Geo-Temporal Networks & Visualization

Sumeet Kumar
sumeetku@cmu.edu

The CASOS Center
School of Computer Science, Carnegie Mellon
Summer Institute 2018

Visualizing Tweets

Analysis of New York Hurricane Sandy data
Agenda

- Load Twitter data
- Aggregate data
- Visualize basics node-metrics over time using measure charts
- Visualize basics network-measures over time
- Geo-spatial visualization
- Use of shape files
- Color shapes by count (like geo-spatial color map)

Raw Data

- Import ‘Sandy_Twitter_anon.xml’. Drag and drop will work.
- Frames Count: 10.5K
- Source information: Geo-tagged tweets from Manhattan
- Date range: 10/26 13:59:06...11/3 23:56:23
- Tweet examples:
  - 10/26: “Sandy strengthens as it moves from Jamaica to Cuba ...only 1 mph below the status of a major Category 3 hurricane.”
  - 11/3: “NBC News reports that the death toll in the U.S. is now 109, including at least 40 in New York City.”
Data in ORA

- We have 1 MetaNetwork/tweet
  - Agents: The person tweeting and the people he/she mentions
  - Knowledge: The hashtags used in the tweet
  - Location: The location of the tweet
- First, we will aggregate the data by 1 day and look at:
  - Basic statistics and network visualization over time
  - Where the tweets came from over time
  - Where #s moved over time
  - Where agents tweeted from over time

Aggregating the data

- Things to think about: What period to use? What do we lose?
Aggregating the data

- Things to think about: What period to use? What do we lose?

Basic Statistics: Node count over time

- We will use measure charts.
Basic Statistics: Node count over time

- We will use measure charts.
Basic Statistics: Node count over time

- What do you expect?

On the 29th, people used hashtags, even though less people were tweeting.

Basic Network Statistics: Density over time

- What do you expect?
How do the networks look like over time?

1. Visualize one day network at a time.
2. Select meta-network, use play in visualizer.
Visualize the #tag network

Observations

- Within the Largest connected component (LCC), a few #s dominated - #sandy, #NYC, #HurricaneSandy
  - When we remove these, we lose a lot of the structure in the network
- Oftentimes, the things outside the LCC were thus unrelated
  - But not always - e.g. 11/1: 
    "#gerritsonbeack in bkln not in zone a on #cbs they say no food no water need help"
- There was a higher density in the LCC as time progressed
- Within the LCC of the AA network, nearly all of the nodes were public accounts
- Bottom line:
  - The “geo-tagged Twitter network” in Manhattan organized around information hubs (public accounts) who were concerned with the earthquake
  - But where were they?
Tracking Tweet volume in Space over time

- We know all of the Tweets we have came from around Manhattan
  - Can we use them to see which areas were most affected?
  - How does this change over time?
- Overview: using a Zillow shapefile, visualize where tweets were coming from over time in the Geospatial visualizer
  - [http://www.zillow.com/howto/api/neighborhood-boundaries.htm](http://www.zillow.com/howto/api/neighborhood-boundaries.htm)
  - Download 'New York Neighborhood Boundaries'

Geospatial Visualization

- Aggregate Network by 1 year
  - Will be left with a single network
- Select Geo-spatial visualization from menu
Geospatial Visualization

• Remove all aggregation

Geospatial Visualization

• Remove all labels
Import shape file

- Import shape files
- Color shapes by tweets count
Import shape file

- Import shape files
- Color shapes by tweets count
Tracking Tweet volume in Space over time

http://www.bbc.co.uk/news/world-us-canada-20151303
Observations

• Many of the tweets came from Midtown, the business hub of the city
  - This isn’t that surprising, given it has a high density

• But we can see that the areas that had relative increases in tweets appear to be places where disaster struck worse

• Was it because people were moving there or because more people from there were tweeting?
  - Let’s use Loom to take a look at movements

Observations

• The most important agents (i.e. those that tweeted the most) tended to stay in the same place when they were tweeting
  - We could compare this to a model of how agents typically move to see if this happened because of the disaster

• However, this suggests that, indeed, proportionally more “residents” may have been tweeting from the more affected areas during the disaster
Analysis of Hashtags “HurricaneSandy” Overtime

Top Hashtags

<table>
<thead>
<tr>
<th>10/26</th>
<th>10/27</th>
<th>10/28</th>
<th>10/29</th>
<th>10/30</th>
<th>10/31</th>
<th>11/1</th>
<th>11/2</th>
<th>11/3</th>
</tr>
</thead>
<tbody>
<tr>
<td>#Sandy</td>
<td>HurricaneSandy</td>
<td>Sandy</td>
<td>Sandy</td>
<td>#nyc</td>
<td>#NYC</td>
<td>#Manhattan</td>
<td>#Manhattan</td>
<td></td>
</tr>
<tr>
<td>NOAA</td>
<td>#Miami</td>
<td>#nyc</td>
<td>#NY</td>
<td>Sandy</td>
<td>SandyNYC</td>
<td>#Sandy</td>
<td>traffic</td>
<td></td>
</tr>
<tr>
<td>NPP</td>
<td>#Florida</td>
<td>#Sandy</td>
<td>HurricaneSandy</td>
<td>#NewYork</td>
<td>#Miami</td>
<td>#HurricaneSandy</td>
<td>Miami</td>
<td>Miami</td>
</tr>
<tr>
<td>VIBS</td>
<td>#nyc</td>
<td>#Frankenstorm</td>
<td>#NewYork</td>
<td>HurricaneSandy</td>
<td>#Florida</td>
<td>#NY</td>
<td>Florida</td>
<td>Florida</td>
</tr>
<tr>
<td>NASA</td>
<td>#water</td>
<td>#Hurricane</td>
<td>traffic</td>
<td>HurricaneSandy</td>
<td>Sandy</td>
<td>#ConEdison</td>
<td>NYC</td>
<td></td>
</tr>
<tr>
<td>Miami</td>
<td>#Sandy</td>
<td>#newyork</td>
<td>#Miami</td>
<td>#NY</td>
<td>#Manhattan</td>
<td>#NY</td>
<td>#nyc</td>
<td>#Stuytown</td>
</tr>
<tr>
<td>Florida</td>
<td>storm</td>
<td>#fb</td>
<td>#Florida</td>
<td>#Miami</td>
<td>#traffic</td>
<td>#Miami</td>
<td>HurricaneSandy</td>
<td>+spam</td>
</tr>
<tr>
<td>New York</td>
<td>#readiness</td>
<td>#Evacuation</td>
<td>#Bott</td>
<td>#Florida</td>
<td>#NY</td>
<td>#Florida</td>
<td>traffic</td>
<td>+1</td>
</tr>
<tr>
<td>NY</td>
<td>#flashlights</td>
<td>#MTA</td>
<td>NY</td>
<td>#austria</td>
<td>#ConEdison</td>
<td>#Manhattan</td>
<td>#NewYork</td>
<td>Sandy</td>
</tr>
<tr>
<td>East New York</td>
<td>#batteries</td>
<td># inundation</td>
<td>#HellzKitchen</td>
<td>#manhattan</td>
<td>#NJ</td>
<td>#recovery</td>
<td>#NY</td>
<td>#WestVillage</td>
</tr>
</tbody>
</table>
Conclusion

- We began by noting that
  - Hashtag count actually decreased during the Hurricane even though number of agents increased, suggesting that information became much more centered around the hurricane
  - The “geo-tagged Twitter network” in Manhattan organized around information hubs (public accounts) who were concerned with the earthquake
- Twitter may not have been the best source of information during the hurricane, but there is evidence that
  - People tweeted proportionately more from affected areas
  - Hashtags became more localized

Thank You!!