Geo-Temporal Networks & Sentiment Mining

Sumeet Kumar

sumeetku@cmu.edu

The CASOS Center

School of Computer Science, Carnegie Mellon

Summer Institute 2017

institute for SOFTWARE RESEARCH

Carnegie Mellon

Center for Computational Analysis of Social and Organizational Systems

http://www.casos.cs.cmu.edu/

Part 1

Analysis of New York Hurricane Sandy data

Raw Data

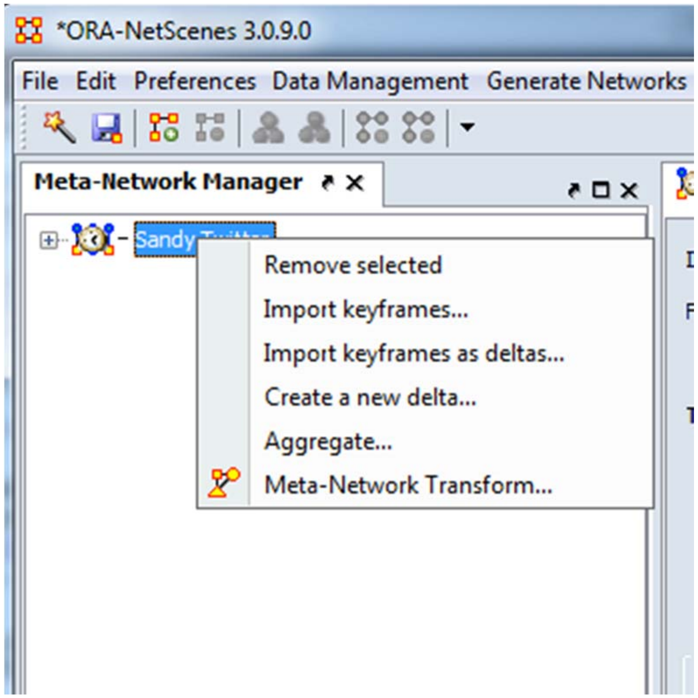
- Count: 10.5K
- Location information: Geo-tagged tweets from Manhattan
- Date range: 10/26 13:59:06…11/3 23:56:23
  – 10/26: “Sandy strengthens as it moves from Jamaica to Cuba …only 1 mph below the status of a major Category 3 hurricane.”
  – 11/3: “NBC News reports that the death toll in the U.S. is now 109, including at least 40 in New York City.”

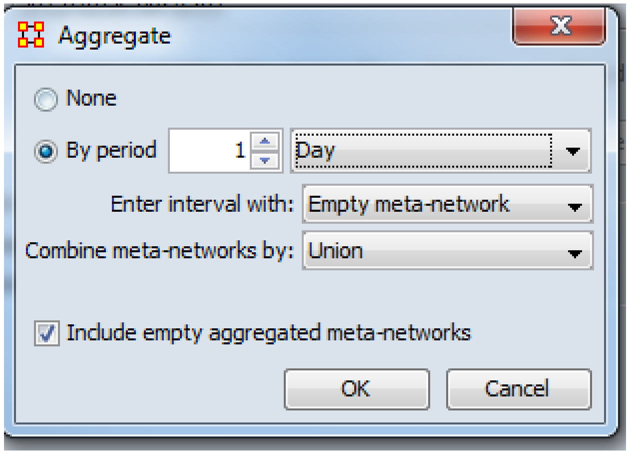
Data in ORA

- We have 1 MetaNetwork/tweet
  - Agents: The person tweeting and the people 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

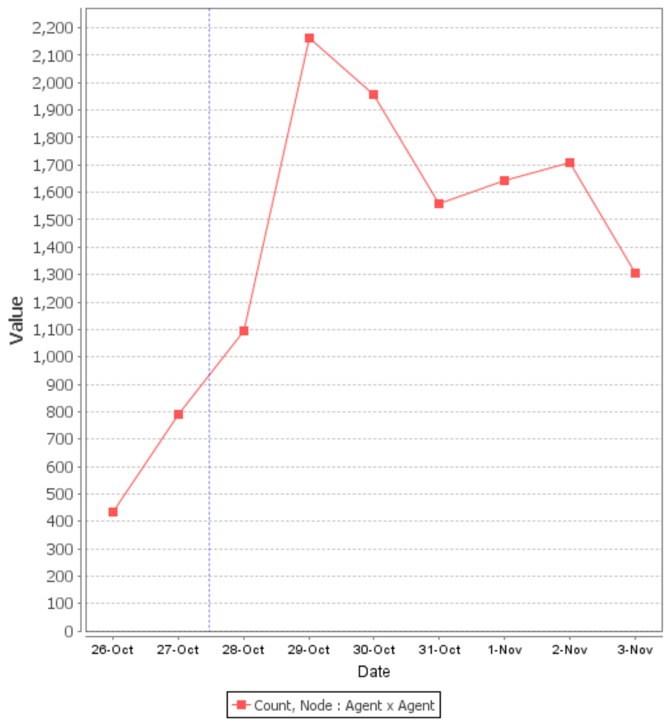
- Things to think about: Why? What do we lose?

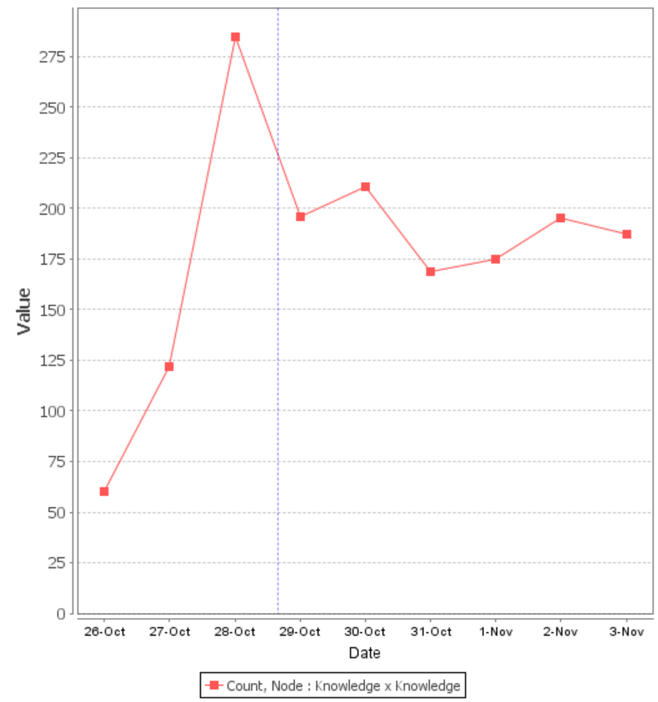




Basic Statistics: Node count over time

- What do you expect?

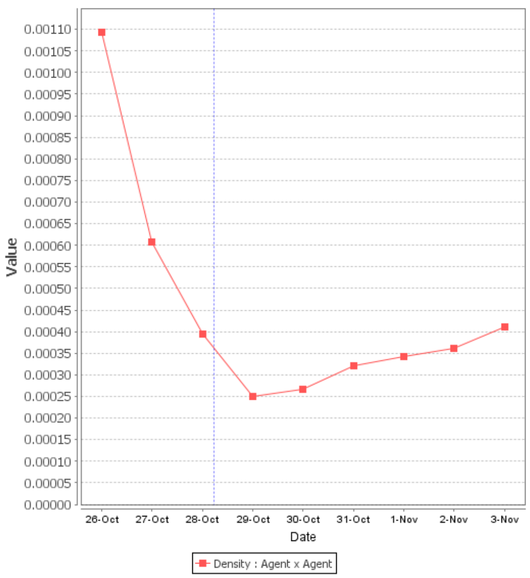


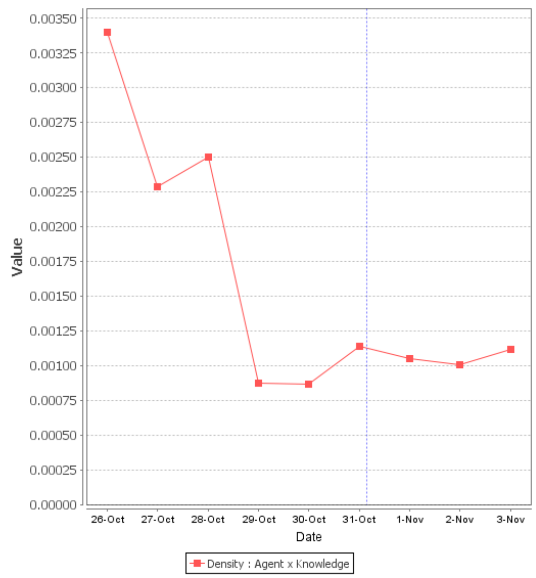


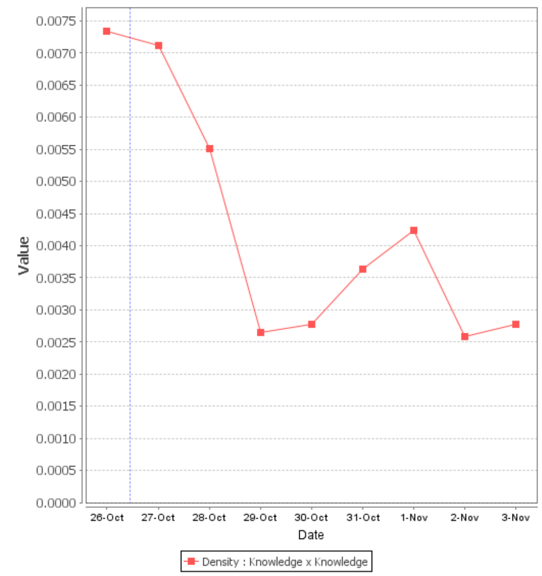
On the 29th, people focused on more specific topics/hashtags, even though more people were tweeting

Basic Statistics: Density over time

- What do you expect?

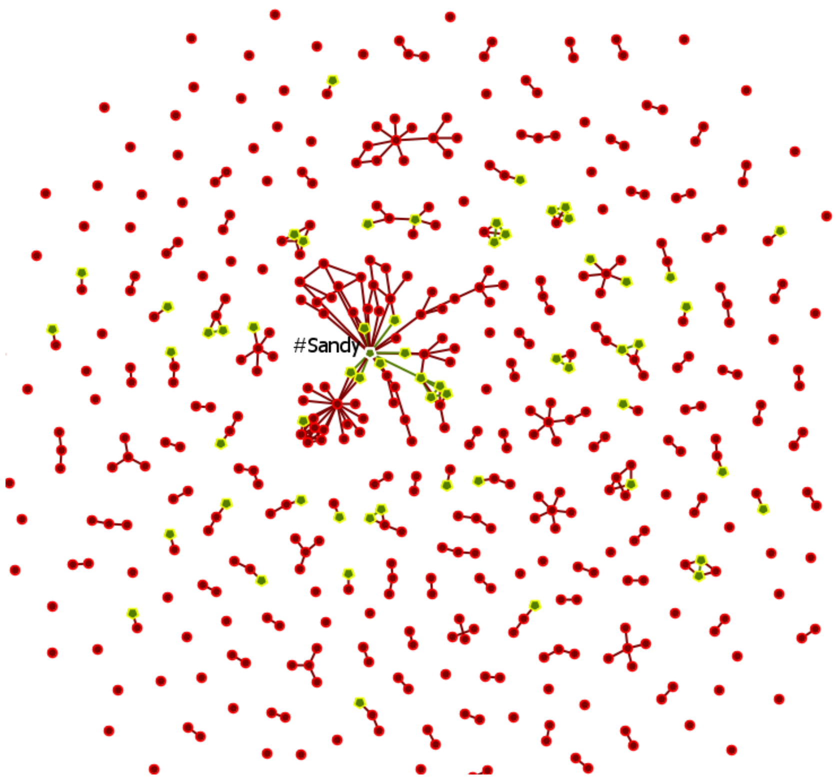


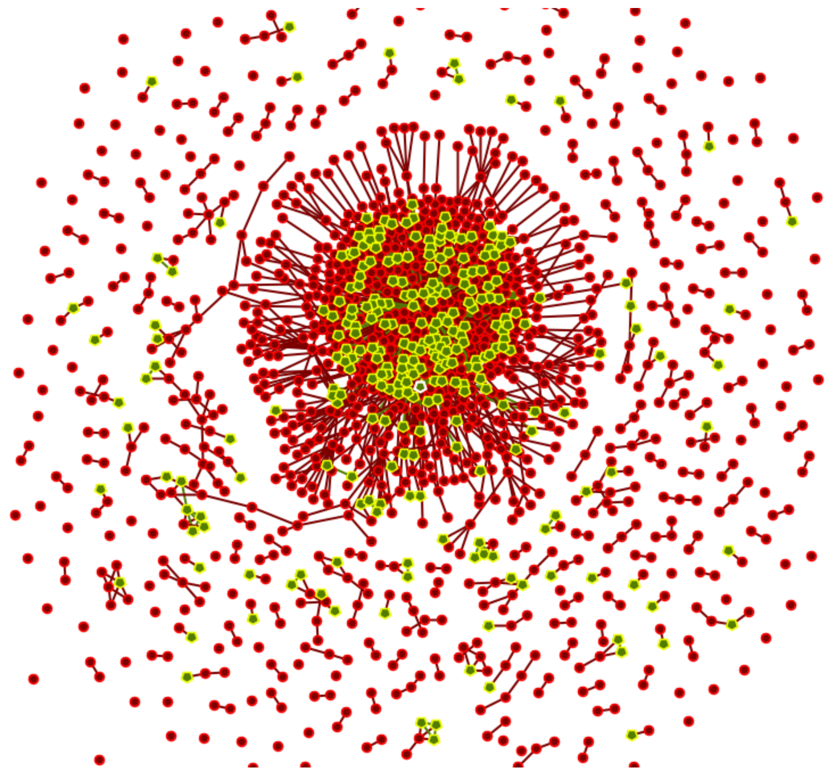




What did the network look like over time?

- What do you expect?



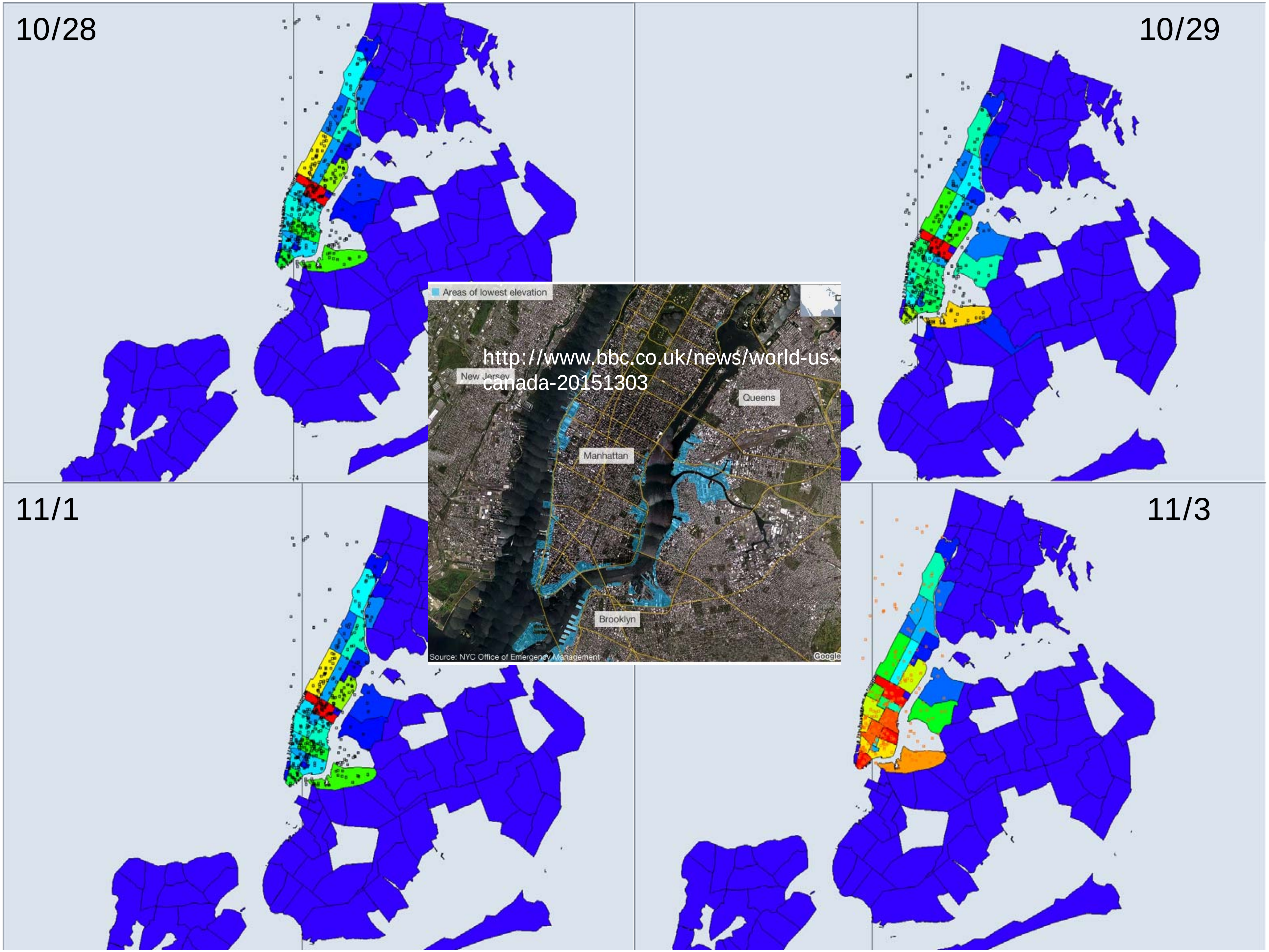


Observations

- Within the 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


10/28
10/29
11/1
11/3
Areas of lowest elevation
http://www.bbc.co.uk/news/world-us-canada-20151303
New Jersey
Queens
Manhattan
Brooklyn
Source: NYC Office of Emergency Management
Google

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

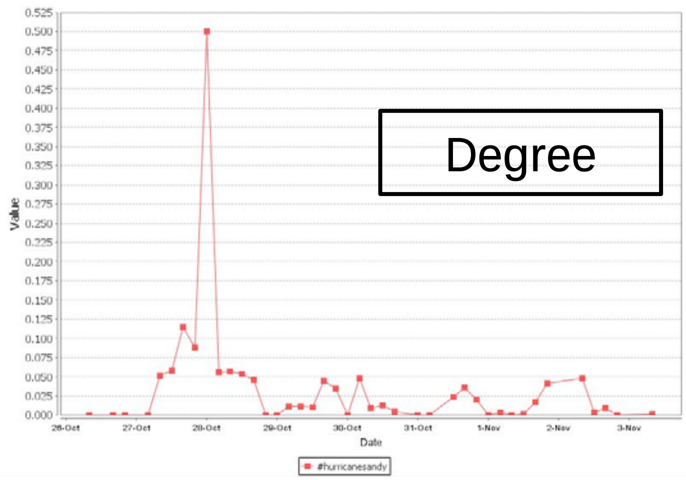
- But we can see that the areas that had relative increases in tweets appear to be places where disaster struck worse

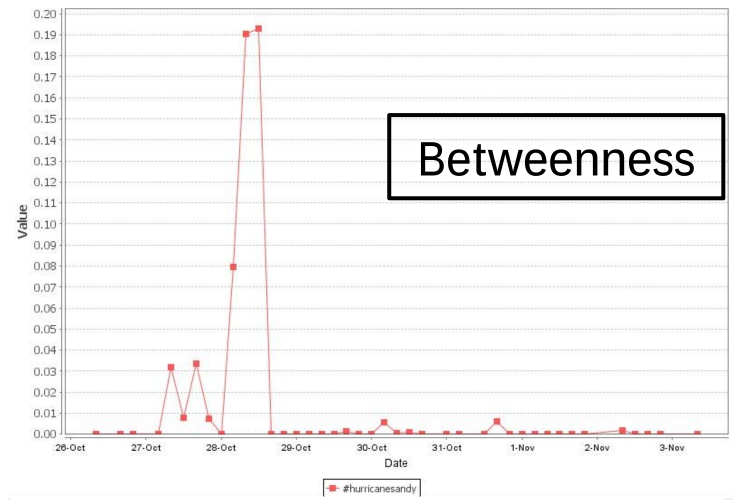
- Was it because people were moving there or because more people from there were tweeting?
  – Let's use Loom to take a look at movements

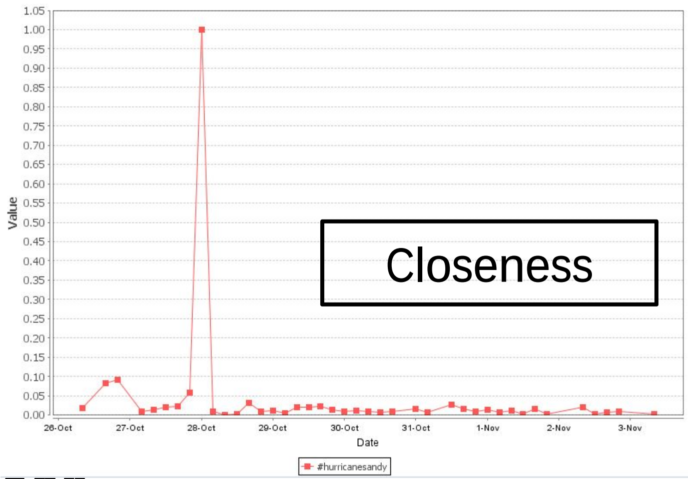
Observations

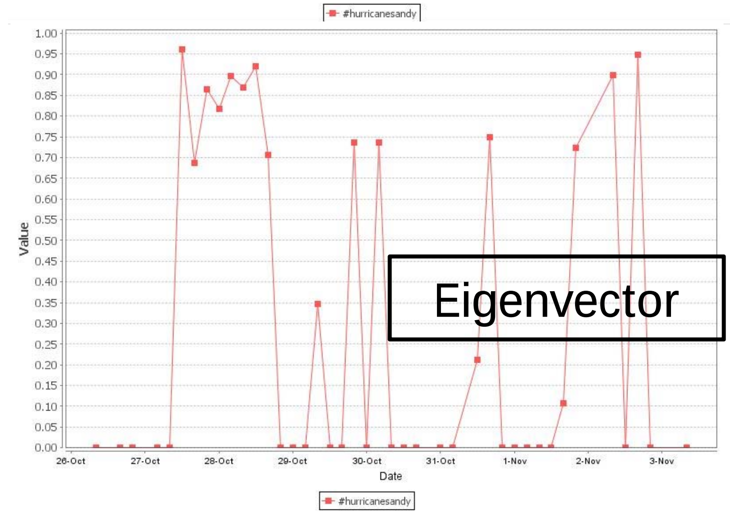
- The most important agents (i.e. those that tweeted the most) tended to stay in the same place when they were tweeting
  – We could compare this to a model of how agents typically move to see if this happened because of the disaster
- However, this suggests that, indeed, proportionally more “residents” may have been tweeting from the more affected areas during the disaster

Analysis of Hashtags “HurricaneSandy” Overtime









Top Hashtags

10/26	10/27	10/28	10/29	10/30	10/31	11/1	11/2	11/3
#Sandy	#hurricanesandy	#hurricanesandy	#Sandy	#Sandy	#nyc	#NYC	#Manhattan	#Manhattan
#NOAA	#Miami	#nyc	#NYC	#NY	#sandy	#SandyNYC	#Sandy	#traffic
#NPP	#Florida	#Sandy	#HurricaneSandy	#Newyork	#Miami	#HurricaneSandy	#Miami	#Miami
#VIIRS	#nyc	#Frankenstorm	#NewYork	#hurricanesandy	#Florida	#NYRNY	#Florida	#Florida
#NASA	#water	#Hurricane	#traffic	#hashNewyork	#hurricanesandy	#sandy	#ConEdison	#NYC
#Miami	#Sandy	#newyork	#Miami	#NYC	#Manhattan	#NY	#nyc	#Stuytown
#Florida	#storm	#fb	#Florida	#Miami	#traffic	#Miami	#hurricanesandy	#pcvst
#Newyork	#readiness	#Evacuation	#tcot	#Florida	#NY	#Florida	#traffic	#1
#NY	#flashlights	#MTA	#NY	#astoria	#ConEdison	#Manhattan	#Newyork	#Sandy
#hashNewyork	#batteries	#brooklyn	#HellsKitchen	#manhattan	#NJ	#recovery	#NY	#WestVillage

Conclusion

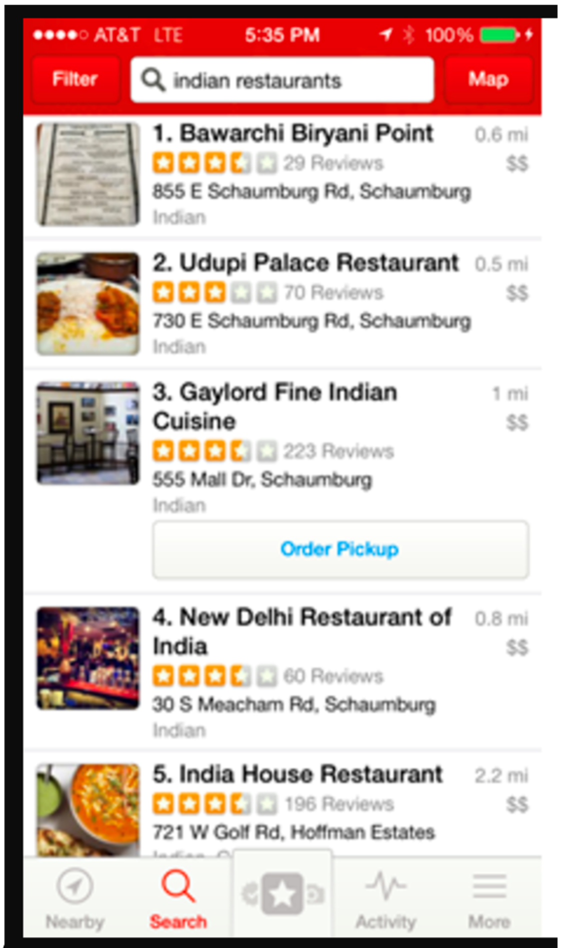
- We began by noting that
  - Hashtag count actually decreased during the Hurricane even though number of agents increased, suggesting that information became much more centered around the hurricane
  - The "geo-tagged Twitter network" in Manhattan organized around information hubs (public accounts) 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**

Part 2

Sentiment Analysis

(Opinion Mining)

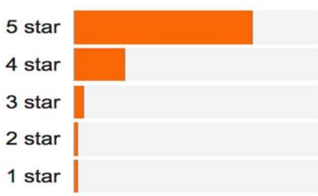
Why Sentiment Analysis?



Why Sentiment Analysis?

Amazon Echo - Black
$100 online, **$180** nearby ★★★★★ 5,137 product reviews

Reviews



4.6
★★★★★
5,137 reviews

ease of use	"Overall the Amazon Echo was easy to set up and use."
sound/audio	"Sound quality amazing!"
value	"Excellent sound and great value."
size	"The speaker is also impressive for its size."
design/style	"Love it amazing easy to use love the look sound amazing"
customer service	"Great customer service"
battery	"Awesome plus made it better by buying the battery!"

★★★★★ **LOVE IT!** – April 3, 2017

Imgiac – Review provided by Bed Bath & Beyond
April 3, 2017

Awesome sound! You must gave Amazon Prime at least to get the benefit of the music which is my main interest. I have Prime so it paid for itself already!

★★★★★ **Echo rocks!!** – December 15, 2016

Techgirlo – Review provided by Bed Bath & Beyond
December 15, 2016

I love Alexa and she is becoming my new BFF! I love it so much that I bought all the men in my life the gift of Alexa for Christmas. Just beginning to scratch the surface with all she can do and improve in my world. Don't wait, get yours now while they are on sales!!

★★★★★ **Amazon Echo** – November 21, 2016

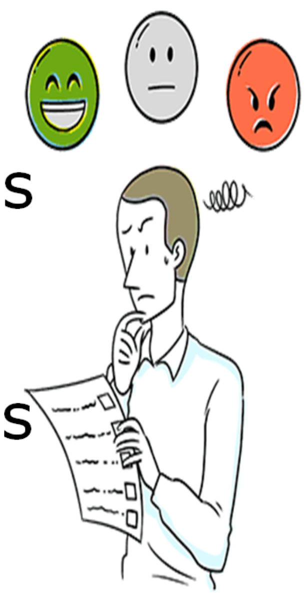
NickP – Review provided by Bed Bath & Beyond
November 21, 2016

We have owned the Echo since it was born. Each day we find another use for the Echo. Looked at originally, as a novelty, it fast became a very useful household tool. Our Granddaughters enjoy it so much. We purchased this one for a Christmas Gift. Extremely useful!

What is Sentiment Analysis?

Goals:

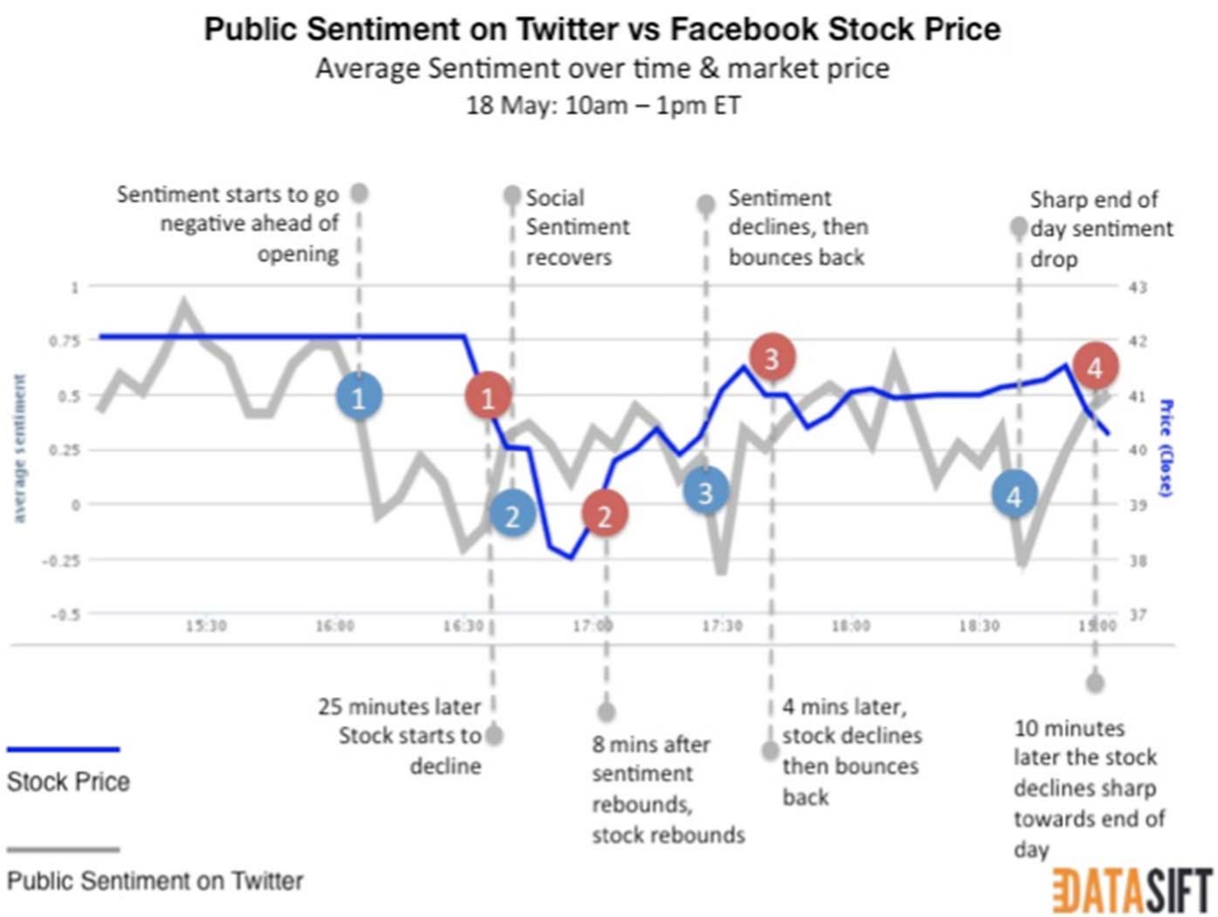
- Elicit emotional responses in internet exchange
- Identify subjective information in text
- Often outputs polarity (-ve, +ve, neutral) or scale (1,..., 5)
- Attitude of the writer towards a topic
- Find source, target and complex emotions



Use Cases:

- Business looking to market their products
- Understanding voters
- Build Networks??

Linking Text Sentiment to Stock Market



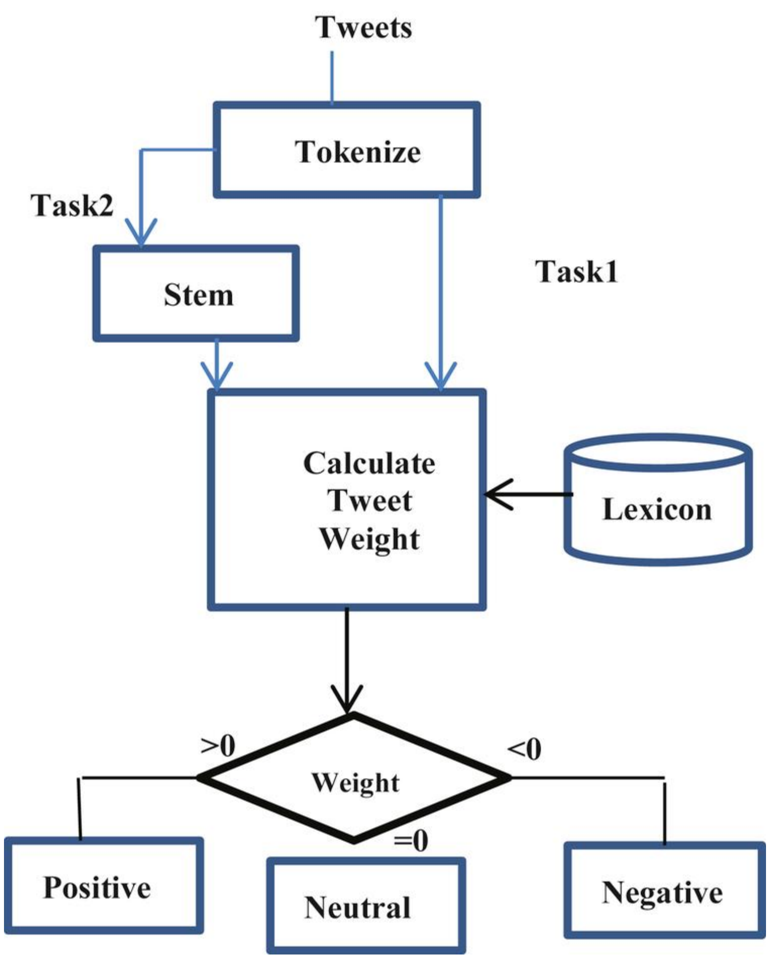
Source: http://socialmediaimpact.com/5-reasons-twitter-becoming-essential-stock-market/

Common Approaches

- Lexicon Based
  - Count based Techniques
  - Rule based Techniques

- Machine Learning/ Statistics Based
  - Naïve Bayes
  - Latent Semantic Analysis
  - Neural Networks (Deep Learning)
  - And More

Sentiment Analysis – Lexicon Based



Many open source Lexicons are available
e.g. SentiStrength

Source: Journal of Intelligent & Fuzzy Systems, vol. 29, no. 1, pp. 107-117, 2015

23

Sentiment Analysis –Naïve Bayes

Assuming POS Tags and n-grams are conditional independence
M = Twitter message, S = sentiment, G = n-grams, T = POS tags

$$P(s|M) = \frac{P(s) \cdot P(M|s)}{P(M)}$$

$$P(s|M) \sim P(G|s) \cdot P(T|S)$$

$$P(s|M) \sim \prod_{g \in G} P(g|s) \cdot \prod_{t \in G} P(t|s)$$

Sentiment Analysis –Naïve Bayes

Let's find the sentiment in text "SI is good" using Naïve Bayes.
Assume: + implies Positive sentiment and - implies Negative sentiment, and P(-) = P (+) = 0.5

$$P(s|M) = \frac{P(s) \cdot P(M|s)}{P(M)}$$

P(+|"SI is good") = P(+)* P("SI is good" | +) / P("SI is good")

P(-|"SI is good") = P(-)* P("SI is good" | -) / P("SI is good")

Divide the last two equation to find the ratio of sentiment:
P(+|"SI is good") / P(-|"SI is good") = P("SI is good" | +) / P("SI is good" | -)

Using independence assumption:

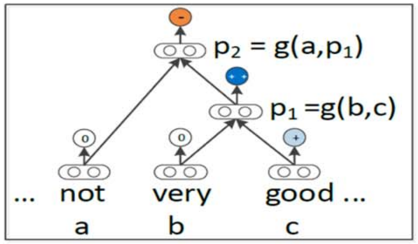
= P("SI"| +) P("good"| +) / P("SI"| -) P("good"| -)
= P("good"| +) / P("good"| -) = .01 / 0.00001 = 1000
=> Sentiment is more positive

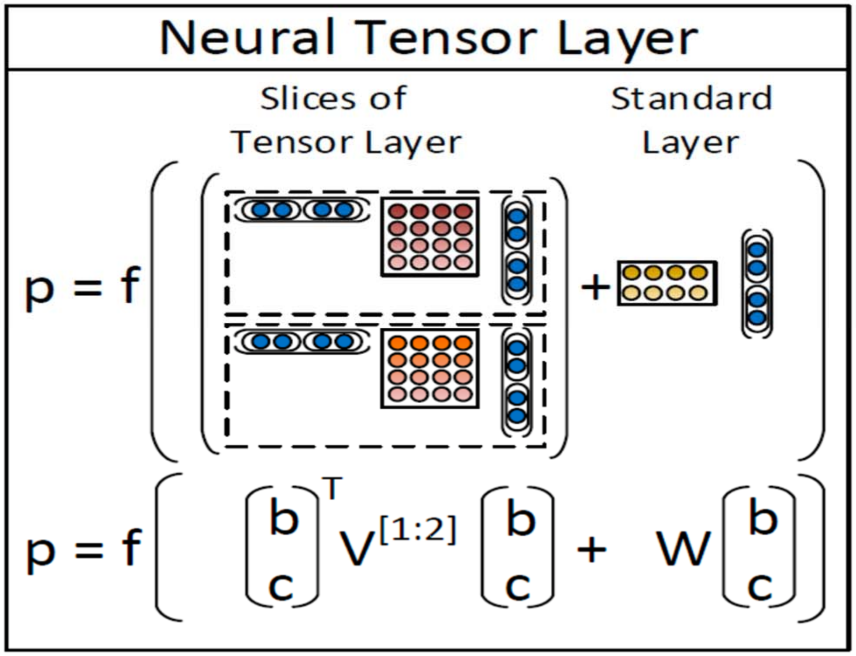
Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Sentiment Analysis –Neural Network

Recursive Neural Tensor Network

Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank
Socher et al. 2013

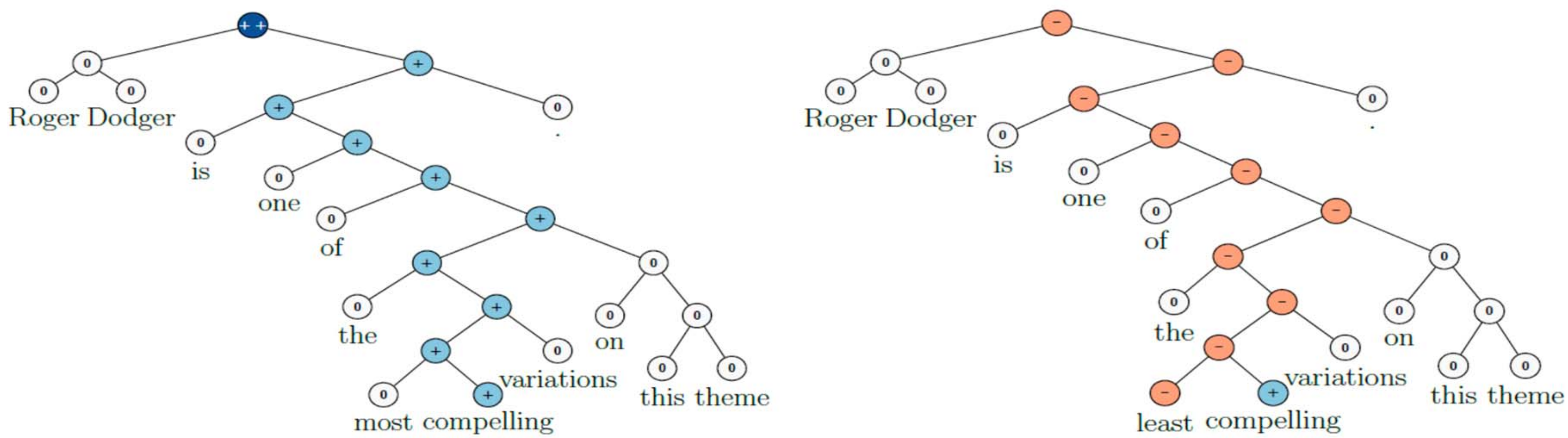




Source: http://cs224d.stanford.edu/lectures/CS224d-Lecture11.pdf

Sentiment Analysis – Neural Network

Negation Results

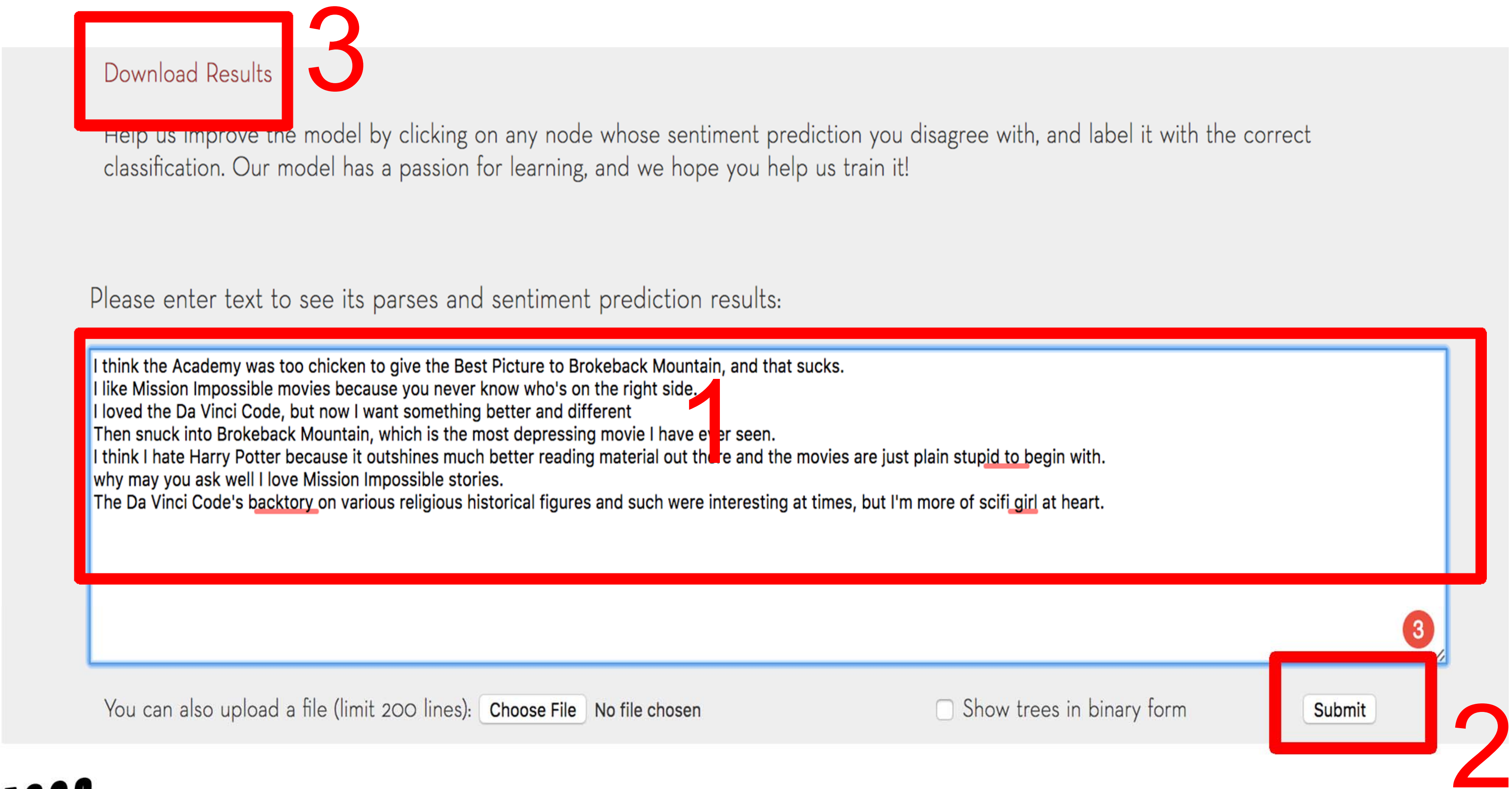


Source: http://cs224d.stanford.edu/lectures/CS224d-Lecture11.pdf

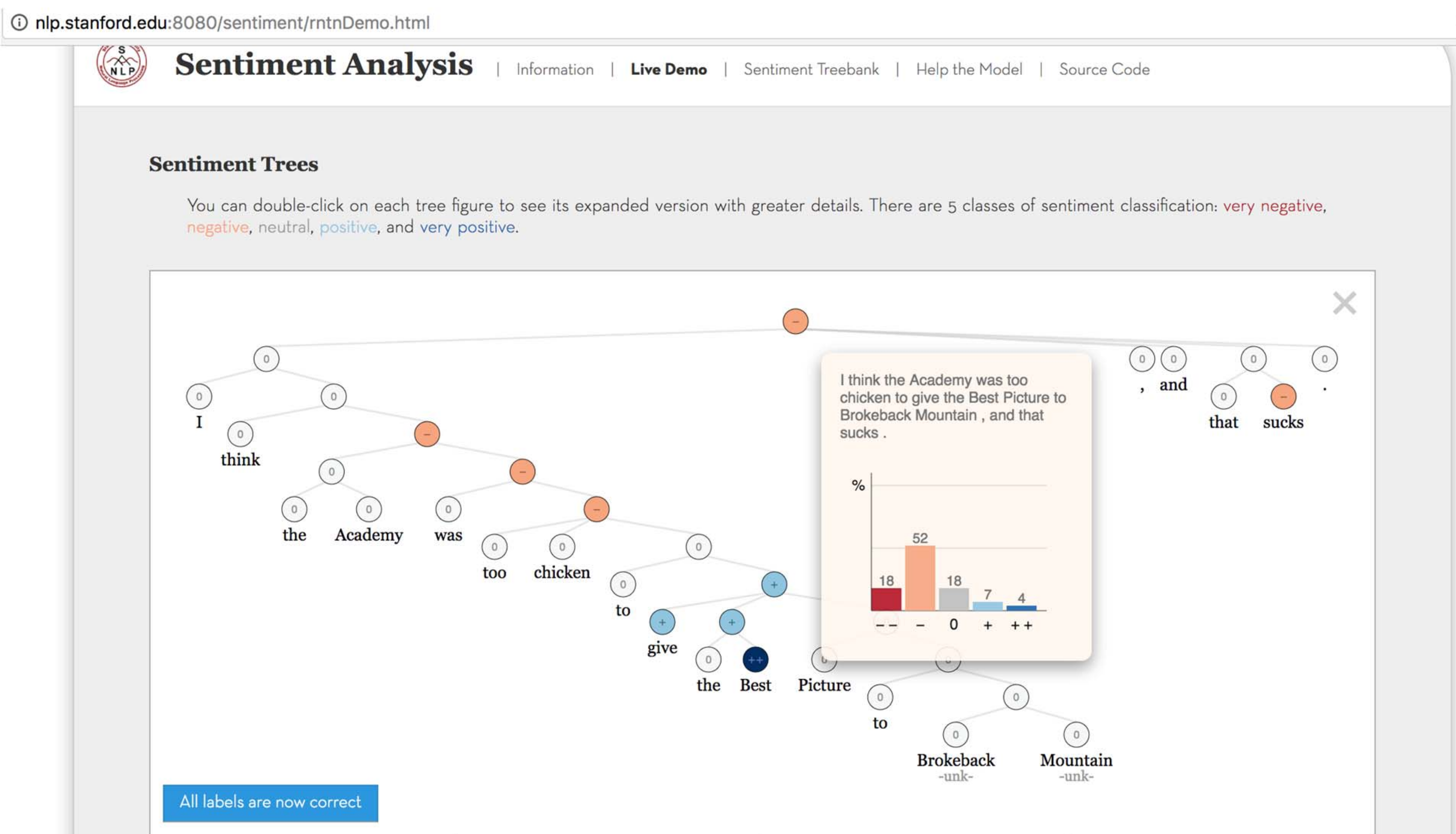
Movie review examples

- I think the Academy was too chicken to give the Best Picture to Brokeback Mountain, and that sucks.
- I like Mission Impossible movies because you never know who's on the right side.
- I loved the Da Vinci Code, but now I want something better and different
- Then snuck into Brokeback Mountain, which is the most depressing movie I have ever seen.
- I think I hate Harry Potter because it outshines much better reading material out there and the movies are just plain stupid to begin with.
- why may you ask well I love Mission Impossible stories.
- The Da Vinci Code's backtory on various religious historical figures and such were interesting at times, but I'm more of scifi girl at heart.

SA using Stanford NN: Input Data



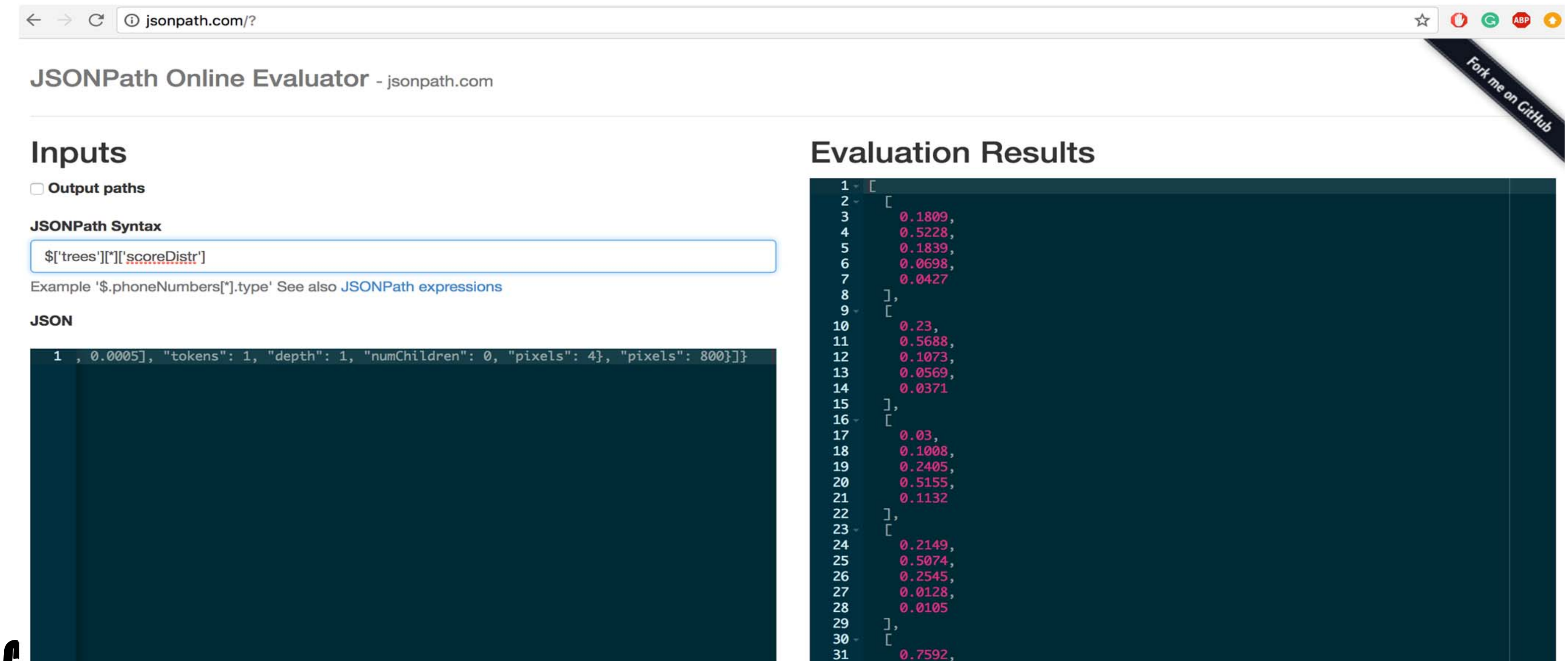
SA using Stanford NN: Input Data



Carnegie Mellon
isr institute for SOFTWARE RESEARCH

SA using Stanford NN: Get Scores

1. **[Go to website](http://jsonpath.com/)** http://jsonpath.com/?
2. Change formula: $['trees'][*]['scoreDistr']
3. Copy output



SA using Stanford NN: Get CSV

1. Got o https://konklone.io/json/
2. Paste output of the last page as input
3. Copy output table

Secure https://konklone.io/json/

Convert JSON to CSV

Click your JSON below to edit. Create a permalink any time. Please report bugs and send feedback on GitHub. Made by @konklone.


```
[
  [
    0.1809,
    0.5228,
    0.1839,
    0.0698,
    0.0427
  ],
  [
    0.23,
    0.5688,
    0.1073,
```


Extremely large files may cause trouble — the conversion is done inside your browser.

Below are the first few rows (7 total). Download the entire CSV, or show the raw data.

0	1	2	3	4
0.1809	0.5228	0.1839	0.0698	0.0427
0.23	0.5688	0.1073	0.0569	0.0371
0.03	0.1008	0.2405	0.5155	0.1132
0.2149	0.5074	0.2545	0.0128	0.0105
0.7592	0.172	0.0559	0.0022	0.0108
0.0339	0.1174	0.1329	0.5841	0.1317
0.1213	0.5455	0.2396	0.0676	0.026

SA using Stanford NN: Use Excel to get the final Sentiment Score

- Use Sentiment Calculator Excel sheet
- Paste the result from the last page
- Sheet uses a formula to calculate the final score

Sno	Text	-2	-1	0	1	2	Score
1	"I think the Academy was too chicken to give the Best Picture to Brokeback Mountain , and that sucks .",	0.1809	0.5228	0.1839	0.0698	0.0427	-0.73
2	"I like Mission Impossible movies because you never know who 's on the right side .",	0.23	0.5688	0.1073	0.0569	0.0371	-0.90
3	"I loved the Da Vinci Code , but now I want something better and different",	0.03	0.1008	0.2405	0.5155	0.1132	0.58
4	"Then snuck into Brokeback Mountain , which is the most depressing movie I have ever seen .",	0.2149	0.5074	0.2545	0.0128	0.0105	-0.90
5	"I think I hate Harry Potter because it outshines much better reading material out there and the movies are just plain stupid to begin with .",	0.7592	0.172	0.0559	0.0022	0.0108	-1.67
6	"why may you ask well I love Mission Impossible stories .",	0.0339	0.1174	0.1329	0.5841	0.1317	0.66
7	"The Da Vinci Code 's backtory on various religious historical figures and such were interesting at times , but I 'm more of scifi girl at heart ."	0.1213	0.5455	0.2396	0.0676	0.026	-0.67

SA using Stanford NN: Useful Links

- http://nlp.stanford.edu:8080/sentiment/rntnDemo.html
- http://www.jsoneditoronline.org/
- http://www.convertcsv.com/json-to-csv.htm
- http://sentistrength.wlv.ac.uk/
- http://boston.lti.cs.cmu.edu/classes/95-865-K/HW/HW3/ (Sentiment datasets)
- http://jsonpath.com/?
  – $['trees'][*]['text']
  – $['trees'][*]['scoreDistr']
- https://konklone.io/json/

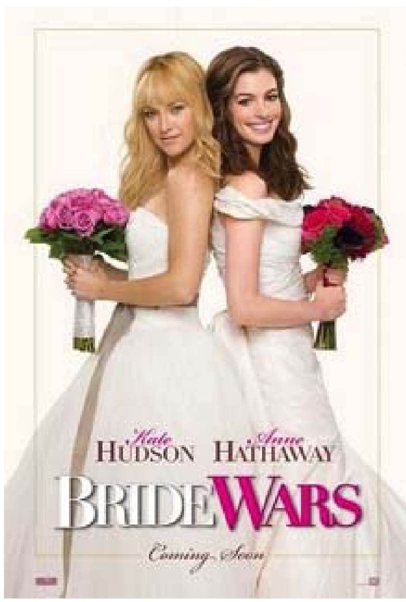
Evaluation

- How well the result agrees with human judgment?
- Human raters typically agree 79% #
- Recursive Tensor Network in around 85%
- Vader Sentiment Analysis Performance : 96% (F1)
  - May be overfitting

#Ogneva, M. "How Companies Can Use Sentiment Analysis to Improve Their Business". Mashable. Retrieved 2012-12-13.

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Anne Hathaway and Warren Buffett have recently been linked in the media—though not romantically, thank god!!



Oct. 3, 2008—Rachel Getting Married opens: BRK.A up .44%

Jan. 5, 2009—Bride Wars opens: BRK.A up 2.61%

Feb. 8, 2010—Valentine's Day opens: BRK.A up 1.01%

March 5, 2010—Alice in Wonderland opens: BRK.A up .74%

Nov. 24, 2010—Love and Other Drugs opens: BRK.A up 1.62%

Nov. 29, 2010—Anne announced as co-host of the Oscars: BRK.A up .25%

Source http://www.cnbc.com/id/42305525

Thank You

37

Backups

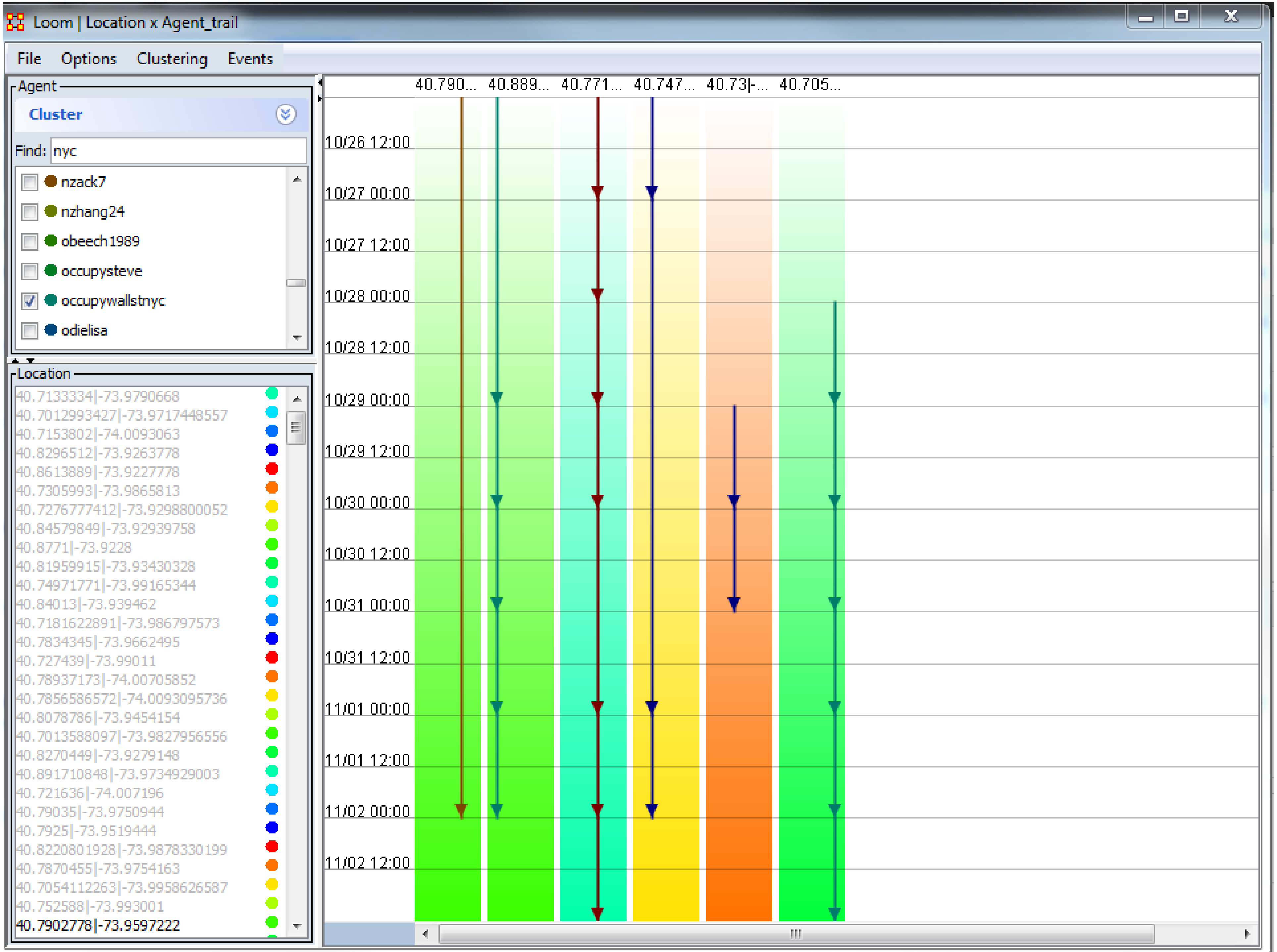
Can We Build Networks from Text Sentiment?

Movement of Agents over time

- First, we have to transpose the LocationxAgent network
- Then, lets find the most important agents in the AA network (Key Entity Report)

Rank	2012/10/26 00:00:00 - Transformed		2012/10/27 00:00:00 - Transformed		2012/10/28 00:00:00 - Transformed		2012/10/29 00:00:00 - Transformed		2012/10/30 00:00:00 - Transformed		2012/10/31 00:00:00 - Transformed		2012/11/01 00:00:00 - Transformed		2012/11/02 00:00:00 - Transformed		2012/11/03 00:00:00 - Transformed	
1	holynamemedctr	0.006	lowereastnyc	0.007	MikeBloomberg	0.001	nyc8675309	6.064e-004	ConEdison	0.002	roz233	9.172e-004	psegdelivers	7.010e-004	ConEdison	0.003	PSEGdelivers	0.002
2	teaneckpatch	0.003	AHurricaneSandy	0.003	irelandinny	0.001	nytimes	4.331e-004	nytimes	0.002	conedison	7.338e-004	NYCMayorsOffice	5.791e-004	mamalukapr	0.001	ansontang	0.002
3	AHurricaneSandy	0.003	spookyshorty	0.002	occupywallstnyc	0.001	cnnbrk	3.754e-004	MikeBloomberg	0.001	downtownnyc	7.338e-004	jackdoylesnyc	5.791e-004	PSEGdelivers	0.001	ConEdison	0.002
4	NJOEM2010	0.002	AP	0.001	hurricannesandy	0.001	MTAInsider	3.754e-004	cimages	0.001	PSEGdelivers	5.503e-004	conedison	5.486e-004	NewYorkPost	0.001	amerigochattin	0.002
5	dwavy	0.002	milesnessuno	0.001	peepcheeks	0.001	spookyshorty	3.176e-004	srubenfeld	0.001	spookyshorty	5.045e-004	mikebloomberg	4.572e-004	NYCMayorsOffice	0.001	downtownnyc	0.001
6	teanecknjgov	0.002	AFrankenStorm	0.001	SandysHurricane	0.001	queenesther	2.888e-004	PSEGdelivers	0.001	earlymorningrun	4.586e-004	portermason	4.572e-004	MichaelSkolnik	0.001	theparlourbk	0.001
7	nyc8675309	0.002	BigFrankenStorm	0.001	jacqui_stafford	7.607e-004	breakingstorm	2.599e-004	mamalukapr	9.370e-004	MTAInsider	4.586e-004	nyrsrelieveny	3.657e-004	SonsOfEssex	0.001	tai_morshed	0.001
8	shrinkthinks	0.002	fdo_cabezas	8.455e-004	elliotthetrainr	7.607e-004	thisisonlytemp1	2.310e-004	ryanhatesthis	8.519e-004	WSJ	4.127e-004	maryjnico	3.048e-004	NYGovCuomo	9.517e-004	MTAInsider	0.001
9	princexo_	0.002	hughlaurious	8.455e-004	dtrotbot	7.607e-004	nynjpaweather	2.310e-004	cnnbrk	7.667e-004	hankishtwit	4.127e-004	JD	3.048e-004	portermason	8.053e-004	mamalukapr	7.660e-004
10	mildlybitter	0.002	cimages	8.455e-004	dopeusername	7.607e-004	hurricannesandy	2.021e-004	harrisgraber	7.667e-004	msecls	4.127e-004	roxannaparadise	2.743e-004	nysportsfan92	7.321e-004	veronicaromm	6.702e-004

- Now, we can see how they've moved



Loom | Location x Agent_trail
File
Options
Clustering
Events
Agent
Cluster
Find: nyc
nzack7
nzhang24
obeech1989
occupysteve
occupywallstnyc
odielisa
Location
40.7133334|-73.9790668
40.7012993427|-73.9717448557
40.7153802|-74.0093063
40.8296512|-73.9263778
40.8613889|-73.9227778
40.7305993|-73.9865813
40.7276777412|-73.9298800052
40.84579849|-73.92939758
40.8771|-73.9228
40.81959915|-73.93430328
40.74971771|-73.99165344
40.84013|-73.939462
40.7181622891|-73.986797573
40.7834345|-73.9662495
40.727439|-73.99011
40.78937173|-74.00705852
40.7856586572|-74.0093095736
40.8078786|-73.9454154
40.7013588097|-73.9827956556
40.8270449|-73.9279148
40.891710848|-73.9734929003
40.721636|-74.007196
40.79035|-73.9750944
40.7925|-73.9519444
40.8220801928|-73.9878330199
40.7870455|-73.9754163
40.7054112263|-73.9958626587
40.752588|-73.993001
40.7902778|-73.9597222
40.790...
40.889...
40.771...
40.747...
40.73|-...
40.705...
10/26 12:00
10/27 00:00
10/27 12:00
10/28 00:00
10/28 12:00
10/29 00:00
10/29 12:00
10/30 00:00
10/30 12:00
10/31 00:00
10/31 12:00
11/01 00:00
11/01 12:00
11/02 00:00
11/02 12:00

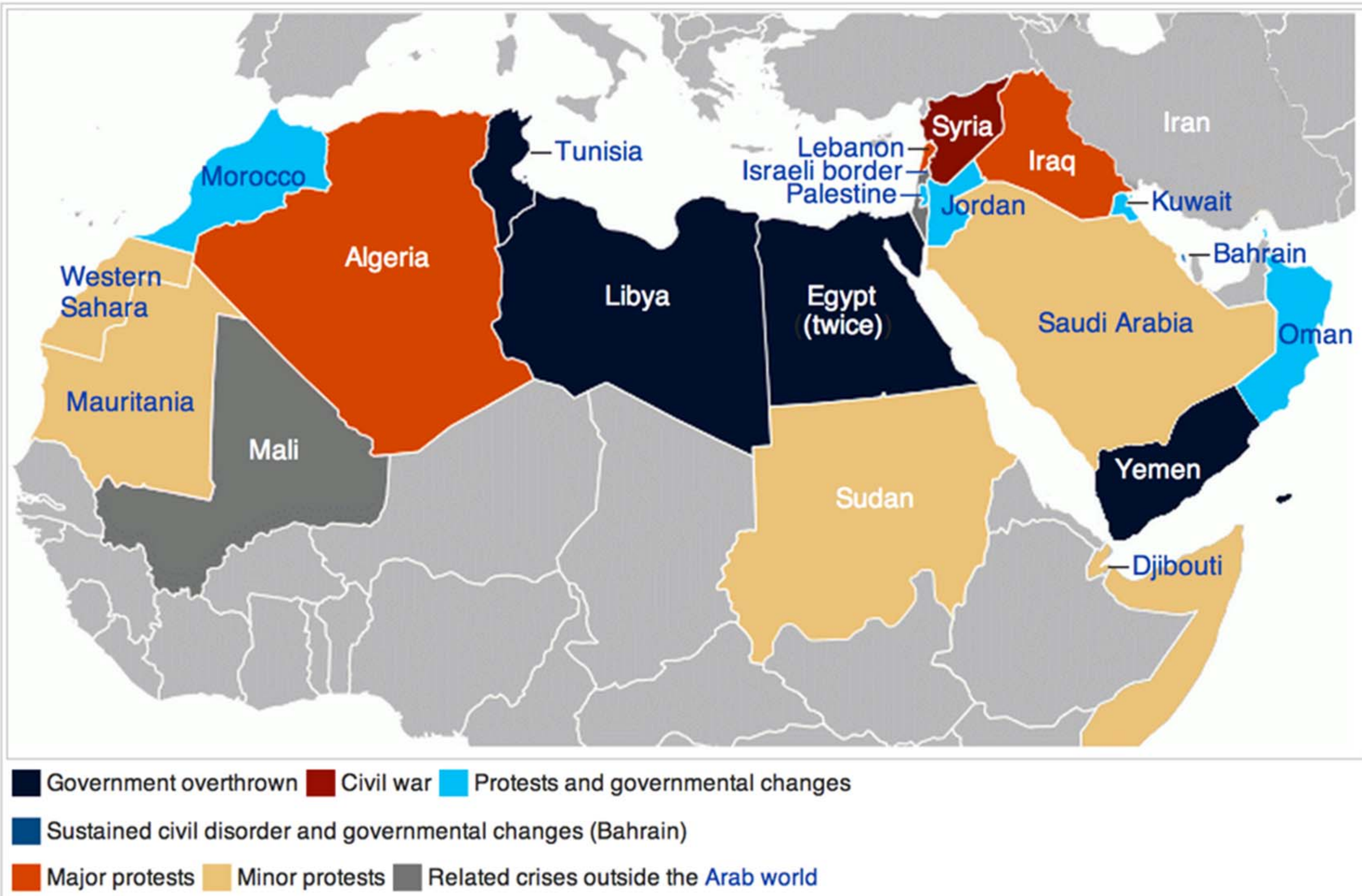
Part 2

Analysis of Arab Spring data

Arab Spring

- A revolutionary wave of demonstrations in the Arab countries
  - Began on 18 December 2010
  - Rulers had been forced from power in Tunisia, Egypt (twice), Libya and Yemen
  - Civil uprisings erupted in various countries.
- Social networks play key roles in the evolutions
  - Twitter/facebook are used to organize demonstrations/protects.
  - “digital democracy” brought by social media.

Revolution Demonstration from Wikipedia

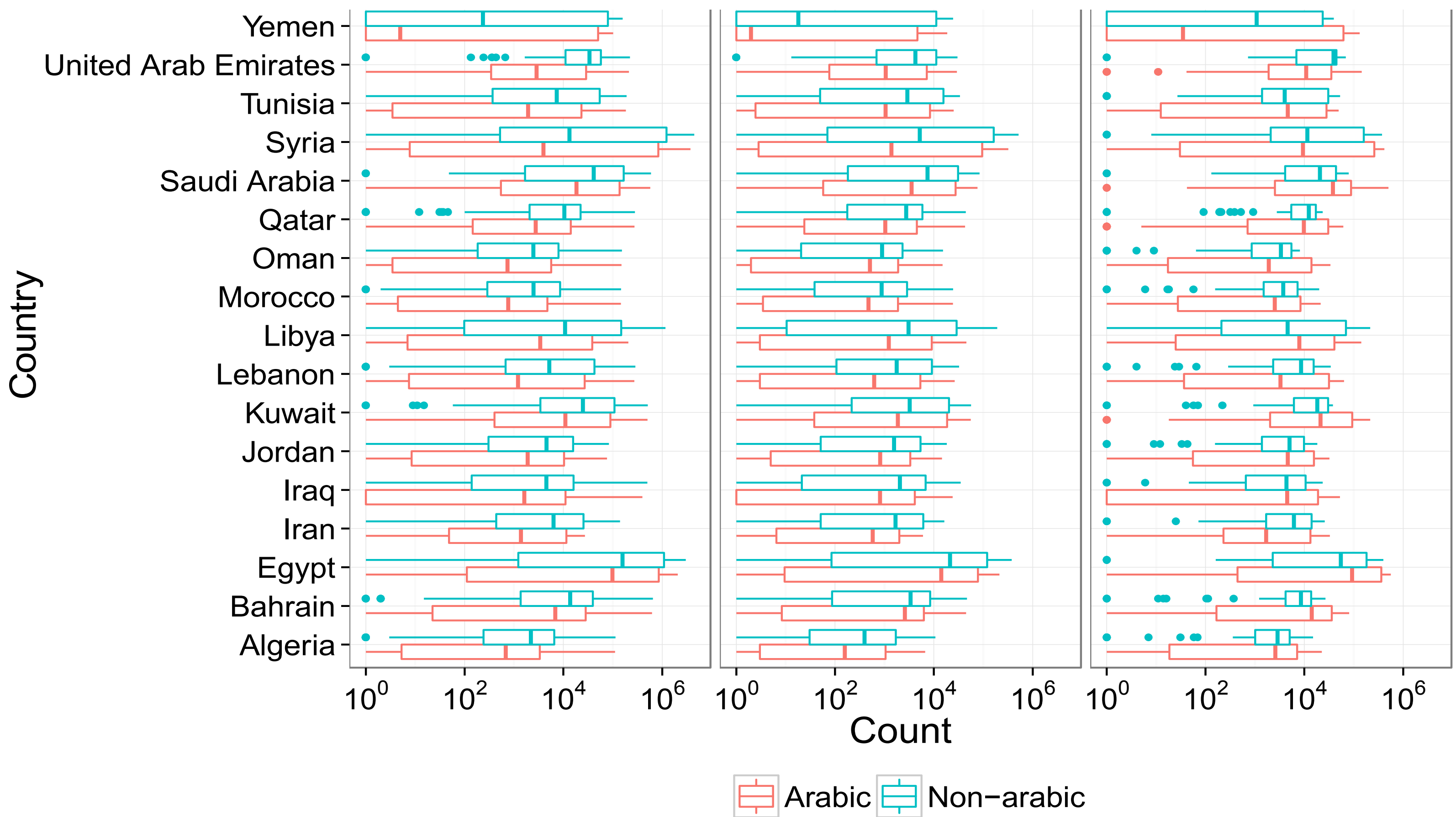


Data Collection

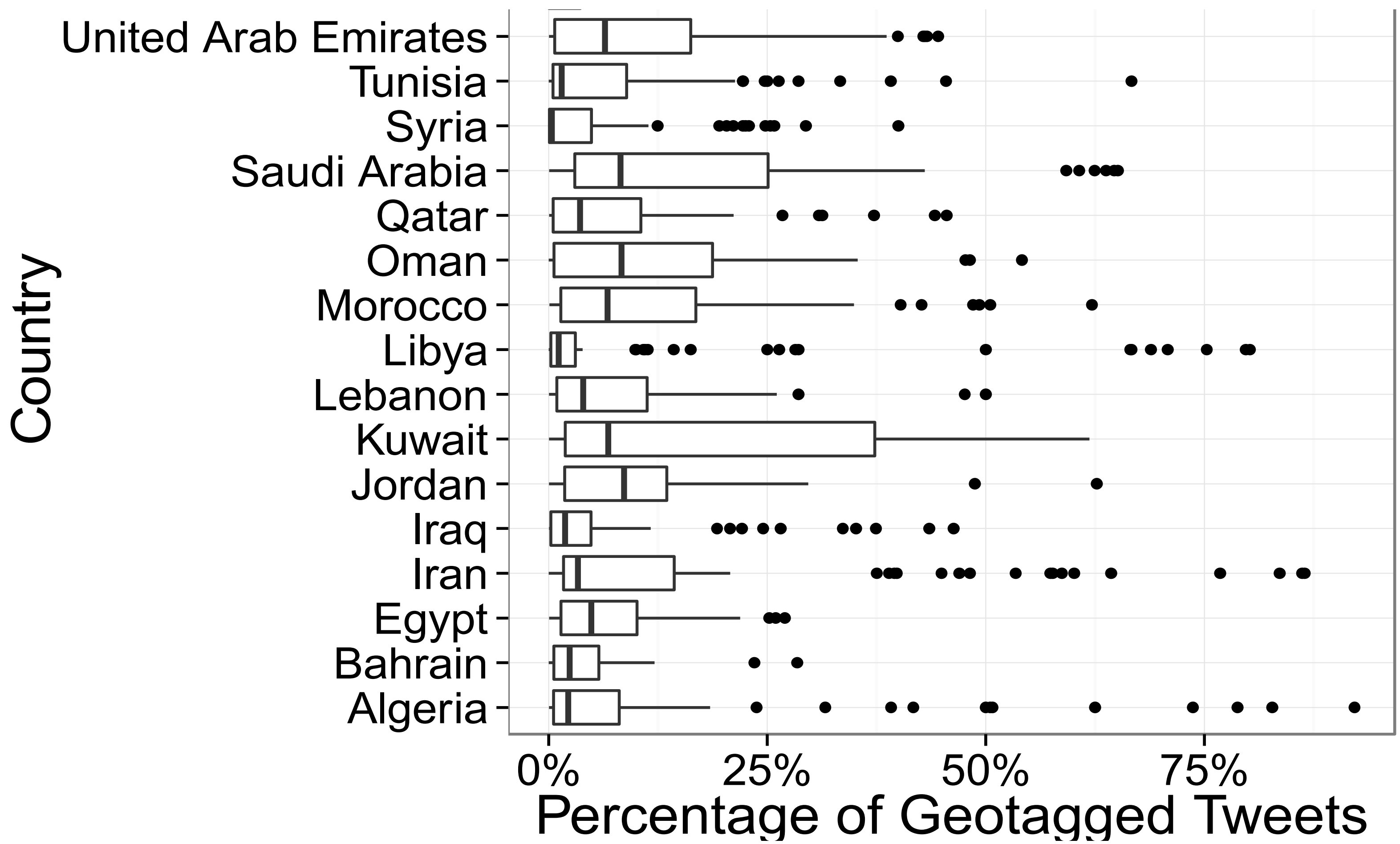
- A huge amount of twitter data has been collected
  - A collector is running continuously to collect data from Apr,2009 to Nov,2013 using a geo location bounding box in the Arab Spring region.
  - Data are collected through public API of twitter.
  - Up to 10% of all the twitter in the geo bounding box is collected.
- Texts are processed using two different ways:
  - Only Arabic texts
  - Only English texts

Carnegie Mellon
institute for SOFTWARE RESEARCH

Num. tweet Statistics by Country



Percentage of Geotagged Tweets by Country



Identifying Topic Groups

- Latent Dirichlet Allocation(LDA) is used to identify topic groups based on texts of the twitter message
  - The tweets of all users are aggregated to form a single "document".
  - "topics" are identifies by looking at following a probabilistic graphical model.
  - Each "topic" is represented by a group of terms.
  - Each "document" is scored over different topics.

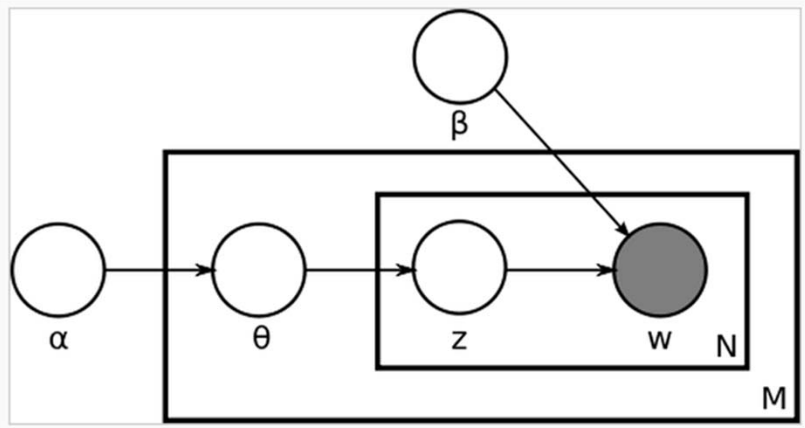
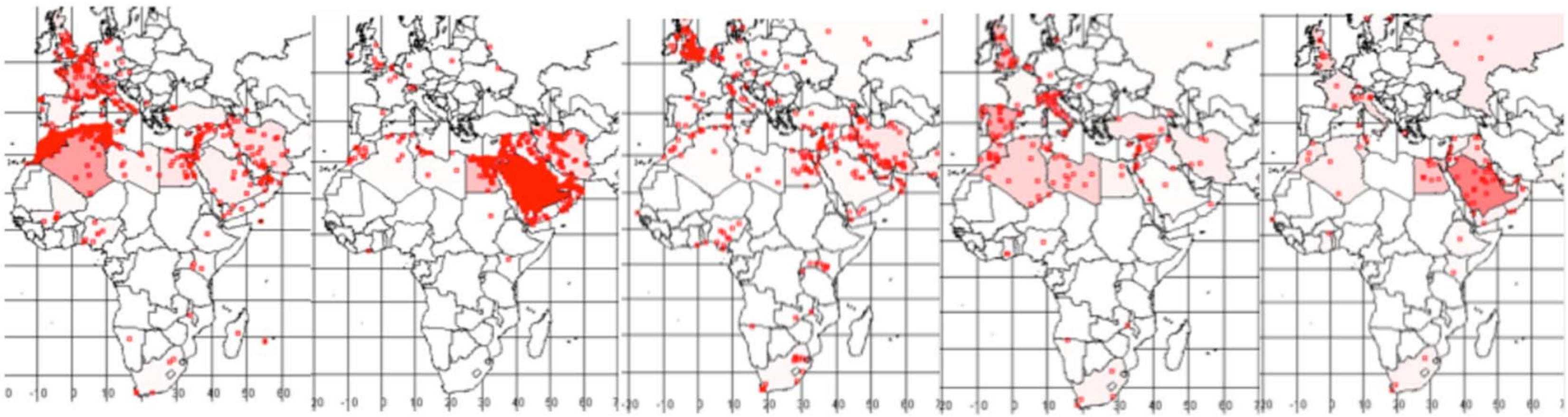


Plate notation representing the LDA model.

Locality of Topic Groups.



Topic 40 (English)	Topic 46 (English)	Topic 91 (Arabic)	Topic 92 (Arabic)	Topic 98 (English)
social	american	Arab	Good	follow
happening	east	Spring	Peace	sporting
african	beirut	Gulf	Roses	mosh
families	jordan	Ali	Special	coast
meridien	middle	Tyrant	Possible	liam

Change of Topics Overtime

Date	Topic	Date	Topic	Date	Topic	Date	Topic	Date	Topic
2010 10	94	2011 5	46	2011 12	46	2012 7	46	2013 2	62
2010 11	94	2011 6	46	2012 1	46	2012 8	46	2013 4	62
2010 12	94	2011 7	46	2012 2	46	2012 9	46	2013 5	62
2011 1	94	2011 8	46	2012 3	46	2012 10	62	2013 9	62
2011 2	74	2011 9	46	2012 4	46	2012 11	62	2013 10	62
2011 3	41	2011 10	46	2012 5	46	2012 12	62	2013 11	62
2011 4	41	2011 11	46	2012 6	46	2013 1	62		

Persistent Topics Overtime

Topic 94 (Arabic)	Topic 41 (Arabic)	Topic 46 (English)	Topic 62 (Arabic)
people	Arabs	american	Egypt
god	people	east	Mursi
life	country	information	Head
solutions	beloved	Beirut	people
even	what is happening	jordan	President

Generating Networks

- An agent X agent network is generated
  - Each agent has a topic score vector generated by the LDA algorithm.
  - Agent to agent relations are generated by calculating a similarity score between the topic score vectors.
  - For scalibility issues, we choose a score that is eficient to calculate:

$$Sim_{i,j} = \frac{\min\left(|v_1|, |v_2|\right)}{\max\left(|v_1|, |v_2|\right)}$$

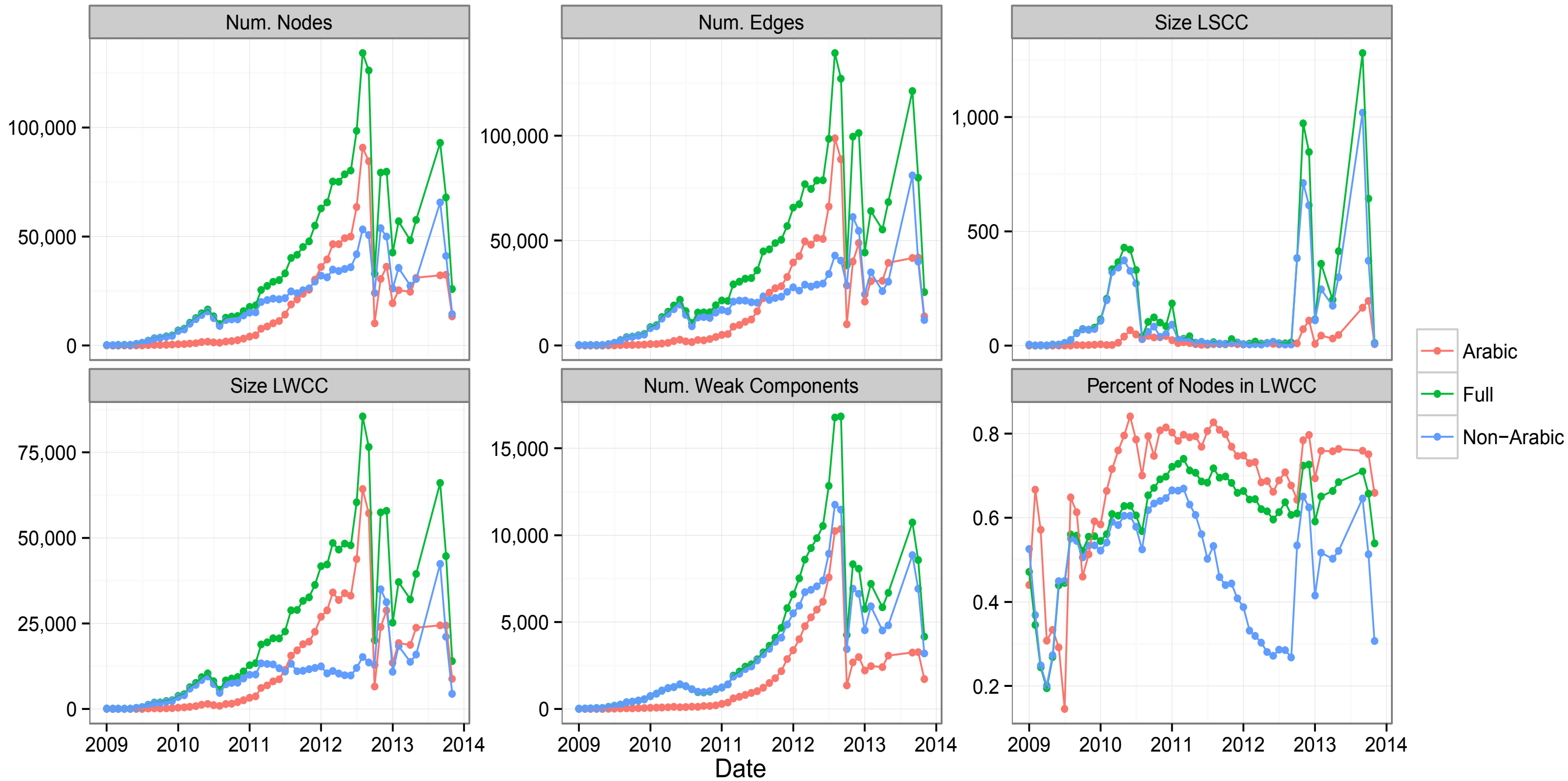
Network Metrics—Original Data Set

Country	nodes	edges	density
Bahrain	4559	206642	0.012
Qatar	6948	378262	0.016
Iraq	1852	42257	0.024
Iran	975	6304	0.013
Libya	4394	110910	0.011
Algeria	780	5913	0.019
Egypt	42060	9490034	0.011
Kuwait	19713	6087116	0.031
Lebanon	6687	226560	0.010
Morocco	5612	258507	0.016
Jordan	3711	79486	0.012
Saudi Arabia	33663	35921282	0.063
Oman	2193	45820	0.019
Syria	40625	8603652	0.010
Yemen	1109	84280	0.137
United Arab Emirates	24417	3542578	0.012
Tunisia	3692	63728	0.009

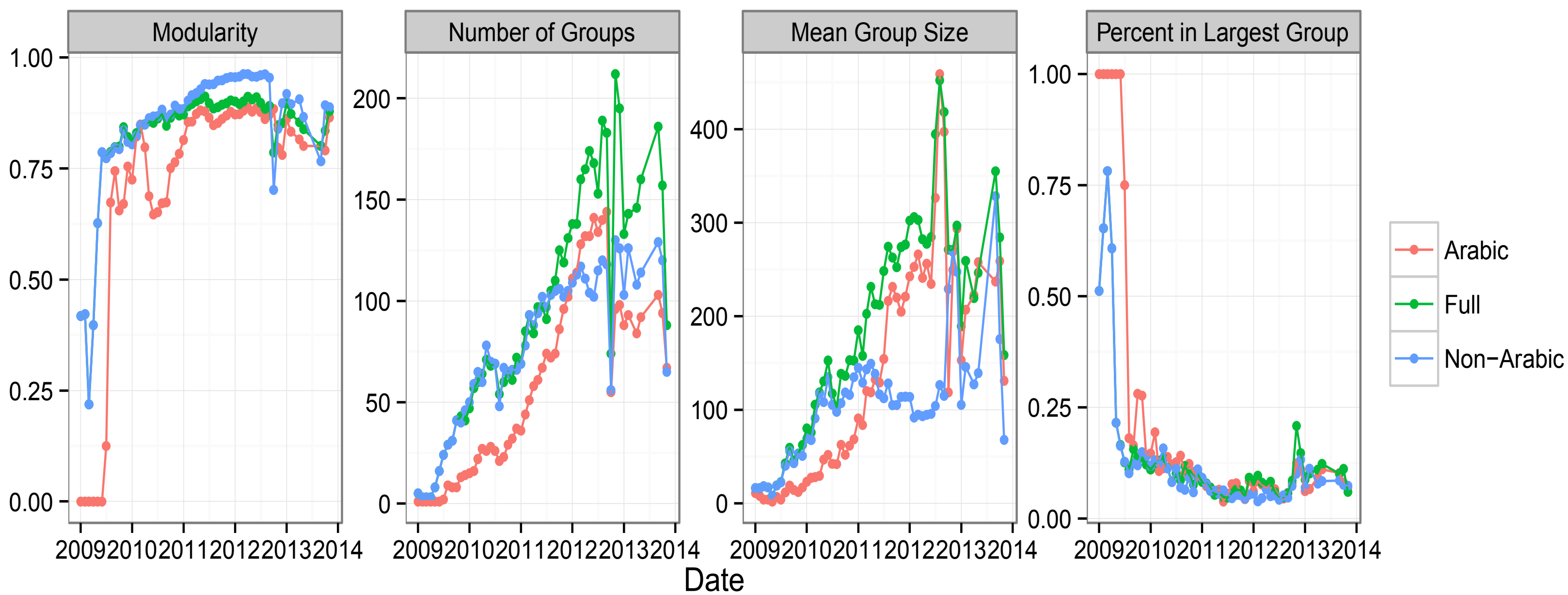
Network Metrics—English Only Data Set

country	nodes	edges	density
bahrain	8698	149612	0.004
qatar	10721	230981	0.004
iraq	3295	23008	0.004
iran	1344	4998	0.006
libya	5259	88827	0.006
algeria	955	5134	0.011
egypt	62653	7964548	0.004
kuwait	45955	4273476	0.004
lebanon	7722	171573	0.006
morocco	6689	157733	0.007
jordan	4887	61438	0.005
saudi_arabia	136543	46843301	0.005
oman	4491	71297	0.007
syria	53350	7042616	0.005
yemen	6000	131767	0.007
uae	33448	3155592	0.006
tunisia	4253	49105	0.005

Component Groups in Twitter Reply Network



Metrics for Network Grouping



Conclusions

- Geo/temporal locality of topic groups
  - Different geo locations have different focuses of topic groups.
  - The overall most popular topic groups change over time.
- Relations in the co-topic network is an good measure of activities in a country.
  - Countries have more relations are more likely to have revolutions.