Social Media Tools for Political & Development Analysis
A Systematic Literature Review, 2007-2016

Gabrielle Cheung
glcheung@usc.edu
Presentation Structure

1. Background
   - Motivations
   - Method

2. Elections
   - Key Results & Findings
   - Discussion

3. Political Mobilization & Regime Transition | Disaster Response & Management | Disease Surveillance
   - Key Results & Findings
   - Discussion

4. Significance of This Review for UNU-CS
   - (Policy) Implications
1. Background: Motivations

(i) Review of Methodological Trends

- Remedy gap in the existing literature, as most review essays tend to focus on substantive trends
- Identify prospects and pitfalls of extant projects operationalized in similar mission areas

(ii) Integration of News Treatments

- Endeavor to discern developments/innovations that remain un(der)reported in extant studies
- Gauge general reception of local & international media outlets toward social media-derived solutions to real-life political/development problems

(iii) Facilitation of Robustness Checks

- Compare & contrast debates, methodological inclinations, and findings across thematic areas
1. Background: Review Method

(i) Selection of Databases, Repositories, and Search Engines
   • Specialist databases (ACM Digital Library, IEEE Xplore Digital Library)
   • Generalist repositories & search engines (JSTOR, Lexis HK, ScienceDirect, Google Scholar & Google Scholar Citations)

(ii) Selection of Sources
   • Selected ICT4D journals (e.g., AJC, EJISDC, ITD)*
   • Selected secondary journals (e.g., East European Politics, Political Analysis)
   • Selected conference proceedings (e.g., P-ISCRAM, P-SIGCHI, P-SWID)#

(iii) Selection of Search Terms
   • Boolean operators & modifiers by thematic area

(iv) Inclusion/Exclusion Criteria

(v) Data Extraction
   • In accordance with the PRISMA Statement (see Moher et al. 2009)

* Asian Journal of Communication, Electronic Journal of Information Systems in Developing Countries, Information Technology for Development
Overview of the Four Subsamples

**Elections** \( (n = 56) \)

**Political Mobilization & Regime Transition** \( (n = 45) \)

**Disaster Response & Management** \( (n = 68) \)

**Disease Surveillance** \( (n = 38) \)

*Note: “F/c” refers to forthcoming publications.*
2. Elections: Results

(i) Three Key Debates

• Utility of Social Media Data for Election Prediction
  − Do online opinions mirror offline political sentiment? (e.g., Tumasjan et al. 2010)
  − How to compute who/which party will win? (e.g., Mahmood et al. 2013)
  − How to improve prediction techniques? (e.g., Kagan et al. 2015)

• Political Implications of Soc. Med. Usage During the Election Cycle
  − How does usage affect perceptions of fairness & impartiality? (e.g., Bailard 2012)
  − Does usage affect chances of electoral success? (e.g., Bühler and Bick 2013)
  − Do soc.med. platforms promote deliberative democracy? (e.g., Best and Meng 2015)

• Significance of Soc. Med.-based Election Monitoring
  − What explains the rise in citizen monitors? (e.g., Moyo 2010)
  − How efficacious are crowdsourced efforts in detecting election fraud? (e.g., Bader 2013)
  − Can monitoring build trust? (Smyth and Best 2013)
(ii) Social Media Platforms

- One platform: 85.7% (n = 48)
- Two to seven platforms: 14.3% (n = 8)
- Twitter: 76.8% (n = 43)
- Others: Facebook (14.3%, n = 8), YouTube (7.1%, n = 4), MySpace (n = 2), blogs (n = 3), Aggie (n = 2), discussion forums (n = 1), Flickr (n = 1), Karta Narusheniy (n = 1), LinkedIn (n = 1), Ushahidi (n = 1).
(iii) Geographical Foci

- Single-country case study: 85.7% \((n = 48)\)
- Multiple-country (2-6) case studies: 14.3% \((n = 8)\)
- Countries: United States \((n = 13)\), Nigeria \((n = 6)\), Germany \((n = 5)\), the United Kingdom \((n = 5)\), the Netherlands \((n = 4)\), Pakistan \((n = 4)\), Indonesia \((n = 3)\), Italy \((n = 3)\), Spain \((n = 3)\), Canada \((n = 2)\), France \((n = 2)\), India \((n = 2)\), Kenya \((n = 2)\), Australia \((n = 1)\), Azerbaijan \((n = 1)\), Ghana \((n = 1)\), Liberia \((n = 1)\), Mexico \((n = 1)\), Palestine \((n = 1)\), Russia \((n = 1)\), Sierra Leone \((n = 1)\), Singapore \((n = 1)\), Tanzania \((n = 1)\), Turkey \((n = 1)\), Zimbabwe \((n = 1)\)

Understudied Regions -
- Elections: Oceania (e.g., NZ, PIs), Nordic countries (e.g., Finland, Iceland, Sweden), South America
- Election Monitoring: World ex. Africa and Russia

(iv) Event Foci

- Presidential elections: 32.1% \((n = 18)\)
- General/national elections: 21.4% \((n = 12)\)
- Remainder: Parliamentary elections, regional (EU-wide) elections, local elections, debates
2. Elections: Results

(v) Actors under Study

• All users of a given social media platform: 33.9% (n = 19)
• Only political elite: 23.2% (n = 13)
• Only geographically relevant users of a platform: 5.4% (n = 3)
• Remainder: Selected elite & mass users; selected language users; selected users with known voting intentions; students; most influential users

(vi) Data Collection & Analysis Methods

• Twitter: Mostly APIs & additional software/systems like Aggie (Best and Meng 2015), Twimemachin (Mahmood et al. 2013), twitteR (Khatua et al. 2015), Twitter crawlers using Perl (Skoric et al. 2012), MySQL databases (Skoric et al. 2012; Song et al. 2014), and tagging systems to store data (Song et al. 2014)

• Blogs/discussion forums: Manual extraction; Nigerian Blog Aggregator (Ifukor 2010)

• Other methods: Online survey (Bühler and Bick 2013); field experiment (Bailard 2012); semi-structured interviews, contextual observations, focus groups (Smyth and Best 2013; Lazarus and Saraf 2015)
(vi) Data Collection & Analysis Methods (Cont’d)

• Automated sentiment analysis (or sentiment scoring) to measure opinion polarity and intensity (e.g., Wegrzyn-Wolska and Bougueroua 2012, Fink et al. 2013, Nooralahzadeh et al. 2013, Razzaq et al. 2014, Ceron et al. 2015, etc.)

• Qualitative content analysis (e.g., Robertson 2011; Ahmed and Skoric 2014)

• Network analysis (GEPHI) (e.g., Mascaro and Goggins 2015)

• Text analysis (incl. methods like multinomial topic modeling; term-co-occurrence retrieval; and software like Luminoso) (e.g., Song et al. 2014; Best and Meng 2015)
(vii) Novel Methodological Innovations

- **TaraTweet** (Soler et al. 2012: 1195)
  - “Web application developed in collaboration between social researchers and computer scientists of the University of Castilla-La Mancha [that] allows the monitoring of social conversations in Twitter through some hashtags defined by the user” and “counts keywords which users have introduced in the creation of a specific experiment defined before [...]”

- **Karta Narusheniya** (aka “Map of Violations”) (Bader 2013)
  - Ushahidi-inspired tool that tracks the spatial distribution of electoral fraud while also making use of social media platforms like YouTube
  - Reportedly engaged “thousands of individuals” (*ibid.*, 521) who contributed to a “database that contains over 13,000 reports” (*ibid.*) during the 2011-2012 election cycle in Russia
  - Flagged up two main types of electoral malpractice:
    (a) Voting fraud: “ballot-stuffing,” “organised group voting with breaches of the secrecy of the vote,” “multiple voting,” and “vote-buying” (*ibid.*, 526);
    (b) Counting fraud: “intentional miscounting of votes,” “protocol tampering,” and “divergence between protocol and official final results” (*ibid.*)
2. Elections: Results

Official Website of Karta Narusheniy

Sample Report with YT Video as Evidence

Source: http://www.kartanarusheniy.org
2. Elections: Results

Compilation of Descriptive Statistics from the March 4, 2012 Election
(Google-Translated Version)

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Violations of the voting day</th>
<th>Violations during the campaign</th>
</tr>
</thead>
<tbody>
<tr>
<td>level elections</td>
<td>other violations</td>
<td>other violations</td>
</tr>
<tr>
<td>federal</td>
<td>gubernatorial</td>
<td>165</td>
</tr>
<tr>
<td>non-federal</td>
<td>1669</td>
<td>165</td>
</tr>
<tr>
<td>total</td>
<td>1669</td>
<td>165</td>
</tr>
<tr>
<td>complaints</td>
<td>Violation of the rights of observers, representatives of the media</td>
<td>Violation of the rights of candidates</td>
</tr>
<tr>
<td>complaint lodged</td>
<td>violations during voting by absentee, &quot;at home&quot; illegal voting</td>
<td>121</td>
</tr>
<tr>
<td>111</td>
<td>354</td>
<td>121</td>
</tr>
<tr>
<td>Administrative resources</td>
<td>479</td>
<td>121</td>
</tr>
<tr>
<td>728</td>
<td>675</td>
<td>121</td>
</tr>
<tr>
<td>111</td>
<td>462</td>
<td>121</td>
</tr>
<tr>
<td>661</td>
<td>284</td>
<td>121</td>
</tr>
<tr>
<td>266</td>
<td>123</td>
<td>121</td>
</tr>
</tbody>
</table>

Source: http://www.kartanarusheniya.org
(vii) Novel Methodological Innovations

- **Aggie 2.0** (Lazarus and Saraf 2015)
  - Deployed in Liberia, Ghana, Kenya for this study
  - Used to “track, aggregate, analyze and respond to user-generated content available over social media during elections” (ibid., 1).
  - Has “better usability” and is able to integrate “formal reporting with social media aggregation” from “trained formal field observers using the ELMO observer platform” (ibid., 2)
  - Still faced “challenges in validating and testing the truthfulness of a few (not all) reports sourced from social media reports” (ibid., 1)
2. Elections: Discussion

Five Key Findings

(i) **Dominance** of Twitter-centric Studies

- 76.8% of studies used Twitter data (vs. 14.3% that used Facebook data)
- Extant explanations: free-to-use data (in most cases); wider range of data collection methods; real-time analysis rendered possible

Cf. FB being the world’s largest social media company & platform, with 5X monthly active users – 1.65 billion vs. 0.31 billion (see Facebook 2016; Twitter 2016)

(ii) Recurrent “Failure” to Recognize User-Voter **Non-interchangeability**

- Frequent systematic conflation between “users” and “voters”
- Limited acknowledgment of the existence of spam bots, paid users, fanatic non-voters
- Lack of discussion on possibility of large-scale “red team attacks” (i.e., professional online opinion manipulators) (see Lazer et al. 2014)
2. Elections: Discussion

(iii) Limited Development of Tools to Overcome Geopolitical Constraints

- Domestic resistance against “security threats” and resultant social media blackouts/shutdowns (e.g., Uganda’s and Chad’s 2016 elections)
- Partisan-driven incentives to misinform the public and/or monitory actors
- Emergence of privacy infringement & political prosecution concerns (“secondary map of informants”)

(iv) Perils of Observational Research

- Inability to ascertain robustness of causal claims
  (Very few experimental studies in subsample; an exception was Bailard’s (2012) field experiment)
- Difficulty to measure impact of social desirability bias on results

(v) Relative Paucity of Normative Discussions

- Desirability of Election Monitoring vs. the Problem(s) of “Autocratic Adaptation” (Sjoberg 2014)
- “The Monitor’s Dilemma”: Who should have the power to monitor elections, and decide which elections get monitored? Is a monitory actor legitimate if it is not regulated/subject to scrutiny?
2. Elections: News Treatments

Monthly Frequency of English News Articles That Mentioned “Social Media” in Their Headline or Lead Paragraph (January 2007 – July 2016)

Source: Lexis HK
2. Elections: News Treatments

2. Elections: News Treatments

News Treatments of ELMO

Sources: OpenDataKit (2013) Link; Guyanese Online (2015) Link
2. Elections: News Treatments

Additional Treatments of ELMO

2. Elections: News Treatments

News Treatments of Aggie

1st Annual African Elections Project Lectures held

Glory by Mavis Obor.

This lecture organised by PanPilgrims is the first in its series that forms part of the African Elections Project interventions to use ICTs and face-to-face dialogues to ensure that practices are adopted to promote free and credible elections across the continent. The African Elections Project established in 2009 by PanPilgrims has organised elections in 13 African countries with a vision of enhancing the ability of journalists, citizens and the news media to provide more timely and relevant election-related information and knowledge while undertaking monitoring of specific and important aspects of governance.

2. Elections: News Treatments

News Treatments of *Karta Narusheniy*

**The Moscow Times**

*Russian Election Monitoring NGO Golos Removed from Foreign Agents List*

By Anna Dolgov | Jul. 23 2015 10:59 | Last edited 10:59

**The Guardian**

*Russian election monitors complain of state harassment*

Golos watchdog found guilty of breaching Russian electoral law in what critics describe as a stepped-up campaign of intimidation

Members of Russia’s only independent election monitoring group have told the Guardian they are being harassed and intimidated by state security officials in the run-up to Sunday’s parliamentary elections.

**Sources:** The Moscow Times (2015) [Link]; The Guardian (2011) [Link]
2. Elections: News Treatments

Non-Exhaustive List of **Additional Projects** Mentioned in the News

(i) The **Electome** Project (U.S., in progress)
   • MIT Lab for Social Machines, Knight Foundation
   • Holistic election-related real-time mapping

(ii) The **Kyeet** Project (Myanmar, 2015)
   • Myanmar ICT for Development Organization, Center for Civic Tech
   • Election-monitoring app
   • Profiled in *Frontier Myanmar, Myanmar Times, Nikkei Asian Review, Yahoo! News/Foreign Policy*, etc.

(iii) The **Every Vote Count** Project (Nigeria, 2015)
   • Mobile Xcetera
   • Election-monitoring app
   • Profiled in *The Guardian, Love World Plus (Nigeria)*

(iv) The **PakVotes** Project (Pakistan, 2013)
   • Bytes for All, U.S. Institute of Peace
   • Election-monitoring app
   • Profiled in *Express Tribune (Pakistan), The World Post* (under Huffington Post)
2. Elections: News Treatments

(i) The Electome

Projects

The Electome: Fueling the Horse Race of Ideas in the 2016 Election

Soroush Vosoughi, Prashanth Vijayaraghavan, Margaret Hughes, William Powers, Andrew Heyward, Russell Stevens and Deb Roy

Project Alumni: Sophie Chou, Ping-Hwa Kung, Eric Chu, Neo (Mostafa) Mohsenvand, Raphael Schoad

The “Electome” is a comprehensive mapping of the content and network connections among the campaign’s three core public sphere voices: the Candidates, the Media, and the Public. This mapping is used to trace the election’s narratives as they form, spread, morph and decline among these three groups – identifying who and what influences these dynamics. We are developing metrics that show how responsive the Candidates and Media are to the issues that the Public find most relevant and important.

LSM’s analytics will serve as the basis of news coverage that LSM will produce in collaboration with major media partners. The ultimate aim is to offer an alternative to the “horse race journalism” that has dominated election news for the last half-century. The Electome will instead focus on surfacing the important issues at stake in the campaign, or the “The Horse Race of Ideas.”

The John S. and James L. Knight Foundation is providing support for the Electome project as part its efforts to advance excellence in journalism and increase civic engagement. To date, LSM’s “Horse Race of Ideas” deployments include:

2. Elections: News Treatments

(ii) Kyeet

2. Elections: News Treatments

(ii) Kyeet

Turning Toward Democracy, Burma Casts Votes In Historic Election

This is precisely what voters now hope to change — and they’re willing to overlook what might have counted as NLD shortcomings under normal conditions. Maung Maung Hla, 39, a driver, proudly his ink-stained pinky Sunday, evidence that he’d cast his vote. He had to wait in line for almost two hours early this morning before he finally got his ballot. “I voted for the candidate who can lead the country,” he said, referring to Suu Kyi. “I voted for change.” Hla is a Muslim, but the NLD’s refusal to include a single Muslim on its candidate list — an apparent reaction to pressure from Buddhist ultranationalists — doesn’t seem to have deterred him. “That doesn’t bother me at all.” It’s not entirely clear that other members of the country’s sizable Muslim minority share that view.

The military has shaped the country’s political structure in such a way that it will still retain enormous influence even if the NLD scores a landslide win. Yet Burmese citizens and civil society organizations are doing their best to make sure that voters will have their say. As the polls closed, a group of young Burmese in identical dark blue T-shirts huddled around computers in an 11th-story office in downtown Rangoon. Members of a group called Kyeet.org (“kyeet” is the Burmese word for “watch”), they were collating data from 2,000 citizen election monitors all over the country who were using a special smartphone app to report election irregularities.

International election observers have cited numerous deficiencies in the voting so far. In the most egregious case, some 4 million overseas Burmese were excluded from the polls due to onerous registration rules. For all its obvious imperfections, though, this election will still offer Burmese citizens their first chance for real political participation in a generation. So far they seem keen to seize the opportunity.

Lam Yik Fei/Getty Images

2. Elections: News Treatments

(iii) Every Vote Count

2. Elections: News Treatments

(iv) PakVotes

Citizen journalists to monitor polls

By Our Correspondent Published: April 23, 2013

In some ways, it was like any other election in Pakistan—violence, corruption, and allegations of rigging. In other ways, it was like nothing this 60-year-old nation had ever seen.

May 11, 2013 marked the first constitutional transfer of power from one civilian government to another in Pakistani history. Fittingly, this big first was captured with another big first—a globally trending hashtag. As the Saturday morning sun rose over sleepy Islamabad, #PakVotes was already surging to the top of Twitter’s trending list all over the world.

Championed by the Pakistani human rights organization “Brave for All,” Pak Votes was a digital campaign seeking to bring “openness” and “inclusion” into the Pakistani political process. On Election Day, Pak Votes replaced pinpoints and predictions with faces—showcasing everyone from rural women voting for the first time in villages to grizzled elders rolling to their local polling stations in wheelchairs. They showed people of various parties, ethnicities, and locations.

2. Elections: Discussion of News Treatments

Summary

(i) Largely **Positive** Portrayal of Election-Monitoring Activities
   - Social media- and app-driven monitoring posited as “solutions” to electoral fraud and other types of irregularities
   - “Novel force for social good”: improves quality of democratic practices, builds trust in society, educates/informs the public
   - Reports identify accusations of “foreign interference” made by government actors (e.g., the incumbent party), but rarely support this narrative

(ii) Relative Sensitivity Toward the **Status** of Monitors
   - Reports do tend to differentiate between professionally trained foreign and domestic monitors, citizen journalists, and lay people
   - BUT they rarely compare and/or pass judgment on whether a certain type of monitor is operationally superior to another, etc.
<table>
<thead>
<tr>
<th>Other Thematic Areas:</th>
<th>A. Political Mobilization &amp; Regime Transition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B. Disaster Management &amp; Response</td>
</tr>
<tr>
<td></td>
<td>C. Disease Surveillance</td>
</tr>
</tbody>
</table>
3A. Political Mobilization & Regime Transition

(i) Five Key Debates

• The Role(s) of Soc. Med. in Large-scale Political Mobilizations
  − How has soc. med. usage affected the formation of collective identities? (e.g., Bennett and Segerberg 2012)
  − Why can soc. med. help overcome the collective action problem, even in the absence of recognized leaders and common goals? (e.g., Bennett et al. 2014; Treré 2015)

• Processes Linking Soc. Med. Usage with Activism and Governance Issues
  − Does soc. med. usage shape how ordinary people learn about protest movements, and plan their involvement/non-involvement? (e.g., Tufekci and Wilson 2012)
  − How do protesters use platforms like Twitter during periods of political upheaval? Does soc. med. usage actually increase the likelihood of on- and/or offline political engagement? (e.g., Valenzuela et al. 2012; Earl et al. 2013; Vissers and Stolle 2014)

• Social Media as the Main Cause of Real-life Protests and Revolutions
  − Did soc. med. cause the Arab Spring? If so, how? If not, what kind of mediating variable did it produce? (Howard et al. 2011; Danju et al. 2013; Alaimo 2015)
(i) Five Key Debates (Cont’d)

• The Rise of “Networked Authoritarianism”
  – Are authoritarian regimes necessarily vulnerable to the mobilizing potential of soc.med.? (E.g., Pearce and Kendzior 2012)
  – Can soc. med. actually serve to empower such regimes? (E.g., Youmans and York 2012)

• Transnational Political Mobilization
  – Can soc. med. serve to promote transnational political activism, as evidenced by the Occupy Movement, the Sunflower Movement, or the Shahbag Movement? (E.g., Pearce and Kendzior 2012; Theocharis et al. 2015; Chen et al. 2015; Raychoudhury et al. 2015)
(iii) Data Collection & Analysis Methods

**Data Collection Methods**

- **Twitter:** APIs; Google Analytics Real-time Scraper; Topsy API; TwapperKeeper; MentionMapp (e.g., Howard et al. 2011; Poell 2014; Raychoudhury et al. 2015)

- **Surveys:** Electronic instrument via soc. med. platforms with a link to Qualtrics (e.g., Zouniga et al. 2014; Chen et al. 2015); direct recruitment during protests (Tufekci and Wilson 2012)

- **Professional polling firm:** Probability sampling (Valenzuela et al. 2012)
(iii) Data Collection & Analysis Methods (Cont’d)

Data Analysis Methods

- **Content Analysis** (e.g., Howard et al. 2011; Pearce and Kendzior 2012; Starbird and Palen 2012; Vallina-Rodriguez et al. 2012)

- **Descriptive/Inferential Statistical Analysis** (e.g., Tufekci and Wilson 2012; Zouniga et al. 2014; Raychoudhury et al. 2015)

- **Interpretative Case Studies** (e.g., Lim 2012; Molaei 2015)

- **Social Network Analysis** (e.g., Theocharis 2013; Huang and Sun 2014; Poell 2014; Nouh and Nurse 2015)

- **Discourse Analysis** (e.g., Liu 2015)

However, methodological innovations were *not* reported by studies in this subsample.
Three Key Debates

• Information-sharing Behavior & Propagation Trends
  – Why do people, esp. non-stakeholders, share disaster info? (e.g., Chen and Sakamoto 2014)
  – How do disaster-implicated victims engage with soc. med.? (e.g., Aisha et al. 2015)
  – Any differences in info-sharing behavior before, during, and after crisis events? (e.g., Kaewikitipong et al. 2016)
  – Why do people partake in “derivative info propagation” despite risk of spreading mere rumors or even misinformation? (Starbird and Palen 2010)

  – Could there be greater coordination among soc. med. platforms? (e.g., White et al. 2009)
  – How to refine knowledge management? (e.g., Yates and Paquette 2011)
  – How to create “disaster soc. med. tools”? (e.g., Houston et al. 2015)

• Maximizing Utility of Individual Platforms in Times of Crisis
  – Can Twitter act as an emergency response tool? (e.g., Mills et al. 2009)
  – How can actionable data be extracted more efficaciously? (e.g., Ashktorab et al. 2014)
  – How can government actors or other professional bodies make use of soc. med.? (e.g., Chatfield and Brajawidagda 2013; Plotnick et al. 2015)
(ii) Actors under Study

- Ordinary users of a given social media platform: 55.9% \((n = 38)\)
- Direct victims: 4.4% \((n = 3)\)
- Others: Government agencies (e.g., Chatfield and Brajawidagda 2013)
  - National weather service (e.g., Chatfield and Brajawidagda 2014)
  - Public officials (e.g., Sutton et al. 2014)
  - Emergency management practitioners (e.g., Calderon et al. 2014)
  - County-level emergency managers (e.g., Plotnick et al. 2015)

(iii) Data Collection & Analysis Methods

- Twitter: Mostly APIs, in addition to software/systems like Topsy (Chatfield and Brajawidagda 2012; Olteanu et al. 2014; Teodorescu 2015) and Tweepy (Chatfield and Brajawidagda 2013)
- HEROIC – a data collection system jointly developed by the University of Colorado-Colorado Springs and UC Irvine, and funded by the National Science Foundation (Thomson et al. 2012; Thomson and Ito 2012)
- NVivo (Takahashi et al. 2015)
- TweetArchivist (Spence et al. 2015; Lachlan et al. 2016)
- One study (Hashimoto et al. 2015) commissioned HottoLink, Inc. to collect Twitter data
(iii) Data Collection & Analysis Methods (Cont’d)

Other Collection Methods:
• Surveys disseminated via social media platforms (e.g., Facebook), MTurk, SurveyGizmo, QuestionPro
• Paper instruments
• Interviews involving disaster victims & experts from HADR organizations, civil society groups, government agencies, telecom companies

Data Analysis Methods:
• Descriptive and/or inferential statistical analysis
• Network analysis
• Sentiment analysis (SentiStrength)
• Experimental tests (e.g., Monte-Carlo simulations)
3B. Disaster Management & Response

(iv) Novel Methodological Innovations

TweetTracker (Kumar et al. 2011)

Source: http://tweettracker.fulton.asu.edu/
3B. Disaster Management & Response

(iv) Novel Methodological Innovations

Hybrid Processing System for Disaster-related Soc. Med. Data (Erskine et al. 2013)

The system aims to:

- Provide “extreme scalability and process ad-hoc information in real time” (ibid., 2)
- Identify unintentional and deliberate misinformation through crowdsourced expert evaluation
- Overcome language barriers faced when analyzing available soc. med. text
- Enable more efficacious allocation of resources
3B. Disaster Management & Response

(iv) Novel Methodological Innovations

**MicroFilters** (Ilyas 2014)

The system aims to:

- Extract image data from tweets that indicate direct damage, and discard cases irrelevant to rescue efforts
- Develop a machine-learned classification strategy that isolates areas with actual, mappable damage and to which rescue teams should be dispatched
- Identify ways to overcome the problem of data sparseness
(vii) Novel Methodological Innovations

Social Media Crisis-Mapping Platform (Middleton et al. 2014)

The platform aims to:

- Utilize a geo-parsing algorithm to scrape disaster-afflicted locations from tweets
- Improve “precision of street-level tweet incident reports” (ibid., 1)
- Evaluate the accuracy of resultant social media crisis maps against NGA-published data
3C. Disease Surveillance

(i) Two Key Debates

- Detection of the Spread and/or Prevalence of Diseases in a Population
  - Could soc. med. platforms be used as “health sensors”? (Achrekar et al. 2011)
  - How do soc. med. spikes in certain disease-related keywords correlate with real-life disease outbreaks as documented by official authorities? (Doan et al. 2012; Oyeyemi et al. 2014; Chew and Eysenbach 2010)

- Prospects of Using Soc. Med. Data to Predict Outbreak Patterns
  - How can soc. med. data facilitate the development of more accurate prediction methods and models? (Lazer et al. 2014)
  - Are data scraped from certain platforms (e.g., Twitter) more accurate than others (e.g., Google Flu Trends) in predicting disease outbreaks? (Broniatowski et al. 2015)
3C. Disease Surveillance

(ii) Actors Under Study

- Ordinary users of a soc. med. platform: 97.4% \((n = 37)\)
- Healthcare professionals: 2.6% \((n = 1)\)

(iii) Data Collection and Analysis Methods

Data Collection Methods

- Twitter and Weibo: Mainly APIs
- Twitter NLP: To eliminate noisy data after identifying their context (e.g., Aramaki et al. 2011; Asamoah et al. 2015)
- FluTrack.org: To collect flu-related tweets (e.g., Chorianopoulos and Talvis 2015)
- Carmen: To collect data on geo-locatable tweets (e.g., Broniatowski et al. 2013, 2015; Dredze et al. 2013; Paul et al. 2014)
(iii) Data Collection and Analysis Methods (Cont’d)

*Data Analysis Methods*

- Descriptive/Inferential Statistical Analysis (e.g., Dredze et al. 2014; Smith et al. 2015)
- Content Analysis (e.g., Chew and Eysenbach 2010; Lee et al. 2014; Oyeyemi et al. 2014)
- Sentiment Scoring and Analysis (e.g., Salathé and Khandelwal 2011)
- Social Network Analysis (e.g., Huesch et al. 2013)
- Topic Modeling: Ailment topic aspect modeling (e.g., Dredze 2012); temporal topic modeling (Chen et al. 2014)
(iv) Novel Methodological Innovations

SNEFT, Its System Architecture, and Crawler (Achrekar et al. 2011: 714)

- A social network-enabled flu trends system that aims to “track and predict flu activity” (ibid., 718)

- Provides a “timely warning to public health authorities for further investigation and response” (ibid., 714)
(iv) Novel Methodological Innovations

ATAM: The Ailment Topic Aspect Model (Dredze 2012: 83)

• A probabilistic graphical model that aims to discover ailments as well as broader correlations from raw Twitter data (ibid., 81).

• Successfully identified a “positive correlation between states with high smoking rates and those with high Twitter message rates about cancer,” among other key correlational trends (ibid., 82).
(iv) Novel Methodological Innovations

Carmen (Dredze et al. 2013)

- A geolocation system that assigns a location to every tweet “from a database of structured location information” (ibid., 2) with purported accuracy of 90%+.

- Developed to facilitate the study of tweets that originated from narrow locations of interest (e.g., a single country like the U.S.), rather than tweets in a language of interest (e.g., English) that may involve multiple vast locations.

Source: GitHub (2013) Link
3. Discussion of Results

Three Key Findings Across Thematic Areas

(i) Rather Consistent **Dominance** of Twitter-centric Studies

- Elections: 85.7% of studies; PM&RT: 42.2%; DR&M: 73.5%; DS: 97.4%
- Twitter as the most frequently used platform for (real-time) monitoring
- Other platforms (notably, YouTube and Flickr) valued for their utility in rendering possible the provision of documentary evidence

(ii) **Varying Sensitivity** toward Data (Unre)liability

- System architecture often designed to scrape data with relevant keywords, but could lack reliability verification mechanism
- Concerns about data unreliability sparingly raised by researchers, but highly salient in news treatments ("Twitter can’t be trusted in a crisis")

(iii) **Questionable Utility** of Latest Innovations

- Regularly observed inability to appeal to originally targeted end users
- NGOs’ reluctance to utilize soc.med.-derived tools has led to their gradual phase-out or broadened (sometimes, nebulous) scope
3. Discussion of Results

Twitter “can’t be trusted in a crisis”

Published on December 2, 2013

The Australian covers a study by Pascal Van Hentenryck and Manuel Cebrian about the accuracy of social media which suggests that only a small proportion of Twitter postings can be depended upon in a natural disaster or major crisis. Read the full story.

Source: The Australian (2013) Link - Subscription Needed
4. Significance of This Review to UNU-CS

The Mission Statement of UNU-CS:

(i) “Lead in investigating and inventing human centered information and communication technologies addressing some of the priorities central to the UN and the world such as: sustainability, development, governance, peace and security, human rights and human dignity.”

(ii) “Impact policymakers, within the UN system and beyond, through actionable knowledge and thought-leadership.”

(iii) “Nurture the next generation of inter-disciplinary computer scientists, social scientists and designers in developing countries.”

(iv) “Embrace the enormous dynamism of the city of Macau and Pearl River Delta region while still working globally.”

4. Significance of This Review to UNU-CS

(Policy) Implications: Three Main Considerations

(i) Addressing Noisier, Messier, and More Competitive Political Environments
   - Anticipatable increase in foreign & domestic actors involved in the “monitoring market”
   - Expected growth in sophistication and savviness of (professional) opinion manipulators both on- and offline

(ii) Developing More Advanced Data Verification Technology
   - Collaboration with scientific communities that have already gained relevant experiential knowledge (e.g., NASA and its “Planet Hunters” project)
   - Eventual establishment of reporting standards for crowdsourced data (e.g., introducing the prerequisite of documentary evidence provision)

(iii) Encouraging Closer Cooperation between System Architects and End Users to Ensure Real-life Applicability of Soc. Med. Tools
   - Introduction of regularized deliberative channels with the objective of eradicating unintentional obsolescence of developed tools
Thank you

The references for this review can be downloaded [here](#) in BibTeX format.