Predicting Spread of Disease from Social Interactions

Adam Sadilek
Henry Kautz
Vincent Silenzio
JazzStarDiva [0.94634229]  
I'm getting sick on this bus  
#literally

Micah_88 [1.14140003]  
oh ok I was sick n still is
Organic Sensor Networks

- 52% of adults use online social networks
- Smartphone access
  - Real time
  - “In the moment”
  - Location aware

Thursday, October 11, 12
Organic Sensor Networks

- Detailed measurements at a population scale
- No active user participation
- Fine granularity
- Timely
- Inference
- Prediction
The Data

- New York City
  - 6K geo-active users
  - 2.5M tweets by geo-active users
    - 103K “follows” relationships
    - 32K “friends” relationships
feeling horrible.. this flu had me up till 3am..... Jesus!
Spread of Disease

- **Identify** people with *symptoms*
- **Quantify** the impact of:
  - Co-location
  - Social ties
- **Predict** contagion with fine granularity
Tweet: “feeling sick”
Cascade SVM

Corpus of 5,128 tweets labeled by human workers

Corpus of 1.6 million machine-labeled tweets

Random sample of 200 million tweets

Corpus of "other" tweets

Corpus of "sick" tweets

Final corpus
## SVM Features

<table>
<thead>
<tr>
<th>Feature</th>
<th>Weight</th>
<th>Feature</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>sick</td>
<td>0.9579</td>
<td>sick of</td>
<td>-0.4005</td>
</tr>
<tr>
<td>headache</td>
<td>0.5249</td>
<td>you</td>
<td>-0.3662</td>
</tr>
<tr>
<td>flu</td>
<td>0.5051</td>
<td>lol</td>
<td>-0.3017</td>
</tr>
<tr>
<td>fever</td>
<td>0.3879</td>
<td>love</td>
<td>-0.1753</td>
</tr>
<tr>
<td>feel</td>
<td>0.3451</td>
<td>i feel your</td>
<td>-0.1416</td>
</tr>
<tr>
<td>coughing</td>
<td>0.2917</td>
<td>so sick of</td>
<td>-0.0887</td>
</tr>
<tr>
<td>being sick</td>
<td>0.1919</td>
<td>bieber fever</td>
<td>-0.1026</td>
</tr>
<tr>
<td>better</td>
<td>0.1988</td>
<td>smoking</td>
<td>-0.0980</td>
</tr>
<tr>
<td>being</td>
<td>0.1943</td>
<td>i’m sick of</td>
<td>-0.0894</td>
</tr>
<tr>
<td>stomach</td>
<td>0.1703</td>
<td>pressure</td>
<td>-0.0837</td>
</tr>
<tr>
<td>and my</td>
<td>0.1687</td>
<td>massage</td>
<td>-0.0726</td>
</tr>
<tr>
<td>infection</td>
<td>0.1686</td>
<td>i love</td>
<td>-0.0719</td>
</tr>
<tr>
<td>morning</td>
<td>0.1647</td>
<td>pregnant</td>
<td>-0.0639</td>
</tr>
</tbody>
</table>
lafemmefatalexx [0.74721562]
My feet hurt sooooooo baddddd I need to be carried home waaaaahhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh never wearing heels again.
Twitter Health

- Aggregate accuracy comparable with:
  - Google Flu Trends ($R = 0.73$)
  - CDC statistics

+ we can model fine-grained interactions between specific individuals
Health Prediction

\[ h_t - 1 \quad \rightarrow \quad h_t \quad \rightarrow \quad h_t + 1 \quad \rightarrow \]

\[ \ldots \]

\[
\begin{align*}
    &h_t - 1 \\
    &h_t \\
    &h_t + 1 \\
\end{align*}
\]

Precision Viterbi
- Precision smoothing
- Recall Viterbi
- Recall smoothing
- Precision baseline
- Recall baseline

Number of days into the future (x)

Thursday, October 11, 12
Health Insights

[Snow, 1855]
Psychological Health

• Effect of connectedness on mental health
  • Twitter data exhibits patterns from animal studies: higher status -> better health

• Can we estimate emotional state from tweets?
  • Tweeters may try to disguise certain feelings

• Interaction of social network and depression
  • Is depression contagious?
  • How does network reduce or reinforce?
Summary

• **Quantify** impact of physical **encounters** and **social ties** on public **health**

• **Predict** future health
  - Real world observed via Twitter
  - Fine granularity
  - Real-time
  - No active user participation