From Social to Medical:
how is digital information being used?

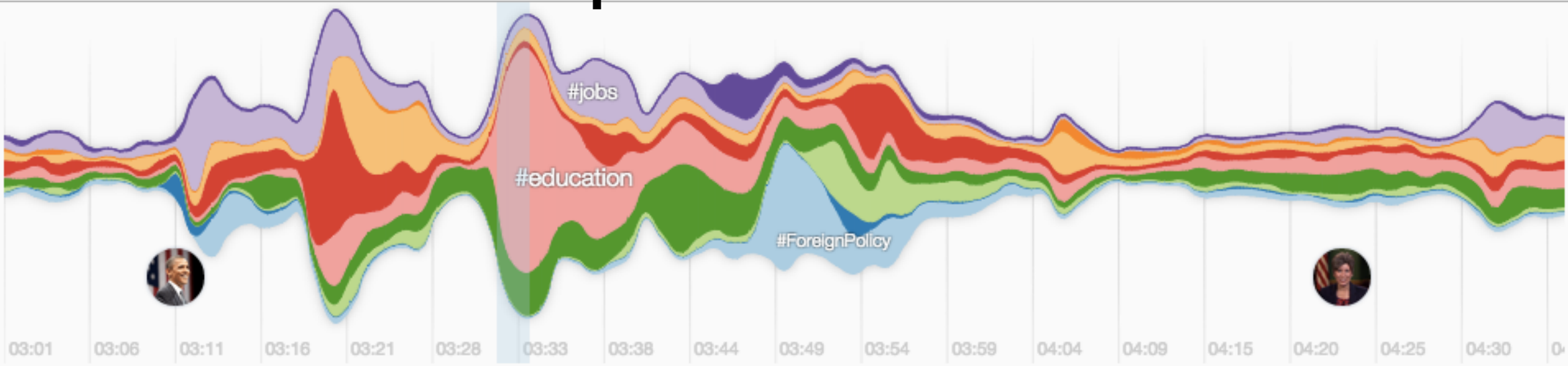
Ciro Cattuto
ISI Foundation
Torino, Italy & New York City

European Medicines Agency
London, 19 September 2016

INSTITUTE
FOR SCIENTIFIC INTERCHANGE
FOUNDATION


user-generated content
Twitter

temporal context

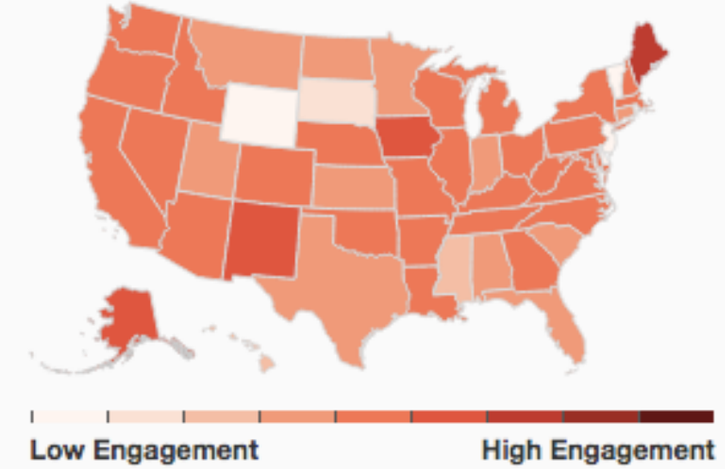


spread that idea all across America, so that two years of college becomes as free and universal in America as high school is today. And I want to work with this Congress, to make sure Americans already burdened with student loans can reduce their monthly payments, so that student debt doesn't derail anyone's dreams.

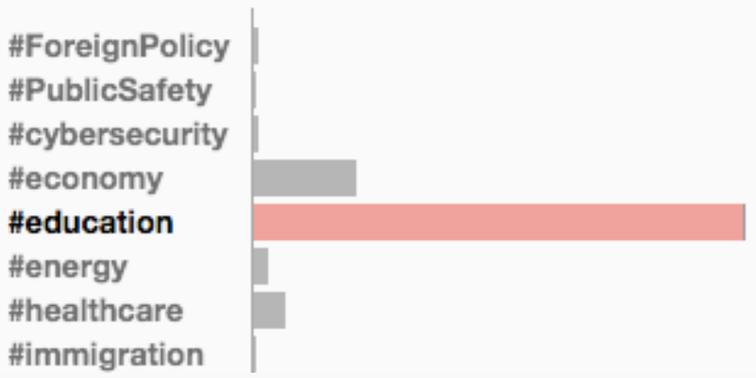
Thanks to Vice President Biden's great work to update our job training system, we're connecting community colleges with local employers to train workers to fill high-paying jobs like coding, and nursing, and robotics. Tonight, I'm also asking more businesses to follow the lead of companies like CVS and UPS, and offer more educational benefits and paid apprenticeships — opportunities that give workers the chance to earn higher-paying jobs even if they don't have a higher education.

And as a new generation of veterans comes home, we owe them every opportunity to live the American Dream they helped defend. Already, we've made strides towards ensuring that every veteran has access to the

Real-time engagement distribution on Twitter for this paragraph



Map for #education



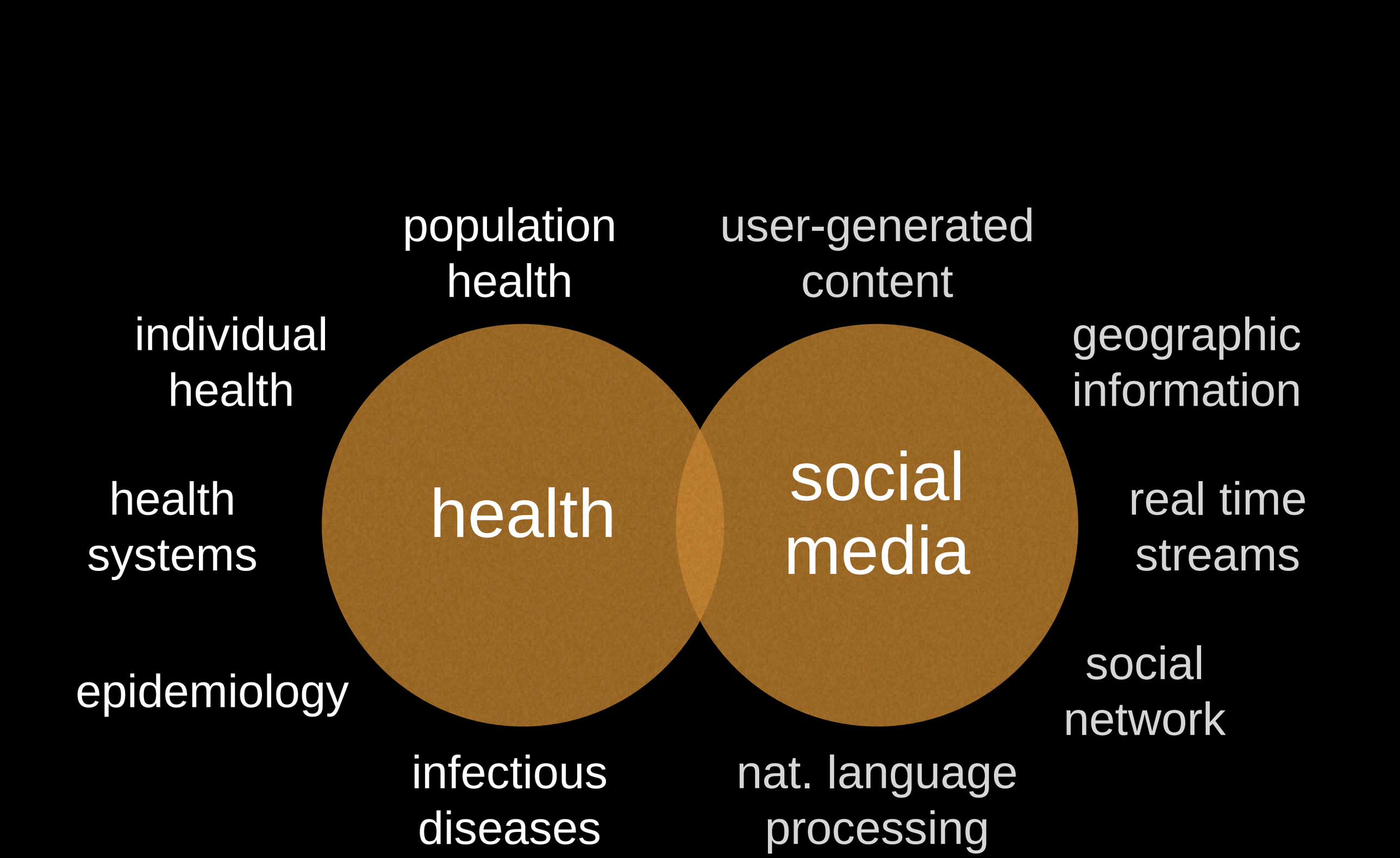
twitter.github.io/interactive/sotu2015

geographic context

social context

digital traces

social media **users as “sensors”**
real-time streams
geo-referenced information
social network context
machine processable textual content
high penetration, good population coverage


population health
user-generated content
individual health
geographic information
health systems
real time streams
health
social media
epidemiology
social network
infectious diseases
nat. language processing

COMPUTATIONAL HEALTH

www2017.com.au

We invite research contributions for the 26th World Wide Web Conference Computational Health Track.

Health and medicine are increasingly searched for, tracked, and delivered via digital means. One of the primary among these is the Internet. This represents opportunities to improve our health and the delivery of medicine, for example, by learning about aspects of people's health that are difficult to otherwise track, by facilitating rapid collection and dissemination of time-critical medical data, and by providing novel interventions to improve health.

Computational Health

Crowdsourcing

Internet Monetisation and Online Markets

Search

Security and Privacy

Semantics and Knowledge

Social Network Analysis and Computational Social Science

Systems and Infrastructure

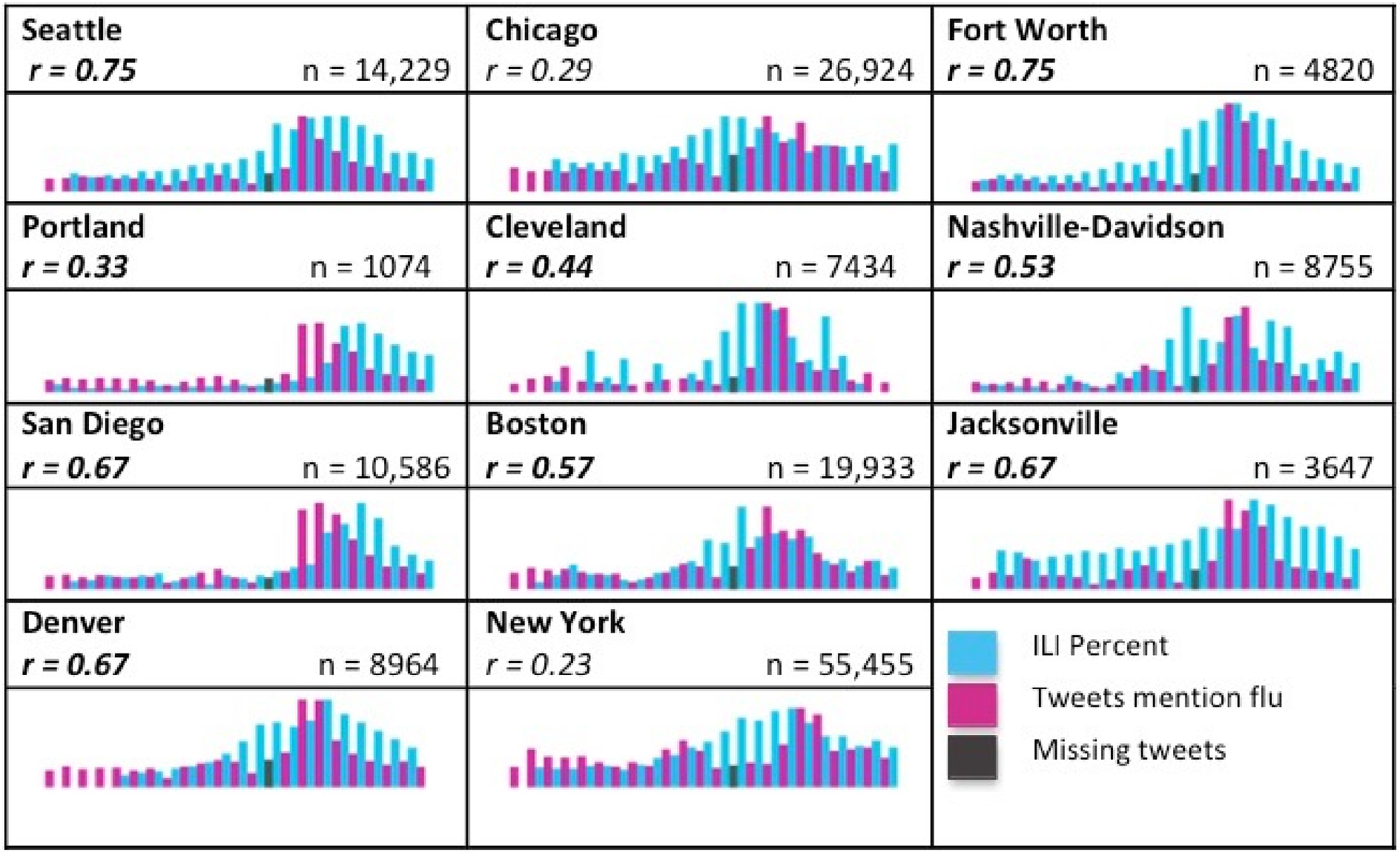
Ubiquitous and Mobile Computing

User Modeling, Personalisation and Experience

use cases

- influenza prevalence
- outbreak investigation
- obesity prevalence
- heart disease mortality
- adverse effects of HIV drug treatment
- spatio-temporal allergy patterns
- vaccination coverage
- mental health
- food poisoning
- Web-based patient communities

influenza prevalence vs Twitter-based indicators



A.C. Nagel *et al.*,
The Complex Relationship of Realspace Events and Messages in Cyberspace: Case Study of Influenza and Pertussis Using Tweets
J Med Internet Res 15(10), e237 (2013)

www.zikatracking.org

ZIKATRACKING

ZIKA AWARENESS FROM PLACE MENTIONS

LAST DAY / LAST HOUR

Last updated on: 18 Sep 17:07

Florida

Health News @health_news11 9m
Florida expands Zika zone in Miami Beach after five new cases.. Related Articles: https://t.co/jIAODZ7uIo
1 retweet since 18 Sep 15:30

DesGlobal @dspalten 27m
#Zika zone triples in #Miami Beach after more cases found https://t.co/AVf10V2Ie0 /article102462592.html #Florida
4 retweets since 18 Sep 06:02

Reuters Top News @Reuters 36m
Florida expands Zika zone in Miami Beach after five new cases https://t.co/ODh57cJbIr
146 retweets since 17 Sep 07:36

Televicentro HN @televicentrohn 2h
Advertencia por el #zika en #Florida: "No se bañen en las playas de South Beach" . https://t.co/HoDLlMBPaM https://t.co/KPQl5QdVqH
1 retweet since 18 Sep 06:15

ちょいエロ吐きます。 @tw56itterpp 2h
Sex ban recommended for British tourists in Florida after Zika outbreak。(ˊ3ˋ)

Inside Code @Elit12345 2h
Wann Rot-Grün offiziell wird. PLUS: Nächste Schritte in Florida. First local transmission of Zika virus confirmed in Koalitionsverhandlungen

zikatracking.org

Number of mentions
10 100
10

OPEN ACCESS Freely available online

PLoS MEDICINE

Health in Action

Surveillance Sans Frontières: Internet-Based Emerging Infectious Disease Intelligence and the HealthMap Project

John S. Brownstein*, Clark C. Freifeld, Ben Y. Reis, Kenneth D. Mandl

The Opportunity

As developed nations continue to strengthen their electronic disease surveillance capacities [1], the parts of the world that are most vulnerable to emerging disease threats still lack essential public health information infrastructure [2,3]. The existing network of traditional surveillance efforts managed by health ministries, public health institutes, multinational agencies, and laboratory and institutional networks has wide gaps in geographic coverage and often suffers from poor and sometimes suppressed information flow across national borders [4]. At the same time, an enormous amount of valuable information about infectious diseases is found in Web-accessible information sources such as discussion sites, disease reporting networks, and news outlets [5,6,7]. These resources can support situational awareness by providing current, highly local information about outbreaks, even from areas relatively invisible to traditional global

Summary Points

- Valuable information about infectious diseases is found in Web-accessible information sources such as discussion forums, mailing lists, government Web sites, and news outlets.
- Web-based electronic information sources can play an important role in early event detection and support situational awareness by providing current, highly local information about outbreaks, even from areas relatively invisible to traditional global public health efforts.
- While these sources are potentially useful, information overload and difficulties in distinguishing "signal from noise" pose substantial barriers to fully utilizing this information.
- HealthMap is a freely accessible, automated real-time system that monitors, organizes, integrates, filters, visualizes, and disseminates online information about emerging diseases.
- The goal of HealthMap is to deliver real-time intelligence on a broad range

department visits with acute respiratory illness [9,10]. This was followed by media reports of a respiratory disease among health care workers in February 2003, all captured by the Public Health Agency of Canada's Global Public Health Intelligence Network (GPHIN) [10,11,12]. In parallel, online discussions on the ProMED-mail system referred to an outbreak in Guangzhou, well before official government reports were issued [13].

These Web-based data sources not only facilitate early outbreak detection, but also support increasing public awareness of disease outbreaks prior to their formal recognition. Through low-cost and real-time Internet data-mining, combined with openly available and

Funding: This work was supported by grants R21LM009263-01 and R01LM007970-01 from the National Library of Medicine, the National Institutes of Health, the Canadian Institutes of Health Research, and a research grant from Google.org. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

840 alerts for Alerts from past week

Outbreaks in Current Location

Sign up for email alerts

314 Respiratory Alerts

SARS (2), MERS (184), Swine Flu H1N1 (12), Coronavirus (34), Avian Influenza H7N9 (32), Influenza (11), Bronchitis/Bronchiolitis (2), Tuberculosis (8), Whooping Cough (13), Avian Influenza H5N1 (8), Hantavirus (4), Legionnaires' (2), Respiratory Illness (2)

186 Vectorborne Alerts

Dengue (125), Chikungunya

www.healthmap.org

Google

Terms of Use

HealthMap About Mobile Projects Disease Daily

Log in

participatory surveillance

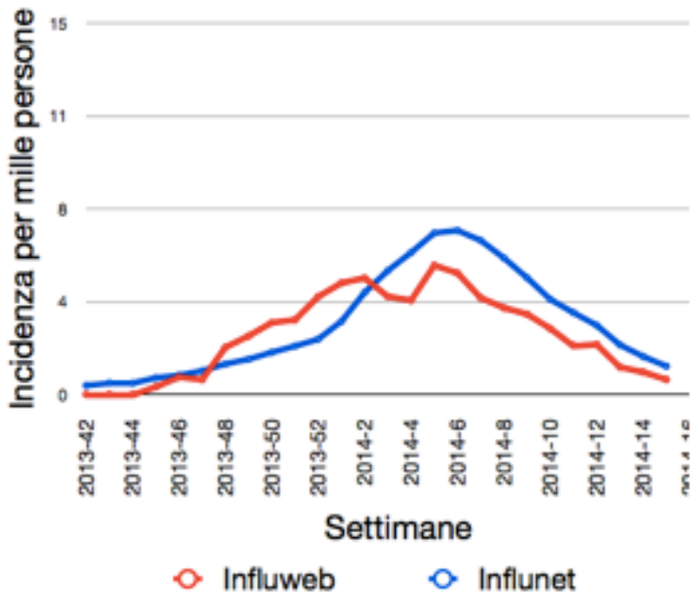
Influenzanet

Influenzanet is a system to monitor the activity of influenza-like ill… …via the internet

INFLUWEB

Home | Notizie | Il Progetto | Cos'è l'influenza | Mappe | Statisti

Incidenza totale della sindrome influenzale rilevata dalle Sorveglianze InfluWeb e Influnet



Ultime notizie

Aggiornamenti sull'influenza per la settimana 15 del 2014

In Italia, l'attività delle sindromi influenzali è definitivamente scesa sotto la soglia epidemica del due per mille, decretando la conclusione della stagione influenzale avvenuta nella tredicesima settimana. Il livello di incidenza in quest'ultima settimana è stato inferiore all'uno per mille. In tutte le regioni italiane è terminato il periodo

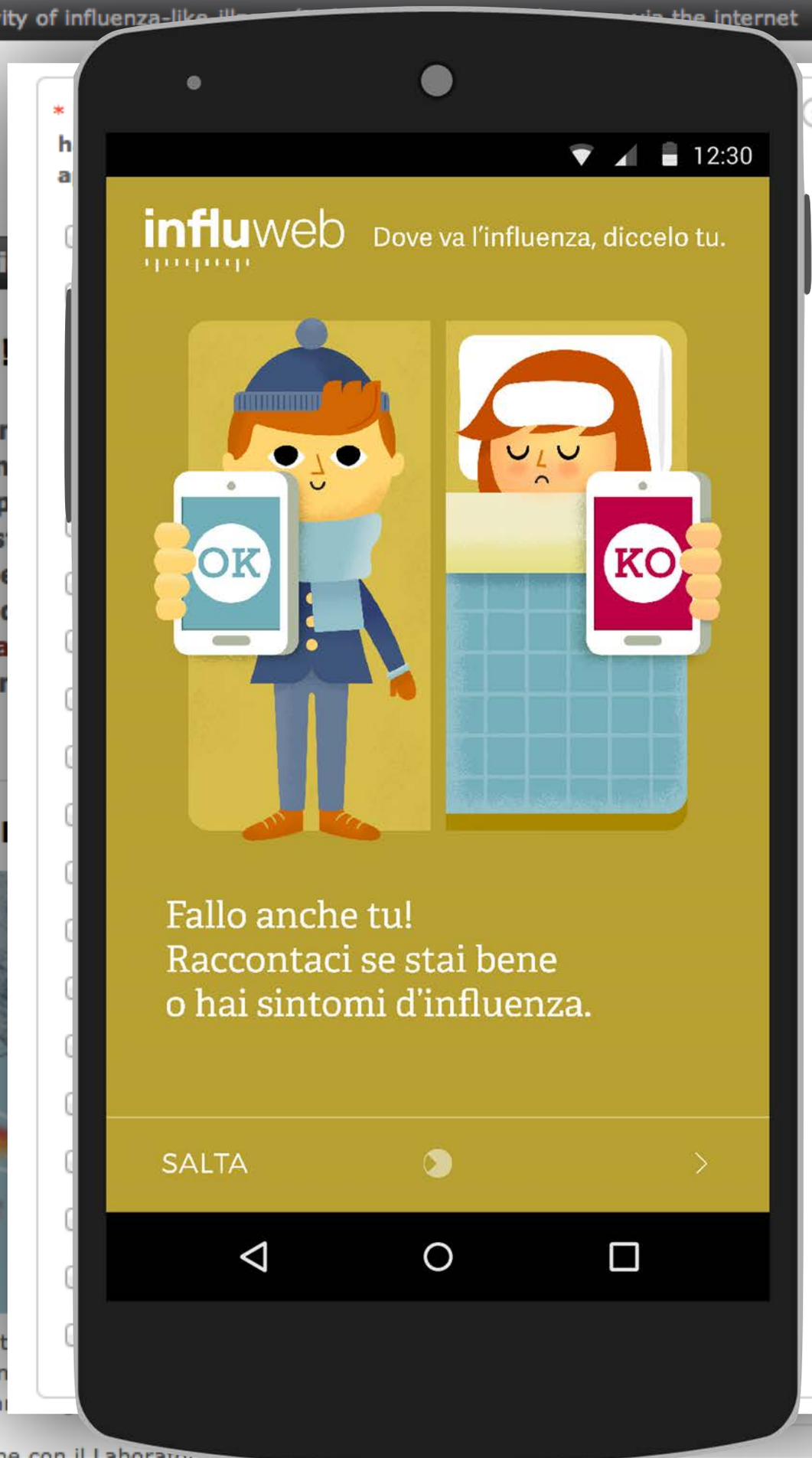
Partecipa anche tu!

Influweb è un progetto scien… informazioni sull'influenza in… il Paese che riportano il prop… settimana durante tutta la s… campagna di raccolta dati pe… puoi registrarti comunque ec… rete europea di sorveglianza… reminder settimanale che ar… raccolta dati.

Il meteo dell'influe…

quotidiana. Ma non sembra essere t… giornale, oltre alle previsioni del ten… casi di influenza ci saranno nelle va…

Il team di Influweb, in collaborazione con il Laborat…



citizen as (flu) sensors

deGroteGriepMeting.nl
B-NL

Influenzanet

www.influensakoll.se
Kartlägger förekomst och spridning av influensa i Sverige

flusurvey

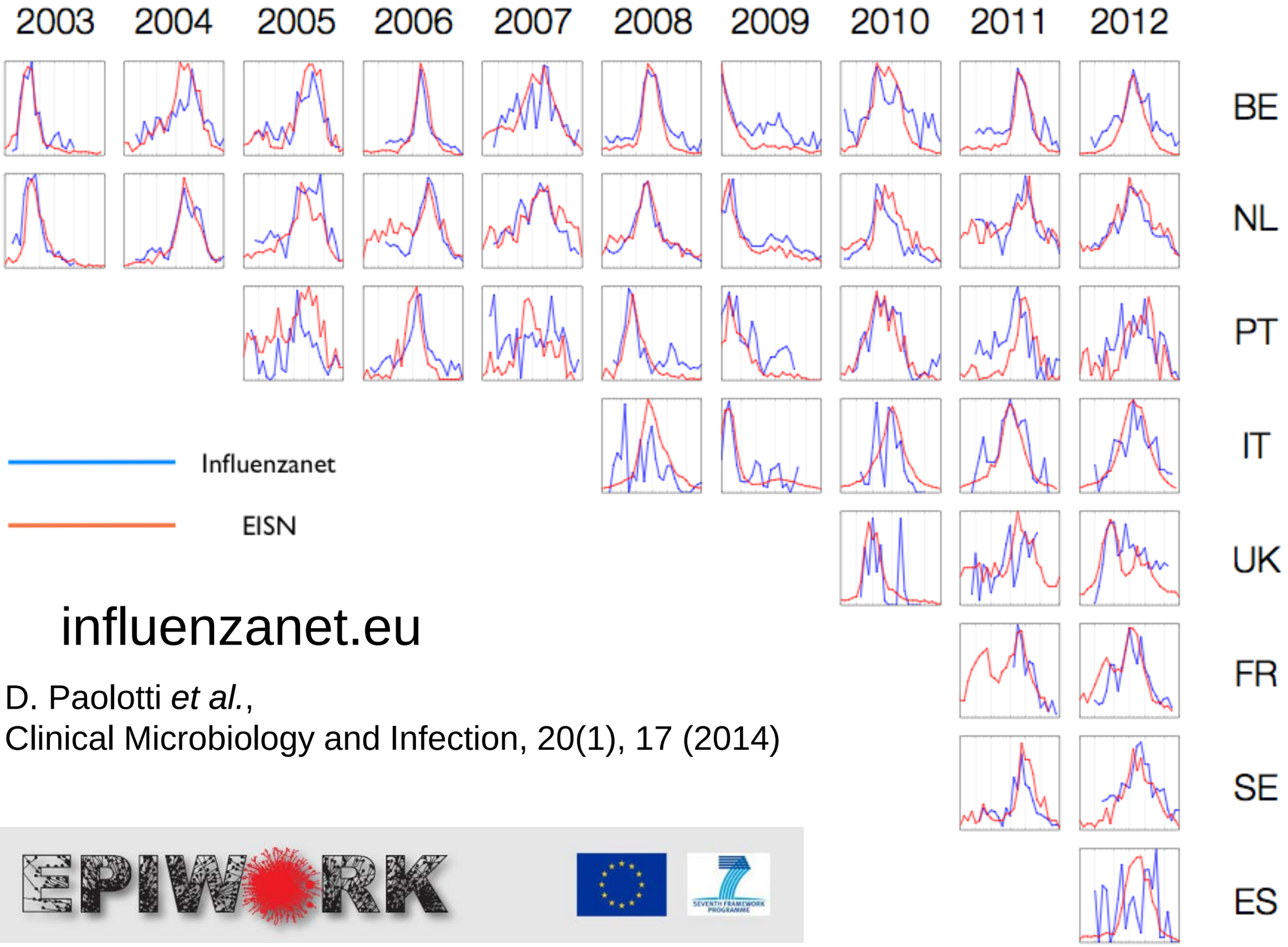
grippenet.fr

gripenet

INFLUWEB

GripeNet.es

EU-wide participatory surveillance



Influenzanet

EISN

influenzanet.eu

D. Paolotti *et al.*,
Clinical Microbiology and Infection, 20(1), 17 (2014)

EPIWORK

Assessing the Online Social Environment for Surveillance of Obesity Prevalence

Rumi Chunara, Lindsay Bouton, John W. Ayers, John S. Brownstein

Published: April 24, 2013 • http://dx.doi.org/10.1371/journal.pone.0061373

R. Chunara et al., PLoS ONE 8(4): e61373 (2013)

Abstract

Background

Understanding the social environmental around obesity has been limited by available data. One promising approach used to bridge similar gaps elsewhere is to use passively generated digital data.

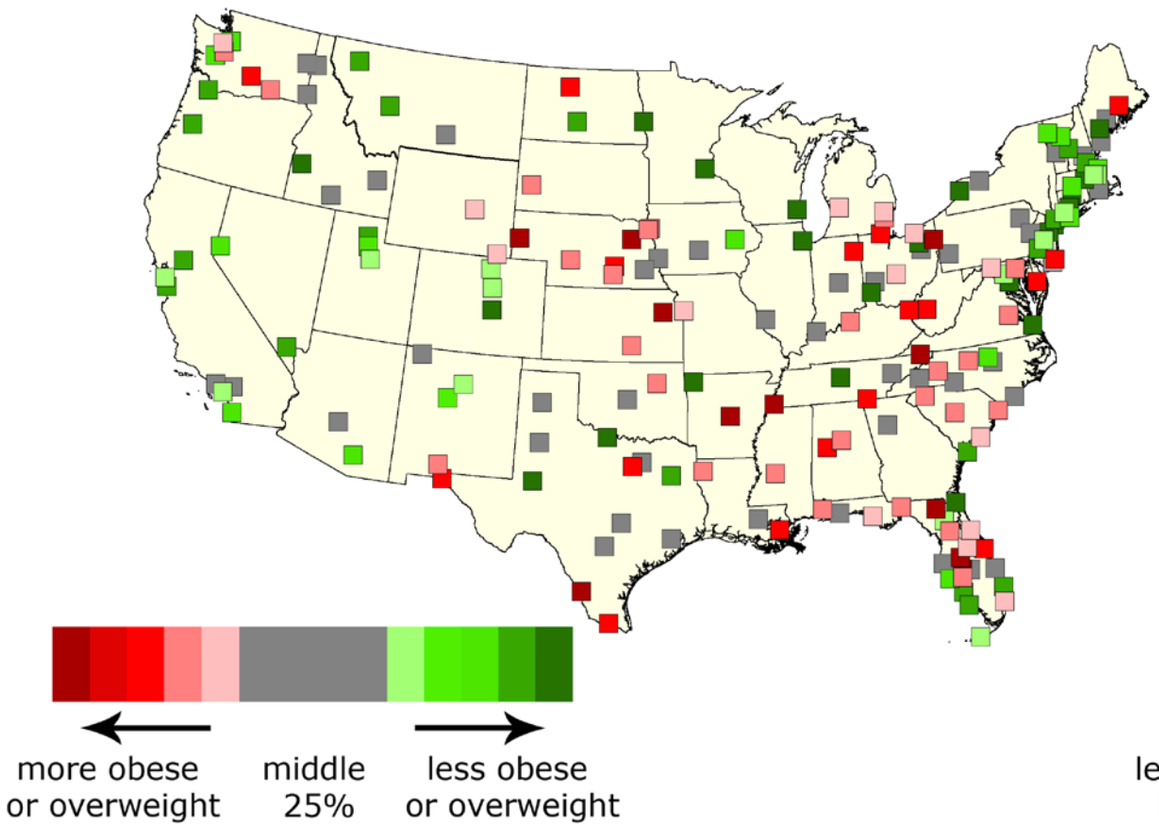
Purpose

This article explores the relationship between online social environment via web-based social networks and population obesity prevalence.

Methods

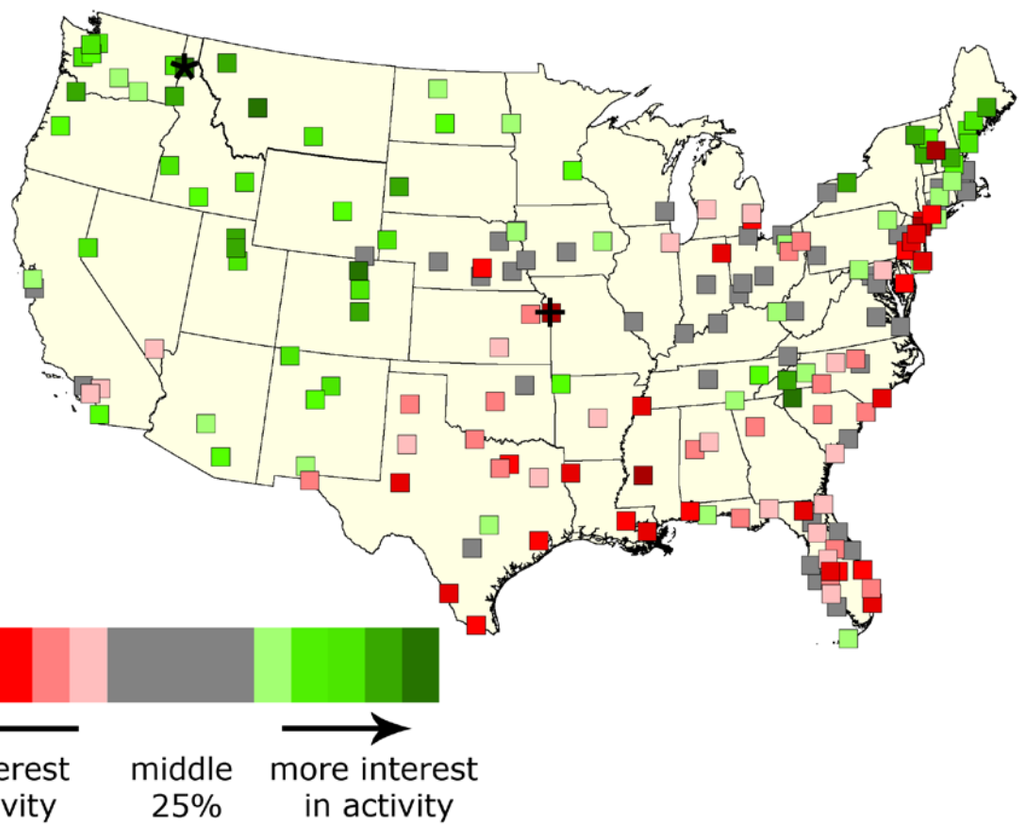
We performed a cross-sectional study using linear regression and cross validation to measure the relationship and predictive performance of user interests on the online social network Facebook to obesity prevalence in metros across the United States of America (USA) and neighborhoods within New York City (NYC). The outcomes, proportion of obese and/or overweight population in USA metros and NYC neighborhoods, were obtained via the Centers for Disease Control and Prevention Behavioral Risk Factor Surveillance and NYC EpiQuery systems. Predictors were geographically specific proportion of users with activity-related and sedentary-related interests on Facebook.

obesity



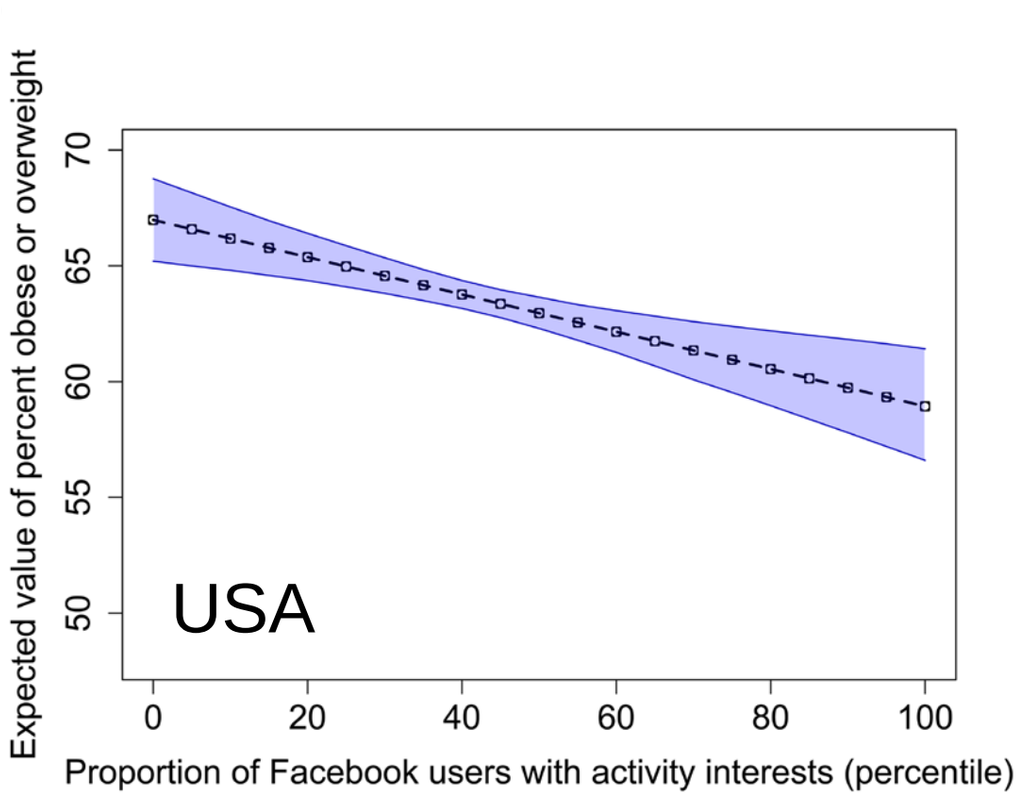
Disease Control and Prevention's Behavioral Risk Factor Surveillance System (BRFSS)
+
Selected Metropolitan/Micropolitan Area Risk Trends (SMART)

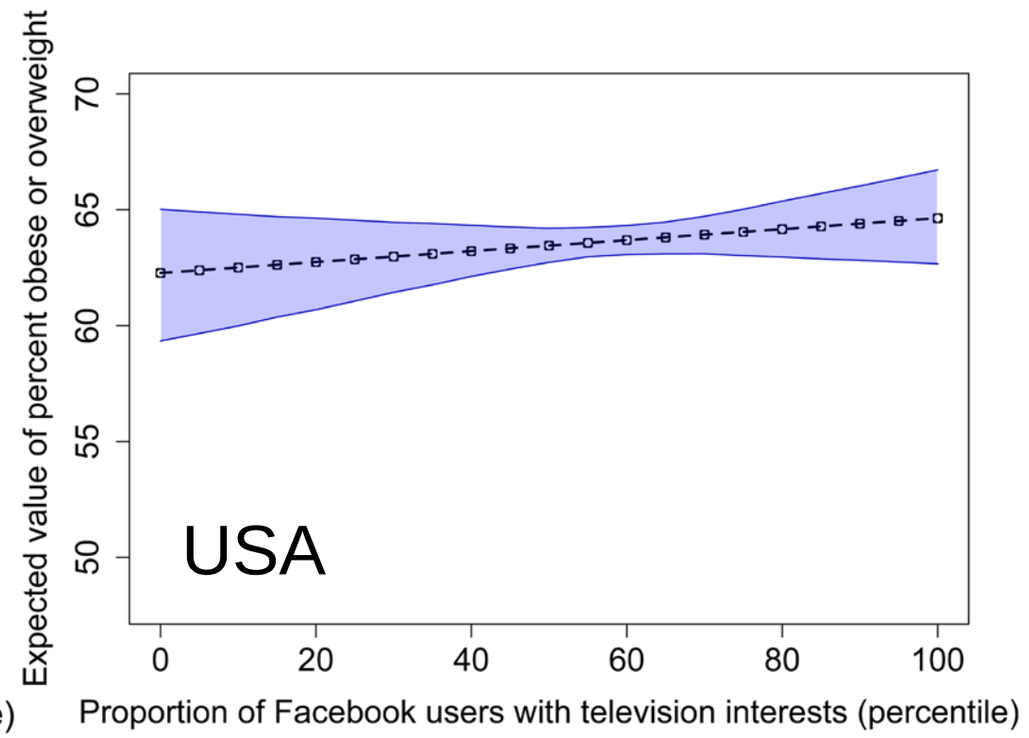
interest in activity

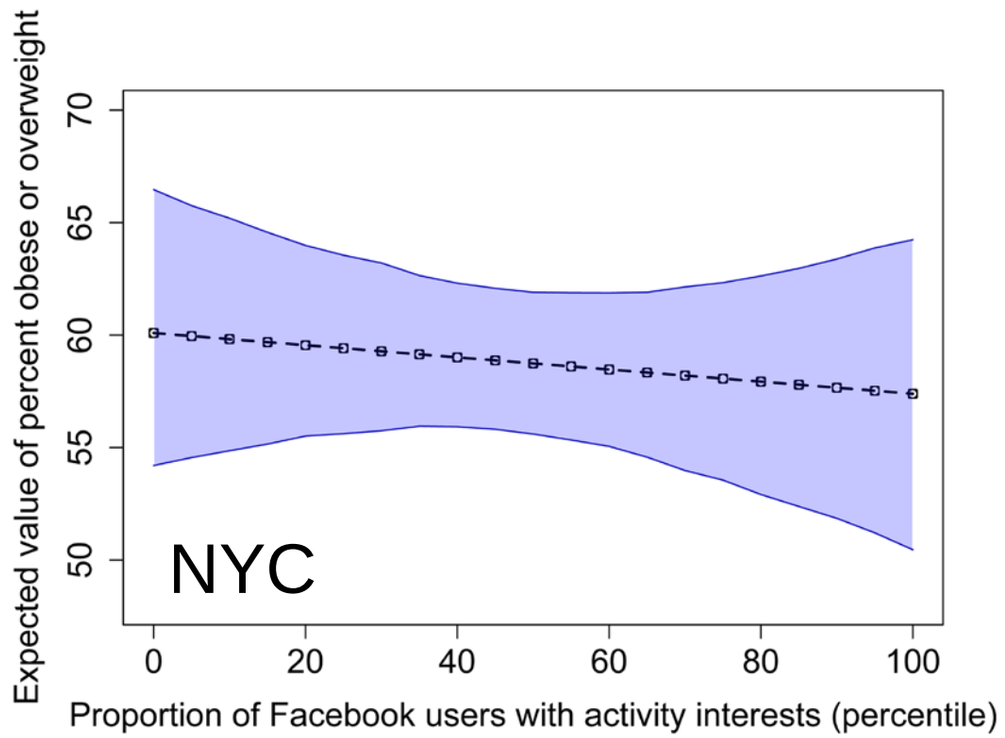


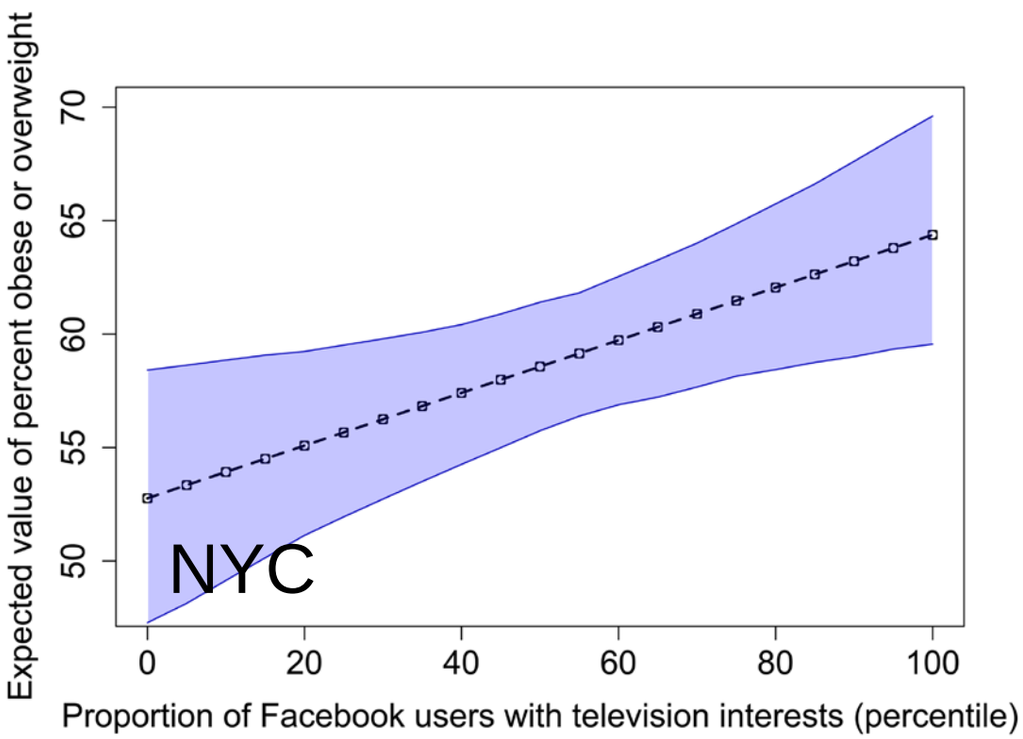
Facebook likes
on pages in the categories
"health and wellness",
"outdoor fitness activities",
"television"

obesity prevalence vs Facebook-based indicators









Psychological SCIENCE
A Journal of the Association for Psychological Science

Psychological Language on Twitter Predicts County-Level Heart Disease Mortality

Johannes C. Eichstaedt[1]

Hansen Andrew Schwartz[1,2]

Margaret L. Kern[1,3]

Gregory Park[1]

Darwin R. Labarthe[4]

Raina M. Merchant[5]

Sneha Jha[2]

Megha Agrawal[2]

Lukasz A. Dziurzynski[1]

Maarten Sap[1]

Christopher Weeg[1]

Emily E. Larson[1]

Lyle H. Ungar[1,2]

Martin E. P. Seligman[1]

[1]Department of Psychology, University of Pennsylvania
[2]Department of Computer and Information Science, University of Pennsylvania
[3]Graduate School of Education, University of Melbourne
[4]School of Medicine, Northwestern University
[5]Department of Emergency Medicine, University of Pennsylvania

« Previous | Next Article »
Table of Contents

This Article

Published online before print January 20, 2015, doi: 10.1177/0956797614557867

Psychological Science **February 2015** vol. 26 no. 2 **159-169**

» Abstract *Free*
Full Text
Full Text (PDF)
Supplemental Material

All Versions of this Article:
» Version of Record - Feb 9, 2015
OnlineFirst Version of Record - Jan 20, 2015
What's this?

Services

- Email this article to a colleague
- Alert me when this article is cited
- Alert me if a correction is posted
- Article Metrics
- Similar articles in this journal
- Similar articles in PubMed
- Download to citation manager
- Request Permissions
- Request Reprints

Citing Articles

- View citing article information

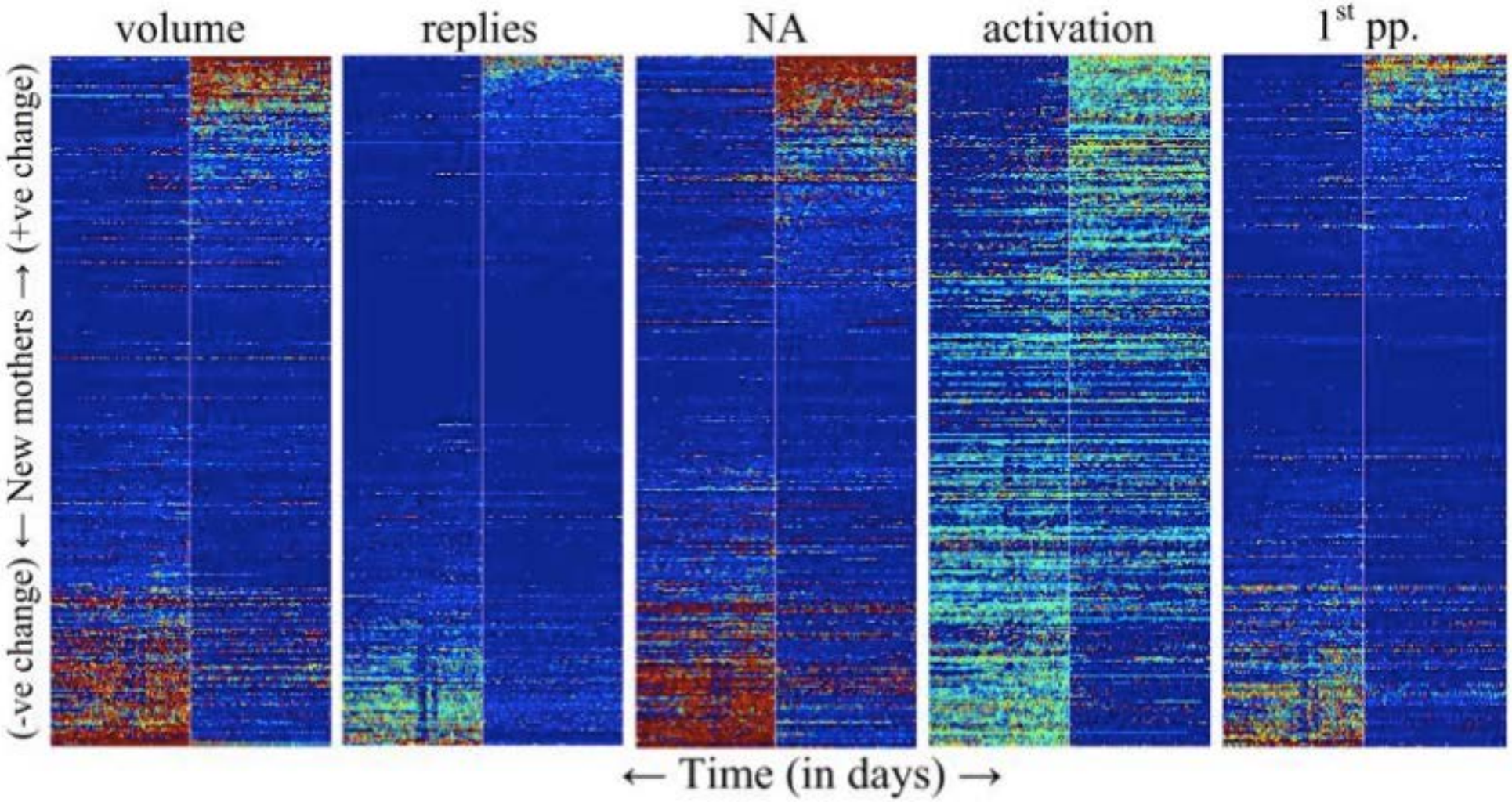
J.C. Eichstaedt *et al.*, Psychological Science 26(2), 159-169 (2015)

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



M. De Choudhury *et al.*, Proc. SIGCHI 2013

ARC symptoms & antihistamine drugs mentions

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

Can Twitter Be a Source of Information on Allergy? Correlation of Pollen Counts with Tweets Reporting Symptoms of Allergic Rhinoconjunctivitis and Names of Antihistamine Drugs

Francesco Gesualdo, Giovanni Stilo, Angelo D'Ambrosio, Emanuela Carloni, Elisabetta Pandolfi, Paola Velardi, Alessandro Fiocchi, Alberto E. Tozzi

Published: July 21, 2015 • DOI: 10.1371/journal.pone.0133706

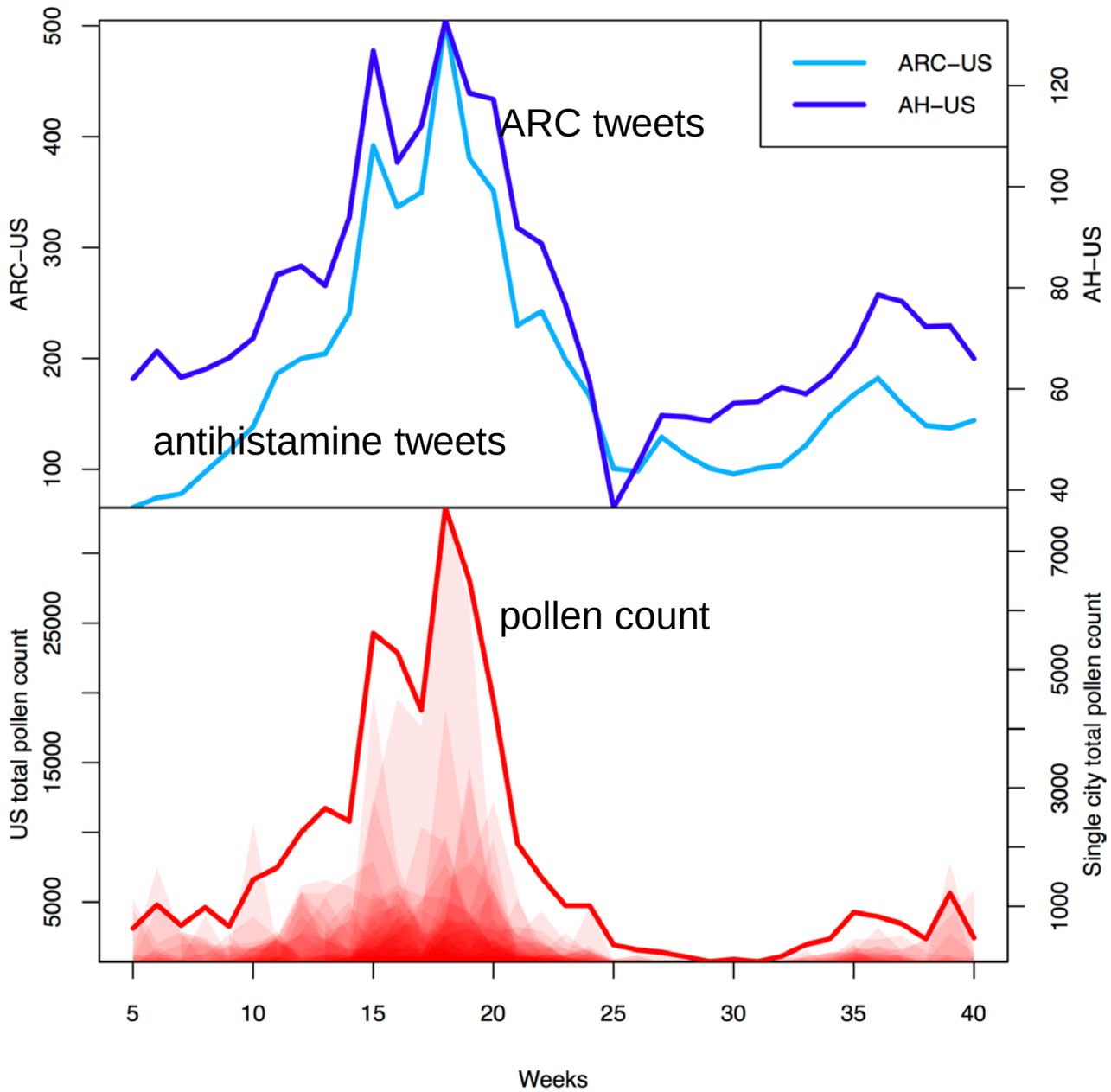
Abstract

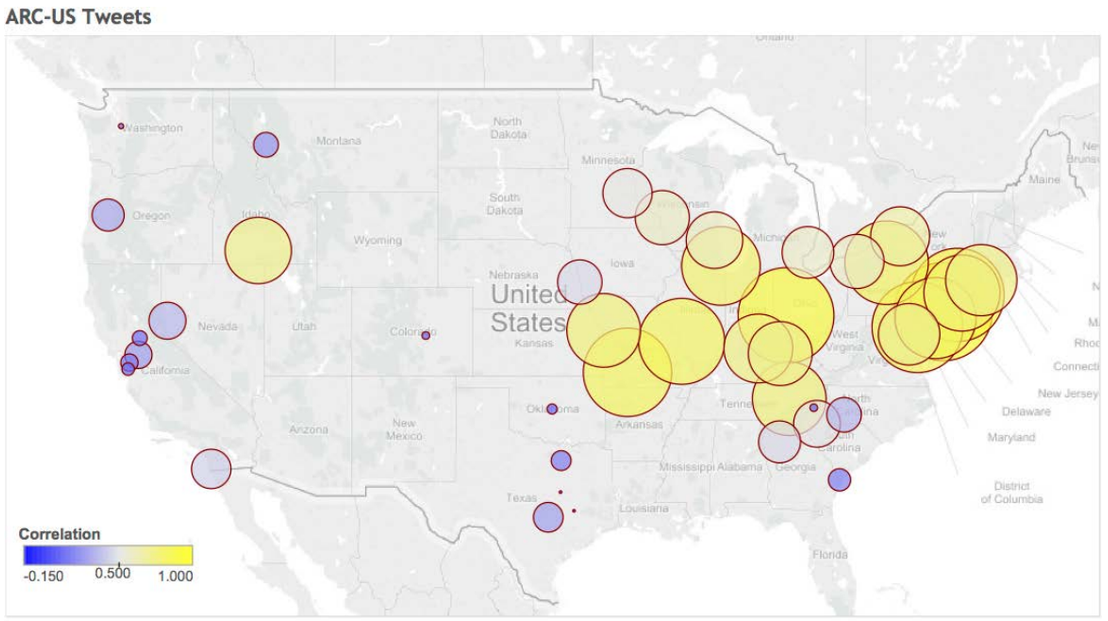
Pollen forecasts are in use everywhere to inform therapeutic decisions for patients with allergic rhinoconjunctivitis (ARC). We exploited data derived from Twitter in order to identify tweets reporting a combination of symptoms consistent with a case definition of ARC and those reporting the name of an antihistamine drug. In order to increase the sensitivity of the system, we applied an algorithm aimed at automatically identifying jargon expressions related to medical terms. We compared weekly Twitter trends with National Allergy Bureau weekly pollen counts derived from US stations, and found a high correlation of the sum of the total pollen counts from each stations with tweets reporting ARC symptoms (Pearson's correlation coefficient: 0.95) and with tweets reporting antihistamine drug names (Pearson's correlation

Subject Areas

Twitter
Pollen
Allergies
Latitude
Allergic diseases

ARC symptoms & antihistamine drugs mentions



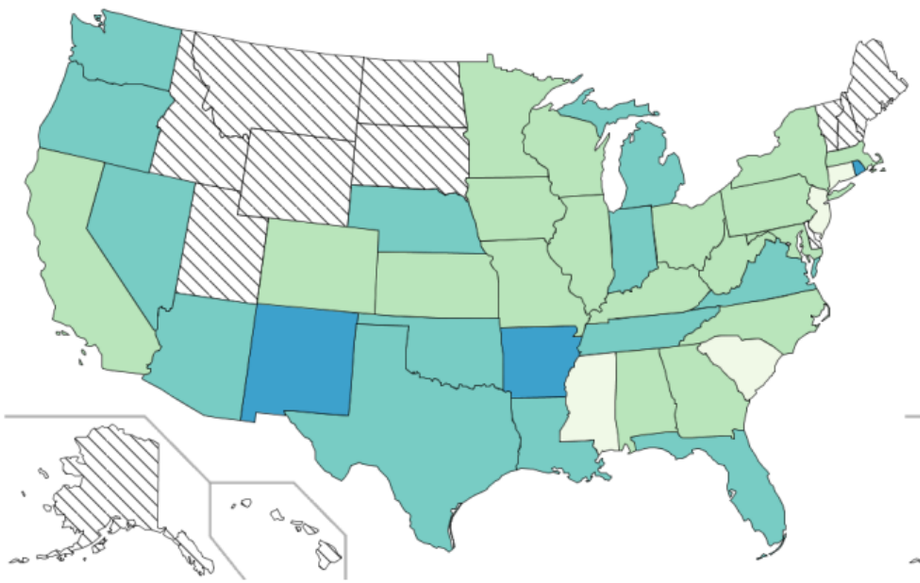


"[…] the findings seem to indicate that the spring pollens (grass and tree pollens) are more associated with antihistamine use than ragweed. If so, the **common wisdom that weeds are the major responsible of allergic rhinitis in the US should be reconsidered**."

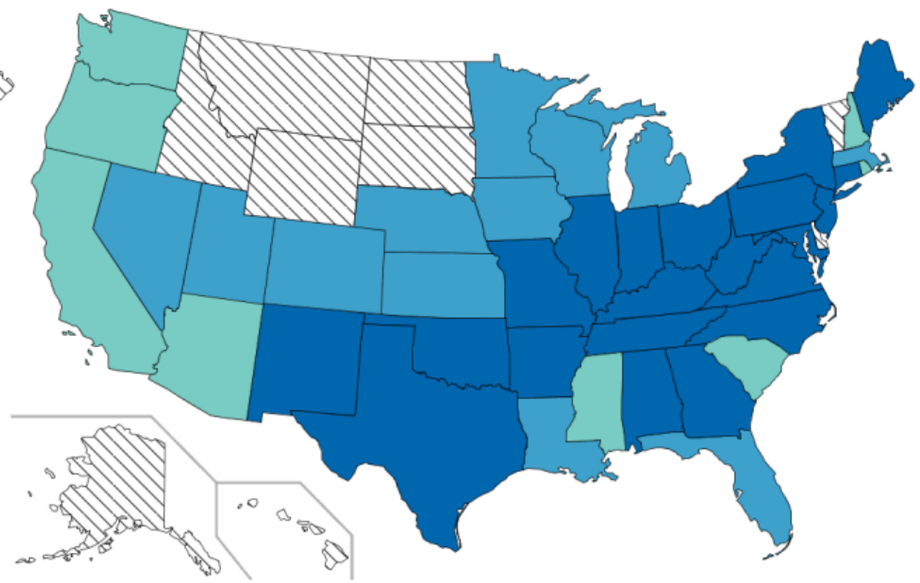
F. Gesualdo *et al.*, PLoS ONE 10(7): e0133706 (2015)

allergy patterns in the US

watery
helsinki
mold
watering
faucet
lolss
sneezes
sneezy
teary
bloom



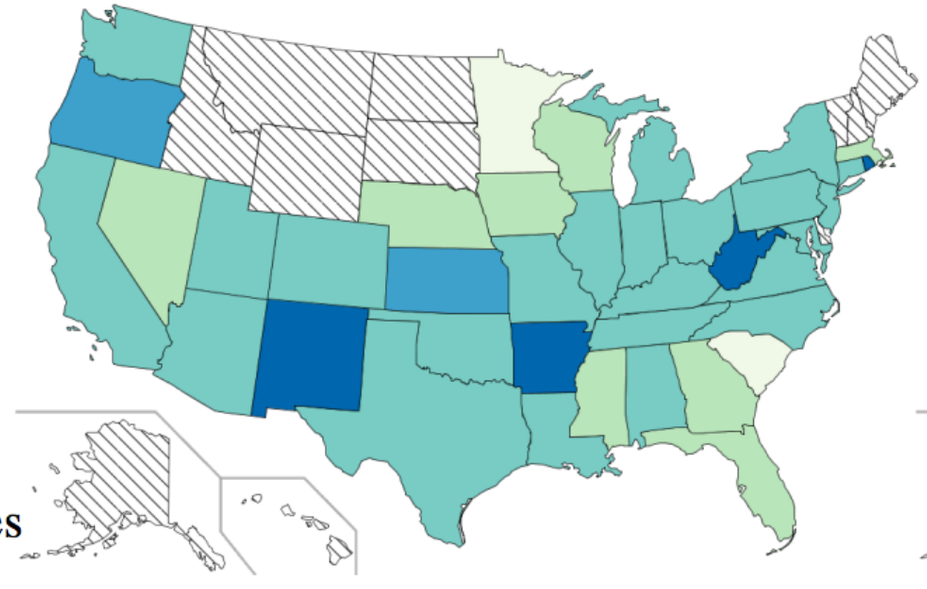
(a) February



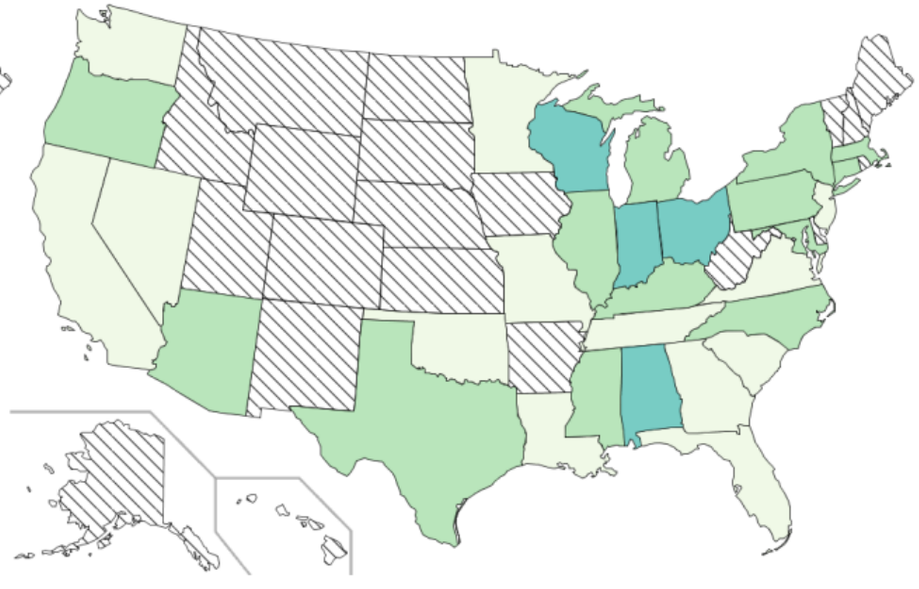
(b) April

pollen
zyrtec
claritin
spring
watering
trees
watery
itching
bloom
grass

hayfever
grass
watering
watery
claritin
humidity
zyrtec
bonkers
mold
antihistamines



(c) June



(d) August

dust
catherine
pinche
buildings
mold
hadd
gato
cessation
meditating
ragweed

M.J. Paul and M. Dredze

You Are What You Tweet: Analyzing Twitter for Public Health

Proc. AAAI 5th Int. Conf. on Weblogs and Social Media (ICWSM11)

Adverse Effects of HIV Drug Treatment

JMIR Publications

MEMBERSHIP SIGN IN

JMIR Public Health and Surveillance

Current Issue Upcoming Issue Top Articles B

Published on 27.07.15 in Vol 1, No 2 (2015): Jul-Dec

This paper is in the following e-collection/theme issue:

Infoveillance, Infodemiology and Digital Disease Surveillance

Medicine 2.0: Social Media, Open, Participatory, Collaborative Medicine Infodemiology and Infoveillance

Pharmacovigilance

Sectio

Article | Cited By (0) | Tweetations (18) | Metrics

Original Paper

Identifying Adverse Effects of HIV Drug Treatment and Associated Sentiments Using Twitter

Cosme Adrover[1], PhD ; Todd Bodnar[1], BSc ; Zhuojie Huang[1], PhD ; Amalio Telenti[2*], MD, PhD ; Marcel Salathé[1*], PhD

[1]Center for Infectious Disease Dynamics, Department of Biology, Penn State University, University Park, PA, United States

[2]J. Craig Venter Institute, La Jolla, CA, United States

*these authors contributed equally

C. Adrover *et al.*, JMIR Public Health Surveill 2015;1(2):e7

Adverse Effects of HIV Drug Treatment

Initial Dataset taken from Sep 2010 - Aug 2013

40 M tweets

containing the following keywords:

Sustiva	Epivir	Norvir	Truvada	HIV treatment
Stocrin	Retrovir	Kaletra	Combivir	HIV drug
Viread	Viramune	Isentress	Kivexa	anti-hiv
FTC	Edurant	Tivicay	Epzicom	triple therapy hiv
Ziagen	Prezista	Atripla	Complera	anti hiv
3TC	Reyataz	Trizivir	Stribild	

Tokenization

If words in tweet do not match keywords, we discard tweet

1.8 M tweets

Removing tweets with "FTC"

Large bias towards references to Federal Trade Commission

.3 M tweets

Removing tweets with "t.co", "bit.ly", "HIV %"

Statistical Sampling
tweets removed based on annotated random samples of tweets

94 k tweets

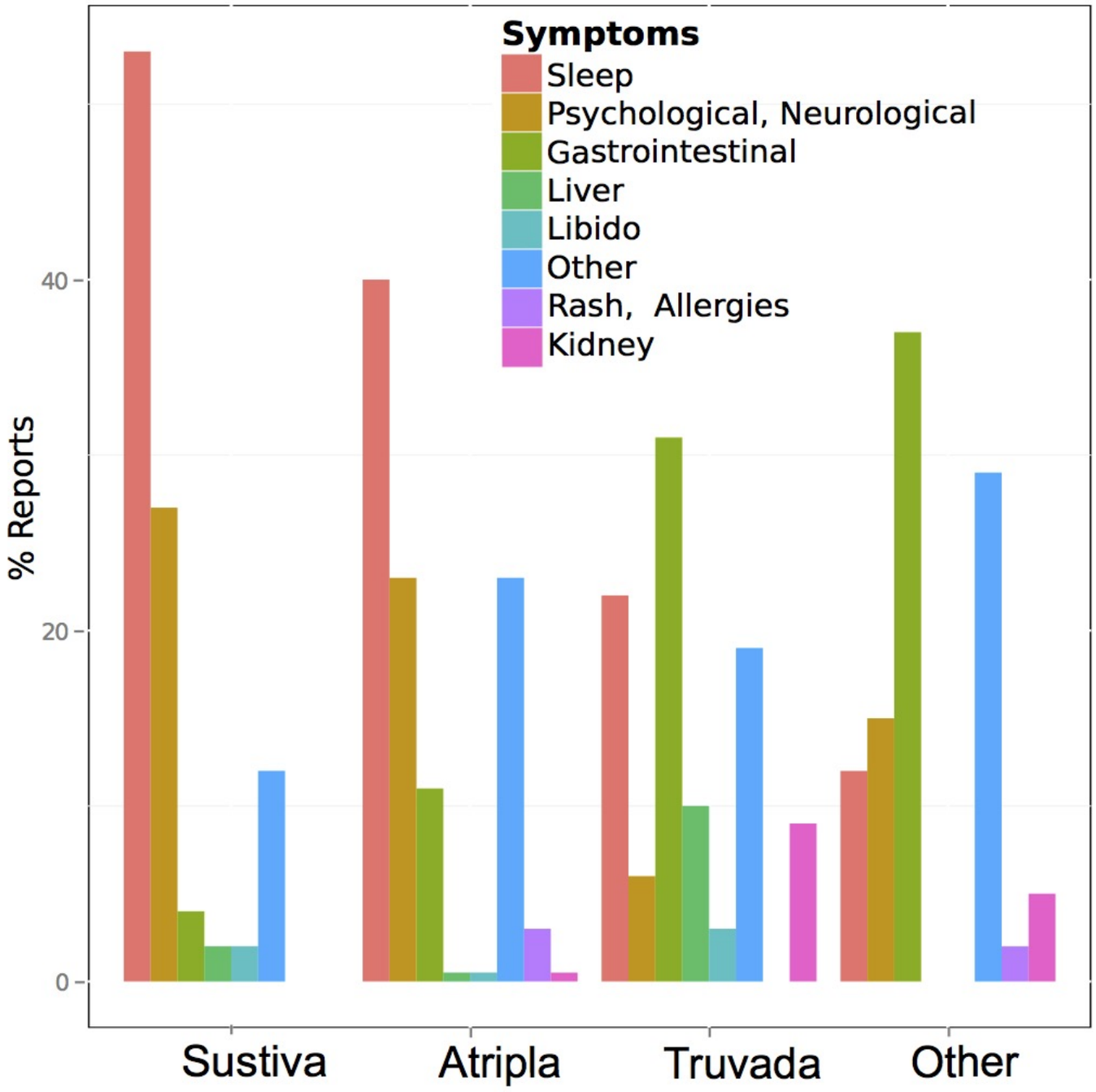
Removing tweets with "http", "news", "buy"

60 k tweets

Removing tweets in other languages

37 k tweets

Adverse Effects of HIV Drug Treatment



"[…] we found a surprisingly large proportion of reports of drug toxicity, a high level of precision on drug-specific effects, and a general negative perception of treatment. We believe that **monitoring of social media will be informative for the field, and broadly applicable to the surveillance of other therapies with long duration of use**, even for those targeted at diseases that are associated with considerable stigma."

C. Adrover *et al.*, JMIR Public Health Surveill 2015;1(2):e7

OPEN ACCESS Freely available online

PLoS COMPUTATIONAL BIOLOGY

Assessing Vaccination Sentiments with Online Social Media: Implications for Infectious Disease Dynamics and Control

Marcel Salathé*, Shashank Khandelwal

Center for Infectious Disease Dynamics, Department of Biology, Penn State University, University Park, Pennsylvania, United States of America

Abstract

There is great interest in the dynamics of health behaviors in social networks and how they affect collective public health outcomes, but measuring population health behaviors over time and space requires substantial resources. Here, we use publicly available data from 101,853 users of online social media collected over a time period of almost six months to measure the spatio-temporal sentiment towards a new vaccine. We validated our approach by identifying a strong correlation between sentiments expressed online and CDC-estimated vaccination rates by region. Analysis of the network of opinionated users showed that information flows more often between users who share the same sentiments - and less often between users who do not share the same sentiments - than expected by chance alone. We also found that most communities are dominated by either positive or negative sentiments towards the novel vaccine. Simulations of infectious disease transmission show that if clusters of negative vaccine sentiments lead to clusters of unprotected individuals, the likelihood of disease outbreaks is greatly increased. Online social media provide unprecedented access to data allowing for inexpensive and efficient tools to identify target areas for intervention efforts and to evaluate their effectiveness.

Citation: Salathé M, Khandelwal S (2011) Assessing Vaccination Sentiments with Online Social Media: Implications for Infectious Disease Dynamics and Control. PLoS Comput Biol 7(10): e1002199. doi:10.1371/journal.pcbi.1002199

Editor: Lauren Ancel Meyers, University of Texas at Austin, United States of America

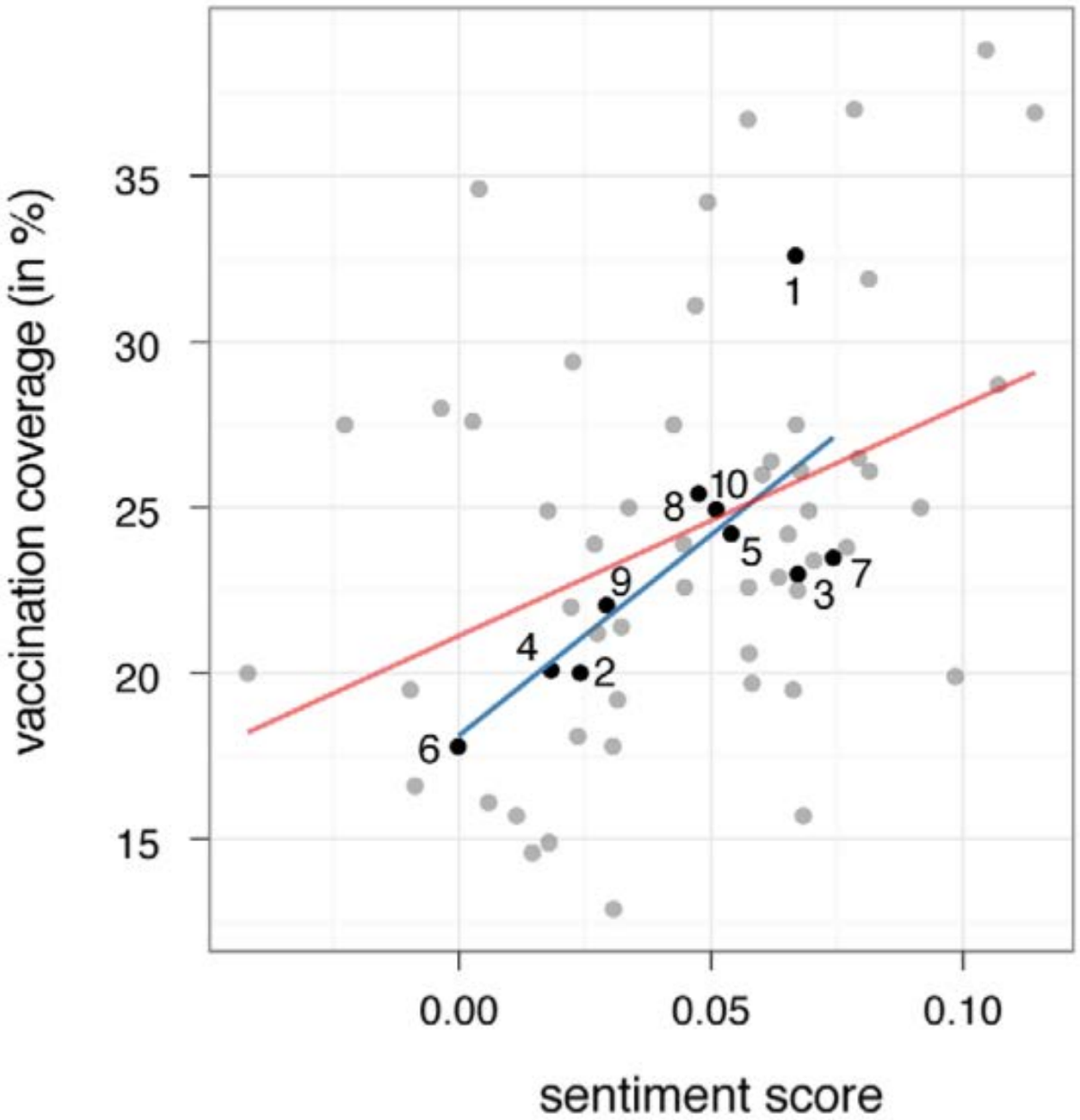
Received May 10, 2011; **Accepted** July 30, 2011; **Published** October 13, 2011

Funding: MS acknowledges funding from Society in Science: the Branco Weiss fellowship. http://www.society-in-science.org/. 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: salathe@psu.edu

Twitter sentiment vs vaccination coverage



M. Salathé, PLoS Comput. Biol. 7(10), e1002199 (2011)

Web-based patient communities

conditions, symptoms, treatments...

Live better, together!™

Making healthcare better for everyone through sharing, support, and research

Join now
(it's free!)

Learn from others
Compare treatments, symptoms and experiences with people like you and take control of your health

Connect with people like you
Share your experience, give and get support to improve your life and the lives of others

Track your health
Chart your health over time and contribute to research that can advance medicine for all

www.patientslikeme.com

350,000 members | 2,500+ conditions | 60+ published research studies | 28 million data points about disease

PatientsLikeMe and the FDA Sign Research Collaboration Agreement

Posted June 15th, 2015 by patientslikeme

WASHINGTON D.C., June 15, 2015–PatientsLikeMe and the U.S. Food and Drug Administration (FDA) have signed a research collaboration agreement to determine how patient-reported data can give new insights into drug safety. Under the collaboration, PatientsLikeMe and the FDA will systematically explore the potential of patient-generated data to inform regulatory review activities related to risk assessment and risk management. The announcement was made at the start of the Drug Information Association's (DIA) annual meeting in Washington D.C.

PatientsLikeMe Co-Founder and President Ben Heywood said the agreement is an unprecedented step toward enhancing post-market surveillance and informing regulatory science. "Most clinical trials only represent the experience of several hundred or at most several thousand patients, making it impossible to anticipate all the potential side effects of drugs in the real world. Patient-generated data give a more complete picture about a drug's safety by providing a window into patients' lives and healthcare experiences over time. We're very encouraged by the FDA's action to evaluate newer sources of data to help identify benefits and risks earlier."

"[…] The company's drug safety initiatives began in 2008 with a pilot program that allowed patients living with Multiple Sclerosis (MS) to report adverse events directly to the FDA. One year later the company launched the first drug safety platform on social media, enabling industry partners to meet their regulatory obligations. In all, PatientsLikeMe has collected more than 110,000 adverse event reports on 1,000 different medications, data that the FDA will now be able to access and analyze as a supplement to traditional sources […]"

thanks

Ciro Cattuto
ISI Foundation
Torino, Italy & New York City

European Medicines Agency
London, 19 September 2016

INSTITUTE
FOR SCIENTIFIC INTERCHANGE
FOUNDATION