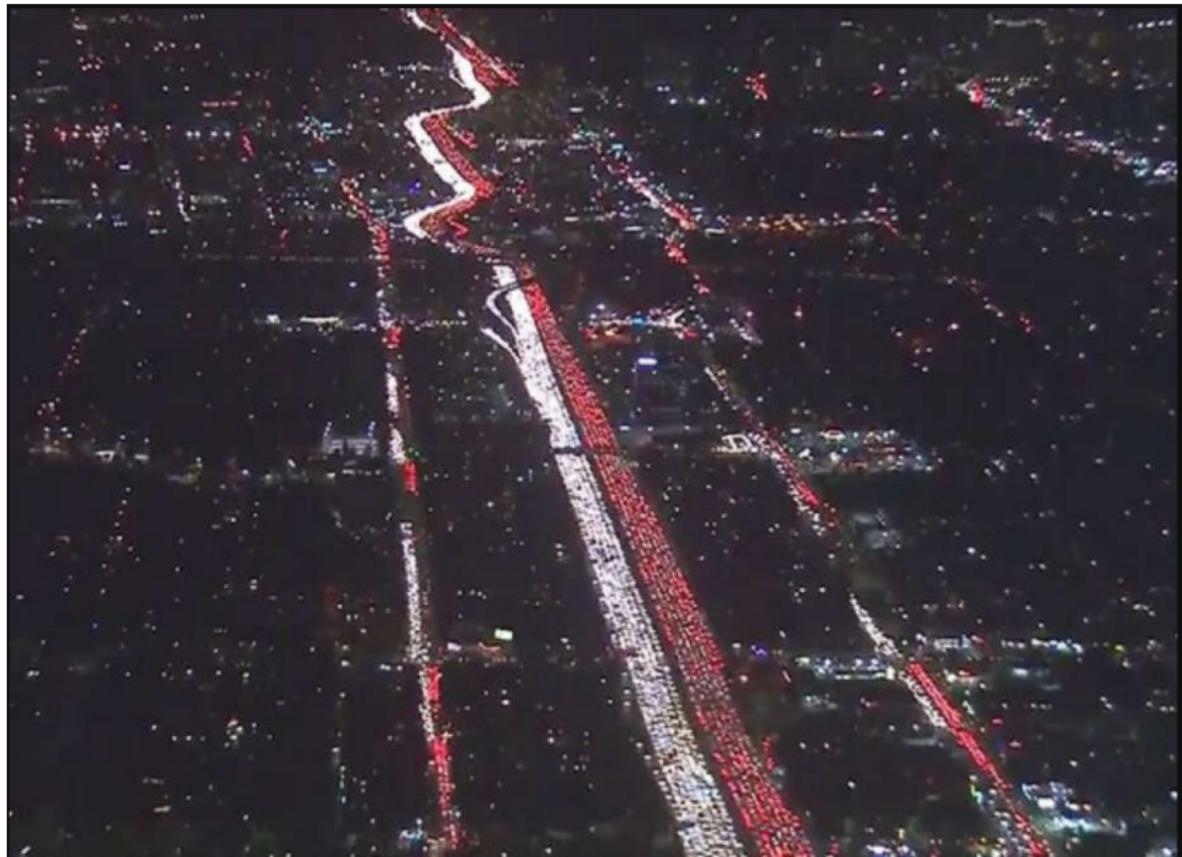

Data Mining on Twitter for Improving Public Safety and Health

Alfred George Francisco Milan III¹

University of California, Santa Cruz

December 1, 2017

Introduction



Introduction(Continued)



Background



Data Collection

location
CM CityMole Chicago - 2015-09-11 10:00:00
IL-59 Northbound Boulevard Accident at IL-64 North [..] cmole.com/1Qtc

traffic event
Chicago Traffic Tracker - 4/10/15
Heavy Traffic on SB Western: Kennedy Expy to Fullerton. 07:05 pm 09/11/2015
goo.gl/fb/cKuRuw

time
Total Traffic LA - TotalTrafficLA - 59
Disabled vehicle, right lane blocked in #WestCovinaSanDimas on I-10 EB at Vila Verde #LAttraffic bit.ly/1OF3951

Accident

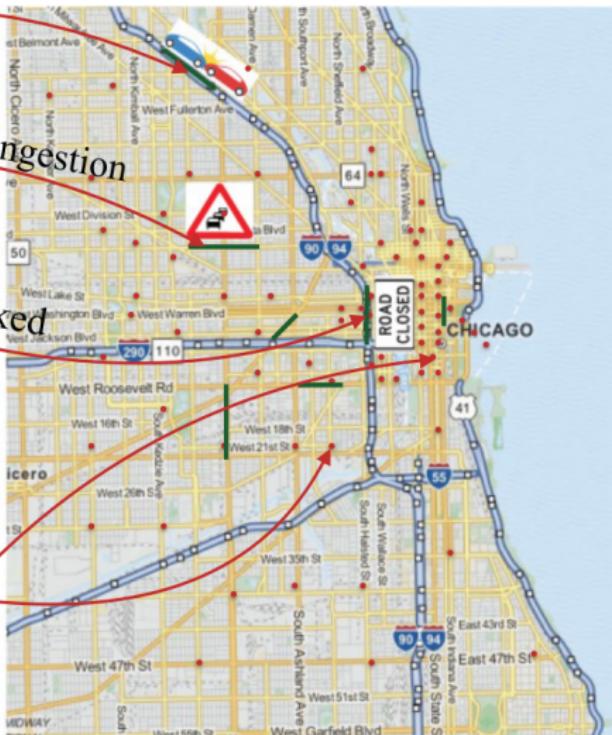
Congestion

Blocked

Traffic Related Twitter Data

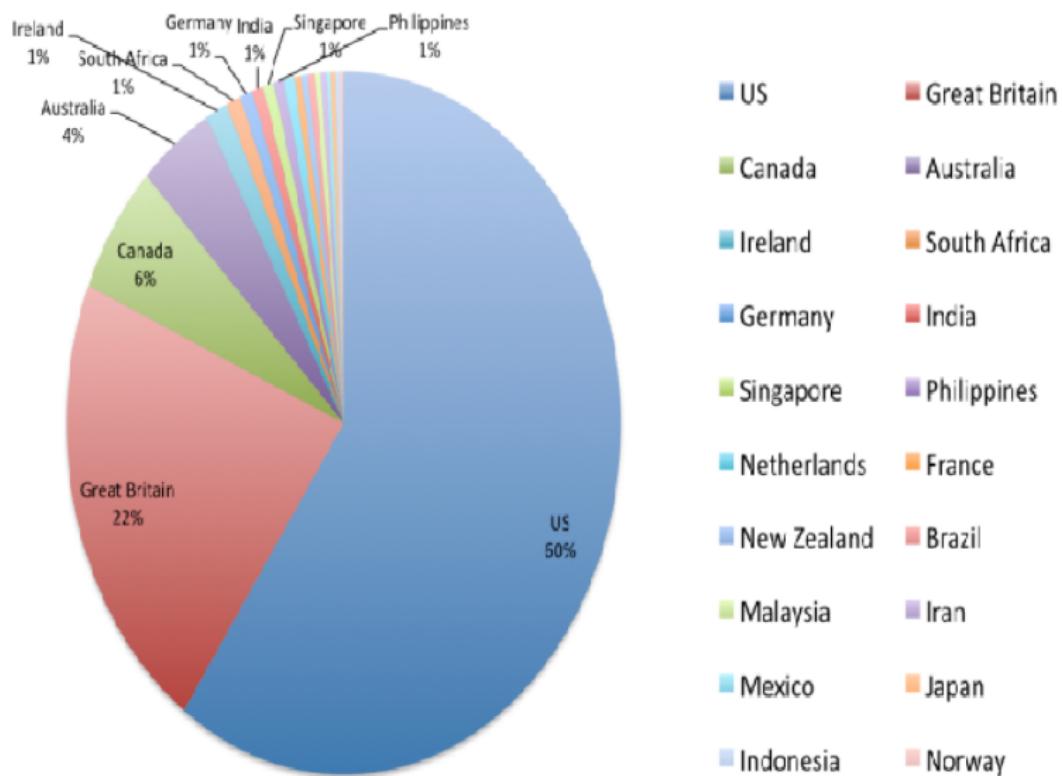
time	location	speed
Fri Dec 12 19:53:13 UTC 2014	41.96803056,-87.74465194,312,0	
Fri Dec 12 20:06:43 UTC 2014	41.97609111,-87.78584528,312,0	
Sun Dec 07 15:24:48 UTC 2014	41.84599972,-87.74363556,0,11	
Tue Dec 02 20:56:43 UTC 2014	41.84628917,-87.7435325,88,0	
Fri Dec 05 20:52:55 UTC 2014	41.76479778,-87.7930775,101,22	
Fri Dec 05 20:58:38 UTC 2014	41.76385667,-87.772915,3,0	
Fri Dec 05 21:01:05 UTC 2014	41.76378139,-87.77299639,26,9	
Thu Dec 04 14:26:03 UTC 2014	41.81607722,-87.62998194,338,45	
Thu Dec 04 14:51:06 UTC 2014	41.86069361,-87.67870111,214,9	
Thu Dec 04 15:19:53 UTC 2014	41.79219333,-87.80153972,178,9	
Thu Dec 04 15:18:10 UTC 2014	41.74818556,-87.79886667,207,9	
Thu Dec 11 14:22:03 UTC 2014	41.80568861,-87.63679417,312,0	
Thu Dec 11 14:38:49 UTC 2014	41.79579528,-87.63783889,85,20	
Thu Dec 11 15:03:29 UTC 2014	41.78994306,-87.63157111,111,16	
Thu Dec 11 18:31:02 UTC 2014	41.75774333,-87.62539139,338,96	
Thu Dec 11 18:41:39 UTC 2014	41.80654917,-87.63945528,212,0	
Thu Dec 11 18:59:28 UTC 2014	41.8089425,-87.63819639,85,95	

Probe GPS Data



Map of Downtown Chicago

Data Collection (Continued)



Total Users Sampled = 140,077

Total Locations Matched = 44,276

Data Collection (Continued)



event time location

Concert added: sch.mp/adPEt - PT
@themizzi The Mizerables has a show on
05/24/2015 at 08:00 PM @ Beat Kitchen in
Chicago, IL h...

event time location

Show added: sch.mp/aclZB - RT
@Sonic7Rock Sonic 7 has a show on Sat
#Indy500 #KingOfTheRoad #BlackRockCity AnneMarie #cbysara #DearLouise JohnNash
7:30-11pm. Outsi...

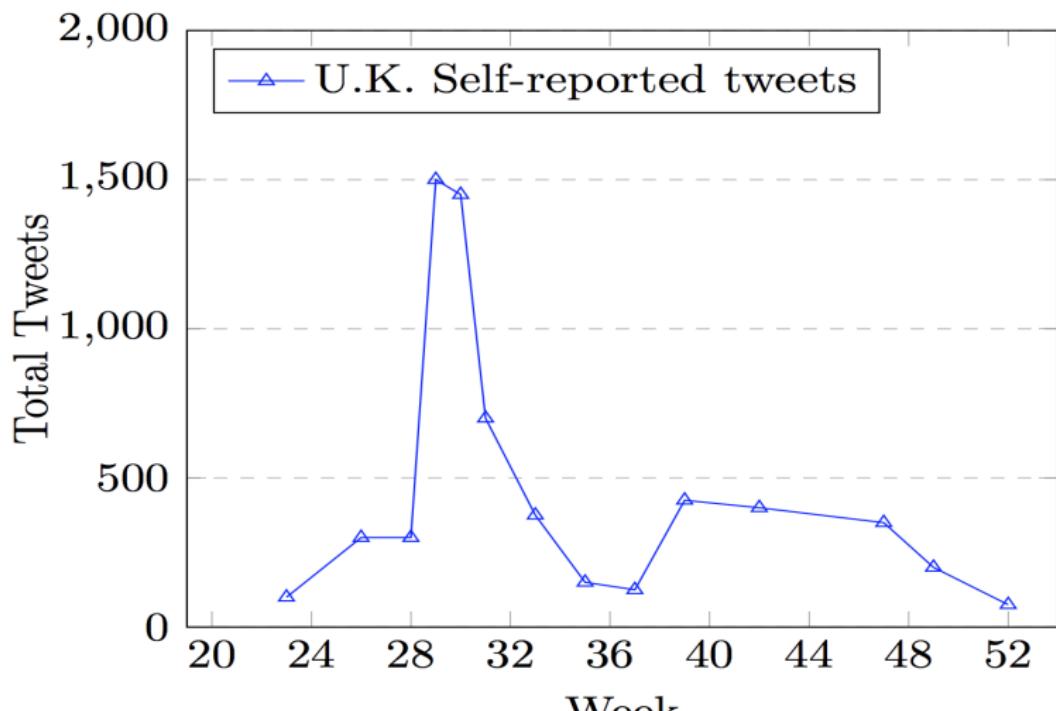
- ▶ Identify two things:
 1. type of traffic event
 2. geocoding

Data Collection (Continued)

$$I(r_i, se_j) = f(loc_{r_i}; loc_{se_j}, \sum_{se_j}) = \frac{1}{2\pi |\sum_{se_j}^{1/2}|} \times e^{-1/2(loc_{r_i} - loc_{se_j})T} \\ \times \sum_{se_j}^{-1} (loc_{r_i} - loc_{se_j})(1)$$

- ▶ r_i = road segment
- ▶ se_j = social event
- ▶ loc_{r_i} = location of the road segment
- ▶ loc_{se_j} is the location of the social event

Experiment



U.K. Self-Reporting Tweets

Experiment (Continued)

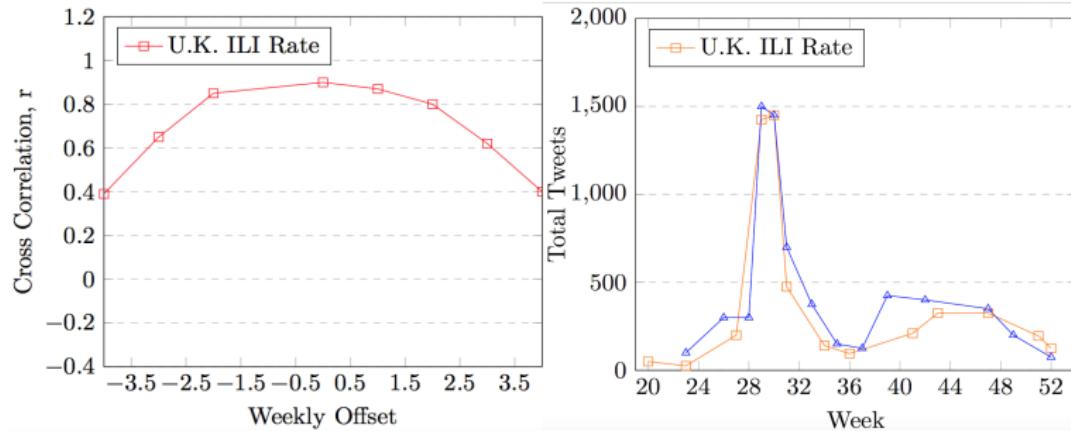
$$r = \frac{\sum_t (x(t) - \bar{x}) \times (y(t-i) - \bar{y})}{\sqrt{\sum_t (x(t) - \bar{x})^2} \times \sqrt{\sum_t (y(t-i) - \bar{y})^2}} \quad (2)$$

- ▶ for $i = 0$, shows the correlation between the two data signals
- ▶ for $i = -1$, shows that the first signal determines the second signal

During the interval (-1, 1):

- ▶ when $r = -1$, signifies opposite signs despite having similar shapes.
- ▶ when $r = 0$, there is no correlation
- ▶ when $r = 1$, there is a strong correlation and near-identical shape between the two signals.

Results



1. Very strong Correlation between Self Reported Tweets in the United Kingdom and H.P.A.
2. 3.8% error among a sample of 1,000 tweets
3. The normalization process used is effective

Conclusion

- ▶ Twitter is an accurate data source
- ▶ GPS Probe data and surveillance data is obsolete compared to Twitter
- ▶ Twitter provides real-time information
- ▶ Twitter allows public officials to react faster to prevent next outbreaks and address traffic issues

References

- ▶ Kostkova, P., Szomszor, M., and St. Louis, C. 2014. swineflu: The use of twitter as an early warning and risk communication tool in the 2009 swine flu pandemic. *ACM Trans. Manage. Inf. Syst.* 5, 2, Article 8 (July 2014), 25 pages.
DOI:<http://dx.doi.org/10.1145/2597892>
- ▶ Senzhang Wang, Xiaoming Zhang, Jianping Cao, Lifang He, Leon Stenneth, Philip S. Yu, Zhoujun Li, and Zhiqiu Huang. 2017. Computing urban traffic congestions by incorporating sparse GPS probe data and social media data. *ACM Trans. Inf. Syst.* 35, 4, Article 40 (July 2017), 30 pages. DOI:
<http://dx.doi.org/10.1145/3057281>