HealthTweets.org
A Platform for Public Health Surveillance using Twitter

United States Flu Rate

United Kingdom Flu Rate

New York State Flu Rate

Make your own

Michael J. Paul
Johns Hopkins University

Renyuan Cheng, Mark Dredze, David Broniatowski
Social Media for Public Health

- Mental health
- Drug use
- Tobacco use
- Emergency preparedness and response
- Vaccination sentiment
- ...
Influenza Surveillance

- Use Web data to track current influenza rates
  - Twitter messages
  - Google Search Queries
  - Wikipedia searches
  - ...
Research to Practice

• Over two dozen papers demonstrating ability to track influenza with Twitter

• How can a public health official make use of these results?

• Gap between research code and timely, ready to use information
Columbia Prediction of Infectious Diseases: Influenza Forecasts

This visualization summarizes influenza incidence in US cities during the first 33 weeks of the 2013-2014 flu season (through Week 30). Data from Google Flu Trends and WHO-INFLUENZA collaborating laboratories are combined to estimate influenza incidence. These data are used recursively to optimize population-based mathematical models of influenza transmission dynamics. The optimized models are then used to forecast influenza incidence for the remainder of the season (lower right panel). Please see instructions on the left for navigation options.

NEWS (6/18): Dr. Shaman and team placed first in CDC’s “Predict the Influenza Season Challenge” at www.cdc.gov/flu/fluactivity/predict-flu-challenge-winner.htm
Sickweather - Sickness Forecasting App

Everyday thousands of people around the globe share their health-related experiences on social media sites like Facebook and Twitter. When you and others are sick, or when you're close to someone who is sick, you can even let Sickweather know. Sickweather scans indicators of illness, allowing you to check for sickness as easily as you can check for the weather.

See How Sickweather Works
Our Goal

- Create a platform for research collaboration with public health officials
  - Deliver timely health trends from social media
  - Receive detailed feedback to facilitate new research
Requirements

• Goal: show users health trends in Twitter data

• Requires
  • Trend identification
  • Tweet geolocation
  • Historical data
Trend Identification

• System uses a variety of approaches to trend identification
  • Keyword based
  • Machine learning classifiers
  • Example: Influenza
  • A series of statistical machine learning classifiers
Tweet Geolocation

- Identify the location of tweets
- Two sources of geolocation information
  - Geotagged tweets (~3% of US data)
  - Automatically inferred locations based on profile information
    - Carmen (Dredze et al, 2013)
Twitter Geolocation

- User profiles
  - Self-reported locations
    - 56% of users fill this in
- Tweet content
  - Language analysis
- More involved – we don’t do this here
Carmen

- Returns structured object for each tweet
  - City, County, State, Country
- Fast and simple
  - 27,000 tweets / sec
- Code available on Github
  - Python and Java versions
Carmen

- Uses GPS data when available
  - Get location information from Yahoo Map API
- Mapping of user profile strings to places
  - e.g. “NYC”, “New York” -> {city: New York, state: NY, …}
  - Manually curated
  - Automatically added aliases using location clusters created from social network structure
    - Bergsma et al, 2013
- 4,811 unique places in our mapping
Evaluation

- Treated GPS locations as ground truth
- Evaluated geolocation from user profiles against the ground truth
  - Accuracy (precision)
  - Coverage (recall)
- Test set: 56,000 tweets (plus 10,000 dev)
Evaluation

• Accuracy:
  • Country: 91%
  • State+Country: 65%
  • Within 250 miles: 75%
  • Within 25 miles: 55%

• Coverage:
  44% (38% without automatic line extensions)
Historical Data

- Trends require a temporal analysis
  - Helpful to have historical data
- Two streaming API data collections
  - Health keyword based collection
  - Public sample
- Normalization: compute per capita rates based on public sample normalization
Demo
Health Tweets

Follow health trends based on tracking Twitter messages. (Learn More)

United States Flu Rate

United Kingdom Flu Rate

New York State Flu Rate

Make your own


©2014 by the Johns Hopkins Social Media and Health Research Group | contact@healthtweets.org

Data on this site is collected from third-party social media. Reported health information may be incomplete or inaccurate.
Create a plot by selecting plot options below.

Select plot resolution.

- Week

Select plot value type.

- Normalized Counts
- Raw Counts

Select plot date range.

From: 11/26/2013  First date  To: 2/24/2014  Last date

Data to plot (maximum 4 plot lines.)

Disease: Influenza (Classifier)
Location: United States

Add data  Plot
Create a plot by selecting plot options below.

Select plot resolution.
- Week

Select plot value type.
- Normalized Counts
- Raw Counts

Select plot date range.
From: 11/26/2013  To: 2/24/2014

Data to plot (maximum 4 plot lines.)

Disease: Influenza (Classifier)
Location: United States

Remove

Disease  Location
- Influenza (Classifier)  Select
  - United Kingdom  No Cities Available

Add data  Plot
New York
For the week containing Tuesday, January 01, 2013:
Influenza (Classifier)
Tweets: 3,698
Total Tweets: 142,256

18 Known Cities:
Albany (Albany County, New York, United States)
Astoria (Queens County, New York, United States)
Batavia (Genesee County, New York, United States)
Binghamton (Broome County, New York, United States)
Bronx (Bronx County, New York, United States)
Brooklyn (Kings County, New York, United States)

Shading granularity (maximum value for range): 0.03275 (68%)
### United States

**Friday, January 11, 2013**

- **Total Tweets:** 210,499
- **Total Health Tweets:** 192,226

<table>
<thead>
<tr>
<th>Disease</th>
<th>Rate</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anthrax</td>
<td>0.0</td>
<td>0</td>
</tr>
<tr>
<td>Bird Flu</td>
<td>0.000180523422914</td>
<td>36</td>
</tr>
<tr>
<td>Cholera</td>
<td>1.42518491774e-05</td>
<td>3</td>
</tr>
<tr>
<td>Common Cold</td>
<td>0.0797153430658</td>
<td>16,780</td>
</tr>
<tr>
<td>Diphtheria</td>
<td>0.0</td>
<td>0</td>
</tr>
<tr>
<td>Flu Vaccine</td>
<td>0.019406267963</td>
<td>4,985</td>
</tr>
</tbody>
</table>

*587 Known Cities*

### Week of January 06, 2013

- **Total Tweets:** 0
- **Total Health Tweets:** 0

<table>
<thead>
<tr>
<th>Disease</th>
<th>Rate</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anthrax</td>
<td>0.0</td>
<td>0</td>
</tr>
<tr>
<td>Bird Flu</td>
<td>0.00137713577e-06</td>
<td>9</td>
</tr>
<tr>
<td>Cholera</td>
<td>0.0</td>
<td>0</td>
</tr>
<tr>
<td>Common Cold</td>
<td>0.000566246001485</td>
<td>533</td>
</tr>
<tr>
<td>Diphtheria</td>
<td>0.0</td>
<td>0</td>
</tr>
<tr>
<td>Flu Vaccine</td>
<td>3.29336323566e-05</td>
<td>21</td>
</tr>
</tbody>
</table>

### Month of January 2013

- **Total Tweets:** 941,287
- **Total Health Tweets:** 11,733

<table>
<thead>
<tr>
<th>Disease</th>
<th>Rate</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anthrax</td>
<td>1.57533229218e-06</td>
<td>699</td>
</tr>
<tr>
<td>Bird Flu</td>
<td>5.85802750651e-05</td>
<td>23,993</td>
</tr>
<tr>
<td>Cholera</td>
<td>1.00514764279e-06</td>
<td>446</td>
</tr>
<tr>
<td>Common Cold</td>
<td>0.0178656985721</td>
<td>7,936,169</td>
</tr>
<tr>
<td>Diphtheria</td>
<td>2.4114526499e-07</td>
<td>107</td>
</tr>
<tr>
<td>Flu Vaccine</td>
<td>0.00379907496172</td>
<td>138,671</td>
</tr>
</tbody>
</table>

### Since Monday, May 25, 2009

- **Total Tweets:** 443,715,809
- **Total Health Tweets:** 95,480,606

<table>
<thead>
<tr>
<th>Disease</th>
<th>Rate</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anthrax</td>
<td>1.57533229218e-06</td>
<td>699</td>
</tr>
<tr>
<td>Bird Flu</td>
<td>5.85802750651e-05</td>
<td>23,993</td>
</tr>
<tr>
<td>Cholera</td>
<td>1.00514764279e-06</td>
<td>446</td>
</tr>
<tr>
<td>Common Cold</td>
<td>0.0178656985721</td>
<td>7,936,169</td>
</tr>
<tr>
<td>Diphtheria</td>
<td>2.4114526499e-07</td>
<td>107</td>
</tr>
<tr>
<td>Flu Vaccine</td>
<td>0.00379907496172</td>
<td>138,671</td>
</tr>
</tbody>
</table>
Future Directions

- Share site with public health officials
- Deliver metrics that are of immediate practical use to decision makers
- Receive feedback for improving these metrics
Thank You

- www.healthtweets.org
- Tell your friends!
- Email for an account: contact@healthtweets.org