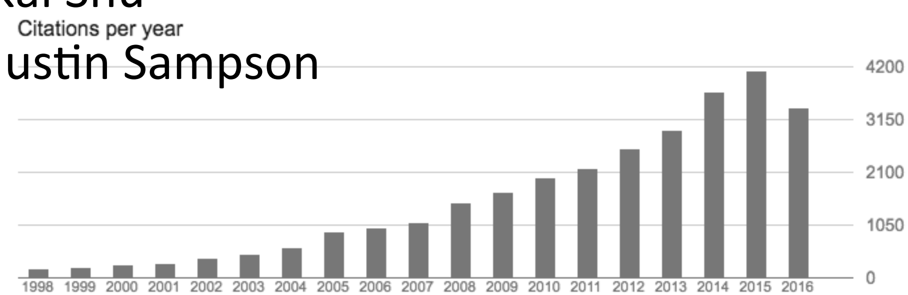
Searching for Credible Information via Social Media Mining

Huan Liu

Data Mining and Machine Learning Lab
Arizona State University

Thanks to Former and Current PhD Students of DMML

- Reza Zafarni, Asst Prof, Syracuse U
- Xia Hu, Asst Prof, Texas A&M U
- Magdiel Galan, Intel
- Shamanth Kumar, Castlight Health
- Pritam Gundecha, IBM Res Almaden
- Jiliang Tang, Asst Prof, MSU
- Huiji Gao, LinkedIn
- Ali Abbasi, Machine Zone
- Salem Alelyani, Asst Prof, King Khalid U
- Xufei Wang, LinkedIn
- Geoffrey Barbier, AFRL
- Lei Tang, Clari
- Zheng Zhao, Google
- Nitin Agarwal, Chair Prof, UALR
- Sai Moturu, PostDoc, MIT Media Lab
- Lei Yu, Assc Prof, Binghamton U, NY
- Robert Trevino, AFRL
- Yunzhong Liu, LeEco, US
- Somnath Shahapurkar, FICO
- Fred Morstatter
- Isaac Jones
- Suhas Ranganath
- Suhang Wang
- Tahora Nazer
- Jundong Li
- Liang Wu
- Ghazaleh Beigi
- Kai Shu
- Justin Sampson



False, Misleading, and Inaccurate Information

- Spam
- Fraud
- Fake News

Disinformation (purposeful)

- Rumor
- Urban Legend
- Gossip

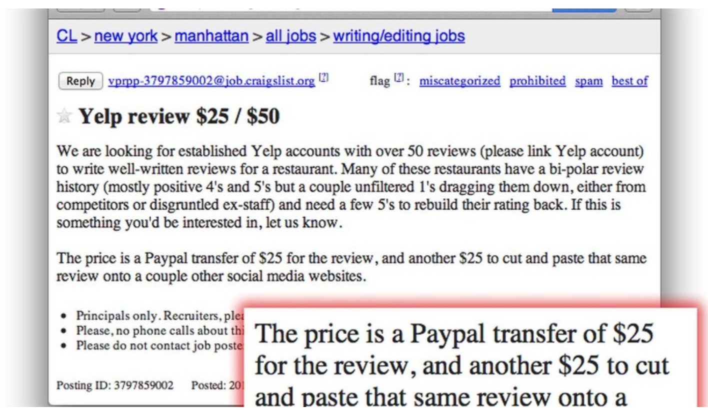
Misinformation (unintentional) & Disinformation

- Information can be: ***true, false,*** *or* ***uncertain***
- Big Data: 6^{th} `V' Everyone Should Know About
  - Vulnerability
  - Social media has all 6 V's

Spam in Social Media

- Unwanted content information generated by spamming users as comments, chat, fake requests that are used to promote products or spread malicious information.

– Fake reviews – Malicious links – Fake requests



Fraud (Scam) in Social Media

- A social media fraud is defrauding and/or taking advantage of social media users with the use of social media services.

– Swindle money

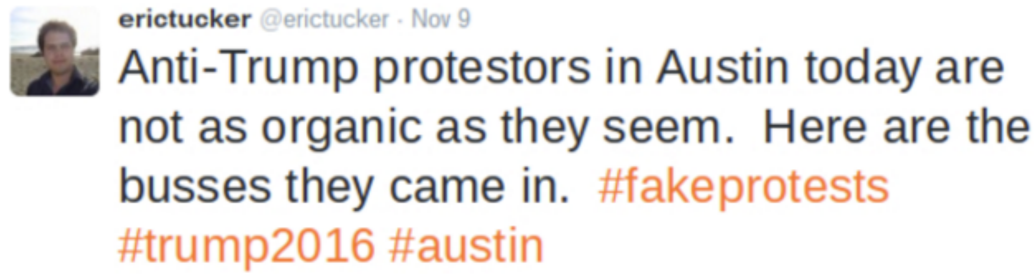
– Steal personal information

Fake News Websites and Social Media

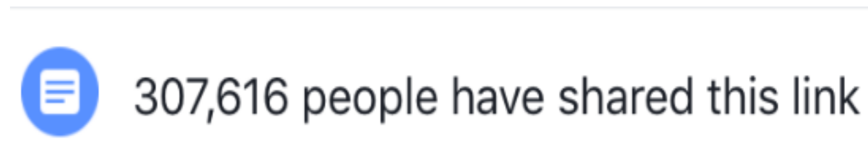
- Fake news websites deliberately publish hoaxes, propaganda, and disinformation to drive traffic *exacerbated by social media*
- Fake news can affect domestic politics, *inflamed by social media*, due to limited resources to check the veracity of claims
  - Easy to "like" and "share", but taking effort to check, albeit just a few clicks away (effort asymmetry)
- Fake news + Social media ➡ Cyberwarfare

Fake News Is Rampant in Social Media

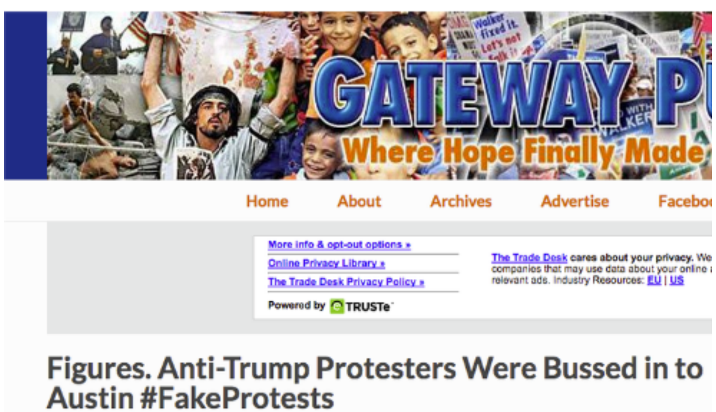
- Fake news spreads on social media
  - Spreads rapidly



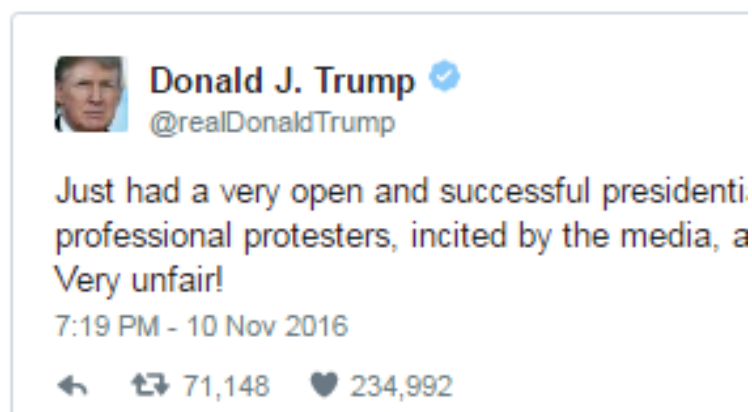
  - Evolves fast



- Crossover to other networks



- Modified content

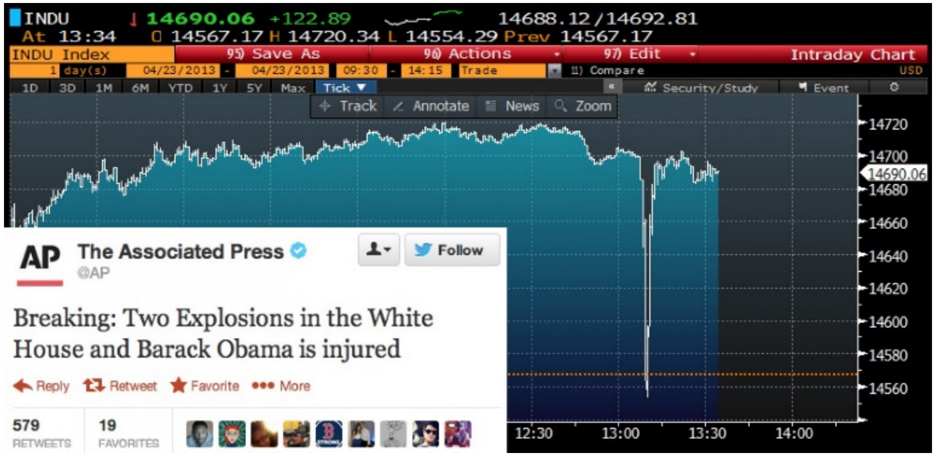


Fake News Can Cause Real Harm

- Pizzagate: stories of fake news from Reddit lead to real shooting

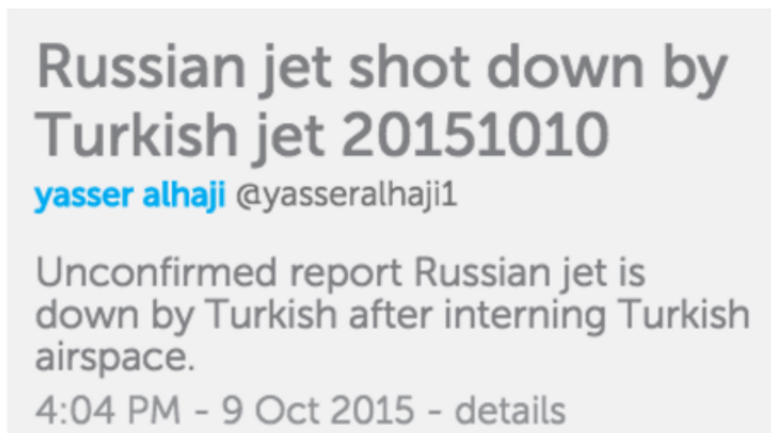
Fake News Onslaught Targets Pizzeria as Nest of Child-Trafficking, New York Times, 2016

- A false rumor erased $136 billion in 10 minutes



Rumors

- Wikipedia: "A tall tale of explanations circulating from person to person and pertaining to an object, event, or issue in public concern".
- Rumors can be ***true*** or ***false***.
  - False rumor



Gossip in Social Media

- Gossip is idle chat and rumor about personal and/or private affairs of others.
- Social media allows for faster, a larger scale of, and more convenient idle chat.

  – Celebrity:

    "Obamas moving to Asheville"

  – Friends:

People "are much more likely to gossip when a story unites a familiar person with an interesting scenario."

Familiarity with Interest Breeds Gossip: Contributions of Emotion, Expectation, and Reputation, PLoS ONE, 2014

Urban Legend in Social Media

- Fictional stories with macabre elements rooted in local popular culture.
  - On social media, it develops faster and spreads wider
    - Urban legend of Fengshui

- *In summary*, it is imperative to study **credibility checking**

On Credibility Checking

- Studying different *types* of *credibility* and the need for different data and information sources in credibility checking
  - We don't have to reinvent wheels in social media mining and can "stand on the shoulder of giants"
  - Machines differ from humans in credibility checking
- About Credibility Checking
  - *Types* of Credibility (social sciences, psychology, CS)
  - *Aspects* of Credibility Checking
  - *Components* of Credibility Checking in Social Media

Four Types of Credibility

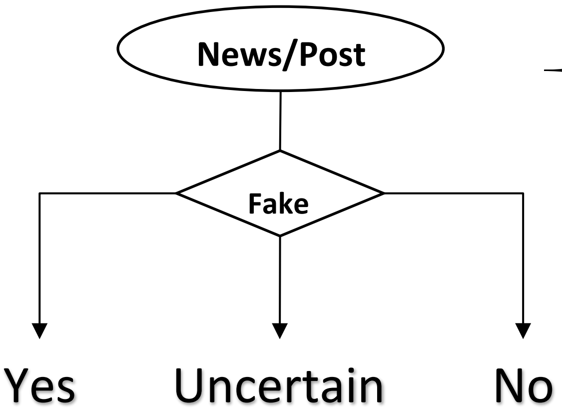
- *Presumed* credibility (general assumptions)
  - “Our friends usually tell truth”
- *Reputed* credibility (based on third parties’ reports)
  - For instance, prestigious awards or official titles
- *Surface* credibility (simple inspection)
  - “People judge a book by its cover”
- *Experienced* credibility (first-hand experience)
  - “Time can tell” （路遥知马力，日久见人心）

Any new type to explore in social media?

Aspects of Credibility Checking (CC)

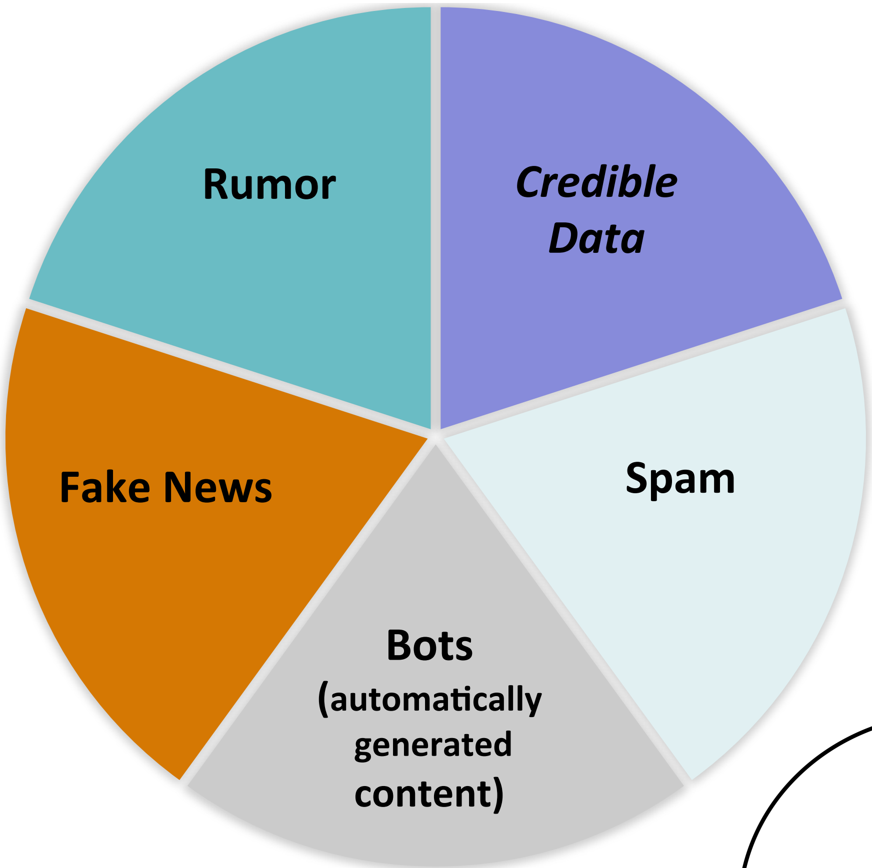
- Can we turn CC into a problem easier for users or AM Turks (without much expertise) to check?
- Issues about Credibility Checking Measures
  - Reputation and History (time)
  - Accuracy and Relevance
  - Transparency and Integrity (consistency)
  - Response from independent sources (consistency)
- Implication or impact assessment
  - Not every piece of fake news is disastrous
  - "Warn or not to warn": how to balance?

Components in Credibility Checking in Social Media



- Recipients
  - Expertise, experience
  - Background, occupation
- Senders
  - Reputation
  - Length of online presence
  - Social networks
- Source of information
  - Provenance
  - Reputation, Curation/Editing
- Content
  - Length
  - Writing style
  - Topics
  - URLs
  - Multimedia
- Network context
  - Topic thread (Outlier detection)
  - Retweets
  - Replies
  - Comments
- Crowdsourcing (fact-checking sites, e.g., Snopes)
- Ground truth (multifaceted, gold standard)

Searching for Credible Information

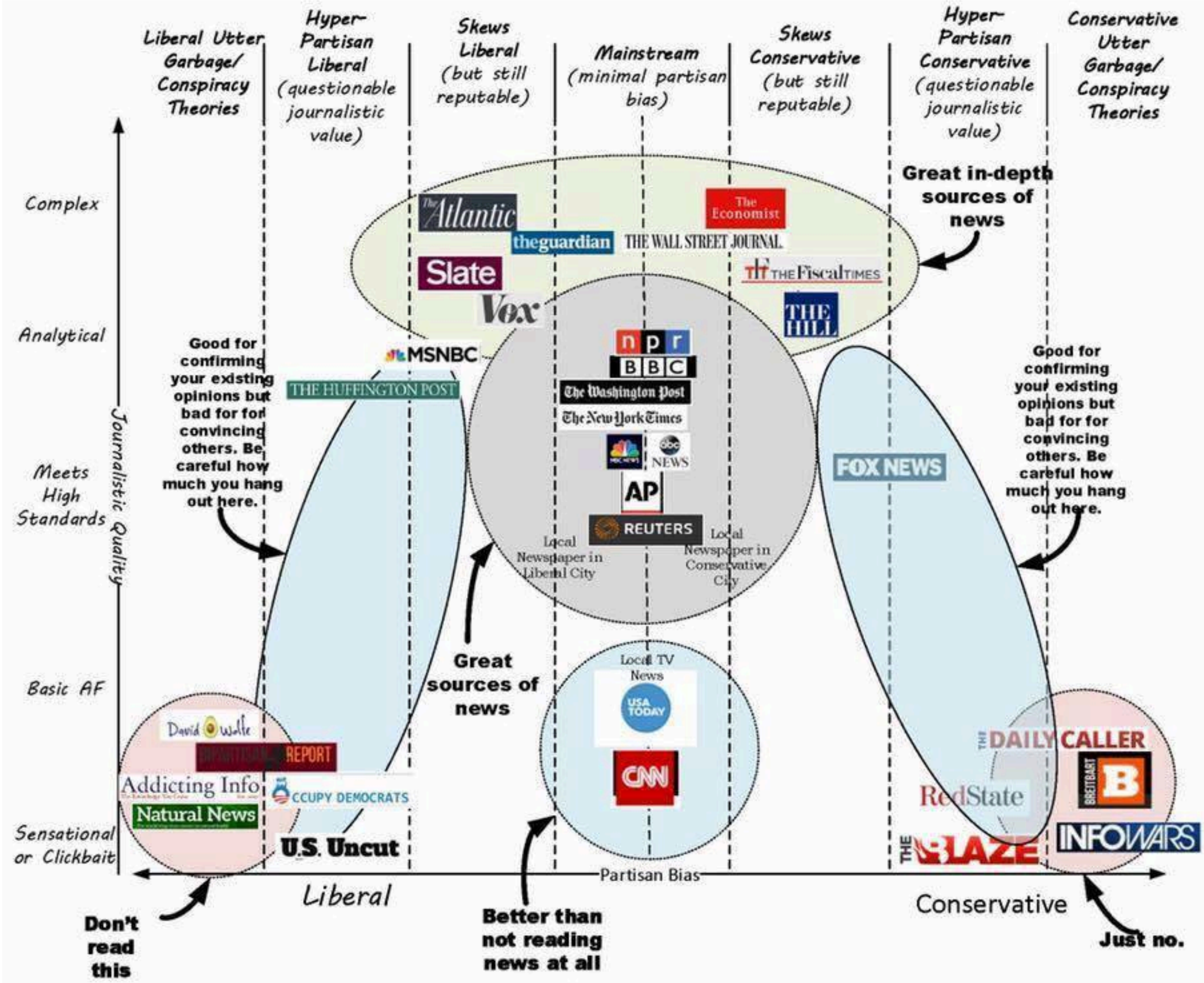


General Elimination Methodology

- A Unique Challenge
  - Ground truth
- Additional Challenges
  - Credibility verification
  - Dynamic change
  - Timeliness
- Alternative Approaches
  - Rumor Detection
  - Spam Detection
  - Bot Detection
  - Inferring Distrust

Using Social Media for Credibility Checking

- Velocity and Volume
  - 6,000 tweets per second, 5 million per day on Twitter
  - 55 million status and 300 million photos per day on FB
- Variety
  - Geo-spatial, textual, pictorial, temporal, social dimensions
  - Cross modality (e.g., geotagged pictures)
- Veracity
  - Truthfulness and accuracy of information
- *Use* big data, multi-source info, and social networks *to compensate for* lack of expertise (以其之矛还其之盾）



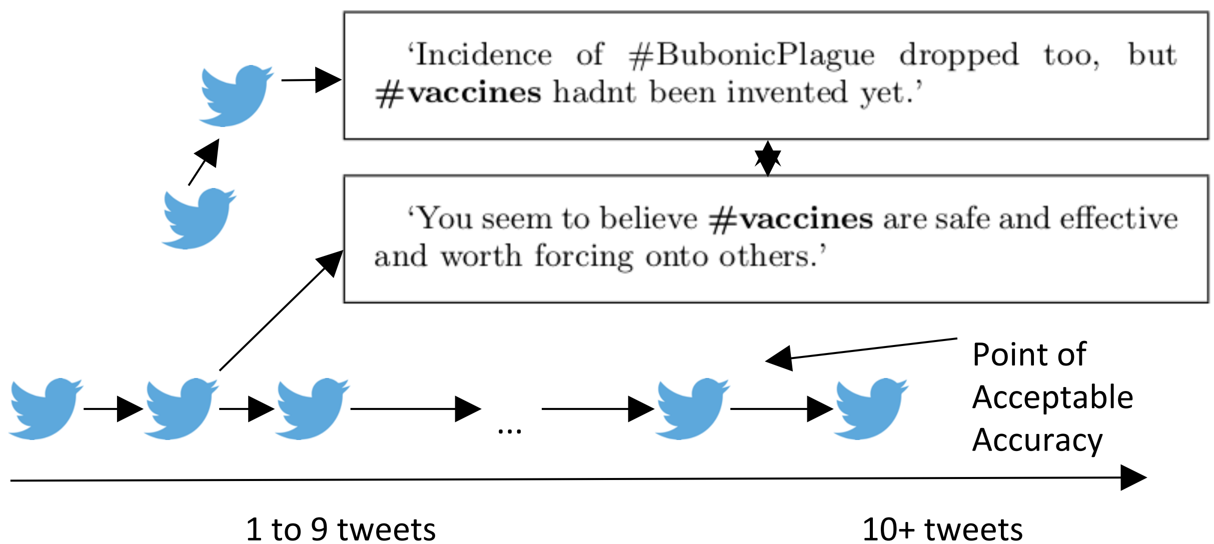
A decent breakdown of all things real and fake news.
http://imgur.com/7xHaUXf

Rumor Detection

- *Rumor*: unverified and relevant information that circulates in the context of ambiguity.
- Goal: detecting emerging rumors with minimum information as early as possible
  - If intervention is not feasible, get early warning or prepared
- Challenges:
  - How to overcome the lack of information in a single tweet?
  - How to detect rumors in their formative stage?

Insufficient Information in a Single Tweet

- A single tweet could be damaging, but contains little information w/o context for detection
- Treat batches of tweets as "conversations"
  - Based on keyword similarities
  - Based on reply chains
- Aggregate conversations
  - Shared hashtags
  - Common links
  - Cosine similarity



Detection of Emerging Rumors

- Emergent detection - link the first tweet in a rumor with those already posted
- Standard rumor classifications are not effective for small conversations
  - Lack of network and statistical data
  - Data sparsity issues
- Implicit linking works effectively for detecting small rumor cascades

Bot Detection

- Bots
  - Innocuous: relay information from official sources
  - Malicious: spread rumors and false information
- Goal: Remove bots from social media data with high Recall
  - WHY?
- Challenges
  - Acquiring ground truth
  - Increasing Recall without significantly reducing Precision

Bots in Social Media

- Bots on Twitter:
  - Twitter claims 5% of 230M users are bots.
  - One study found 20M bot accounts = 9%**.
  - 24% of all tweets are generated by bots***.
- 5-11% of Facebook accounts are fake****.

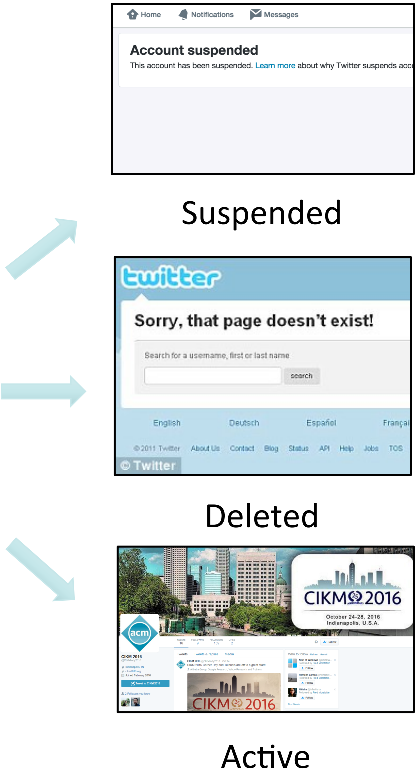
* http://blogs.wsj.com/digits/2014/03/21/new-report-spotlights-twitters-retention-problem/
** http://www.nbcnews.com/technology/1-10-twitter-accounts-fake-say-researchers-2D11655362
*** https://sysomos.com/inside-twitter/most-active-twitter-user-data
**** http://thenextweb.com/facebook/2014/02/03/facebook-estimates-5-5-11-2-accounts-fake/

Finding Ground Truth

Status on Twitter as a labeling mechanism

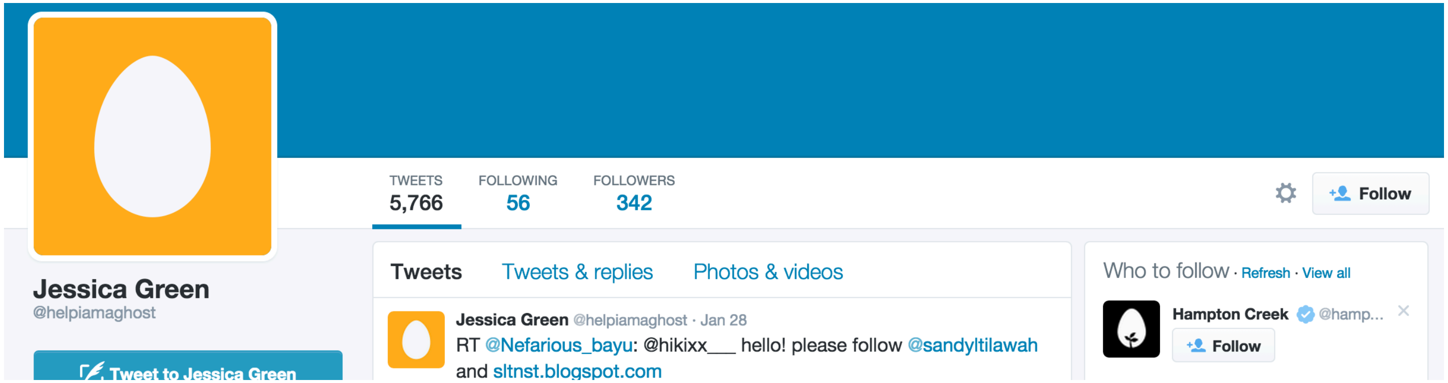
- Three states of a Twitter user:
  - Active
  - Suspended
  - Deleted
- **Idea:**
  - Use these states as labels
  - Two snapshots of each user is taken

Initial Crawl
- Finds seed set of users.
- Crawls Profile, Network, …



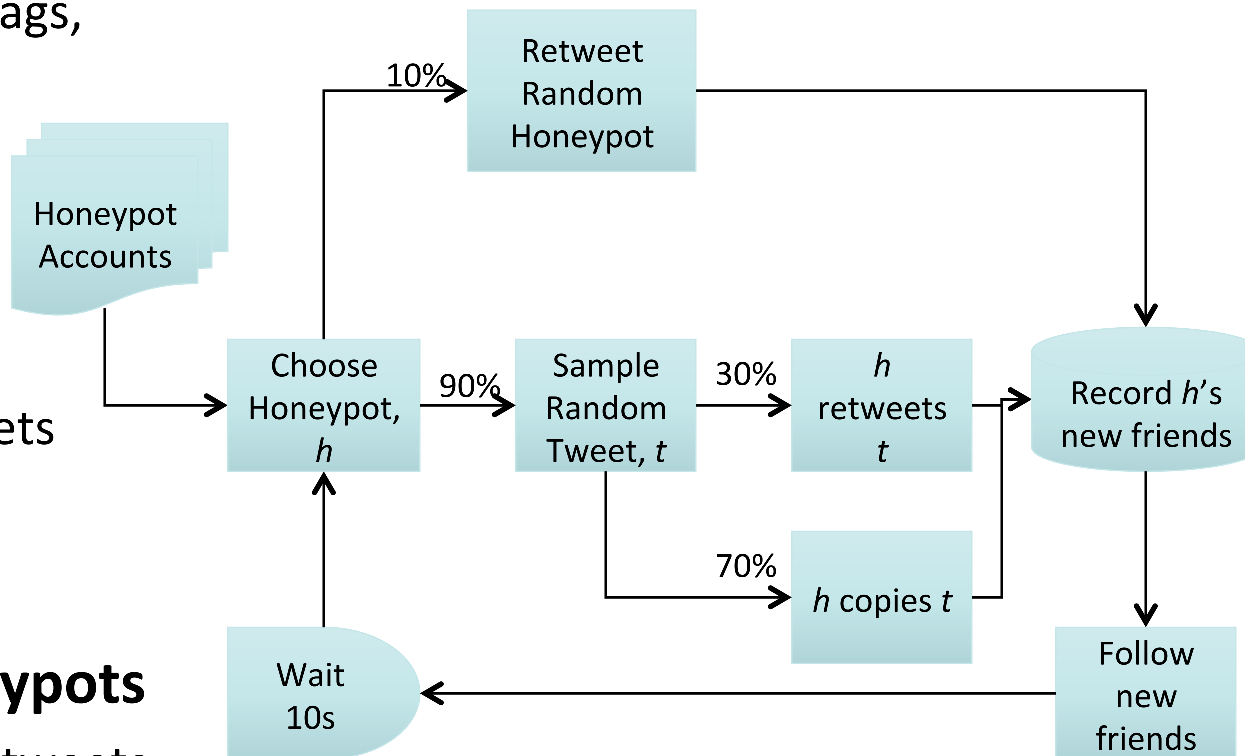
Ground Truth - Honeypots

- Act as obvious bot accounts
- Attract other bot accounts
- Bots are identified when they follow our account
- **Assumption:** Real users do not follow bots



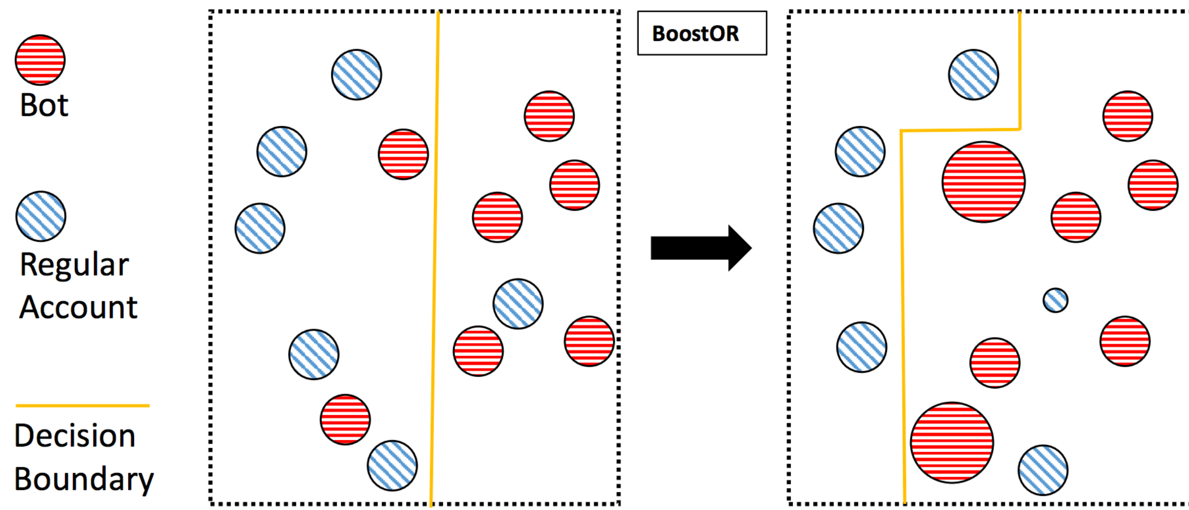
Honeypots - Logic

- **Post "Luring" Content**
  - Post content that will be seen
  - trending topics, hashtags, "famous" tweets...
- **Maintain Network Connections**
  - "Follow back", Retweets
  - Fame begets fame
- **Promote Other Honeypots**
  - Retweet each other's tweets
  - Mention each other



BoostOR

- Based on AdaBoost
- Try to increase Recall without drastic decrease in Precision
- Iteratively update the weight of instances:
  - Unchanged
    if correctly classified
  - Decreased
    if false negative
  - Increased
    if false positive



Trust-Distrust Prediction

- Goal
  - Trust and distrust relations can play an important role in helping online users collect reliable information
  - Finding trustworthy users and reliable information is of significant importance
  - How to predict trust relations between users?

- Challenges
  - Trust relations are extremely sparse
  - Distrust relations are even sparser than trust ones
  - Finding *substitute features* indicative of trust and distrust

Trust and Emotions

- According to psychology, user's emotions can be strong indicators of trust and distrust relations
- Emotional information is more available than that of trust/ distrust
- There exists a correlation between emotions and trust/ distrust relations

Modeling Emotional Information

- Users with positive (negative) emotions are more likely to establish trust (distrust) relations
- Users with high positive (negative) emotion strengths are more likely to establish trust (distrust)
- The Emotional Trust Distrust framework ETD
  - Low-rank matrix factorization
  - Emotional information regularization

Studying Bias in Social Media Data

- Twitter shares its data
  - "Firehose" feed - 100% - costly
  - "Streaming API" feed - 1% - free
- We usually obtain data via sampling
  - Is the sampled data from the Streaming API representative of the true activity on Twitter's Firehose?

- Challenges
  - How to determine if the sample is biased when we do not have access to the whole data?
  - How to obtain an unbiased sample?

Twitter's Streaming API vs. Firehose

- Data from Firehose and Streaming API has been collected for specific period of time to perform analysis
- More than 90% of all geotagged tweets are available via Streaming API and there is not significant difference in location distribution
- Based on in-degree centrality and betweenness centrality in user-user retweet networks, the Streaming API finds ~50% of the key users

Mitigating Bias in Twitter's Streaming API

Can we find bias without the Firehose?

Estimating Bias from Streaming API:

- Obtain trend of hashtag from Sample API and Streaming API
- Bootstrap Sample API to obtain confidence intervals
- Mark regions where Streaming API is outside of confidence intervals

Mitigating Bias:

- Leverage multiple crawlers to maximize data for each query
- Round Robin Splitting

Time-Critical Information in Crisis Response

- Social media is used to request for immediate assistance during crisis
- Time-critical posts demand immediate attention
- Addressing these queries promptly can help in emergency response
- How can these posts be distinguished from others?

- What Is Required in *Finding Time-Critical Responses?*
  - Users with expertise or knowledge
  - Fast response
  - Relevant answers

Finding Time-Critical Responses

- Many questions asked during crisis should be immediately attended
- Many responders are busy
- How can we find a prompt responder who can provide a relevant answer?

- Challenges of Identifying Prompt Responders
  - How do we estimate the *reply time* of users to identify prompt responders?
  - Timeliness and relevance: how do we integrate timeliness with relevance to rank candidate responders?

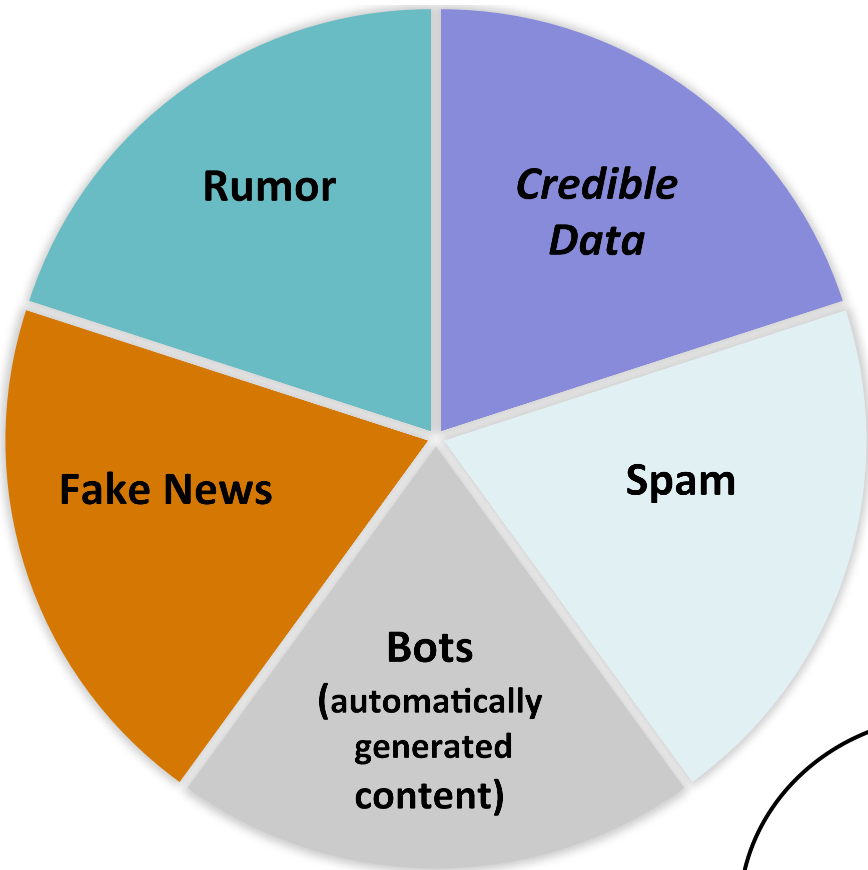
Information Seeking in Social Media

- Social media is used to request for help during crisis
- Addressing these queries promptly can help in emergency response

Identifying Candidate Responders

- Timeliness
  - The user can respond more quickly if she is available soon after the question is posted. It can be estimated using the previous posting times
  - A user responds to questions faster if she has replied promptly to similar questions in the past
- Relevance
  - Users whose previous content is similar to the question have higher relevance and their response is more likely to be a relevant answer
- Timeliness and relevance are integrated by combining the ranking scores

Searching for Credible Information



- A Unique Challenge
  - Ground truth
- Additional Challenges
  - Credibility verification
  - Dynamic change
  - Timeliness
- Alternative Approaches
  - Rumor Detection
  - Spam Detection
  - Bot Detection
  - Inferring Distrust

General Elimination Methodology

以其之矛还其之盾

Thank You All

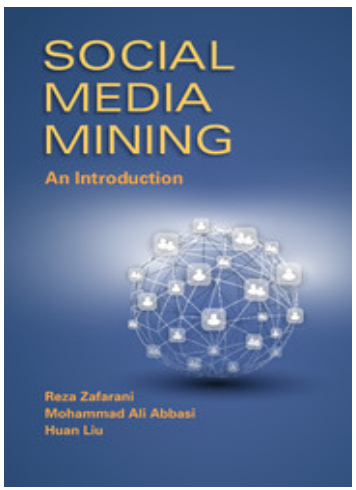
- Professor Yang's kind invitation and warm hospitality
- Funding support from ONR, NSF, ARO, among others
- DMML Lab former and current members, and Liang Wu for helping with the preparation of this presentation

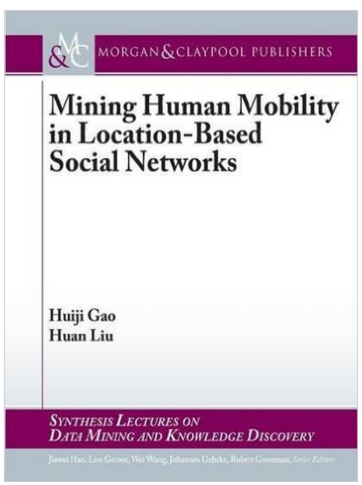
Search for "Huan Liu" for more information about DMML

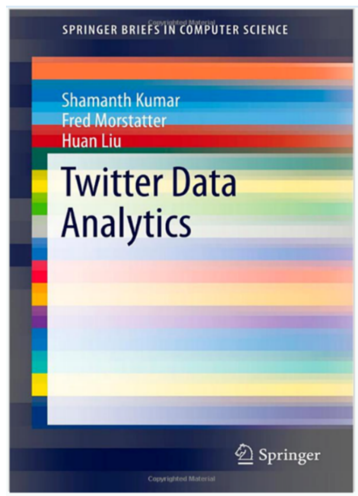
H Liu, F Morstatter, J Tang, and R Zafarani. **``The good, the bad, and the ugly: uncovering novel research opportunities in social media mining",** in Trends of Data Science, International Journal on Data Science and Analytics, Springer International Publishing Switzerland. September, 2016. DOI 10.1007/s41060-016-0023-0

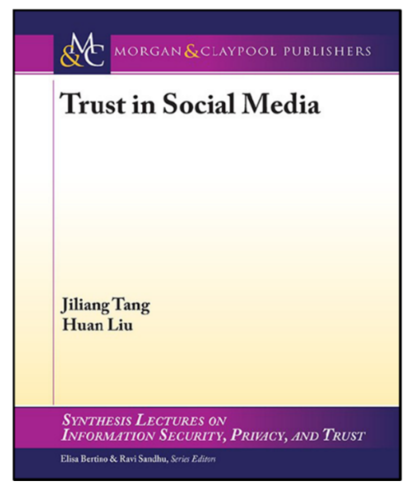
Repositories and Recent Books

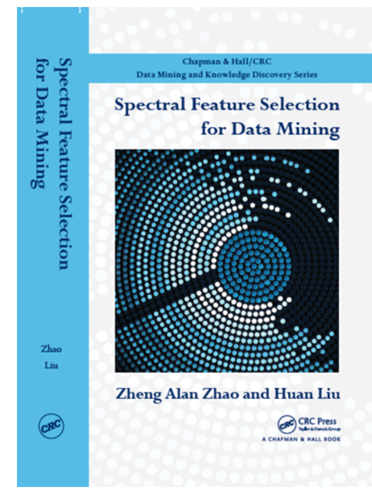
- scikit-feature – an open source feature selection repository in Python
- Social Computing Repository











Social Media Mining

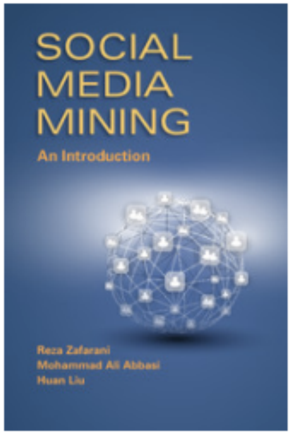
Home | Download Book | Slides/Tutorials | Table of Contents | Errata | How to Cite

Social Media Mining

An Introduction

A Textbook by Cambridge University Press

Reza Zafarani — *Syracuse University*
Mohammad Ali Abbasi — *Machine Zone*
Huan Liu — *Arizona State University*



PDF DOWNLOAD

Accessed 90,000+ times
from 160+ countries and 1200+ Universities

Cambridge University Press | amazon.com | Barnes & Noble Booksellers | eBooks.com | Turing

> *The growth of social media over the last decade has revolutionized the way individuals interact and*

http://dmml.asu.edu/smm/

References

1. [Beigi SDM'16] Ghazaleh Beigi, Jiliang Tang, Suhang Wang, and Huan Liu. "Exploiting Emotional Information for Trust/Distrust Prediction". SIAM International Conference on Data Mining (SDM16), May 5-7, 2016. Miami, Florida.
2. [Morstatter ASONAM'16] Fred Morstatter, Liang Wu, Tahora H. Nazer, Kathleen M. Carley, and Huan Liu. "A New Approach to Bot Detection: Striking the Balance Between Precision and Recall", IEEE/ACM International Conference on Advances in Social Network Analysis and Mining (ASONAM2016), August 18-21, San Francisco, CA.
3. [Morstatter WWW'14] Fred Morstatter, Jürgen Pfeffer, Huan Liu. When is it Biased? Assessing the Representativeness of Twitter's Streaming API", WWW Web Science 2014.
4. [Morstatter ICWSM'13] Fred Morstatter, Jürgen Pfeffer, Huan Liu, Kathleen M Carley. Is the Sample Good Enough? Comparing Data from Twitter's Streaming API with Twitter's Firehose", ICWSM 2013.
5. [Sampson CIKM'16] Justin Sampson, Fred Morstatter, Liang Wu and Huan Liu. "Leveraging the Implicit Structure within Social Media for Emergent Rumor Detection", short paper, ACM International Conference of Information and Knowledge Management (CIKM2016), October 24-28, 2016. Indianapolis, Indiana.
6. [Sampson ICDM'15] Justin Sampson, Fred Morstatter, Reza Zafarani, and Huan Liu. "Real-Time Crisis Mapping Using Language Distribution". Demo. In Proceedings of IEEE International Conference on Data Mining (ICDM2015), November 14 - 17, 2015. Atlantic City, NJ.