

Twitter & Disease Surveillance:

- a) crowdsourcing disease surveillance
- b) health behavior assessment

Email: salathe@psu.edu

Twitter: [@marcelalathe](https://twitter.com/marcelalathe), [@salathegroup](https://twitter.com/salathegroup)

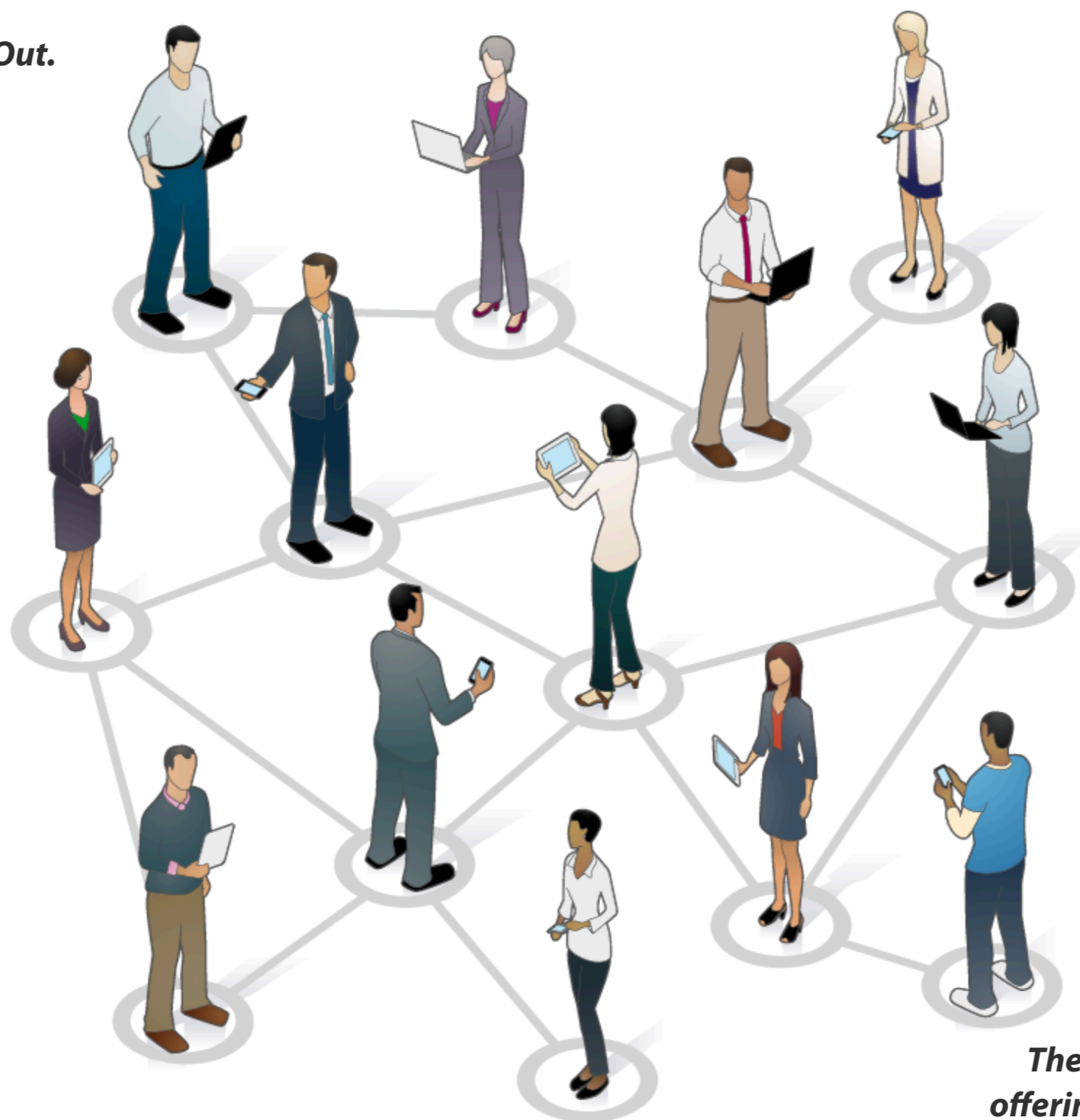
Center for Infectious Disease Dynamics (CIDD)

Department of Biology

Department of Computer Sciences and Engineering



**What Can You Do To Resist The
U.S. H1N1 "Vaccination"
Program? Help Get Word Out.
The H1N1 "Vaccine" Is
DIRTY.DontGetIt.**

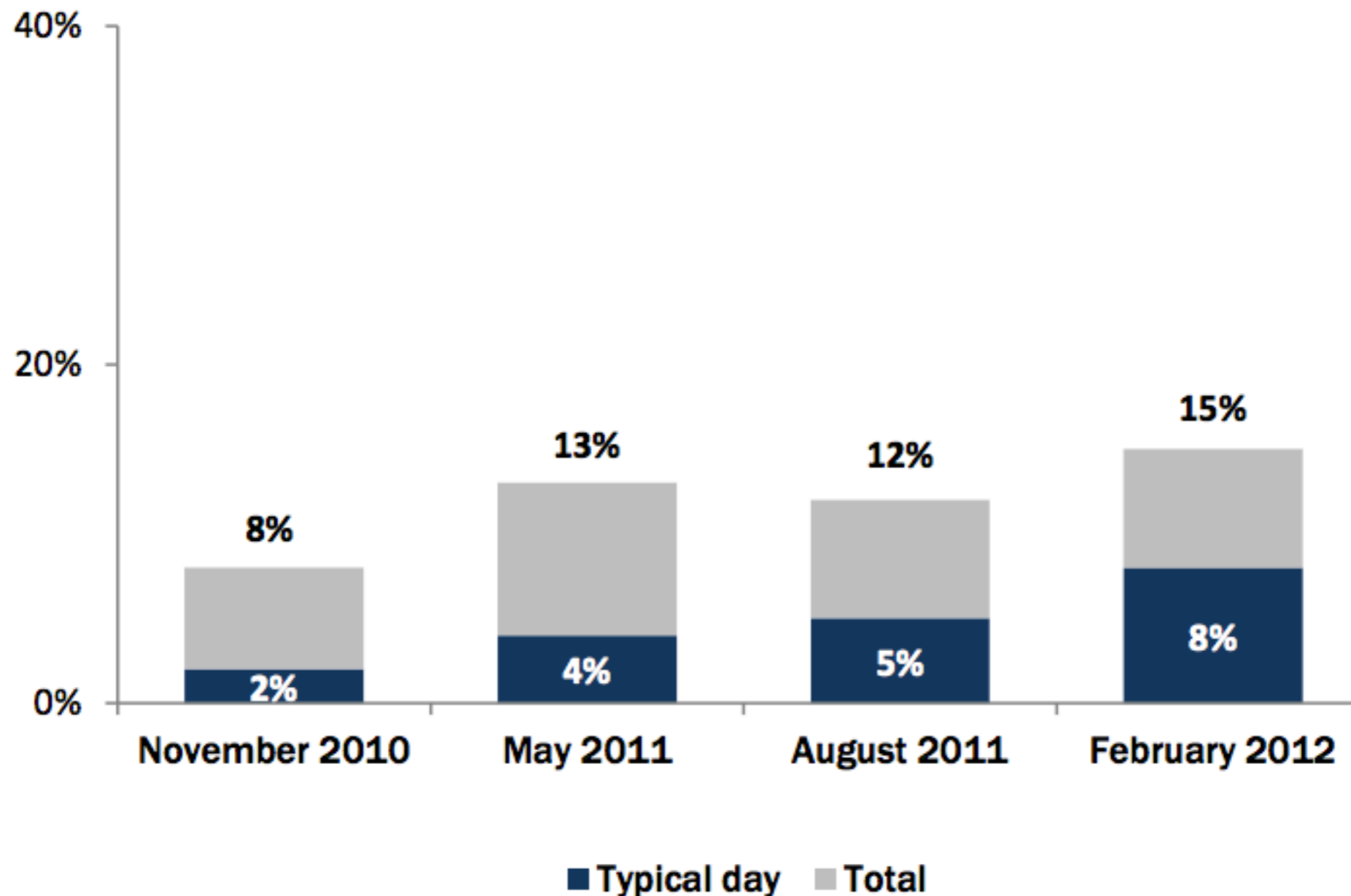


***off to get swine flu
vaccinated before work***

***The Health Department will be
offering the seasonal flu vaccine for
children 6 months - 19 yrs. of age
starting on Monday, Nov. 16.***

Twitter usage over time

% of internet users who use Twitter



Source: Pew Research Center's Internet & American Life Project Winter 2012 Tracking Survey, January 20-February 19, 2012. N=2,253 adults age 18 and older, including 901 cell phone interviews. Interviews conducted in English and Spanish. Margin of error is +/-2.7 percentage points for internet users (n=1,729).

Who uses Twitter?

% of internet users within each group who use Twitter

All adult internet users (n=1729)	15%
Men (n=804)	14
Women (n=925)	15
Age	
18-29 (n=316)	26**
30-49 (n=532)	14
50-64 (n=521)	9
65+ (n=320)	4
Race/ethnicity	
White, Non-Hispanic (n=1229)	12
Black, Non-Hispanic (n=172)	28**
Hispanic (n=184)	14
Annual household income	
Less than \$30,000/yr (n=390)	19
\$30,000-\$49,999 (n=290)	12
\$50,000-\$74,999 (n=250)	14
\$75,000+ (n=523)	17
Education level	
No high school diploma ² (n=108)	22
High school grad (n=465)	12
Some College (n=447)	14
College + (n=698)	17
Geographic location	
Urban (n=520)	19**
Suburban (n=842)	14**
Rural (n=280)	8

Source: Pew Research Center's Internet & American Life Project Winter 2012 Tracking Survey, January 20-February 19, 2012. N=2,253 adults age 18 and older, including 901 cell phone interviews. Interviews conducted in English and Spanish. The margin of error is +/-2.7 percentage points for internet users. **Represents significant difference compared with all other rows in group.

Twitter adoption by age, 2010-2012

% of internet users in each group who use Twitter

	November 2010	May 2011	February 2012
All adults	8%	13%	15%
18-24	16	18	31
25-34	9	19	17
35-44	8	14	16
45-54	7	9	9
55-64	4	8	9
65+	4	6	4

Sources: Pew Research Center's Internet & American Life Project tracking surveys. 2012 data based on January 20-February 19, 2012 Tracking Survey. N=2,253 adults age 18 and older, including 901 cell phone interviews, margin of error is +/-2.7 percentage points based on internet users (n=1729).

“Typical day” Twitter use by age, 2010-2012

% of internet users in each group who use Twitter on a typical day

	November 2010	May 2011	February 2012
All adults	2%	4%	8%
18-24	4	9	20
25-34	5	5	11
35-44	2	6	9
45-54	2	3	3
55-64	1	2	4
65+	<1	<1	1

Sources: Pew Research Center’s Internet & American Life Project tracking surveys. 2012 data based on January 20-February 19, 2012 Tracking Survey. N=2,253 adults age 18 and older, including 901 cell phone interviews, margin of error is +/-2.7 percentage points based on internet users (n=1729).

05 Jun 2012

Tweet counts

Software	Number of Tweets
Twitter Locations Stream	1148401
Twitter ILI Keywords Stream	496434

Tweeter counts

Software	Number of Tweeters
Friends Collector	5559

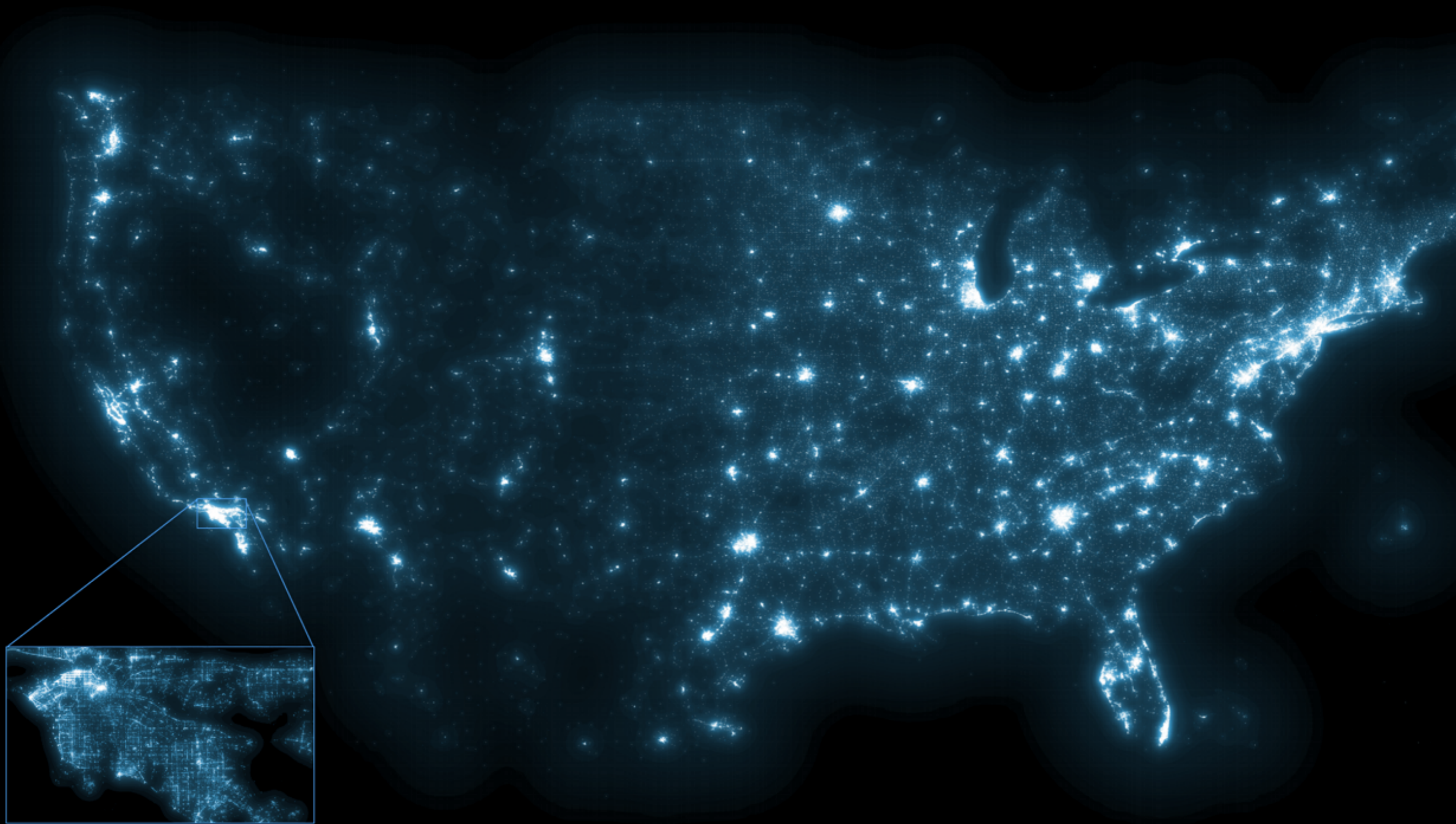
Total Tweets Over All Time

Twitter Locations Stream	429,124,075
Twitter ILI Keywords Stream	164,519,170
Total	593,643,245

Machine Information

Disk Space

meme	80Gi / 298Gi (27%)
machinelearning	86Gi / 298Gi (29%)
datamining	206Gi / 298Gi (70%)



crowdbreaks.com



Select the geographic area of interest, and find out which diseases are trending.

crowdbreaks

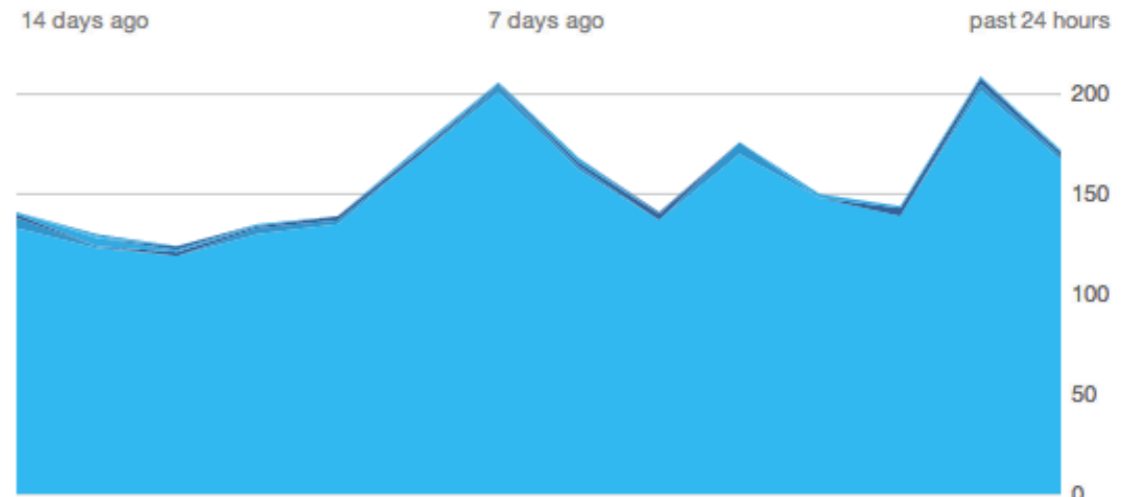
Disease surveillance, #crowdsourced. **Powered by you.**

Geographic Map [Select US state\(s\)](#)



▼ 1. Common cold

change prev. day: **-17.70%** | change prev. week: **+10.69%**



Keywords (counts last day):

All keywords (172)

cold (167), uri (2), runny nose (2), colds (0), nasal congestion (0), pharyngitis (1), common cold (0)

Help us make crowdbreaks even better!

With a single click, you can help us improve our disease detection algorithm. Simply keep answering the questions below. [\[How does this work?\]](#)

► 2. Sexually transmitted diseases

change prev. day: **-11.54%** | change prev. week: **+11.41%**


Is the following message about the common cold?

Really wish these bad dreams would just go away!! Sick of waking up in a cold sweat and not being able to breathe!!

Yes, it is.

No, it isn't.

I'm not sure.

Message content from  [Twitter](#)


Is the following message about the common cold?

Ughhhh I wish I didn't have this cold.

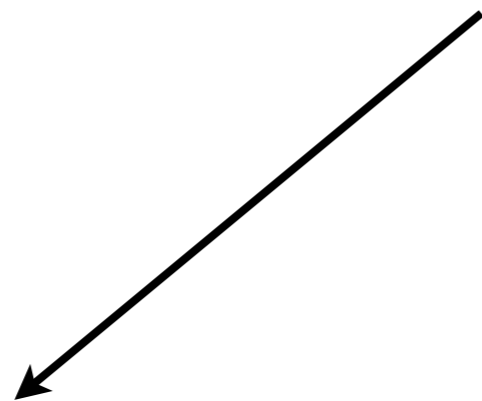
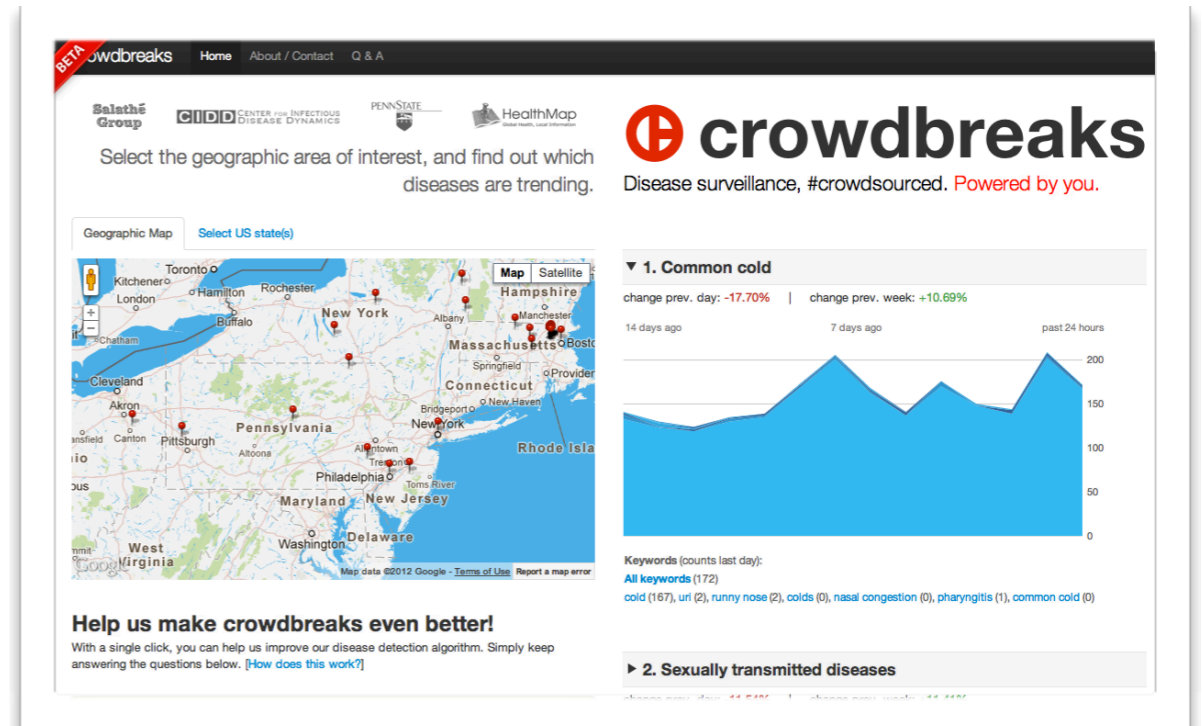
Yes, it is.

No, it isn't.

I'm not sure.

Message content from  [Twitter](#)

Crowd data



Crowd feedback
(incl. experts)



Machine learning
(even that could be
crowdsourced,
e.g. kaggle.com)



We're making data science a sport.™

Participate in competitions

Kaggle is an arena where you can match your data science skills against a global cadre of experts in statistics, mathematics, and machine learning. Whether you're a world-class algorithm wizard competing for prize money or a novice looking to learn from the best, here's your chance to jump in and geek out, for fame, fortune, or fun.

[Sign up as a competitor](#)

[\(Need convincing?\)](#)

Create a competition

Kaggle is a platform for data prediction competitions that allows organizations to post their data and have it scrutinized by the world's best data scientists. In exchange for a prize, winning competitors provide the algorithms that beat all other methods of solving a data crunching problem. Most data problems can be framed as a competition.

[See how it works](#)

Recruitment Competitions

[Browse all »](#)



Facebook Recruiting Competition

JOB

Data Scientist at **Facebook**
Multiple

 Ends 34 days

 25 teams

 Jobs



[View available positions »](#)

Featured Competitions

[Browse all »](#)



Heritage Health Prize

Identify patients who will be admitted to a hospital within the next year, using historical claims data.

 Ends 10 months

 1086 teams

 \$3 million



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

Copyright: © 2011 Salathé, Khandelwal. 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: 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.

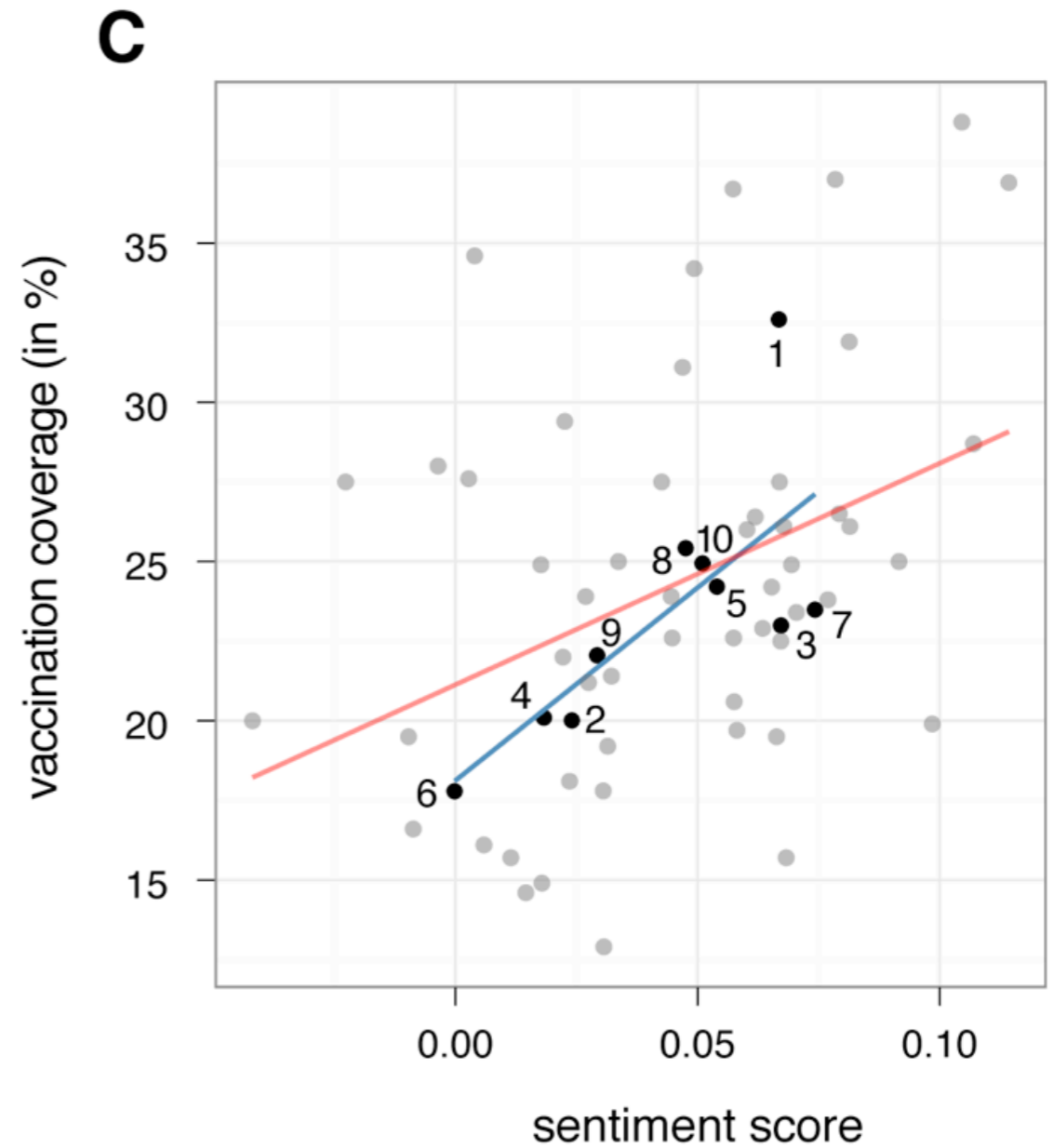
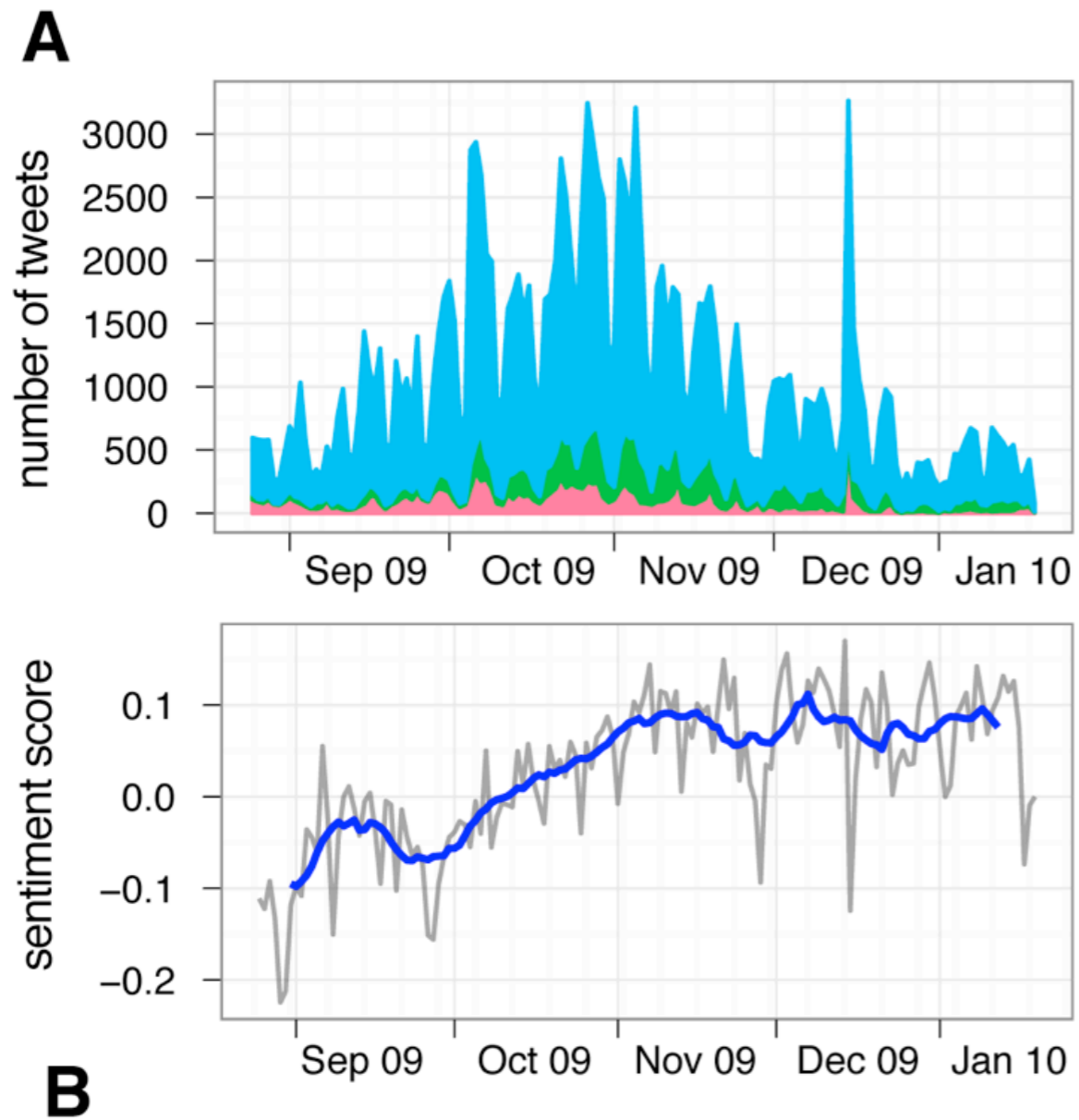
Competing Interests: The authors have declared that no competing interests exist.

* E-mail: salathe@psu.edu

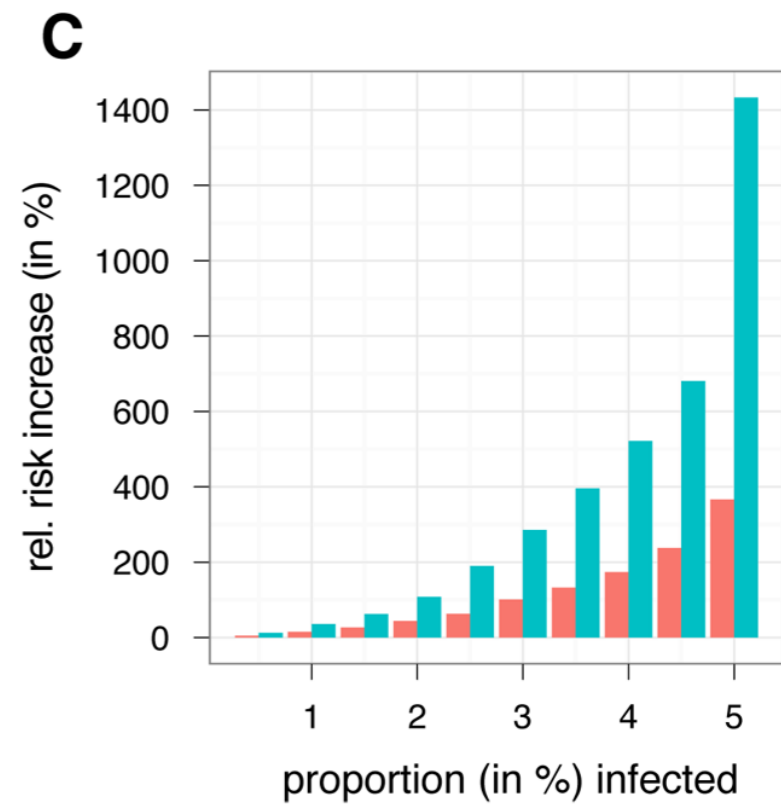
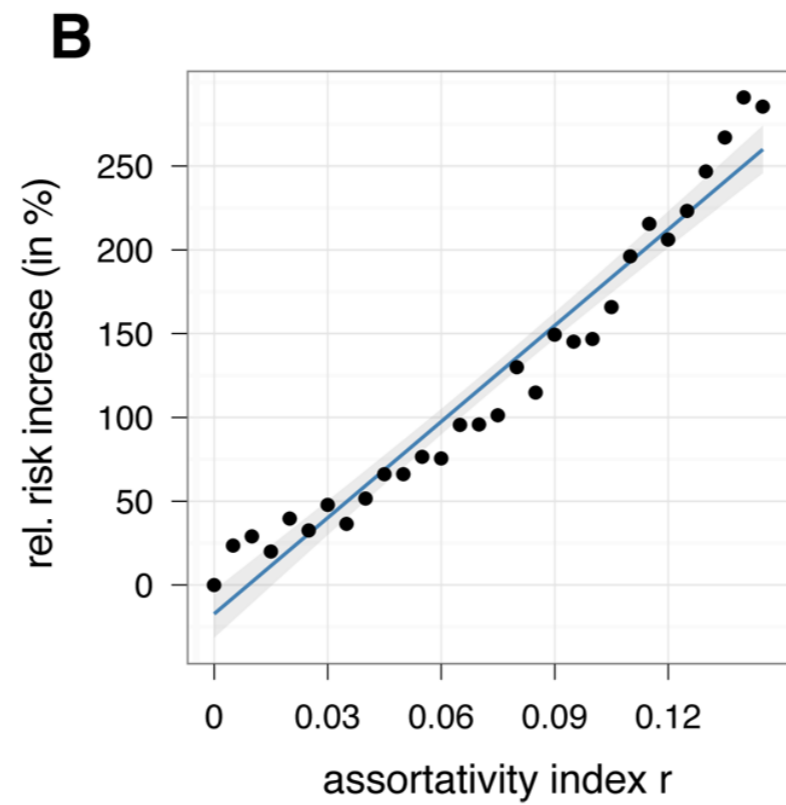
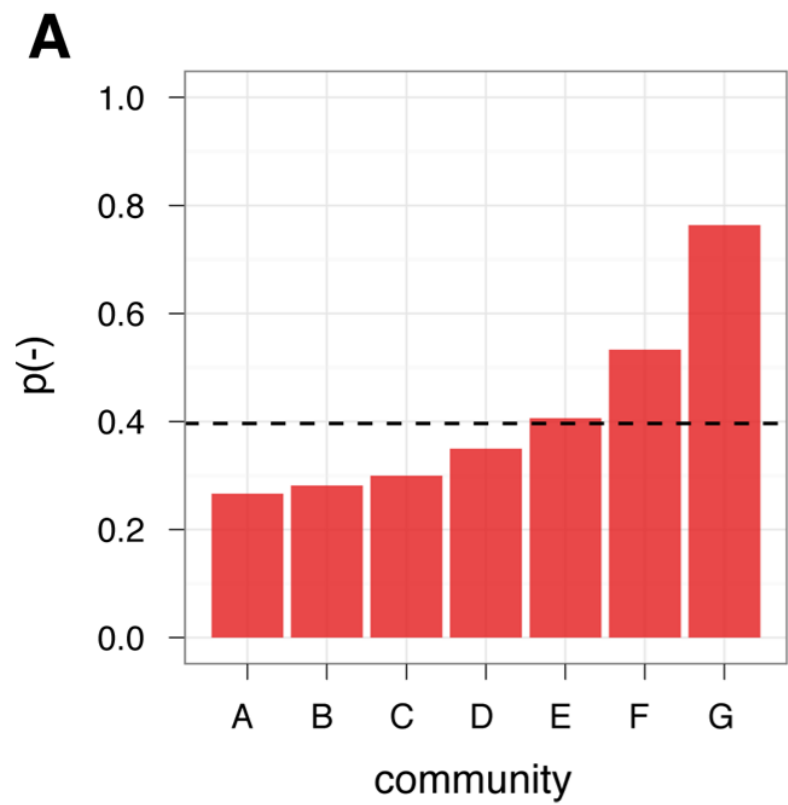
Introduction

Outbreaks of vaccine preventable diseases are a major public health issue. Outbreaks are more likely to occur if either overall vaccination rates decline [1], or if communities with very low vaccination rates increase in frequency or size [2]. An individual

time, pandemic influenza A(H1N1) was spreading nationwide but a vaccine became widely available only very late in the year. We collected practically all publicly available text messages on Twitter (so called “tweets”) containing English keywords relating to vaccination as well as location information provided by the



(~10% Tweets were assessed by humans, rest by machine learning algorithms)



Thank you !

www.salathegroup.com

