What were the Tweets about? 
Topical Associations between Public Events and Twitter Feeds

Yuheng Hu (@hyheng) Arizona State Univ. 
Ajita John  Avaya Labs 
Dorée Seligmann  Avaya Labs 
Fei Wang  IBM T.J Watson Research
Social media has changed our way of responding to events

triangles not foaf in online social networks, interesting! #icwsm

How do we make sense of the responses?
Research goals

• Make sense of a crowd’s responses to a media event
  – What these tweets were about?
  – How did the event influence the tweeting behavior of the crowd?
  – Did these tweets refer to specific parts of the event and if so, what parts?

• Related work
  – Monitoring Twitter usage patterns during debate [Shamma et al 2009]
  – Sentimental analysis of tweets to understand the events [Diakopoulos et al 2010]
Preliminary understanding

- Event: Republican Primary debate on Sept 7 2011 20:00-22:00
- Data: over 110,000 tweets tagged #ReaganDebate
Three conjectures about the crowd's responses

- **Before the event**, responses tend to be mostly tangential or very general
- **During the event**, responses tend to be very specific to parts of the event
- **After the event**, responses tend to be slightly more specific to the event from the ones posted before the event
More insights

• A tweet’s content can be either weakly or strongly influenced by the debate’s content

• Tweets with weak correlations used words mostly about the general topics of the debate (steady tweets)

• Tweets with strong correlations used words mostly related to specific topics discussed in the debate (episodic tweets)
Our approach

A steady tweet draws words mostly from general topics.

General Topics remain constant through transcript.

Each segment has specific topics.

An episodic tweets draws words mostly from specific topics and thus refers to particular segments.
ET-LDA in a nutshell

Event

- Determine event segmentation

Tweets

- Determine which segment a tweet (word) refers to
- Determine tweet (word) type
- Determine word’s topic in event
- Tweets word’s topic

\[ \alpha_\delta \]
\[ \delta^{(s)} \]
\[ \alpha_\theta \]
\[ \theta^{(s)} \]
\[ Z_s^i \]
\[ W_s^i \]
\[ N_s \]
\[ S^{(t)} \]
\[ C^{(t)} \]
\[ \gamma^{(t)} \]
\[ \lambda^{(t)} \]
\[ \psi^{(t)} \]
\[ \alpha_\gamma \]
\[ \alpha_\lambda \]
\[ \alpha_\psi \]

\[ S_T \]
\[ M_T \]
\[ K \]
Experiments

• Setup
  – Tweets for President Obama’s speech on the Middle East on May 19 2011 (#MESpeech, 11,988) and Republican Primary debate in the US on Sept 7 2011 (#ReaganDebate, 112,414)
  – Expanding tweets by search snippets
  – Event transcripts from New York Times
  – Model settings: Gibbs sampling and pick #topics by maximizing log-likelihood
General/Specific topics from #MESpeech

<table>
<thead>
<tr>
<th>S</th>
<th>Specific</th>
<th>Top Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>Human rights</td>
<td>Rights Transition People Power</td>
</tr>
<tr>
<td></td>
<td>Foreign policy</td>
<td>Secure Mideast Arab Clinton State</td>
</tr>
<tr>
<td>S2</td>
<td>Terrorism</td>
<td>Bin Laden Mass Murderer Cry</td>
</tr>
<tr>
<td></td>
<td>People</td>
<td>Dignity Power Street Square people</td>
</tr>
<tr>
<td>S3</td>
<td>Arab democracy</td>
<td>Democracy Yemen Syrian Bahrain</td>
</tr>
<tr>
<td></td>
<td>Egypt revolution</td>
<td>Mubarak Resign Policy Reform</td>
</tr>
<tr>
<td>S4</td>
<td>Youth</td>
<td>Promote Untapped Arab talent youth</td>
</tr>
<tr>
<td></td>
<td>Free speech</td>
<td>Open Internet Mind Access Paper</td>
</tr>
<tr>
<td>S5</td>
<td>Economics</td>
<td>Aid Trade Debt Billion Dollar</td>
</tr>
<tr>
<td></td>
<td>Reform</td>
<td>Egypt Reform Support Resource</td>
</tr>
<tr>
<td>S6</td>
<td>Border problem</td>
<td>Israel Palestine Borders State Jordan</td>
</tr>
<tr>
<td></td>
<td>Peace treaty</td>
<td>Palestine Peace Jewish Agreed treaty</td>
</tr>
<tr>
<td>S7</td>
<td>Human rights</td>
<td>Rights Transition People Power</td>
</tr>
<tr>
<td></td>
<td>Foreign policy</td>
<td>Secure Mideast Arab Clinton State</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>General</th>
<th>Top Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arab spring</td>
<td>Arabia Bahrain Iran Mosques</td>
</tr>
<tr>
<td></td>
<td>Syrian Leader Government Stepped</td>
</tr>
<tr>
<td>Israel &amp; Palestine</td>
<td>Israel Palestine Borders Lines Hamas</td>
</tr>
<tr>
<td></td>
<td>Negotiate Permanent Occupation</td>
</tr>
<tr>
<td>Obama</td>
<td>President Job Tough Critique</td>
</tr>
<tr>
<td></td>
<td>Jews Policies Attacking Weakness</td>
</tr>
</tbody>
</table>

**MESpeech**: specific topics are sensitive to the event’s context and keep evolving as the event progresses
ReaganDebate: specific topics are rather disordered and occur repeatedly.
Examples of Steady/Episodic tweets

• **ReaganDebate**
  - **Steady**
    - Something the #GOP candidates won't mention about Reagan - Reagan grew the size of the federal government tremendously. #reagandebate
    - Yes, we need to talk about jobs and teachers needing jobs! #Reagandebate
  - **Episodic**
    - Huntsman said Ronnie!! Take a shot! #GOPDebate #tcot #ReaganDebate
    - Wow, Ron Paul. Really, you think airlines would give a rip about security? Free market nonsense. #reagandebate
Evolution of Episodic tweets over the event’s timeline

**MESpeech:** Few responses were episodic initially. The percentage of episodic tweets rises up/drops down as the event begins and as it is ending, it remained mostly stable during the event.

**ReaganDebate:** More responses are strongly influenced by the event. The crowd was less excited during the opening. The percentage of episodic tweets rises significantly during the debate.
Distribution of Segments Referred to by Episodic tweets (#MESpeech)

- Horizontally, the crowd’s attention tended to shift from one segment to the next as the event progressed.
- Vertically, episodic tweets also refer to the segments whose topics were expected to be discussed later on in the event.

<table>
<thead>
<tr>
<th>Segment</th>
<th>Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>Foreign policy</td>
</tr>
<tr>
<td>S2</td>
<td>Anti-Terrorism</td>
</tr>
<tr>
<td>S3</td>
<td>Arab democracy</td>
</tr>
<tr>
<td>S4</td>
<td>Youth &amp; Free speech</td>
</tr>
<tr>
<td>S5</td>
<td>Economics reform</td>
</tr>
<tr>
<td>S6</td>
<td>Israel &amp; Palestine</td>
</tr>
<tr>
<td>S7</td>
<td>Foreign policy</td>
</tr>
</tbody>
</table>
Subjective Evaluation

- Participants: 31 graduate students
- Method: questionnaire
- Participants asked to assess quality of topics and soundness of episodic tweets by ET-LDA
- ET-LDA performed consistently better than baselines
Our findings

1. Crowd’s responses tended be steady before the event and after the event, while during the event, they were more episodic. (our conjectures are confirmed!)

2. Crowd showed different levels of engagement in different kinds of events

3. Topical context of the tweets did not always correlate with the timeline of the event
Questions?

• For details: yuheng@asu.edu
• Web: http://www.public.asu.edu/~yuhenghu/
• Twitter: @hyheng

Thank you!