5-2015

Adapting the Standard SIR Disease Model in Order to Track and Predict the Spreading of the EBOLA Virus Using Twitter Data

Armin Smailhodzic
Western Kentucky University, Armin.Smailhodzic465@topper.wku.edu

ADAPTING THE STANDARD SIR DISEASE MODEL IN ORDER TO TRACK AND PREDICT THE SPREADING OF THE EBOLA VIRUS USING TWITTER DATA

A Thesis
Presented to
The Faculty of the Department of Physics and Astronomy
Western Kentucky University
Bowling Green, Kentucky

In Partial Fulfillment
of the Requirement for the Degree
Master of Science

By
Armin Smailhodzic

May 2015

ADAPTING THE STANDARD SIR DISEASE MODEL IN ORDER TO TRACK AND PREDICT THE SPREADING OF THE EBOLA VIRUS USING TWITTER DATA

Date Recommended 4/24/2015

Keith Andrew

Keith Andrew, Director of Thesis

Phillip Womble

Lance Hahn

Dean, Graduate School 4-25-15

Date

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Keith Andrew, for his mentorship, guidance, patience, time, and input throughout my undergraduate and graduate studies at Western Kentucky University. I would also like to thank Dr. Phillip Womble and Dr. Lance Hahn for their assistance and constructive criticism.

Also, I would like to thank Sharon Windham, secretary for the Physics and Astronomy Department, for putting up with all my crazy requests from buying me popsicles to helping me fill out all my paperwork throughout undergraduate and then graduate studies. I will miss our gossip breaks.

Thank you Tara Wink and Jason Smith, the two of you have pushed me to continuously challenge myself. The two of you were always there for me to vent when things were rough and provided me an escape at your home when I needed to just sit down and work numerous days away to make progress. (Kitty, thanks for keeping me entertained)

Lastly I would like to thank my parents and sister. My parents gave up many things to ensure that my sister and I would have the opportunities we have in life. They moved us halfway around the globe for us to have a chance at a better future through gaining an education. Their sacrifices have molded me into the man I am and there aren't enough words to explain how thankful I am for them. I love you mom and dad. Sis, thanks for always getting better grades and being smarter than me (even if it wasn't in Physics), I enjoyed the competition.

To all my other friends, you know who you are, thank you for putting up with me going super nerd and not having time to be social. Thanks for the memories!

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ADAPTING THE STANDARD SIR DISEASE MODEL IN ORDER TO TRACK AND PREDICT THE SPREADING OF THE EBOLA VIRUS USING TWITTER DATA

Armin Smailhodzic May 2015 47 Pages

Directed by: Dr. Keith Andrew, Dr. Phillip Womble, and Dr. Lance Hahn

Department of Physics and Astronomy Western Kentucky University

A method has been developed to track infectious diseases by using data mining of active Twitter accounts and its efficacy was demonstrated during the West African Ebola outbreak of 2014. Using a meme based n-gram semantic usage model to search the Twitter database for indications of illness, flight and death from the spread of Ebola in Africa, principally from Guinea, Sierra Leone and Liberia. Memes of interest relate disease to location and severity and are coupled to the density of Tweets and re-Tweets. The meme spreads through the community of social users in a fashion similar to nonlinear wave propagation- like a shock wave, visualized as a spike in Tweet activity. The spreading was modeled as a system isomorphic to a modified SIR (Susceptible, Infected, Removed disease model) system of three coupled nonlinear differential equations using Twitter variables. The nonlinear terms in this model lead to feedback mechanisms that result in unusual behavior that does not always reduce the spread of the disease. The resulting geographic Tweet densities are coupled to geographic maps of the region. These maps have specific threat levels that are ported to a mobile application (app) and can be used by travelers to assess the relative safety of the region they will be in.

INTRODUCTION

Ebola

According to the article “The First 9 Months of the Epidemic and Forward Projections,” published in the New England Journal of Medicine, Ebola hemorrhagic fever is an infectious and deadly disease. It was named after a river in the Democratic Republic of the Congo in 1976. “By September 14, 2014, a total of 4507 confirmed and probable cases of Ebola Virus Disease (EVD), as well as 2296 deaths from the virus, had been reported from five countries in West Africa: Guinea, Liberia, Nigeria, Senegal, and Sierra Leone (64)”. (1)

The current EVD outbreak, beginning in the early months of 2014, is larger than all previous ones combined and has demonstrated much higher mortality and morbidity rates. The true numbers of cases and deaths are expected to be higher than what is being reported. Current problems causing incorrect reporting include: symptomatic persons not seeking diagnosis and treatment, laboratory diagnoses not being included in national databases, and suspected EVD patients that were buried before they could be diagnosed. (8, 64)

Guinea, Liberia, and Sierra Leone are currently at the center of the outbreak and face challenges in implementing control measures to stop the current outbreak (8). The Ebola virus is spread mainly through contact with the body fluids of patients demonstrating symptoms. Transmission can be prevented by a combination of early diagnosis, patient isolation, infection control, and safe burial practices. (17, 69,70)

Twitter

Jack Dorsey, Evan Williams, Biz Stone and Noah Glass created Twitter in 2006 and the site was launched in 2006. According to Internet Live Stats, there are 7,532 Tweets sent per second and around 300 million active users (62). ”The Beginners Guide to Twitter,” states that Twitter is a social networking and micro-blogging service that allows its users to answer the question, “What are you doing?” by sending short text messages 140 characters in length, called “Tweets”, to ones friends, or “followers.” Users can send messages directly using the Twitter website, as a single text message, or via a third-party application. Tweets can be viewed on users profile pages, on the home page of each follower, and on the Twitter public timeline. (62)

Memes

A meme is a uniquely occurring word or set of words (n-gram). Single word memes (1-gram) appear more often than two (2-gram) or three (3-gram) set memes and so on. This pattern mimics Zipf’s Law, which predicts the frequency of a word occurring, which is directly proportional to the length of the word. This is also true of memes, the shorter the meme set is the more likely it is to occur; therefore the longer the meme set is (more specific) its occurrence should decrease. Using this law, memes were made longer in order to get more specific and detailed results, which in turn yield more accurate model data. (25)

The initial meme concept arises from an idea of Dawkins (73) in his development of language word constructions that have cultural and social features similar to genes and embody word change (36). Memes transfer information; cluster (20) and can enter into conventional usage on short time scales. To more accurately capture the nature of a given meme we use the n-gram definition where an n-gram is a meme consisting of n related words. N-grams can be characterized by their relative entropy (26) and nearby word associations (25), some word combinations are much more probable and often certain pairings can be attributed to geolocation (43,77).

Topsy

Topsy Labs was founded in 2007 and has access to every Tweet since 2006, which offers more than 425 billion data points. Using Topsy's software, paid users and limited edition users can "instantly analyze any topic, term or hashtag across years of conversations on millions of web sites" (66). The company was founded by Vipul Ved Prakash and Duncan Greatwood (66). According to Rayson, in " 6 Reasons to use Topsy", "Topsy search is more powerful than the standard and advanced Twitter search, due to the fact that one can use a wide range of search operators. One can also search for Tweets from a person, referencing a person, containing links to a site or even search for conversations between two people. By using Topsy's free social analytics tools over time one can: review all the shares of a piece of content including shares by key influencers, find influencers on a specific topic, find all content from a person or influencer on a specific topic, find all Tweets from a user linking to a specific site and the links they are sharing, find sentiment scores for a brand and track over time, and undertake detailed analytics tracking for keywords or sites (51)".

Google Charts

Google Charts is a tool that allows users to add visualization for data shown on ones website. Visualizations range from simple line charts to complex tree maps. Within Google Charts the chart gallery provides a large number of ready-to-use chart templates with code examples. Google Charts uses JavaScript, HTML, and CSS that users can embed on their web page. The final output is rendered using HTML5, which allows it to be viewed on Windows/Android/Apple devices. All chart types are created by populating the DataTable class, by doing so it makes it easy to switch between chart types to find the ideal appearance for the data one wishes to display. (67)

SIR Disease Model

Mathematical models of epidemics are created under the major assumptions that the population being observed can be divided into three subsets. These subsets are analogous to three types of people in the population. The simplest model, which was described by Kermack and McKendrick in 1927, consists of three subsets: susceptible (S), infected (I) and recovered (R). (31)

Susceptible - Any individuals that are able to be infected by the disease.

Infected - Infected individuals can spread the disease to susceptible individuals.

Recovered - Individuals assumed to be immune for life after full recovery.

The SIR model is quantified using ordinary differential equations (ODEs), which assumes that there is no randomness involved, the same starting conditions would yield the same result, and there is no time limit with continuous time. The rate of new infections defined as βSI, where β is a parameter for infectivity. The recovery rate is detonated by r, which is a constant per capita recovery rate and therefore the overall rate of recovery is rI. Based on the assumptions by Kermack and McKendrick the scheme of the model can be drawn as:

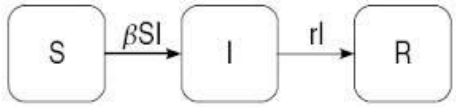


Figure 2 Scheme of the basic SIR model, boxes represent compartments, and arrows indicate flux between the compartments.

The scheme can also be translated into a set of differential equations:

$$\frac{dS}{dt} = -\beta\frac{SI}{N} \quad (1)$$

$$\frac{dI}{dt} = \beta\frac{SI}{N} - \gamma I \quad (2)$$

$$\frac{dR}{dt} = \gamma I \quad (3)$$

$$N = S + I + R \quad (4)$$

.

This model was adapted to use Twitter and population data to create a mathematical spreading model that included social media trends. (38, 31)

Mathematica

Wolfram Mathematica states that, "Mathematica is a software program used in many scientific, engineering, mathematical and computing fields, based on symbolic mathematics (75)". Wolfram Research founded by Stephen Wolfram of Champaign, Illinois, developed the software. Mathematica is split into two parts: the kernel, which interprets expressions and returns result expressions; and the front end, which allows the creation and editing of Notebook documents containing program code. All contents and formatting can be generated algorithmically and interactively edited. (75)

Present Work

Using Twitter platforms to search for memes of different lengths, location specific data was collected for West Africa. This location data was used in order to determine the current location of Ebola virus outbreaks, how quickly the data/news about Ebola was being spread and how the virus progressed through West Africa in 2014. Most of the Twitter accounts observed belonged to either doctors or news crews who were in the regions to either help the local medical infrastructure or to report on the ongoing spread of the virus. The data gathered was analyzed and used for predicting the spread/cases of the ongoing outbreak. The data obtained/collected in the early months of 2014, before the CDC/WHO began releasing situation reports was used to determine the initial spreading parameters using power functions. These parameters were then used in calculating the overall new case counts as the outbreak continued throughout 2014. The information obtained from this analysis was converted into interactive graphs using the free Google Visualization Charts, which allowed anyone to see what areas were affected by the Ebola virus and how many known cases were within each country. The data is then interpreted through a modified SIR model with Mathematica in order to visualize the spreading in the population.

Thesis Overview

This work will attempt to model disease spreading of Ebola through the use of social media. Twitter interactions will be focused on since these interactions are limited to 140 characters and have a possibility of being geo-located by the user. By creating longer and more specific memes one is able to get data for a region of interest. Doing so allows one to see the increase and decrease of data being shared on a daily basis. This would allow observation of trends by looking for peaks in the data mined. By selecting to observe specific Twitter users, one could observe the movement of said user accounts as they move around and continue to share via social media.

MATERIALS AND METHODS

Related Work

In 2010, Dr. Zhiuan Cheng of Texas A&M University wrote a paper on geo-locating Twitter users. In his paper, “You Are Where You Tweet: A Content-Based Approach to Geo-locating Twitter Users”, Cheng states that a large majority of Twitter users’ location information is absent from their profiles and to overcome the lack of location information they proposed and researched a framework for estimating a Twitter user’s city-level location based on the content of the user’s Tweets. In their research they found that they could place 51% of Twitter users within 100 miles of their actual location. (11)

Much of the early work on web-based disease surveillance relied on query logs and click-through data from search engines (12), most famously Google’s Flu Trends service. Other sources of information include articles from the news media and online mailing lists. The most recent work was conducted by Johns Hopkins University in 2013 wherein they used Twitter data from databases to compare Twitter users sharing information about being infected with the flu to CDC findings during that same time period. This information was then plotted with precision-recall curves in order to compare the findings to provided CDC data. (38)

Data Collection

Memes of different lengths were analyzed using Topsy. A single word meme (1-gram meme), such as Ebola, had more hits than memes of longer composition. The 1-gram meme used to get a baseline of information shared was Ebola. Over a one-month period the word Ebola appeared over 6.5 million times on Twitter, during July and

August in 2014. Peaks usually corresponded to global news agencies discussing the outbreak or updated information about new cases of Ebola and the total death toll.

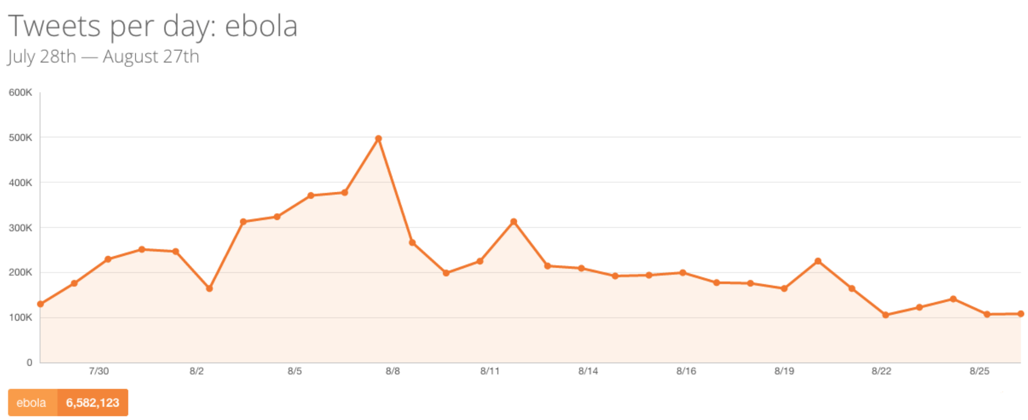


Figure 2 Results for 1-gram meme, showing the daily Tweets for the meme during a one-month period in 2014. The peak corresponds to new outbreak confirmations.

As the meme begins to get longer and in turn more specific the amount of occurrences reduces as is evident for all three countries affected by the Ebola virus. By making the meme longer and preserving the word order, one was able to get more detailed information for the region of interest. The observed peaks corresponded more to new confirmed cases and deaths instead of outbreak claims. By going through each peak and looking at the shared data one was able to determine when new cases were appearing in each country of interest along with new deaths per country.

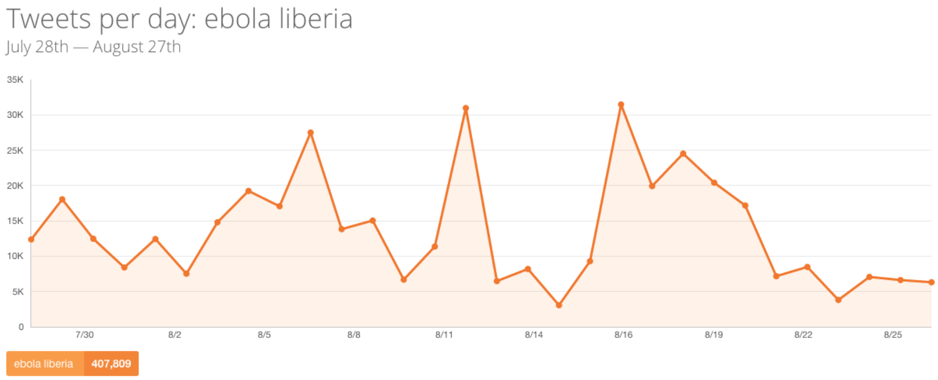


Figure 3 Result for 2-gram meme, showing the daily Tweets for the meme Ebola Liberia in 2014. Peaks were identified as events where an outbreak was discussed or the number of confirmed cases or deaths was mentioned.

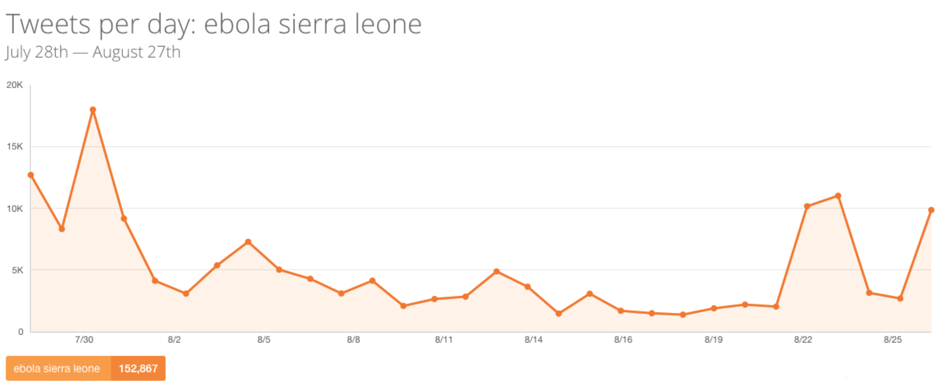


Figure 4 Result for 2-gram meme, showing the daily Tweets for the meme Ebola Sierra Leone in 2014. Peaks were identified as events where an outbreak was discussed or the number of confirmed cases or deaths was mentioned.

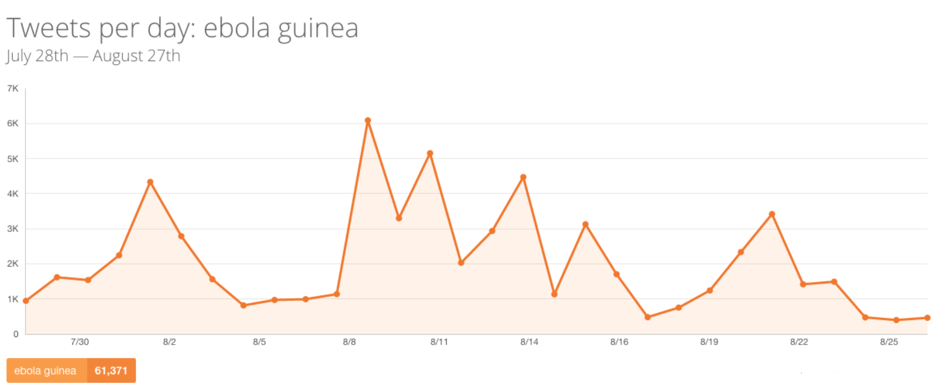


Figure 5 Result for 2-gram meme, showing the daily Tweets for the meme Ebola Guinea in 2014. Peaks were identified as events where an outbreak was discussed or the number of confirmed cases or deaths was mentioned.

Comparing all three of these countries one can see how the information travels and when information is similar/relevant for each country. When all three countries peak at the same time, it was found that information was being shared about the impact of the outbreak going on in Western Africa. When peaks were not identical for all three regions, the information shared was about the number of new confirmed cases and deaths per region.

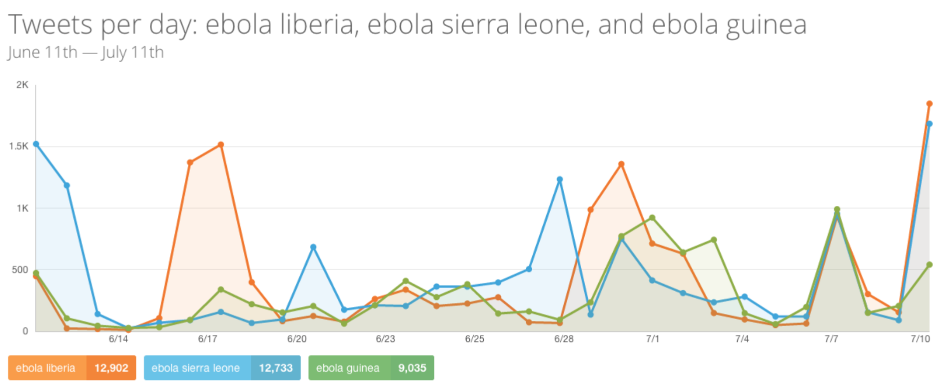


Figure 6 Comparison of the daily results for the 2-gram meme set, Ebola + Country, in 2014.

As the meme continues to get longer the occurrences decrease but the data becomes more accurate for calculations. For the following images death was added at the end of the meme. However, when using the term die or deaths at the end of the meme the results per day were dramatically lower.

Peaks in the 3-gram meme set corresponded to national health centers declaring a new case within the country or updating the information on the amount of confirmed cases and deaths within each country.

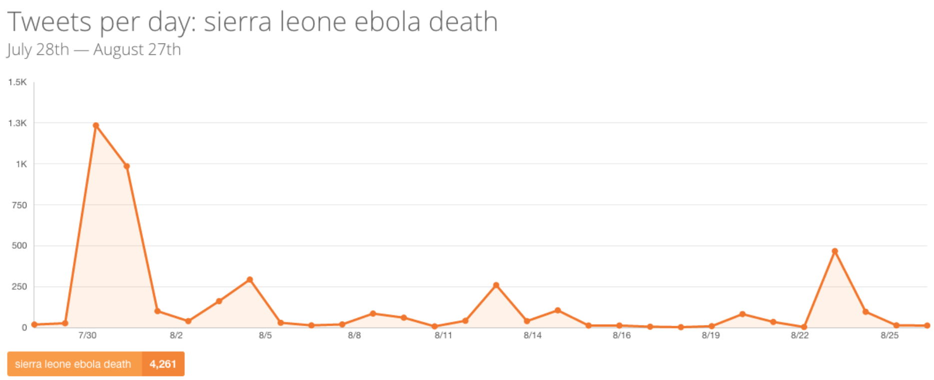


Figure 7 Result for 3-gram meme, showing the daily Tweets for the meme Sierra Leone Ebola Death in 2014.

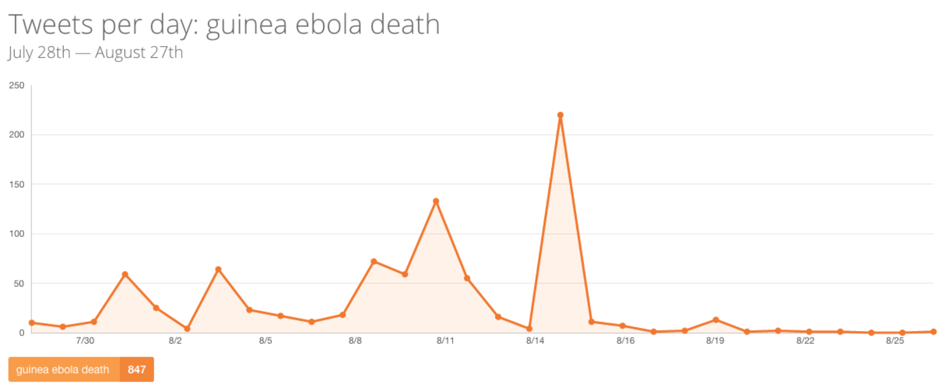


Figure 8 Result for 3-gram meme, showing the daily Tweets for the meme Guinea Ebola Death in 2014.

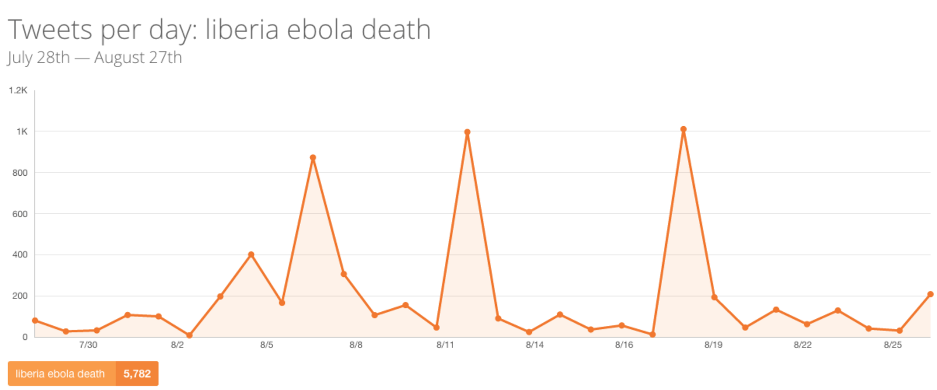


Figure 9 Result for 3-gram meme, showing the daily Tweets for the meme Liberia Ebola Death in 2014.

Mapping

Maps were created using Google Visualization software with jsfiddle coding support. Data collected for the amount of confirmed infected and confirmed deaths found on Twitter was plotted with the following code:

```
<script type="text/javascript"
src="https://www.google.com/jsapi?autoload={'modules':[{'name':'visualization','
version':'1','packages':['geochart']}]}"></script>
<div id="regions_div" style="width: 700px; height: 500px;"></div>
google.setOnLoadCallback(drawRegionsMap);
function drawRegionsMap() {
var data = google.visualization.arrayToDataTable([
['Country', 'Confirmed Cases', 'Confirmed Deaths'],
['Guniea', 510, 377],
['Liberia', 670, 355],
['Nigeria', 12, 3],
['Sierra Leone', 783, 334],
['Congo', 531, 402],
['Gabon', 65, 53]
]);
var options = {
region: '002',};
var chart = new
google.visualization.GeoChart(document.getElementById('regions_div'));
chart.draw(data, options, options, options);}
```

This code yielded the following interactive map, which allows users to scroll across highlighted regions in order to determine the country and information on cases and deaths per country since the year 2000.

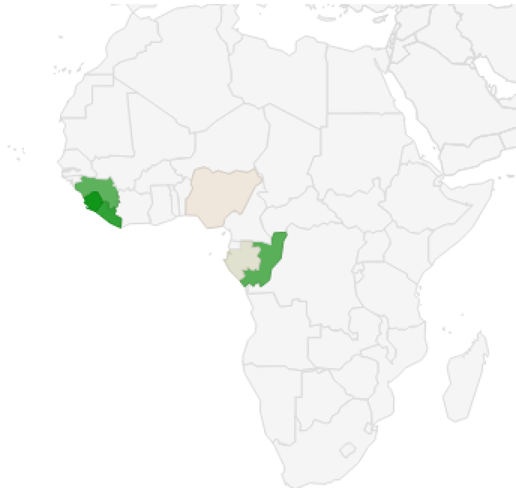


Figure 10 Map of Africa, depicting the countries, which have been affected by Ebola since the year 2000 through 2014.

This process was then repeated for the three countries affected with the 2014 outbreak. The code had to be modified to give results in marker output instead of regional output. The code to do so is as follows:

```
<script type='text/javascript' src='https://www.google.com/jsapi'></script>
<script type='text/javascript'>
google.load('visualization', '1', {
'packages': ['geochart']});
</script>
<div id="chart_div" style="width: 900px; height: 500px;"></div>
google.setOnLoadCallback(drawMarkersMap);
function drawMarkersMap() {
var data = google.visualization.arrayToDataTable([
['Region', 'Area Infected in Square Km'],
['Grand Campe Mount', 1993],
['Bomi', 1932],
['Montserrado', 1909],
['Margibi', 2616],
['Grand Bassa', 7936],
['Rivercess', 5594],
['Nimba', 11551],
['Bong', 8754],
['Lofa', 9982],
['Monrovia', 100] ]);
var options = {
region: 'LR',
displayMode: 'markers', };
```

```
var chart = new
google.visualization.GeoChart(document.getElementById('chart_div'));
chart.draw(data, options);}
```

The output allows users to see the affected areas within the country with the outbreak. This code tells users the area that is affected in square kilometers.

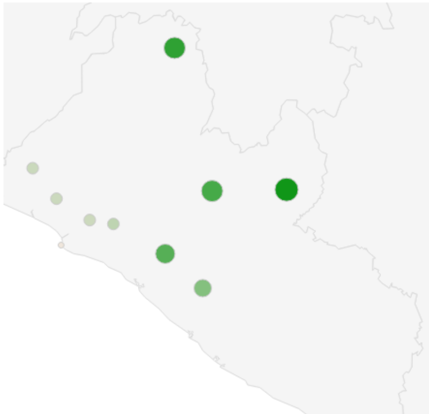


Figure 11 Map of Liberia with the affected regions and square kilometers per region during the first 7 months of the outbreak in 2014.

This process was then repeated and applied to create maps of Guinea and Sierra Leone.

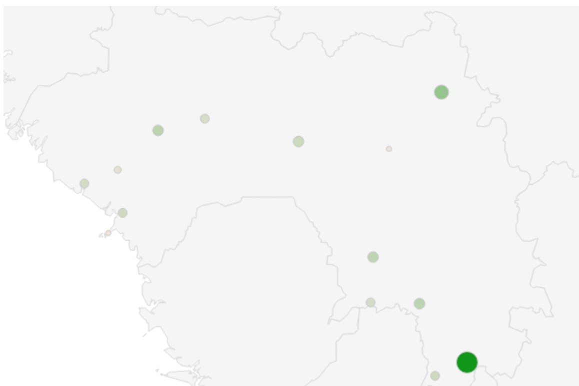


Figure 12 Map of Guinea with the affected regions and square kilometers per region during the first 7 months of the outbreak in 2014.

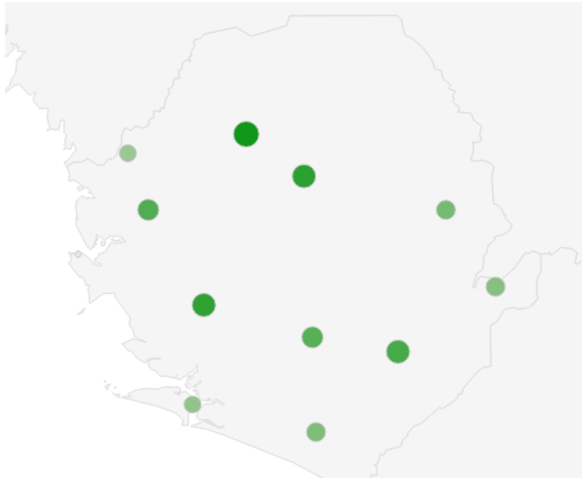


Figure 13 Map of Sierra Leone with the affected regions and square kilometers per region during the first 7 months of the outbreak in 2014.

Mathematica

The following Mathematica code is a representation of the adapted SIR model. Beta and gamma values are calculated before being placed into the equation, and can be adjusted depending on the infection rate and information propagation. Time was set to a 30-day interval and the resulting graphs represented how S, R and I progressed during this 30-day interval. Beta and gamma can be adjusted depending on the information distribution variable and the infectivity of the virus. The code allots for the observation of chaos, if it is present in the data set.

```
s = NDSolve[{x'[t] == -.67 ((x[t] y[t]) / ((x[t] + y[t] + z[t]))),
    y'[t] == 0.67 ((x[t] y[t]) / ((x[t] + y[t] + z[t]))) - 0.087 y[t],
    z'[t] == 0.087 y[t], x[0] == 1, z[0] == 0, y[0] == 0.0093}, {x, y, z},
   {t, 0, 30}, MaxSteps → ∞]
{{x → InterpolatingFunction[{{0., 30.}}, <>],
  y → InterpolatingFunction[{{0., 30.}}, <>],
  z → InterpolatingFunction[{{0., 30.}}, <>]}}
ParametricPlot3D[Evaluate[{x[t], y[t], z[t]} /. s], {t, 0, 1},
 PlotPoints -> 1000, ColorFunction → (Hue[#2] &)]
Plot[x[t] /. s, {t, 0, 30}]
Plot[y[t] /. s, {t, 0, 30}]
Plot[z[t] /. s, {t, 0, 30}]
```

The following charts represent the data comparing Ebola Tweets to Ebola Sierra Leone Tweets for the month of August, modeled using Mathematica. This data was then compared to the SIR model of disease spreading to ensure accuracy on the modeling. The x-axis represents the time interval while the y-axis represents the population.

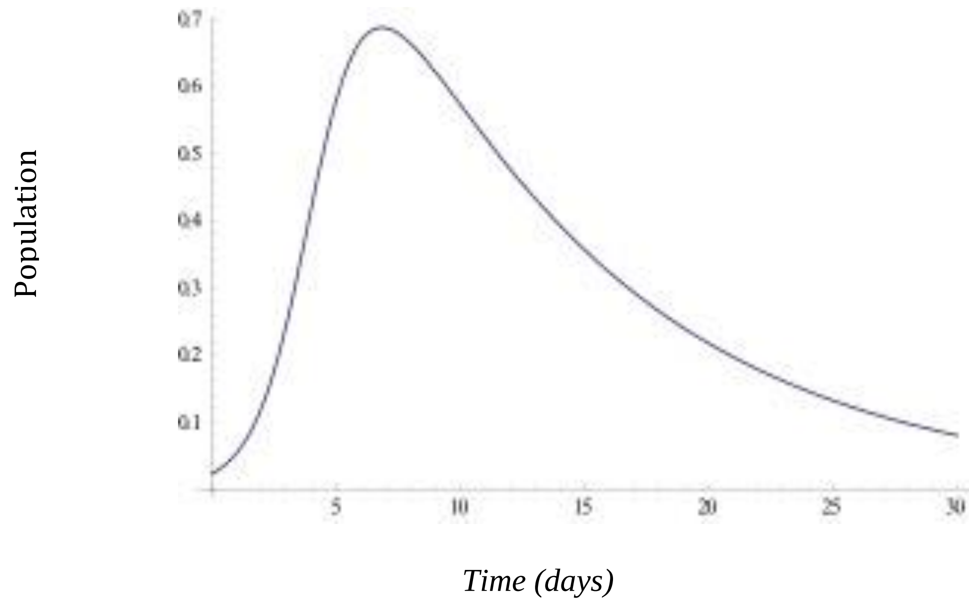


Figure 14 I vs. Time

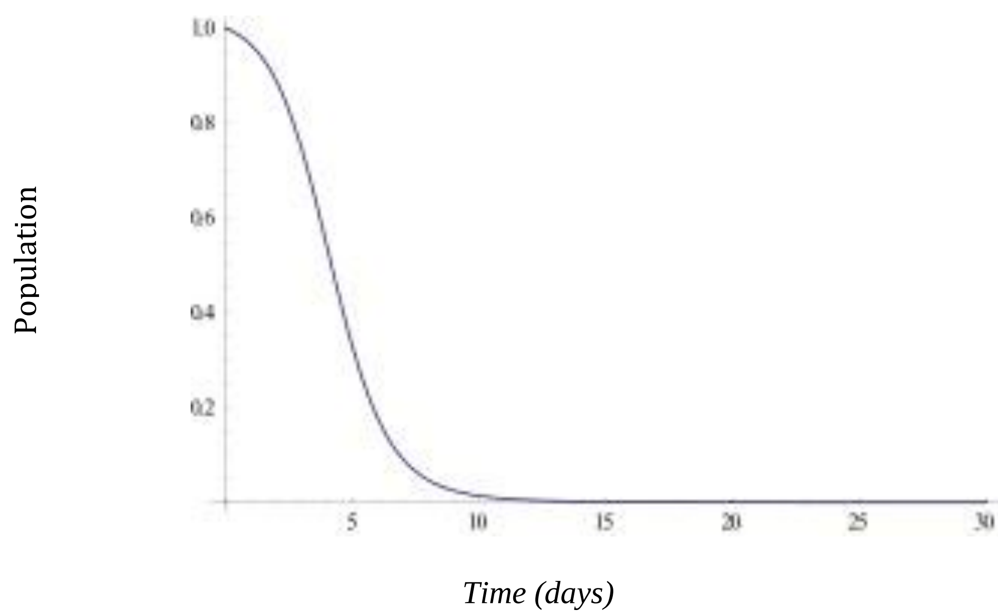


Figure 15 S vs. Time

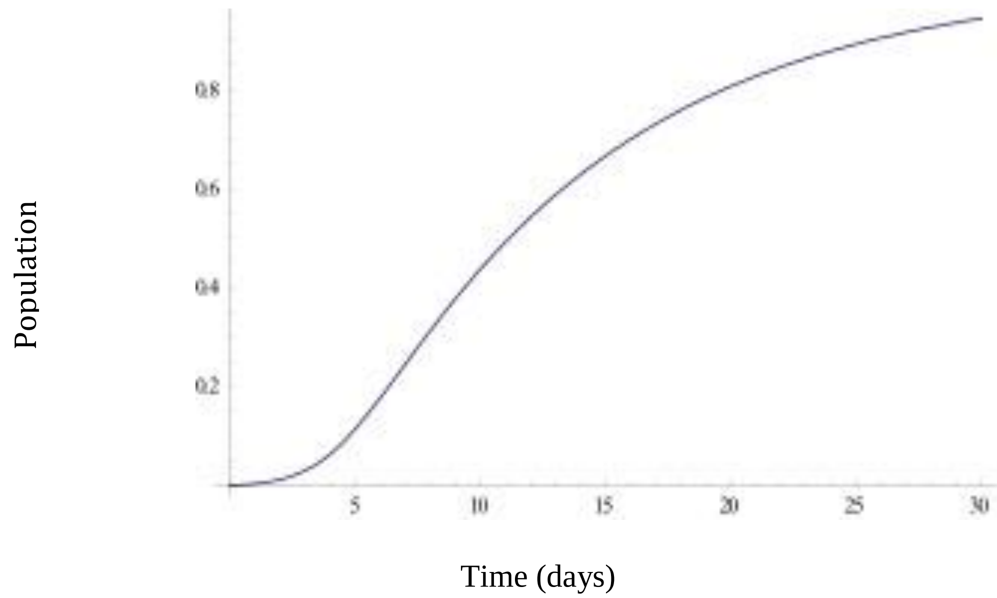


Figure 16 R vs. Time

SIR Model

The SIR model had to be adapted to model Twitter data received through the Topsy analytical software. In the original model S stood for Susceptible population, I for Infected population and R for Recovered population. These variables were changed to match the Twitter data. S became the total number of Tweets about Ebola (1-gram meme), I became the total number of 2-gram memes and R was set to 0 since the Tweets were set for a one-day period. Once the variables were determined the constants, beta, β, and gamma, γ, had to be calculated. Beta is the information distribution variable, which was determined to be between one and three days. Beta was changed depending on the population of the country and the amount of Tweets per country. The lower the beta constant, the lower number of Tweets in the area. Gamma is the incubation period of the disease one wants to model, in the case of Ebola, the incubation period lasts between two and twenty-one days, thus an incubation period of ten days was selected. Beta and gamma are calculated inversely, yielding a beta constant between one-third and one depending on the country and a gamma constant of one-tenth. (59)

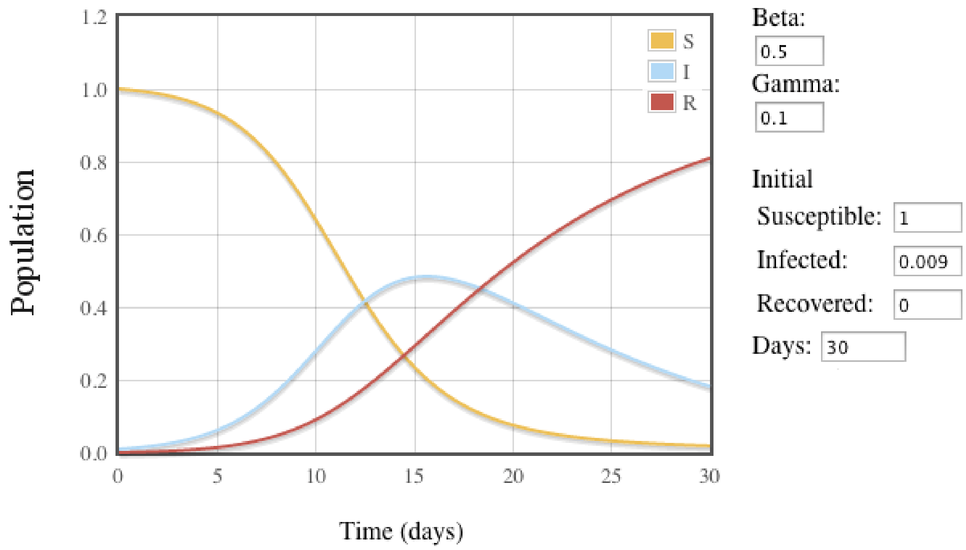


Figure 17 SIR model for Ebola Guinea Tweets (I) during August 2014.

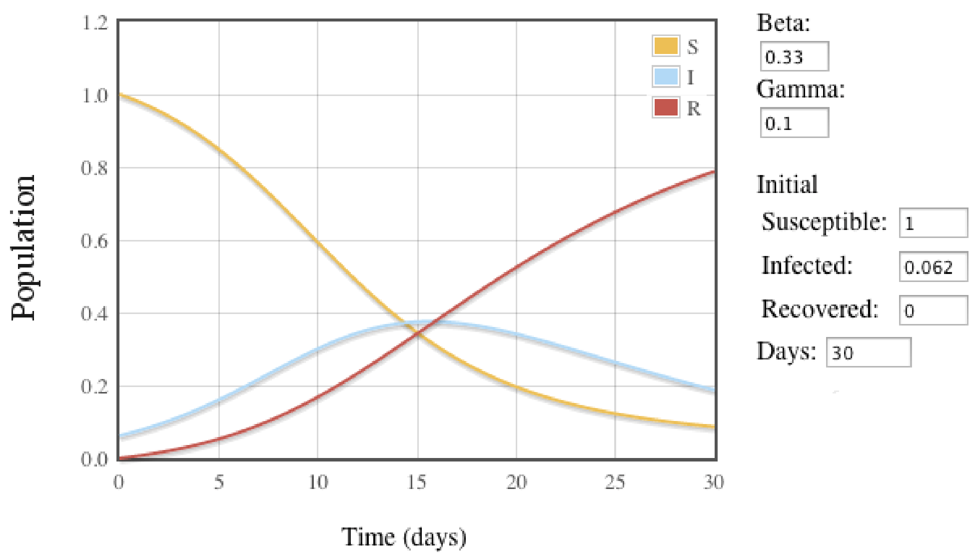


Figure 18 SIR model for Ebola Liberia Tweets (I) during August 2014.

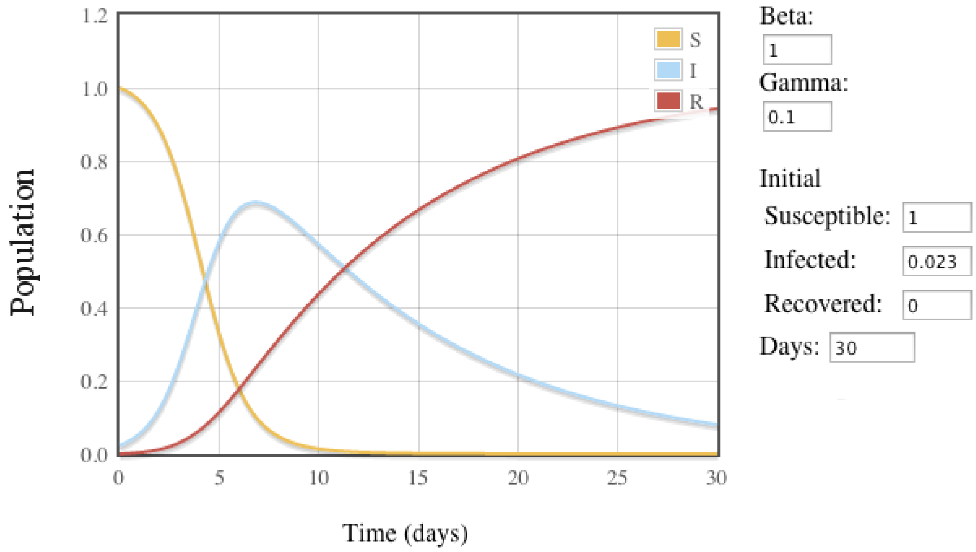


Figure 19 SIR model for Ebola Sierra Leone Tweets (I) during August 2014.

The following are SIR model results when comparing the 1-gram meme Ebola to the 3-gram meme Country + Ebola + Death. Gamma remained constant for all three SIR models regardless of the country. Beta was either 0.5 or 1 depending on the country.

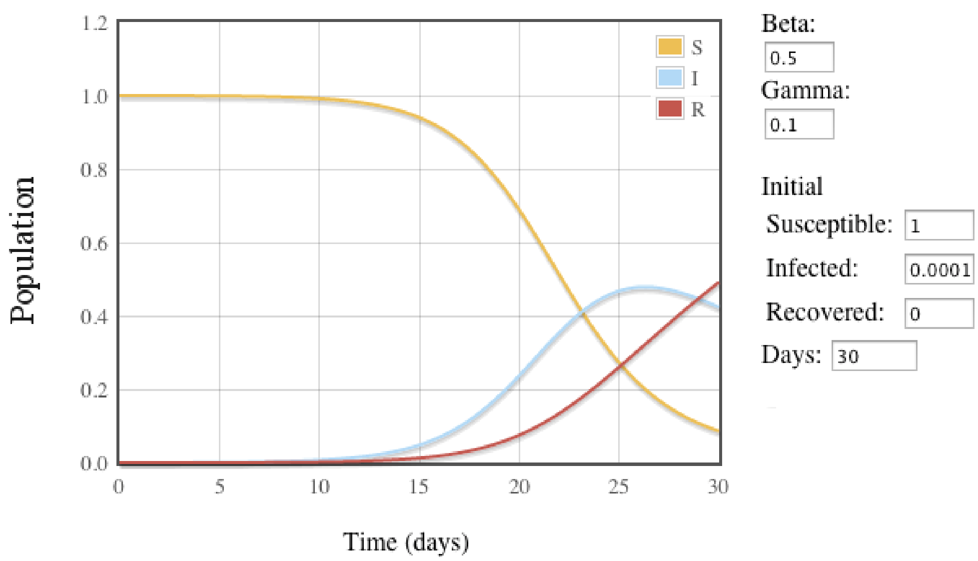


Figure 20 SIR model for Guinea Ebola Death Tweets (I) during August 2014.

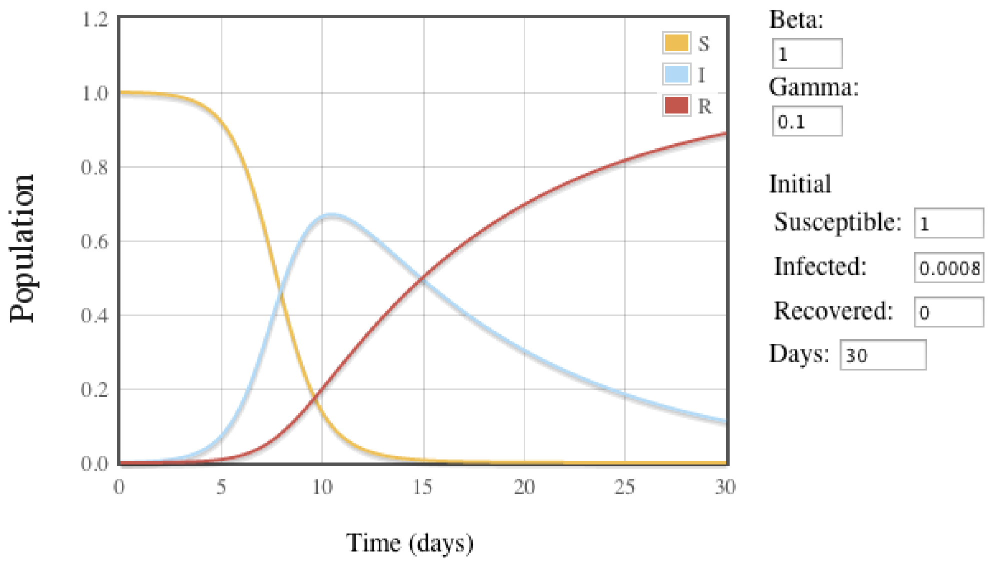


Figure 21 SIR model for Liberia Ebola Death (I) during August 2014.

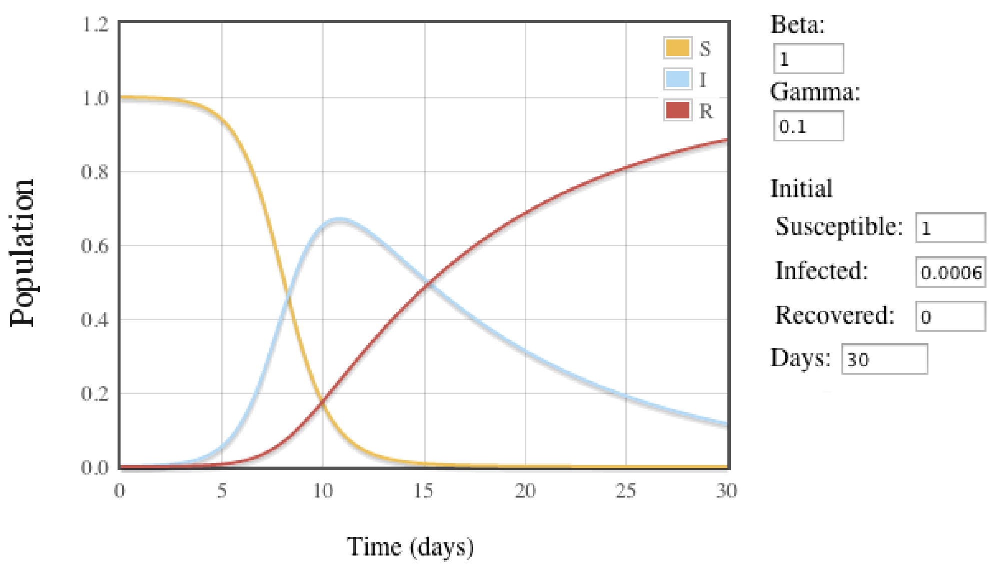


Figure 22 SIR model for Sierra Leone Ebola Death (I) during August 2014.

iBuildApp

iBuildApp is a free online product that allows one to create an app that can be published on the Apple and Google store. The product allows the developer to create a template, select the features of the menu, select a background, select the splash screen and manage the content of each menu item. The following is an image of what appears when the application is loaded on a cellular device.

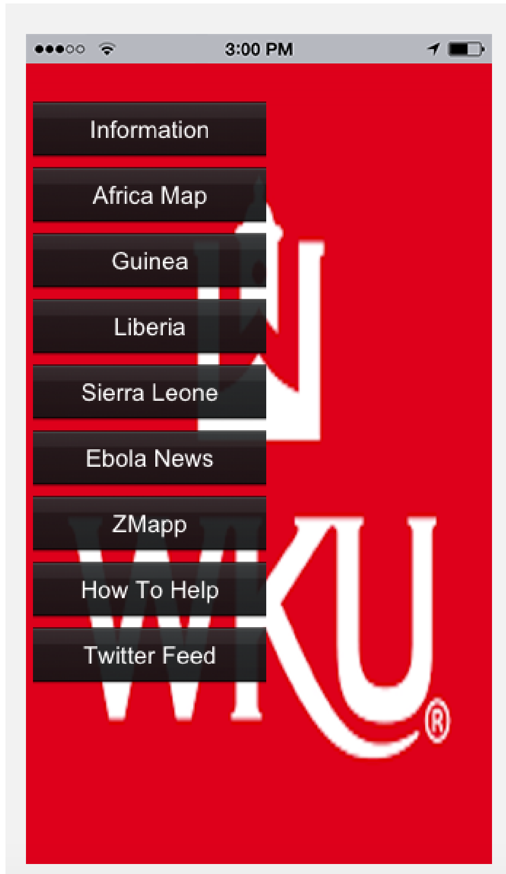


Figure 23 The screen and options available once the application is opened on any cellular device, iPad or iPod.

The information menu provides users with the definition of the virus, information on the risk of exposure, prevention methods, signs and symptoms of the virus, information on the transmission of the virus, how the virus is diagnosed and treated.

The maps tabs allow users to look at the interactive maps created through the use of jsfiddle and Google charts.

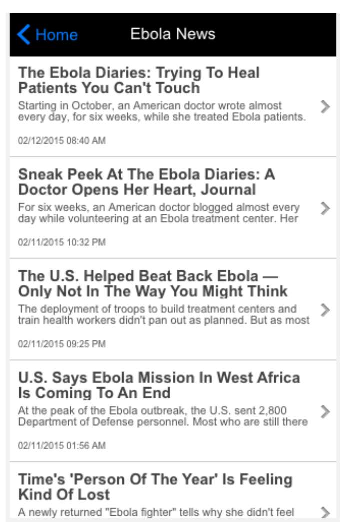
The Ebola News tab, is updated daily and contains relative information about how the virus is spreading, new cases and death counts, what local governments are doing to prevent the spreading of the virus and what international governments are doing to aid the effort of ending the spread of the virus as well as helping provide medical attention to the people infected by the virus.

The SIR Model tab allows users to look at the images created for different time periods and comparing 1-gram memes vs. 2-gram and 3-gram memes.

The Twitter Feed tab allows users to follow and see what the account Ebola Alert is sharing. This account shares the latest news as well as holding daily online conversations on how to go about preventing the spread of the disease.

The ZMapp tab allows user to get the latest information on the production of the vaccine and any side effects that patients may experience from this vaccine.

The How To Help tab takes people the US-SL Ebola Health Care Task Force Coalition webpage which instructs anyone who wants to help on how to do so. It states which supplies are needed the most to combat the virus and where these supplies can be shipped to ensure they reach the front lines.



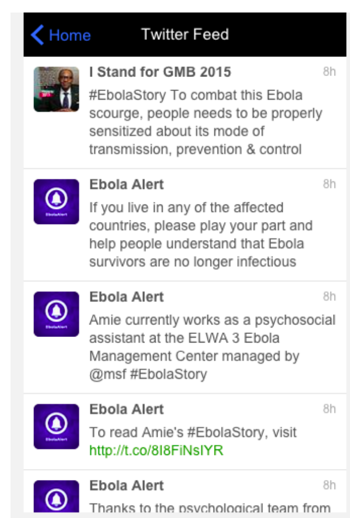


Figure 24 Information

Figure 25 Ebola News

Figure 26 Twitter Feed

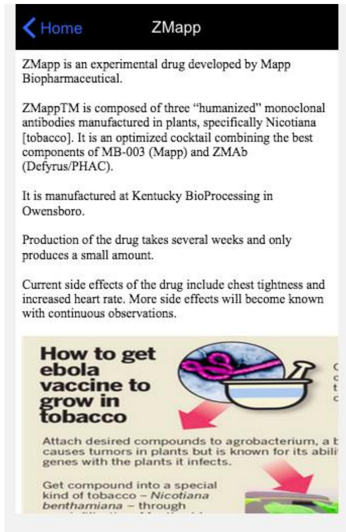


Figure 27 ZMapp Info

Figure 28 How to Help

RESULTS

Through the data mining process it became evident that a number of doctors and hospitals were sharing information about new cases in the area as well as death counts in their communities. While digging further into some of the information shared, spreadsheets were found that kept a running tally of total cases, new cases, and deaths within the three most affected countries. Some of the data collected was shared before even the CDC and WHO began sharing the information in their respective situation reports. Using this data preliminary calculations were done in order to determine the fit coefficients for a power function in order to predict future cases. The following three graphs depict individual power fit results for the three main countries affected by the outbreak: Guinea, Liberia, and Sierra Leone.

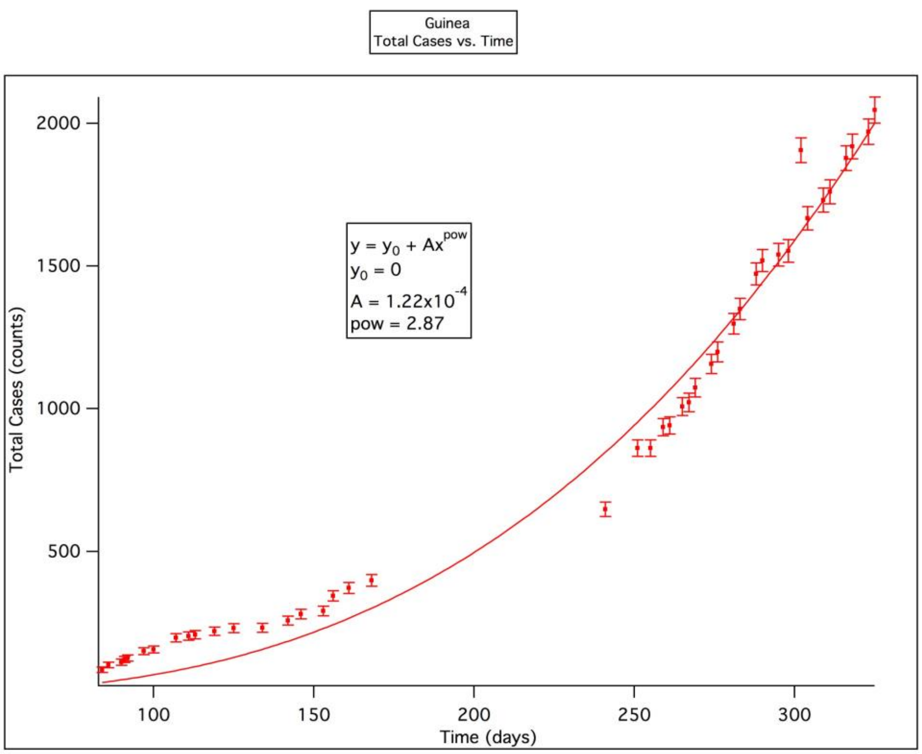


Figure 29 Total Cases vs. Time for Guinea since the start of the Outbreak in 2014.

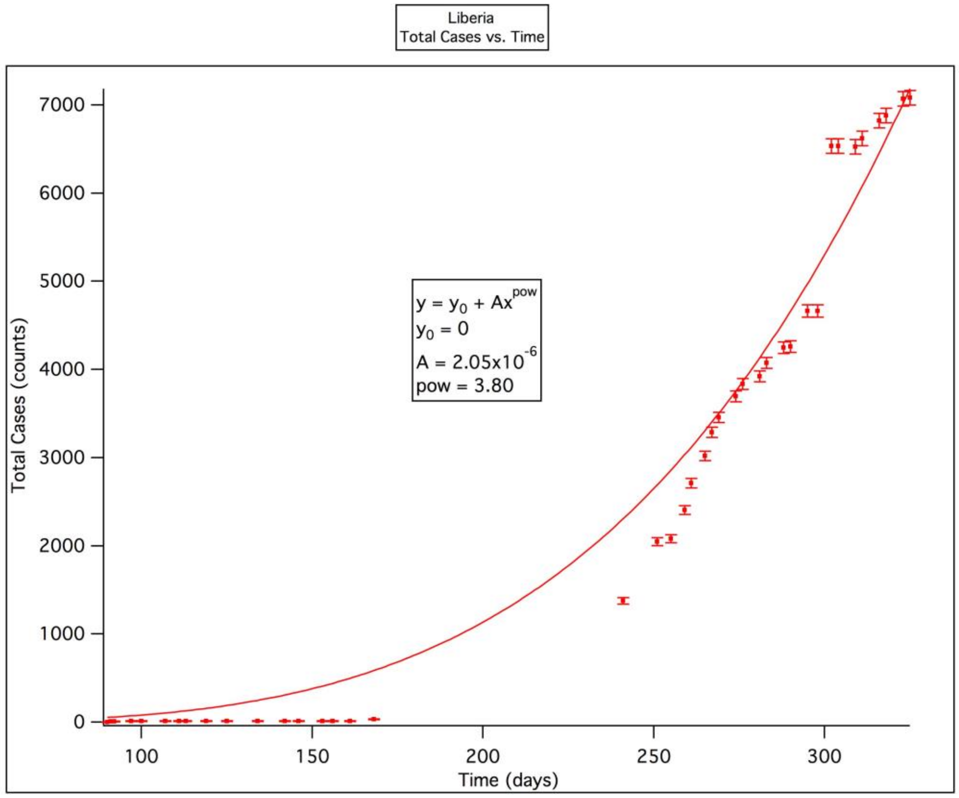


Figure 30 Total Cases vs. Time for Liberia since the start of the Outbreak in 2014.

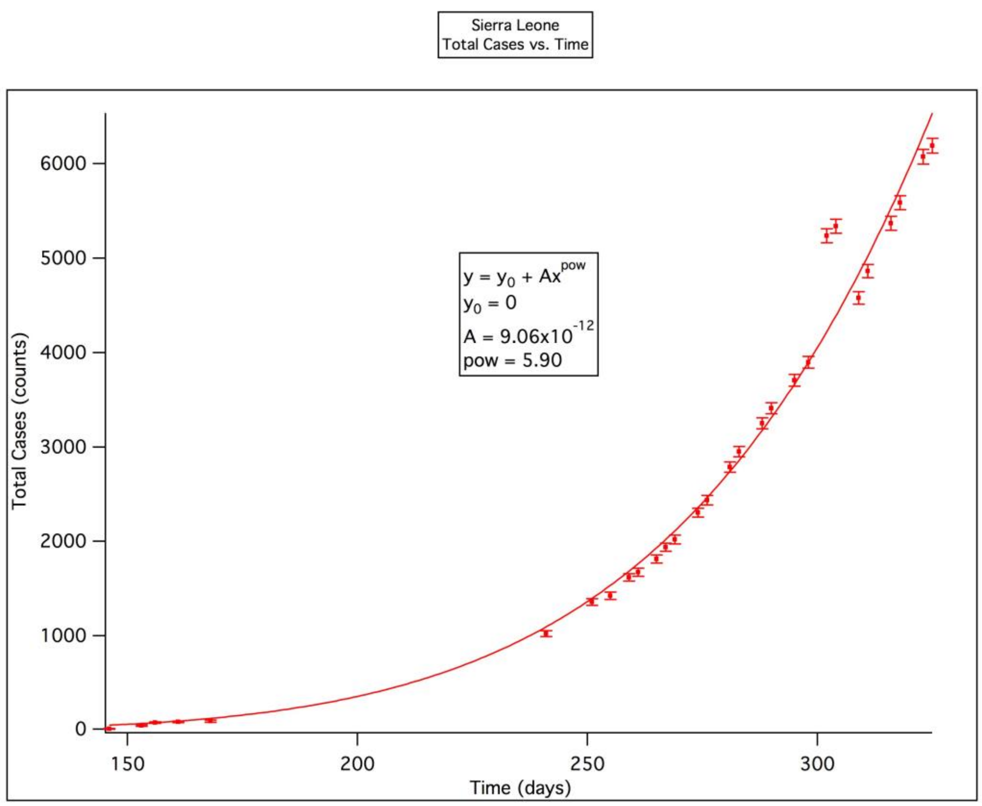


Figure 11 Total Cases vs. Time for Sierra Leone since the start of the Outbreak in 2014.

As is evident by the previous graphs, there is a 100-day gap between points being graphed. This dead space corresponds to the CDC/WHO not sharing any new case information and the previous doctors and hospitals also not sharing anymore information via twitter. This could be due to the fact that during this time period West Africa was working creating a situation report that would in return allow them to receive global assistance in their continuing fight versus the Ebola outbreak.

These three initial graphs give boundary conditions that were used in order to predict the total new cases in the West Africa region. The average power function that was used in order to predict the total cases throughout the first 10 months of the outbreak was Total Cases = $9.8*10^{-7}(x^{4.05})$. Using this new fit parameter a prediction graph was created, this graph was then compared to the data provided by the CDC and WHO in their respective situation reports. The following graph displays these results. Error bars have been placed representing a 10% error. This was done since the true numbers of people whom have died of Ebola is unknown even by the CDC/WHO. This is due to the burial practices and the lack of workers being able to identify the cause of death before a body is buried. By incorporating this 10% error with the CDC/WHO data when comparing the predicted number versus the accepted number, in this case the data provided by the CDC/WHO, it is evident that the predicted cases fall within the 10% error bars. The predictions shift from being over the accepted value to below the accepted value around the 280th day of the observed outbreak.

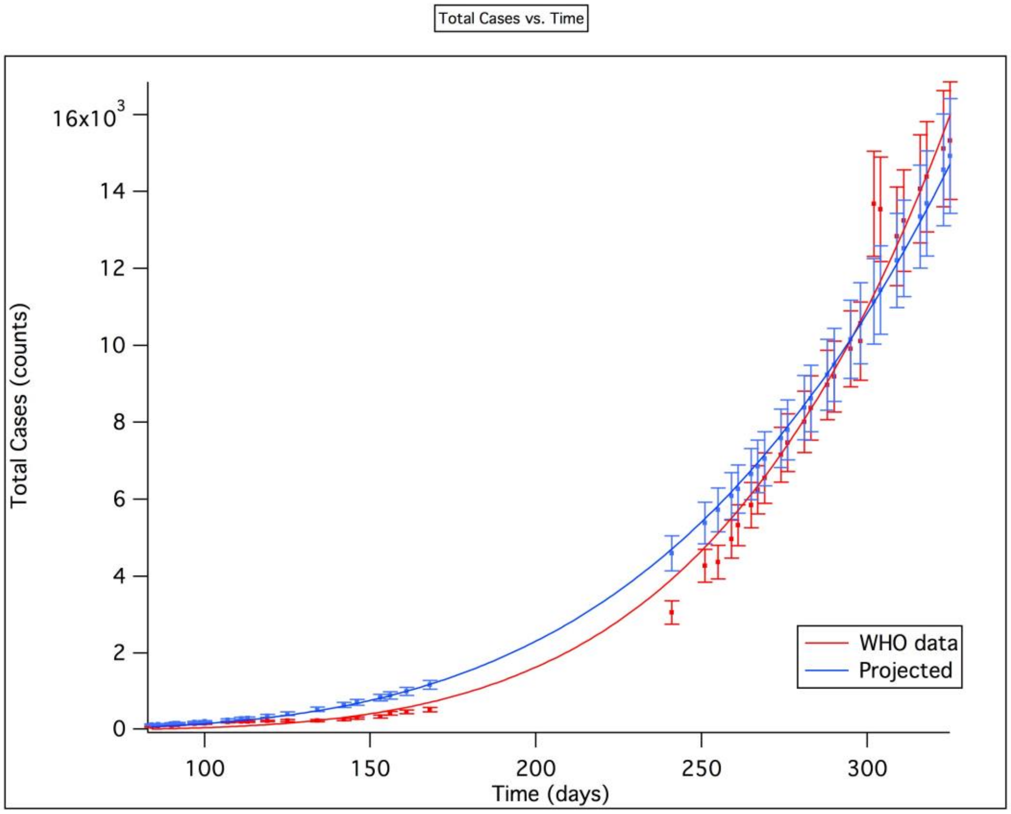


Figure 32 Total Cases vs. Time comparing the Projected values to the accepted values given by the WHO since the start of the Outbreak in 2014.

The data has been interpreted using a static model of disease spreading, for future projects one should consider using a dynamic model to both interpret and predict the spreading of a disease. As was seen with the current model, the power equation was a combination of the results for each individual country affected by the Ebola outbreak. In the future, one could dynamically plot the predictions based on where a majority of the cases were showing up. The larger the initial population the larger the potential infection rate, allowing for a more dynamic approach as time goes on. By allowing for a dynamic change within the prediction stage one would be able to more precisely and accurately predict the affect of a disease.

The following data table depicts a segment of the comparison data. For comparison purposes the root mean square deviation (RMSD) was taken as well as the accuracy between the predicted and accepted total case values.

Days	Predicted Cases	Actual Cases	RMSD	% Accuracy
267	6853	6250	603	91
269	7057	6553	504	93
274	7587	7157	430	94
276	7808	7470	338	96
281	8381	8011	370	96
283	8619	8376	243	97
288	9237	8973	264	97
290	9493	9191	302	97
295	10158	9911	247	98
298	10573	10114	459	96

Table 1 Comparison for the Projected vs. Actual Case Values. Using a RMSD comparison along with a percent accuracy comparison.

The next important step was sharing the information gathered through the creation of the application. The creation of the application has allowed us to create contacts within each country that assist in providing updates which we later confirm with the updates from the CDC to account for accuracy of the information, which to date have been within 10%. Through the interactions with these contacts, it has been well documented that there is a need for better practices of dealing with patients affected by the virus in order to keep the healthcare workers from contracting the virus. As the outbreak continues more healthcare workers are being affected which has a direct correlation to the number of healthcare workers volunteering to go assist in West Africa.

Since the creation of the application in early August 2014, there have been 40+ updates to the application. These updates include the addition of case counts and death counts, changing maps or adding information to current maps, adding maps for cases outside of West Africa, changing information in tabs, and making the application overall more user-friendly and appealing. The application has been downloaded by over 2300 users, combination of Apple and Google store purchases, since its publication date and has been downloaded in the United States, Canada, Europe, Asia Pacific, Latin America and The Caribbean, Africa, The Middle East, and India. The latest update has made the application more static in nature allowing for less information updates but still providing travelers with information needed to remain education on the virus.

The application has received international recognition and has been discussed in numerous newspapers including the Washington Times and Al Jazeera. Most recently, travelpulse.com named the application one of the 5 Travel Apps You Must Own.

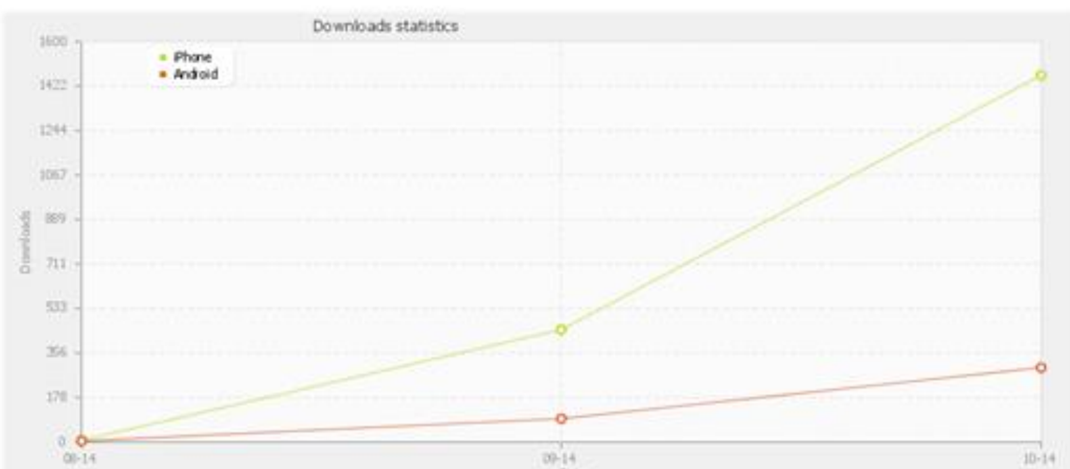


Figure 33 Daily application downloads for a three month period in 2014.

The large increase in download statistics and application launches is a direct correlation to the first Ebola virus case in the United States. The growth was also aided in multiple nationally acknowledged Ebola scares.

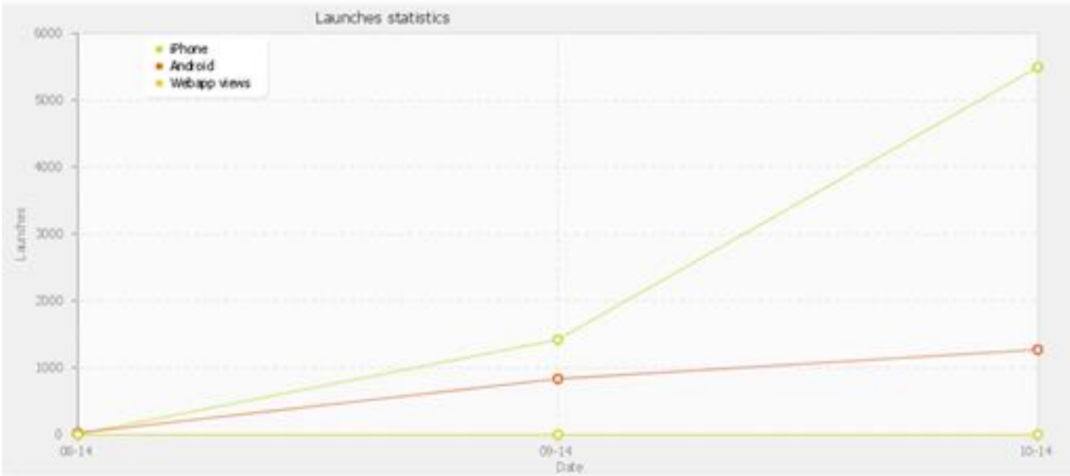


Figure 34 Daily application launches for a three month period in 2014.

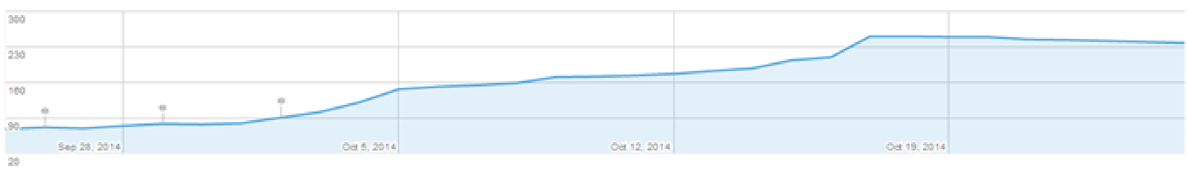


Figure 35 Amount of application installations from the Google Play store in 2014.

Users in the United States on the Google Play store only downloaded the application until the end of September, after which it began international, downloads. The peak corresponds to a press release that received national attention and therefore sparked an interest in the application. On the Apple Store the application was downloaded from the beginning of its publication and has seen increases and decreases in weekly downloads that correlate to Ebola events in the news.

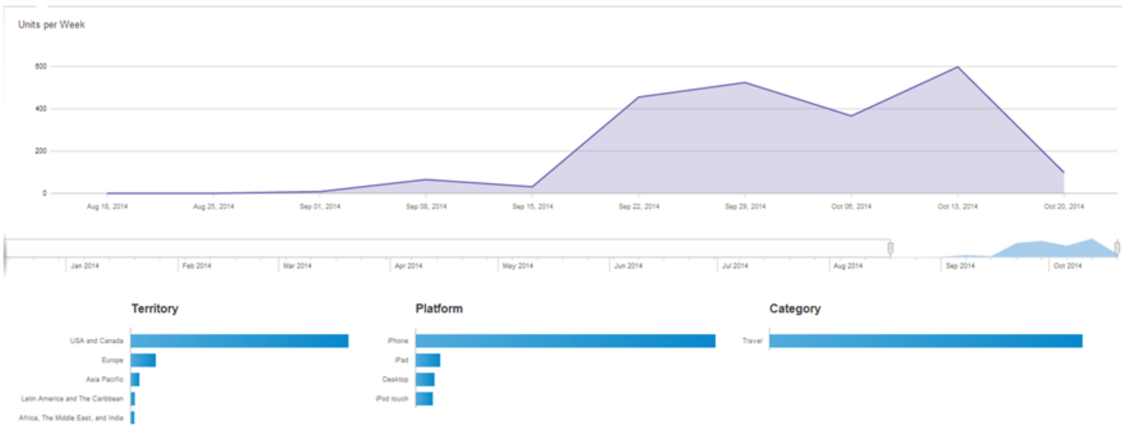


Figure 36 Amount of application installations from the Apple store in 2014.

The addition of ads within the application allows us to generate revenue without having to charge users to download the application. This change, increased application downloads by 700%. The following graph shows the revenue generated since ads were added to the application.

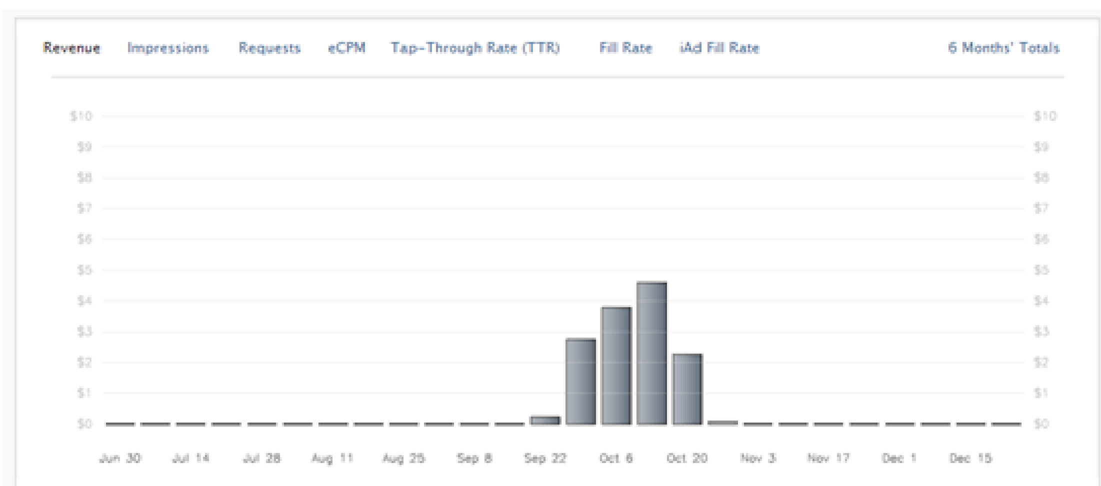


Figure 37 Revenue generated through ads within the application in 2014.

Ads within the application allow for continuous revenue generation throughout the lifetime of the application. As long as users are opening the application to look at updated information revenue is being generated.

Currently the application is still available on both the Apple and Google store for download and has recently reached 2500+ downloads as well as 12000+ application views. We continue to monitor and update the application with the most up to date case confirmations and deaths throughout the West African region. We are maintaining relationships with nursing associations in order to share more up to date information and provide people with information on how to help fight the Ebola virus.

CONCLUSION

These results were used in the continuous prediction process of the spreading of the Ebola virus in West Africa. The predictions were used in order to determine the spread of the virus into surrounding countries as well as predictions on the new amount of cases each week. These predictions were then compared to the values posted by the Centers for Disease Control and Prevention and the World Health Organization. The predictions during the first couple of months for the spread of the Ebola virus were within 10% of the accepted value, which was the value given by the Centers for Disease Control and Prevention and the World Health Organization. (59)

Since the Ebola virus has crossed continental borders the background noise has increased exponentially and in turn has created difficulties in the data mining process. As an example before the virus was feared of spreading throughout the United States, the average amount of data mined was 6.5 million Tweets; but after fear of the virus spreading throughout the United States the average amount of data mined was 20.3 million Tweets. (59)

The reduction background noise over time will continue to improve the data set and in turn increase the accuracy of the predictions. The one single factor that would increase the accuracy of the case predictions and possible tracking of any virus would be for all Twitter users to enable GPS location services, since currently only 3% of Twitter users have GPS location services enabled a plethora of data is not being used for analysis. (59)

By enabling GPS services more data could be collected allowing for more accurate tracking and possible prevention of any virus spreading throughout a population.

Current software allows for 78% accuracy on predicting where a Tweet originated from, this is based on the country. For example if one were to try and data mine Tweets in the United States based on states that 78% accuracy would drop to 24%. New efforts are underway to understand the nature of the Ebola spread and resiliency. (59)

Models are being developed to examine community interactions, the effects of several strains, time dependent couplings, genuine wave equation interactions, etc. Many of these efforts are based upon first person accounts collected by workers in the region who had firsthand knowledge and exposure to the community during the efforts to institute treatments. (59)

The nature of the Ebola carriers has also complicated matters with containment where it appears that several species of bats, especially the common fruit bat, have a genome that makes them excellent carriers. Also of great service is the development of a vaccine now in clinical trials that will impact the spread of the disease and the nature of modeling the epidemic across the world. The social media archives will provide a rich data set to assist in the reconstruction of our understanding of the spread of the disease and wills serve as a permanent and accessible aid in our understanding of the transmission of information. (59)

We have been able to create an application that users can download for free from both the Apple and Google store in order to follow along with the progress of the Ebola outbreak. The application also provides users with information on how the precautions they need to take when traveling to any area affected by the Ebola virus as well as information on the symptoms they need to look for if they have traveled to an area affected by the Ebola virus. The application will continue to get updated weekly until August 2015, after which it will be made into a static informational tool.

Future Work

New efforts are underway to understand the nature of the Ebola spread and resiliency. Models are being developed to examine community interactions, the effects of several strains, time dependent couplings, genuine wave equation interactions, etc. Many of these efforts are based upon first person accounts collected by workers in the region who had firsthand knowledge and exposure to the community during the efforts to institute treatments. The nature of the Ebola carriers has also complicated matters with containment where it appears that several species of bats, especially the common fruit bat, have a genome that makes them excellent carriers. Also of great service is the development of a vaccine now in clinical trials that will impact the spread of the disease and the nature of modeling the epidemic across the world. The social media archives will provide a rich data set to assist in the reconstruction of our understanding of the spread of the disease and wills serve as a permanent and accessible aid in our understanding of the transmission of information. (59)

Currently we are trying to use the model proposed by Weng, Wang and Xu (59, 71) to turn our ordinary partial differential equation into a partial differential equation that incorporates both temporal and spatial features into the modeling. This model has been described in my previously published paper, "Sample NLPDE and NLODE Social-Media Modeling of Information Transmission for Infectious Diseases: Case Study Ebola (59)."

REFERENCES

1. About Ebola Virus Disease. (2014, December 8). Retrieved April 6, 2015, from http://www.cdc.gov/vhf/ebola/about.html
2. Aggarwal, A., & Ponnurangam, K. (2014, August 7). Followers or Phantoms? An Anatomy of Purchased Twitter Followers. Retrieved April 6, 2015, from http://arxiv.org/abs/1408.1534
3. Aggarwal, A., & Ponnurangam, K. (2014, November 16). What They Do in Shadows: Twitter Underground Follower Market. Retrieved April 6, 2015, from http://arxiv.org/abs/1411.4299
4. Agreste, S., De Meo, P., Ferrara, E., Piccolo, S., & Provetti, A. (2014, February 7). Analysis of a heterogeneous social network of humans and cultural objects. Retrieved April 6, 2015, from http://arxiv.org/abs/1402.1778
5. Aguiar, M., & Stollenwerk, N. (2007, April 24). A new chaotic attractor in a basic multi-strain epidemiological model with temporary cross-immunity. Retrieved April 6, 2015, from http://arxiv.org/abs/0704.3174
6. Aspirot, L., Mordecki, E., & Rubino, G. (2011, December 1). Fluid Limits Applied to Peer to Peer Network Analysis. Retrieved April 6, 2015, from https://hal.inria.fr/hal-00789636/en
7. Bakshy, E., Hofman, J., Mason, W., & Watts, D. (2011, January 1). Everyone's an influencer: Quantifying influence on twitter. Retrieved April 6, 2015, from http://dl.acm.org/citation.cfm?id=1935845
8. Barry, M., Traore, F., Sako, F., Bah, E., Poncin, M., Keita, S., . . . Toure, A. (2014, October 23). Ebola outbreak in Conakry, Guinea: Epidemiological, clinical, and

outcome features. Retrieved April 6, 2015, from http://www.ncbi.nlm.nih.gov/pubmed/25391486

9. Bortolussi, L., & Hillston, J. (2012, September 20). Fluid Approximation of CTMC with Deterministic Delays. Retrieved April 6, 2015, from http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6354633
10. Cataldi, L., Di Caro, L., & Schifanella, C. (2010, January 1). Emerging topic detection on Twitter based on temporal and social terms evaluation. Retrieved April 6, 2015, from http://dl.acm.org/citation.cfm?id=1814249
11. Cheng, Z., Caverlee, J., & Lee, K. (2010, October 26). You are where you tweet: A content-based approach to geo-locating twitter users. Retrieved April 6, 2015, from http://dl.acm.org/citation.cfm?id=1871535
12. Chew, C., & Eysenbach, G. (2010, November 29). Pandemics in the Age of Twitter: Content Analysis of Tweets during the 2009 H1N1 Outbreak. Retrieved April 6, 2015, from http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0014118
13. Conover, M., Davis, C., Ferrara, E., McKevley, K., Menczer, F., & Flammini, A. (2013, June 23). The Geospatial Characteristics of a Social Movement Communication Network. Retrieved April 6, 2015, from http://arxiv.org/abs/1306.5473
14. Conover, M., Ferrara, E., Menczer, F., & Flammini, A. (2013, May 29). The Digital Evolution of Occupy Wall Street. Retrieved April 6, 2015, from http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0064679
15. Dawkins, R. (2006). The selfish gene (New ed., Vol. 199). Oxford: Oxford University Press.

16. De Meo, P., Ferrara, E., Abel, F., Aroyo, L., & Houben, G. (2013, October 16). Analyzing User Behavior across Social Sharing Environments. Retrieved April 6, 2015, from http://arxiv.org/abs/1310.4399
17. Del Rio, C., Mehta, A., Lyon, M., & Guarner, J. (2014, November 18). Ebola Hemorrhagic Fever in 2014: The Tale of an Evolving EpidemicEbola Hemorrhagic Fever in 2014 FREE. Retrieved April 6, 2015, from http://annals.org/article.aspx?articleid=1897363
18. Dye, C. (2003, June 20). Modeling the SARS epidemic. (Epidemiology). Science, 1884-1885.
19. Ferrara, E., & Fiumara, G. (2012, February 2). Topological Features of Online Social Networks. Retrieved April 6, 2015, from http://arxiv.org/abs/1202.0331
20. Ferrara, E., JafariAsbagh, M., Varol, O., Qazvinian, V., Menczer, F., & Flammini, A. (2014, November 3). Clustering memes in social media streams. Retrieved April 6, 2015, from http://arxiv.org/abs/1411.0652
21. Finn, S., Takis Metaxas, P., & Mustafaraj, E. (2014, November 13). Investigating Rumor Propagation with TwitterTrails. Retrieved April 6, 2015, from http://arxiv.org/abs/1411.3550
22. Franca, U., Sayama, H., McSwiggen, C., Daneshvar, R., & Bar-Yam, Y. (2014, November 3). Visualizing the "Heartbeat" of a City with Tweets. Retrieved April 6, 2015, from http://arxiv.org/abs/1411.0722
23. Fusaroli, R., Perlman, M., Mislove, A., Paxton, A., Matlock, T., & Dale, R. (2014, October 29). Timescales of Massive Human Entrainment. Retrieved April 6, 2015, from http://arxiv.org/abs/1410.8105

24. Hahn, L. (2011, January 1). The End is Near: The Statistical Regularities of Sentence Endings. Retrieved April 6, 2015.

25. Hahn, L. (2012, June 1). Measuring local context as context-word probabilities. Retrieved April 6, 2015, from http://www.ncbi.nlm.nih.gov/pubmed/21960060

26. Hahn, L., & Sivley, R. (2011, September 1). Entropy, semantic relatedness and proximity. Retrieved April 6, 2015, from http://www.ncbi.nlm.nih.gov/pubmed/21557006

27. Haustein, S., Lariviere, V., Thelwall, M., Amyot, D., & Peters, I. (2014, October 2). Tweets vs. Mendeley readers: How do these two social media metrics differ? Retrieved April 6, 2015, from http://arxiv.org/abs/1410.0569

28. How to Build an Iphone App in 5 min. (2011, January 1). Retrieved April 6, 2015, from http://ibuildapp.com/howto/

29. Hu, Y., Manikonda, L., & Kambhampati, S. (2014, January 1). What we Instagram: A First Analysis of Instagram Photo Content and User Types. Retrieved April 6, 2015, from http://www.public.asu.edu/~yuhenghu/paper/icwsm14.pdf

30. Jin, F., Dougherty, E., Saraf, P., Cao, Y., & Ramakrishnan, N. (2013, January 1). Epidemiological Modeling of News and Rumors on Twitter. Retrieved April 6, 2015, from http://people.cs.vt.edu/naren/papers/news-rumor-epi-snakdd13.pdf

31. Kermack, W., & McKendrick, A. (1927). A Contribution To The Mathematical Theory Of Epidemics. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 115, 700-721.

32. Khan, A., Tshioko, K., & Heymann, D. (1999, February 1). The reemergence of Ebola hemorrhagic fever, Democratic Republic of the Congo, 1995. Retrieved April 6, 2015, from http://www.ncbi.nlm.nih.gov/pubmed/9988168
33. Khan, M. (2014). A Computer Virus Propagation Model Using Delay Differential Equations with Probabilistic Contagion and Immunity. International Journal of Computer Networks & Communications, 111-128.
34. King, A., De Celles, M., Magpantay, F., & Rohani, P. (2014, December 2). Avoidable errors in the modeling of outbreaks of emerging pathogens, with special reference to Ebola. Retrieved April 6, 2015, from http://arxiv.org/abs/1412.0968
35. Kuhn, T., Perc, M., & Helbing, D. (2014, January 1). Inheritance Patterns in Citation Networks Reveal Scientific Memes. Retrieved April 6, 2015, from http://www.academia.edu/9461966/Inheritance_Patterns_in_Citation_Networks_Reveal_Scientific_Memes
36. Kulkarni, V., Al-Rfou, R., Perozzi, B., & Skiena, S. (2014, November 12). Statistically Significant Detection of Linguistic Change. Retrieved April 6, 2015, from http://arxiv.org/abs/1411.3315
37. Kwak, H., Lee, C., Park, H., & Moon, S. (2010, January 1). What is Twitter, a social network or a news media? Retrieved April 6, 2015, from http://dl.acm.org/citation.cfm?id=1772751
38. Lamb, A., Paul, M., & Dredze, M. (2013, January 1). Separating Fact from Fear: Tracking Flu Infections on Twitter. Retrieved April 6, 2015, from http://site.uottawa.ca/~nat/Courses/csi5387_Winter2014/paper15.pdf

39. Lehmann, J., Goncalves, B., Ramasco, J., & Cattuto, C. (2011, November 8). Dynamical Classes of Collective Attention in Twitter. Retrieved April 6, 2015, from http://arxiv.org/abs/1111.1896

40. Li, J., & Cardie, C. (2013, September 17). Early Stage Influenza Detection from Twitter. Retrieved April 6, 2015, from http://arxiv.org/abs/1309.7340

41. Li, J., & Cardie, C. (2013, September 27). Timeline Generation: Tracking individuals on Twitter. Retrieved April 6, 2015, from http://arxiv.org/abs/1309.7313

42. Li, J., Ritter, A., & Jurafsky, D. (2014, November 11). Inferring User Preferences by Probabilistic Logical Reasoning over Social Networks. Retrieved April 6, 2015, from http://arxiv.org/abs/1411.2679

43. Liu, J., Zhao, K., Khan, S., Cameron, M., & Jurdak, R. (2014, December 1). ArXiv.org cs arXiv:1412.0327. Retrieved April 6, 2015, from http://arxiv.org/abs/1412.0327

44. Luksza, M., Bedford, T., & Lassig, M. (2014, November 6). Epidemiological and evolutionary analysis of the 2014 Ebola virus outbreak. Retrieved April 6, 2015, from http://arxiv.org/abs/1411.1722

45. Manikonda, L., Hu, Y., & Kambhampati, S. (2014, October 29). Analyzing User Activities, Demographics, Social Network Structure and User-Generated Content on Instagram. Retrieved April 6, 2015, from http://arxiv.org/abs/1410.8099

46. Mkhatshwa, T., & Mummert, A. (2010, October 15). Modeling Super-spreading Events for Infectious Diseases: Case Study SARS. Retrieved April 6, 2015, from http://arxiv.org/pdf/1007.0908.pdf

47. Morales, A., Borondo, J., Losado, J., & Benito, R. (2014, November 3). Efficiency of Human Activity on Information Spreading on Twitter. Retrieved April 6, 2015, from http://arxiv.org/abs/1411.0473

48. Nematzadeh, A., Ferrara, E., Flammini, A., & Ahn, Y. (2014, January 7). Optimal network modularity for information diffusion. Retrieved April 6, 2015, from http://arxiv.org/abs/1401.1257

49. Newman, M. (2002). Spread of epidemic disease on networks. Physical Review E, 66(1), 016128.

50. Rajtmajer, S., Griffin, C., Mikesell, D., & Squicciarini, A. (2014, August 12). A cooperate-defect model for the spread of deviant behavior in social networks. Retrieved April 6, 2015, from http://arxiv.org/abs/1408.2770

51. Rayson, S. (2013, September 3). 6 Reasons To Use Topsy - The Free Twitter Social Analytics Tool -. Retrieved April 6, 2015, from http://anderspink.com/6-reasons-to-use-topsy-the-free-twitter-social-analytics-tool/

52. Robin, J., & Mammadov, M. (2014, December 4). Dynamics of Ebola epidemics in West Africa 2014. Retrieved April 6, 2015, from http://arxiv.org/abs/1412.1579

53. Schoen, H., Gayo-Avello, D., Metaxas, P., Mustafaraj, E., Strohmaier, M., & Gloor, P. (2013). The power of prediction with social media. Internet Research, 23(5), 528-543.

54. Segel, L. (2003). J. D. Murray: Mathematical Biology (3rd Ed), Volume I (An Introduction) and Volume II (Spatial Models and Biomedical Applications). Mathematical Medicine and Biology, 377-378.

55. Seth, A., Purohit, H., Jadhav, A., Kapanipathi, P., & Chen, L. (2010). Understanding Events through Analysis of Social Media. Retrieved April 6, 2015, from http://corescholar.libraries.wright.edu/cgi/viewcontent.cgi?article=2164&context=knoesis

56. Shah, D., & Zaman, T. (2011). Rumors in a Network: Who's the Culprit? IEEE Transactions on Information Theory, 57(8), 5163-5181.

57. Shulgin, B. (1998). Pulse vaccination strategy in the SIR epidemic model. Bulletin of Mathematical Biology, 60(6), 1123-1148.

58. Signorini, A., Segre, A., Polgreen, P., & Galvani, A. (2011). The Use of Twitter to Track Levels of Disease Activity and Public Concern in the U.S. during the Influenza A H1N1 Pandemic. PLoS ONE, 6(5), E19467-E19467.

59. Smailhodzic, A., Andrew, K., Hahn, L., Womble, P., & Webb, C. (2014, December 31). Sample NLPDE and NLODE Social-Media Modeling of Information Transmission for Infectious Diseases:Case Study Ebola. Retrieved April 10, 2015, from http://arxiv.org/abs/1501.00198

60. Smailhodzic, A., Taylor, M., Andrew, K., Hahn, L., Womble, P., Webb, C., & Thompson, B. (2014). A System of Coupled ODEs as a Cyber Model for Analyzing Wavelike Information Transmission from Data Mining Tweets. Bulletin of the American Physical Society, 59.

61. Small, M., Tse, C., & Walker, D. (2006). Super-spreaders and the rate of transmission of the SARS virus. Physica D: Nonlinear Phenomena, 215(2), 146-158.

62. Smith, B. (2012, June 5). The Beginner's Guide to Twitter. Retrieved October 9, 2014, from http://mashable.com/2012/06/05/twitter-for-beginners/

63. Takis Metaxas, P., Mustafaraj, E., Wong, K., Zeng, L., O'Keefe, M., & Finn, S. (2014, November 13). Do Retweets indicate Interest, Trust, Agreement? Retrieved April 6, 2015, from http://arxiv.org/abs/1411.3555

64. Team, W. (2014). Ebola Virus Disease in West Africa — The First 9 Months of the Epidemic and Forward Projections. New England Journal of Medicine, 371(16), 1481-1485.

65. Thelwall, M., Haustein, S., Larivière, V., Sugimoto, C., & Bornmann, L. (2013). Do Altmetrics Work? Twitter and Ten Other Social Web Services. PLoS ONE, 5, E64841-E64841.

66. Topsy Labs. (n.d.). Retrieved October 29, 2014, from https://www.crunchbase.com/organization/topsy-labs

67. Using Google Charts. (2013, May 15). Retrieved October 29, 2014, from https://developers.google.com/chart/interactive/docs/index

68. Varol, O., Ferrara, E., Ogan, C., Menczer, F., & Flammini, A. (2014, June 27). Evolution of Online User Behavior During a Social Upheaval. Retrieved April 6, 2015, from http://arxiv.org/abs/1406.7197

69. Venkatramanan, S., & Kumar, A. (2010, February 6). New Insights from an Analysis of Social Influence Networks under the Linear Threshold Model. Retrieved April 6, 2015, from http://arxiv.org/abs/1002.1335

70. Venkatramanan, S., & Kumar, A. (2014, May 28). Influence Spread in Social Networks: A Study via a Fluid Limit of the Linear Threshold Model. Retrieved April 6, 2015, from http://arxiv.org/abs/1405.7096

71. Weng, F., Wang, H., Xu, K., Wu, J., & Jia, X. (2013, July 8). Characterizing Information Diffusion in Online Social Networks with Linear Diffusive Model. Retrieved April 6, 2015, from http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=6681600&url=http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6681600
72. Weng, L., Flammini, A., & Vespignani, A. (2013, October 1). Modeling Information Diffusion in Online Social Networks with Partial Differential Equations. Retrieved April 6, 2015, from http://arxiv.org/abs/1310.0505
73. Weng, L., Flammini, A., Vespignani, A., & Menczer, F. (2012). Competition among memes in a world with limited attention. Scientific Reports, 1-8.
74. Weng, L., Ratkiewicz, J., Perra, N., Goncalves, B., Castillo, C., Bonchi, F., & Schifanella, R. (2013, February 25). The Role of Information Diffusion in the Evolution of Social Networks. Retrieved April 6, 2015, from http://arxiv.org/abs/1302.6276
75. Wolfram Mathematica. (n.d.). Retrieved October 28, 2014, from http://www.wolfram.com/mathematica/
76. Wright, C. (2014). Using Statistical Methods to Determine Geolocation Via Twitter. Masters Theses & Specialist Projects, Paper 1372.
77. Wright, C., Hahn, L., Womble, P., & Andrew, K. (2013). Geolocation Analysis and Social Median Data Acquisition via Twitter Content. Bulleting of the American Physical Society, 58.
78. Yang, J., & Counts, S. (2010). Predicting the Speed, Scale, and Range of Information Diffusion in Twitter. ICWSM, 10(2010), 355-358.

79. Zaman, T. R., Herbrich, R., Van Gael, J., & Stern, D. (2010, December). Predicting information spreading in twitter. In Workshop on Computational Social Science and the Wisdom of Crowds. NIPS, 104(45), 17599-17601.