



CS 6474 Social Computing: News, Trends and Forecasting

Munmun De Choudhury

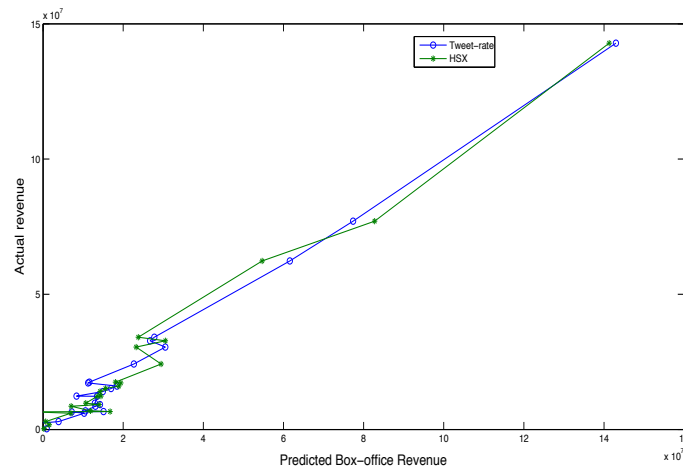
munmund@gatech.edu

Week 13 | November 16, 2016

Predicting the Future With Social Media

Summary

- The article examines if social media can be used to predict real-world outcomes
- One of the earliest “predicting the present” papers using social media
- Method – simple tweet volume model outperformed market predictors of box office revenues
 - Compared with an HDX model – Hollywood Stock Exchange
- Potential shown how Twitter sentiment may be tapped



Characterizing Debate Performance via Aggregated Twitter Sentiment

Summary

- The article develops an analytical methodology and visual representations to help understand temporal dynamics of event sentiment as they unfold over time on social media
 - Particularly analyze the first Presidential debates of the 2008 US Presidential elections
- Help a journalist or public affairs people
- Data – authors queried the Twitter Search API for #current, #debate08 and #tweetdebate
 - 1,820 tweets from 664 people during the 97-minute debate and 1,418 tweets from 762 people in the 53 minutes following the debate
- Identify controversial topics and how sentiment changed for both candidate on a moment to moment basis

What is Twitter, a Social
Network or a News Media?

Summary

- Examine the potential of Twitter as a medium of information sharing
- Data was all of Twitter – crawl was made in 2009 – 41.7 million user profiles, 1.47 billion social relations, 4, 262 trending topics, and 106 million tweets
- Findings:
 - In its follower-following topology analysis authors found a non-power-law follower distribution, a short effective diameter, and low reciprocity
 - Retweet popularity based influence is different from follower count based influence
 - 85% trending topics are headline news
 - Quick diffusion of retweets occur, independent of the follower count of the originator

Before you read the paper, did you really agree that Twitter is a news media?

Current Issue > vol. 107 no. 41 > Sharad Goel, 17486–17490, doi: 10.1073/pnas.1005962107



Predicting consumer behavior with Web search

Sharad Goel¹, Jake M. Hofman¹, Sébastien Lahaie¹, David M. Pennock¹, and Duncan J. Watts¹

Author Affiliations

Edited* by Simon A. Levin, Princeton University, Princeton, NJ, and approved August 10, 2010 (received for review April 29, 2010)

Abstract Full Text Authors & Info Figures Metrics

Abstract

Recent work has demonstrated that Web search volume can “predict the present,” meaning that it can be used to accurately track outcomes such as unemployment levels, auto and home sales, and disease prevalence in near real time. Here we show that what consumers are searching for online can also predict their collective future behavior days or even weeks in advance. Specifically we use search query volume to forecast the opening weekend box-office revenue for feature films, first-month sales of video games, and the rank of songs on the Billboard Hot 100 chart, finding in all cases that search counts are highly predictive of future outcomes. We also find that search counts generally boost the performance of baseline models fit on other publicly available data, where the boost varies from modest to dramatic, depending on the application in question. Finally, we reexamine previous work on tracking flu trends and show that, perhaps surprisingly, the utility of search data relative to a simple autoregressive model is modest. We conclude that in the absence of other data sources, or where small improvements in predictive performance are material, search queries provide a useful guide to the near future.

culture | predictions

This Issue



October 12, 2010
vol. 107 no. 41
Masthead (PDF)
Table of Contents

◀ PREVIOUS ARTICLE

NEXT ARTICLE ▶

Don't Miss



PNAS Full-Text iOS App
Download the app for free from iTunes today!

Article Tools

- Article Alerts
- Export Citation
- Save for Later
- Request Permission

Share



Published online before print
September 27, 2010, doi:



"I Wanted to Predict Elections with Twitter and all I got was this Lousy Paper" Balanced Survey on Election Prediction using Twitter Data

Daniel Gayo-Avello

(Submitted on 28 Apr 2012)

Predicting X from Twitter is a popular fad within the Twitter research subculture. It seems both appealing and relatively easy. Amongst all the things that can be predicted from Twitter, electoral prediction is maybe the most attractive, and at this moment there is a growing body of literature on such a topic. This is not just a simple research problem but, above all, it is extremely difficult. However, most of the authors seem to be more interested in claiming positive results than in providing sound and reproducible methods. It is also especially worrisome that many recent papers seem to only acknowledge those who have used the idea of Twitter predicting elections, instead of conducting a balanced literature review showing both sides of the matter. After reading many of these papers I have decided to write such a survey myself. Hence, in this paper, every study relevant to the matter of electoral prediction using Twitter data is commented. From this review it can be concluded that the predictive power of Twitter regarding elections has been greatly exaggerated. Many interesting research problems still lie ahead.

Comments: 13 pages, no figures. Annotated bibliography of 25 papers regarding electoral prediction from Twitter data

Subjects: **Computers and Society (cs.CY)**; Computation and Language (cs.CL); Social and Information Networks (cs.SI); Physics and Society (physics.SI)

Cite as: [arXiv:1204.6441](https://arxiv.org/abs/1204.6441) [cs.CY]

(or [arXiv:1204.6441v1](https://arxiv.org/abs/1204.6441v1) [cs.CY] for this version)

Submission history

Why Social Media Can't Predict Elections

- Post-hoc analysis, not *real* prediction
- Demographics not considered
- The people who tweet may not be the people who vote
- There's no way to count votes on Twitter – even your neighbor's dog has a Twitter profile
- Chance is not a valid baseline because incumbency tends to play a major role in most of the elections
- All the tweets are assumed to be trustworthy. That is, the presence of rumors, propaganda, misleading information, sarcasm, humor is ignored.
- Self-selection bias is simply ignored. People tweet on a voluntary basis and, therefore, data is produced only by those politically active.

[Google.org home](#)

[Dengue Trends](#)

Flu Trends

[Home](#)

United States ▾

National ▾

[Download data](#)

[How does this work?](#)

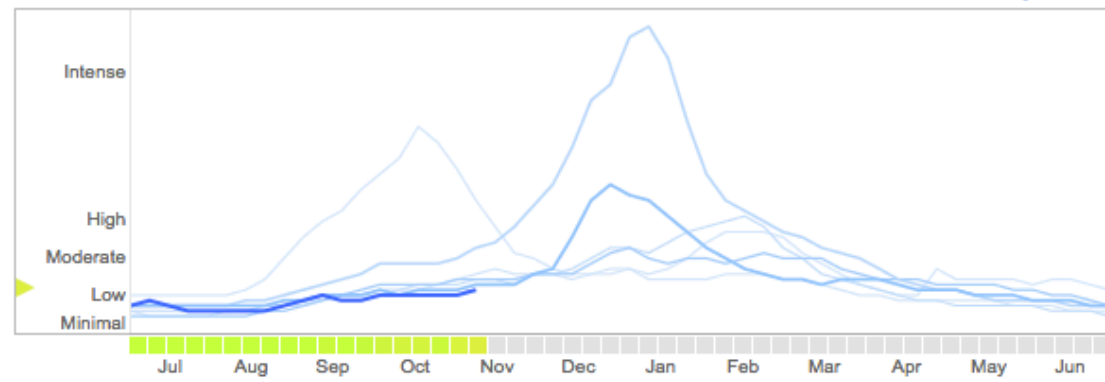
[FAQ](#)

Explore flu trends - United States

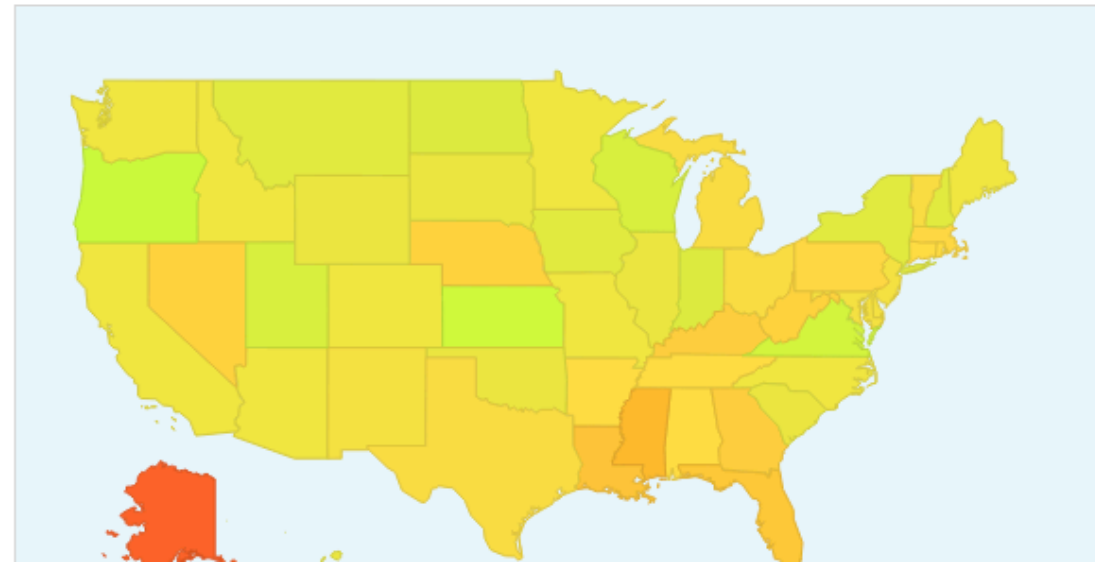
We've found that certain search terms are good indicators of flu activity. Google Flu Trends uses aggregated Google search data to estimate flu activity. [Learn more »](#)

National

● 2014-2015 ● Past years ▾



States | [Cities](#) (Experimental)





Steven Salzberg
Contributor

FOLLOW

Fighting Pseudoscience
full bio →

Opinions expressed by Forbes
Contributors are their own.



6
COMMENTS

PHARMA & HEALTHCARE 3/23/2014 @ 9:00AM | 47,371 views

Why Google Flu Is A Failure

+ Comment Now + Follow Comments

It seemed like such a good idea at the time.

People with the flu (the influenza virus, that is) will probably go online to find out how to treat it, or to search for other information about the flu. So [Google](#) GOOG +1.04% decided to track such behavior, hoping it might be able to predict flu outbreaks even faster than traditional health authorities such as the Centers for Disease Control (CDC).

Instead, as the authors of [a new article in Science](#) explain, we got “big data hubris.” David Lazer and colleagues explain that:

“Big data hubris” is the often implicit assumption that big data are a substitute for, rather than a supplement to, traditional data collection and analysis.

The folks at Google figured that, with all their massive data, they could outsmart anyone.



Edition: U.S. ▾



Like 182k



Follow



Newsletters



Huffington Post Search

FRONT PAGE

WOMEN

TASTE

GOOD NEWS

PARENTS

STYLE

POST50

RELIGION

WEDDINGS

HUFFPOST LIVE

ALL SECTIONS

[Healthy Living](#) • [Health And Fitness](#) • [GPS for the Soul](#) • [Health News](#) • [Sleep](#) • [Healthy Living Videos](#) • [Moments Not Milestones](#) • [Lifestyle](#) • [Third Metric](#) • [Thrive](#)



10 Ways Introverts Interact Differently With The World



7 Things You Should Never Apologize For

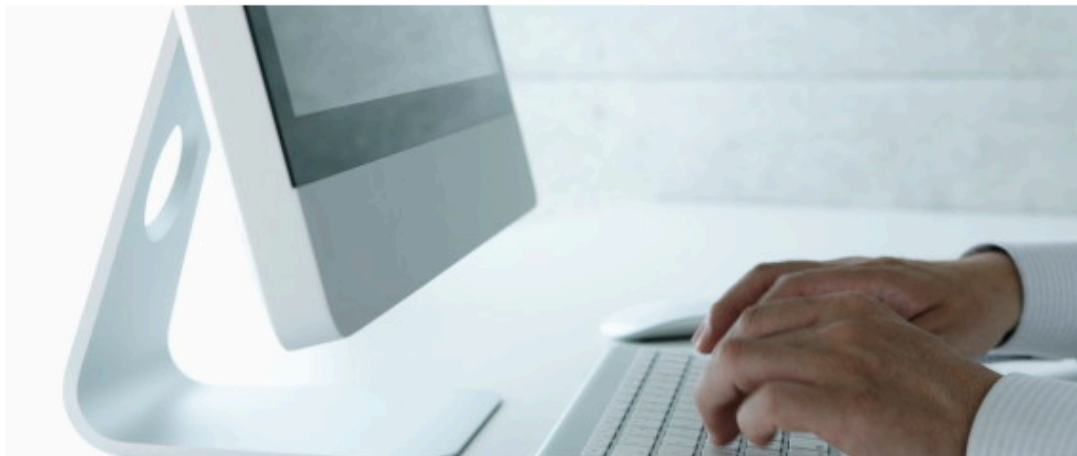


When Being Beautiful Might Count Against You



Data Fail! How Google Flu Trends Fell Way Short

Posted: 03/16/2014 8:12 pm EDT | Updated: 03/16/2014 8:59 pm EDT



MORE: [Flu Trends](#), [Google Flu Trends Inaccurate](#), [Google Flu Trends Fail](#), [Google Flu Trends Data](#)

Science The World's Leading Journal of Original Scientific Research, Global News, and Commentary.

Science Home

Current Issue

Previous Issues

Science Express

Science Products

My Science

About the Journal

Home > Science Magazine > 14 March 2014 > Lazer *et al.*, 343 (6176): 1203–1205

Science 14 March 2014:
Vol. 343 no. 6176 pp. 1203–1205
DOI: 10.1126/science.1248506

< Prev | Table of Contents | Next >

Read Full Text to Comment (1)

ADVERTISEMENT

Science Webinar SERIES

Register for upcoming webinars.

Access Now!

Article Views

> Summary

> Full Text

> Full Text (PDF)

> Figures Only

> Supplementary Materials

> Podcast Interview

POLICY FORUM

BIG DATA

The Parable of Google Flu: Traps in Big Data Analysis

David Lazer^{1,2,*}, Ryan Kennedy^{1,3,4}, Gary King³, Alessandro Vespignani^{5,6,3}

± Author Affiliations

*Corresponding author. E-mail: d.lazer@neu.edu.

In February 2013, Google Flu Trends (GFT) made headlines but not for a reason that Google executives or the creators of the flu tracking system would have hoped. *Nature* reported that GFT was predicting more than double the proportion of doctor visits for influenza-like illness (ILI) than the Centers for Disease Control and Prevention (CDC), which bases its estimates on surveillance reports from laboratories across the United States (1, 2). This happened despite the fact that GFT was built to predict CDC reports. Given that GFT is often held up as an exemplary use of big data (3, 4), what lessons can we draw from this error?

[Read the Full Text](#)

The editors suggest the following Related Resources on *Science* sites

In *Science Magazine*

Article Tools

> Save to My Folders

> Download Citation

> Alert Me When Article is Cited

> Post to CiteULike

> Article Usage Statistics

> E-mail This Page

> Rights & Permissions

> Commercial Reprints and E-Prints

ADVERTISEMENT

AAAS
ADVANCING SCIENCE. SERVING SOCIETY

&

Nationwide
Insurance

Exclusive Discounts

What is the primary challenge of social media based predictions over traditional predictions?

Asur and Huberman's paper primarily uses observational data for prediction. Note all focus on retrospective prediction. What are the problems with this approach? How to fix this problem?

Kwak et al. found that 85% trending topics are news. To what extent the design of Twitter (and its trending topic) algorithm is responsible for it? Is it still true (given the Twitter of today?)

What makes a social media a news media?
Guided by Kwak et al's methodology, is
Facebook a news media (Mark Z.'s statement)

Conversely, is the New York Times a "social
media" now that you can comment on articles?

Facebook fires trending team, and algorithm without humans goes crazy

Module pushes out false story about Fox's Megyn Kelly, offensive Ann Coulter headline and a story link about a man masturbating with a McDonald's sandwich

● [I worked on Facebook's Trending team - the most toxic work experience of my life](#)

Sam Thielman in New York

@samthielman

Monday 29 August 2016
12.48 EDT



This article is 2 months old

Shares 20k
Comments 734

Save for later



The trending module was meant to have 'learned' from the human editors' curation decisions. Photograph: Bloomberg/Bloomberg via Getty Images

What utility or unique dimension does social media bring to event and news analytics?
What are the challenges? (Compare with traditional forms of news coverage)

Dikapoulos and Shamma focused on understanding the dynamics of a planned event with social media. To what extent would it be possible or useful to examine the patterns of unplanned events in this manner?