\right)$$

1.2.2 Graph Wavelets

Graph wavelets are a form of graphical models bringing three kinds of benefits: (a) they can represent the social network (structure), (b) they perform inference between nodes/edges; and (c) they can help capture the properties of the social network. We employ graphical models here for spatial information propagation and specifically graph wavelets to study absenteeism. The classic wavelet has been referred to as a mathematical microscope since it is capable of showing signal abnormality with different scales. Wavelets help analyze signals which contain features that vary in time, space, and frequency (scale). Graph wavelets are particularly suited to study complex networks, as they 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.

1.2.3 Epidemiological Models

Epidemiological models provide a foundational approach in social network analysis since they elucidate the embedded information diffusion process. These models typically divide the total population into several compartments which reflect the status of an individual. For instance, common compartments denote susceptible (S), exposed (E), infected (I), and recovered (R) individuals. Individuals transit from one compartment to another, with certain probabilities that have to be estimated from data. The simplest model, SI, has two states; susceptible (S) individuals get infected (I) by one of their neighbors and stay infected thereafter. While conceptually easy to understand, it is unrealistic for practical situations. The SIS model is popular in infectious disease modeling wherein individuals can transition back and forth between susceptible (S) and infected (I) states (e.g., think of allergies and the common cold); this model is often used as the baseline model for more sophisticated approaches. The epidemic model SIR was firstly proposed to simulate the disease spreading on population groups in 1927 [48], which enables individuals to recover (R) but is not suited for modeling news cascades on Twitter since there is no intuitive mapping to what ‘recovering’ means. The SEIZ model (susceptible, exposed, infected, skeptic) proposed by Bettencourt et al. [5] takes the interesting approach of introducing an exposed state (E). Individuals in such a state take some time before they begin to believe (I) in a story (i.e., get infected).

1.3 Goals of the Dissertation

The overall aim of this dissertation is to identify modeling approaches and strategies that identify novel information propagate patterns as motivated earlier. We propose four mass movement topics here.

Topic 1: Mass Protest Adoption in Social Networks Modeling the movement of information within social media outlets, like Twitter, is key to understanding to how ideas spread but quantifying such movement runs into several difficulties. Two specific areas that elude a clear characterization are (i) the intrinsic random nature of individuals to potentially adopt and subsequently broadcast a Twitter topic, and (ii) the dissemination of information via non-Twitter sources, such as news outlets and word of mouth, and its impact on Twitter propagation. These distinct yet inter-connected areas must be incorporated to generate a comprehensive model of information diffusion. We propose a bispace model to capture propagation in the union of (exclusively) Twitter and non-Twitter environments. To quantify the stochastic nature of Twitter topic propagation, we combine principles of geometric Brownian motion and traditional network graph theory. We apply Poisson process functions to model information diffusion outside of the Twitter mentions network. We discuss techniques to unify the two sub-models to accurately model information dissemination. We demonstrate the novel application of these techniques on real Twitter datasets related to

mass protest adoption in social communities.

Topic 2: Protests Detection from Group Abnormality Event detection in online social media has primarily focused on identifying abnormal spikes, or bursts, in activity. However, disruptive events such as socio-economic disasters, civil unrest, and even power outages, often result in abnormal troughs involving group absenteeism of activity. We present the first study, to our knowledge, that models absenteeism and uses detected absenteeism as a basis for event detection in location based social networks (LBSN) such as Twitter. Our framework addresses the challenges of (i) early detection of absenteeism, (ii) identifying the point of origin, and (iii) identifying groups or communities underlying the absenteeism. Our approach uses the formalism of graph wavelets to represent the spatiotemporal structure and user activity in a LBSN. This formalism affords multiscale analysis, enabling us to detect anomalous behavior at different graph resolutions, which in turn allows identification of event location and anomalous groups underlying the network. We introduce a systematic two-pass detection method using graph wavelets to detect group absenteeism and then check if there is a subsequent activity spike.

Topic 3: Causality Inference to Climate Related Protest To infer climate protest causality, we need to develop a classifier which is able to separate out climate related protests from others. By analyzing historical climate protest events, we identify that different climate disasters cause related protests with different time span, depends the climate disaster influence and frequency. From constructing knowledge graph to represent link relationships between entities, we discloses protest causalities in Latin American countries, illustrate the pathways from climate disasters to climate protests. This paper also identifies the climate related protest patterns, discover the coherent relationship among different protests demanding.

Topic 4: Real Movement Distinguish in Social Networks Quantifying information diffusion on social network has been an interesting and unresolved problem for several years now. A better understanding of information diffusion, especially how news and rumors propagate through a network empower us to design strategies that can enhance spreading of news and curbing of rumors. Epidemic models have been used in the past to study information diffusion based on an assumption that rumor/news spreading is no different than the propagation of a contagious disease.

We use an enhanced epidemic model SEIZ that has been specifically designed for information diffusion. The model introduces one more compartment called exposed (E), which refers to the individuals who has been exposed to a story but have still not adopted/rejected it. We use five true news stories and three rumors from varied geographical locations and topics. We also introduce a one-step graph transfer model that can mimic step by step information propagation on Twitter. Our experimental results prove that SEIZ model is far more accurate in describing information diffusion than the other baseline epidemic models.

Further, our one-step graph transfer model imitates information cascades of the stories with a very reasonable error.

1.4 Organization of the Dissertation

The remainder of the dissertation proposal is organized as follows.

In Chapter 2, we address the problem of multiple spaces information dissemination, such as via social networks and outlets such as word of mouth. Specifically, we introduce a trust function to simulate how users are influenced by their friends through direct mention using the ‘@’ symbol. We present how our bispace model can capture propagation in the union of (exclusively) Twitter and non-Twitter environments.

Chapter 3 defines social network movements by an undirected, weighted graph. We detect the group abnormality not only by observing the user activity, but also consider the whole network structure. We propose to use graph wavelet to detect the group abnormality from a global viewpoint. We pay attention to user activity vectors and model their behaviors on graphs and uses detected abnormality as a basis for event detection.

Section 4 describes how we deal with problem of identifying linkages between climate change related phenomena and climate protests. We build a climate protest classifier which is able to separate out protests directly or indirectly resulting from a major climatic, severe weather, or environmental event. By analyzing large historical protest reports, we make use of knowledge graph to represent the link relationships between entities, and further locate and identify the causality of most climate protests.

In Chapter 5, we investigate the problem of distinguishing real movements from rumors in social networks. Here we design strategies that can enhance the spreading of news and the curbing of rumors. We present how to simulate the ‘doubt’ and ‘believe’ sentiment propagations. We also introduce a one-step graph transfer model that can mimic step by step information propagation on Twitter. Finally, we test the models using five true news stories and three rumors from varied geographical locations and topics. We also study the problem of misinformation propagation in the era of Ebola. All the experiments are conducted on Ebola-related rumors and all the evaluations are based on real-world data.

Chapter 2

Modeling Mass Protest Adoption in Social Network Communities using Geometric Brownian Motion

2.1 Introduction

This chapter was published in the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD 2014) [44].

In recent years social networking sites such as Twitter and Facebook have provided not just a platform for communication but also a means of mobilization and strategic interaction between key players of social movements, e.g., protests. Traditionally social movements occur within a subset of the population and have spread through on-the-ground communities and unions. With the advent of leaner communication technologies like Twitter, the way such movements form and spread through modern society has changed. With Twitter, in particular, traditional slogans have transformed into hashtags which can offer a consistent way of communicating the reason and motivation of social movements like protests and uprisings.

In this chapter, we focus on Twitter’s user networks during protests and similar civil unrest activities in Latin America. Our goals are to model the propagation and growth of contagion-like protest waves within a social network and to understand the social and structural dynamics underlying such phenomena. The key problem is understanding the nature of information propagation among motivated users of a social network. We have observed that such mass protests emerge very swiftly and sharply. In Twitterspeak, they would be considered trending but most such trends quickly decline on the social network even if not in the physical world. Modeling protest-related topic propagation on networks involves several challenges.



Figure 2.1: Mexico teacher protest events from Sep 1 to Sep 7, 2013. The blue pins denote protest cities; the numbers in red denote the sequence of protests as they spread across the country.

First, social protest propagation through online media can spread over large areas more quickly than traditional methods since users are geographically distributed. For example, on September 1, 2013, the Mexican government's education reform bill drew the wrath of teachers country-wide who opposed the reform (which required regular assessments of their performance as educators). Twitter was a virtual loudspeaker, providing a platform for organization and strategization for teachers to put forth their arguments against the bill. A series of mass teacher protests erupted and spread from city to city. As shown in Fig. 2.1, we see the movement spreading over time to different locations with no obvious visual mobilization pattern. The second challenge is that Twitter's user network embodies many subgraphs based on social ties which might afford different propagation rates due to subgraph-specific structures.

Thus identifying how the cause of a protest is adopted by Twitter users and how mobilization happens in the underlying network is a difficult task. To address this problem we present an integrated framework with new theoretical models as well as empirical validation on real Twitter data for actual protests witnessed in the recent past. Our key contributions are:

- We model the inherent heterogeneity in propagation using a bispace model, comprised of the Twitter mentions network (where both globally and locally influential neighbors contribute to a user's recruitment) and a latent space (where external exposure to protest-related information is captured).
- We focus on the role of community-driven information propagation over the bispace

model. We use geometric Brownian motion (GBM) over the mentions network and Poisson processes over the latent space to model information propagation during mass social movements.

- We illustrate the effectiveness of our approach in modeling several key mass protest adoption scenarios in multiple countries of Latin America, viz. Argentina, Brazil, Colombia, Mexico, Uruguay, and Venezuela.

The rest of this chapter is organized as follows. Section 2 covers related work in the areas of social movements, information diffusion in networks, external influences, and Brownian motion. Section 3 proposes the geometric Brownian motion propagation mechanism. Section 4 introduces the bispace propagation model, especially the model of propagation in latent space. In Section 5, we present our dataset and experimental setup, followed by initial experimental findings. Section 6 discusses the evaluation results for our approach followed by a brief discussion in Section 7.

2.2 Related Work

We briefly review related work next, which comes from multiple areas.

Social movements: Oliver and Myers [67] develop a foundation for theoretical insights of social movements and describe the limitations of simplified models. The Arab Spring of 2010 served as a context for many researchers [29, 7, 89, 19, 74] to study the role social networking sites play in the spread and recruitment of participants in protests. A detailed anatomy of modern social protests is described by Saad-Filho [74] with the June 2013 anti-government protests in Brazil as a context. In this work, we study the processes and sociological impacts of protests in the modern era, fortified by online social networks and the communities in and around them.

Information diffusion in networks: Previous studies have approached the modeling of information propagation and diffusion in social networks through several means, e.g., contagion models (SIR [13] SISa [35]), diffusion based threshold and cascade models [47], rise-and-fall patterns [56], coverage models [83], and survival theory [70]. A good survey of different models of information diffusion is presented in [30].

External influences: We believe that the effects of influences that originate external to the observed diffusion network, such as mass media and offline spread of information, can impact the way in which information flows within the online network. Myers et al. [61] study the emergence of URLs on Twitter with a probabilistic generative process using both internal and external exposure curves in a contagion-like model. Similar attention to the role of external factors is paid by Crane and Sornette [20] for tracking the popularity of YouTube videos using a diffusion model. Iwata et al. [40] use a shared cascade Poisson process

model to discover latent influences in social activities such as item adoption. Using shared parameters among multiple Poisson processes, they were able to simulate sequences of item adoption events.

Brownian motion: Zhou and colleagues (e.g., [103, 31, 104]) develop the notion of Brownian motion on networks which they use to discover communities of hierarchical structure both locally and globally. We extend this approach in this chapter to formulate a propagation algorithm based on geometric Brownian motion (GBM). Borrowed from statistical physics, GBM has been used heavily in finance to model stock price movements. Scale invariance and the ability to model abrupt bumps along propagation paths are the primary motivations for using GBMs to model stochastic processes [86].

Our work builds on the concepts introduced in [31, 40, 103, 104] but differs from the other diffusion models described earlier by considering both the role of communities of users and the abrupt nature of propagation of volatile information such as mass social protests. We include the notion of bispace where both latent (attributed to external influences) and observed user network influences are considered. We infer propagation rates for communities in the observed network and allow implicit recruitment of users into protest actions through a Poisson process.

2.3 Formalisms

2.3.1 Basics

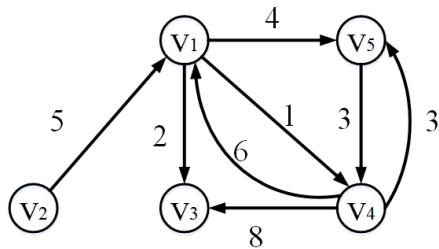


Figure 2.2: An example mentions network. Nodes denote Twitter users, directed edges denote direction of mentions between users, and edges are labeled with mention frequency.

We model Twitter activity as a network $G(V, E)$ of mentions. Here, each vertex $v \in V$ represents a Twitter user. There is a directed edge from user v_i to user v_j if v_i mentions v_j in a tweet. We define ω_{ij} to be the number of tweets in which user v_i mentions user v_j . Note that ω_{ij} is not necessarily equal to ω_{ji} . Key players such as celebrities and politicians are more likely to be mentioned by other users, rather than the other way around. As can

be seen in Fig. 2.2, the mentions network is a directed graph. Weight w_{14} is the number of times Twitter user v_1 mentions user v_4 , which is 1, while w_{41} is 6. Note that w_{21} is 5, while w_{12} is 0 (not shown).

We define the neighborhood $N(v_i)$ of a user v_i as the set of all users mentioned by v_i , i.e., those for whom there is a directed edge from v_i . For each user $v_j \in N(v_i)$, we define the Brownian distance from user v_i to v_j to be

$$d_{ij} = \frac{1}{(\omega_{ij} + 1)(\omega_{ji} + 1)^\gamma(\eta_{ij} + 1)^\gamma} \quad (2.1)$$

Here, η_{ij} is the number of common direct neighbors shared by user v_i and user v_j [104]. In Fig. 2.2, node v_1 and v_4 share two common direct neighbors— v_3 and v_5 —and hence η_{14} is 2.

We use the bias coefficient $\gamma \geq 1$ to heuristically weigh mentions that carry more impact. If v_i mentions v_j , meaning that $\omega_{ij} > 0$, we believe this expresses v_i 's intention to propagate information to v_j . Since v_j may not know or care about v_i and consequently may seldom or never mention v_i , the return mentions, measured by ω_{ji} , are (up)weighted by γ . Furthermore, if v_i and v_j share neighbors in the mentions network, the two users may have a closer relationship than other users with no shared mentioned Twitter users, and thus this component is weighted by γ as well. A Laplacian-style (+1) correction is used when there are no counter mentions or no mutual mentions. Note that for $\gamma = 1$, d_{ij} is an unbiased Brownian distance since ω_{ij} , ω_{ji} , and η_{ij} will have the same weight.

2.3.2 Trust functions and GBM

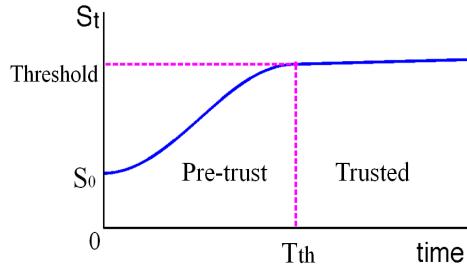


Figure 2.3: Trust function. A threshold defines the transition between the pre-trust and trusted period.

Next we introduce the notion of a trust function S_t which we use to model an individual user's agreement with an idea as expressed in tweets. (The trust function S_t is a function of the two entities between whom trust is modeled, but in this section we simplify the notation for ease of exposition.) We divide the trust process into a pre-trust period and a trusted period. In the pre-trust period, as a user receives new information, that user's trust, S_t , increases exponentially until S_t reaches the trust threshold at time T_{th} and enters the

trusted period. In the trusted period, new information increases S_t linearly. For simplicity, an individual user cannot revoke trust once this threshold has been crossed. In our Twitter mentions network, a user's trust in a topic crosses the threshold when they have tweeted about it. During the pre-trust period, we model the trust function as follows (the coefficient μ accounts for change in the average value of this stochastic process):

$$\frac{dS_t}{S_t} = \mu dt \quad (2.2)$$

We then add a Wiener process W_t to account for stochasticity. According to the properties of a Wiener process [66], dW_t is essentially Gaussian white noise and contributes to our equation as:

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t \quad (2.3)$$

In this way, we modeled the trust function S_t as a geometric Brownian motion (GBM) process which is a continuous-time stochastic process [66]. Per convention, we call μ the drift and σ the volatility. The drift represents deterministic trends while the volatility refers to the influence of unpredictable events in this model [93]. For simplicity, we consider μ and σ to be constant during the pre-trust period in this chapter. (Our concern here primarily is with this period.)

According to Itō's theorem [66], given the initial value S_0 , the above stochastic differential equation has the following analytic solution:

$$S_t = S_0 \exp \left(\left(\mu - \frac{\sigma^2}{2} \right) t + \sigma W_t \right) \quad (2.4)$$

The above solution for S_t is a log-normally distributed random variable with expected value and variance given as [66]:

$$E(S_t) = S_0 e^{\mu t} \quad (2.5)$$

$$Var(S_t) = S_0^2 e^{2\mu t} \left(e^{\sigma^2 t} - 1 \right) \quad (2.6)$$

S_t is a geometric Brownian motion stochastic process, which is typically denoted as $\mathcal{B}(\mu, \sigma)$. In this chapter we use an initial trust of $S_0 = 1$ without loss of generality.

2.3.3 GBM propagation

Suppose that user v_i posts a protest-related tweet at time t_0 which indicates that v_i has been recruited or infected. Whether v_i will infect its neighbor v_j depends on v_j 's trust function with v_i . For instance, if v_j is a close friend of v_i , then it is more likely that v_j will be infected in a short time because of v_j 's trust in v_i . But if v_j is not a very close friend of v_i , then it might take a long time to build v_j 's trust with v_i and to accept v_i 's status. Only after v_j 's trust with v_i crosses some threshold, v_j gets infected.

For better quantitative analysis, we consider d_{ij} to be the trust threshold. After crossing this threshold, v_j will agree with v_i 's opinion. According to the properties of GBMs, the trust function S_t grows continuously over time. This implies that, if some user is infected, all of that user's neighbors will eventually get infected given enough time for diffusion.

Since we assume a user cannot revoke trust, his or her status will never change once infected. Based on the above assumptions, we now detail our process for GBM propagation through the mentions network; see Algorithm 1. Since GBM is a time-continuous stochastic process, we discretize time using time steps of duration δt each. At the start of the simulation, all infected users are considered as newly infected users. Assume that the complete mass protest propagation duration is T . Once a user v_i becomes infected, the node is marked as a newly infected user, and the new status begins to affect the statuses of the neighbors, i.e., $N(v_i)$. For each user $v_j \in N(v_i)$, we use $t_{ij} = 0$ to initialize the time instant from which v_i begins to affect v_j . After all the time variables t_{ij} of $N(v_i)$ are so initialized, user v_i 's status is updated to reflect that v_i is no longer a newly infected user, to avoid duplicate initializations.

Suppose that at current time t , v_j 's trust with v_i is denoted as S_t^{ij} . According to the GBM properties, $\ln(S_t^{ij})$ is a Gaussian variable given by:

$$\ln(S_t^{ij}) \sim \mathcal{N}\left(\left(\mu - \frac{\sigma^2}{2}\right)t, \sigma^2 t\right) \quad (2.7)$$

If at time t , $\ln(S_t^{ij}) \geq d_{ij}$, this means that v_j gets infected since v_j 's trust with v_i is bigger than the distance d_{ij} . Now v_j begins to affect his or her own neighbors. Instead at time t , if $\ln(S_t^{ij}) < d_{ij}$, then at the next time step, $t + \delta t$, the trust is still a Gaussian variable, but with higher expectation and variance:

$$\ln(S_{t+\delta t}^{ij}) \sim \mathcal{N}\left(\left(\mu - \frac{\sigma^2}{2}\right)(t + \delta t), \sigma^2(t + \delta t)\right) \quad (2.8)$$

2.3.4 GBM parameter estimation

We use past protest events in which Twitter played a significant role in propagation to train our GBM model parameters. For each user who gets infected we record their Brownian distance and infection time. Suppose v_j gets infected by v_i after time t_{ij} ; then as per our propagation model, we claim that v_j 's trust function S_t^{ij} with v_i holds:

$$\ln(S_t^{ij}) \geq d_{ij} \quad (2.9)$$

where d_{ij} is the Brownian distance from v_i to v_j . For the convenience of parameter estimation, we can assume that $\ln(S_t^{ij}) = d_{ij}$. It then follows that d_{ij} is a normally distributed random variable which can be expressed as:

$$d_{ij} \sim \mathcal{N}\left(\left(\mu - \frac{\sigma^2}{2}\right)t_{ij}, \sigma^2 t_{ij}\right) \quad (2.10)$$

Because during the parameter estimation process, for each infected user v_j , we are not interested in exactly which user gets v_j infected, we use $x_j = d_{ij}$, and $\tau_j = t_{ij}$ in the following part of this section for simplicity. The set of n users that are infected during the infection process have independent infection rates, and we get the following likelihood function:

$$\mathcal{L}(\theta, \sigma^2 | v_1, \dots, v_n) = \prod_{j=1}^n \frac{1}{\sigma \sqrt{2\pi\tau_j}} \exp\left(-\frac{(x_j - (\mu - \frac{\sigma^2}{2})\tau_j)^2}{2\sigma^2\tau_j}\right)$$

The optimal estimators can be obtained by maximizing the above likelihood function. We differentiate the natural logarithm of the likelihood function above in terms of μ and σ , and set them to zeros. By solving the two equations simultaneously, we obtain the optimal estimators $\hat{\mu}$ and $\hat{\sigma}^2$.

2.4 Bispase Propagation Model

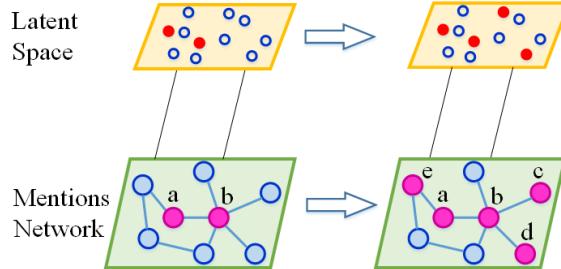


Figure 2.4: Bispase propagation model. In the latent space, users infections are explained by a Poisson model, and the red nodes denote the infected users from one time step to another. In the mentions network space, users are infected according to the GBM model. Here, the purple nodes (a, b, c, d, e) denote user infections explained by the GBM model.

Many information diffusion models assume that propagation occurs over a single domain. However, it is hard to build a complete, exhaustive network of interactions. For instance, consider building a network based only on which Twitter users follow which other users. This network will miss interactions such as retweets and mentions and the effect of influences originating outside of Twitter. Therefore, considering only a single space will make it difficult to account for all possible factors that influence the spread of information. In this study, we propose a bispase diffusion model that accounts for two domains of diffusion: the observed social network and the latent space, as can be seen in Fig. 2.4. In our case, the observed user space is the Twitter mentions network, whereas the latent space refers to any interactions outside of this network. To account for varying diffusion dynamics, each space is intended to have its own propagation model. As described earlier, we model propagation through the

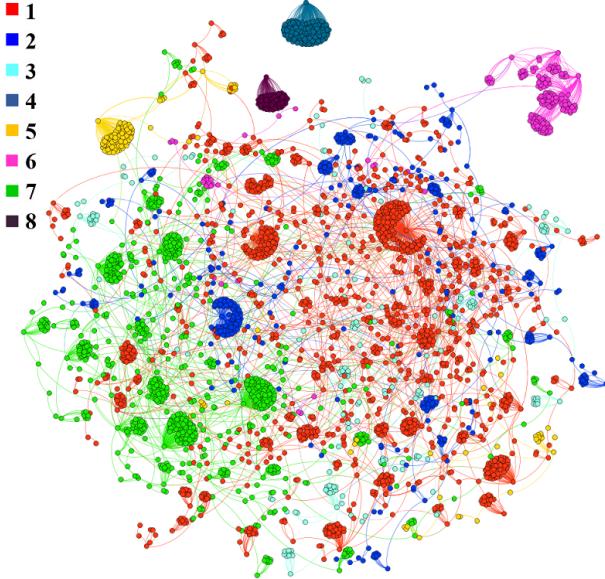


Figure 2.5: Major communities of teacher protest events (Sep 1 to Sep 12, 2013, Mexico).

Twitters mentions network as Geometric Brownian motion. We use the Poisson distribution to describe information propagation in the latent space.

2.4.1 GBM with Communities

Within networks, a community refers to the appearance of densely connected groups of vertices, with sparse connections between each group [65]. Instead of treating the whole network as a single propagation space, we use network structure to further split the network into communities. For our mentions network we use the Louvain method [6] for community detection to split the network into groups of users. For each community of users we can calculate classical graph features such as average degree, diameter, density, and clustering coefficient with which we can characterize them. In Fig. 2.6 we plot several features for each of the 8 communities found in the case study of Mexican teachers protest of 2013. Diameter $r = \max_{i,j} \text{dist}(v_i, v_j)$ is the length (in number of edges) of the longest geodesic path between any nodes v_i and v_j [64]. The clustering coefficient c_i is the proportion of node v_i 's neighbors that are connected. Graph density is defined as $\frac{2|E|}{|V|(|V|-1)}$ where E is the number of edges and V is the number of nodes [79]. As shown in Fig. 2.6, diameter and graph density vary considerably.

With the observed network further split into several communities, each community is intended to have its own model parameters for GBM. In GBM, $\ln(S_t^{ij})$ is a Gaussian distribution $\mathcal{N}((\mu - \frac{\sigma^2}{2})t, \sigma^2 t)$. We assume that each user within a community shares the same μ and σ so

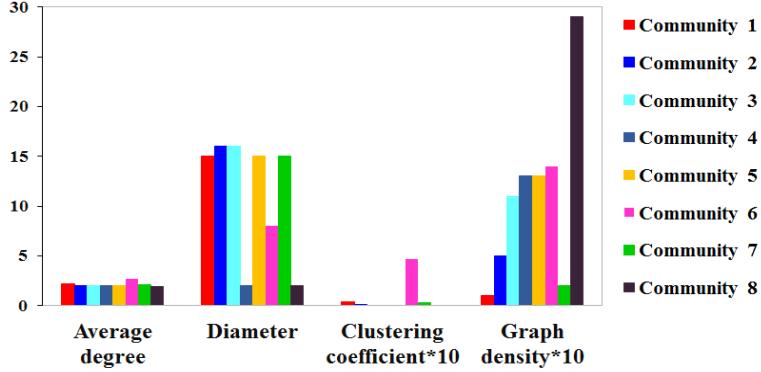


Figure 2.6: Key graph properties of communities underlying the Mexican teacher protest events.

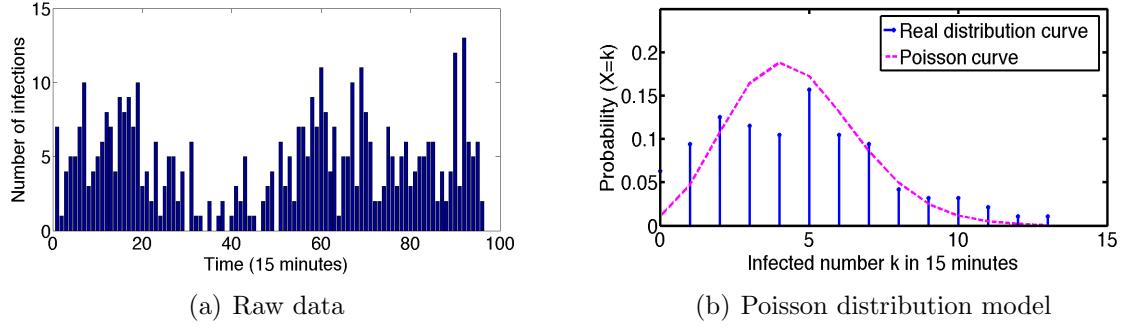


Figure 2.7: Poisson distribution in latent space propagation. (a) shows the raw data outside of the mentions network of teacher protest events on Sep 3, 2013. (b) shows the probability distribution of the number of infections.

that each community has its characteristic μ and σ . As information propagates through the mentions network, it may pass through different communities. For an infected user v_i and one of the non-infected neighbors $v_j \in N(v_i)$, we assume the following propagation strategy:

- If v_i and v_j are in the same community c_i , the propagation process will follow $\mathcal{B}_{c_i}(\mu_{c_i}, \sigma_{c_i}^2)$.
- Propagation from one community to another happens as per the source community's model parameters. For instance, for propagation from community c_i to community c_j , we will use the source community c_i 's GBM parameters.
- After information propagates into a different community, it will spread according to the new community's parameters. Once the information has entered community c_j from community c_i , subsequent infections henceforth will use community c_j 's parameters.

At each time step we use the μ and σ of any given node's current community for propagation from that node.

Table 2.1: Mass protests studied in this chapter.

No.	Event	Hashtags	Country	Affected cities	Event date(s)
1.	YoSoy132 student movement	#LaMarchaYoSoy132, #YoSoy132, #132, #soy132	Mexico	Nationwide	2012-05-17 to 2012-05-25
2.	Anti-government protests against tax reform and other policies pursued by President Juan Manuel Santos	#CacerolazoPaSantos, #5D	Colombia	Nationwide	2012-12-05
3.	Education reform protests by teachers	#ReformaEducativa	Mexico	Nationwide	2013-09-01
4.	Social protests against violence and crime	#UruguayosIndignados, #HartosDeLaViolencia	Uruguay	Montevideo	2012-05-14
5.	Protests against the “media law”	#LorenzettiNoMeFalles, #MediosBuitres	Argentina	Buenos Aires	2012-11-27
6.	Protests against Senate President Renan Calheiros’s election	#STFjulgueRenan, #SocorroJoaquim, #ForaRenan	Brazil	Nationwide	2013-02-22 to 2013-02-26
7.	Anti-government student protests against abuse of public media for election campaign	#ConatelCareTabla	Venezuela	Caracas	2013-03-20

2.4.2 Propagation in Latent Space

As mentioned before, in the latent space, we are modeling unobserved interactions of users. Since there are so many factors that might affect the dissemination of information, such as news outlets, word-of-mouth, it is reasonable to assume that the probability of the number of newly infected users in a given time interval satisfies the Poisson distribution [32] in the latent space.

For each node in the mentions network, it can only be infected by the GBM process. However, for those isolated users outside the mentions network, it is only possible that they get infected via the mechanics of the Poisson process. (Recall that in the GBM process, users get infected primarily via their neighbors.) We use X to represent the number of infected users with time interval δt and so the probability of the infected users is given by:

$$\Pr(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} \quad (2.11)$$

To obtain an estimator of λ , we can only use information about Twitter users who are outside the mentions network as our training dataset. We count the infected users outside the mentions network with time interval of 15 minutes during the Mexican teachers protest, and plot them as shown in Fig. 2.7(a). Adequately modelable by a Poisson distribution, we use the average value as the estimate of λ . Fig. 2.7(b) depicts the Poisson distribution fit with $\hat{\lambda} = 4.18$. If there are M_0 isolated users, the probability of each of these users to get infected in time interval δt is λ/M_0 . To summarize, for any user not in the mentions network, infection is only possible via the Poisson process. For a user who is already in the mentions network, infection can only happen via the GBM process over the mentions network, as described earlier.

2.5 Experiments

2.5.1 Dataset description

The study described in this chapter uses two datasets: (i) a gold standard report (GSR) of social unrest events in Latin America provided by MITRE that we use to define major mass protest events, and (ii) tweets collected over 14 months from May 2012 to September 2013 from 20 Latin American countries.

The GSR documents each civil unrest event by location, date, type of protest, and specifies the national news articles that first reported the event. For protests that were prominent on Twitter, the GSR news articles often report hashtags which were used by protestors on social media. We selected only those GSR events for which we were able to find such hashtags. This process resulted in 64 unique hashtags related to 40 different protest that occurred in Latin America since May 2012. In Table 1 we list a few of these events from our study.

Our Twitter dataset was built by querying Datasift's streaming API. Each tweet payload includes crucial metadata along with the tweet's content. Though tweets from GPS-enabled devices include geographic coordinates, the percentage of such tweets in the collected sample was too low to be useful.

For this study, we further filtered tweets by removing those that do not contain hashtags relevant to a specific protest. Since most tweets do not have location data, we estimate their location by geocoding the tweet based on each tweet's content and properties of its user. We developed our own geocoding library that uses the World Gazetteer (<http://archive.is/srm8P>) database to lookup location names and geographic coordinates. Tweets can be geocoded to the user's location at the time of tweeting or a location of interest about which the user is tweeting. We focused on event geolocation, which looks for location or landmark names, such as *Plaza de la Independencia* or *Quito, Ecuador*, in a tweet's text. We generated a list of 2000 landmarks by extracting place names mentioned in GSR events which had high mutual information to civil unrest. In cases where no event location was found in a tweet's text, we use geo-coordinates or self-reported location string in the tweet's metadata.

Using the above pipeline we were able to extract and geolocate 20,227,830 unique users to build our mentions network from the filtered tweets that were spread over daily sub-networks.

2.5.2 GBM Diffusion Model

For each of our mass protest events, we filter by its specific keywords (hashtags) to obtain a set of relevant tweets and construct a mentions network from those tweets. We assume that information propagates from an initial infected user to other users through the network from one node to its neighbors. We build an adjacency matrix based on the mentions network and simulate the propagation using the GBM diffusion process as follows:

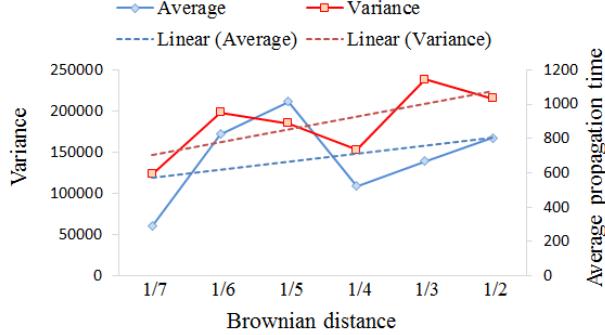


Figure 2.8: Brownian distance vs propagation time for teacher protest events.

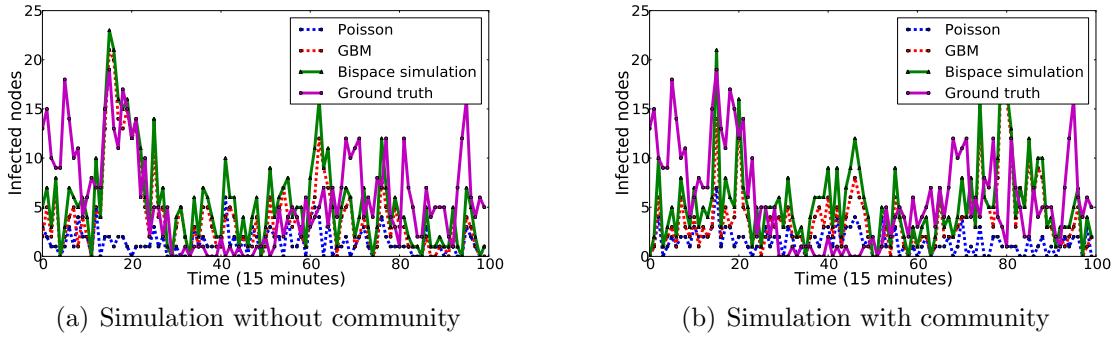


Figure 2.9: GBM and Poisson propagation simulation for Yosoy protests (Mexico) on May 19, 2012.

- 1. Brownian distance:** The Brownian distance is intended to have an inverse relationship with mention frequency. As Fig. 2.8 shows, users with smaller Brownian distance have greater mention frequencies resulting in shorter mean propagation times with less variance. From Fig. 2.8, we can see that infection time and variance generally both increase with an increase in Brownian distance. Heuristically, more frequent mentions indicate stronger ties which leads to easier adoption of information.
- 2. Propagation speed:** To evaluate our dynamic GBM infection process assumptions, we estimate the GBM parameters for different protest events and depict the GBM propagation curves in Figs. 2.9, 2.10, and 2.11. The blue curve depicts the Poisson propagation in latent space. The red curve depicts GBM propagation through the mentions network. The green curve is the overall simulation result while the magenta curve depicts the ground truth of the protest events process. By comparing the green and magenta curves, we can evaluate the effectiveness of our bispaces model in simulating the mass protest events. As shown, we find that, given a mentions network, our bispaces model can simulate the propagation speed at a reasonable scale, at the right magnitudes. As seen in Figs. 2.9(a), 2.10(a) and 2.11(a), we find that we can

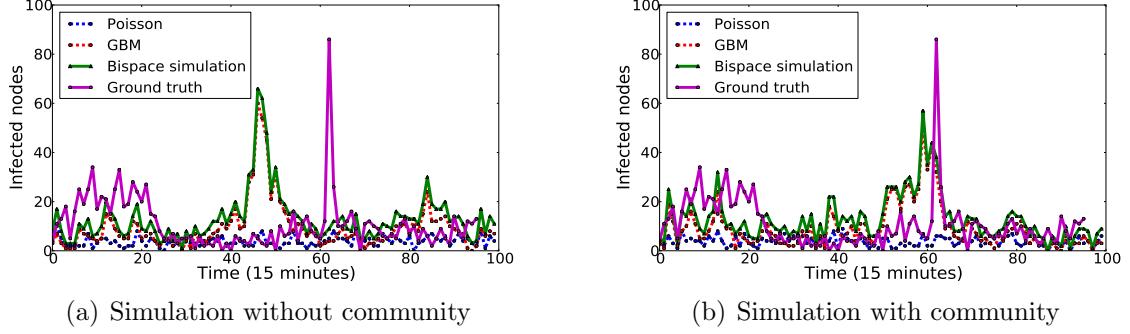


Figure 2.10: GBM and Poisson propagation simulation for teacher protests (Mexico) on Sep 1, 2013.

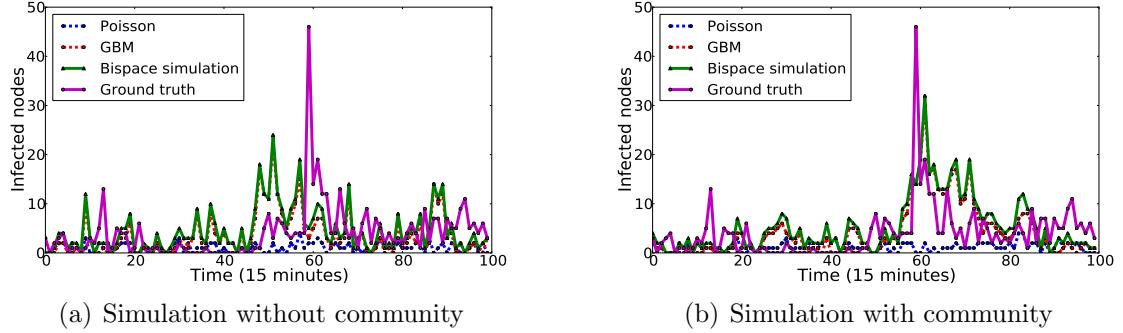


Figure 2.11: GBM and Poisson propagation simulation for Colombia protests on Dec 4, 2012.

capture the burst of activity at the same time point as the ground truth during protest propagation.

2.5.3 GBM Diffusion with Communities

We were also able to observe the variation in μ and σ as community structure varies. In particular, community features like graph density and diameter as shown earlier in Fig. 2.6 may impact GBM propagation. We experimented with two modeling approaches: (i) one set of parameters for the whole network and (ii) different parameters for each community in the network. We ran simulations for both these situations, and plotted the results of the whole network vs. community-specific approach in Figs. 2.9, 2.10, and 2.11. Comparing these simulation results, we find the community approach performs better, especially at capturing peak values. Taking a closer look at Fig. 2.12, we observe that propagation time and speed of infection are different for each community and we are able to simulate local propagation more accurately, which can be seen, e.g., from Fig. 2.10(b), where the GBM with community

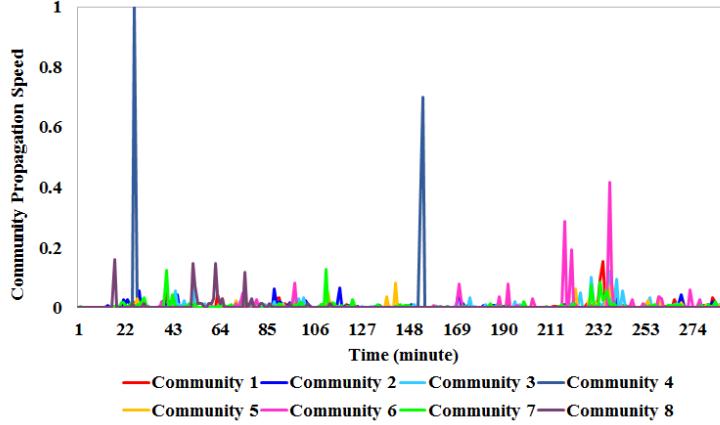


Figure 2.12: Normalized mass-protest propagation speed for major communities during teacher protest events.

method can simulate the burst propagation effectively, while the general GBM method (see Fig. 2.10(a)) fails to capture the exact peak time.

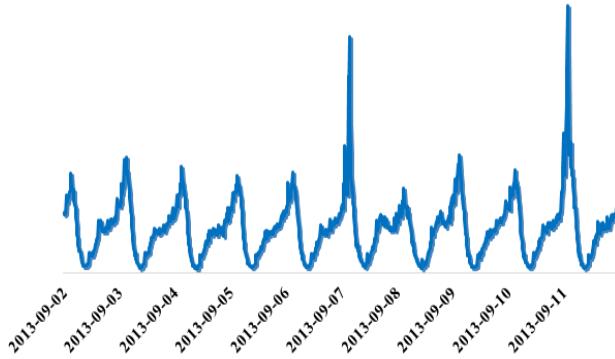


Figure 2.13: Total tweets over time from Sep 2 to Sep 11, 2013 (Mexico).

2.5.4 Latent Space Diffusion Model

We use the following steps to calculate the properties of the latent space for each event.

1. **Latent space:** The intent is to consider all possible external influences and latent interactions in this space. We split Twitter data into unique 15 minute intervals and count the total number of infected users in each interval.
2. **Normalize:** Twitter user activity varies based on time of day and day of week (see Fig. 2.13). For each 15 minute window from Step 1, we find the average number of

tweets over a 4 week period and use this value to normalize the count. This baseline count of tweets over time in the latent space is close to the Poisson distribution. Fig. 2.7(a) shows an example of this baseline.

3. **Train:** Using one week’s data split into 15 minute intervals, we train the Poisson distribution parameters. Fig. 2.7(b) shows that the training curve and ground truth curve can be matched quite well.

2.6 Evaluation results

Table 2.2: GBM simulation results for teacher protest events on Sep 2, 2013.

	Average degree	Diameter	Graph density	Connected components	Average clustering coefficient	Average path length
Simulation	1.791	11	0.002	183	0.083	4.786
Ground truth	1.726	18	0.002	204	0.008	6.261

We present an exhaustive evaluation of our bispace simulation approach alongside various dimensions next:

- **How effective is the performance of the bispace model?**

Recall that the bispace modeling is comprised of two independent process: the GBM simulation in the mentions network, and the Poisson process within the latent space. Given an initial mentions network, after training the GBM parameters of μ and σ , we proceed to conduct the GBM simulation. After estimating the Poisson parameter λ , we are able to do the Poisson simulation within latent space. We see that the GBM model is capable of capturing many mass protest scenarios, to the order of magnitude. Even though it cannot simulate the propagation speed accurately at every time point, the method is effective at capturing the total number of infected nodes with an accuracy of [0.78, 0.95], as shown in Fig. 2.14.

- **How adept is the bispace model at capturing surge/burst moments? How reliable are the simulation results?**

Fig. 2.9 depicts the analysis of the YoSoy132 student movement, whose Twitter activity is generally tortuous, and the curve is full of surges and bursts. From Fig. 2.9 we see that the bispace model is capable of simulating the general surge trends. Comparing the bispace simulation results with ground truth, we can see at many time points, the bispace simulation matches the ground truth. Fig. 2.11 shows the second protest of people protesting against the government in Colombia; here the Twitter activity depicts a burst at a single time point which is hard to capture. We can see the bispace model did show there is a burst, but not at the precise time point, one of its current limitations.

- Is the performance of the model better taking into account community structure?

After numerous experiments, we plot the accuracy distribution of both approaches for all our mass protest situations in Fig. 2.14. Although the accuracies are sometimes interspersed, we can see that in overall the community model generally has a higher accuracy.

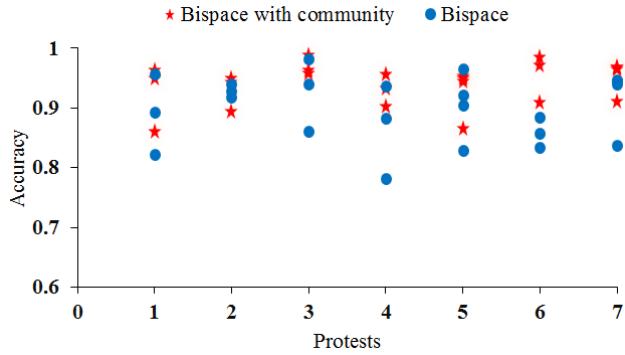


Figure 2.14: Performance accuracy of the bispace model for the 7 protest scenarios considered here, with and without community structure.

- Can the bispace model simulate the propagation path?

In addition to comparing the simulated counts of tweets over time with ground truth values, we can also compare the propagation path generated by the simulation against the actual propagation path through the mentions network. In Fig. 2.15 we can obtain a sense of the type of infection network bispace modeling creates as compared with the actual network. The simulation produces networks with relatively accurate paths and relevant characteristics as shown in Table 2.2. The component to which a user belongs is that of neighbors who can be reached from connected paths running along edges of the graph [64].

- Between the geometric Brownian model and Poisson propagation approaches, which model is more dominant during the simulation process?

From Figs. 2.9, 2.10, 2.11, by observing the blue dashed line (Poisson) and red dashed line (GBM), we can see that the Poisson process shows a mild activity, while the GBM model serves as the dominant component which can capture the moments of key surges.

2.7 Discussion

In this chapter, we have characterized mass protest propagation using a bispace model comprising an observed mentions network space and a latent space. We have introduced a

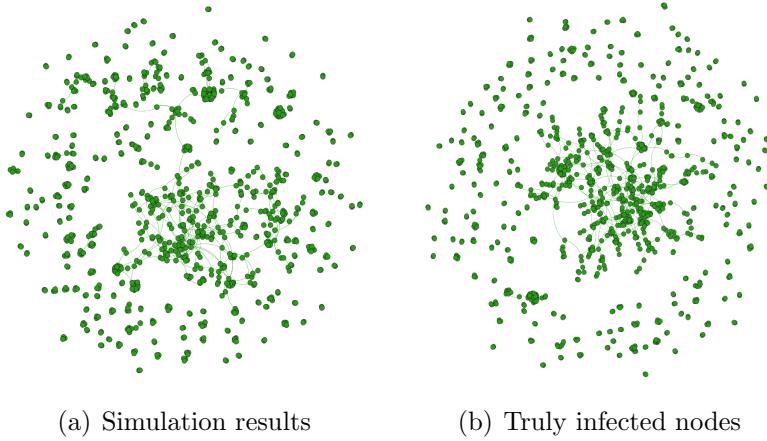


Figure 2.15: Bispace model simulation results compared against ground truth infected nodes for the Mexican teacher protest events.

trust function to simulate propagation in observed space using a geometric Brownian motion diffusion process which can be further extended to support communities with different propagation parameters per community. We considered the latent space of all interactions outside the mentions network to be a Poisson distribution process. We have shown how the GBM diffusion model offers a new approach for modeling propagation through social networks like Twitter. Through our experiments, we find that the time required for spread of protest information through such networks is dependent on the network's substructures. Furthermore, we find that modeling the diffusion process on a community basis provides better results than the assumption that all nodes in the network spread information in the same way.

In future work, we hope to further characterize the hidden network with the goal of uncovering specific latent variables. Additionally, we envision applying the GBM model to other networks, such as the Twitter follower network, to identify those paths most susceptible to information dissemination. Finally, we desire to compare propagation of mass protest language against other themes, such as celebratory events, to aid in determining correlations between topic or sentiment and the resulting social media diffusion.

Chapter 3

Protest Detection from Group Abnormality

3.1 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, Twitter has more than 255 million average monthly active users (78% from mobile) as of March 31, 2014, and an estimated increase of 25% per year¹. Various studies have shown that Twitter is viable as a social “sensor”, and holds great promise for detecting and forecasting significant societal events [9, 75]. In recent years, a significant body of research [1, 36, 52, 53, 75, 77, 91, 92, 98] has focused on modeling bursts and increases of user activity in social media.

However, real world events are not only correlated with burst signals, but can also exhibit unusually low levels of activity in social networks. As shown in Figure 1, a protest in the city of Natal, Brazil began at 5:00 PM (local time) at the Museum of the Republic, with people gradually joining the demonstration. On Twitter, there was an uncharacteristic lull in activity or *group absenteeism* behavior from 6:00 PM—8:00 PM on the same day.

Investigating this phenomenon of unusually calm behavior online holds enormous potential for understanding localized, disruptive societal events. In this paper we focus on absenteeism based event detection, 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 often precedes a major burst in re-activity). This paper presents the first study to systematically investigate group absenteeism in LBSNs. Using graph wavelet techniques, we pose this problem as one of group anomaly detection. To

¹<http://solomozone.com/tag/revenues/>

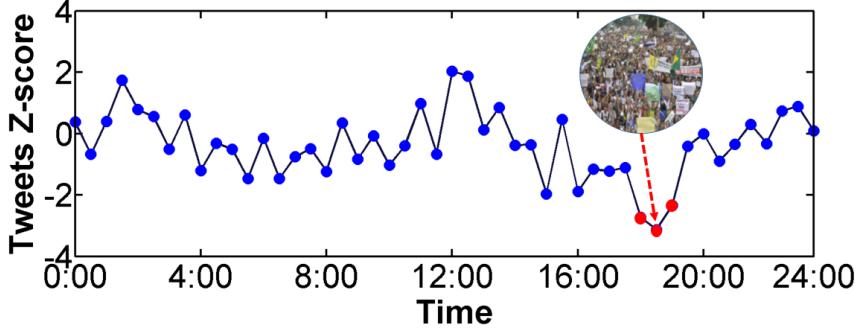


Figure 3.1: Detected group absenteeism in Natal, Brazil beginning at 6:00 PM on June 17, 2013. This absenteeism event coincides with a large protest that happened in the region.

appropriately incorporate absenteeism concepts into our detection approach, we must first address the following questions:

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

Graph wavelets display two outstanding advantages to study the above questions: scalability and low computational complexity. 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, and we are able to select absenteeism groups at different scales. Lastly, we propose a two-pass group anomaly detection method that first detects absenteeism, and then checks if there is a subsequent burst in activity within a specific time period. By comparing correlations between the wavelet coefficients of both of these groups, we are able to relate observed absenteeism to a possible real world event.

Our contributions are thus:

- To the best of our knowledge this is the first study to modeling group absenteeism as a basis for event detection.
- We incorporate graph wavelets as a mechanism to detect the most anomalous subgraphs at different scales. We demonstrate how this is a powerful technique for social media analytics.

- We propose a novel two-pass event detection method that uses correlation scores between the group depicting *absenteeism* and the group demonstrating increased activity to probabilistically determine the likelihood of an event.

The rest of the paper is organized as follows. Section 3.2 reviews related work and existing methodologies and Section 3.3 formalizes the research problem. In Section 3.4, we first discuss the graph wavelet formalism for group absenteeism detection, and subsequently demonstrate how it can be used for two-pass event detection. Section 3.5 presents extensive experiments for event detection, and the paper concludes with a summary of the research in Section 3.6.

3.2 Related Work

Event detection based on LSBNs is a research area that has attracted significant attention in the last years. Traditional approaches focus on capturing spatiotemporal burstiness of keywords [52, 53], Kalman filtering to track the geographical trajectories of hot spots of tweets related to earthquakes [75]; detecting topics of interest that are coherent in geographic regions [24, 36, 98]; applying clustering-based approaches search for emerging clusters of documents or terms using predefined similarity metrics that consider factors such as term co-occurrences and social interactions [1, 77, 91, 92]; and using the notion of compactness of a graph [72] to detect events. Several statistical methods have also been used, based on Kulldroffs spatial scan statistic [50], to detect spatial outliers [16] and have been applied to a wide variety of domains including transportation networks, civil unrest forecasting [101], and heterogeneous social media graphs [17].

Our approach to event detection problem is conceptually different from above mentioned studies. It includes a graph-theoretic framework to detect absenteeism related anomalies and correlate them with future events. Although group absence behavior has been widely studied in the area of organizational behavioral studies [25, 80], it remains unexplored in the area of social network analysis. Resembling closely to group anomaly detection in complex networks, our detection approach is further distinguished by its focus on groups rather than individuals. Existing approaches to group anomaly detection include building generative models of group anomalies [96, 99] where the goal is to automatically infer the groups and detect group anomalies in a social network. Typical to mixture models such methods suffer from high computational complexity due to the size of data and are heavily parameterized.

One of key challenges of our research problem is adapting the detection procedure for both missing and bursty activity groups. For this purpose, we incorporate spectral graph wavelets [33] into our algorithm. This strategy has been quite effectively used in multiscale community mining [87]. Wavelet methods based on spectral graph theory have been applied in a wide array data mining areas such as community detection, anomaly detection [12] and

other machine learning tasks [21, 26, 73, 82]. By constructing wavelets over graphs we are able take advantage of local information encoded in graph structure and then cluster and identify nodes which are similar in a scale-dependent fashion.

3.3 Preliminaries

In this section, we formalize our approach to event detection. We first describe the accompanying notations in section 3.3.1 which will be used throughout the paper. Then we formally present our research problem statement, provide a brief comparison of our approach to a conventional solution, and review the challenging issues that are relevant to an event detection problem.

3.3.1 Notations

Let's assume we are given an undirected, weighted graph $\mathbf{G}(V, E, W; f)$, where $V = \{v_0, v_1, \dots, v_{N-1}\}$ represents the set of N cities, E refers to the connections between neighboring cities, and W is a vector of non-negative weights associated with each edge $e_{ij} \in E$ as a function of geographical distance between a pair of vertices (v_i, v_j) . The function, $f : V \rightarrow \mathbb{R}^N$ maps the vertices of graph \mathbf{G} , and $f(n)$ stands for the value on the vertex v_n . Graph \mathbf{G} 's adjacency matrix \mathbf{A} is of size $N \times N$, where each element a_{ij} is represented as:

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

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 $\mathbf{D} = \text{diag}\{d_1, d_2, \dots, d_N\}$. 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 $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_{N-1} = \lambda_{\max}$ and a complete set of \mathcal{L} 's normalized eigenvectors [4] χ_i for $i = 0, 1, 2, \dots, N - 1$ described by:

$$\mathcal{L}\chi_i = \lambda_i\chi_i \quad (3.2)$$

Obviously, $\chi_o(n) = \frac{\vec{1}}{\sqrt{N}} = \frac{1}{\sqrt{N}}\{1, 1, \dots, 1\}$, and is called direction component of \mathbf{G} .

Absenteeism Score Although the raw volume of user interactions is a good indicator of users' online activities, it can be noisy and often exhibits a strong temporal dependence [18]. For example, in Twitter, the number of user interactions tend to peak later in the day. This can be attributed to the fact users tweet typically at home after work or school. In order to differentiate event-related absenteeism from such noisy signals that arise from the periodicity of users' daily activities, we need to remove these artifacts from our time series

data. Empirically, tweeting locations can be modeled as a Gaussian distribution, which allows us to transform the raw counts of tweets from a given city, v_i , at time interval, t , into a *z-score* measure, and in turn to calculate its *absenteeism score* as:

$$f^t(i; T) = (X_i^t - \mu)/\sigma \quad (3.3)$$

, where i denotes the index of the vertex, X_i^t is the tweeting volume at time interval t , μ is the trailing T -day moving average of the volume at time t , and σ is the standard deviation of that average volume. Here, a positive absenteeism score indicates a high levels of user activity, while a negative score indicates lower levels in activity. For the experiments described later in the paper, we set the value of $T = 30$ days. The notation for absenteeism function $f^t(i; T)$ is simplified to f when t and T are obvious from the context.

3.3.2 Graph Anomaly

According to Equation 3.2, eigenvalues of Laplacian matrix \mathcal{L} can be presented as:

$$\lambda_l = \chi_l^T \mathcal{L} \chi_l = \sum_{e_{mn} \in E} w_{mn} [\chi_l(m) - \chi_l(n)]^2 \quad (3.4)$$

λ_l summarizes all the eigenvector deviations on any directly connected vertex 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. As the eigenvalue increases, larger differences in neighboring components of the graph Laplacian eigenvectors may be present. Hence, for larger λ_l , its corresponding eigenvector, $\chi_l(n)$, has larger deviation among connected vertices [?]. For this reason, we call $\{(\lambda_l; \chi_l)\}$ the graph anomaly pattern (also called Fourier frequency by some researchers) of \mathbf{G} . As mentioned above, $\chi_0(n)$ is the direct component of \mathbf{G} since $\chi_0(i) = \frac{1}{\sqrt{N}}$ for any $v_i \in V$.

It is useful to analyze f by taking into account the intrinsic geometric structure of the graph \mathbf{G} . In order to identify and exploit structure of $f \in \mathbb{R}^N$, the spectral graph $\sigma(\mathcal{L}) := \{\chi_l\}_{l=0}^{N-1}$ can be used as a dictionary of atoms [21]. Thus, f can be decomposed as a linear combination of $\{\chi_l\}_{l=0}^{N-1}$ as

$$f(n) = \sum_{l=0}^{N-1} \hat{f}(l) \chi_l(n) \quad (3.5)$$

, where

$$\hat{f}(l) := \sum_{n=0}^{N-1} \chi_l^*(n) f(n) \quad (3.6)$$

$\hat{f}(l)$ is the inner product of f and anomaly pattern χ_l , and is called the graph anomaly degree in this paper, and is also called the corresponding Fourier coefficient.

3.3.3 Generalized Graph Anomaly

Equation 3.6 gives a clear representation of the anomaly patterns in $f(n)$ based on graph. As discussed in section 3.3.2, λ_l only summarizes all deviations among all the direct connected vertices in graph \mathbf{G} . However, in many applications, deviations among vertices which are not connected directly might also carry important values. Taking social media network for instance, human behavior is not only being affected by his/her direct connected friends, but also by some “far distance” friends in the network. Generalizes graph anomaly considers the deviations among all vertex pairs, which even not being connected directly, as long as they are close enough to each other.

Let $d_G(m, n)$ denote the minimum number of edges for any paths connecting v_m and v_n in graph \mathbf{G} , and $d_G(m, n)$ can be written as:

$$d_G(m, n) = \arg \min_p \{k_1, k_2, k_3, \dots, k_p\} \quad (3.7)$$

subject to

$$m = k_1, n = k_p, \text{ and } w_{k_r, k_{r+1}} > 0 \text{ for } 1 \leq r < p \quad (3.8)$$

Note that d_G disregards the values of the edge weights.

Lemma 1 *Let \mathbf{G} be a weighted graph, \mathcal{L} the graph Laplacian and $p \geq 0$ an integer. For any two vertices v_m and v_n in graph \mathbf{G} , if $d_G(m, n) > p$ then $\mathcal{L}^p(m, n) = 0$.*

The comprehensive proof of lemma 1 can be found in [33]. $\mathbf{G}^p(V^p, E^p)$ denotes the graph with Laplacian matrix \mathcal{L}^p , and ρ_{mn} denotes the weight of edge e_{mn} , where $e_{mn} \in E^p$. Obviously, $V^p = V$. According to the properties of positive semi-definite, the eigenvalues and eigenvectors of \mathcal{L}^p are $\{(\lambda_l^p; \chi_l)\}$, where $0 \leq l \leq N - 1$. According to Equation 3.4, for \mathcal{L}^p , we have

$$\lambda_l^p = \chi_l^T \mathcal{L}^p \chi_l = \sum_{e_{mn} \in E^p} \rho_{mn} [\chi_l(m) - \chi_l(n)]^2 \quad (3.9)$$

Further, according to lemma 1, if $d_G(m, n) > p$, then $\mathcal{L}^p(m, n) = 0$, which equivalently means $\rho_{mn} = 0$. Hence, λ_l^p only summarizes deviations among all vertex pairs which are closer than p edges in graph \mathbf{G} . For this reason, we call $\{(\lambda_l^p; \chi_l)\}$ the generalized graph anomaly pattern of \mathbf{G} .

3.3.4 Problem Statement

In this paper, we focus on the problem of event 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, $\mathbf{G}(V, E, W; f^t)$ at time interval t , select a subset $\Sigma \subseteq V$, such that

$$\Sigma = \arg \min_{P \subseteq V, P \text{ is compact}} \sum_{v_k \in P} f(k) \quad (3.10)$$

A general solution to this problem is using a combinatorial optimization technique, where by defining a constrained objective function over a network one can identify subset of vertices which maximize the corresponding function [72]. Therefore, Equation 3.10 can be modified as:

$$\Sigma = \arg \min_{P \subseteq V} \sum_{v_k \in P} f(k) + \lambda \mu(P) \quad (3.11)$$

, where $\mu(P)$ is the compactness penalty function of P (e.g., the sum of distances among all pairs of the vertices in P [72]), and λ is the regularization parameter. Such methods suffer from the following issues:

1. To define and measure the compactness of subset $P \subseteq V$ is challenging, considering the exponential varieties of complex graphs.
2. To determine a suitable regularization parameter λ in the objective function is ambiguous, because simply combining multiple physically different concepts in the objective function makes the optima sensitive to λ .
3. To solve this objective function is often a **NP-hard** problem, which makes it unpractical in many real world applications. Sometimes, even the approximate solutions are of high computation complexity, if there are any.

In contrast, our approach proposes a novel, absenteeism based events detection algorithm in social networks using spectral graph wavelet theory. The graph wavelets focus on the intrinsic geometric structure of the graph by transversing each vertex $v_i \in V$, and mining the topological information of both local and globally centered vertices supports the ability to conduct a multiscale analysis. In addition, the graph wavelet approach does not introduce any “subjective” objective functions or other compactness concepts, and thus provides a fair and low computational method in terms of complexity for identifying abnormal group behavior in a wide variety of application scenarios.

3.4 Two-pass Event Detection Model

This section describes the modeling approach and the implementation details of our event detection algorithm. We begin by presenting concepts of graph wavelets and their properties in Section 3.4.1. Section 3.4.2 discusses applications of graph wavelets in group anomaly detection. Lastly, Section 3.4.3 describes the group absenteeism based two-pass event detection algorithm in detail.

3.4.1 Graph Wavelets

Classic wavelet is called mathematical microscope since it is capable of showing signal abnormality 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 3.6, the anomaly pattern $\hat{f}(l)$ presents the anomaly components of f based on \mathbf{G} 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. [33] proposed constructing wavelet transforms of 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 $v_a \in V$ 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 v_a is constructed as:

$$\psi_{s,a}(n) = \sum_{l=0}^{N-1} g(s\lambda_l)\chi_l^*(a)\chi_l(n) \quad (3.12)$$

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

$$W_f(s, a) = \langle \psi_{s,a}, f \rangle = \sum_{l=0}^{N-1} g(s\lambda_l)\hat{f}(a)\chi_l(n) \quad (3.13)$$

Similar to classical wavelets, graph wavelets provide following three properties, which are presented in detail in [33].

1. **Reconstruction.** When generating 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

$$\hat{h}(\omega) = \sqrt{\int_{\omega}^{\infty} \frac{|\hat{g}(\omega')|^2}{\omega'} d\omega'} \quad (3.14)$$

, 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:

$$\phi_a(n) = \sum_{l=0}^{N-1} h(\lambda_l) \chi_l^*(a) \chi_l(n) \quad (3.15)$$

Accordingly, the scaling coefficients are defined as

$$S_f(a) = \langle \phi_a, f \rangle \quad (3.16)$$

Using scale set $\Theta := \{s_j\}_{j=1}^J$, the discretized graph wavelet set $\{\psi_{s_j,a}\}_{j=1}^J$, and scaling function set $\{\phi_a\}_{a=0}^{N-1}$ constitute a frame [33]. So there will be NJ wavelet coefficients in the frame. According to frame theory [23], $f \in \mathbb{R}^N$ can be reconstructed by a limited number of scaling and graph wavelet coefficients. A detailed algorithm and treatment concerning the choice of Θ can be found in [33].

- 3. Localization in vertex domains.** Given a central vertex v_a and its graph wavelet $\psi_{s,a}$, suppose the kernel function g is $K+1$ times continuously differentiable, satisfying $g^{(r)} = 0$ for all $r < K$ and $g^{(K)} = C \neq 0$. If there is $s' > 0$ such that $|g^{(K+1)}(\lambda)| \leq B$ for all $\lambda \in [0, s'\lambda_{max}]$, let v_n be a vertex of \mathbf{G} such that $d_G(n, a) > K$, then there exist constants D and s'' , such that

$$\frac{\psi_{s,a}(n)}{\|\psi_{s,a}\|} \leq Ds \quad (3.17)$$

for all $s < \min(s', s'')$. $d_G(n, a)$ is the geodesic or shortest path distance, which is the minimum number of edges in any path that connect vertices v_n and v_a [33]. Equation 3.17 shows for any vertex v_n that is far away from center vertex v_a ($d_G(n, a) > K$), its wavelet value $\psi_{s,a}(n)$ is upper bounded by Ds . In other words, for vertex v_n which is far away from vertex v_a , its wavelet value is linearly attenuated by scale s . When the scale s is small, their wavelet coefficients will be vanished quickly.

3.4.2 Group Anomaly Detection

Recall that in Equation 3.12, $\|\psi_{s,a}\|$ depends on scale s because of the kernel function g . Thus for fair comparison we normalize the graph wavelet coefficient as:

$$W'_f(s, a) = \langle \psi'_{s,a}, f \rangle = \langle \frac{\psi_{s,a}}{\|\psi_{s,a}\|}, f \rangle \quad (3.18)$$

$W'_f(s, a)$ represents the topological distribution of $f(n)$. And for $\psi_{s,a}$ its kernel vertices denoted by $\mathcal{K}(\psi_{s,a})$, defines a set of vertices such that $d_G(n, a) \leq K$ and $0 \leq \psi_{s,a}(n)$. The other vertices are called as marginal vertices. When f has a uniformly large positive/negative values on the kernel vertices, $W'_f(s, a)$ is going to be a large positive/negative value as well.

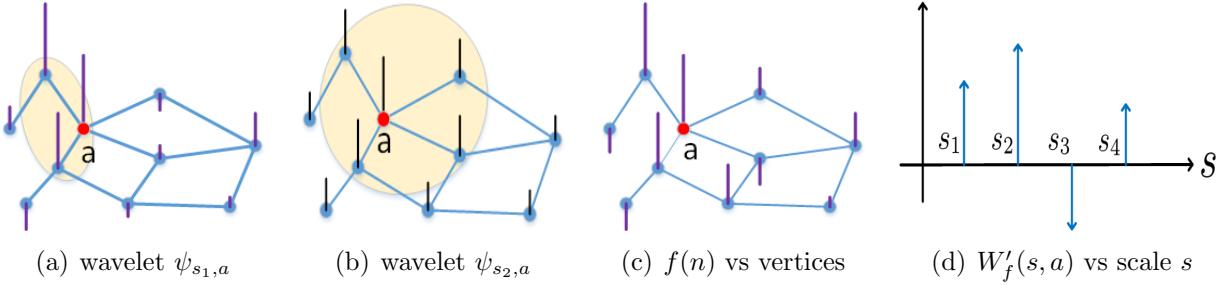


Figure 3.2: Using graph wavelets for abnormal group identification.

Otherwise, $|W'_f(s, a)|$ will be greatly attenuated. These weights are linearly controlled by scale s . According to Equation 3.17, when s is small, the weights of the marginal vertices are severely attenuated, and hence the weights of the kernel vertices are enhanced since $\psi_{s,a}$ is normalized. Essentially, $W'_f(s, a)$ is equivalent to the sum of f with large weights on kernel vertices, and small weights on marginal vertices. When f is of uniformly large negative/positive value on kernel vertices, then $W'_f(s, a)$ will be a large negative/positive value with scale s . We call the wavelets with minimal and maximal $W'_f(s, a)$ absenteeism wavelet and burst wavelet, respectively.

The localization property of graph wavelet makes it appropriate for group anomaly detection since it automatically identifies the kernel vertices from V . These kernel vertices form a compact subset since each one of them is close to the same center vertex v_a . Thus, under this definition of compactness in graph wavelet, the problem of Equation 3.10 can then be converted to the following problem: *given a graph and absenteeism score vector, $\mathbf{G}(V, E, W; f^t)$ at time interval t , select a graph wavelet $\psi_{s,a}$ and its kernel vertices Σ , such that*

$$\psi_{s,a} = \arg \min_{s_j \in \Theta, v_n \in V} W'_f(s_j, n) \quad (3.19)$$

, and

$$\Sigma = \mathcal{K}(\psi_{s,a}) \quad (3.20)$$

As Equation 3.19 avoids the compactness constrain condition, thus its complexity is greatly reduced.

Figure 3.2 shows two graph wavelets centered on the same vertex v_a , but with two different scales, $\psi_{s_1,a}$ and $\psi_{s_2,a}$, where $s_1 < s_2$. The length of the black bar on each vertex denotes its graph wavelet value. The highlighted areas denote the kernel vertices with larger graph wavelet values. Figure 3.2(c) is f 's distribution along each vertex, whose value is denoted as the vertical bar. As we can see, with the graph wavelet pattern of $\psi_{s_2,a}$, f shows a larger value on most of the kernel vertices, and $W'_f(s_2, a)$ a large value as well. That also means $\psi_{s_2,a}$ optimally partitions the set V into two distinct groups - kernel and marginal vertices. As shown in Figure 3.2(d), $W'_f(s_3, a)$ is where the absenteeism is most significant, and $W'_f(s_2, a)$ is where bursty behavior is observed.

Remarks

1. As graph wavelet and scaling functions 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 select vertices that are close to the central vertex v_a , and attenuate the impact of other marginal vertices that are far away from v_a . Unlike other conventional methods, $\psi_{s,a}$ is automatically scalable, and maintains the graph's topological information.
3. The graph wavelet does not introduce any objective functions and constraints. In addition, the scale $\{s_j\}_{j=1}^J$ set is numerically small, and once the eigenvalues and eigenvectors of \mathbf{G} are known, the computation complexity of graph wavelet coefficients is $O(NJ)$, which makes it easily adaptable to wide variety of application scenarios.

3.4.3 Group Absenteeism Based Event Detection

This section proposes a novel two-pass absenteeism based event detection algorithm. The underlying rationale of this algorithm is based on the following concepts.

1. As discussed in section 3.4.1, distribution of f can be well reconstructed by the J scaling and NJ wavelet coefficients. Each of those normalized wavelet coefficients $W'_f(s, a)$ represents a distribution pattern of f on \mathbf{G} . It is equivalent to saying that $\psi_{s,a}$ represents a special distribution pattern, which shares a large and uniform value around the central vertex with scale s .
2. When a significant event occurs, preceded by group absenteeism behavior in social networks, such as a severe earthquake or a massive protest, it is likely to be succeeded by a spike or burst in online user activity. With this observation, we can represent an absenteeism behavioral pattern as ψ_{s_l, a_l} at time l centering at vertex v_{a_l} , and a burst related pattern as ψ_{s_τ, a_τ} at time τ centering at vertex v_{a_τ} . We assume the burst pattern happens within the time window size of L after absenteeism pattern is identified. Further, a notion of response time can be represented using the time difference $t_{rsp} = \tau - l$.
3. Both absenteeism and burst signal must show a strong correlation, especially if they occur in close proximity spatially and temporally.

$$\rho(\psi_{s_l, a_l}, \psi_{s_\tau, a_\tau}) = \frac{\langle \psi_{s_l, a_l}, \psi_{s_\tau, a_\tau} \rangle}{\|\psi_{s_l, a_l}\| \cdot \|\psi_{s_\tau, a_\tau}\|} \quad (3.21)$$

Algorithm 1 Two-Pass Absenteeism Event Detection

```

1: Input: graph and absenteeism score vector  $\mathbf{G}(V, E, W; f^l)$  at time interval  $l$ , and time window size  $L$ .
2: Output: correlation  $\rho$  and response time  $t_{rsp}$ .
3: compute the spectral  $\sigma(\mathcal{L})$  of graph  $\mathbf{G}$ ;
4: set the graph wavelets  $\psi_{s,a}$  and scales set  $\{s_j\}_{j=1}^J$ ;
5: compute  $W'_f(s_j, a)$  for all  $v_n \in V$  and  $s_j \in \{s_j\}_{j=1}^J$ ;
6: detect the most absenteeism pattern  $\psi_{s_l, a_l}$  of  $f^l$ 
7: for all  $\tau$  from  $l + 1$  to  $l + L$  do
8:   calculate the most burst wavelet  $\psi_{s_\tau, a_\tau}$  of  $\mathbf{G}(V, E, W; f^\tau)$  at time interval  $\tau$ ;
9:   calculate correlation coefficients  $\rho(\psi_{s_l, a_l}, \psi_{s_\tau, a_\tau})$ ;
10: end for
11: identify the maximal coefficients  $\rho_{max} = \text{argmax}_\tau \rho(\psi_{s_l, a_l}, \psi_{s_\tau, a_\tau})$ ;
12: return  $\rho_{max}$  and  $t_{rsp}$ .

```

Based on these concepts, the higher the correlation, the higher probability that burst patterns is caused by the preceding group absenteeism. When ρ is above the threshold (threshold is set at 0.5), we infer that an event occurred and that it evolved on social networks into distinct phases: first group absenteeism, followed by a spike or burst in user activity.

In summary, the two-pass absenteeism event detection can be summarized as shown in Algorithm 1. Because the computation of the graph spectral graph $\sigma(L)$ is a one-time computation, the two-pass algorithm has $O(NJ)$ complexity.

1. Compute \mathbf{G} 's spectral graph $\sigma(L)$. Because $\sigma(L)$ is independent of the time interval, it is computed only once and can be solved by classical matrix factorization methods.
2. Set scale set $\{s_j\}_{j=1}^J$ according to algorithms in [33], and the NJ normalized graph wavelet coefficients $W'_f(s_j, a)$ at current time interval l .
3. Identify the most negative $W'_f(s_l, a_l)$, and determine the corresponding ψ_{s_l, a_l} as the absenteeism pattern.
4. For all the time interval τ from $l+1$ to $l+L$, compute all the $W'_f(s_j, a)$, detect the most positive one and identify the corresponding graph wavelet ψ_{s_τ, a_τ} as the burst pattern at time τ , and the correlation score ρ_τ between absenteeism pattern ψ_{s_l, a_l} and burst pattern ψ_{s_τ, a_τ} .
5. Detect the largest ρ_τ as ρ_{max} ; return ρ_{max} and response time t_{rsp} .

3.5 Experimental Results

This section discusses the applications of our approach for detecting group absenteeism. We begin by briefly describing the datasets we use for our experiments in Section 3.5.1.

Then we discuss in Section 3.5.2 how we assemble the graph \mathbf{G} and construct the graph wavelets $\psi_{s,n}$ for absenteeism based event detection. Finally, in Section 3.5.3, we describe the three case studies that illustrate how the two-pass event detection method is able to identify absenteeism events.

3.5.1 Data Collection and Preprocessing

The study described in this paper uses tweets from 22 countries in Latin America that were collected over 12 months (May 2013 to May 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 [69]. 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 tweet’s 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. Of those, we focused on 1290 cities that had a daily average tweet volume greater than 100.

3.5.2 Experimental Setup

In our experiments, for each Latin American country we create a graph $G(V, E, W; f)$ where each node $v \in V$ describes a city, f denotes the Z-score distribution along each node, and each edge connects the nearby cities. To reduce the graph complexity, we employ a 10-nearest neighbor algorithm to construct edges in the graph. By traversing each v the edge set E is determined, where weight of edge e_{ij} is computed as:

$$w_{ij} = \exp(-d_{ij}^2/2\sigma^2) \quad (3.22)$$

, where $d_{ij} = \sqrt{(lat_i - lat_j)^2 + (lon_i - lon_j)^2}$, and $\sigma^2 = 1$. lat_i and lon_i denotes the geographical coordinate pair representing each city. Our choice for the wavelet generating kernel function, $g(x)$, is motivated by our goal to achieve scale-dependent localization.

$$g(x) = \begin{cases} x & \text{for } x < 1 \\ s(x) & \text{for } 1 \leq x \leq 2 \\ 2x^{-1} & \text{for } x > 2 \end{cases} \quad (3.23)$$

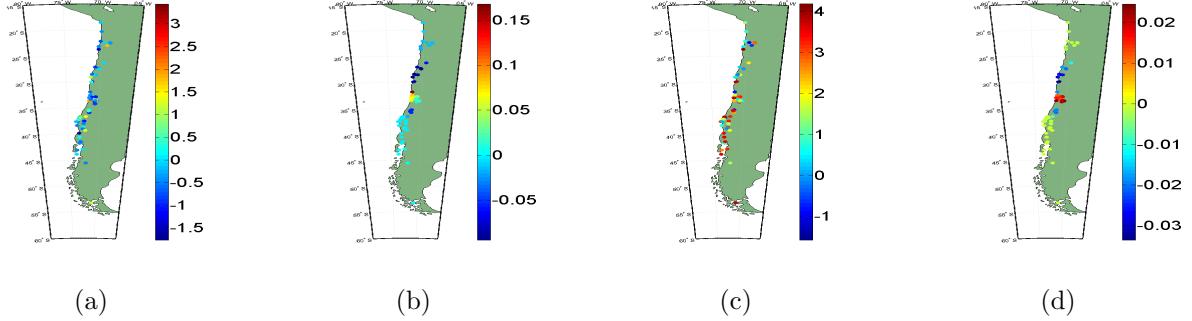


Figure 3.3: Iquique Earthquake, Chile. Above plots show differences in distributions of absenteeism score and wavelet coefficients calculated at 8:45 PM, April 1, 2014 (a-b) involving group absenteeism and later when burst in activity is captured at 11:00 AM, April 2, 2014 (c-d), respectively.

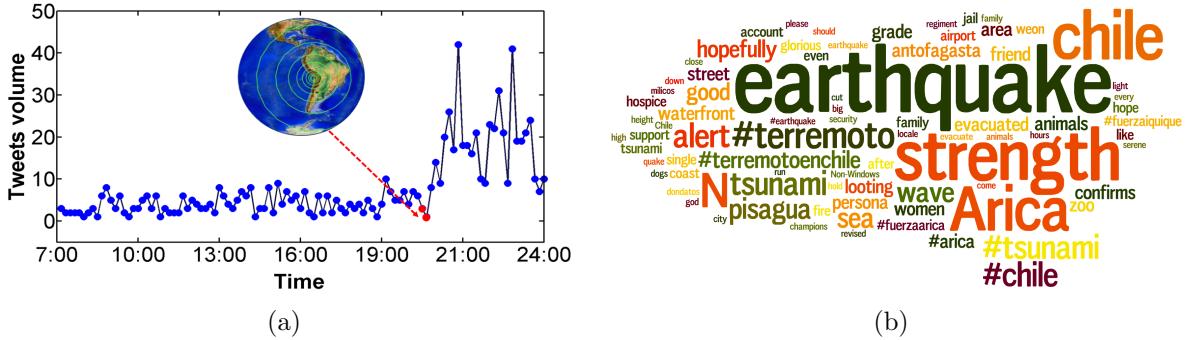


Figure 3.4: Iquique earthquake, Chile, April 1, 2014. (a) Tweet time series for Iquique on April 1, 2014. (b) Word cloud of tweets which mention ‘Iquique’.

where $s(x) = -5 + 11x - 6x^2 + x^3$. 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 [33].

3.5.3 Case Studies

Case study 1: Iquique Earthquake, Chile. On April 1, 2014 around 8:46 PM (local time) a large earthquake struck off the coast of Chile, northwest of the port city of Iquique. We show the distribution of absenteeism scores and normalized wavelet coefficient values of the graph wavelets from the beginning of this event and over a 24 hour period. As shown in Figure 3.3(a) we observe an absenteeism behavior, where the scores are dominated by very low (blue spectrum) of Z-score values (indicating high absenteeism). Likewise in Figure 3.3(b), we witness low coefficients values for the northern regions of Chile, where the impact of the earthquake was most significant. As the news of earthquake spread throughout

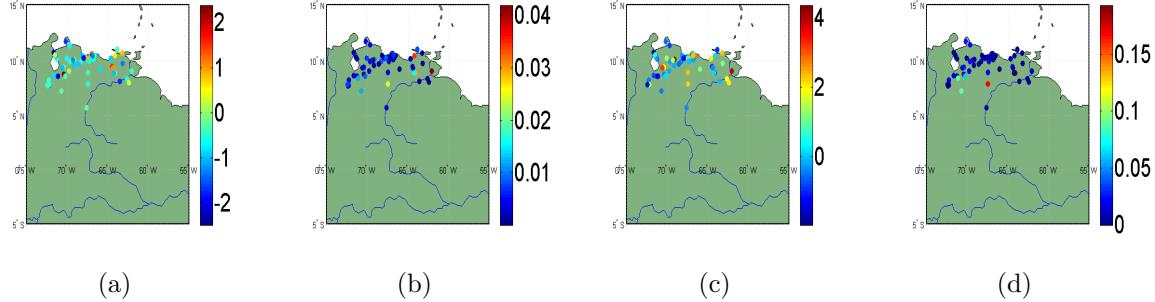


Figure 3.5: Power Outage in Venezuela. Above plots show differences in distributions of absenteeism score and wavelet coefficients calculated at 7:40 PM, December 2, 2013 (a-b) involving group absenteeism and later when burst in activity is captured at 8:45 PM in the same day (c-d), respectively.

the next day, user activity on social media increased. This bursty behavior is seen on April 2nd, around 11:00 AM. We can observe from Figure 3.3(c) that our z-scores have increased (red spectrum) significantly and the coefficient value distribution (see figure 3.3(d)) of graph wavelets, for northern regions of Chile is now in red spectrum. From the graph wavelet distributions in Figs. 3.3(b) 3.3(d), we can see that the kernel area of the absenteeism/burst wavelets cover most large negative/positive values. In this way, the wavelet identifies the abnormal negative/positive groups in absent/burst time intervals, respectively. Furthermore, a high correlation score of 0.726 was calculated for the wavelets from absenteeism and bursty periods of this episode. As a result, we note that there is a strong connection between the burst in activity and the previously observed absenteeism, signaling an event was detected.

From the graph wavelets generated in absenteeism time period 3.3(b), we found the central node to be the city of ‘Iquique’. We study the timeseries (Fig. 3.4(a)) of Twitter activity for Iquique and word clouds (see Fig.. 3.4(b)) generated from their content, to see how events unfolded during the course of the earthquake. Strong absenteeism is observed from 8:45 PM to 9:20 PM. We also checked user mobility through geotagged tweets from city of Iquique, on April 1, 2014 and found that the user mobility fraction had increased by 15.4%.

Case Study 2: Massive power outage in Venezuela. A massive power outage in Venezuela plunged several major cities including the capital city, Caracas in to darkness around 7:40 PM (local time) on December 2, 2013. News media reported ², that the power outage lasted for 10-15 minutes, and the people of Caracas soon went out in the streets to protest. This action at the beginning of the episode coincides with the absenteeism period detected by our algorithm. The scatter plots showing distribution of absenteeism scores and wavelet coefficients (Figs. 3.5(a), 3.5(b)) indicate that most of the low values are less than 0. Shortly after the absenteeism, we detected a huge burst in activity around 8:45 PM, signaled by the increased z-scores (low absenteeism) and coefficient values (Figs. 3.5(c), 3.5(d)). A

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

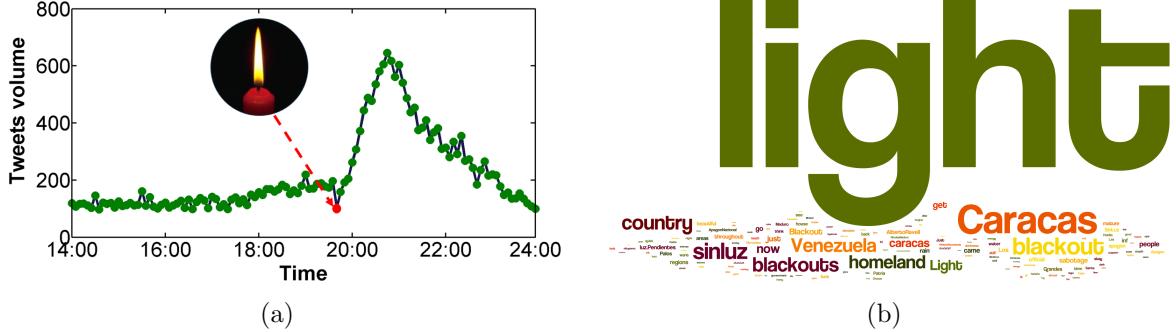


Figure 3.6: Caracas, Venezuela power outage, December 2, 2013. (a) Time series of tweets volume. (b) Word cloud of tweets mentioning ‘Caracas’.

correlation score of 0.617 was calculated on comparing the graph wavelets from both absentee and burst period.

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

Case Study 3: Christmas Day. As mentioned earlier, an absenteeism behavior may not always lead to a spike in activity. In this case, our model detected strong absenteeism in social media activity for major holidays such as Christmas day, however it was not followed by a bursty period in Twitter activity. One explanation of this behavior is that people tend to travel back to visit family during the holidays. This is supported by low values of z-scores or high absenteeism in Figure 3.7(a) and wavelet coefficients in Figure 3.7(b) with respect to Argentinian tweets on December 25, 2013. Hence, no subsequent burst period was detected for this event.

Interestingly as the Christmas day approached we observed (see Figure 3.7(c)) that user mobility gradually increases and z-score decreases signaling greater absenteeism. We used Pearson’s correlation coefficient to measure the two time series and found a correlation score of -0.94.

3.5.4 Why Absenteeism Group Detection?

Previous research has demonstrated the importance of burst detection in Twitter. In our study, we argue that group absenteeism can also be vital for detecting disruptive societal events. Modeling absenteeism is crucial, because it can serve as a surrogate signal for event detection. For example, in the case of the Iquique earthquake, where our algorithm detected

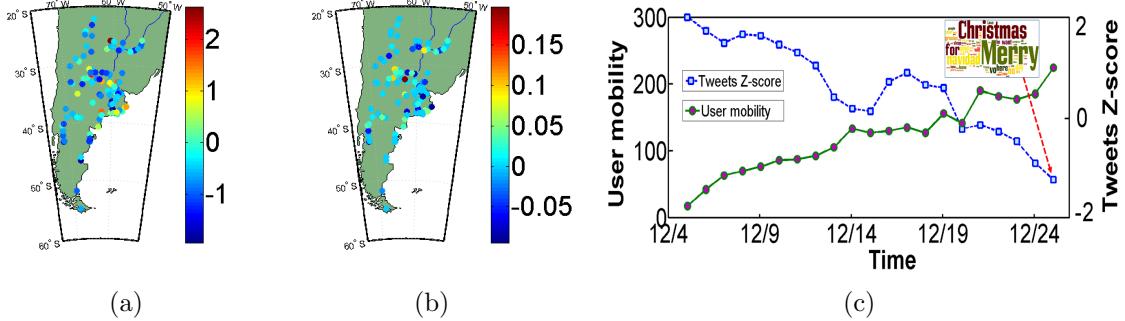


Figure 3.7: The Christmas Day in Argentina: Above plots shows distributions of (a) absenteeism score and (b) wavelet coefficients calculated on December 25, 2013. (c) Time series comparing absenteeism score and user mobility corresponding to tweets between December 5, 2013 and December 25, 2013.

an absenteeism behavior on Twitter followed by a spike in user activity. Unlike traditional event detection methods which identify real time events only after they have occurred i.e., once the burst signal has been identified; the absenteeism signal can be observed much earlier, and it renders a foresight or view into the future events. Our approach thus offers a significant advantage over current strategies that focus solely on modeling spike or burst related patterns for event detection.

3.6 Conclusion

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

In future work, we plan to extend our detection model to capture extent of an event's influence over network. Another interesting extension of our work would be to include absenteeism as feature to classify event of different nature (disruptive vs non-disruptive).

Chapter 4

Causality Inference to Climate Related Protest

4.1 Introduction

Climate change, extreme weather, and the state of the environment directly impact the availability of food [2], [3], energy [58], and shelter [90]. As finite resources become scarce, the residual impacts on local economies can have disastrous and long-lasting effects on the fundamental livelihoods of inhabitants for decades [54]. The examples of this occurring are numerous. The extended drought in Syria in 2011 is cited as one of the principle causes of civil war [27, 46]. In a smaller scale example, the environmental impact of lead contamination in the drinking water in the United States led to protests in 2016. The extreme weather event, Hurricane Manuel, that devastated the western coasts of Mexico led to subsequent protests over resources at points as long as 1 year after the initial event.

Of course, the occurrence of either a shift in climate, extreme weather, or environmental catastrophe is not sufficient to guarantee that civil unrest is likely to follow. In general the causal mechanisms leading to civil unrest are very complex, and there is no easy way to determine a linear pathway to protest. However, to date, little quantitative analysis has been performed on the residual effects of changes resulting from climate, extreme weather, and the environment using a large volume of data. In this analysis, we focus on the breadth of the climate events by looking at events generated from a large Gold Standard Report (GSR) containing all of the protests that have occurred in Latin America from 2011-2015.

GSR is a gold standard report of protests organized by MITRE, using human analysts, to survey newspapers for reportings of civil unrest. The GSR includes many features, as shown in Figure 4.1, such as protest location, event date, protest type, status, crowd size, headline, date, population, protest description, first reported links, etc.. The description feature is brief description of the protest, generally, it tells us who, where, why and when protest. As

Figure 4.1 shows, the protest description is ‘small farmers want the bank to forgive their debts due to the drought, which has hampered production’.

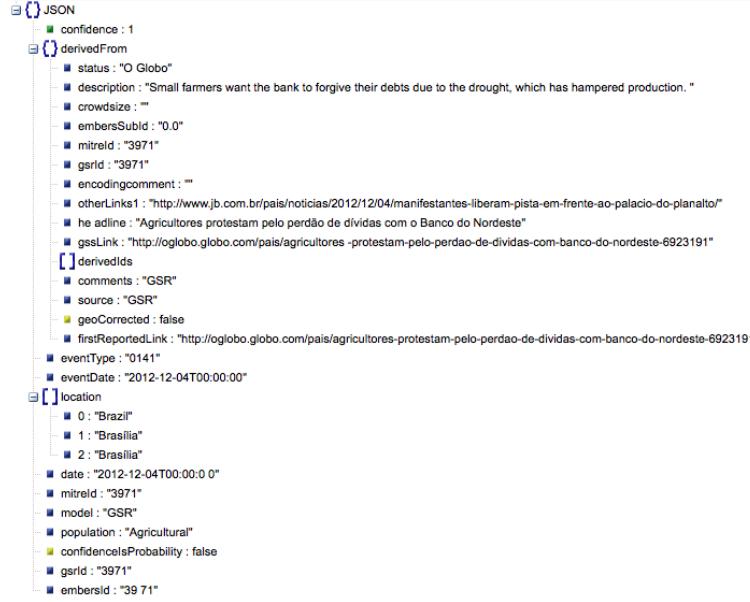


Figure 4.1: Gold standard report (GSR) format.

We address three foundational problems: first, the identification of climate related protest. By developing a logistic regression classifier, 25352 GSR civil unrest events were classified as either being climate or non-climate related using terms in the description of the event. Second, from analyzing large historical climate protests records, we look into the text description of protests and extract the climate protest category for each country. For each major climate category, we adopt the knowledge graph approach to define linkage relationship between entities, and study the possible protest causalities. Third, we find that the massive climate protests have coherent correlations within them, we also show that certain protest types are more prone to associate with certain other protest types, such as lack of water is highly linked with power shortage, and so on. Generally, the main contribution of this paper can be summarized as:

1. We develop a logistic regression classifier, which can classify climate protests from non-climate protests automatically based on protest event descriptions.
2. We analyze the climate protest spikes and disclose its relationship with climate disasters. For instance, the time span caused by storm and hurricane events in Mexico last much longer. However, for drought events in Brazil, the protests being initiated more swiftly, also last much shorter.
3. We figure out the proportion of protest causality. By studying some major climate

disasters, we also discover each protest category's evolution pattern, thus how does the climate disasters lead to armed climate protests.

4. We investigate the climate co-occurrence. For instance, the water related protests are often accompanies with electricity shortage, while land ownership protests are often associate with farmers.

4.2 Related research

The path from climate, extreme weather and environmental effects to civil unrest is causally complex [37, 78] and involves various combinations of climate change [10], natural resources, human security, and social stability. In general, sensitivities to climate change, exposure to climate change, and the ability of a society to adapt are indicators of whether or not violence will erupt [38]. A commonly studied pathway is the effect of climate on food prices which then induces civil unrest. An examples of this occurrence is the Arab Spring uprisings in 2011, and how weather effects food prices [45]. The pathway to civil unrest is also not limited to a local region, where one study shows the Chinese drought effecting the supply wheat causing prices to rise in the Egyptian break market leading to protest [85]. The pathways of food prices to protest have also been studied in the global south [11], Africa, and Asia [94, 34]. However, even this path of climate effects on income level leading to conflict is not eminently clear [76].

4.3 Climate protest classifier

The classifier is designed to label text documents into two or more predefined categories. In this work, we only have two categories: climate or non-climate related protest. By sample analysis, more than 90% records belongs to non-climate related protest, thus the dataset can be ascribed as un-balanced dataset. So we consider majority assign classification as baseline, adopt other four classical classification methods: K-Nearest Neighbor, Naive Bayes, weighted SVM and Logistic regression.

4.3.1 Majority assign

Majority assign method is taken as a baseline for the unbalanced classification dataset. It first calculates the climate related protest rate with the training data-set as p , and non-climate related protest as $1 - p$, and then uses this distribution to randomly assign each testing event. Suppose there are N testing events, by this algorithms, the true-positive would be Np^2 , false-positive and false-negative would both be $Np(1 - p)$, on average. Hence

the precision, recall and F -measure would all be p , and the accuracy would be $p^2 + (1-p)^2$. For unbalanced data-set, since $p \ll (1-p)$, the accuracy approximately equals to $(1-p)^2$, while the F -measure is p .

4.3.2 K Nearest Neighbor

To classify a class-unknown document X , the K-Nearest Neighbor (KNN) classifier algorithm ranks the document's neighbors among the training document vectors, and uses the class labels of the k most similar neighbors to predict the class of the new document. The classes of these neighbors are weighted using the similarity of each neighbor to X , where similarity is measured by Euclidean distance or the cosine value between two document vectors [55].

KNN has been used in statistical estimation and pattern recognition already in the beginning of 1970s as a non-parametric technique. A case is classified by a majority vote of its neighbors, with the case being assigned to the class most common amongst its K nearest neighbors measured by a distance function. If $K = 1$, then the case is simply assigned to the class of its nearest neighbor. Choosing the optimal value for K is best done by first inspecting the data. We first manually identify 100 climate-related protest events as the training sets. In our protest filter design, text similarity measures play a fundamentally important, where apply Corpus-Based similarity for distance computation between different event descriptions. In our experiment, we set K to be 100.

4.3.3 Naive Bayes

Essentially, Naive Bayes is maximum a posteriori classifier, which can be represented as $c = \text{argmax}_c p(c|e)$. e is the protest description, and consists of multiple words w_i , and can be denoted as $e = \langle w_0, w_1, \dots \rangle$. $c = \{\text{climateprotest}, \text{non-climateprotest}\}$. However, there is no trivial solution to measure the joint probability distribution for e, c considering the extremely complex underlying structures among w_i . Naive Bayes circumvents this problem by assuming the independency among w_i . Hence, the probability of each protest e being class c can be simplified as:

$$p(c|e) \propto p(c) \prod_i p(w_i|c)$$

, where $p(w_i|c) = \frac{f_{w_i}^c}{f_w^c}$ is the conditional probability of term w_i that appears in the description of e . $f_{w_i}^c$ is the occur frequencies of w_i in class c , and f_w^c is the total word number in class c . If a new term w_i does not occur in the training dataset, then $p(w_i|c) = 0$. $p(w_i|c)$ measures how much likeness of being c for the existence of term w_i . To mitigate the zeroing affects, *Laplace – smoothing* modifies $p(w_i|c)$ as

$$p(w_i|c) = \frac{f_{w_i}^c + 1}{f_w^c + W}$$

, where W is the total word number for climate and non-climate protests together. Usually, the conditional probability is small which might results in float point underflow. In reality, it is converted as:

$$c = \text{argmax}_p(c|e) = \text{argmax}_c \{ \log(p(c)) + \sum_i \log(p(w_i)|c) \}.$$

4.3.4 Weighted Support Vector Machine

The training data consists of N pairs $(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$, with $x_i \in \mathbb{R}^p$, and $y_i \in \{-1, +1\}$. By introducing a hyperplane of $P := \{x | x^T \beta + \beta_0 = 0\}$, the classification rule is defined as $G(x) = \text{sign}[x^T \beta + \beta_0]$. To find the hperplane P for unseperable sets, it is often converted into the following quadratic convex optimization problem by defining the slack variables $\zeta = (\zeta_1, \zeta_2, \dots, \zeta_N)$.

$$\begin{aligned} & \underset{\beta, \beta_0}{\text{minimize}} \quad \frac{1}{2} \|\beta\|^2 + C \sum_{i=1}^N \zeta_i \\ & \text{subject to} \quad y_i(x_i^T \beta + \beta_0) \geq 1 - \zeta_i, \quad \forall i \\ & \quad \zeta_i \geq 0, \quad \forall i \end{aligned} \tag{4.1}$$

C is the penalty parameter. For separable sets, C corresponds to ∞ .

The problem with above classifier is that the penalty for misclassification are the same. However, there are a lot of cases the miss alram should have a much higher cost than the false alarms. To considet those scenarios, we introduce two different penalty for miss alarm and false alarm. For simplicity, I and J denotes the subscript of positive and negative set. Thus, the problem of 4.1 can be re-formulated as:

$$\begin{aligned} & \underset{\beta, \beta_0}{\text{minimize}} \quad \frac{1}{2} \|\beta\|^2 + C_1 \sum_{i \in I} \zeta_i + C_2 \sum_{j \in J} \eta_j \\ & \text{subject to} \quad \begin{aligned} x_i^T \beta + \beta_0 & \geq 1 - \zeta_i, \forall i \in I \\ x_j^T \beta + \beta_0 & \leq -1 + \eta_j, \forall j \in J \\ \zeta_i & \geq 0, \forall i \in I \\ \eta_j & \geq 0, \forall j \in J \end{aligned} \end{aligned} \tag{4.2}$$

The Lagrange function of 4.3 is

$$\begin{aligned}
L_p = & \frac{1}{2} \|\beta\|^2 + C_1 \sum_{i \in I} \zeta_i + C_2 \sum_{j \in J} \eta_j \\
& - \sum_{i \in I} \alpha_i [x_i^T \beta + \beta_0 - (1 - \zeta_i)] - \sum_{i \in I} \mu_i \zeta_i \\
& + \sum_{j \in J} \theta_j [x_j^T \beta + \beta_0 - (1 - \eta_j)] - \sum_{j \in J} \tau_j \eta_j
\end{aligned} \tag{4.3}$$

The input of the SVM experiment is GSR json file, and the input file includes lots of features. Among these features, several of them are more important than others, including status, description, crowdsize, headline, eventType, eventDate, location, date, population etc. The description feature is brief description of the events, which plays a dominant role in the entire dataset.

In order to apply SVM algorithm on GSR dataset, we need to vectorize text data in the dataset. First of all, we construct a word corpus which includes every word shown in the dataset (including non-words). We accept non-words because most coinages come from Internet and some of them might be important for the events. As we accept non-words, the corpus might be large than our corpus vocabulary. Then, for every fields, the content is converted to a corresponding vector based on the corpus. If one word exists in specific field, a calculated value will be assign to the corresponding element in the vector. Other numerical values will also be added in the vector. By this way, we could convert the dataset into a huge matrix.

Minimizing (2) can be rewritten as a constrained optimization problem with a differentiable objective function in the following way.

For each $i \in \{1, \dots, n\}$ we introduce the variable ζ_i , and note that $\zeta_i = \max(0, 1 - y_i(w \cdot x_i + b))$ if and only if ζ_i is the smallest nonnegative number satisfying $y_i(w \cdot x_i + b) \geq 1 - \zeta_i$.

Thus we can rewrite the optimization problem as follows

$$\text{minimize } \frac{1}{n} \sum_{i=1}^n \zeta_i + \lambda \|w\|^2$$

,

subject to $y_i(x_i \cdot w + b) \geq 1 - \zeta_i$ and $\zeta_i \geq 0$, for all i .

This is called the "primal" problem. After the preprocessing, we could apply SVM algorithm on GSR matrix. We randomly label 1000 events from GSR dataset and use ten-fold cross validation to train out SVM. Figure 1 shows the result of SVM experiment.

4.3.5 Logistic regression classifier

The logistic regression classifier is mathematically characterized as:

$$P(Y = 0|X = x) = \frac{1}{1 + \exp(\sum_{i=1}^N k_i x_i + b)}$$

$$P(Y = 1|X = x) = \frac{\exp(\sum_{i=1}^N k_i x_i + b)}{1 + \exp(\sum_{i=1}^N k_i x_i + b)}$$

The GSR includes many features, such as status, description, crowd size, headline, event Type, event Date, location, date, population etc. To reduce the computation complexity, we only apply the description text as input to the logistic regression classifier. First of all, we construct a bag of words from the training data-set descriptions by deleting meaningless stop-words, like "the", "a/an", "at", and etc.. The bag of words is composed of M words denoted as $[w_1, w_2, \dots, w_M]$. Each GSR description X is considered as a vector of length M . If word w_i occurs in its description, then $X(i)$ will be assigned with 1; otherwise 0. Further each protest in the training dataset is assigned $Y = 1$ as climate protest, or $Y = 0$ as non-climate protest by manually checking its description meaning. In this way, each GSR record is converted to a corresponding vector based on the bag of words. Second of all, we estimate k_i and b based on maximum likelihood criterion. This process is usually converted to convex optimization problem with efficient solutions [41]. Once the coefficients of k_i and b are estimated, if the probability of $P(Y = 1|X = x)$ is larger than 0.5, then the protest event will be classified as climate protest.

4.3.6 Evaluation

We manually labelled 1700 GSR protest records as climate or non-climate protests. Using 70% dataset as training, and the rest 30% as test. To ensure we have a trustworthy classification results, we evaluate the performance carefully by cross evaluation. The evaluation criteria are precision (positive predictive value), recall (true positive rate), F-measure (a measure that combines precision and recall) and accuracy (the proportion of true results both true positives and true negatives among the total number of cases examined). We compare with four well-known classification methods: majority assign, K-nearest neighbor, Naive Bayes, and weighted support vector machine (SVM). Since the climate events account for a small portion of all the events, which make it an unbalanced classification problem, so we change the traditional support vector machine into weighted SVM, by adding more importance to the climate protest events (we set the class weight to be 100). From Table 4.1, we prove logistic regression method outperforms other methods uniformly.

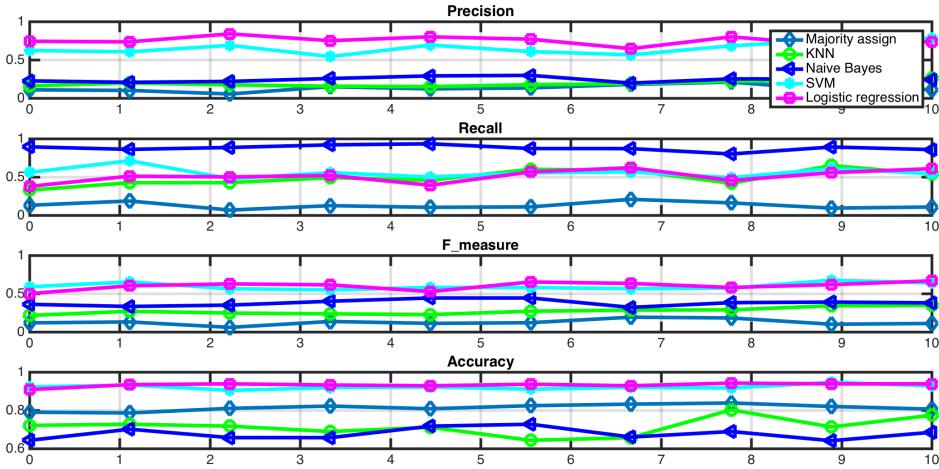


Figure 4.2: Classification methods comparison.

Table 4.1: Classification methods comparison.

	Precision	Recall	F_ measure	Accuracy
Majority assign	0.1274	0.1289	0.1258	0.8136
KNN	0.1906	0.4913	0.2723	0.7154
Naive Bayes	0.2432	0.8779	0.3798	0.6777
Weighted SVM	0.6543	0.5565	0.5966	0.9218
Logisitic Regression	0.7513	0.5102	0.6018	0.9322

4.4 Climate Motivated Protests

There were a total of 25352 recorded civil unrest events in Latin American countries from July 2011 to March 2015 that were included in our dataset. Using our climate protest classifier, we were able to separate out protests directly or indirectly resulting from a major climatic, severe weather, or environmental event. In the subsequent analysis, these three categories of event types are labeled with a common definition of “climate event”. Of the candidate civil unrest events, 991 (3.9%) events are classified as climate-motivated across all Latin American countries for that time period. In the subsequent sections, we conduct a multi-dimensional analysis of these protests to understand potential implications of the breadth of impact resulting from climate motivated protests.

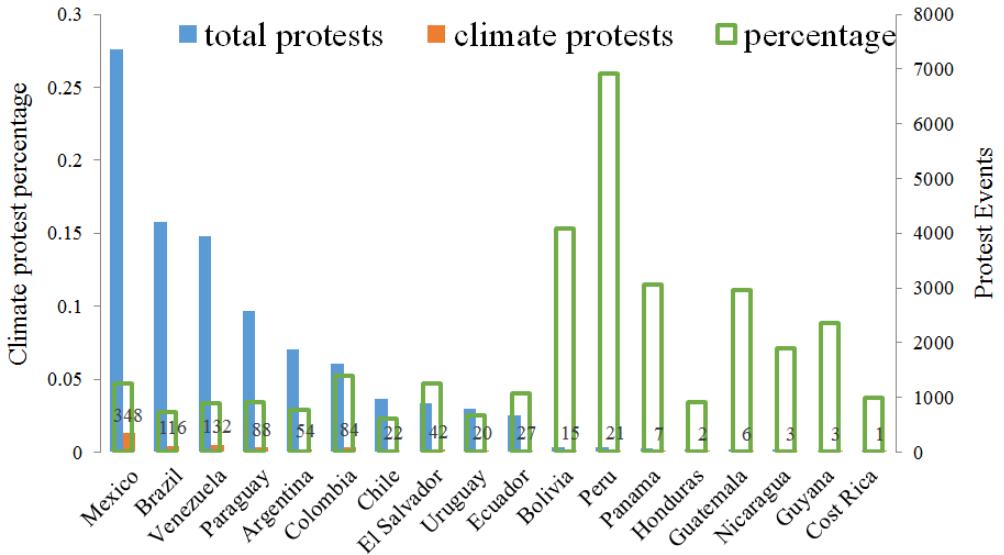


Figure 4.3: Blue bar shows all the GSR protest events, yellow bar shows climate related protest events, green area shows the climate protest percentage over all the Latin American countries, from July 2012 to March 2015.

4.4.1 Frequency Analysis by Country

The first analysis we conduct is a comparison of the representative number of protests within and across each country. The results of our classifier selection show the total number of protests and the percentage of those that are climate motivated in Figure 4.3. The country with the most protests overall is Mexico, and Costa Rica has the least. A similar trend is also seen in terms of climate motivated protests. As evidenced by the climate to non-climate protest ratio, the portion of protests related to climate remains fairly constant across countries with the exception of Peru. In this particular case, there were numerous protests centered on mining and its effect on the environment that dominate the overall protest landscape. As the number of total protests decrease, we see more variability in the ratio as expected. For these countries, which typically have smaller populations, the significance of a single type of protest has more of an impact on the measure than larger countries.

To show the effect of the population on the number of climate protests, we plot the result of a linear regression in Figure 4.5. The result of this shows an $R^2 = 0.64$, showing a slight linear relationship. However, the interesting part of this analysis lies in the residual errors. The set of countries including Mexico, Venezuela, Paraguay, and Colombia, all demonstrate the occurrence of more climate protests than would be expected given the entire dataset. On the contrary, Brazil has fewer climate protests given the size of their population. There could be a number of reasons for these findings such as socio-political stability, environmental sensitivity, and the type of climate events. All of these are potential avenues for further causal or anecdotal studies. In the following, however, we choose Brazil, Mexico, and Venezuela for

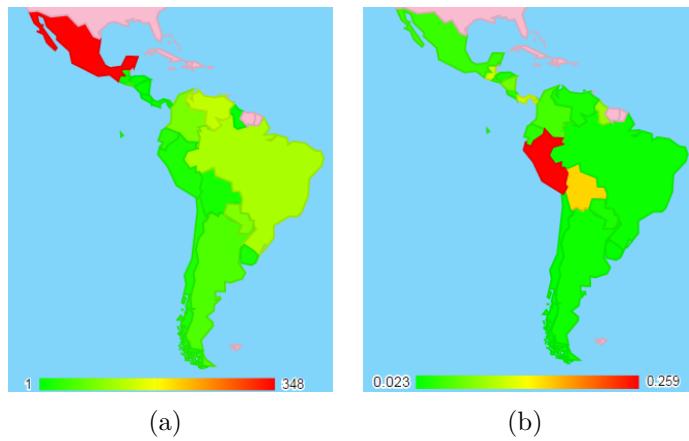


Figure 4.4: (a) Climate related protests events numbers; (b) Climate related protests percentage in Latin American countries, from July 2012 to March 2015.

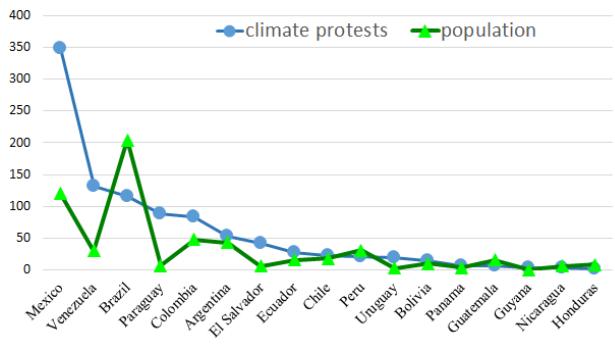


Figure 4.5: Climate protest events and population (million) of each country. The two series have a Pearson correlation coefficient 0.64.

further analysis into overall trends of climate protests, and how these are shaped in the data recovered by the classifier.

We investigate the protest event time series in South American. As shown in Figure 4.6, on average, February, June, July, and August see the most climate-related protest events. We can also see the for the none-climate protest, the temporal distribution is different since it see most protests in March.

4.4.2 Spatial Distribution of Climate Motivated Protests

In this manuscript we are defining the climate protest as being different from a regular civil-unrest event by a relation to an climate event. Next, we investigate if there is any fundamental difference in terms of where these protests occur in relation to protests in general. For this analysis we use Mexico, Brazil, and Venezuela which all have many protest

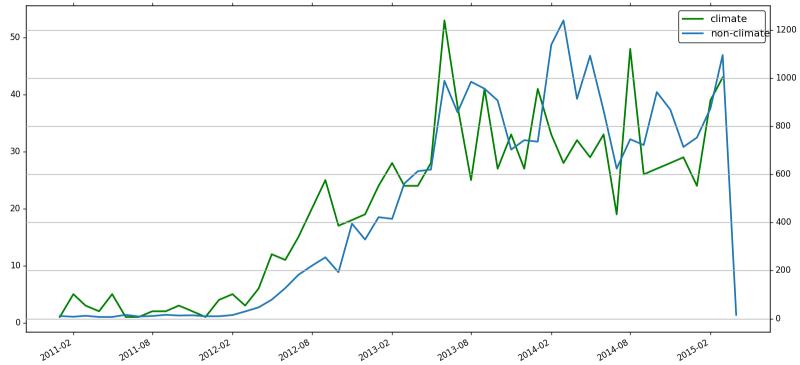


Figure 4.6: climate protest and non-climate protest time series from 2011 to March 2015.

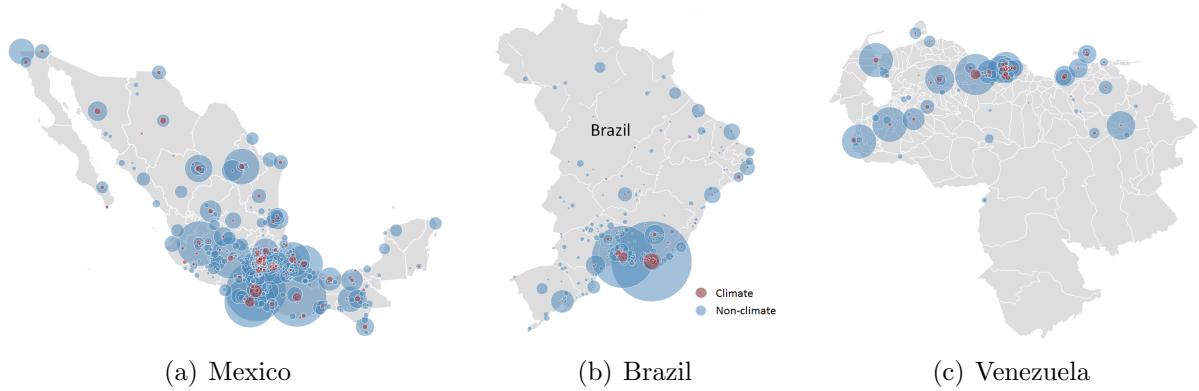


Figure 4.7: Climate and non-climate protests from July 2012 to March, 2015. Red circle represents climate related protest events, and blue circle represents non-climate related protests.

events, and the percentage of those that are related to the climate are all at about 4%. The spatial distribution of events is shown in Figure 4.7. Both the total number of protests and those that are climate motivated are shown and represented by the size of the blue and red shaded circles, respectively.

In both Brazil and Venezuela, many of the protests appear at or near their coastal boundaries, and Mexico has more inland activity. However, we have already established a connection between population and protests. This is no different for the spatial distribution, where much of the population of Brazil and Venezuela is located in coastal regions. The protests in Brazil mainly center at two major cities Sao Paulo and Rio de Janeiro. In Mexico and Venezuela climate protests have a more uniform distribution across the cities. Therefore, there is no particularly strong evidence to suggest that certain regions of these countries are

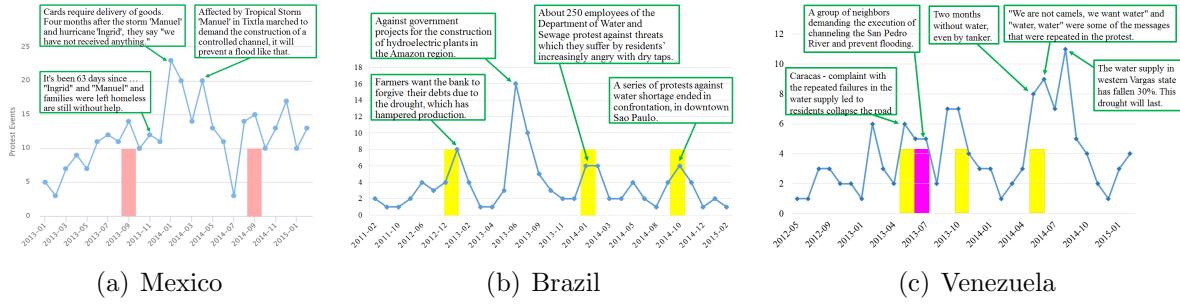


Figure 4.8: (a) Mexico climate disasters and climate protests. The blue time series shows the climate related protest events, and light red vertical lines show two storm diasters in Mexico, storm Manuel in September 17, 2013 and hurricane Odile in September 15, 2014 respectively. (b)Brazil climate disasters and climate protests. The blue time series shows the climate related protest events, and yellow vertical lines show three drought diasters in Brazil, drought in Feb 2012, Heat wave in Feb 2014, and drought in Oct 2014, respectively. (c) Venezuela climate disasters and climate protests. The blue time series shows the climate related protest events, and rose vertical line show local area flood diaster, and yellow vertical lines drought disasters.

more prone to protest with respect to the climate than they would normally be willing to protest in general. In terms of the climate events defined in this study, effects of climate, the environment, and extreme weather are not regionally exclusive to certain populations. Through complex channels such as food supply, the effects of climate impact can ripple across spatial networks.

4.4.3 Temporal Dependency on Climate Events

The temporal dependency of climate protest occurrences is analyzed for each country. As with the spatial domain, the effects of climate events are non-local in time in some cases. The ground truth for the events was established for extreme weather only, as the event itself is more local in time than climate and environmental changes. This data is available by combining the following sources: International Disaster Database EMDAT¹, World Disasters Timeline² and European Commission’s Humanitarian Aid and Civil Protection department (ECHO)³. The official climate disaster report for each country is shown with climate related protests in Figure 4.8.

¹<http://www.emdat.be/database>

²<http://www.mapreport.com/>

³<http://ec.europa.eu/echo/>

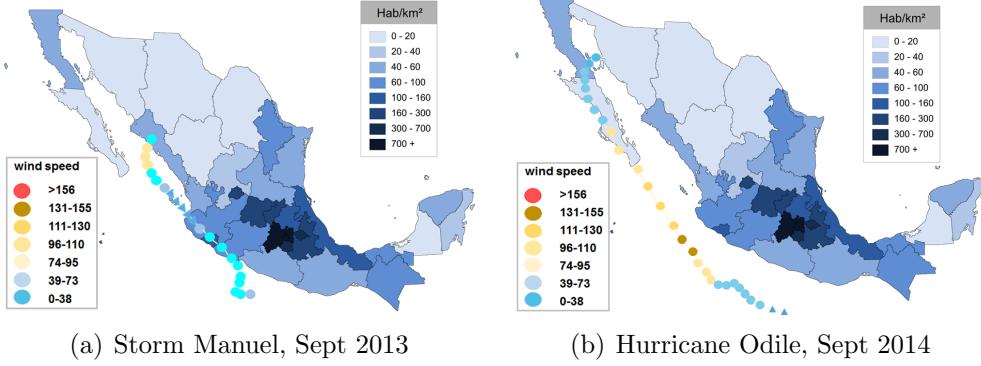


Figure 4.9: The map shows population density of all Mexico's 32 states. The track shows Tropical Storm Manuel of 2013 and Hurricane Odile of 2014, points in different color represent the wind speeds.

Mexico climate disasters Figure 4.8(a) shows the Mexican extreme weather events and protests where the blue time series represents the climate protests events, and the two red bars shows the occurrence of two storms. The first storm is the combined tropical storm Manuel (category 1) and hurricane Ingrid in September 17, 2013. The track maps can be seen in Figure 4.9(a). Tropical storm Manuel crossed the west coast of Mexico and resulted in more than 23,000 people fleeing their homes due to heavy rains spawned by what had been Hurricane Ingrid. Of those displaced 9,000 went to emergency shelters. In terms of infrastructure, at least 20 highways and 12 bridges had been damaged⁴. After the storm, related protests and other civil unrest events broke out and lasted for more than 17 months because the government's response had been inadequate. The storm related protests reached a climax in January 2014, and second climax in April 2014. On November 19, 2013, there was report saying “it’s been 63 days since the onslaught of ‘Ingrid’ and ‘Manuel’ and families were left homeless are still without help”⁵ Four months after the storm Manuel and the effects of Hurricane Ingrid, they say “we have not received anything”. On April 7, protest descriptions said “Affected by Tropical Storm ‘Manuel’ in the municipal head of Tixtla marched to demand the construction of a controlled channel, it will prevent a flood like that caused the overflow from the Black Lagoon in September 2013”. The last protest event we have on record from the climate protest classifier occurred 17 months after the original event. This demonstrates that the residual capacity of these events to impact the livelihoods of people is not guaranteed to be local in time. As we show, the range of impact can extend even beyond the occurrence of other storms.

In Figure 4.8(a), the second red bar shows hurricane Odile. It is a category 3 storm that occurred in 2014, and the track of the storm’s path is shown in Figure 4.9(b). Despite hurricane Odile being a more intense storm, there were not many protests related to the

⁴<https://weather.com/storms/hurricane/news/tropical-storm-manuel-hurricane-ingrid-hit-mexico-opposite-coasts-20130916>

⁵Quotes are translated from the native language of the country.



Figure 4.10: Word cloud of Mexico storm Manuel, Sept 13, 2013.

event. Comparing the storm's paths in Figure 4.9, Tropical Storm Manuel hit Mexico's mainland, which caused more destruction. Hurricane Odile 2014 had less of an impact on the Mexican mainland, even though it crossed the state of Baja California. However, this is the second smallest Mexican state by population. This can explain why storm 2013 lead to tremendous protests, while hurricane 2014 does not.

Brazil climate disasters Figure 4.8(b) shows the relationship between protests classified by our algorithm and actual extreme weather events in Brazil. The three yellow bars show three separate drought events in Brazil, which resulted in drought related protests almost immediately. The drought in February 2012 hampered production, which caused farmers to protest. The heat wave in February in 2014, and drought in October 2014 resulted in water shortages, causing civil unrest. The biggest spike in June 2013 described protests against government's projects for the construction of hydroelectric plants in the Amazon region⁶ and is more of an environmental impact type of event. In general, for these events we see predominantly local relationships in time between the protest and the preceding event. For Brazil in particular, the extreme weather event matches fairly well with the onset of drought.

Venezuela climate disasters In Figure 4.8(c), the climate motivated events are shown in relation to relevant extreme weather events for Venezuela. The pink bar represents sudden onslaught of rain in June 2013 that caused a heightened risk of flooding and landslides in the densely populated communities on the outskirts of Caracas. It triggered a small portion of protests to prevent flooding. The yellow bars denote drought disasters. The drought in May 2014 triggered rationing of tap water in the capital, Caracas, where residents formed lines lasting hours to fill jugs of water⁷. This drought disaster lasted so long that related protests reached a climax in September 2014. Unlike Brazil, the data in Venezuela on droughts proved tough to ascribe to a particular drought event. They occur rather frequently and

⁶<http://www.bloomberg.com/news/articles/2013-06-05/protests-over-brazil-hydropower-leads-to-delays-and-boosts-costs>

⁷<http://www.breitbart.com/national-security/2014/05/31/severe-scarcity-prompts-venezuelan-government-to-ration-water/>

there is a substantial amount of overlap in the residual protest events that it was difficult to distinguish to which it was referring.

4.5 Climate protests causality

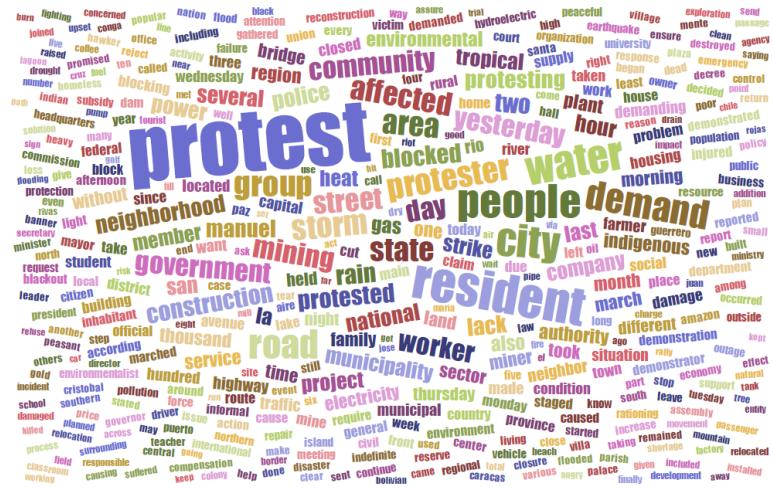


Figure 4.11: Word cloud of all the climate related protests, from GSR descriptions.

4.5.1 Word cloud

Of the climate related protests, we are interested in what are the protesters demanding. To have a birds view of climate protests, we extract all the climate protest descriptions and plot the word cloud, as shows in Figure 4.11. We can see words like ‘water’, ‘storm’, ‘mining’, ‘rain’, ‘construction’, ‘power’, ‘heat’, ‘gas’, ‘environment’, ‘electricity’, and other weather, environment related keywords are dominant, which gives us a general idea of what protesters are demanding.

4.5.2 Analysis of Protest Descriptions

As stated previously, we are not blind to the realization that the causes of climate motivated protests are in general complex. In the following, we analyze the descriptions of the protest events in order to gain insight into the general pathways by which protests within our corpus have occurred. Shown in Figure 4.12 is a weighted Sankey diagram showing the bipartite graph of the most common keywords in the descriptions of protests from each country. Apparently, many of the protests identified by the classifier in one way or another have something to do with lack of water followed by climateal effects in general. Other prominent

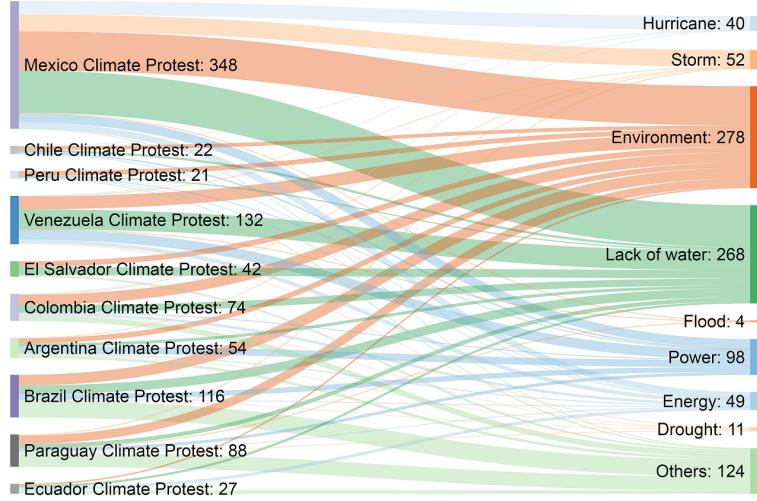


Figure 4.12: Climate protest causality diagram. Left bar shows ten countries' climate protest numbers, and right bar shows nine climate event categories which cause climate protests.

keywords include mentions of power and energy issues. Each country also exhibits its own protest keyword categories. In Mexico, the most notable protest keywords involve lack of water, environmental concern, storm and hurricane. In Venezuela, apart from lack of water and environment problems, the dominant keywords are blackout and energy issues. In Peru, more than half of climate protests are about a mining project, which is an environment concern. While in Argentina, 35% events protest against blackout issues. We expand on these observations in the following where we analyze several dimensions of the keywords to extract details about pathways to protest.

4.5.3 Pathways to Climate Motivated Protest

For some severe and dominant climate events, such as storm, hurricane, flood, and drought events, we employ the knowledge graph to represent the link relationships between entities. By matching the object or subject with climate related keywords, and predicate to be causality relationship like result of, cause by, lead to, blamed, accused of, demanding, against, request, we can locate and further identify the causality of most protest descriptors.

Figure 4.13(a) shows the storm caused protests demands in Mexico, which generally falls into four categories: supply, home, government and reconstruction. In the supply related protest, the causality includes but not limited to: lack of drinking water, lack of good support, and power outage. The second category is about home, they protest either because of lost homes, or request to relocate to avoid storm, or request to reconstruct homes. Another protest type targets at government, they either fight because government did not take action, or blame government's indifference to damages, or request finance compensation to the damages. In

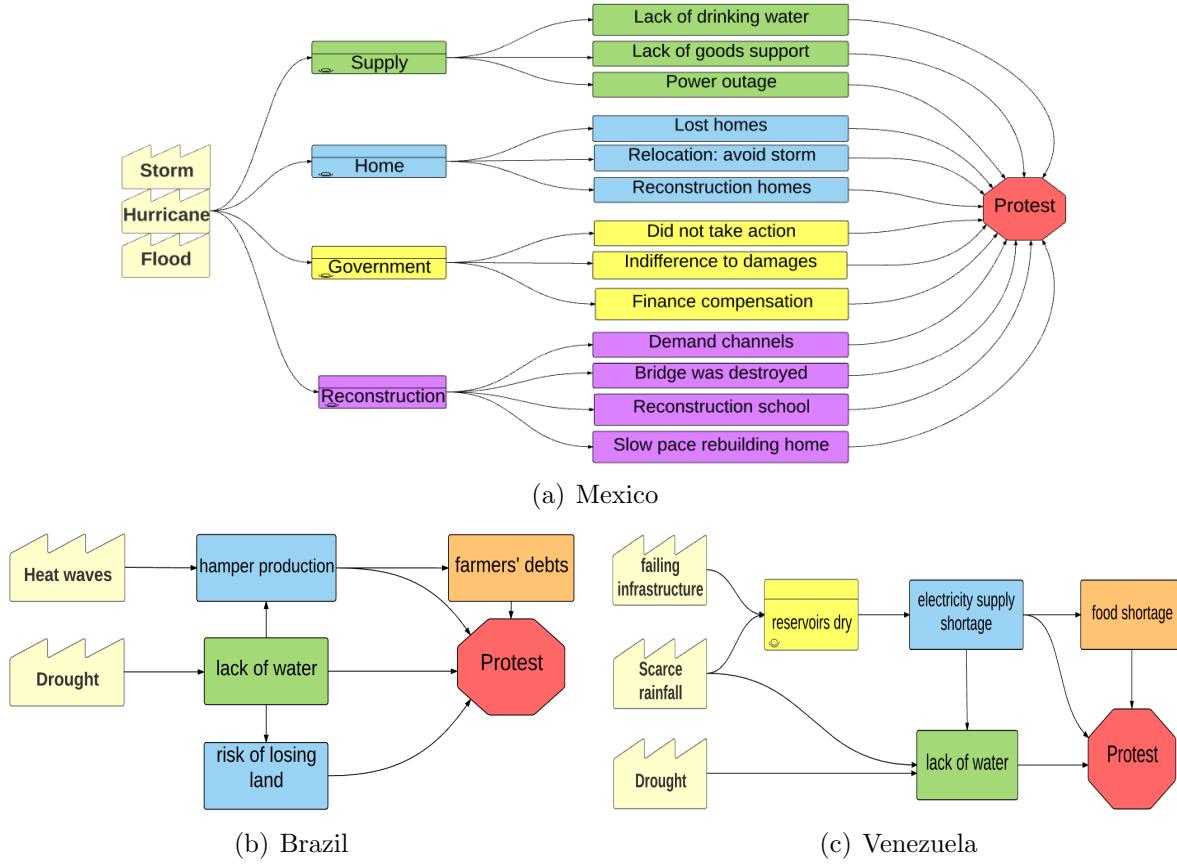


Figure 4.13: Climate protest causality diagram

the reconstruction category, residents demand reconstruct channels to avoid more storms, request to reconstruct bridges, roads, schools, or unsatisfied with the slow pace of rebuilding homes. Figure 4.13(b) describes the causality of Brazil climate related protests. One line is heat wave hampered production, which cause farmers' protests, the other line is drought causing residents lack of drinking water thus lead to protest, and the third line is lack of water causing farmers facing the risk of losing land, which result in protest. In Venezuela, the protests are more water-electricity centralized, as shown in Figure 4.14(c). Scarce rainfall, drought plus failing infrastructure, which makes water shortage and blackout is an everyday fact of life in Venezuela. The electricity shortage deteriorates water shortage, leads to food shortage, and worsens food quality, and so forth. All those situation touches off climate related protests.

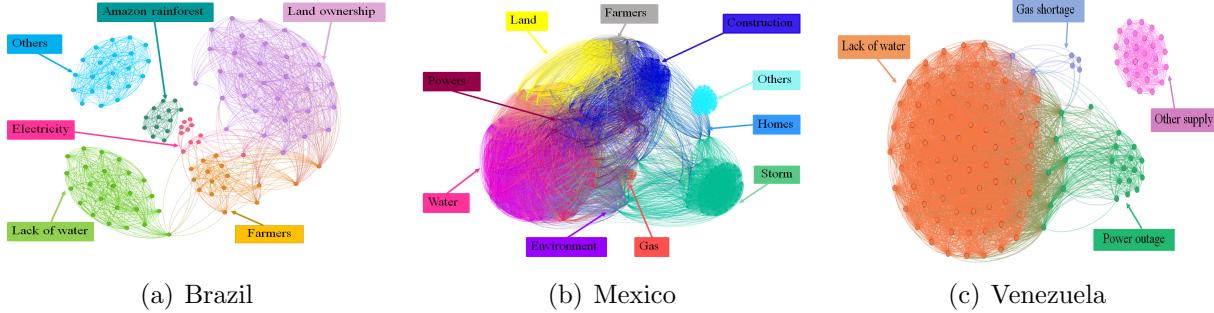


Figure 4.14: Climate protest clustering results

4.6 Climate protest pattern

The above analysis shows the general protest causality, however, we intend to further discover the coherent correlations of protest reasons, hoping to answer questions like is there any protest pattern, or are there some protests associate with others. We treat each protest event as a node and connect two nodes with weight based on their protest description text similarity. Specially, we pay attention to the protest themes or protest demands, if two descriptions have the same protest demanding, their weight will be very high, otherwise, their connection weight tends to be 0. In this way, we build a weighted undirected network $G(V, E, W)$, with each protest as node V , and their connection as edge E , their weight as W . If the weight between two nodes is 0, there will be no edge. We employ Louvain method [6] to split the network into several clusters.

We show in Figure 4.14 the climate protest clustering results that provides the protest proportion and coherent correlations among different protest types. Figure 4.14(a) illustrates Brazil's climate protest pattern, the results shows that in Brazil, the largest protest cluster is about land ownership which accounts for 26.7%, the second cluster lack of water takes up 20.7%, farmers cluster occupies 13.8%, of which, one interesting discover is land and farmers clusters are closely coherent, and lack of water is also closely bind with farmers. Amazon rainforest is another striking protest which is responsible for 11.3%. Figure 4.14(b) shows the protest pattern of Mexico, which has the most climate protest events and complex patterns. We can see the rose red cluster which denotes lack of water is the most dominant protest, accounts for 20.5%, the green cluster represents tropical storm is the second largest protest type, takes up 19.0%, the dark blue cluster construction accounts for 17%, and the yellow cluster land is responsible for 11.8%. We find, water protest is intertwined with environment protest and power protest, land protest is closely related with farmers, while construction cluster is coherent with baby blue cluster which denotes homes (2.6%). Figure 4.14(c) gives the overview clustering results of Venezuela climate protests. We can see, the yellow cluster which represents lack of water protests takes up the largest portion, as high as 55.8%, the green cluster which denotes power outage accounts for second part, 22.1%, and the blue cluster which stands for gas shortage accounts for 5%, the purple cluster shows the rest

climate protest portion, which include food shortage, medicine shortage, water tank robbery behavior, etc.. Clearly, as expected, lack of water protest is intertwined with power outage protest, which corresponds to the fact that lack of water and power shortage is everyday life in Venezuela.

4.7 Climate protests in Twitter

We are also interested in climate events influence on social media, such as Twitter. Using keywords list we are able to filter tweets, then cluster tweets into different partitions based on similarity among tweets using distance function, taking tweets content, geolocation and other features into consideration.

Events Clustering is used to separate events happened at same place or at same time, or the separate different events happened simultaneously on local and entire country. By measure the distance among tweets based on similarity, tweets collection can be clustered into subsets in which tweets are exactly related and similar. Each partition includes similar tweets stand for a specific event. Without events clustering, different events will be mixed. As shown in Figure 4.15(a), Mexico Hurricane were mixed with severe drought happen in Culiacan, with the aid of event clustering, we are able to distinguish those distinct extreme weather events, even though they may happen at the same time. For each event, we plot the related Tweets word cloud besides the flag. Figure 4.15(b) illustrates four drought events in Brazil, on May 2012.

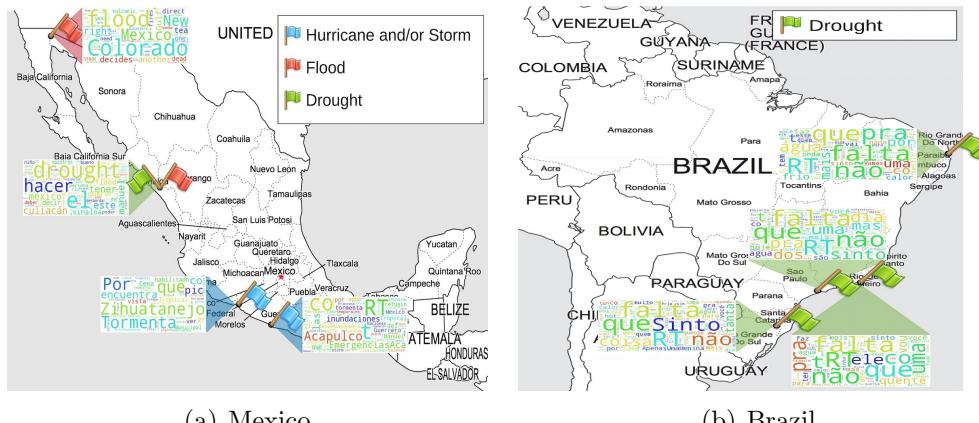


Figure 4.15: Climate protest events in Mexico, Sept 2013 and Brazil, May 2012. Different flag represents different climate disasters. The adjacent word cloud shows Twitter discussion as per that event.

4.8 Discussion

Climate changes, extreme weather and environmental catastrophes can all exert a devastating amount of harm to people around the world. To better understand this process, we show different pathways to protest following severe events in Latin America from 2011 to 2015. Our analysis differs from those previously published in that we consider the breadth of climate protests over a wide spatial and temporal domain. This is accomplished by identifying climate related protests using a logistic regression classifier acting over keyword vectors of protests descriptions in our protest GSR dataset. We found this approach achieved an F-score of 0.60 and accuracy of 0.93, which was the best performing of other common binary classifiers. The results of the classifier indicate a number of broad properties about climate related protests.

From our analysis, we found different climate disasters may cause related protests with different time span, for instance, the Mexico storm Manuel aroused climate related protest as long as 17 months, while in Venezuela, the protests caused by one drought always overlap with the other drought. This paper discloses protest causalities in Latin American countries, illustrate the pathways from climate disasters to climate protests. This paper also identifies the climate related protest patterns, discover the coherent relationship among different protests demanding, such as in Venezuela, the majority protests are against lacking of water, which has high co-occurrence with protests against power outage.

Chapter 5

Distinguish Real Movement from Rumor on Twitter

5.1 Introduction

This chapter was published in the ACM SIGKDD Workshop on Social Network Mining and Analysis (SNAKDD 2013) [42].

Online social networks have become a staging ground for modern movements, with the Arab Spring being the most prominent example. Nine out of ten Egyptians and Tunisians responded to a poll indicating that they used Facebook to organize protests and spread awareness. As a precautionary measure, governments have taken to blocking social networking websites, showcasing the importance of understanding this phenomenon.

Interestingly, the role of social networks is not limited to helping organize the activities of disruptive elements. Many key government and news agencies have also begun to embrace Twitter and other social platforms to disseminate information. After the tragic 2013 explosions at the Boston Marathon, the FBI resorted to online social networks to broadcast crucial information about the suspects. The viral diffusion of information provided them with vital information about the suspects. At the same time it is well known that online activity on sites such as Reddit led to mistaken identification of some individuals and the spread of several rumors.

We were motivated to apply the latest in epidemiological modeling to understand information diffusion on Twitter, in relation to the spread of both news and rumors. Epidemiological models provide a classical approach to study how information diffuses. These models typically divide the total population into several compartments which reflect the status of an individual. For instance, common compartments denote susceptible (S), exposed (E), infected (I), and recovered (R) individuals. Individuals transit from one compartment to another,

with certain probabilities that have to be estimated from data. The simplest model, SI, has two states; susceptible (S) individuals get infected (I) by one of their neighbors and stay infected thereafter. While conceptually easy to understand, it is also unrealistic for practical situations. The SIS model is popular in infectious disease modeling wherein individuals can transition back and forth between susceptible (S) and infected (I) states (e.g., think of allergies and the common cold); this model is often used as the baseline model for more sophisticated approaches. The SIR model enables individuals to recover (R) but is not suited for modeling news cascades on Twitter since there is no intuitive mapping to what ‘recovering’ means. The SEIZ model (susceptible, exposed, infected, skeptic) proposed by Bettencourt et al. [5] takes the interesting approach of introducing an exposed state (E). Individuals in such a state take some time before they begin to believe (I) in a story (i.e., get infected). While the authors of [5] used this approach to model the adoption of Feynman diagrams by communities of physicists, our work explores their use in modeling news and rumors on Twitter.

The key contributions of this chapter are:

- Our work is the first to employ the SEIZ model to model real Twitter datasets. We employ non-linear least squares optimization of the underlying systems of ODEs over tweet data, and demonstrate how this model is better at modeling rumor and news diffusion than the traditional SIS model.
- We analyze eight representative stories (four true events and four rumors) across a range of topics (politics, terrorism, entertainment, and crime) and over several geographic regions (USA, Mexico, Venezuela, Cuba, Vatican). While not an exhaustive list, this demonstrates the wide applicability of the proposed model.
- We demonstrate the capability of the SEIZ model to quantify compartment transition dynamics. We showcase how such information could facilitate the development of screening criteria for distinguishing rumors from real news happenings on Twitter.

5.2 Related work

Related work falls in three categories.

Information Diffusion Significant work has gone into research on information diffusion on social media, e.g., see [15, 51, 71, 97]. Recently, Matsubara etc. [56] conducted research on the rise and fall patterns of information diffusion, and managed to capture the power-law fall pattern and periodicities inherent in such data. Gomez-Rodriguez et al. [28] built a cascade transmission model to track cascading process taking place over a network; they

traced overall blogs and news for a one-year period and found that the top 1000 media sites and blogs tend to have a core-periphery structure.

Epidemiological models Mathematical modeling of disease spread not only provides vital information about the propagation of the disease in a human network, but also offers insight into the strategies that can be used to control them. The classification of the human population into different groups forms the basic premise of using epidemiological models for modeling information diffusion. The two widely used such models are SIR (Susceptible, Infected, Recovered) and SIS (Susceptible, Infected, Susceptible) models. Newman et al. [63] showed that a large class of standard epidemiological models, viz. the SIR models, can be solved exactly on a wide variety of networks, and confirmed the correctness of solutions with numerical simulations of SIR epidemics on networks. Kimura et al. [49] proposed the application of the SIS model to study information diffusion where the nodes can be activated multiple times. Zhao et al. [102] proposed an SIHR (Spreaders, Ignorants, Hibernators, Removed) rumor spreading model, with forgetting and remembering mechanisms to simulate rumor spreading in inhomogeneous networks. Xiong et al. [95] proposed a diffusion model with four different states: susceptible, contacted, infected, and refractory (SCIR) to identify the threshold value of the spreading rate approaches almost zero. Bettencourt et al. [5] proposed the SEIZ (susceptible, exposed, infected, skeptic) model to capture the adoption of Feynman diagrams by using the publication counts after World War II. They extract the general features for idea spreading and estimate the idea adoption process. Their result showed that the SEIZ model can fit the long term idea adoption process with reasonable error, but does not demonstrate whether this model can be applied on large scale datasets, or whether can be applied on Twitter, where the story unfolds in real-time.

Rumor modeling As far as we know, Daley [22] first proposed the similarity between epidemics and rumors using mathematical analysis. Some researchers have studied rumor propagation modeling in different network topologies [62, 100]; however, they do not provide any discussion of propagation differences between news and rumors. Shah et al. [81] detect rumor sources in network using maximum likelihood modeling. In [8], Budak et al. prove that minimizing the spread of the misinformation (i.e., rumors) in social networks is an NP-hard problem and also provide a greedy approximate solution. Castillo et al. [14] delve into twitter content modeling, such as sentiment analysis and hashtags to identify rumors, while Qazvinian et al. [68] try to address this issue using broader linguistic methods, to learn possible features of rumor and determine whether a twitter user believes a rumor or not. More related work appears in [39, 88]. Our goal is to develop an understanding of these processes using diffusion models.

Table 5.1: Twitter datasets studied in this chapter.

No.	Dataset	Date	Area	Type	Country	#Tweets	Response ratio	Keywords & Hashtag
1	Boston	04-15-2013	terrorism	news	USA	501259	68.3%	Marathon, (#)bostonmarathon
2	Pope	02-11-2013	religion	news	Vatican	31365	56.75%	Pope, (#)Benedict
3	Amuay	08-25-2012	accident	news	Venezuela	49015	62.89%	Amuay, refinery, explosion
4	Michelle	02-24-2013	entertainment	news	USA	3762	54.45%	Michelle Obama, Oscars
5	Obama	04-23-2013	politics	rumor	USA	791	46.14%	White House, explosions
6	Doomsday	12-21-2012	mythology	rumor	Global	11833	52.19%	Doomsday, Mayan, doom
7	Castro	10-16-2012	politics	rumor	Cuba	3862	54.45%	Fidel Castro, Dr. Marquina
8	Riot	09-05-2012	crime	rumor	Mexico	4631	47.17%	Antorcha Campesina, Nezahualcoyotl

5.3 Datasets

We focus on twitter datasets that have reliable coverage of the events being studied; the volume of tweets ranges from as low as 791 to nearly three orders of magnitude greater. As described in Table 5.1, the news and rumors studied were drawn from a variety of regions and across a diversity of topics. Data collection was aimed at gathering tweets highly related to the events under study. We employed customized sets of keywords and hashtags pertaining to each incident. Finally, date range restrictions were used to define relevant tweets for each event. It is also pertinent to note that the tweets analyzed spanned a variety of languages: English, Spanish, Italian, and Portuguese.

5.3.1 News topics

Boston Marathon Bombings. Two pressure cooker bombs exploded near the finish line of 2013 Boston Marathon on April 15, 14:49:12 local time, killing three people and injuring more than 264 others. The FBI released photographs and surveillance videos on online social networks which spread like wildfire and provided crucial leads for identifying the suspects¹.

Pope Resignation. Pope Benedict XVI announced his resignation on the morning of February 11, 2013. In nearly 6 centuries, this was the first time a pope has stepped down from his office. This news received reactions from all across the world².

Amuay Refinery Explosion. Propane and butane gas leakage caused an explosion at the Amuay refinery in Venezuela on August 25, 2012 1:11 am local time. The blast killed 48 people, injured 151 others and damaged 1600 homes³.

Michelle Obama at the 2013 Oscars. In the 2013 Oscar awards ceremony, a big surprise was the appearance of US first lady Michelle Obama for presenting the ‘Best Picture’ award⁴.

¹<http://www.cnn.com/2013/04/15/us/boston-marathon-explosions>

²<http://www.cnn.com/2013/02/11/world/europe/pope-resignation-q-and-a>

³<http://www.cnn.com/2012/08/25/world/americas/venezuela-refinery-blast>

⁴<http://www.mediaite.com/tv/michelle-obama-makes-cameo-at-the-oscars-announces-best-picture-winner/>

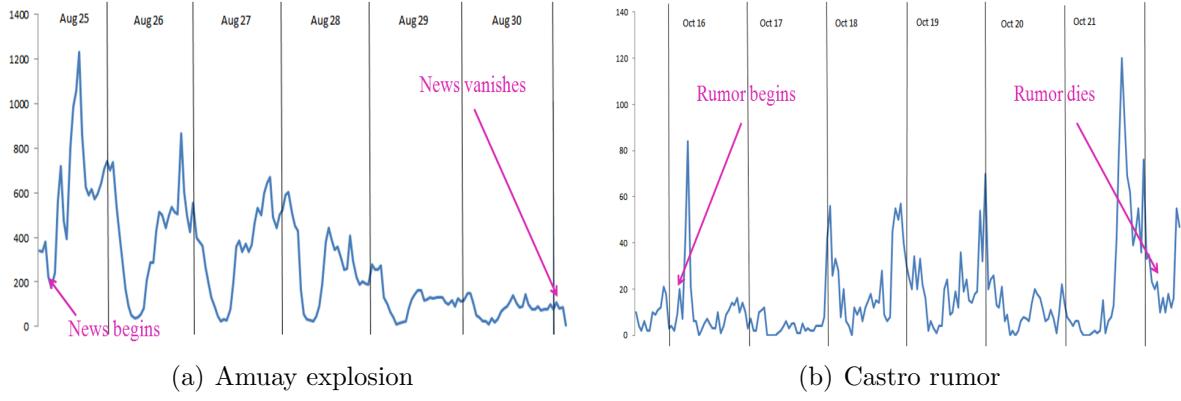


Figure 5.1: Tweet volume.

5.3.2 Rumors

Obama injured. A fake associated press (AP) tweet originated on April 23, 2013 that President Obama was hurt in White House explosions which caused a brief period of instability in financial markets. The information was false and it was determined that the Twitter account was hacked.

Doomsday. December 21, 2012 was rumored to be the Doomsday as it marked the end date of a 5126 year long cycle in the Mesoamerican long count calendar. This rumor spread like wildfire and social networks were flooded with panic and anxiety posts. Considering that we are still alive, Doomsday turned out to be nothing more than a rumor on a massive scale⁵.

Fidel Castro's death. On October 16, 2012 a Naples doctor claimed that former Cuban leader, Fidel Castro suffered a cerebral hemorrhage and is near a neurovegetative state. However, on October 21, 2012, these rumors were denied by Elias Jauva, former Venezuelan vice president, who released pictures of him meeting Castro a few days back⁶.

Riots and shooting in Mexico. A very interesting example that highlights the perils of rumor spreading on social networks pertains to the false reports of violence and impending attack in Nezahualcoyotl, Mexico. (False) rumors spreading on Twitter and Facebook about shootouts caused (real) panic and chaos in Mexico City on September 5, 2012. Interestingly, authorities themselves turned to Twitter to deny these rumors⁷.

⁵<http://en.wikipedia.org/wiki/Doomsday>

⁶<http://www.inquisitr.com/371007/fidel-castro-allegedly-appears-in-public-after-stroke-rumors/>

⁷<http://www.foxnews.com/world/2012/09/08/tweets-false-shootouts-cause-panic-in-mexico-city/>

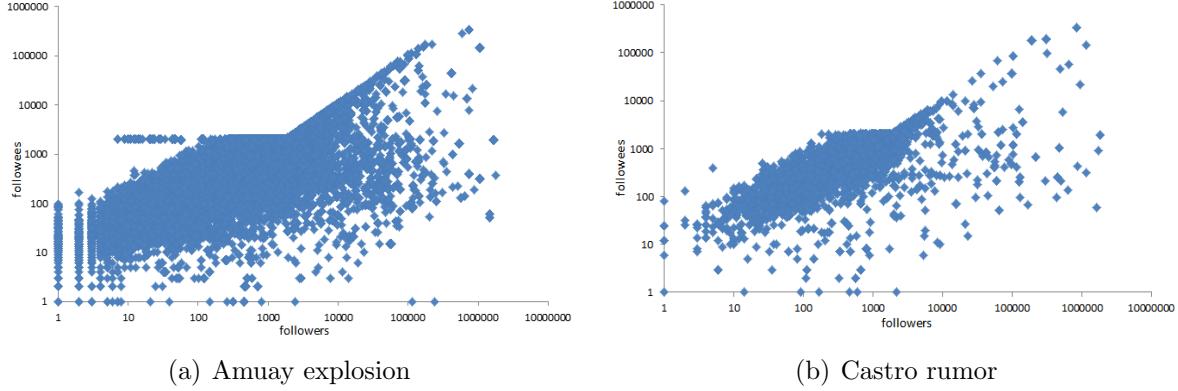


Figure 5.2: Followers/followees distributions. Followers: people who follow the person; Followees: people who are followed by the person.

5.3.3 Preliminary Analysis

We compare the basic properties of news and rumor propagation, by characterizing tweet volume over time, follower/followee distributions, the ‘response ratio’ of a story, and the retweet cascades. In order to maintain brevity, we show results from only two stories in this section: one from our news collection (the Amuay explosion) and one from our rumor collection (Fidel Castro’s purported death).

Tweet Volume. For both examples, we plot the tweet volume over time from the beginning of the story. Figure 5.1(a) shows the activity for the 2012 Amuay refinery explosion example. An activity burst was formed immediately after the news was made public. The number of tweets dropped progressively as the days went by. This activity trend displays attributes similar to breaking news propagation as described by Mendoza et al. [57]. In contrast, Figure 5.1(b) depicts the volume of tweets about a rumor regarding the health of the former Cuban leader Fidel Castro. Here we see occasional spikes of tweet volume; note the increase in tweet volume around October 21st, when the rumors were officially denied.

Followers and Followees Distributions. Figure 5.2(a) is a log-log scatter plot of the followers/followees distribution about the Amuay explosion news, and Figure 5.2(b) is the corresponding plot about Fidel Castro’s death rumor. There is no significant qualitative or quantitative difference in this case; in particular both plots show that the number of followees is less than the number of followers.

Response Ratio. A tweet can either be a post made by the user’s initiative, or a responsive post to some other user’s post (e.g., retweets and replies). As Starbird et al. [84] discuss, retweets reveal how information propagates through a social network: the ‘deeper’ a retweet, the more relevant the tweet is for the community. Based on this idea, we define the response ratio of a story as the fraction of responsive tweets to the total number of tweets in the story. Table 5.1 lists the response ratio for all the 8 stories. As we can see, response ratios for news are higher than that for the rumors.

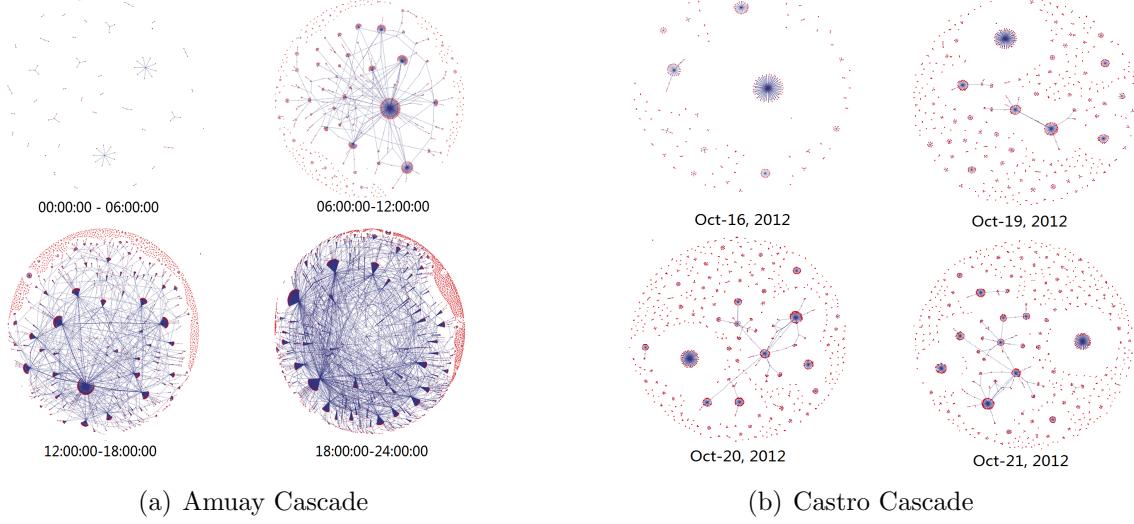


Figure 5.3: Retweet cascade for the Amuay Explosion news and Castro rumor. Each node is a user id, and each edge connects the retweet user to the original user.

Retweet Cascades. A retweet cascade reflects how the social media network propagates information. Figure 5.3 depicts the evolution of the retweet graphs for the Amuay news and Castro rumor dataset. For Amuay news, we plot four graphs with intervals of 6 hours, depicting that a burst has been formed during 6am-12am, only 5 hours after the accident. Fig. 5.3(b) shows the retweet graphs of the rumor for several days. We can see even after one day, there is no burst of tweets related to this rumor. Compared with the network between the news and rumors, we find several features about the rumor. 1) The network for the news instance is more complex and users can obtain news from many sources, while users obtain the rumor information only from limited information centers. 2) There is an immediate burst after a news is made public while there is no obvious burst for the rumors.

5.4 Our Approach

As stated earlier, we used compartmental population models to quantify the propagation of news and rumors on Twitter, focusing primarily on the SIS and SEIZ models.

5.4.1 SIS

As described earlier, this model divides the population into two compartments, or classes: susceptible and infected. Note that in this model, infected individuals return to the susceptible class on recovery because the disease confers no immunity against reinfection.

In order to adapt this model for Twitter, we have given new meaning to these terms. An individual is identified as infected (I) if he posts a tweet about the topic of interest, and susceptible (S) if he has not. A consequence of this interpretation is that an individual posting a tweet is retained to the infected compartment indefinitely; hence, he can not propagate back to the susceptible class as is possible in an epidemiological application. At any given time period t , we use $N(t)$ to denote the total population size, $S(t)$ the susceptible population size, and $I(t)$ the infected population size, such that $N(t) = I(t) + S(t)$. As shown in Figure 5.4, the SIS spreading rule can be summarized as follows:

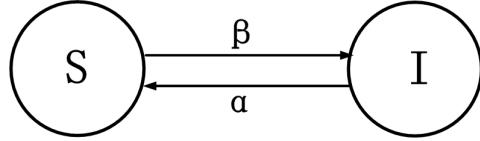


Figure 5.4: SIS model framework

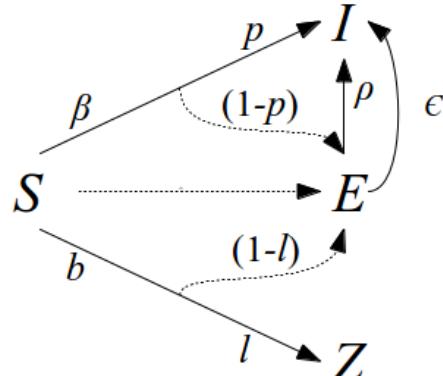


Figure 5.5: SEIZ model framework

- An individual that tweets about a topic is regarded as infected.
- A susceptible person has not tweeted about the topic.
- A susceptible person coming into contact with an infected individual (via a tweet) becomes infected himself, thus immediately posting a tweet.
- Susceptible individuals remain so until coming into contact with an infected person.

The SIS model is mathematically represented by the following system of ordinary differential equations (ODEs) [60]:

$$\frac{d[S]}{dt} = -\beta SI + \alpha I \quad (5.1a)$$

$$\frac{d[I]}{dt} = \beta SI - \alpha I \quad (5.1b)$$

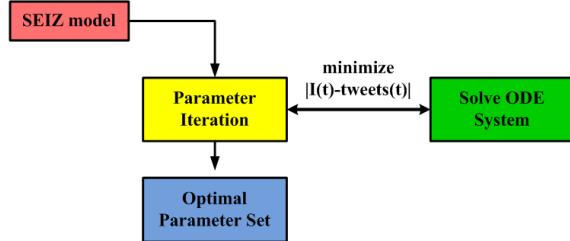


Figure 5.6: Numerical implementation work-flow.

5.4.2 SEIZ

One drawback of the SIS model is that once a susceptible individual gets exposed to disease, he can only directly transition to infected status. In fact, especially on Twitter, this assumption does not work well; people's ideologies are complex and when they are exposed to news or rumors, they may hold different views, take time to adopt an idea, or even be skeptical to some facts. In this situation, they might be persuaded to propagate a story, or commence only after careful consideration themselves. Additionally, it is quite conceivable that an individual can be exposed to a story (i.e. received a tweet), yet never post a tweet themselves.

Based on this reasoning, we considered a more applicable, robust model, the SEIZ model which was first used to study the adoption of Feynman diagrams [5]. In the context of Twitter, the different compartments of the SEIZ model can be viewed as follows: Susceptible (S) represents a user who has not heard about the news yet; infected (I) denotes a user who has tweeted about the news; skeptic (Z) is a user who has heard about the news but chooses not to tweet about it; and exposed (E) represents a user who has received the news via a tweet but has taken some time, an exposure delay, prior to posting. We note that referring to the Z compartment as skeptics is in no way an implication of belief or skepticism of a news story or rumor. We adopt this terminology as this was the nomenclature used by the original authors of the SEIZ model [5].

A major improvement of the SEIZ model over the SIS model is the incorporation of exposure delay. That is, an individual may be exposed to a story, but not instantaneously tweet about it. After a period of time, he may believe it and then be promoted to the infected compartment. Further, it is now possible for an individual in this model to receive a tweet, and not tweet about it themselves. As shown in Figure 5.5, SEIZ rules can be summarized as follows:

- Skeptics recruit from the susceptible compartment with rate b , but these actions may result either in turning the individual into another skeptic (with probability l), or it may have the unintended consequence of sending that person into the exposed (E) compartment with probability $(1 - l)$.
- A susceptible individual will immediately believe a news story or rumor with proba-

Table 5.2: Parameter definitions in SEIZ model[5]

Parameter	Definition
β	S-I contact rate
b	S-Z contact rate
ρ	E-I contact rate
ϵ	Incubation rate
$1/\epsilon$	Average Incubation Time
bl	Effective rate of S ->Z
$\beta\rho$	Effective rate of S ->I
$b(1-l)$	Effective rate of S ->E via contact with Z
$\beta(1-p)$	Effective rate of S ->E via contact with I
l	S->Z Probability given contact with skeptics
$1-l$	S->E Probability given contact with skeptics
p	S->I Probability given contact with adopters
$1-p$	S->E Probability given contact with adopters

bility p , or that person will move to the exposed (E) compartment with probability $(1 - p)$.

- Transitioning of individuals from the exposed compartment to the infected class can be caused by one of two separate mechanisms: (i) an individual in the exposed class has further contact with an infected individual (with contact rate ρ), and this additional contact promotes him to infected; (ii) an individual in the exposed class may become infected purely by self-adoption (with rate ϵ), and not from additional contact with those already infected.

The SEIZ model is mathematically represented by the following system of ODEs. A slight difference of our implementation of this model is that we do not incorporate vital dynamics, which includes the rate at which individuals enter and leave the population N (represented by μ [5]). In epidemiological disease applications, this encompasses the rate at which people become susceptible (e.g. born) and deceased. In our application, a Twitter topic has a net duration not exceeding several days. Thus, the net entrance and exodus of Twitter users over these relatively short time periods is not expected to noticeably impact compartment sizes and our ultimate findings⁸.

⁸<http://www.statisticbrain.com/twitter-statistics/>

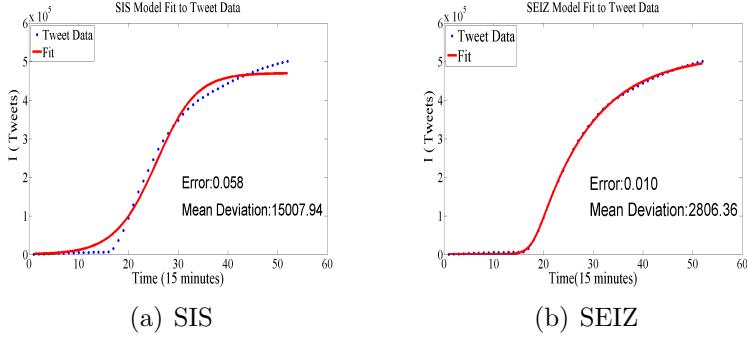


Figure 5.7: Best fit modeling for Boston news.

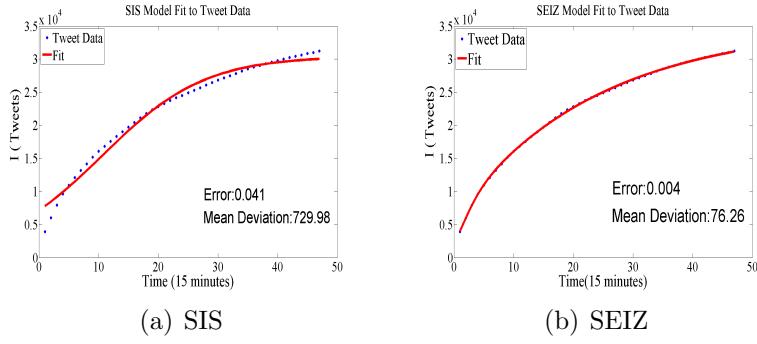


Figure 5.8: Best fit modeling for Pope news.

$$\frac{d[S]}{dt} = -\beta S \frac{I}{N} - bS \frac{Z}{N} \quad (5.2a)$$

$$\frac{d[E]}{dt} = (1-p)\beta S \frac{I}{N} + (1-l)bS \frac{Z}{N} - \rho E \frac{I}{N} - \epsilon E \quad (5.2b)$$

$$\frac{d[I]}{dt} = p\beta S \frac{I}{N} + \rho E \frac{I}{N} + \epsilon E \quad (5.2c)$$

$$\frac{d[Z]}{dt} = lbS \frac{Z}{N} \quad (5.2d)$$

5.4.3 Practical Issues

During our adoption of the SIS and SEIZ models to understand Twitter datasets, we were constrained by several factors. The first constraint was the unknowns in the models. For example, we do not know the transition rates between the compartments nor the initial sizes of the compartments.

Another constraint is the inability to quantify the total population size. This value appears to simply be the total number of Twitter accounts; however the value that we truly want is the number of individuals **who could be exposed to the news or rumor topic**. This

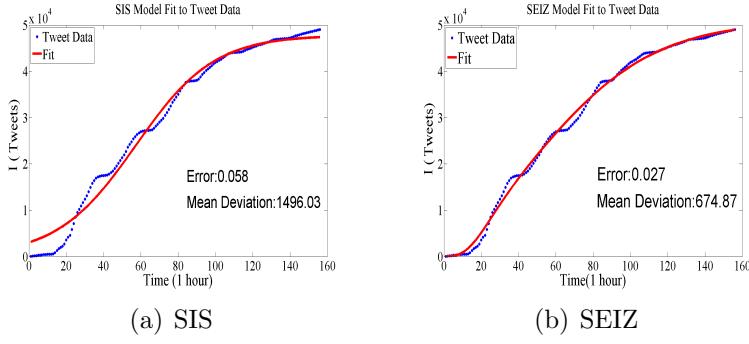


Figure 5.9: Best fit modeling for Amuay news.

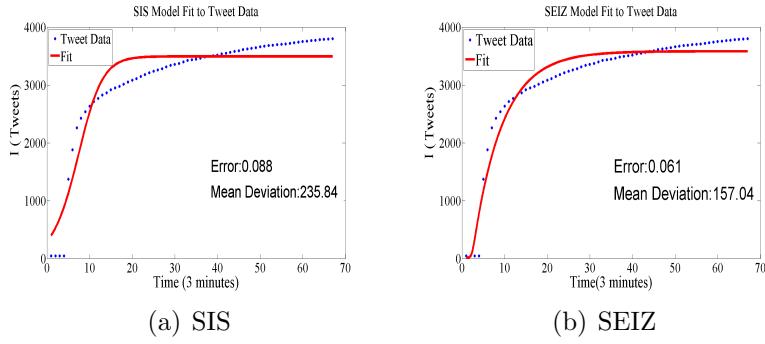


Figure 5.10: Best fit modeling for Michelle news.

value shows to be very different from the total number of Twitter accounts. Consider the ~ 175 million (M) registered Twitter accounts. Of these, (i) ~ 90 million have no followers, and (ii) ~ 56 million follow no one⁹. To further complicate the matter, there exists an abundance of “fake” Twitter accounts, which are never used by any real person. They are simply sold to users wishing to enhance their perceived popularity. Coupling these facts with sporadic Twitter usage due to night-time inactivity and user “unplugging”, it is clear that establishing a reliable estimate of users who could receive a tweet is quite difficult.

Synthesizing all of these factors, we assume the following in our SEIZ model implementation:

1. We do not have reliable population specifics.
 - (a) We do not know N , total population size.
 - (b) We do not know $S(t_0)$, $E(t_0)$, $I(t_0)$, or $Z(t_0)$, the initial values of each population compartment.
 2. **Infected** individuals (I) submit a tweet.
 3. **Skeptics** (Z) have been exposed to story, but do not tweet.

⁹<http://www.businessinsider.com/chart-of-the-day-how-many-users-does-twitter-really-have-2011-3>

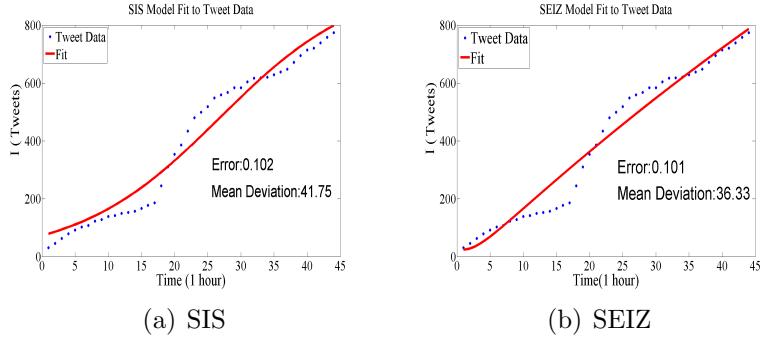


Figure 5.11: Best fit modeling for Obama news.

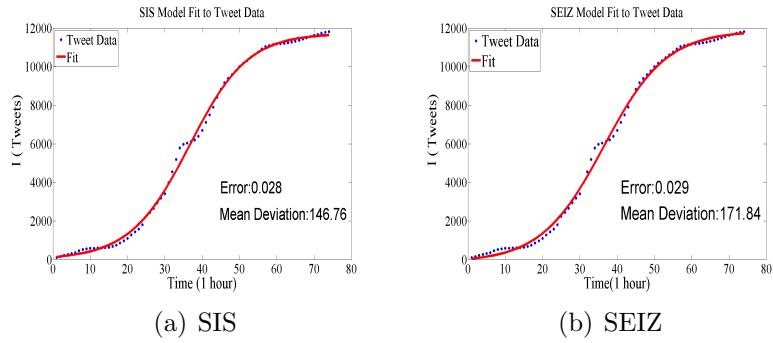


Figure 5.12: Best fit modeling for Doomsday rumor.

4. Vital dynamics do not contribute to the overall population size. Thus, N is a constant.

The implication of these assumptions is that total population size N and initial population sizes for each compartment $S(t_0)$, $E(t_0)$, $I(t_0)$, and $Z(t_0)$ are viewed as unknowns. They are therefore treated as parameters in the parameter fit routine, and fit along with the other model parameters [5].

5.4.4 Parameter Identification

For each of the population models (SIS and SEIZ), represented by equation sets 5.1 and 5.2, we performed a nonlinear least squares fit of the model to Twitter data. As shown in Figure 5.6, each step of this fitting process involved selecting a set of parameter values (rate constants and probabilities in equations 5.1 and 5.2, and initial compartment sizes), and numerically solving the system of ODEs with these parameter values. The set of parameter values that minimized $|I(t) - \text{tweets}(t)|$ was identified as the optimal parameter set.

The experimental implementation was done in Matlab. The **lsqnonlin** function performed the least squares fit. The ODE systems were solved with a forward Euler function that we developed. This algorithm was selected due to its computational efficiency, and used a time-step of no more than 0.05. This threshold demonstrated to be numerically stable; in

Table 5.3: Fitting error of SIS and SEIZ models

	Boston	Pope	Amuay	Michelle	Obama	Doomsday	Castro	Riot	Average
<i>SIS</i>	0.058	0.041	0.058	0.088	0.102	0.028	0.082	0.088	0.068
<i>SEIZ</i>	0.010	0.004	0.027	0.061	0.101	0.029	0.073	0.093	0.050

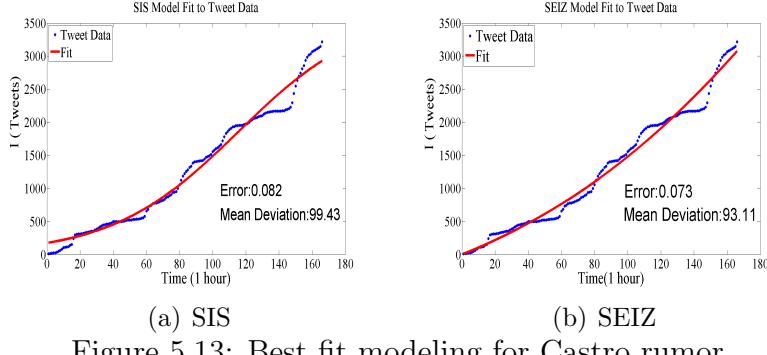


Figure 5.13: Best fit modeling for Castro rumor.

several instances we compared the forward Euler solution to those generated by Matlab's ode45 (5^{th} -order Explicit Runge-Kutta with embedded 4^{th} -order error control), and observed nearly identical solutions.

5.5 Experimental Results

5.5.1 Fitting Results

For each of the Twitter datasets, we were interested in quantifying the transitions of users through the different compartments of the SIS and SEIZ models. Figures 5.7 - 5.14 display the results for the best fit of SIS and SEIZ models (Equations 5.1 and 5.2) to the eight Twitter stories. Also displayed for each figure are the relative error in 2-norm

$$\frac{\|I(t) - \text{tweets}(t)\|_2}{\|\text{tweets}(t)\|_2}$$

and the mean error deviation

$$\frac{\sum_{i=1}^n |I(t_i) - \text{tweets}(t_i)|}{n},$$

where n is the number of data points.

The error metrics for these eight stories clearly indicate that the SEIZ model fits the Twitter data much more accurately than the SIS model. Furthermore, the low relative error of the SEIZ model fit suggests that this model accurately represents the Twitter data for each of the eight stories; see Table 5.3. A common observation about all the eight stories is that

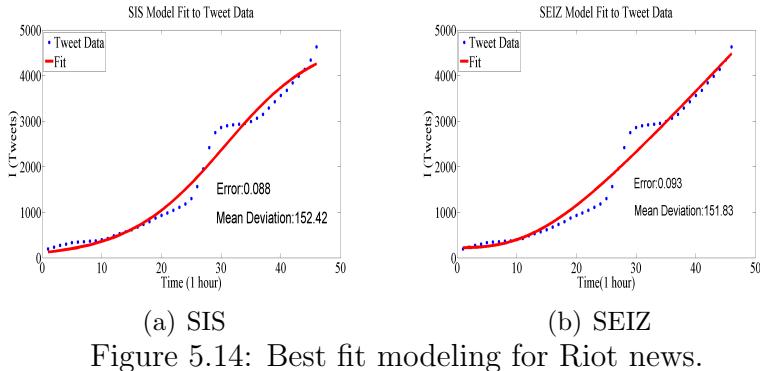


Figure 5.14: Best fit modeling for Riot news.

the SEIZ model is far more accurate in modelling the initial spread of the news on Twitter as compared with the SIS model. This behaviour can be explained by the delay caused by individuals in the “Exposed” category taking some time before posting a story themselves [5].

Given that the SEIZ model is superior to the SIS model in this application, and that the SEIZ model demonstrates an accurate representation of information diffusion on Twitter, a natural question arises “How can this model help us?” The answer is really simple. Since we have a mathematical model for the Twitter data, we can study solutions to some of the constraints as mentioned in the “Practical Issues” section. A well fitted SEIZ model provides values for all contact rates and transition probabilities as defined by Equation 5.2. These parameters empower us to investigate the dynamics of news and rumor spread on Twitter in a fashion that is not possible without a mathematical model. Table 5.2 specifies the SEIZ model parameters that we can now examine to assess news and rumor propagation on Twitter.

5.5.2 Boston Marathon Bombing Analysis

To demonstrate a line of analysis that is now possible with the SEIZ mathematical model, we use quantities from the SEIZ model fit of the Boston Marathon bombing Twitter data (Table 5.2). Results are summarized in Table 5.4.

Here we discuss the dynamics of all 4 compartments, so we specially show all 4 compartments in the SEIZ time-course plot only for Boston Marathon bombing (Figure 5.15(a)). These results suggest that the effective rate of susceptible individuals becoming skeptics is much greater than those that becoming infected. The decrease in $S(t)$ occurs directly with an increase in $Z(t)$, and $S(t)$ becomes stable at the same time that $Z(t)$. $I(t)$ does increase as $S(t)$ decreases, but its rate of change is much slower, and the majority of $I(t)$ increase occurs after $S(t)$ has stabilized to a minimal value, demonstrating that the continued change in the infected compartment has no further influence on the change in the susceptible compartment.

Table 5.4 also demonstrates that the skeptics compartment is more influential on transitioning susceptible users to the exposed class than does infected users. Figure 5.15(a) shows

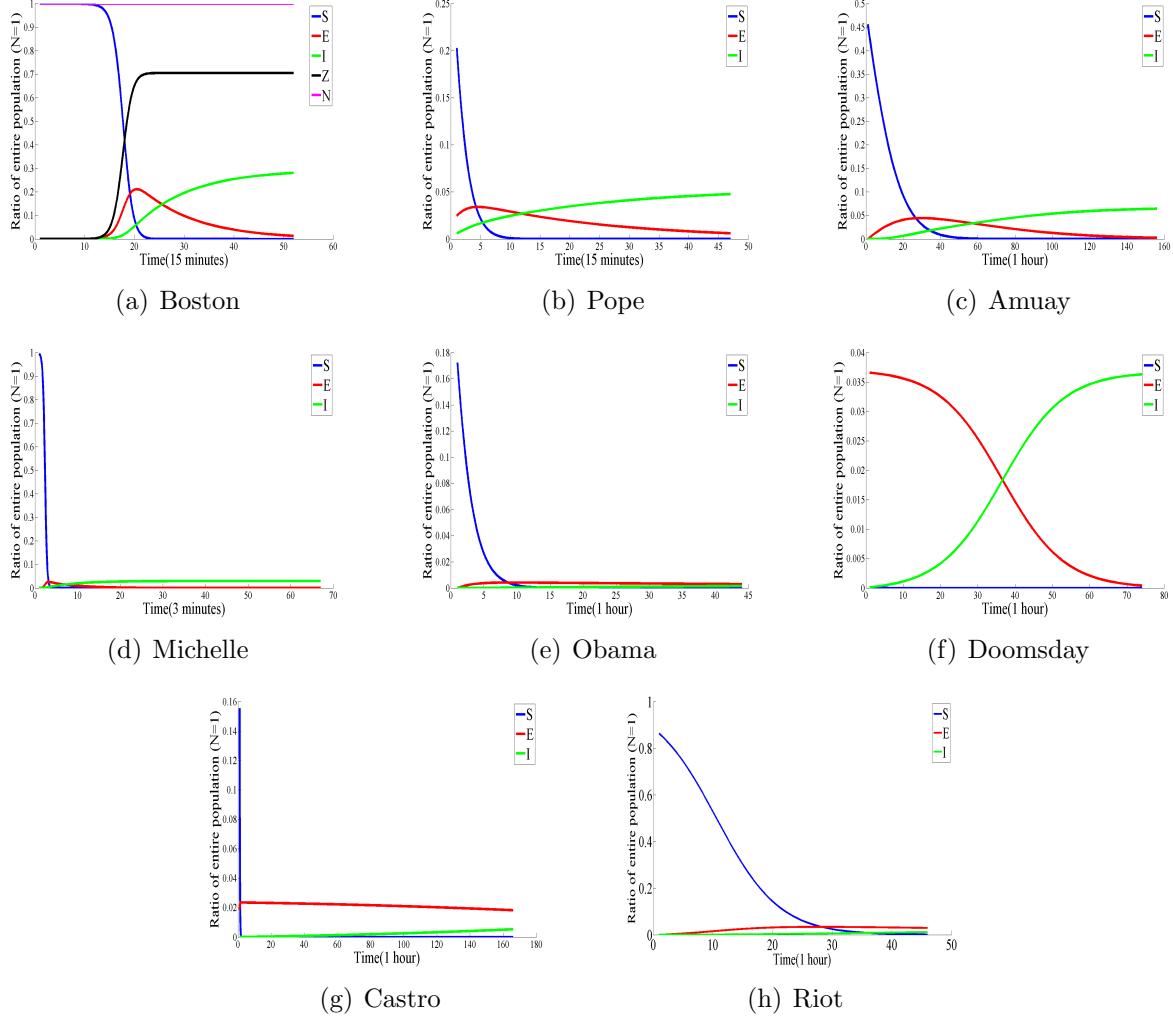


Figure 5.15: SEIZ compartment time-course results.

this as the increase in $E(t)$ is strongly correlated with the increase in $Z(t)$. $E(t)$ also peaks as $Z(t)$ peaks, and $E(t)$ begins to decrease at a rate negative to that of the $I(t)$ increase. In fact, the increase in $I(t)$ directly coincides with a comparable decrease in $E(t)$. These data suggest that the increase in infected users is not due in large part by recruitment of susceptible users, but rather from the natural transition to the infected compartment by exposed individuals.

Putting this all together, we can deduce that virtually all individuals are initially in the susceptible compartment. Most susceptible users become skeptics from interaction with skeptics, and those susceptible users that do transition to the exposed class do so by their interaction with skeptics. The infected compartment increases predominately from the exposed class, and not from direct recruitment of susceptible individuals. Thus, these findings suggest that it was in-fact non-Twitter mediums that most greatly aided in the generation

Table 5.4: Ratios of SEIZ model for Boston dataset.

$\frac{bl}{\beta p}$	3.1E5
$\frac{b(1-l)}{\beta(1-p)}$	1.0E4
$\frac{\epsilon}{\rho}$	7.8

of Twitter propagation! Further, the $\frac{\epsilon}{\rho}$ ratio indicates that the exposed users became infected more so due to information incubation and self-adoption, and not so much from direct contact with infected users.

The remaining instances of SEIZ time-course plots are shown in Figure 5.15, we can see how S, E, and I dynamic change over time. These analyses exemplify the types of analyses that can be used to study Twitter dynamics via the SEIZ population model.

5.5.3 Rumor Detection

We next examined if our implementation of the SEIZ model, applied to our Twitter examples, could be utilized to facilitate the discrimination of true news from rumors. We began by assembling an equation to relate the key parameters of the SEIZ model. In our first attempt at performing this, we restricted our attention to the exposed compartment; this class has direct or indirect interconnections between the other three compartments, and is a key path to the infected compartment. To exemplify this, consider the extreme case where susceptible individuals are attempted to be recruited by skeptics, and ultimately end up in the infected compartment (Figure 5.5). This can only be accomplished by passing through the exposed compartment.

We quantify a ratio through E as the ratio of the sum of the effective transition rates *entering* this compartment (from S) to the sum of the transition rates *exiting* this compartment (to I). We define this ratio as R_{SI} , using the subscripts to denote the contributions from the susceptible and infected compartments in this quantity:

$$R_{SI} = \frac{(1-p)\beta + (1-l)b}{\rho + \epsilon} \quad (5.3)$$

R_{SI} possesses all rate constants and probability values of the SEIZ model and relates them to the exposed compartment with a kind of flux ratio, viz. the ratio of effects entering E to those leaving E. A R_{SI} value greater than 1 implies that the influx into the exposed compartment is greater than the efflux. Similarly, a value less than 1 indicates that members

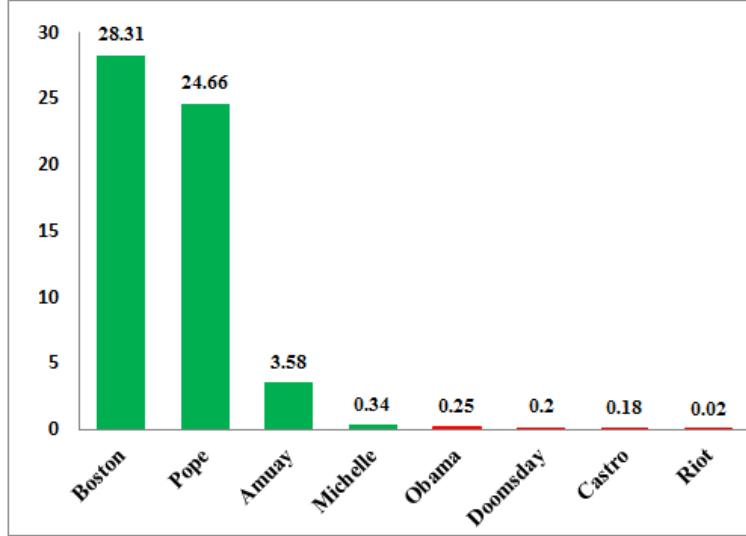


Figure 5.16: Ultimate R_{SI} values for eight Twitter datasets.

are added to the exposed group more slowly than they are removed. We hypothesized that this measure could potentially aid in the distinction of rumor topics from news topics; all parameters of the SEIZ fit are utilized in this measure, and they are related via the R_{SI} value to a key compartment of this model. If a distinction between rumors and true news stories is to be seen with the SEIZ model, we identify the R_{SI} measure to be a probable candidate in aiding this process.

Ultimate R_{SI} value We then computed R_{SI} using the specific parameter values attained from our model fits of the eight cases (Figure 5.16). Here we can see that the true news about the Boston Marathon bombing, Pope resignation, and Amuay refinery explosion do in fact have much higher R_{SI} values than the rumor topics: Doomsday, Fidel Castro death, Mexico City riots, and Barrack Obama injury which each have much lower R_{SI} values. However, the Michelle presence at the Oscars, which is classified as true news, has a very low R_{SI} value. This particular case is interesting since Michelle did not really show up to the 2013 Oscar Awards Ceremony. She simply participated remotely via video telecast. It is thus arguable that this topic could have been discussed in the media in terms similar to rumors.

Dynamic R_{SI} value For one story, if we collect related Tweets at the very beginning stage, then the R_{SI} value would be a time series. Figure 5.17 shows the dynamic R_{SI} values for Castro rumor and Boston bombing real news. We can see for the Castro rumor, at the beginning stage, the R_{SI} value was pretty high, which indicate considerable people believe this story. With time pass by, the R_{SI} values decreased sharply. However, for the Boston Marathon bombing real news, the R_{SI} value was very low at beginning, but increased rapidly

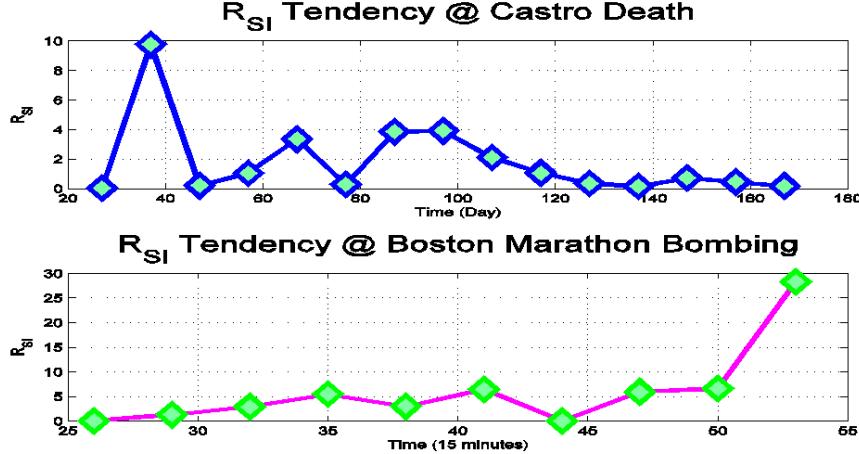


Figure 5.17: Dynamic R_{SI} values for Castro rumor and Boston bombing real news.

within several hours and reached around 28.31 very soon. The R_{SI} time series indicate people's confidence with those stories is dynamically changing.

These findings suggest, for these specific topics, that the parameters in the SEIZ model can potentially aid in the challenge of distinguishing rumor versus true news. We are not claiming that the R_{SI} value is the unique measure to accomplish this, nor are we claiming that the SEIZ model itself is the sole tool to do this. As is suggested by our findings, we postulate that a fit of a compartmental model, in the spirit of the SEIZ model, to Twitter data provides valuable propagation information that can be coupled with other data analysis strategies (e.g., content modeling) to augment the accuracy and reliability of true news story and rumor topic discrimination.

5.6 Conclusion

In this chapter, we have demonstrated how true news and rumor stories being propagated over Twitter can be modelled by epidemiologically-based population models. We have shown that the SEIZ model, in particular, is accurate in capturing the information spread of a variety of news and rumor topics, thereby generating a wealth of valuable parameters to facilitate the analysis of these events. We then demonstrated how these parameters can also be incorporated into a strategy for supporting the identification of Twitter topics as rumor or news. As of now, we are modeling propagation over static data. In future, we plan to adapt this model for capturing news and rumors in real-time.

Chapter 6

Misinformation Propagation on Twitter in the age of Ebola

6.1 Introduction

This chapter was published in IEEE Computer, Volume 47, Issue 12, pages 90-94, Dec 2014 [43].

Mark Twain is credited with saying that a lie can travel halfway around the world while the truth is putting on its shoes. As the Ebola disease rages on in West Africa, the only other epidemic being talked about is the rapid spread of misinformation on social media about the disease, its origins and impact, and response strategies. We sought to characterize the spread of both news and rumors on Twitter about the deadly disease with a view to understanding the prevalence of misinformation.

For context, although Ebola is not a new disease, the current outbreak happening in West Africa is believed to be more than three times worse than all the previous Ebola outbreaks in history combined. The three countries that have the most widespread transmission, viz. Guinea, Liberia, and Sierra Leone, are also those where public health experts fear massive under-reporting due to a variety of social considerations. Even syndromic surveillance strategies, e.g., social media mining and participatory surveillance, are not effective in these countries due to poor penetration of Internet use, and lack of roads and communication infrastructure where Ebola is most prevalent.

Social media has become one of the primary sources by which people learn about worldwide developments so it is instructive to study the spread of Ebola related information on Twitter. Most of the current chatter on Twitter about Ebola reached its peak during late Sep-mid Oct (2014) during which period there have been Ebola-related developments in the US and Europe. (In contrast, Twitter penetration in the three specific West African countries is

low.)

A brief timeline of these developments will help in the discussion that follows. On September 30, 2014, the CDC confirmed the first importation of Ebola into the United States when Thomas Eric Duncan traveled from Liberia to visit family in Dallas. On October 6, in Madrid, Spain, Teresa Romero, a nurse, was reported to be the first person to have contracted the disease outside of West Africa.

On October 8, back in the US, Duncan succumbed to Ebola. A few days later, a healthcare worker at Texas Presbyterian Hospital in Dallas who provided care for Duncan, tested positive for Ebola. On the morning of Oct. 14, a second healthcare worker, who also provided care for Duncan, reported to the hospital with a low-grade fever and was isolated. This healthcare worker also tested positive for Ebola subsequently.

Many states and cities began making contingency plans and issuing travel advisories and guidelines. Lawmakers called for screening passengers and proposed travel bans for Ebola-stricken countries. On October 23, Craig Spencer, a doctor returning from work in Guinea, was rushed to Bellevue Hospital Center with a 100.3 fever and became New York City's first Ebola patient.

6.2 Rumors on Twitter



Figure 6.1: Word clouds constructed from Ebola-related tweets.

The period from end of Sep to mid-late Oct, when Ebola activity happened in the US, is also the period when conspiracy theories, innuendo and rumors began to propagate wildly on Twitter. We gathered tweets during this period and filtered them by either the mention of the keyword ‘ebola’ or relevant hashtags such as #ebola, #EbolaVirus, #EbolaOutbreak, #EbolaWatch, #EbolaEthics, #EbolaChat, #nursesfightebola, #ebolafacts, #StopEbola, #FightingEbola, and #UHCRevolution.

From the gathered tweets, we removed stopwords for further processing. Figure 6.1 depicts word clouds constructed from the tweets for specific days. As can be seen, on 2014-09-29, when there was no Ebola incident in the US, people's attention were primarily focused on

Rumor no.	Content	Label
1	Ebola vaccine only works on white people	White
2	Ebola patients have risen from the dead	Zombie
3	Ebola could be airborne in some cases	Airborne
4	Health officials might inject Ebola patients with lethal substances	Inject
5	There will be no 2016 election and complete anarchy	Vote
6	The US government owns a patent on the Ebola virus	Patent
7	Terrorists will purposely contract Ebola and spread it around	Terrorist
8	The new iPhone 6 is infecting people with Ebola	iPhone
9	There is a suspected Ebola case in Kansas City	Kansas
10	Ebola has been detected in hair extensions	Hair

Figure 6.2: Top 10 Ebola-related rumors (by volume; from 09/28/2014 to 10/18/2014).

Liberia and other African countries. On 2014-09-30, after CDC confirmed that Mr. Duncan in Dallas tested positive for Ebola, related keywords rose to the fore.

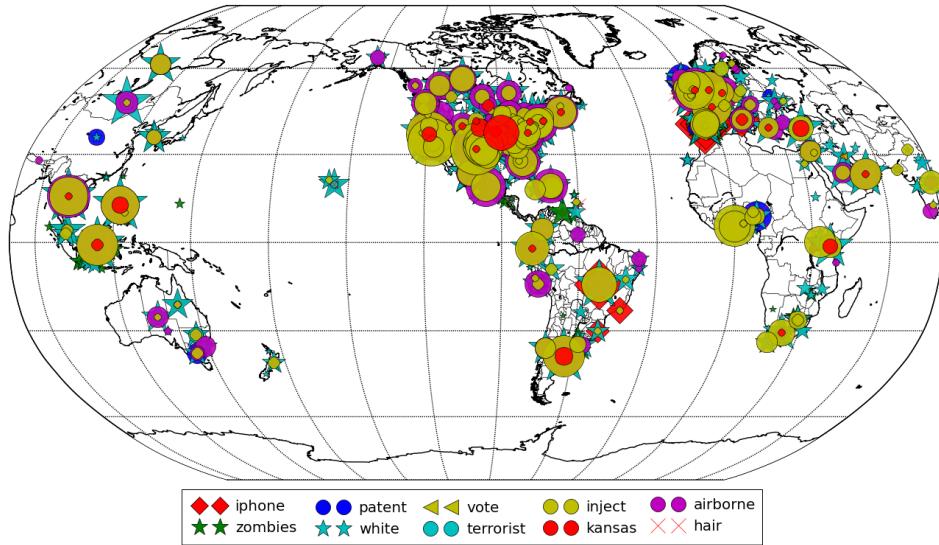


Figure 6.3: Distribution of top-10 rumors obtained from geolocated tweets. Data from 10/08/2014 is used for this plot. Icon size is proportional to the logarithm of the tweet volume.

Next, we studied information cascades in our collection of tweets with a view toward identifying misinformation and spread of falsehoods. We identified several widespread rumors circulating on Twitter, the top 10 of which are shown in Figure 6.2. (We focus on rumors in English only for our study.) Most are self-explanatory as to their intent and interpretation.

Two other rumors are note worthy. The ‘snake’ rumor (which originated at least as early as late summer 2014) asserts that Ebola came across the border from Guinea to Sierra Leone

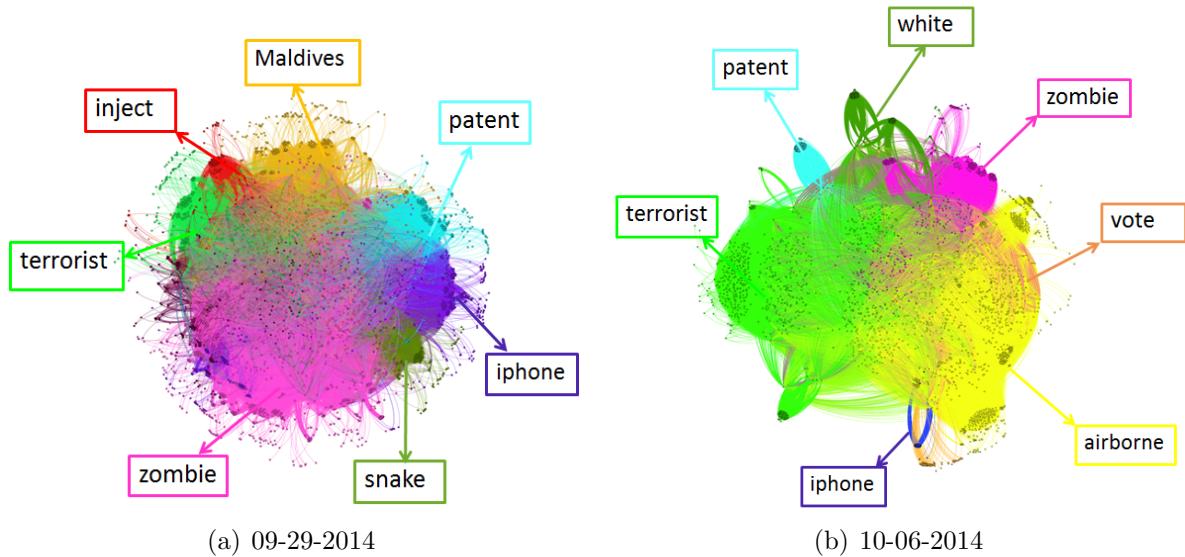


Figure 6.4: How rumors cluster: (a) 09/29/2014. (b) 10/06/2014. Rumors are color coded consistently across the two frames.

via a snake in a bag. As stated in [4], “a lady had a snake in a bag. When somebody opened the bag, that made the lady die.” The Maldives rumor pertains to an uncorroborated report that Ebola patients have been reported (and quarantined) in the Maldives.

For each of these rumors, we geocoded tweets participating in the spread of such rumors with a view to understanding their geographical scope. As Figure 6.3 shows, the “airborne” and “inject” rumors were most prevalent in the US with specific other rumors (e.g., “patent”) being prevalent in other parts of the world.

Next, we employed a dynamic query expansion model [101] to study the rumors in greater detail. The DQE model begins with a seed set of keywords (e.g., “ebola”, “rumor”), identifies tweets that mention these keywords, and iteratively expands them into a larger set of keywords. By conducting a modularity-based optimization over the underlying network of expanded tweets connected by shared keywords, DQE can identify specific localized instantiations of rumors.

As shown in Figure 6.4, on 09/29/2014 (when there was no incidence of Ebola in the US), the dominant rumor is the zombie rumor. By 10/06/2014, other rumors pertaining to how Ebola can be airborne and that it is a potential terrorist weapon gained hold.

Although Figure 6.4 might suggest that rumors are quite rampant, it is important to keep in perspective that they are but a small fraction of all information propagation related to Twitter. Figure 6.5 and 6.6 compare the time-indexed spread of the ‘patent’ rumor versus a true news story (about the first US incidence of Ebola in Dallas). Here different colors denote different communities participating in information propagation, not different rumors/news

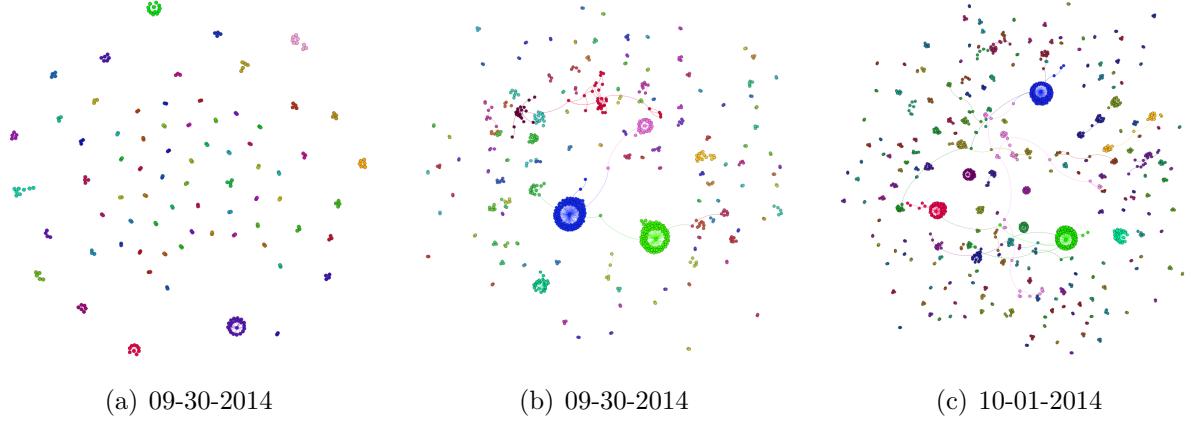


Figure 6.5: Ebola-related patent rumor propagation over time.

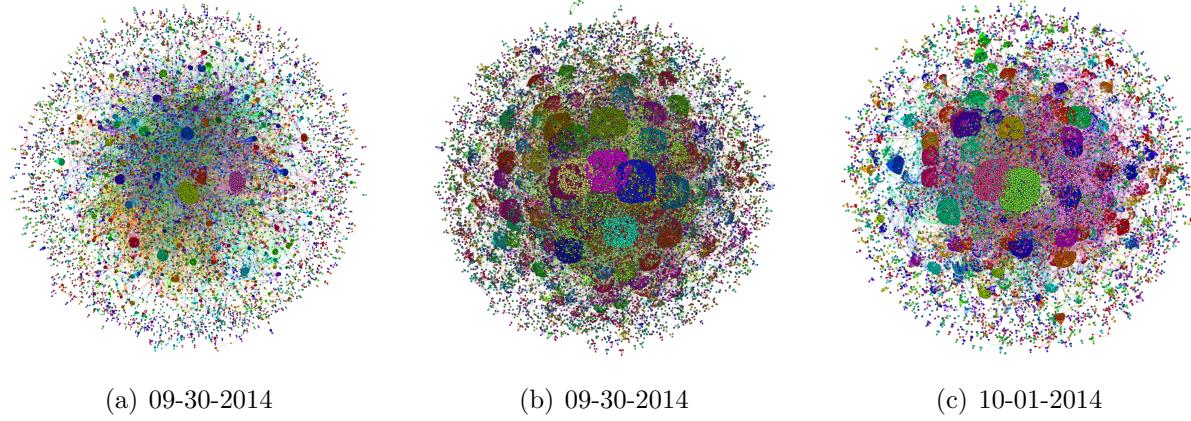


Figure 6.6: Ebola-related Dallas news propagation over time.

stories. Each node in these graphs denotes a Twitter user, and an edge between nodes denotes a reply or retweet relationship. As can be seen, news stories permeate better whereas rumors are more localized, distributed, and comparatively smaller in permeation.

6.3 Epidemiological modeling of rumors

Another way to study the spread of rumors (versus news) is from an epidemiological modeling standpoint. An epidemiological model helps capture the likelihood of an individual getting infected with a virus or, here, of adopting an idea that he or she has been exposed to.

In earlier work [42] we demonstrated how we can accomplish this objective using the SEIZ epidemiological model that was originally proposed to study the adoption of ideas [5]. The

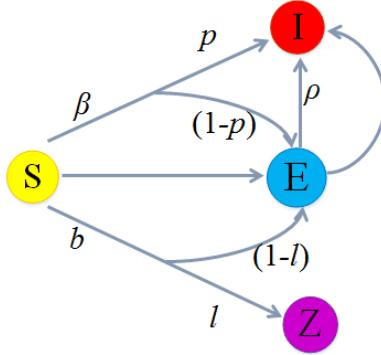


Figure 6.7: The SEIZ compartmental model. The various states denote: (S) Susceptible. (I) Infected. (E) Exposed. (Z) Skeptic.

SEIZ model is particularly suited to studying rumor propagation as it captures distinctions in how people respond to ideas: whether they adopt it readily or are initially skeptical.

The idea in the SEIZ model is to compartmentalize a population into four categories, denoted as S, E, I, and Z. We interpret these categories with specific reference to Twitter propagation. Susceptible (S) represents a user who has not heard the information; infected (I) denotes a user who has (re)tweeted about the information; skeptic (Z) denotes a user who has heard about the information but chooses not to (re)tweet about it; and exposed (E) represents a user who has received the information via a tweet but has taken some time, an exposure delay, prior to reposting or sharing that information.

The transitions between these states are modeled as shown in Fig 6.7. We caution that referring to the Z compartment as a skeptic is in no way an implication of the underlying truth or falsehood of the information; it simply helps capture whether users readily adopt an idea or take some time to adopt it.

Model fits of SEIZ to the different rumors and time course information for each of the state variables is given in Figure 6.8. As can be seen the SEIZ model is capable of capturing a variety of information spread patterns: quasi-linear (e.g., patent), sigmoidal (white), and other non-linear patterns (zombies and airborne).

Time course results from the SEIZ compartmental model as shown in Figure 6.8 depict broadly similar patterns. Here N denotes the total size of the population (distinct Twitter users). High values of S rapidly decrease with a relatively comparable increase in Z, and a gradual increase in I that continues as E decreases. However, the patent rumor time-course data has a noticeably different response profile than the other rumor examples.

Here, the initial value of the S group begins with less than half of the total population size, and only slightly higher than the initial values of the Z and E groups. Second, the Z group is essentially constant, meaning that the number of skeptics does not change throughout the propagation time course. Third, the decrease in S does not correspond to a change in Z, as

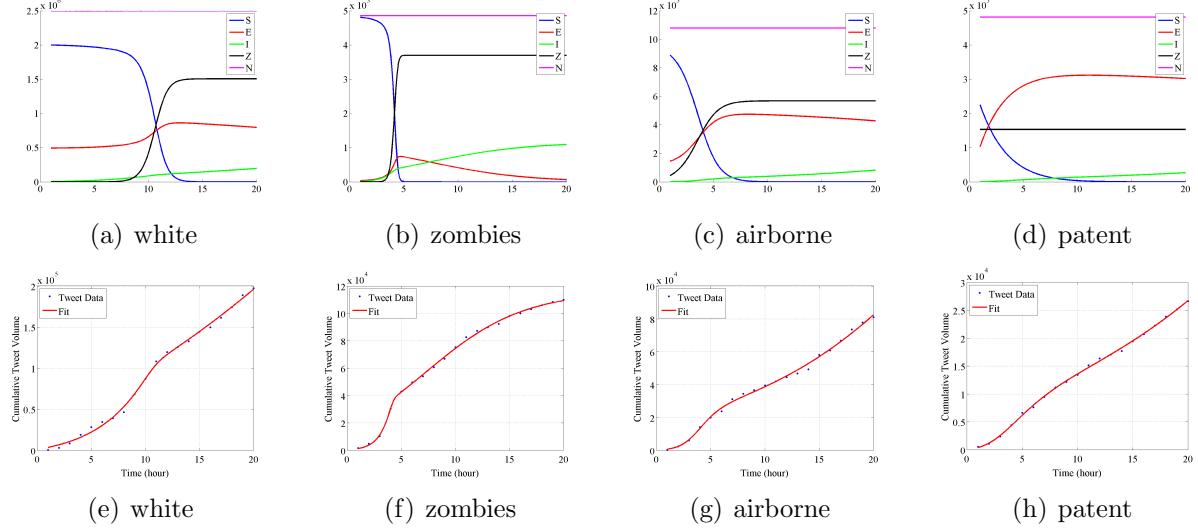


Figure 6.8: Model fits of SEIZ to different rumors: (from left to right) white, zombies, airborne, and patent. (top row) Fitting results. (bottom row) time-course profiles of different compartments.

is observed in the other rumor examples. Rather, the drop in S is met with a near identical increase in E.

These findings hint that a large influx into E without a corresponding efflux to I combined with a stagnant Z group will produce an elevated response ratio. In other words, there is a large exposure to the rumor topic without significant change in skepticism.

In our earlier work on characterizing rumors [42], we defined the notion of a response ratio which quantifies transitions through the exposed compartment. The response ratio provides a relative measure of the population influx into the E compartment versus the efflux from this compartment. We hypothesize that this ratio could be one of the factors useful in discriminating rumors from true news, with larger response ratios associated with factual news topics.

To compare response ratios across rumor and news, we select three breaking (true) news stories pertaining to Ebola: ‘Dallas’ refers to the story of the first Ebola patient (Duncan) identified in the US; ‘NYC’ refers to the first confirmation of an Ebola patient (Spencer) in New York City; and ‘Spencer’ refers to the specific symptoms and travel activities of Spencer in the days before he was diagnosed.

Ultimate R_{SI} value The ultimate response ratios for these three news stories and other rumors are shown in Figure 6.9. It can be seen that all three news stories (blue bars) have response ratios higher than 25, with a mean value of approximately 38, while eight of the 10 rumors stories (red bars) have a response ratio less than or equal to 6.4, with a mean of

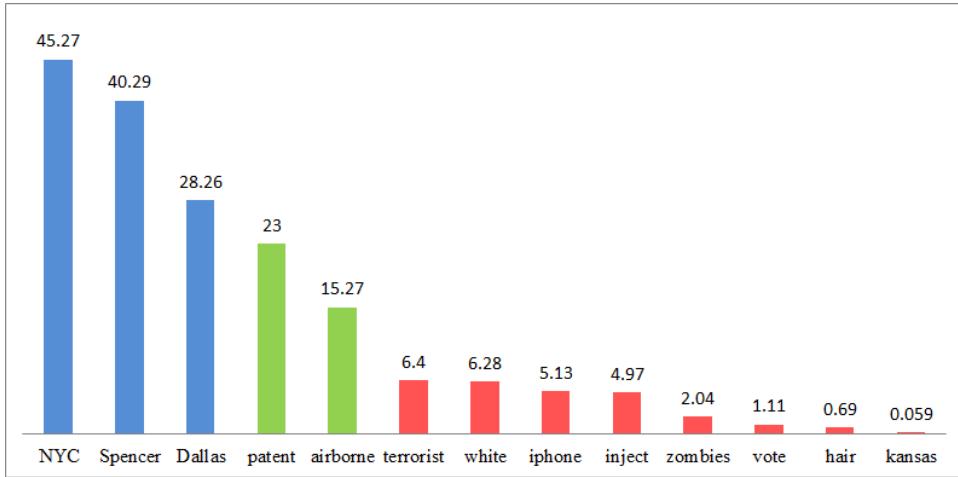


Figure 6.9: Ultimate response ratios for 3 news stories (left) and 10 rumors related to Ebola.

only 3.33. Two of the 10 rumors (green bars; ‘paten’ and ‘airborne’) have elevated response values, suggesting that there was greater belief associated with these topics than the other eight rumors.

Dynamic R_{SI} value Figure 6.10 illustrates the dynamic response ratio of Zombie rumor and Dallas real news. We can see the response ratio for Ebola Zombie rumor swings between 0 to 4, which means people’s believe extent to Zombie rumor has been very low. However, the dynamic response ratio for Dallas real news kept increasing and reached as high as nearly 100, which reveals more and more people accept this real news.

The study here has shown that propagation of misinformation can sometimes have the same characteristics as genuine newsworthy developments. In an age where many consumers receive their news from real-time social media platforms, it is imperative that rumors and half-truths be characterized as such and able to be distinguished from news. The tools presented here can support the quantitative evaluation of information spread as it happens.

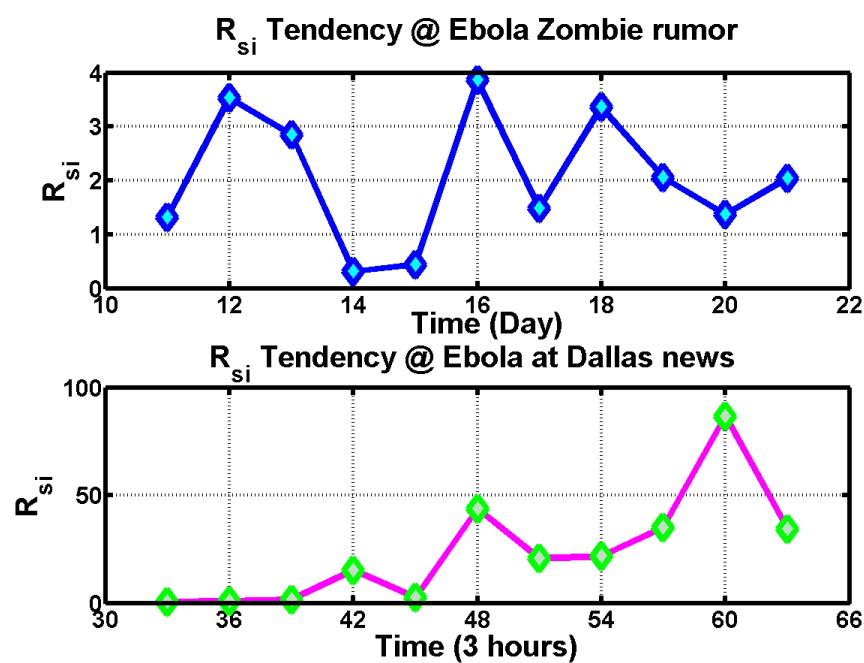


Figure 6.10: Dynamic R_{SI} values for Zombie rumor and Dallas real news.

Chapter 7

Concluding Remarks

As social media (e.g., Twitter) continues to increase in popularity, it is becoming employed as a social sensor into real-world mass movement event detection. Modeling and studying their adoption patterns gives us insight into investigating social and physical aspects of those events and precursors. This dissertation has presented several approaches and strategies with the goal of detecting and predicting mass movements and further inferring its causality, with given information mixed with real news and rumors. Those include techniques to capture information propagation across multiple spaces, as well as a graph wavelet approach that broadens predictive capabilities to capture group abnormality within dynamic changing networks. Numerous forms of mass movements have been investigated and diverse aspects of modeling and detecting have been addressed.

Using social media as indicators for real-word event detection is indeed helpful tool, however, they do possess limitations, perhaps most notably when applied to a specific event type, such as mass movement studied here. First, modeling protest-related topic propagation on networks is never trivial. One challenge is social protest propagation through online media can spread over large areas more quickly than traditional methods since users are geographically distributed, the other challenges include mass protest information can be spread by multiple social medias and lot of paths, like word of mouth, TV and news broadcast. Second, detect the group abnormality on social media is challenging. One challenge is Twitter's user network embodies many subgraphs based on social ties which is dynamically changing the graph structures since users are active. The other challenge includes real world events are not only correlated with burst signals, but can also exhibit unusually low levels of activity in social networks. Despite these restrictions, graph wavelet have in fact provided powerful capacity in capturing graph abnormality (considering burst behavior and absenteeism behavior), even on dynamic changing networks.

A fundamental objective of this dissertation has been to model mass movement adoption behavior, and in doing so, several significant advantages are gained beyond the target. One contribution is the ability to model information propagation across multiple networks/spaces,

and capture the propagation speed and possible propagation paths, which is demonstrated in Chapter 2. Another benefit that enhanced the mass movement detection is though group abnormality detection, as introduced in chapter 3. Graph wavelet provides appropriate definition of group abnormality which can cover both burst and absenteeism with different scales, thereby increasing the probability of capturing protest behaviors. Another by-product is the capability of quantify compartment transition dynamics using epidemic model SEIZ, and facilitate the development of screening criteria for distinguishing real movements from rumors happenings on Twitter, as demonstrated in Chapter 5.

Understanding information propagation over dynamic social network is highly-popular for addressing real-world problems in social network analysis. This dissertation analyzes several fundamental questions underlie the propagation-like processes, such as mass movement adoption, rumors transmission. These methodologies can be extended to other applications such as infectious diseases, public health, marketing, and so on. Given current advancements in information propagation, this is a revolutionary time for research in real-world applications using social network analysis.

Bibliography

- [1] C. C. Aggarwal and K. Subbian. Event detection in social streams. In *Proc. SDM'12*, volume 12, pages 624–635, 2012.
- [2] L. Akil, H. A. Ahmad, and R. S. Reddy. Effects of climate change on salmonella infections. *Foodborne pathogens and disease*, 11(12):974–980, 2014.
- [3] P. Antwi-Agyei, E. D. Fraser, A. J. Dougill, L. C. Stringer, and E. Simelton. Mapping the vulnerability of crop production to drought in ghana using rainfall, yield and socioeconomic data. *Applied Geography*, 32(2):324 – 334, 2012.
- [4] R. B. Bapat. *Graphs and matrices*. 2010.
- [5] L. Bettencourt, A. Cintrón-Arias, D. I. Kaiser, and C. Castillo-Chávez. The power of a good idea: Quantitative modeling of the spread of ideas from epidemiological models. *PHYSICA A*, 364:513–536, 2006.
- [6] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, 2008.
- [7] R. M. Bond, C. J. Fariss, J. J. Jones, A. D. Kramer, C. Marlow, J. E. Settle, and J. H. Fowler. A 61-million-person experiment in social influence and political mobilization. *Nature*, 489(7415):295–298, 2012.
- [8] C. Budak, D. Agrawal, and A. El Abbadi. Limiting the spread of misinformation in social networks. In *Proc. WWW'11*, pages 665–674. ACM, 2011.
- [9] 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*, 5(1):77–85, 2013.
- [10] M. Burke, S. M. Hsiang, and E. Miguel. Climate and conflict. Technical report, National Bureau of Economic Research, 2014.
- [11] R. BUSH. Food riots: Poverty, power and protest1. *Journal of Agrarian Change*, 10(1):119–129, 2010.

- [12] 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.
- [13] H. Castellini and L. Romanelli. On the propagation of social epidemics in social networks under SIR model. *eprint arXiv:nlin/0703053*, Mar. 2007.
- [14] C. Castillo, M. Mendoza, and B. Poblete. Information credibility on twitter. In *Proc. WWW’11*, pages 675–684. ACM, 2011.
- [15] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi. Measuring user influence in twitter: The million follower fallacy. In *ICWSM’10*, volume 14, page 8, 2010.
- [16] D. Chen, C.-T. Lu, Y. Kou, and F. Chen. On detecting spatial outliers. *Geoinformatica*, 12(4):455–475, 2008.
- [17] F. Chen and D. B. Neill. Non-parametric scan statistics for event detection and forecasting in heterogeneous social media graphs. In *Proc. KDD’14*, pages 1166–1175. ACM, 2014.
- [18] E. Cho, S. A. Myers, and J. Leskovec. Friendship and mobility: user movement in location-based social networks. In *Proc. KDD’11*, pages 1082–1090. ACM, 2011.
- [19] M. D. Conover, E. Ferrara, F. Menczer, and A. Flammini. The digital evolution of occupy wall street. *PloS one*, 8(5):e64679, 2013.
- [20] R. Crane and D. Sornette. Robust dynamic classes revealed by measuring the response function of a social system. *PNAS*, 105(41):15649–15653, 2008.
- [21] D. I Shuman, B. Ricaud, and P. Vandergheynst. Vertex-frequency analysis on graphs. *Applied and Computational Harmonic Analysis*, 2013.
- [22] D. J. Daley and D. G. Kendall. Epidemics and rumours. *Nature*, 204(4963):1118–1118, 1964.
- [23] I. Daubechies et al. *Ten lectures on wavelets*, volume 61. 1992.
- [24] 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.
- [25] A. P. Gaudine and A. M. Saks. Effects of an absenteeism feedback intervention on employee absence behavior. *Journal of Organizational Behavior*, 22(1):15–29, 2001.
- [26] 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.

- [27] P. H. Gleick. Water, drought, climate change, and conflict in syria. *Weather, Climate, and Society*, 6(3):331–340, 2014.
- [28] M. Gomez-Rodriguez, J. Leskovec, and A. Krause. Inferring networks of diffusion and influence. In *Proc. SIGKDD’10*, pages 1019–1028. ACM, 2010.
- [29] S. González-Bailón, J. Borge-Holthoefer, A. Rivero, and Y. Moreno. The dynamics of protest recruitment through an online network. *Scientific reports*, 1, 2011.
- [30] A. Guille, H. Hacid, C. Favre, and D. A. Zighed. Information diffusion in online social networks: A survey. *ACM SIGMOD Record*, 42(1):17–28, 2013.
- [31] H. Zhou. Network landscape from a brownian particle’s perspective. *Physical Review E*, 67(4):041908, 2003.
- [32] F. A. Haight and F. A. Haight. *Handbook of the Poisson distribution*. Wiley New York, 1967.
- [33] 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.
- [34] C. Hendrix, S. Haggard, and B. Magaloni. 1 grievance and opportunity: Food prices, political regime, and protest, 2009.
- [35] A. L. Hill, D. G. Rand, M. A. Nowak, and N. A. Christakis. Emotions as infectious diseases in a large social network: the sisa model. *Proc. Royal Society B: Biological Sciences*, 277(1701):3827–3835, 2010.
- [36] L. Hong, A. Ahmed, S. Gurumurthy, A. J. Smola, and K. Tsoutsouliklis. Discovering geographical topics in the twitter stream. In *Proc. WWW’12*, pages 769–778, 2012.
- [37] S. M. Hsiang, K. C. Meng, and M. A. Cane. Civil conflicts are associated with the global climate. *Nature*, 476(7361):438–441, 2011.
- [38] IPCC. *Climate Change 2007: Climate Change Impacts, Adaptation and Vulnerability*. Cambridge University Press, 2007.
- [39] V. Isham, S. Harden, and M. Nekovee. Stochastic epidemics and rumours on finite random networks. *PHYSICA A*, 389(3):561–576, 2010.
- [40] T. Iwata, A. Shah, and Z. Ghahramani. Discovering latent influence in online social activities via shared cascade poisson processes. In *Proc. KDD’13*, pages 266–274, 2013.
- [41] G. James, D. Witten, T. Hastie, and R. Tibshirani. *An introduction to statistical learning*, volume 112. Springer, 2013.

- [42] F. Jin, E. Dougherty, P. Saraf, Y. Cao, and N. Ramakrishnan. Epidemiological modeling of news and rumors on twitter. In *Proceedings of the 7th Workshop on Social Network Mining and Analysis*, page 8. ACM, 2013.
- [43] F. Jin, R. P. Khandpur, N. Self, E. Dougherty, S. Guo, F. Chen, B. A. Prakash, and N. Ramakrishnan. Modeling mass protest adoption in social network communities using geometric brownian motion. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1660–1669. ACM, 2014.
- [44] F. Jin, W. Wang, L. Zhao, E. Dougherty, Y. Cao, C.-T. Lu, and N. Ramakrishnan. Misinformation propagation in the age of twitter. *Computer*, (12):90–94, 2014.
- [45] S. Johnstone and J. Mazo. Global warming and the arab spring. *Survival*, 53(2):11–17, 2011.
- [46] C. P. Kelley, S. Mohtadi, M. A. Cane, R. Seager, and Y. Kushnir. Climate change in the fertile crescent and implications of the recent syrian drought. *Proceedings of the National Academy of Sciences*, 112(11):3241–3246, 2015.
- [47] D. Kempe, J. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *Proc. KDD’03*, pages 137–146. ACM, 2003.
- [48] W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 115, pages 700–721. The Royal Society, 1927.
- [49] M. Kimura, K. Saito, and H. Motoda. Efficient estimation of influence functions for sis model on social networks. In *Proc. IJCAI’09*, pages 2046–2051. Morgan Kaufmann Publishers Inc., 2009.
- [50] M. Kulldorff. A spatial scan statistic. *Communications in Statistics-Theory and methods*, 26(6):1481–1496, 1997.
- [51] H. Kwak, C. Lee, H. Park, and S. Moon. What is twitter, a social network or a news media? In *Proc. WWW’10*, pages 591–600. ACM, 2010.
- [52] T. Lappas, B. Arai, M. Platakis, D. Kotsakos, and D. Gunopoulos. On burstiness-aware search for document sequences. In *Proc. KDD’09*, pages 477–486, 2009.
- [53] T. Lappas, M. R. Vieira, D. Gunopoulos, and V. J. Tsotras. On the spatiotemporal burstiness of terms. *Proceedings of the VLDB Endowment*, 5(9):836–847, 2012.
- [54] P. Le Billon. The political ecology of war: natural resources and armed conflicts. *Political geography*, 20(5):561–584, 2001.

- [55] Y. Liao and V. R. Vemuri. Using text categorization techniques for intrusion detection. In *USENIX Security Symposium*, volume 12, pages 51–59, 2002.
- [56] Y. Matsubara, Y. Sakurai, B. A. Prakash, L. Li, and C. Faloutsos. Rise and fall patterns of information diffusion: model and implications. In *Proc. KDD’12*, pages 6–14. ACM, 2012.
- [57] M. Mendoza, B. Poblete, and C. Castillo. Twitter under crisis: Can we trust what we rt? In *Proc. SOMA ’10*, pages 71–79. ACM, 2010.
- [58] C. C. Mitigation. Ipccl special report on renewable energy sources and climate change mitigation. 2011.
- [59] P. Mörters and Y. Peres. *Brownian motion*, volume 30. Cambridge University Press, 2010.
- [60] J. D. Murray. *Mathematical biology*, volume 2. Springer, 2002.
- [61] S. A. Myers, C. Zhu, and J. Leskovec. Information diffusion and external influence in networks. In *Proc. KDD’12*, pages 33–41, 2012.
- [62] M. Nekovee, Y. Moreno, G. Bianconi, and M. Marsili. Theory of rumour spreading in complex social networks. *PHYSICA A*, 374(1):457–470, 2007.
- [63] M. E. Newman. Spread of epidemic disease on networks. *Physical review E*, 66(1):016128, 2002.
- [64] M. E. Newman. The structure and function of complex networks. *SIAM review*, 45(2):167–256, 2003.
- [65] M. E. Newman. Modularity and community structure in networks. *PNAS*, 103(23):8577–8582, 2006.
- [66] B. Øksendal. *Stochastic differential equations*. Springer, 2003.
- [67] P. E. Oliver and D. J. Myers. Diffusion models of cycles of protest as a theory of social movements. *Presented at the Congress of the International Sociological Association*, 1998.
- [68] V. Qazvinian, E. Rosengren, D. R. Radev, and Q. Mei. Rumor has it: Identifying misinformation in microblogs. In *Proc. EMNLP’11*, pages 1589–1599. Association for Computational Linguistics, 2011.
- [69] 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*, pages 1799–1808, 2014.

- [70] M. G. Rodriguez, J. Leskovec, and B. Schölkopf. Modeling information propagation with survival theory. In *Proc. ICML'13*, pages 666–674, 2013.
- [71] D. M. Romero, B. Meeder, and J. Kleinberg. Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In *Proc. WWW'11*, pages 695–704. ACM, 2011.
- [72] P. Rozenshtein, A. Anagnostopoulos, A. Gionis, and N. Tatti. Event detection in activity networks. In *Proc. KDD'14*, pages 1176–1185, 2014.
- [73] R. Rustamov and L. Guibas. Wavelets on graphs via deep learning. In *Proc. NIPS'13*, pages 998–1006, 2013.
- [74] A. Saad-Filho. Mass protests under ‘left neoliberalism’: Brazil, june-july 2013. *Critical Sociology*, 39(5):657–669, 2013.
- [75] 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.
- [76] H. Sarsons. Rainfall and conflict, 2011.
- [77] H. Sayyadi, M. Hurst, and A. Maykov. Event detection and tracking in social streams. In *ICWSM'09*, 2009.
- [78] J. Scheffran, M. Brzoska, J. Kominek, P. M. Link, and J. Schilling. Climate change and violent conflict. *Science*, 336(6083):869–871, 2012.
- [79] J. Scott and P. J. Carrington. *The SAGE handbook of social network analysis*. SAGE publications, 2011.
- [80] B. C. Seamonds. Stress factors and their effect on absenteeism in a corporate employee group. *Journal of Occupational and Environmental Medicine*, 24(5):393–hyphen, 1982.
- [81] D. Shah and T. Zaman. Rumors in a network: Who’s the culprit? *Information Theory, IEEE Transactions on*, 57(8):5163–5181, 2011.
- [82] 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.
- [83] Y. Singer. How to win friends and influence people, truthfully: influence maximization mechanisms for social networks. In *Proc. WSDM'12*, pages 733–742. ACM, 2012.
- [84] K. Starbird and L. Palen. (how) will the revolution be retweeted?: information diffusion and the 2011 egyptian uprising. In *Proc. ACM'12*, pages 7–16. ACM, 2012.
- [85] T. Sternberg. Chinese drought, bread and the arab spring. *Applied Geography*, 34:519 – 524, 2012.

- [86] P. Tankov. *Financial modelling with jump processes*. CRC Press, 2004.
- [87] N. Tremblay and P. Borgnat. Graph wavelets for multiscale community mining. 2014.
- [88] D. Trpevski, W. K. Tang, and L. Kocarev. Model for rumor spreading over networks. *Physical Review E*, 81(5):056102, 2010.
- [89] Z. Tufekci and C. Wilson. Social media and the decision to participate in political protest: Observations from Tahrir square. *Journal of Communication*, 62(2):363–379, 2012.
- [90] K. Warner, C. Ehrhart, A. d. Sherbinin, S. Adamo, T. Chai-Onn, et al. In search of shelter: Mapping the effects of climate change on human migration and displacement. *In search of shelter: mapping the effects of climate change on human migration and displacement*, 2009.
- [91] 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*, pages 2541–2544, 2011.
- [92] J. Weng and B.-S. Lee. Event detection in twitter. *ICWSM’11*, pages 401–408, 2011.
- [93] U. F. Wiersema. *Brownian motion calculus*. Wiley. com, 2008.
- [94] G. Wischnath and H. Buhaug. On climate variability and civil war in asia. *Climatic Change*, 122:709–721, 2014.
- [95] F. Xiong, Y. Liu, Z.-j. Zhang, J. Zhu, and Y. Zhang. An information diffusion model based on retweeting mechanism for online social media. *Physics Letters A*, 376(30-31):2103–2108, 2012.
- [96] L. Xiong, B. Póczos, J. G. Schneider, A. Connolly, and J. VanderPlas. Hierarchical probabilistic models for group anomaly detection. In *Proc. AISTATS’11*, pages 789–797, 2011.
- [97] J. Yang and S. Counts. Predicting the speed, scale, and range of information diffusion in twitter. *Proc. AAAI’10*, 2010.
- [98] Z. Yin, L. Cao, J. Han, C. Zhai, and T. Huang. Geographical topic discovery and comparison. In *Proc. WWW’11*, pages 247–256, 2011.
- [99] R. Yu, X. He, and Y. Liu. Glad: group anomaly detection in social media analysis. In *Proc. KDD’14*, pages 372–381. ACM, 2014.
- [100] D. H. Zanette. Dynamics of rumor propagation on small-world networks. *Physical Review E*, 65(4):041908, 2002.

- [101] 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*, 9(10):e110206, 2014.
- [102] L. Zhao, H. Cui, X. Qiu, X. Wang, and J. Wang. Sir rumor spreading model in the new media age. *PHYSICA A*, 392(4):995–1003, 2012.
- [103] H. Zhou. Distance, dissimilarity index, and network community structure. *Physical Review E*, 67(6):061901, 2003.
- [104] H. Zhou and R. Lipowsky. Network brownian motion: A new method to measure vertex-vertex proximity and to identify communities and subcommunities. In *Proc. ICCS'04*, pages 1062–1069. Springer, 2004.