leveraging social networks for digital drug detection

*lessons from surveys and biosurveillance*

Nicholas Genes, MD, PhD
Michael Chary, PhD

November 8, 2013

@nickgenes  @macacharya
disclosures

• Studied chondrocyte biomechanics before emergency medicine & informatics
• Founding Partner, Medgadget.com
• Editorial Board, Medscape EM
• Editorial Board, Emergency Medicine Practice
• Editorial Board, VisualDx
• Senior Editor, Emergency Physicians Monthly
disclosures

• bath salts caught me by surprise
Miami “Face Eater” May Have Been High On “Bath Salts”


A naked man who chewed off the face of another man in what is being called a zombie-like attack may have been under the influence of “bath salts,” a drug referred to as the new LSD, according to reports from CNN affiliates in Miami.

Why ‘bath salts’ are dangerous
Emergency Department Visits After Use of a Drug Sold as "Bath Salts" --- Michigan, November 13, 2010--March 31, 2011

Weekly
May 20, 2011 / 60(19);624-627

On May 18, this report was posted as an MMWR Early Release on the MMWR website (http://www.cdc.gov/mmwr).

On February 1, 2011, in response to multiple news reports, the Michigan Department of Community Health (MDCH) contacted the Children's Hospital of Michigan Poison Control Center (PCC) regarding any reports of illness in the state caused by the use of recreational designer drugs sold as "bath salts." Unlike traditional cosmetic bath salts, which are packaged and sold for adding to bath water for soaking and cleaning, the drugs sold as "bath salts" have no legitimate use for bathing and are intended for substance abuse. These products can contain stimulant compounds such as 3,4-methylenedioxypyrovalerone (MDPV) or 4-methylmethcathinone (mephedrone). The PCC told MDCH that, earlier in the day, the PCC had learned that numerous persons had visited the local emergency department (ED) in Marquette County with cardiovascular and neurologic signs of acute intoxication. This report summarizes the subsequent investigation, which identified 35 persons who had ingested, inhaled, or injected "bath salts" and visited a Michigan ED during November 13, 2010--March 31, 2011. Among the 35 patients, the most common signs and symptoms of toxicity were agitation (23 patients [66%]), tachycardia (22 [63%]), and delusions/hallucinations (14 [40%]). Seventeen patients were hospitalized, and one was dead upon arrival at the ED. The coordinated efforts of public health agencies, health-care providers, poison control centers, and law enforcement agencies enabled rapid identification of this emerging health problem. Mitigation of the problem required the execution of an emergency public health order to remove the toxic "bath salts" from the marketplace. Lessons from the Michigan experience could have relevance to other areas of the United States experiencing similar problems.

From November 2010 to January 2011, the Marquette County ED treated seven patients who arrived at the ED with hypertension, tachycardia, tremors, motor automatisms, mydriasis, delusions, and paranoia. Some patients were violent, placing increased demand on ED staff members. Responding to the cluster also placed additional demands on local law enforcement and foster care, because many patients had young children who needed care while their parents were incapacitated. The patients reported using "bath salts" purchased at a local store for about $20 a package and labeled "not intended for human consumption." By February 3, a total of 13 cases in Marquette County and one death had been reported to the PCC. Efforts by the local ED, law enforcement, and prosecuting attorney's office led to the execution of an emergency public health order on February 4 by the Marquette County Health Department. The proprietor of the store was ordered to immediately remove from sale and turn over to government authorities any and all products known as White Rush, Cloud Nine, Ivory Wave, Ocean Snow, Charge Plus, White Lightning, Scarface, Hurricane Charlie, Red Dove, White Dove, and Sextasy. The Michigan Department of State Police laboratory tested the White Rush seized from the store and detected the presence of MDPV.
The Toxicology Investigators Consortium Case Registry--the 2011 experience.

Wiegand TJ, Wax PM, Schwartz T, Finkelstein Y, Gorodetsky R, Brent J; Toxicology Investigators Consortium Case Registry Investigators.

Abstract
In 2010, the American College of Medical Toxicology established its Case Registry, the Toxicology Investigators Consortium (ToxIC). ToxIC is a prospective registry, which exclusively compiles suspected and confirmed toxic exposure cases cared for at the bedside by medical toxicologists at its participating sites. The Registry aims to fulfill two important gaps in the field: a real-time toxicosurveillance system to identify current poisoning trends and a powerful research tool in toxicology. ToxIC allows extraction of information from medical records making it the most robust multicenter database on chemical toxicities in existence. All cases seen by medical toxicologists at participating institutions were entered in a database. Information characterizing patients entered in 2011 was tabulated. 2010 data was also included so that cumulative total numbers could be described as well. The current report is a summary of the data collected in 2011 in comparison to 2010 entries and also includes cumulative data through December 31st, 2011. During 2011, 28 sites with 49 specific institutions contributed a total of 6,456 cases to the Registry. The total number of cases entered into the registry at the end of 2011 was 10,392. Emergency departments remained the most common source of consultations in 2011, accounting for 53% of cases. The most common reason for consultation was for pharmaceutical overdoses, which occurred in 48% of patients, including intentional (37%) and unintentional (11%) exposures. The most common classes of agents were sedative-hypnotics (1,492 entries in 23% of cases), non-opioid analgesics (1,368 cases in 21% of cases), opioids (17%), antidepressants (16%), stimulants/sympathomimetics (12%), and alcohol (8%). N-acetylcysteine was the most commonly administered antidote during 2011, similar to 2010, followed by the opioid antagonist naloxone, sodium bicarbonate, physostigmine and flumazenil. Anti-crotalid Fab fragments (CroFab) were administered in 106 out of 131 cases in which an envenomation occurred. There were 35 deaths recorded in the Registry during 2011. The most common associated agents, including when reported as sole agent or in combination with other agents, were opioids and analgesics (codeine, oxycodone, oxymorphone, hydrocodone, and heroin) in 10 of the ten opioid-related deaths and heroin in three. Acetaminophen was the most common single agent reported overall being identified in all eight of the death cases attributed to analgesics. There were significant trends identified during 2011. Cases involving designer drugs including psychoactive bath salts and synthetic cannabinoids increased substantially from 2010 to 2011. The psychoactive bath salts were responsible for a large increase in stimulant/sympathomimetic-related cases reported to the Registry in 2011 with overall numbers doubling from 3% of Registry entries in 2010 to 12% in 2011. Entries involving psychoactive drugs of abuse also increased twofold from 2010 to 2011 jumping from 3 to 6%, primarily due to increasing frequency of synthetic cannabinoid ("K2") related intoxications as 2011 progressed. The 2011 Registry included over 600 ADR's (10% of Registry Cases) with 115 agents causing at least 2 ADR's. This is up from only 3% of cases (116 total cases) in 2010. The ToxIC Case Registry continues to grow. At the end of 2011, over 10,000 cases had been entered into the Registry. As demonstrated by the trends identified in psychoactive bath salt and synthetic cannabinoid reports, the Registry is a valuable toxicosurveillance and research tool. The ToxIC Registry is a unique tool for identifying and characterizing confirmed cases of significant or potential toxicity or complexity to require bedside consultation by a medical toxicologist.
The police found five additional gallons of liquid PCP in an East Harlem housing project that was raided this week, with officials calling it a dug-in hub for a drug thought to be an obsolete relic.

“I remember it from the 1970s,” Police Commissioner Raymond W. Kelly, a former police officer, said Friday of the hallucinogen. Since that era, New York has been known, in terms of drugs, as more of a “cocaine and heroin town,” he said.

The additional PCP brought the total amount seized to 7.5 gallons. Mr. Kelly said that one liquid ounce of the drug would produce about 500 doses, which each would sell for $10 in the courtyard near the Milbank-Frawley housing complex, which was raided during the week. The 7.5 gallons would have yielded about 480,000 doses, making the total amount seized worth roughly $4.8 million.

“The scope of what they were doing was tremendous,” Mr. Kelly said. “The quantity of stuff was larger than I had fully appreciated.”

The Manhattan district attorney’s office identified several local pending cases involving people believed to be on PCP.

On Dec. 16, a police officer saw a 2-year-old girl shivering in her a stroller on East 115th Street, her mother passed
new drugs of abuse

• wide variety of symptoms, risks
• less likely to interface with healthcare
• marked variation at regional / local levels
  – changing legal, economic and social landscape
new drugs of abuse

- wide variety of presentations, risks
- marked variation at regional / local levels
  - changing legal, economic and social landscape

- DAWN, NSDUH data lags, misses variants, can’t capture efficacy of new rules

- so what are our sources of information?
  - news media, PCC seminars, word-of-mouth
  - we ought to be able to do better
NSDUH origins

• 1971 (sample size: 3000, Marihuana & Cocaine)
• 1990: annual surveys began
• 1999: pencil/paper to computer-assisted / private interviews
• 2002: incentives to interviewees

• modifications
  – increase reporting, complicate trends
DAWN origins

- designed by DEA and NIDA in 1974
- became part of HHS in 1980
- joined SAMHSA in 1992
- 2011: 5 million ED visits reviewed from 223 hospitals (229k cases) (actual ED visits: 123m)
- despite protections, confusion persists over visits that “involve” illicit drugs vs. visits where patients mention illicit drugs as part of history
## Emergency Department Case Report

### 1. Facility

### 2. Date of Visit
- MONTH:  [ ]
- DAY: [ ]
- YEAR: [ ]

### 3. Time of Visit
- HOUR: [ ]
- MINUTE: [ ]
- a.m.
- p.m.
- military

### 4. Age
- Less than 1 year: [ ]
- Not documented: [ ]

### 5. Patient's Home ZIP Code
- [ ]
- Not documented: [ ]

### 6. Sex
- Male: [ ]
- Female: [ ]
- Not documented: [ ]

### 7. Race/Ethnicity
- White: [ ]
- Black or African American: [ ]
- Hispanic or Latino: [ ]
- Asian: [ ]
- American Indian or Alaska Native: [ ]
- Native Hawaiian or Other Pacific Islander: [ ]
- Not documented: [ ]

### 8. Diagnosis
List up to 4 diagnoses noted in the patient’s chart. Do not list ICD codes.
1. [ ]
2. [ ]
3. [ ]
4. [ ]

### 9. Case Description
Beginning with the presenting complaint, describe how the drug(s) was related to the ED visit. Copy verbatim from the patient's chart when possible.

### 10. Substance(s) Involved
Using available documentation, list substances that caused or contributed to the ED visit. Record substances as spelled out as possible (i.e., brand [trade] name preferred over generic name preferred over chemical name, etc.). Do not record the same substance by two different names. Do not record current medications unrelated to the visit.

#### Alcohol involved?
- Yes: [ ]
- No/Not documented: [ ]

### 11. Type of Case
Using the Decision Tree, select the first category that applies:
- Suicide attempt: [ ]
- Seeking detox: [ ]
- Alcohol only (age < 21): [ ]
- Adverse reaction: [ ]
- Overmedication: [ ]
- Malicious poisoning: [ ]
- Accidental ingestion: [ ]
- Other: [ ]

### 12. Disposition
Select one:
- Treated and released: [ ]
- Discharged home: [ ]
- Released to police/jail: [ ]
- Admitted to this hospital: [ ]
- ICU/Critical care: [ ]
- Surgery: [ ]
- Transferred: [ ]
- ICU/Critical care: [ ]
- Hospitalized: [ ]
- Other inpatient unit: [ ]
- Died: [ ]
- Referral: [ ]
- Outpatient detox/treatment: [ ]
- Psychiatric unit: [ ]
- Other: [ ]

### 13. Comments
Enter here any questions or issues you have about this case. Do not include information that could identify the patient.
is there another way?

• Post 9/11: “biosurveillance” enters lexicon
  – also based on health encounters?

  – anthrax is ideal for biosurveillance – rare, vague, but potentially very treatable if caught early

  – by the time healthcare is interfaced, too late
    • search for patient activity before doctor’s visit
Early statistical detection of anthrax outbreaks by tracking over-the-counter medication sales

Anna Goldenberg*, Galit Shmueli††, Richard A. Caruana*, and Stephen E. Fienberg*††

*Center for Automated Learning and Discovery and †Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213-3890

Contributed by Stephen E. Fienberg, February 27, 2002

The recent series of anthrax attacks has reinforced the importance of biosurveillance systems for the timely detection of epidemics. This paper describes a statistical framework for monitoring grocery data to detect a large-scale but localized bioterrorism attack. Our system illustrates the potential of data sources that may be more timely than traditional medical and public health data. The system includes several layers, each customized to grocery data and tuned to finding footprints of an epidemic. We also propose an evaluation methodology that is suitable in the absence of data on large-scale bioterrorist attacks and disease outbreaks.

biosurveillance | time series analysis | grocery data

We describe a statistical system designed for biosurveillance that is part of a larger project investigating ways to use information technology to improve clinical preparedness for bioterrorism (1). Our goal is evaluating the possible use of non-public health data, and in particular grocery sales, for the early detection of a bioterrorism attack. The potential of these data for timely detection lies in the earlier manifestation of an attack in grocery and over-the-counter (OTC) medication sales, and in their high level of detail.

We begin in the next section by providing background and a characterization of an outbreak of a bioagent, focusing on anthrax. Then we describe traditional data collected from medical and public health sources and their ability to detect attacks programs. It is especially difficult to predict, detect, or prevent a bioterrorism attack (3).

Known outbreaks of inhalational anthrax include the recent October 2001 mail-delivered anthrax envelopes in Florida, New York, Washington, DC, and New Jersey; the 1979 Sverdlovsk, Russia accident; the 1995 releases in a Tokyo subway by the terrorist group Aum Shinrikyo (3); and the 1959 outbreak in New Hampshire. From these incidents, we have learned about the fatal results and the importance of timely detection and treatment in cases of bioterrorism attacks.

The 1979 Sverdlovsk outbreak is believed to have been caused by an accidental release of Bacillus anthracis spores from a military microbiology facility nearby to where the victims lived and worked (4). This release resulted in at least 79 cases of anthrax infection and 68 documented deaths (3, 5). According to the pathologists who made the diagnosis for 42 autopsies, healthy people died within 1–4 days from contracting the bacteria (6).

An earlier outbreak occurred during a study conducted at the Arms Textile Mill in Manchester, NH in 1959. After the deaths of several workers between 1957–1959 from cutaneous anthrax, a controlled experiment was conducted at the Arms Textile Mill and at three other mills in the northeastern states. The experimental group was vaccinated against anthrax, whereas the control group received a placebo. Several months after the study began an outbreak of inhalational anthrax occurred, and because the vaccination proved to be effective, all workers were vacci-
Rapid Deployment of an Electronic Disease Surveillance System in the State of Utah for the 2002 Olympic Winter Games

Per H. Gesteland, MD $^1$, Michael M. Wagner, MD, PhD $^{2,3}$, Wendy W. Chapman $^2$, PhD$^2$, Jeremy U. Espino, MD $^2$, Fu-Chiang Tsui, PhD $^{2,3}$, Reed M. Gardner, PhD $^1$, Robert T. Rolfs, MD, MPH $^{1,4}$, Virginia Dato, MD, MPH, $^4$Brent C. James, MD, MStat $^{1,5}$, Peter J. Haug, MD $^{1,5}$

$^1$Department of Medical Informatics, University of Utah, Salt Lake City, UT
$^2$The RODS Laboratory, Center for Biomedical Informatics, University of Pittsburgh, PA
$^3$Biomedical Security Institute, University of Pittsburgh and Carnegie Mellon, PA
$^4$Utah Department of Health, $^5$Intermountain Health Care, Salt Lake City, UT

ABSTRACT

The key to minimizing the effects of an intentionally caused disease outbreak is early detection of the attack and rapid identification of the affected individuals. The Bush administration's leadership in advocating for biosurveillance systems capable of monitoring for bioterrorism attacks suggests that we should move quickly to establish a nationwide early warning biosurveillance system as a defense against this threat. The spirit of collaboration and unity inspired by the events of 9-11 and the 2002 Olympic Winter Games in Salt Lake City provided the opportunity to demonstrate how a

RODS is an electronic public health surveillance system that has been deployed since 1999 in Western Pennsylvania. The system captures clinical data from multiple and competing health systems under shared data agreements. RODS receives data in real time from existing clinical information systems in the form of HL7 messages. It then automatically analyzes data from all patients presenting for acute care with chief complaints of diarrhea, rash, respiratory illness, viral illness and other key symptoms and looks for changes in the usual patterns. A web-based interface provides access to a geographic information system, graphical analytical tools, and other analytical tools that facilitate rapid
Health Information Exchange, Biosurveillance Efforts, and Emergency Department Crowding During the Spring 2009 H1N1 Outbreak in New York City

Jason S. Shapiro, MD, MA
Nicholas Genes, MD, PhD
Gilad Kuperman, MD, PhD
Kevin Chason, DO
The New York Clinical Information Exchange (NYCLiX) Clinical Advisory Committee H1N1 Working Group*
Lynne D. Richardson, MD

From the Department of Emergency Medicine (Shapiro, Genes, Chason, Richardson) and Center for Biomedical Informatics (Shapiro), Mount Sinai School of Medicine, New York, NY; the New York Presbyterian Hospital, New York, NY (Kuperman); the Department of Biomedical Informatics, Columbia University, New York, NY (Kuperman); and the New York Clinical Information Exchange, New York, NY (Shapiro, Kuperman).

Novel H1N1 influenza spread rapidly around the world in spring 2009. Few places were as widely affected as the New York metropolitan area. Emergency departments (EDs) in the region experienced daily visit increases in 2 distinct temporal peaks, with means of 36.8% and 60.7% over baseline in April and May, respectively, and became, in a sense, the “canary in the coal mine” for the rest of the country as we braced ourselves for resurgent spread in the fall. Biosurveillance efforts by public health agencies can lead to earlier detection, potentially forestalling spread of outbreaks and leading to better situational awareness by frontline medical staff and public health workers as they respond to a crisis, but biosurveillance has traditionally relied on manual reporting by hospital administrators when they are least able: in the midst of a public health crisis. This article explores the use of health information exchange networks, which enable the secure flow of clinical data among otherwise unaffiliated providers across entire regions for the purposes of clinical care, as a tool...
Figure 1. Daily ED visits across all 11 NYCLIX sites during the spring 2009 H1N1 outbreak. Also superimposed is a 7-day moving average and key events in local media coverage of H1N1.
parallel efforts for biosurveillance

• more widespread EHR, HIE has helped biosurveillance – but also shows limitations

• renewed focus on personal actions, instead of patient encounters
  – improved consumer technology helps efforts
  – social networks + smartphones (& fitness devices?)
Infodemiology: Tracking Flu-Related Searches on the Web for Syndromic Surveillance

Gunther Eysenbach MD MPH 1) 2)
1) Centre for Global eHealth Innovation, University Health Network, Toronto M5G2C4, Canada, and 2) Department of Health Policy, Management and Evaluation, University of Toronto, Canada.
Email: geysenba@uhnres.utoronto.ca

Abstract
Background: Syndromic surveillance uses health-related data that precede diagnosis and signal a sufficient probability of a case or an outbreak to warrant further public health response.
Objective: While most syndromic surveillance systems rely on data from clinical encounters with health professionals, I started to explore in 2004 whether analysis of trends in Internet searches can be useful to predict outbreaks such as influenza epidemics and prospectively gathered data on Internet search trends for this purpose.
Results: There is an excellent correlation between the number of clicks on a keyword-triggered link in Google with epidemiological data from the flu season 2004/2005 in Canada (Pearson correlation coefficient of current week clicks with the following week influenza cases \( r = .91 \)). The “Google ad sentinel method” proved to be relevant for assessing syndromic surveillance use and seeking health information \(^2\). An interesting question is whether tracking health information seeking behaviour of populations over time can be used for public health purposes, particularly syndromic surveillance. The CDC defines syndromic surveillance as “surveillance using health-related data that precede diagnosis and signal a sufficient probability of a case or an outbreak to warrant further public health response.” While most syndromic surveillance systems rely on data from clinical encounters with health professionals, monitoring for example sick-leave prescriptions, house calls, hospital- or pharmacy-based data \(^4\), \(^5\), there have also been previous experiments with unconventional methods to use preclinical “health information seeking” data for syndromic surveillance, for example monitoring calls to a “NurseLine” such as NHS Direct \(^6\)–\(^8\). However, there does not seem to be any prior evaluation of the use of Internet search data for syndromic surveillance.
Eysenbach 2006

• do searches predict subsequent ILI?
• Google doesn’t share search data to public
• bought ads on Google for flu-related keywords
  – cost $365.64 for 2004/5 Canada flu season
  – ad campaign simply asked searchers about flu sx
• compared to FluWatch Sentinel labs, docs
Figure 2 (above). Scatter plot (with regression line) showing the excellent correlation between “clicks” and flu cases in the following week (*)

Figure 3 (above). Scatter plot (with regression line) showing the considerably worse correlation between sentinel physicians’ ILI reports and flu cases in the following week (**)
Explore flu trends - United States

We've found that certain search terms are good indicators of flu activity. Google Flu Trends uses aggregated Google search data to estimate flu activity. Learn more »

National

[Graph showing flu trends over time]

States | Cities (Experimental)
what if you don’t need to ask Google about your flu symptoms?

social networks could capture more:
  sx, rx, sentiment, contacts ... prognosis?
The Use of Twitter to Track Levels of Disease Activity and Public Concern in the U.S. during the Influenza A H1N1 Pandemic

Alessio Signorini¹, Alberto Maria Segre¹, Philip M. Polgreen²,³*

1 Department of Computer Science, University of Iowa, Iowa City, Iowa, United States of America. 2 Department of Epidemiology, College of Public Health, University of Iowa, Iowa City, Iowa, United States of America. 3 Department of Internal Medicine, Carver College of Medicine, University of Iowa, Iowa City, Iowa, United States of America

Abstract

Twitter is a free social networking and micro-blogging service that enables its millions of users to send and read each other's "tweets," or short, 140-character messages. The service has more than 190 million registered users and processes about 55 million tweets per day. Useful information about news and geopolitical events lies embedded in the Twitter stream, which embodies, in the aggregate, Twitter users' perspectives and reactions to current events. By virtue of sheer volume, content embedded in the Twitter stream may be useful for tracking or even forecasting behavior if it can be extracted in an efficient manner. In this study, we examine the use of information embedded in the Twitter stream to (1) track rapidly-evolving public sentiment with respect to H1N1 or swine flu, and (2) track and measure actual disease activity. We also show that Twitter can be used as a measure of public interest or concern about health-related events. Our results show that estimates of influenza-like illness derived from Twitter chatter accurately track reported disease levels.

Citation: Signorini A, Segre AM, Polgreen PM (2011) The Use of Twitter to Track Levels of Disease Activity and Public Concern in the U.S. during the Influenza A H1N1 Pandemic. PLoS ONE 6(5): e19467. doi:10.1371/journal.pone.0019467

Editor: Alison P. Galvani, Yale University, United States of America

Received September 20, 2010; Accepted April 7, 2011; Published May 4, 2011

Copyright: © 2011 Signorini et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Support for this research was provided by a National Institutes of Health Career Award NIAID K01 AI75089 (PMP), by the National Institutes of Health grant NIAID-R21-AI081164 (AMS, PMP), and by the Robert Wood Johnson Foundation Pioneer Portfolio. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: philip-polgreen@uiowa.edu
social media biosurveillance

• potential to detect outbreak early

• traditional media can confounds both ends
  – sharing news reports over networks
  – worried well seeking reassurance, stockpiling Rx

• beyond keywords: capture sentiment?
tweetfeel

Try some Twitter trends:

😊     😞  =  0%
why Twitter?

• the most versatile of the social platforms
  – public, anonymous but trackable over time + place
  – short messages but heavy usage, hashtags
    • graffiti analogy

• Facebook = identified (pros and cons)

• Instagram, YouTube, Google+
The Geography of Happiness: Connecting Twitter Sentiment and Expression, Demographics, and Objective Characteristics of Place

Lewis Mitchell, Morgan R. Frank, [...], and Christopher M. Danforth

Additional article information

Abstract

We conduct a detailed investigation of correlations between real-time expressions of individuals made across the United States and a wide range of emotional, geographic, demographic, and health characteristics. We do so by combining (1) a massive, geo-tagged data set comprising over 80 million words
Mitchell 2013

• looked at 10 million geo-tagged tweets
• Mechanical Turk categorization of 10,000 words
  – “lol” = happy  “earthquake” = unhappy
  – no attempt to derive meaning, context
• compared w/ census, satisfaction surveys, murder
  – “café” correlated w/ education, “love” anticorrelated
  – happiness correlated with wealth, against profanity
• Spanish tweets with “con” & “sin” = unhappy
Using Twitter to Examine Smoking Behavior and Perceptions of Emerging Tobacco Products

Mark Myslin¹; Shu-Hong Zhu², PhD; Wendy Chapman³, PhD; Mike Conway³, PhD

¹Department of Linguistics, University of California, San Diego, La Jolla, CA, United States
²Department of Family and Preventive Medicine, University of California, San Diego, La Jolla, CA, United States
³Department of Medicine, University of California, San Diego, La Jolla, CA, United States

Corresponding Author:
Mike Conway, PhD

Department of Medicine
University of California, San Diego
9500 Gilman Drive
La Jolla, CA, 92093
United States
Phone: 1 858 822 4931
Fax: 1 858 822 1934
Email: mconway [at] ucsd.edu

ABSTRACT

Background: Social media platforms such as Twitter are rapidly becoming key resources for public health surveillance applications, yet little is known about Twitter users’ levels of informedness and sentiment toward tobacco, especially with regard to the emerging tobacco control challenges posed by hookah and electronic cigarettes.

Objective: To develop a content and sentiment analysis of tobacco-related Twitter posts and build machine learning classifiers to detect tobacco-relevant posts and sentiment towards tobacco, with a particular focus on new and emerging products like hookah and electronic cigarettes.

Methods: We collected 7362 tobacco-related Twitter posts at 15-day intervals from December 2011 to July 2012. Each tweet was manually classified using a triaxial scheme, capturing genre, theme, and sentiment. Using the collected data, machine-learning classifiers were trained to detect tobacco-related vs irrelevant tweets as well as positive vs negative sentiment, using Naïve Bayes, k-nearest neighbors, and Support Vector Machine (SVM) algorithms. Finally, phi contingency coefficients were computed between each of the categories to discover emergent patterns.

Results: The most prevalent genres were first- and second-hand experience and opinion, and the most frequent themes were hookah, cessation, and pleasure. Sentiment toward tobacco was overall more positive (1939/4215, 46% of tweets) than negative (1349/4215, 32%) or neutral among tweets mentioning it,
"i hate cigarettes with a passion" input tweet

convert to ngrams

unigrams

stopword removal

Porter stemmer

bigrams

trigrams

convert lexical variants to a canonical form (e.g. "smoking", "smoked" → "smok")

MACHINE LEARNING CLASSIFIER (SUPPORT VECTOR MACHINE)

output class ...

output class ...

"i hate cigarettes with a passion" TOBACCO RELATED
beyond sentiment

• sentiment can inform public health campaigns

• can sentiment reveal disease?
Depressive Moods of Users Portrayed in Twitter

Minsu Park  
KAIST  
373-1 Guseond-dong  
Deajeon, Korea  
mansumansu@kaist.ac.kr

Chiyoung Cha  
Ewha Womans University  
82-2 Daehyeon-dong  
Seoul, Korea  
chiyoung@ewha.ac.kr

Meeyoung Cha  
KAIST  
373-1 Guseond-dong  
Deajeon, Korea  
meeyoungcha@kaist.ac.kr

ABSTRACT

Potential benefits of using online social network data for clinical studies on depression are tremendous. In this paper, we present a preliminary result on building a research framework that utilizes real-time moods of users captured in the Twitter social network and explore the use of language in describing depressive moods. First, we analyzed a random sample of tweets posted by the general Twitter population during a two-month period to explore how depression is talked about in Twitter. We found remarkable activities related to depression in Twitter. A large number of tweets contained detailed information about depressed feelings, status, as well as treatment history. Going forward, we conducted a study on 69 participants to determine whether the use of sentiment words of depressed users differed from a typical user. We found that the use of words related to negative emotions and anger significantly increased among Twitter users with major depressive symptoms compared to those otherwise. However, no difference was found in the use of words related to positive emotions between the two groups. Our work provides several evidences that online social networks provide meaningful data for capturing depressive moods of users.

Depression is currently the most commonly diagnosed mental disorder in many developed countries, accounting for 75% of psychiatric admissions and affecting between 9 and 30% of the adult population each year. The costs associated with depression and mental disorder have grown rapidly over recent years and the National Institute of Mental Health reports that in the U.S. major mental disorders cost at least $193 billion annually in lost earnings alone, which is comparable to that spent for cancer treatment [16]. Despite increasing public knowledge and awareness, many individuals with depression go undetected and untreated, leading to a serious public health problem, because as many as half of undetected depressed patients later meet the criteria for major depression [8].

A number of public programs have been proposed to decrease the prevalence of undiagnosed depression including the National Depression Screening Day (NDSD) and the National Anxiety and Depression Awareness Week. These programs raise awareness of depression and offer free depression screening to the general population [11]. Besides screening depression, they also provide information about treatment and promote public discussion. While these public programs are an important step towards solving depression, their main limitation lies in the selection bias of people they can reach, because the programs are participation-oriented.

Categories and Subject Descriptors
H.4 [Information Systems Applications]: Miscellaneous;
J.8 [Family and Medical Sciences]: Health; J.4 [Social and Behavioral Sciences]: Psychology
Figure 3: Average positive and negative sentiment scores in tweets depending on the groups classified by CES-D score
Predicting Postpartum Changes in Emotion and Behavior via Social Media

Munmun De Choudhury          Scott Counts          Eric Horvitz
Microsoft Research, One Microsoft Way, Redmond WA 98052
{munmund, counts, horvitz}@microsoft.com

ABSTRACT
We consider social media as a promising tool for public health, focusing on the use of Twitter posts to build predictive models about the influence of childbirth on the forthcoming behavior and mood of new mothers. Using Twitter posts, we quantify postpartum changes in 376 mothers along dimensions of social engagement, emotion, social network, and linguistic style. We then construct statistical models from a training set of observations of these measures before and after the reported childbirth, to forecast significant postpartum changes in mothers. The predictive models can classify mothers who will change significantly following childbirth with an accuracy of 71%, using observations about their prenatal behavior, and as accurately as 80-83% when additionally leveraging the initial 2-3 weeks of postnatal data. The study is motivated by the opportunity to use social media to identify mothers at risk of postpartum depression, an underreported health concern among large populations, and to inform the design of low-cost, privacy-sensitive early-warning systems and intervention programs aimed at promoting wellness postpartum.

more extreme changes. According to the CDC\(^1\), between 12 and 20 percent of new mothers report postpartum depression (a 13\% incidence rate in a meta-analysis report [20]), a form of depression that typically begins in the first month after giving birth and is characterized by symptoms including sadness, guilt, exhaustion, and anxiety [16].

We examine social media as a tool in public health. Social media is a source of population data about behaviors, thoughts, and emotions, and can serve as record and sensor for events in peoples’ lives. Whether in the form of explicit commentary, patterns of posting, or in the subtleties of language used, social media posts bear the potential to offer evidence as to how a person is affected by life events. Within this context, we investigate the feasibility of forecasting future behavioral changes of mothers following the important life event of childbirth. We extend our prior research that examines the value of harnessing social media signals to characterize changes in new mothers, along three dimensions: patterns of posting, linguistic style, and emotional expression [8]. These measures were used to explore the behavioral changes of a cohort of new mothers who showed large postpartum changes, including those
De Choudhury, 2013

• judged 376 birth announcements from women

• looked at their tweets, 3 months forward/back
  – replies, retweets, links, question marks
  – LIWC lexicon of validated emotional categories
  – LIWC style: personal pronouns, qualifiers, aux verbs
<table>
<thead>
<tr>
<th>Measures</th>
<th>C1</th>
<th>C0</th>
<th>Measures</th>
<th>C1</th>
<th>C0</th>
</tr>
</thead>
<tbody>
<tr>
<td>volume</td>
<td>-0.878</td>
<td>0.838</td>
<td>verbs</td>
<td>-0.911</td>
<td>0.951</td>
</tr>
<tr>
<td>Replies</td>
<td>-0.983</td>
<td>1.417</td>
<td>aux-verbs</td>
<td>-0.889</td>
<td>0.914</td>
</tr>
<tr>
<td>RT</td>
<td>-0.847</td>
<td>2.188</td>
<td>adverbs</td>
<td>3.025</td>
<td>-0.678</td>
</tr>
<tr>
<td>Links</td>
<td>-0.791</td>
<td>2.347</td>
<td>preposition</td>
<td>-0.899</td>
<td>1.020</td>
</tr>
<tr>
<td>Questions</td>
<td>2.110</td>
<td>-0.749</td>
<td>conjunction</td>
<td>-0.890</td>
<td>1.042</td>
</tr>
<tr>
<td>PA</td>
<td>-0.798</td>
<td>0.428</td>
<td>negate</td>
<td>2.867</td>
<td>-0.690</td>
</tr>
<tr>
<td>NA</td>
<td>1.726</td>
<td>-0.593</td>
<td>quantifier</td>
<td>-0.896</td>
<td>0.950</td>
</tr>
<tr>
<td>activation</td>
<td>-0.743</td>
<td>0.525</td>
<td>swear</td>
<td>3.258</td>
<td>-0.751</td>
</tr>
<tr>
<td>dominance</td>
<td>-0.717</td>
<td>0.543</td>
<td>tentative</td>
<td>3.710</td>
<td>-0.700</td>
</tr>
<tr>
<td>Followers</td>
<td>-0.941</td>
<td>1.451</td>
<td>certain</td>
<td>-0.997</td>
<td>1.386</td>
</tr>
<tr>
<td>Followees</td>
<td>-0.928</td>
<td>1.013</td>
<td>inhibition</td>
<td>3.425</td>
<td>-0.722</td>
</tr>
<tr>
<td>functional words</td>
<td>4.574</td>
<td>-0.865</td>
<td>inclusive</td>
<td>-0.914</td>
<td>1.048</td>
</tr>
<tr>
<td>1st pp.</td>
<td>1.924</td>
<td>-0.426</td>
<td>exclusive</td>
<td>-0.822</td>
<td>2.304</td>
</tr>
<tr>
<td>2nd pp.</td>
<td>-0.901</td>
<td>0.872</td>
<td>assent</td>
<td>-0.904</td>
<td>1.073</td>
</tr>
<tr>
<td>3rd pp.</td>
<td>-0.894</td>
<td>0.990</td>
<td>non-fluency</td>
<td>3.028</td>
<td>-0.724</td>
</tr>
<tr>
<td>indefinite pronoun</td>
<td>-0.810</td>
<td>2.082</td>
<td>filler</td>
<td>2.929</td>
<td>-0.786</td>
</tr>
<tr>
<td>article</td>
<td>3.940</td>
<td>-0.701</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Median changes in measures for two classes of new mothers with C1 corresponding to extreme-changing mothers and C0 corresponding to mothers showing smaller changes. Directionality of changes reflects for each measure whether extreme-changing mothers are considered to be changing in increasing or decreasing directions. *Volume* decreases postpartum for C1, but *negative affect* (NA) increases [6,8].
following individuals over time

• suggests changes to mental health, behavior

• add location to reveal contagious disease?
Modeling Spread of Disease from Social Interactions

Adam Sadilek  
Department of Computer Science  
University of Rochester  
Rochester, NY 14627  
sadilek@cs.rochester.edu

Henry Kautz  
Department of Computer Science  
University of Rochester  
Rochester, NY 14627  
kautz@cs.rochester.edu

Vincent Silenzio  
School of Medicine and Dentistry  
University of Rochester  
Rochester, NY 14627  
v.m.silenzio@rochester.edu

Abstract

Research in computational epidemiology to date has concentrated on coarse-grained statistical analysis of populations, often synthetic ones. By contrast, this paper focuses on fine-grained modeling of the spread of infectious diseases throughout a large real-world social network. Specifically, we study the roles that social ties and interactions between specific individuals play in the progress of a contagion. We focus on public Twitter data, where we find that for every health-related message there are more than 1,000 unrelated ones. This class imbalance makes classification particularly challenging. Nonetheless, we present a framework that accurately identifies sick individuals from the content of online communication. Evaluation on a sample of 2.5 million geo-tagged Twitter messages shows that social ties to infected, symptomatic people, as well as the intensity of recent co-location, sharply increase one’s likelihood of contracting the illness in the near future. To our knowledge, this work is the first to model the interplay of social activity, human mobility, and the spread of infectious disease in a large real-world population. Furthermore, we provide the first quantifiable estimates of the characteristics of disease transmission on a large scale without active user participation—a step towards our ability to model and predict the emergence of global epidemics from day-to-day interpersonal interactions.
Sadilek, 2012

- 16m tweets, 630k users: 1 in 30 NYC residents
- GPS-encoded tweeters about illness? 2047
- people colocated within 100 m in brief windows

- keyword searches minus retweets, rated by humans, trained SVM to weight features
  - “sick” without “sick of”
5a: colocation with 40 sick people over 1 hour = 80 sick people over 4 hours = 20% chance of getting sick

5b: more sick friends increases one’s odds of getting sick (unrelated to total # of friends)

Modifying factors?

high risk restaurants
location, time & age for toxicology?
Tweaking and Tweeting: Exploring Twitter for Nonmedical Use of a Psychostimulant Drug (Adderall) Among College Students

Carl L Hanson, PhD, Scott H Burton, [...], and Bret Hansen, BS(Current)

Abstract

Background

Adderall is the most commonly abused prescription stimulant among college students. Social media provides a real-time avenue for monitoring public health, specifically for this population.
Hanson, 2013

- queried Twitter API for Adderall +GPS tweets
  - Removed @Pharm, @Adderall – sourced tweets
- put tweets into “college clusters” of 150 miles
- divided Adderall tweets by student body pops
Hanson, 2013

• Adderall tweets frequently mention alcohol (4.8%), other stimulants (4.7%), cocaine (0.9%)

• Side effects included sleep deprivation (5.0%) or loss of appetite (2.6%) – n/v/d rare
My Being Awake Right Now is brought to you in part by Adderall, with additional promotional consideration by Crippling Anxiety.

Adderall Receives Honorary Degree From Harvard

Do you use ADHD medications like Adderall for "work-performance enhancement"? Interesting read in the @theatlantic inside.pn/9t50QY

Adderall in the morning, Ambien in the evening, vodka in the water bottle

What's @nytimes' etiquette expert @StevenPetrow's productivity secret? Adderall, actually. goo.gl/HrJ8G

Beyond caffeine: The benefits and dangers of using Adderall theatin.tc/1azyFx

Florida Woman Gives 6-Year-Old Son Adderall As "Experiment" | fark.com/goto/8001065/h...

Woman accused of child abuse; gave Adderall to son, 6, as experiment

A West Boca woman is facing a child abuse charge after deputies say she emailed her 6-year-old's teacher to say she'd given the boy Adderall as an experiment, and asked for updates on any behavior...
toward digital drug detection
problems of demographics, location

• who is underage or particularly at-risk?
  – diction / grammar / usage may distinguish
  – content (school-related, pop-related)
  – characteristics of network

• where are these people?
  – GPS (1-3% sometimes enough)
  – location identifiers in profile, or messages
  – location of social network
problems of meaning

• more vexing

• sarcasm, jokes, music lyrics, lying

• maybe any reference means something

• NSDUH was refined over time
approaches to meaning

• format rules?
  – no links (probably a news story)
  – delete duplicates, or crossmatch against lyrics
  – only conversations
  – filter out spambots (based on followers, frequency)
approaches to meaning

• rules about what to exclude
• machine learning techniques

  – trained on human ratings
  – set loose, prospectively evaluate
  – periodic recalibration

  – not to detect sentiment, but implied usage
I don't always smoke marijuana,... but when I... oh who am I kidding? I smoke everyday. #StonerLife

87-year-old Denver man accused of growing 400 marijuana plants t.co/mCfi6z3g well damn o___O
I SELL COCAINE
AND COCAINE ACCESSORIES
digital drug detection

• borrows from increasingly sophisticated biosurveillance techniques
  – inferences on location, network characteristics

• still limited by keyword / sentiment analysis

• NLP techniques needed to derive meaning
next steps (for Twitter toxicology)

• correlate drug tweets with survey or crime data (by location, where available)

• evaluate effectiveness of policy, outreach

• follow user’s tweets over time to assess risk, influence, resource utilization, recidivism