

# **CS 6474/CS 4803 Social Computing: Benefits/Applications of Social Computing Systems - Politics**

*Munmun De Choudhury*

[munmund@gatech.edu](mailto:munmund@gatech.edu)

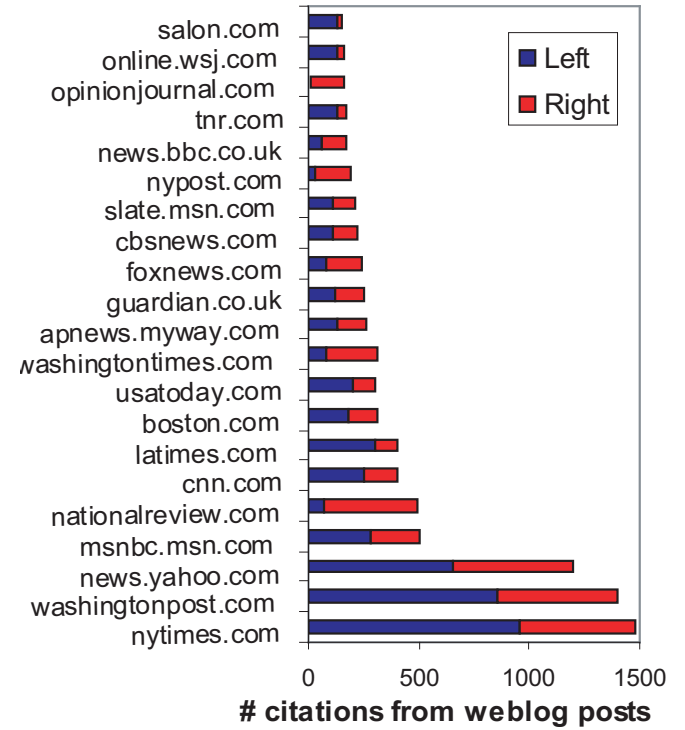
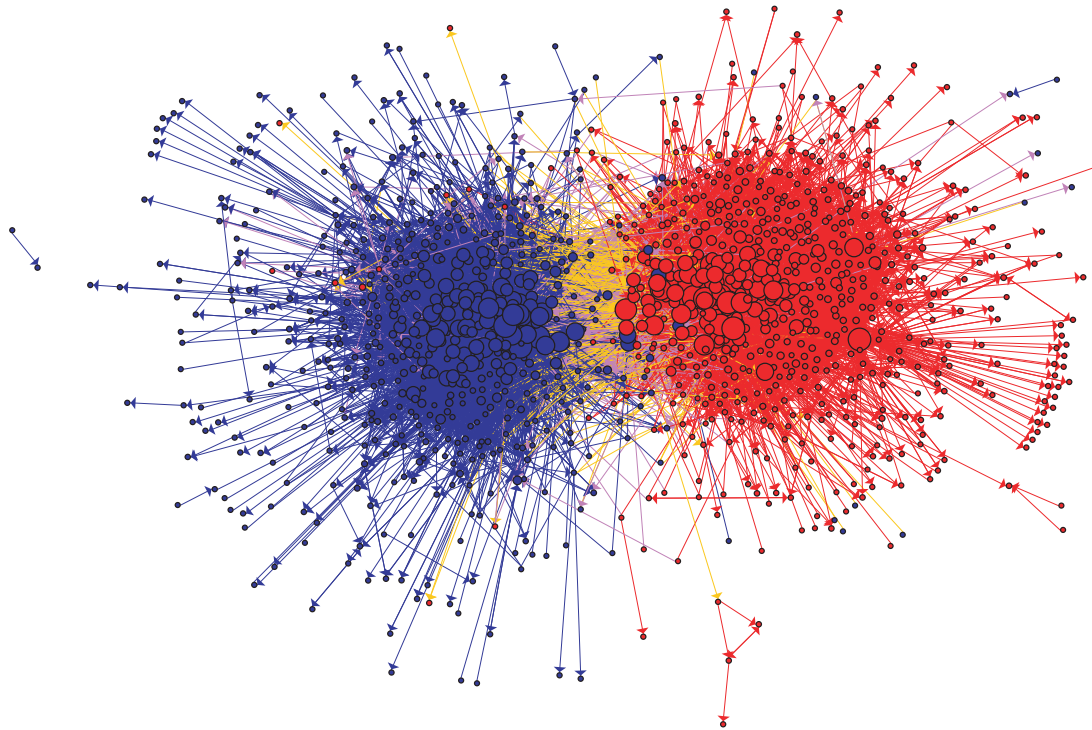
Week 8 | February 28, 2024

Monday, March 4– In-person, Guest  
Lecture by Prof Soya Park (Emory)

Wednesday, March 6 – No class,  
recorded video

# The Political Blogosphere and the 2004 U.S. Election: Divided They Blog

# Summary



Adamic and Glance only analyzed *connections* between conservatives and liberals. Could language analysis of blog content revealed something different?

Would we expect any contrast if the study was done for the 2016 or 2020 US Presidential elections?

The clustered nature of the network – is that a mere reflection of the society or what is manifested online is relatively more exaggerated/muted?

Adamic and Glance only analyzed *connections* between conservatives and liberals. Could language analysis of blog content revealed something different?

Would we expect any contrast if the study was done for the 2016 or 2020 US Presidential elections?

The clustered nature of the network – is that a mere reflection of the society or what is manifested online is relatively more exaggerated/muted?

# Characterizing social media manipulation in the 2020 U.S. presidential election

by Emilio Ferrara, Herbert Chang, Emily Chen,  
Goran Muric, and Jaimin Patel

## Abstract

Democracies are postulated upon the ability to carry out fair elections, free from any form of interference or manipulation. Social media have been reportedly used to distort public opinion nearing election events in the United States and beyond. With over 240 million election-related tweets recorded between 20 June and 9 September 2020, in this study we chart the landscape of social media manipulation in the context of the upcoming 3 November 2020 U.S. presidential election. We focus on characterizing two salient dimensions of social media manipulation, namely (i) automation (*e.g.*, the prevalence of bots), and (ii) distortion (*e.g.*, manipulation of narratives, injection of conspiracies or rumors). Despite being outnumbered by several orders of magnitude, just a few thousands of bots generated spikes of conversations around real-world political events in all comparable with the volume of activity of humans. We discover that bots also exacerbate the consumption of content produced by users with their same political views, worsening the issue of political echo chambers. Furthermore, coordinated efforts carried out by Russia, China and other countries are hereby characterized. Finally, we draw a clear connection between bots, hyper-partisan media outlets, and conspiracy groups, suggesting the presence of systematic efforts to distort political narratives and propagate disinformation. Our findings may have impactful implications, shedding light on different forms of social media manipulation that may, altogether, ultimately pose a risk to the integrity of the election.

## Contents

[Introduction](#)  
[Methodology](#)  
[Data analysis](#)  
[Distortion](#)  
[Conclusions](#)

Adamic and Glance only analyzed *connections* between conservatives and liberals. Could language analysis of blog content revealed something different?

Would we expect any contrast if the study was done for the last US Presidential election?

The clustered nature of the network – is that a mere reflection of the society or what is manifested online is relatively more exaggerated/muted?



Beyond online...

# It's Not the Technology, Stupid: How the 'Echo Chamber' and 'Filter Bubble' Metaphors Have Failed Us

---

AXEL BRUNS  
DIGITAL MEDIA RESEARCH CENTRE  
QUEENSLAND UNIVERSITY OF TECHNOLOGY  
BRISBANE, AUSTRALIA  
A.BRUNS@QUT.EDU.AU  
[HTTP://SNURB.INFO/](http://snurb.info/)  
[@SNURB\\_DOT\\_INFO](https://twitter.com/SNURB_DOT_INFO)

## Introduction

Following the surprise victories of Brexit and Trump in 2016, even the outgoing U.S. president Barack Obama (2017) warned in his farewell speech that “it’s become safer to retreat into our own bubbles”, thereby linking increased electoral volatility and ideological polarisation with concepts such as “echo chambers” (Sunstein 2001) and “filter bubbles” (Pariser 2011). The politicians, journalists, and scholars who support these concepts suggest that, with online and social media as the primary sources of information for a growing percentage of the public (Newman *et al.* 2016), echo chambers and filter bubbles are chiefly responsible for the emergence of communities that espouse contrarian and counterfactual perspectives and ideologies, and for their disconnection from mainstream public debates.

Echo chambers are said to enable these groups to reinforce their views by connecting with like-minded others; filter bubbles to shield them from encountering contrary perspectives. Such disconnection from and ignorance of alternative perspectives is assumed to result from a combination of individual choice, in selecting the news sources to consult or the social media accounts to follow, and the algorithmic shaping of such

Revisiting “The clustered nature of the network – is that a mere reflection of the society or what is manifested online is relatively more exaggerated/muted?”

# Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment

# From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series

**Brendan O'Connor**<sup>†</sup> **Ramnath Balasubramanyan**<sup>†</sup> **Bryan R. Routledge**<sup>§</sup> **Noah A. Smith**<sup>†</sup>  
brenocon@cs.cmu.edu   rbalasub@cs.cmu.edu   routledge@cmu.edu   nasmith@cs.cmu.edu

<sup>†</sup>School of Computer Science  
Carnegie Mellon University

<sup>§</sup>Tepper School of Business  
Carnegie Mellon University

## Abstract

We connect measures of public opinion measured from polls with sentiment measured from text. We analyze several surveys on consumer confidence and political opinion over the 2008 to 2009 period, and find they correlate to sentiment word frequencies in contemporaneous Twitter messages. While our results vary across datasets, in several cases the correlations are as high as 80%, and capture important large-scale trends. The results highlight the potential of text streams as a substitute and supplement for traditional polling.

## Introduction

If we want to know, say, the extent to which the U.S. population likes or dislikes Barack Obama, an obvious thing to do is to ask a random sample of people (i.e., poll). Survey and polling methodology, extensively developed through the 20th century (Krosnick, Judd, and Wittenbrink 2005), gives numerous tools and techniques to accomplish representative public opinion measurement.

With the dramatic rise of text-based social media, millions of people broadcast their thoughts and opinions on a

statistics derived from extremely simple text analysis techniques are demonstrated to correlate with polling data on consumer confidence and political opinion, and can also predict future movements in the polls. We find that temporal smoothing is a critically important issue to support a successful model.

## Data

We begin by discussing the data used in this study: Twitter for the text data, and public opinion surveys from multiple polling organizations.

## Twitter Corpus

Twitter is a popular microblogging service in which users post messages that are very short: less than 140 characters, averaging 11 words per message. It is convenient for research because there are a very large number of messages, many of which are publicly available, and obtaining them is technically simple compared to scraping blogs from the web.

---

# Predicting US Primary Elections with Twitter

---

## **Lei Shi**

R&D, Opera Solutions  
12230 El Camino Real Suite 330,  
San Diego, CA, US  
lshi@operasolutions.com

## **Neeraj Agarwal**

R&D, Opera Solutions  
Express Trade Towers, Plot 15-16,  
Floor 5-6, Noida (UP), INDIA  
neeraj.agarwal@  
operasolutions.com

## **Ankur Agrawal**

R&D, Opera Solutions  
Express Trade Towers, Plot 15-16,  
Floor 5-6, Noida (UP), INDIA  
ankur.agrawal@  
operasolutions.com

## **Rahul Garg**

R&D, Opera Solutions  
Express Trade Towers, Plot 15-16,  
Floor 5-6, Noida (UP), INDIA  
rahul.garg@  
operasolutions.com

## **Jacob Spoelstra**

R&D, Opera Solutions  
12230 El Camino Real Suite 330,  
San Diego, CA, US  
JSpoelstra@operasolutions.com

## **Abstract**

Using social media for political analysis is becoming a common practice, especially during election time. Many researchers and media are trying to use social media to understand the public opinion and trend. In this paper, we investigate how we could use Twitter to predict public opinion and thus predict American republican presidential election results. We analyzed millions of tweets from September 2011 leading up to the republican primary elections. First we examine the previous methods regarding predicting election results with social media and then we integrate our understanding of social media and propose a prediction model to predict the public opinions towards Republican Presidential Elections. Our results highlight the feasibility of using social media to predict public opinions and thus replace traditional polling.

# A System for Real-time Twitter Sentiment Analysis of 2012 U.S. Presidential Election Cycle

**Hao Wang\***, **Dogan Can\*\***, **Abe Kazemzadeh\*\***,  
**François Bar\*** and **Shrikanth Narayanan\*\***

Annenberg Innovation Laboratory (AIL)\*

Signal Analysis and Interpretation Laboratory (SAIL)\*\*

University of Southern California, Los Angeles, CA

{haowang@, dogancan@, kazemzad@, fbar@, shri@sipi}.usc.edu

## Abstract

This paper describes a system for real-time analysis of public sentiment toward presidential candidates in the 2012 U.S. election as expressed on Twitter, a micro-blogging service. Twitter has become a central site where people express their opinions and views on political parties and candidates. Emerging events or news are often followed almost instantly by a burst in Twitter volume, providing a unique opportunity to gauge the relation between expressed public sentiment and electoral events. In addition, sentiment analysis can help explore how these events affect public opinion. While traditional content analysis takes days or weeks to complete, the system demonstrated here analyzes sentiment in the entire Twitter traffic about the election, delivering results instantly and continuously. It offers the public, the media, politicians and scholars a new and timely perspective on the dynamics of the electoral process and public opinion.

have developed a tool for real-time analysis of sentiment expressed through Twitter, a micro-blogging service, toward the incumbent President, Barack Obama, and the nine republican challengers - four of whom remain in the running as of this writing. With this analysis, we seek to explore whether Twitter provides insights into the unfolding of the campaigns and indications of shifts in public opinion.

Twitter allows users to post tweets, messages of up to 140 characters, on its social network. Twitter usage is growing rapidly. The company reports over 100 million active users worldwide, together sending over 250 million tweets each day (Twitter, 2012). It was actively used by 13% of on-line American adults as of May 2011, up from 8% a year prior (Pew Research Center, 2011). More than two thirds of U.S. congress members have created a Twitter account and many are actively using Twitter to reach their constituents (Lassen & Brown, 2010; TweetCongress, 2012). Since October 12, 2012, we have gathered over 36 million tweets about the 2012 U.S. presidential candidates, a quarter million per day on average. During one of the key political events, the Dec 15, 2011 primary debate in Iowa, we collected more

# Predicting the 2011 Dutch Senate Election Results with Twitter

**Erik Tjong Kim Sang** and **Johan Bos**

Alfa-informatica

University of Groningen

Groningen, The Netherlands

{e.f.tjong.kim.sang, johan.bos}@rug.nl

## **Abstract**

To what extent can one use Twitter in opinion polls for political elections? Merely counting Twitter messages mentioning political party names is no guarantee for obtaining good election predictions. By improving the quality of the document collection and by performing sentiment analysis, predictions based on entity counts in tweets can be considerably improved, and become nearly as good as traditionally obtained opinion polls.

tweets, with the general aim of developing natural language processing tools for automatically analyzing the content of the messages in this new social medium, which comes with its own challenges. When the Dutch Senate elections took place in 2011, we took this as an opportunity to verify the predictive power of tweets.

More concretely, we wanted to test whether by simply counting Twitter messages mentioning political party names we could accurately predict the election outcome. Secondly, we wanted to investigate factors that influence the predictions based on the Dutch tweets.

## Studying political microblogging: Twitter users in the 2010 Swedish election campaign

▼ Larsson, Anders Olof

Uppsala University, Disciplinary Domain of Humanities and Social Sciences, Faculty of Social Sciences, Department of Informatics and Media.

▼ Moe, Hallvard

Universitetet i Bergen. (Institutt for informasjons- og medievitenskap)

2012 (English)

*In: New Media and Society*, ISSN 1461-4448, E-ISSN 1461-7315, Vol. 14, no 5, 729-747 p.

### Article in journal (Refereed) Published

#### Abstract [en]

Among the many so-called microblogging services that allow their users to describe their current status in short posts, Twitter is probably among the most popular and well known. Since its launch in 2006, Twitter use has evolved and is increasingly used in a variety of contexts. This article utilizes emerging online tools and presents a rationale for data collection and analysis of Twitter users. The suggested approach is exemplified with a case study: Twitter use during the 2010 Swedish election. Although many of the initial hopes for e-democracy appear to have gone largely unfulfilled, the successful employment of the internet during the 2008 US presidential campaign has again raised voices claiming that the internet, and particularly social media applications like Twitter, provides interesting opportunities for online campaigning and deliberation. Besides providing an overarching analysis of how Twitter use was fashioned during the 2010 Swedish election campaign, this study identifies different user types based on how high-end users utilized the Twitter service. By suggesting a novel approach to the study of microblogging and by identifying user types, this study contributes to the burgeoning field of microblog research and gives specific insights into the practice of civic microblogging.

#### Place, publisher, year, edition, pages

2012. Vol. 14, no 5, 729-747 p.

#### National Category

Social Sciences

Identifiers

## Open Access in DiVA

No full text

## Other links

[Publisher's full text](#)

## Search in DiVA

### By author/editor

[Larsson, Anders Olof](#)

### By organisation

[Department of Informatics and Media](#)

### In the same journal

[New Media and Society](#)

### On the subject

[Social Sciences](#)

## Search outside of DiVA

[Google](#)

[Google Scholar](#)

## Citations

Web of Science®: 141

## Altmetric score





# Tweets and Votes: A Study of the 2011 Singapore General Election

Marko Skoric  
Nanyang Technological University, Singapore  
[Marko@ntu.edu.sg](mailto:Marko@ntu.edu.sg)

Nathaniel Poor  
Brooklyn, NY, USA  
[natpoor@gmail.com](mailto:natpoor@gmail.com)

Palakorn Achananuparp  
Singapore Management University,  
Singapore  
[palakorna@smu.edu.sg](mailto:palakorna@smu.edu.sg)

Ee-Peng Lim  
Singapore Management University,  
Singapore  
[eplim@smu.edu.sg](mailto:eplim@smu.edu.sg)

Jing Jiang  
Singapore Management University,  
Singapore  
[jingjiang@smu.edu.sg](mailto:jingjiang@smu.edu.sg)

## Abstract

*This study focuses on the uses of Twitter during the elections, examining whether the messages posted online are reflective of the climate of public opinion. Using Twitter data obtained during the official campaign period of the 2011 Singapore General Election, we test the predictive power of tweets in forecasting the election results. In line with some previous studies, we find that during the elections the Twittersphere represents a rich source of data for gauging public opinion and that the frequency of tweets mentioning names of political parties, political candidates and contested constituencies could be used to make predictions about the share of votes at the national level, although the accuracy of the predictions was significantly lower than in the studies done in Germany and the UK. At the level of constituency the predictive power of tweets was much weaker, although still better than chance. The findings suggest that the context in which the elections take place also matters, and that issues like media freedoms, competitiveness of the elections and specifics of the electoral system may lead to certain over- and under-estimations of voting sentiment. The implications for future research are discussed.*

Spring” [18] and during the 2009 Iran election protests [4, 12, 21], and at one point the US State Department asked Twitter to delay scheduled maintenance of the service, which would have required downtime, in order to allow Iranians to continue using the service [21].

What is the role of Twitter in political life? Aside from perhaps providing us with direct real-time access to information on demonstrations, protests and revolutions in authoritarian countries, can the Twittersphere help us understand the climate of public opinion around the world?

Given that Twitter is increasingly appropriated for both conversation and collaboration [11], and that tweets can be seen as an electronic word-of-mouth communication [13], it is likely that we can learn something about political sentiment by eavesdropping on these conversations. Typical uses of Twitter, including daily chatter, information sharing, reporting news and conversing [13], can all contain indicators of political opinion and sentiment. This is particularly true during the times of elections, when citizens are more likely to discuss politicians, parties and political issues online. Scholars have argued that Twitter can be used as a “social sensor” to make predictions about electoral outcomes [17]. With 65 million tweets a day by June, 2010 [24], Twitter represents a substantial

# More Tweets, More Votes: Social Media as a Quantitative Indicator of Political Behavior

Joseph DiGrazia<sup>1\*</sup>, Karissa McKelvey<sup>2</sup>, Johan Bollen<sup>2</sup>, Fabio Rojas<sup>1</sup>

**1** Department of Sociology, Indiana University, Bloomington, Indiana, United States of America, **2** School of Informatics and Computing, Indiana University, Bloomington, Indiana, United States of America

## Abstract

Is social media a valid indicator of political behavior? There is considerable debate about the validity of data extracted from social media for studying offline behavior. To address this issue, we show that there is a statistically significant association between tweets that mention a candidate for the U.S. House of Representatives and his or her subsequent electoral performance. We demonstrate this result with an analysis of 542,969 tweets mentioning candidates selected from a random sample of 3,570,054,618, as well as Federal Election Commission data from 795 competitive races in the 2010 and 2012 U.S. congressional elections. This finding persists even when controlling for incumbency, district partisanship, media coverage of the race, time, and demographic variables such as the district's racial and gender composition. Our findings show that reliable data about political behavior can be extracted from social media.

**Citation:** DiGrazia J, McKelvey K, Bollen J, Rojas F (2013) More Tweets, More Votes: Social Media as a Quantitative Indicator of Political Behavior. PLoS ONE 8(11): e79449. doi:10.1371/journal.pone.0079449

**Editor:** Luis M. Martinez, CSIC-Univ Miguel Hernandez, Spain

**Received:** April 18, 2013; **Accepted:** September 22, 2013; **Published:** November 27, 2013

**Copyright:** © 2013 DiGrazia 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:** The authors gratefully acknowledge support from the National Science Foundation (grants SBE 0914939 and CCF 1101743) and the McDonnell Foundation (grant 2011022). The funders' websites can be accessed at <http://nsf.gov> and <http://jsmf.org>, respectively. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** Co-author on this paper, Johan Bollen, is a PLOS ONE Editorial Board member. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

\* E-mail: [jdigrazi@indiana.edu](mailto:jdigrazi@indiana.edu)

# **Social Media Analysis and Public Opinion: The 2010 UK General Election**

## **Abstract**

Social media monitoring in politics can be understood by situating it in theories of public opinion. The multi-method study we present here indicates how social media monitoring allow for analysis of social dynamics through which opinions form and shift. Analysis of media coverage from the 2010 UK General Election demonstrates that social media are now being equated with public opinion by political journalists. Building on this, we use interviews with pollsters, social media researchers and journalists to examine the perceived link between social media and public opinion. In light of competing understandings these interviews reveal, we argue for a broadening of the definition of public opinion to include its social dimension.

**Keywords:** Elections, Grounded Theory, Public Opinion, Social Media, Twitter, United Kingdom.

# Discussion Point

What is it about social media, particularly Twitter, that makes it suitable for understanding political opinions, over other platforms? Would you pick a different platform for this purpose and why?

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

Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon

Department of Computer Science, KAIST  
335 Gwahangno, Yuseong-gu, Daejeon, Korea  
{haewoon, chlee, hosung}@an.kaist.ac.kr, sbmoon@kaist.edu

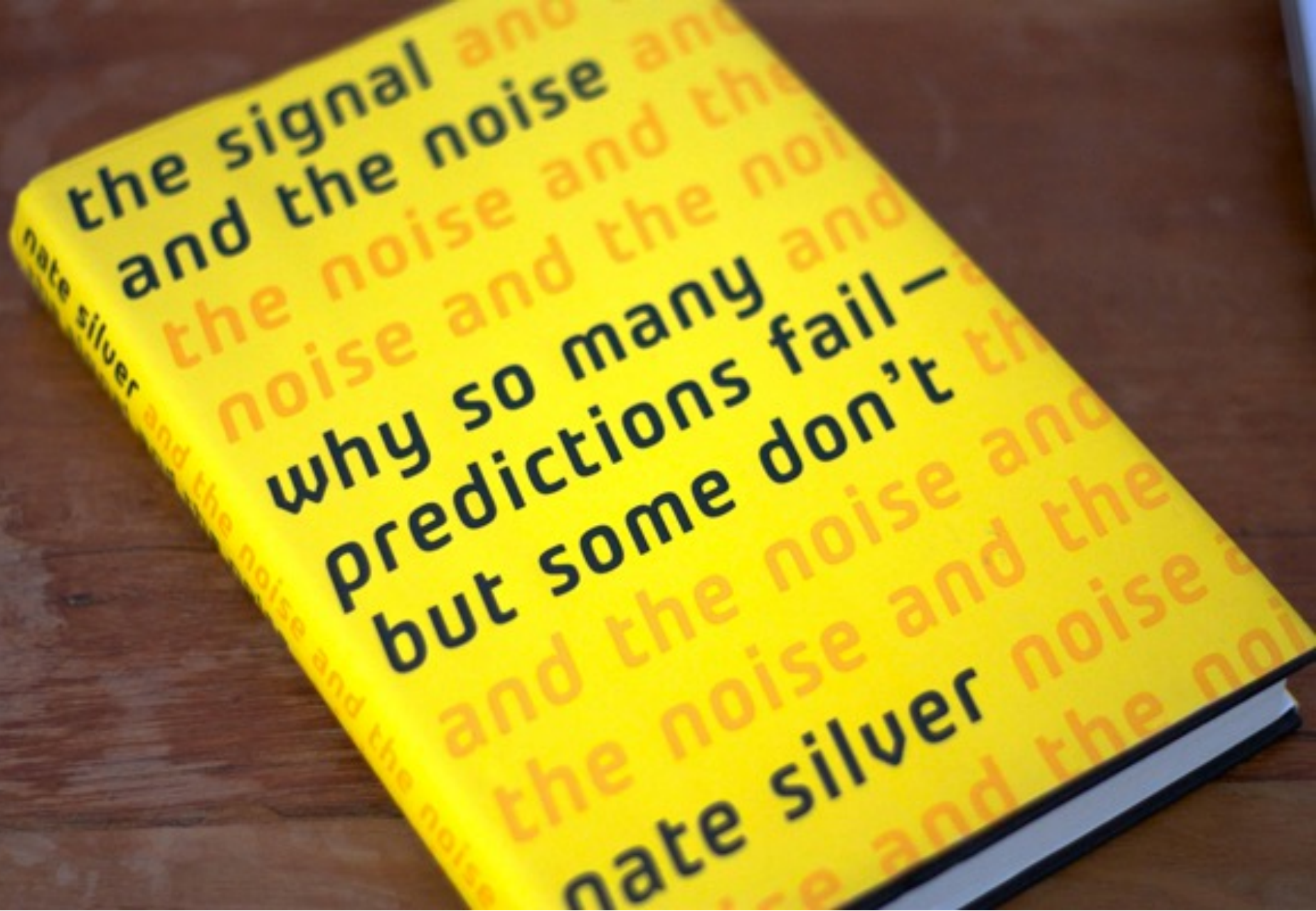
## ABSTRACT

Twitter, a microblogging service less than three years old, commands more than 41 million users as of July 2009 and is growing fast. Twitter users tweet about any topic within the 140-character limit and follow others to receive their tweets. The goal of this paper is to study the topological characteristics of Twitter and its power as a new medium of information sharing.

We have crawled the entire Twitter site and obtained 41.7 million user profiles, 1.47 billion social relations, 4,262 trending topics, and 106 million tweets. In its follower-following topology analysis we have found a non-power-law follower distribution, a short effective diameter, and low reciprocity, which all mark a deviation from known characteristics of human social networks [28]. In order to identify influentials on Twitter, we have ranked users by the number of followers and by PageRank and found two rankings to be similar. Ranking by retweets differs from the previous two rankings, indicating a gap in influence inferred from the number of followers and that from the popularity of one's tweets. We have analyzed the

## 1. INTRODUCTION

Twitter, a microblogging service, has emerged as a new medium in spotlight through recent happenings, such as an American student jailed in Egypt and the US Airways plane crash on the Hudson river. Twitter users follow others or are followed. Unlike on most online social networking sites, such as Facebook or MySpace, the relationship of following and being followed requires no reciprocation. A user can follow any other user, and the user being followed need not follow back. Being a follower on Twitter means that the user receives all the messages (called *tweets*) from those the user follows. Common practice of responding to a tweet has evolved into well-defined markup culture: RT stands for retweet, '@' followed by a user identifier address the user, '#' followed by a word represents a hashtag. This well-defined markup vocabulary combined with a strict limit of 140 characters per posting conveniences users with brevity in expression. The *retweet* mechanism empowers users to spread information of their choice beyond the reach of the original tweet's followers.



Hypothesis testing, statistical significance, descriptive methods, experimental approach

# Chance of winning

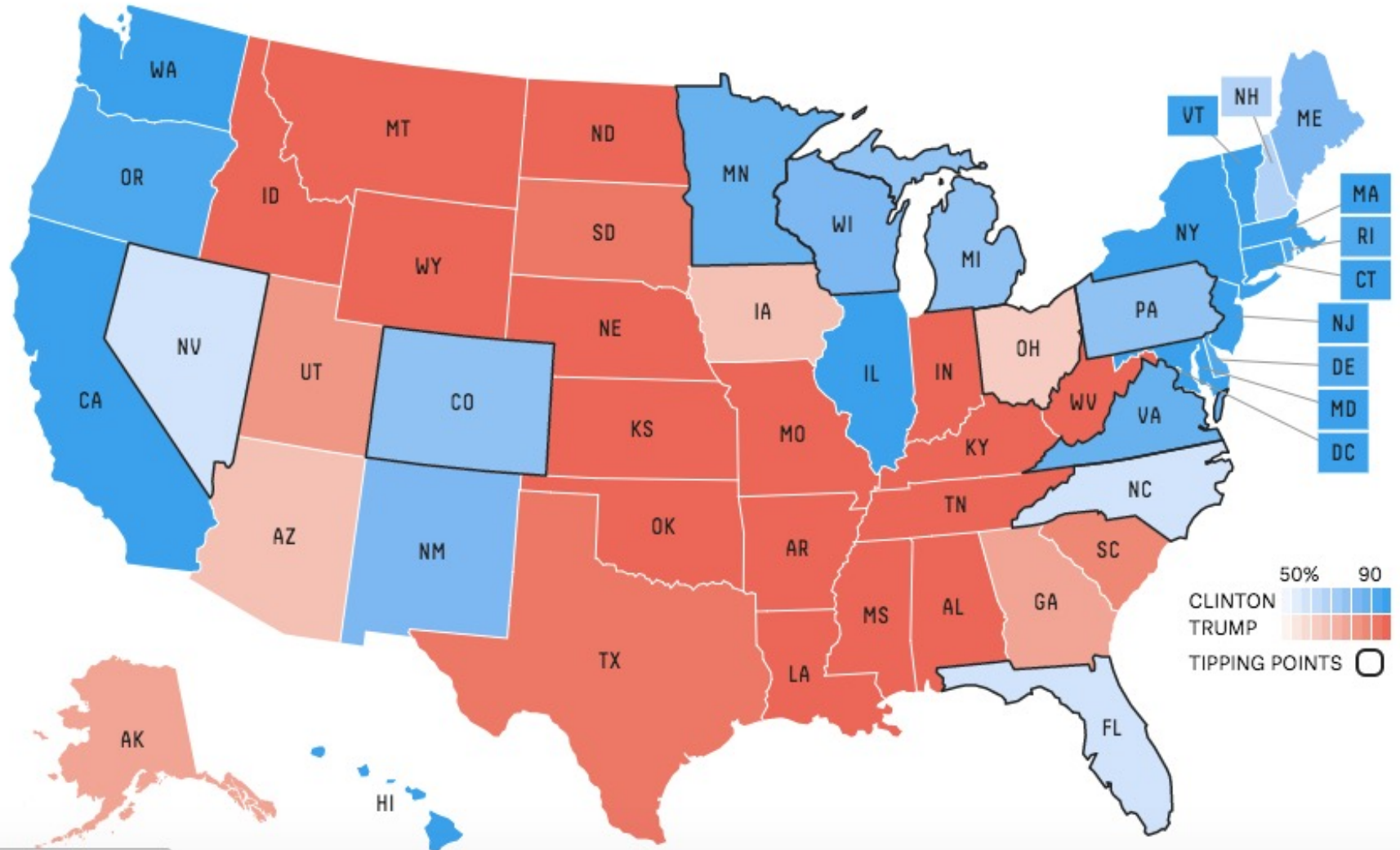


Hillary Clinton

**71.4%**

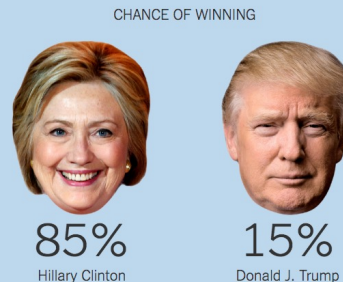
Donald Trump

**28.6%**



# Hillary Clinton has an 85% chance to win.

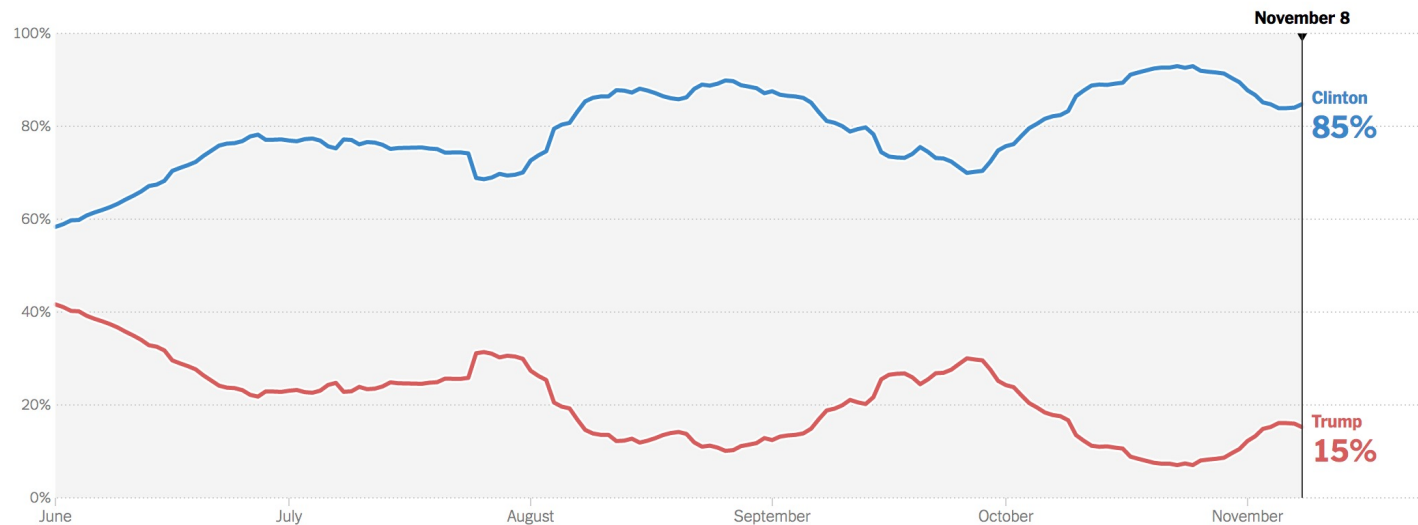
Last updated Tuesday, November 8 at 10:20 PM ET



[Forecast history](#) \ 
 [Recent changes](#) \ 
 [State by state](#) \ 
 [Other forecasts](#) \ 
 [Likely scenarios](#) \ 
 [Explore paths](#) \

The Upshot's elections model suggests that Hillary Clinton is favored to win the presidency, based on [the latest state and national polls](#). A victory by Mr. Trump remains possible: Mrs. Clinton's chance of losing is about the same as the probability that [an N.F.L. kicker misses a 37-yard field goal](#).

For months, we've been updating our estimates with each new poll. Today, it's Election Day, what we've all been waiting for, and there will be no more updates. You can chart different paths to victory below. Here's how our estimates have changed over time:







# "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. Among 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 a research problem but, above all, it is extremely difficult. However, most of the authors seem to be more interested in claiming positive results than providing sound and reproducible methods. It is also especially worrisome that many recent papers seem to only acknowledge those studies that support the idea of Twitter predicting elections, instead of conducting a balanced literature review showing both sides of the matter. After reading several 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. More 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

# Can Collective Sentiment Expressed on Twitter Predict Political Elections?

**Jessica Chung\*** and **Eni Mustafaraj**

Computer Science Department

Wellesley College

jchung, emustafa@wellesley.edu

## Abstract

Research examining the predictive power of social media (especially Twitter) displays conflicting results, particularly in the domain of political elections. This paper applies methods used in studies that have shown a direct correlation between volume/sentiment of Twitter chatter and future electoral results in a new dataset about political elections. We show that these methods display a series of shortcomings, that make them inadequate for determining whether social media messages can predict the outcome of elections.

## Introduction

Many studies have shown the promise of using social media communication to predict the future. The microblogging and social networking service Twitter, which allows its users to publish short, 140-character messages, has been used as a data source for successfully predicting box office revenue for movies (Asur and Huberman 2010), as well as predicting stock market performance (Bollen, Mao, and Zeng 2010). In the realm of politics, however, the existing work relevant to the predictive power of Twitter chatter volume is conflicting. In Germany, the share of tweets alone, could accurately

Coakley		Brown	
#tweets	%	#tweets	%
52,116	53.86	44,654	46.14

Table 1: The share of tweets for each candidate in the MAsen10 data set, in a six day period before election day.

## Data, Results, Analysis

### Data Collection

The 2010 US Senate special election in Massachusetts (“MAsen10”) was held on January 19th, 2010 between the democratic candidate, Martha Coakley and the republican candidate, Scott Brown. Using the Twitter streaming API, we collected tweets that contained either or both candidates’ names. There were 234,697 tweets from 56,165 unique users collected from January 13 to January 20, 2010. The collected data was passed through a series of preprocessing steps in order to remove extraneous material. Hashtags, account names and links to web sites were removed. Contractions were replaced by their full form and emoticons such as “:)” were replaced by named tags, e.g. <happy>.

# Don't Turn Social Media Into Another 'Literary Digest' Poll\*

Daniel Gayo-Avello<sup>†</sup>

September 26, 2011

Department of Computer Science, University of Oviedo (SPAIN)

## **Abstract**

User generated content has experienced an explosive growth both in the diversity of applications and the volume of topics covered by its users. Content published in micro-blogging systems like Twitter is thought to be feasibly data-mined in order to “take the pulse” of society. Recently, a number of positive studies have been published praising the goodness of relatively simple approaches to sampling, opinion mining, and sentiment analysis. This paper will attempt to play devil’s advocate by detailing a study in which such simple approaches largely overestimated Obama’s victory in the 2008 U.S. Presidential Elections. A thorough post-mortem of that experiment has been conducted and several important lessons have been extracted.

# Debate Exercise

## Side 1

Social media is a powerful instrument for political (election) prediction. And here's why...

## Side 2

Using social media for political (election) prediction is flawed. And here's why...

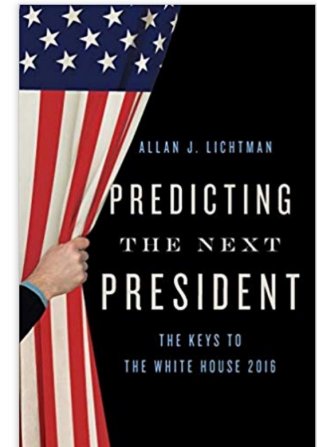
# Why Social Media Can't Predict Elections

Books > New, Used & Rental Textbooks > Social Sciences Predicting the Next President: The Keys to the White House and millions of other books are available for Amazon Kindle. Learn more

# Predicting the Next President 2016th Edition

by Allan Lichtman (Author)  
★★★★☆ 17 ratings

Look inside



ISBN-13: 978-1442269200  
ISBN-10: 1442269200  
Why is ISBN important?

Have one to sell? Sell on Amazon

Add to List

Share

- Kindle \$23.29
- Paperback** \$20.52 - \$26.00
- Other Sellers See all 5 versions

Buy used ✓prime \$20.52

**Buy new** ✓prime **\$26.00**

FREE delivery: **Wednesday**  
Order within 7 hrs 15 mins [Details](#)

📍 Munmun - Smyrna 30080 Qty: 1

**In Stock.**

Ships from and sold by Amazon.com.

✓prime

**Note:** Available at a lower price from [other sellers](#), potentially without free Prime shipping.

[Add to Cart](#)  
[Buy Now](#)

**More Buying Choices** 33 used & new from \$11.24

16 New from \$24.83 | 17 Used from \$11.24

[See All Buying Options](#)

**Note:** Available at a lower price from [other sellers](#), potentially without free Prime shipping.

In Predicting the Next President political analyst and historian Allan J. Lichtman presents thirteen historical factors, or "keys" (four political, seven performance, and two personality), that determine the outcome of presidential elections.

[Report incorrect product information.](#)

Prime Members choose one FREE Kindle book each month

# The professor who correctly predicted elections for the past 30 years...

<https://www.washingtonpost.com/news/the-fix/wp/2016/09/23/trump-is-headed-for-a-win-says-professor-whos-predicted-30-years-of-presidential-outcomes-correctly/>

## Trump is headed for a win, says professor who has predicted 30 years of presidential outcomes correctly



Who will win the 2016 presidential election? This professor has predicted correctly for 32 years

Allan Lichtman created his "13 Keys to the White House" more than 30 years ago – and he's ready to predict who will win in 2016 (Peter Stevenson/The Washington Post)

# Ideal Prediction Engine

- Predict (nearly) anything – General purpose
- Accurate
- Real-time
- Cost effective
- Rewards information, not
  - Raw computational power
  - Persuasion, power, conviction



# Combining Various Data

## Passive

- Fundamental: E.g. statistical model based on past election results, incumbency, presidential approval ratings, economic indicators, etc.
- Social media: Twitter, Facebook
- Corporate: search, page-views, comments, etc.

## Active

- Polls
- **Prediction Markets**
- Experts

# Communicating Predictions

Prediction = Probability

- ✘ *Hillary Clinton will win the 2016 US Presidential election*
- ✔ **71% chance** *Hillary Clinton will win the 2016 US Presidential election*

The papers we read primarily use retrospective data for prediction. Essentially they all “predict the past”. How to fix this problem if we want to actually predict future events?

# Reflections

- Tough to generalize successes – specific cases, particular platforms. (How) would this work for:
  - Multi-state process (e.g., US Primaries)?
  - General elections?
  - Surveillance state/government censorship?
- Despite ongoing challenges, social media will surely play a key role in the future of accurate election prediction