Mathematical Modeling of Trending Topics on Twitter

Jonathan Skaza

April 14, 2015
Comparison of Trending Topics

“Window” refers to the moving sum period (e.g., each point represents count in past 200s)
Project Objectives

• Quantify the diffusion of information on Twitter
Project Objectives

- Quantify the diffusion of information on Twitter
- Compare and contrast different trending topics
Project Objectives

• Quantify the diffusion of information on Twitter
• Compare and contrast different trending topics
• Compare spread of information to spread of infectious disease
Project Objectives

• Quantify the diffusion of information on Twitter
• Compare and contrast different trending topics
• Compare spread of information to spread of infectious disease
• Create a reproducible output product
Agenda

1. Twitter overview, facts, and figures
Agenda

1. Twitter overview, facts, and figures
2. Methodological strategy
Agenda

1. Twitter overview, facts, and figures
2. Methodological strategy
3. Previous studies
1. Twitter overview, facts, and figures
2. Methodological strategy
3. Previous studies
4. Results and discussion
About Twitter

• Created in 2006
About Twitter

- Created in 2006
- Incorporated in 2007

Source: about.twitter.com
About Twitter

• Created in 2006
• Incorporated in 2007
• 288 million monthly active users – 500 million Tweets per day
About Twitter

• Created in 2006
• Incorporated in 2007
• 288 million monthly active users – 500 million Tweets per day
• Source: about.twitter.com
Anatomy of a Tweet

Fun fact: Since field was expanded to 64 teams, average sum of seeds in Final Four has been 11 (this year, it's 10) #MarchMadness

12:24 PM - 31 Mar 2015

1 RETWEET
Twitter Application Programming Interface (API)

- Two different flavors: REST and Streaming
Streaming API

User makes request

Server pulls processed result from data store and renders view

Server opens streaming connection

Twitter accepts connection

Recieves streamed Tweets, performs processing and stores result

Tweets streamed as they occur

Connection closes

Connection closes

Source: dev.twitter.com
# Streaming API Request Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>delimited</td>
<td>locations</td>
</tr>
<tr>
<td>stall_warnings</td>
<td>count</td>
</tr>
<tr>
<td>filter_level</td>
<td>with</td>
</tr>
<tr>
<td>language</td>
<td>replies</td>
</tr>
<tr>
<td>follow</td>
<td>stringify_friend_id</td>
</tr>
<tr>
<td>track</td>
<td></td>
</tr>
</tbody>
</table>

Source: Twitter Developers Documentation

```python
stream.filter(track=['#'])
```
Data Collection

Raw Tweet

{
"created_at":"Fri Mar 27 18:16:52 +0000 2015","id":581520276292280320,"id_str":"581520276292280320","text":
"Loving the #NCAA #MarchMadness? Find out fun facts like which states listened most, overall listening hours
and more! http://t.co/DWfDTDnDg8","source":"\u003cahref="http://twitter.com\" rel="nofollow">
Twitter Web Client\u003c/a\u003e","truncated":false,"in_reply_to_status_id":null,"in_reply_to_status_id_str":null,"in_reply_to_user_id":null,"in_reply_to_user_id_str":null,"in_reply_to_screen_name":null,"user":{"id":1694596596,"id_str":"1694596596","name":"Westwood One","screen_name":"WestwoodOne","location":"In your speakers","url":"http://www.westwoodone.com","description":"Westwood One offers audio products and content to
reach listeners whenever, wherever they are. #powerofsound","protected":false,"verified":true,"followers_count":1123,"friends_count":337,"listed_count":24,"statuses_count":1923,"created_at":
Data Collection

Processed Tweet

Fri Mar 27 18:16:52 +0000 2015, ['NCAA', 'MarchMadness']
Methodology

Sample Twitter

ID Topics

Fit Models

Disease

Hashtags

Prediction

Analyze Results

Jonathan Skaza

Mathematical Modeling of Trending Topics on Twitter
SIR Model

Developed by Kermack and McKendrick (1927)

**Disease**: Proximal to infected individual → Catch disease → Recover from disease

**Meme**: Twitter user → Tweet about topic → Move on in life
Use Markov Chain Monte Carlo (MCMC) simulation techniques to estimate $\beta$, $\gamma$, initial $S$, and initial $I$ (Coelho, Codeco, and Gomes, 2011)
Previous Work

- Infectious Disease
  - Measles (McGilchrist et al., 1996; Grais et al., 2006; Tuckwell and Williams, 2007; Kuniya, 2006)
  - Influenza (Tuckwell and Williams, 2007; Li, Li, and Ghosh, 2009; Hooten, Anderson, and Waller, 2010; Coelho, Codeco, and Gomes, 2011)

- Other Applications
  - Feynman diagrams (Bettencourt et al., 2006)
  - News and rumors on Twitter (Jin et al., 2013)
  - Spread of rumors in social networks (Zhao et al., 2012)
  - Zombie apocalypse (Witkowski and Blais, 2013)
Previous Work

- Infectious Disease
  - Measles (McGilchrist et al., 1996; Grais et al., 2006; Tuckwell and Williams, 2007; Kuniya, 2006)

- Other Applications
  - Feynman diagrams (Bettencourt et al., 2006)
  - News and rumors on Twitter (Jin et al., 2013)
  - Spread of rumors in social networks (Zhao et al., 2012)
  - Zombie apocalypse (Witkowski and Blais, 2013)
Previous Work

• **Infectious Disease**
  - Measles (McGilchrist et al., 1996; Grais et al., 2006; Tuckwell and Williams, 2007; Kuniya, 2006)
  - Influenza (Tuckwell and Williams, 2007; Li, Li, and Ghosh, 2009; Hooten, Anderson, and Waller, 2010; Coelho, Codeco, and Gomes, 2011)

• **Other Applications**
  - Feynman diagrams (Bettencourt et al., 2006)
  - News and rumors on Twitter (Jin et al., 2013)
  - Spread of rumors in social networks (Zhao et al., 2012)
  - Zombie apocalypse (Witkowski and Blais, 2013)
Previous Work

- Infectious Disease
  - Measles (McGilchrist et al., 1996; Grais et al., 2006; Tuckwell and Williams, 2007; Kuniya, 2006)
  - Influenza (Tuckwell and Williams, 2007; Li, Li, and Ghosh, 2009; Hooten, Anderson, and Waller, 2010; Coelho, Codeco, and Gomes, 2011)
- Other Applications
  - Feynman diagrams (Bettencourt et al., 2006)
  - News and rumors on Twitter (Jin et al., 2013)
  - Spread of rumors in social networks (Zhao et al., 2012)
  - Zombie apocalypse (Witkowski and Blais, 2013)
Previous Work

- **Infectious Disease**
  - Measles (McGilchrist et al., 1996; Grais et al., 2006; Tuckwell and Williams, 2007; Kuniya, 2006)
  - Influenza (Tuckwell and Williams, 2007; Li, Li, and Ghosh, 2009; Hooten, Anderson, and Waller, 2010; Coelho, Codeco, and Gomes, 2011)

- **Other Applications**
  - Feynman diagrams (Bettencourt et al., 2006)
  - News and rumors on Twitter (Jin et al., 2013)
  - Spread of rumors in social networks (Zhao et al., 2012)
  - Zombie apocalypse (Witkowski and Blais, 2013)
Previous Work

- **Infectious Disease**
  - Measles (McGilchrist et al., 1996; Grais et al., 2006; Tuckwell and Williams, 2007; Kuniya, 2006)
  - Influenza (Tuckwell and Williams, 2007; Li, Li, and Ghosh, 2009; Hooten, Anderson, and Waller, 2010; Coelho, Codeco, and Gomes, 2011)

- **Other Applications**
  - Feynman diagrams (Bettencourt et al., 2006)
  - News and rumors on Twitter (Jin et al., 2013)
Previous Work

- **Infectious Disease**
  - Measles (McGilchrist et al., 1996; Grais et al., 2006; Tuckwell and Williams, 2007; Kuniya, 2006)
  - Influenza (Tuckwell and Williams, 2007; Li, Li, and Ghosh, 2009; Hooten, Anderson, and Waller, 2010; Coelho, Codeco, and Gomes, 2011)

- **Other Applications**
  - Feynman diagrams (Bettencourt et al., 2006)
  - News and rumors on Twitter (Jin et al., 2013)
  - Spread of rumors in social networks (Zhao et al., 2012)
## Previous Work

- **Infectious Disease**
  - Measles (McGilchrist et al., 1996; Grais et al., 2006; Tuckwell and Williams, 2007; Kuniya, 2006)
  - Influenza (Tuckwell and Williams, 2007; Li, Li, and Ghosh, 2009; Hooten, Anderson, and Waller, 2010; Coelho, Codeco, and Gomes, 2011)

- **Other Applications**
  - Feynman diagrams (Bettencourt et al., 2006)
  - News and rumors on Twitter (Jin et al., 2013)
  - Spread of rumors in social networks (Zhao et al., 2012)
  - Zombie apocalypse (Witkowski and Blais, 2013)
Application of Methodology to #Obama
Specify Prior Probability Distributions

<table>
<thead>
<tr>
<th>Prior Probability Distributions</th>
</tr>
</thead>
<tbody>
<tr>
<td>• $\beta \sim U(0, 2)$</td>
</tr>
<tr>
<td>• $\gamma \sim U(0, 2)$</td>
</tr>
<tr>
<td>• $S_0 \sim U(30, 5000)$</td>
</tr>
<tr>
<td>• $I_0 \sim U(0, 10)$</td>
</tr>
</tbody>
</table>
Example of MCMC Parameter Estimation

**Code Snippet**

```python
model = MCMCModel(sim, beta = Uniform(0,2))
model.run_mcmc(10000)
```

Jonathan Skaza  
Mathematical Modeling of Trending Topics on Twitter
Run #Obama Simulation

Simulate 500 times, drawing from posterior probability distributions

Code Snippet

```python
for i in range(500):
    model.draw()
    sim.run(0,191)
    plot(sim.t,sim.l,'g-',alpha=.1)
```
Simulation Results for #Obama
Jonathan Skaza
Mathematical Modeling of Trending Topics on Twitter
#thecwalingdead Parameter Estimation

![Graph showing parameter estimation for #thecwalingdead]
Mathematical Modeling of Trending Topics on Twitter
#CWC15 Parameter Estimation

![Graph showing parameter estimation for #CWC15](image)
2013-14 U.S. Flu Season (September 1st - April 6th)

Source: Google Flu Trends
Flu Parameter Estimation

2013-14 Flu

 Jonathan Skaza

 Mathematical Modeling of Trending Topics on Twitter
Comparison of Model Parameters

Be wary of comparing different parameter estimates.
Prediction Using #Obama

Fit model to training set

![Graph showing tweet count over time.](image_url)
Prediction Results for #Obama

Run simulation over longer timescale

![Chart showing tweet count over time](chart.png)
Prediction Results for #Obama

Comparison to actual (i.e., validation) data
Prediction with a Larger Training Set

Notice that the training set now captures the peak
Prediction Results for #Obama
Prediction Results for #Obama

![Graph showing the prediction results for #Obama](image-url)
Potential Future Work

- Dynamical model tweaks
Potential Future Work

- Dynamical model tweaks
- Create better identification (i.e., hashtag selection) tools
Potential Future Work

- Dynamical model tweaks
- Create better identification (i.e., hashtag selection) tools
- Optimize “window” selection
Potential Future Work

- Dynamical model tweaks
- Create better identification (i.e., hashtag selection) tools
- Optimize “window” selection
- More prediction applications
Potential Future Work

• Dynamical model tweaks
• Create better identification (i.e, hashtag selection) tools
• Optimize “window” selection
• More prediction applications
• Develop interactive display
Potential Future Work

• Dynamical model tweaks
• Create better identification (i.e., hashtag selection) tools
• Optimize “window” selection
• More prediction applications
• Develop interactive display
• Compare to stochastic modeling strategy
Acknowledgments

Advisor: Brian Blais, PhD

Editorial Reviewer: John Quinn, PhD

Honors Program Coordinator: James Segovis, PhD
Questions?

Contact Info

Email: jonathan.skaza@gmail.com
Twitter: @SkazaSays