

Epidemiological Modeling of News and Rumors on Twitter

Fang Jin*, Edward Dougherty ‡, Parang Saraf*, Yang Cao*‡, Naren Ramakrishnan*

*Department of Computer Science

‡Genetics, Bioinformatics, and Computational Biology Department

Virginia Tech, Blacksburg, VA 24061

*{jfang8, parang, ycao, naren}@cs.vt.edu, ‡edougherty@vt.edu

ABSTRACT

Characterizing information diffusion on social platforms like Twitter enables us to understand the properties of underlying media and model communication patterns. As Twitter gains in popularity, it has also become a venue to broadcast rumors and misinformation. We use epidemiological models to characterize information cascades in twitter resulting from both news and rumors. Specifically, we use the SEIZ enhanced epidemic model that explicitly recognizes skeptics to characterize eight events across the world and spanning a range of event types. We demonstrate that our approach is accurate at capturing diffusion in these events. Our approach can be fruitfully combined with other strategies that use content modeling and graph theoretic features to detect (and possibly disrupt) rumors.

Categories and Subject Descriptors

H.2.8 [Database Management]: Database Applications—*Data Mining*; I.2.6 [Artificial Intelligence]: Learning—*Knowledge acquisition; Parameter learning*

General Terms

Experimentation, Performance

Keywords

SIS, SEIZ, Epidemiological modeling, Rumor detection.

1. INTRODUCTION

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.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

The 7th SNA-KDD Workshop '13 (SNA-KDD'13), August 11, 2013, Chicago, United States.

Copyright 2013 ACM 978-1-4503-2330-7 ...\$5.00.

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. [1] **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 [1] 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 paper 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 show-case how such information could facilitate the development of screening criteria for distinguishing rumors from real news happenings on Twitter.

2. RELATED WORK

Information Diffusion.

Significant work has gone into research on information diffusion on social media, e.g., see [4, 9, 16, 21]. Recently, Matsubara et al. [10] 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. [6] 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. [14] 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. [8] proposed the application of the SIS model to study information diffusion where the nodes can be activated multiple times. Zhao et al. [23] 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. [20] 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. [1] 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 [5] first proposed the similarity between epidemics and rumors using mathematical analysis. Some researchers have studied rumor propagation modeling in different network topologies [13, 22]; however, they do not provide any discussion of propagation differences between news and rumors. Shah et al. [17] detect rumor sources in network using maximum likelihood modeling. In [2], 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. [3] delve into twitter content modeling, such as sentiment analysis and hashtags to identify rumors, while Qazvinian et al. [15] 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 [7, 19]. Our goal is to develop an understanding of these processes using diffusion models.

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 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.

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/>

Table 1: Twitter datasets studied in this paper.

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	4838	47.17%	Antorcha Campesina, Nezahualcoyotl

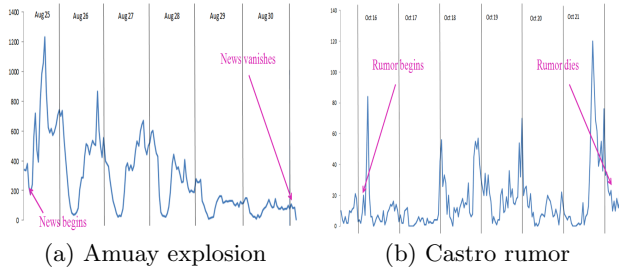


Figure 1: Tweet volume.

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.

3.3 Preliminary Analysis

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/>

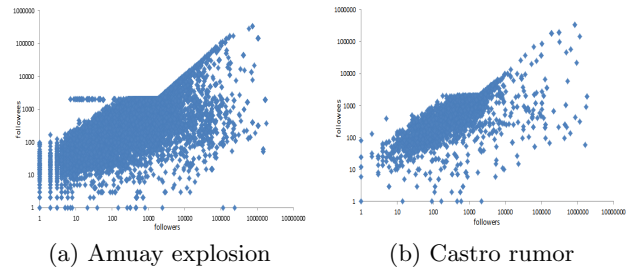


Figure 2: Followers/followees distributions. Followers: people who follow the person; Followees: people who are followed by the person.

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 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. [11]. In contrast, Figure 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 2(a) is a log-log scatter plot of the followers/followees distribution about the Amuay explosion news, and Figure 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. [18] discuss, retweets reveal how information propagates through a social network: the 'deeper' a retweet, the more relevant

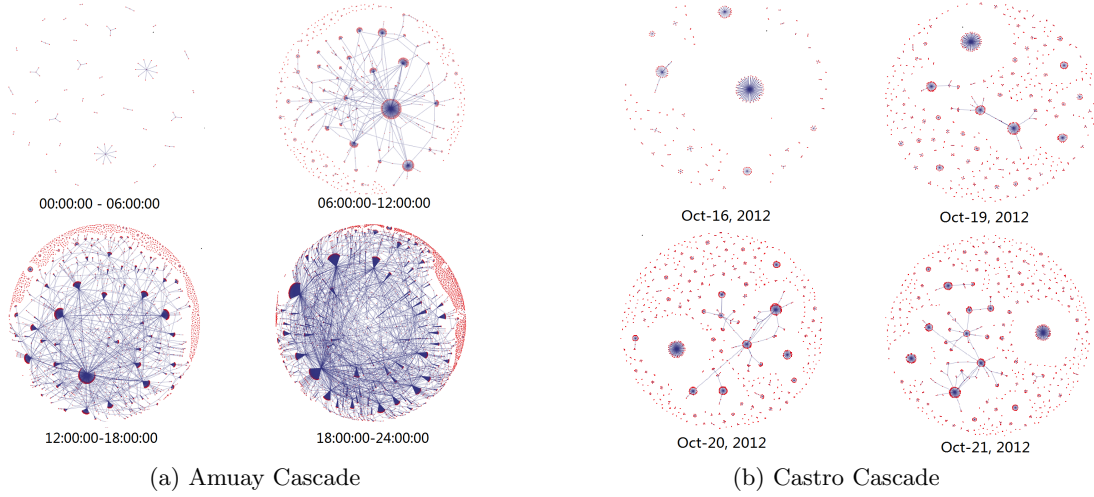


Figure 3: Retweet cascade for the Amuay Explosion news and Castro rumor. Each node is a user id, and each edge connects the retweet user to the original user.

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 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.

Retweet Cascades. A retweet cascade reflects how the social media network propagates information. Figure 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. 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.

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.

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 4, the SIS spreading rule can be summarized as follows:

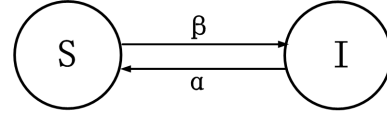


Figure 4: SIS model framework

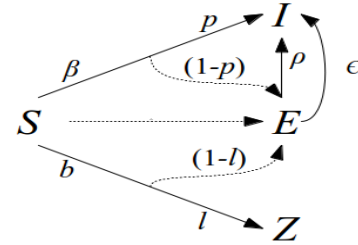


Figure 5: SEIZ model framework

- 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) [12]:

$$\frac{d[S]}{dt} = -\beta SI + \alpha I \quad (1a)$$

$$\frac{d[I]}{dt} = \beta SI - \alpha I \quad (1b)$$

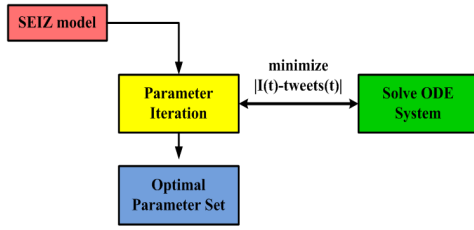


Figure 6: Numerical implementation work-flow.

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 [1]. 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 [1].

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, 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 probability 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 con-

Table 2: Parameter definitions in SEIZ model[1]

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 \rightarrow Z
$\beta\rho$	Effective rate of S \rightarrow I
$b(1-l)$	Effective rate of S \rightarrow E via contact with Z
$\beta(1-p)$	Effective rate of S \rightarrow E via contact with I
l	S \rightarrow Z Probability given contact with skeptics
$1-l$	S \rightarrow E Probability given contact with skeptics
p	S \rightarrow I Probability given contact with adopters
$1-p$	S \rightarrow E Probability given contact with adopters

tact 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 [1]). 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.

$$\frac{d[S]}{dt} = -\beta S \frac{I}{N} - bS \frac{Z}{N} \quad (2a)$$

$$\frac{d[E]}{dt} = (1-p)\beta S \frac{I}{N} + (1-l)bS \frac{Z}{N} - \rho E \frac{I}{N} - \epsilon E \quad (2b)$$

$$\frac{d[I]}{dt} = p\beta S \frac{I}{N} + \rho E \frac{I}{N} + \epsilon E \quad (2c)$$

$$\frac{d[Z]}{dt} = lbS \frac{Z}{N} \quad (2d)$$

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 value shows to be very different from the total number of Twitter accounts. Consider the ~ 175 million (M) registered Twitter accounts. Of

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

Table 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	0.096	0.069
<i>SEIZ</i>	0.010	0.004	0.027	0.061	0.101	0.029	0.073	0.053	0.045

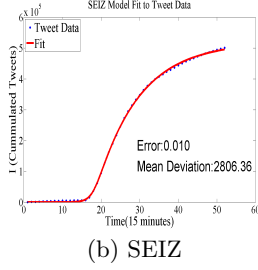
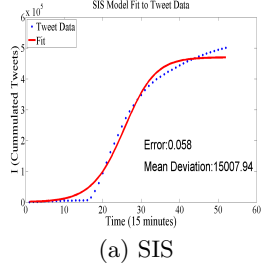


Figure 7: Best fit modeling for Boston news.

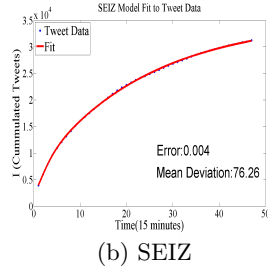
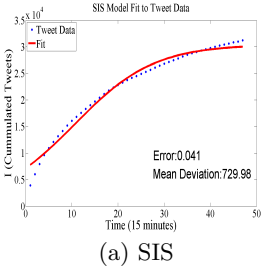


Figure 8: Best fit modeling for Pope news.

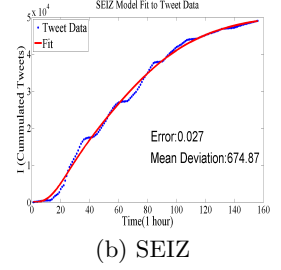
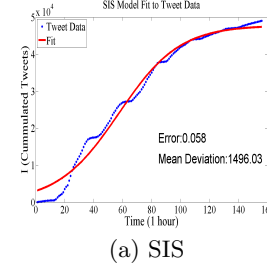


Figure 9: Best fit modeling for Amuay news.

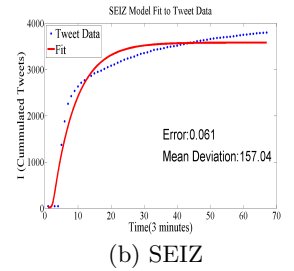
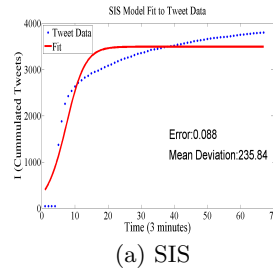


Figure 10: Best fit modeling for Michelle news.

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.
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 [1].

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

4.4 Parameter Identification

For each of the population models (SIS and SEIZ), represented by equation sets 1 and 2, we performed a nonlinear least squares fit of the model to Twitter data. As shown in Figure 6, each step of this fitting process involved selecting a set of parameter values (rate constants and probabilities in equations 1 and 2, and initial compartment sizes), and numerically solving the system of ODEs with these parameter values. The set of parameter values that minimized $|I(t) - tweets(t)|$ 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 several instances we compared the forward Euler solution to those generated by Matlab’s `ode45` (5th-order Explicit Runge-Kutta with embedded 4th-order error control), and observed nearly identical solutions.

5. EXPERIMENTAL RESULTS

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 7 - 14 display the results for the best fit of SIS and SEIZ models (Equations 1 and 2) to the eight Twitter stories. Also

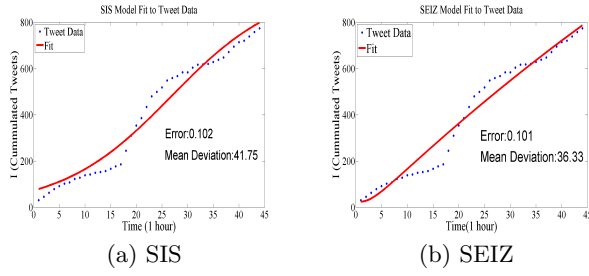


Figure 11: Best fit modeling for Obama news.

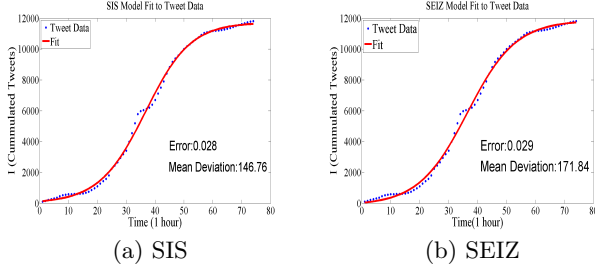


Figure 12: Best fit modeling for Doomsday rumor.

displayed for each figure are the relative error in 2-norm

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

and the mean error deviation

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

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 3. A common observation about all the eight stories is that 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 [1].

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 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 2 specifies the SEIZ model parameters that we can now examine to assess news and rumor propagation on Twitter.

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

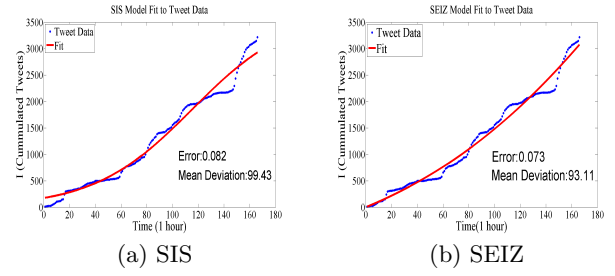


Figure 13: Best fit modeling for Castro rumor.

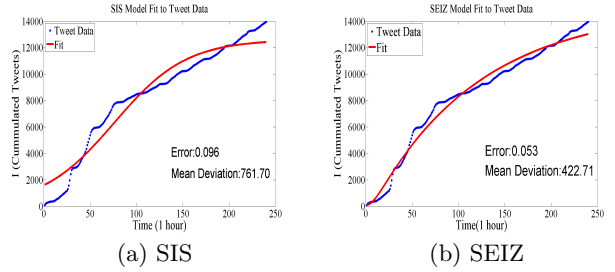


Figure 14: Best fit modeling for Riot news.

SEIZ model fit of the Boston Marathon bombing Twitter data (Table 2). Results are summarized in Table 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 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 4 also demonstrates that the skeptics compartment is more influential on transitioning susceptible users to the exposed class than does infected users. Figure 15(a) shows 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 of Twitter propagation! Further, the $\frac{\epsilon}{\rho}$ ratio indicates that the exposed users became infected more so due to infor-

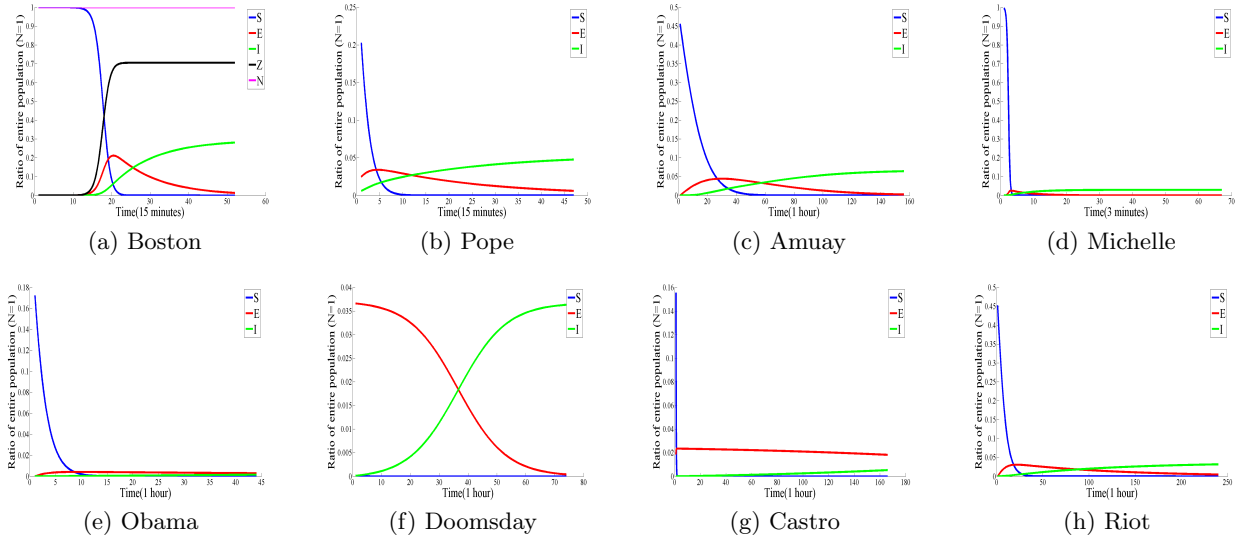


Figure 15: SEIZ compartment time-course results.

Table 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

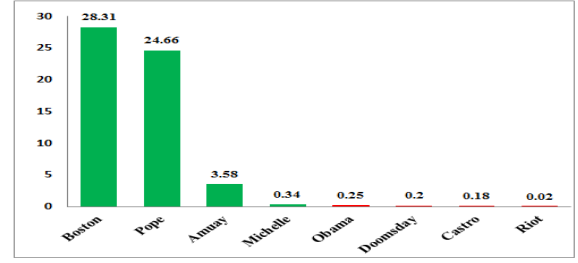


Figure 16: R_{SI} values for eight Twitter datasets.

mation 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 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.

5.3 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). 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:

$$R_{SI} = \frac{(1-p)\beta + (1-l)b}{\rho + \epsilon} \quad (3)$$

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 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.

We then computed R_{SI} using the specific parameter values attained from our model fits of the eight cases (Figure 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.

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.

6. CONCLUSION

In this paper, we have demonstrated how true news and rumor stories being propagated over Twitter can be modeled 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.

Acknowledgments

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

7. REFERENCES

- [1] 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.
- [2] 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.
- [3] C. Castillo, M. Mendoza, and B. Poblete. Information credibility on twitter. In *Proc. WWW’11*, pages 675–684. ACM, 2011.
- [4] 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.
- [5] D. J. Daley and D. G. Kendall. Epidemics and rumours. *Nature*, 204(4963):1118–1118, 1964.
- [6] M. Gomez-Rodriguez, J. Leskovec, and A. Krause. Inferring networks of diffusion and influence. In *Proc. SIGKDD’10*, pages 1019–1028. ACM, 2010.
- [7] V. Isham, S. Harden, and M. Nekovee. Stochastic epidemics and rumours on finite random networks. *PHYSICA A*, 389(3):561–576, 2010.
- [8] 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.
- [9] 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.
- [10] Y. Matsubara, Y. Sakurai, B. A. Prakash, L. Li, and C. Faloutsos. Rise and fall patterns of information diffusion: model and implications. In *Proc. SIGKDD’12*, pages 6–14. ACM, 2012.
- [11] 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.
- [12] J. D. Murray. *Mathematical biology*, volume 2. springer, 2002.
- [13] M. Nekovee, Y. Moreno, G. Bianconi, and M. Marsili. Theory of rumour spreading in complex social networks. *PHYSICA A*, 374(1):457–470, 2007.
- [14] M. E. Newman. Spread of epidemic disease on networks. *Physical review E*, 66(1):016128, 2002.
- [15] 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.
- [16] 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.
- [17] D. Shah and T. Zaman. Rumors in a network: Who’s the culprit? *Information Theory, IEEE Transactions on*, 57(8):5163–5181, 2011.
- [18] 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.
- [19] D. Trpevski, W. K. Tang, and L. Kocarev. Model for rumor spreading over networks. *Physical Review E*, 81(5):056102, 2010.
- [20] 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.
- [21] J. Yang and S. Counts. Predicting the speed, scale, and range of information diffusion in twitter. *Proc. AAAI’10*, 2010.
- [22] D. H. Zanette. Dynamics of rumor propagation on small-world networks. *Physical Review E*, 65(4):041908, 2002.
- [23] 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.