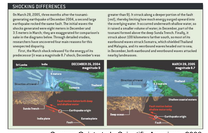
Towards Earthquake-Tsunami Causality via Data Science: Giraph-Derived Credibility Scores for Data from Twitter

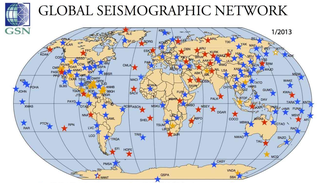
L. I. Lumb ianlumb@yorku.ca, ian.lumb@brightcomputing.com J. R. Freemantle james.freemantle@rogers.com

Motivation

- Earthquake (cause) - tsunami (effect)
  - Relationship uncertain
    - *Limited reaction time*
- Approach from a data-science perspective
  - Objective and subjective data sources
    - Twitter
      - Real-time, location-based, intensity-aware and mobile-enabled



Source: Geist et al., Scientific American, 2006

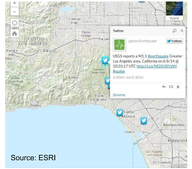


Source: http://www.iris.edu/

Perl script prototype

- Acquires tweets with the keyword "earthquake"

```
use Net::Twitter::Lite::WithAPIv1_1;
my $nt = Net::Twitter::Lite::WithAPIv1_1->new(
      consumer_key     => 'xxxx...xxxxxxx',
      consumer_secret  => 'xxxxxx.....xxxxxxxxxx',
  access_token         => 'xxxxx....xxxxxxxxxxx',
  access_token_secret  => 'xxxxx.....xxxxxxxxxxx',
                  ssl  => 1
);
my $result = $nt->search("earthquake");
for my $status(@{$result->{statuses}} ) {
  print "$status->{text}\n";
}
```



Source: ESRI

Created at: Mon Jun 09 18:29:43 +0000 2014
#Sismo #Earthquake M 2.4, 30km SE of Yosemite Valley, California http://t.co/a0YrmP4jTb
Created at: Mon Jun 09 18:28:54 +0000 2014
@dentmay I might be much of An Earthquake Expert. Don't know what's going to happen? Oh well I'll calm myself down..
Created at: Mon Jun 09 18:28:54 +0000 2014
Md 2.4 CENTRAL CALIFORNIA | Details: http://t.co/gEleS8D4X1 via EMSC©
Created at: Mon Jun 09 18:28:52 +0000 2014
mb 3.7 KURIL ISLANDS | Details: http://t.co/y9OMSjmG5 via EMSC©
Created at: Mon Jun 09 18:28:19 +0000 2014
Prepared for situations like an earthquake?How about your car?Stop by our New West branch and check out these items! http://t.co/5dMT1RfjjX
Created at: Mon Jun 09 18:28:16 +0000 2014
UkTuniTrance Team140+ Pres. Earthquake Sound System 061 (Cristian Ketelaars Guestmix) by UkTuniTranceTM http://t.co/4mbrucjjTj on #SoundCloud
Created at: Wed Jun 04 20:47:16 +0000 2014
@simplic10 more like I've been on the internet for over a decade and know that I've heard very few stories of "Quake player murdered"
Created at: Wed Jun 04 20:30:41 +0000 2014
After my brother-in-law experienced an #earthquake in Peru yesterday his wife discovered just how many #BC has too http://t.co/xaAjj2Tn8s

"Earthquake" Tweets

- Social tweets
- Information tweets
- Ad/spam tweets
- Other tweets
  - The game "Quake"
  - The movie "San Andreas"

Created at: Wed Jun 04 20:29:33 +0000 2014
5.0 earthquake! Thu Jun 05 02:04:27 GMT+09:00 2014 near 84km SW of Iquique, Chile http://t.co/mmFokGQWTT #earthquake
Created at: Wed Jun 04 20:30:13 +0000 2014
The #earthquake continues: Latest via @Spectator_CH /@YouGov -#Labour 36 #Tories 32%, LD 8%, #Ukip 14%. Implied Labour majority- 42 ..
Created at: Wed Jun 04 20:31:35 +0000 2014
#teremoto ML 2.7 CENTRAL ITALY. Magnitude ††ML 2.7 Region ††CENTRAL ITALY Date time ††2014-06-04 20:01:33.9 UTC... http://t.co/Y141OvuK6P
Created at: Wed Jun 04 20:31:23 +0000 2014
RT @MackHanson: An earthquake hitting while I'm pooping is easily in my Top Ten Fears. Along with a dinosaur attack while I'm pooping,
Created at: Tue Jun 10 12:22:42 +0000 2014
@BaddieKuntds I doubt it. Once I was like I hope you the baby and Josie are okay after the earthquake.
Created at: Tue Jun 10 12:22:34 +0000 2014
RT @TheRock: Just wrapped a massive post earthquake scene for SAN ANDREAS. To the hundreds of background actors/extras.. THANK U for all yo...

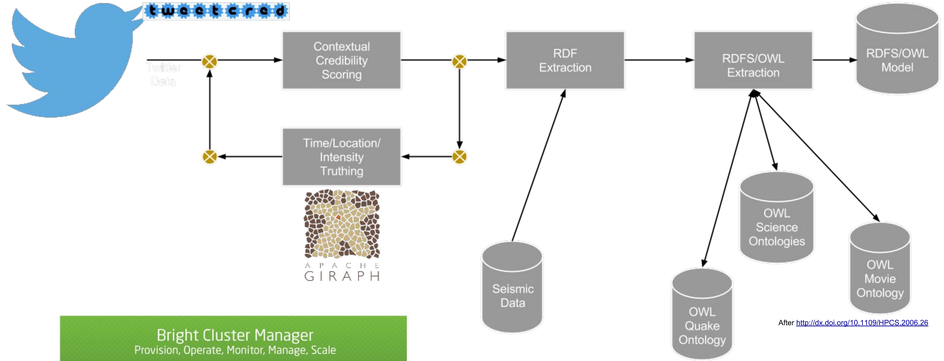
Credible Twitter Data

- Make use of TweetCred
  - `Inherent' credibility
- Introduce contextual credibility
  - Iterative processing of Twitter data via Giraph
    - `Truth' observations extracted from Twitter in time, location and intensity
      - Validate observations at a distance from the source
        - Correlate to the source location
- Incorporate into knowledge-based model
  - Credible Twitter data added to RDFS/OWL model
    - Social networking data complements data from scientific instruments

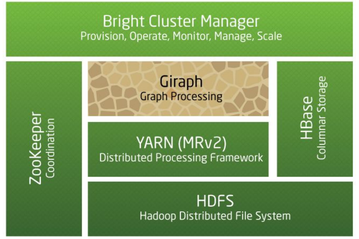
The `Twitter Opportunity'

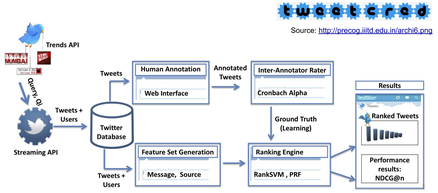
- Complements observations made with scientific instruments
  - Investigate **propagation**, not triangulation
    - Credible Twitter data
      - Is `real' time
      - Is location constrained
        - Dense coverage opportunities via mobility
        - Conveys relative intensity
- Validation through
  - Re-analysis (aka. information forensics)
    - Apply approach to historical data
  - Prediction
    - Apply approach in real time for predictive purposes

After http://classconnection.s3.amazonaws.com/754/flashcards/1160754/png/locating_an_epicenter1432882064652.png



After http://dx.doi.org/10.1109/HIPCS.2006.26





Source: http://precog.iiitd.edu.in/archi6.png

YORK UNIVERSITY

Bright Computing