KASICH RISING
(And other highlights of the Presidential Debate Conversation)
November 9, 2015

A PEORIA Project Report
Associate Professors Michael Cornfield and Lara M. Brown
Data Scientist Jamie Chandler
The Graduate School of Political Management
The George Washington University

In Partnership with Zignal Labs

INTRODUCTION

Welcome (back) to the **PEORIA Project**, the GSPM's harnessing of Zignal Labs's realtime, cross media story-tracking platform to analyze the "public echoes" arising from the 2016 presidential campaigns. While others during the "invisible primary" will investigate each candidate's poll standing, dollars raised and spent, and endorsements won, we track and measure **words**—the chatter about the candidates and the echo of their campaign messages in both mainstream and social media.

PEORIA is an acronym for **P**ublic **E**choes **O**f **R**hetoric **I**n **A**merica, chosen as an allusion to the old vaudeville and marketing phrase "will it play in Peoria?"

Our fundamental premise is that how candidates and their messages play on the trail with the media and the public both affect and reflect the voters' presidential preferences. When a candidate says and stages it right, it resonates positively with the public, creating an echo that benefits the campaign. Of course, the opposite can also occur with negative echoes. From positive to negative, people respond to crafted messages,

In this, our fourth report, we analyze the presidential campaign conversation in mainstream and social media between September 21 and November 1, 2015. This time period featured three events in which talk about confrontations involving some of the candidates could have advanced or reversed their electoral fortunes: two debates (Democratic, October 13; Republican, October 28) and Clinton's testimony before the House Select Committee on Benghazi (October 22).

The accompanying slides may be found here. They are discussed below in the order they appear.

METHODOLOGY

Here we present information on the Zignal Labs platform and our own metrics as applied to the data the platform contains.

A) Zignal Labs

What does the data universe contain?

SOCIAL MEDIA: Every single tweet, every single mention in social/online video (YouTube, Vimeo, MediaBistro), 30+ million blogs.

NEWS and MAINSTREAM MEDIA: news stories from more than 100,000 online outlets including licensed content, all LexisNexis News Content (print news, magazines/journals/newspapers,etc), all television closed caption content from 900 channels in every media market in the US.

Who qualifies as a presidential candidate?

The database indexes information about major candidates whose names appear in the leading polls and political insider publications. For the period March 15 to May 15, 2015, eight candidates announcing within that time period were analyzed as a set. For the period May 16 to July 20, 2015, the twelve candidates announced within that time period were analyzed as a set.

What counts as a mention?
How is share of voice calculated?
How are sentiment classifications (positive, negative, and neutral) determined?

A mention consists of a tweet, news story, blog, video, LexisNexis story or broadcast clip (closed-captioning) that matches a query (a query is a combination of certain keywords or phrases). For this project and related ones, Zignal has built a custom database with real-time

continuous queries of the presidential candidates' names. Multiple mentions within a content unit or "document" are not counted extra.

Share of voice is calculated by summing up the mentions (across all media types) in each candidate profile and taking the ratio of that candidate's total to the entire set of candidates.

Sentiment is determined using natural language processing technology (NLP). Zignal's NLP algorithm assigns a positive, negative, or neutral score to every document to provide an overall sentiment rating. Frequency, intensity, and sentence structure are factored into the model. For example, "love" has a higher score than "like", but an overall negative prediction will still occur if negations such as "not" or "neither/nor" are present within the sentence. Adverbs also serve as multipliers, with phrases like "very good" scoring higher than "good." The backbone of this algorithm is a Recursive Neural Tensor Network, a type of deep learning algorithm that allows us to continually modify and fine-tune our model as time goes on. Unfortunately, sentiment detection is still not an exact science, and NLP fares poorly when sarcasm is present or the overall diction is ambiguous.

Over the course of the project, GWU has the opportunity to manually override and/or correct sentiment which helps train and improve the models' performances. In addition, the project only reports "**net sentiment**," positive less negative or vice versa as the case may be. This move assumes that erroneous classifications are randomly distributed, and that the directionality of sentiment is a fairer albeit thinner indicator than reporting percentages from all three categories.

How are other indicators determined?

Popular Tweets: number of retweets that a tweet gets

Top Issues: a second/third level of filtering. Profile queries (the candidates' names) control what gets ingested into the platform, and issues are tags that categorize the data ingested. Top issues is thus a sorted list of the most frequent tags by candidate.

B) The PEORIA Project

What is an echo?

We use the word "echo" to denote the aggregate verbatim appearances (i.e. mentions) in the database of words and phrases issued by a campaign. An echo is not the same as "resonance," the absorption of campaign messages into the mindset of listeners and speakers.

What are Announcement Echo Dates?

The number of days between the mentions peak on announcement day and the leveling off of mentions before any subsequent rises.

What is a Web Site Share?

A mention that includes a url of a page on one of the campaign's web sites. A candidate's web site serves as a hub where the campaign tries to convert attention or interest into favorable attitudes and enduring support, particularly through the use of landing pages to collect visitors' e-mails and other contact information and to solicit and collect donations.

What are the mainstream and social media?
See the data universe answer under the Zignal Labs section.

What is the GSPM Echo Rating?
A summary assessment combining the objective data described above and in the slides with our subjective sense as longtime students of presidential campaigns (slide 27).

If the candidacy announcement is the political equivalent of a fanfare, then the echo varies from:

11 - Historic. *When the books are written and standards are invoked for future announcements, this will be up there.*
7 - Memorable. *Recall association with the candidate; effective branding of the person in the race of these times.*
4 - Noticed. *The intended brand echoed in the news and social media spheres.*
1 - Crickets. *Virtually imperceptible echo in society, albeit searchable as reaction data, and therefore more than zero.*

FINDINGS

Part One - Presidential Debate Chatter

The Democratic and Republican debates have attracted large audiences, prompting considerable discussion. Since the debates have no scoring system, anyone can have a say in deciding who's won and who's lost. (The same goes for Clinton's Committee hearing appearance in so far as it affects the election.)

We looked at the chatter among the judges made prominent by mainstream media --journalists, pundits, the candidates and their spin doctor spokespersons and surrogates-- as well as judges who surfaced in social media.

First, however, an overview of the time period containing these events. (See the Slide "Conversation Leaders.") The field of candidates divides into three tiers with respect to how many times each name surfaced in the media, and the share of voice each commanded relative

to the others. Donald Trump could have been classified in a tier unto himself; his name appeared more than twice as often as the other two candidates we grouped with him, Hillary Clinton and Bernie Sanders. These three accounted for nearly seven out of every ten candidate mentions in the discourse. For all the infinite technical capacity of contemporary media, a power law or oligopoly of attention prevailed.

In the second tier we placed Ben Carson, Ted Cruz, Jeb Bush, and Marco Rubio. This quartet of Republicans combined for a 33.7% share among those competing for the GOP nomination. Trump netted 54%, leaving the remaining nine Republicans to scrap over 12% of mentions. (We did not include the four candidates from both parties who dropped out in this time period.)

The slide "Trump Still Dominates" displays the distribution in a bar chart. The ensuing slide, with pie charts for each party, shows how Clinton and Sanders exercised a veritable duopoly of attention among declared Democratic candidates and even undeclared, with Joe Biden accounting for a 7% share of voice as the speculation about him climaxed on October 21 with his announcement that he would not run.

The slide "Democratic Time Series" illustrates the shares chronologically. It shows that Clinton's appearance before the Benghazi Committee, while a notable subject of conversation, garnered her 300,000 mentions compared with 700,000 on debate day...and Sanders topped 800,000 that day. Neither event created a lasting buzz in public conversation.

Turning to net sentiment, a calculation of positive less negative inflected mentions with neutrals excluded we see a marked change before and after the Democratic debate only in the case of Biden. The natural language processing algorithm does not adequately distinguish between the negative of disappointment and anger. But clearly a significant number of commenters were not happy with the decision.

The "Republican Time Series" slide discloses a fascinating phenomenon: both Trump and Mike Huckabee attracted more media attention for live-tweeting the Democratic debate than while participating in the Republican debate. Ahead, we will excerpt several of the most re-tweeted Huckabee comments as the Democrats debated. For now, we note that this result reflects the convergence of two social trends in America today: partisan polarization and trolling, or online expressions of hostility directed at another person.

On the slide "Republican Conversation" in order to show changes in mentions in the periods before and after the Republican debate, we literally had to take Trump off the chart. As might be expected, Ted Cruz and Marco Rubio fared quite well on this metric, enjoying sizeable jumps after their stand-out debate performances. Ben Carson remained high in mentions, although in the short post-debate period we include in this report he dropped below Cruz and Rubio (a status unlikely to have lasted much longer in view of the current controversy over the veracity of some of his published and spoken statements). John Kasich and Chris Christie also won greater shares in media conversation after the debate.

"Republican Net Sentiment" disclosed that Jeb Bush and Christie garnered more negative than positive comments both before and after the debate. Bobby Jindal plunged into the negative column, and Kasich and Carly Fiorina also dropped considerably on this metric. None of the Republicans showed improvement, not even those conventionally deemed winners.

When we differentiated Mainstream and Social Media, Clinton and Sanders performed similarly throughout the period. As in previous reports, mainstream media attention was distributed slightly more equitably than in social media, perhaps reflecting the presence of a journalistic "equal time" norm. James Webb generated some social media traffic with his departure from the contest, more than Lincoln Chafee did and indeed more than Martin O'Malley, who did not depart. See the three "Democrats: Who is Talking About Whom" slides for this data.

On the Republican side, being in the undercard debate led to much fewer mentions in both mainstream and social media. Trump's huge advantage was dampened considerably in mainstream media, with a 33% share compared with 58% in social media. Removing Trump from the third "Republicans: Who is Talking About Whom" slide shows that Bush dropped from third to fourth after the GOP debate.

We focused on a week's time span in which Biden made his announcement and Clinton testified to Congress. As can be seen in "The Week That Was For Democrats," the Biden news won more mentions with little run-up or aftermath, but the longer anticipatory and discussion ramps around Clinton's testimony results in a larger share of voice about her. Sanders lost mainstream media oxygen for the duration, but retained social media presence. Clinton's attention gained a boost from her joint appearance with Katy Perry and the candidate's extensive campaign network of supporters, who tweeted in solidarity and even triumph regarding her testimony. That said, as depicted in "Clinton's Top 3 Tweets During Hearings," the network was not all that active during the daytime hours of the live appearance before the committee. Indeed, neither were the Republicans, as shown in the tiny retweet numbers on the slide "Top GOP Responses to Benghazi Hearings.'

Part Two - The Candidates and Social Media Activity

Talk is one thing, talk that leads to campaign engagement and action another. What we call the "echo conversion rate" equals the total number of social media mentions with campaign web site shares in them as a proportion of the total number of social media mentions. It indicates the presence of voices attempting to bring people into the campaign network. As the question mark in the title of slide "Trump on Top? Echo Conversion Rates" suggests, Trump's campaign has not used this engagement tool; far from it. The leader here is Cruz. Kasich and Fiorina also impressed during this time period, while Clinton and Sanders underperformed relative to previous time periods.

The next slide, "Democratic Candidates Social Media Activity," has its main story along the vertical axis: the scale of social media mentions is five times larger on the right-hand chart, for

the post-debate period, than on the pre-debate left-hand chart. The proportions of the candidates to each other remained roughly the same, but the volume skyrocketed.

Another indicator of social media engagement lies in the retweet. As we have seen, candidates tweet during the debates of the opposing party as well as their own. This presents the possibility of interesting and significant contrasts. For the "CNN Debate: Democrats Top Responses" concerned criminal justice, suggesting the involvement and/or courting of #blacklivesmatter. On "Huckabee Live-Trolls Democratic Debate," as noted earlier, Huckabee enjoyed success by commenting astringently as the Democrats performed. The tactic does bring out opposition as can be seen in the fifth retweet captured in the screen shot on the right side of the slide, a comeback at Huckabee in kind. The main point, however, lies in the high volume of #demdebate hashtages featuring Huckabee's at-sign tag.

However, Huckabee did not win a spot in the "Top 3 Retweeted GOP Responses" to the Democratic Debate. As usual, Trump swept the honors.

Pulling back to compare both of our social media engagement metrics, we can see on the slide "Republican Candidates Social Media Activity" that Cruz was way ahead of all his rivals, including poll leaders Trump and Carson. Kasich registered second in the long pre-GOP debate period.

The next slide, "Kasich Works His Network," displays evidence of Kasich's attuning to a campaign strategy in social media mentions. A word cloud has "results now," a campaign slogan, prominent. All five of Kasich's top hashtags reference his campaign and the New Hampshire primary (FITN is an acronym for First In The Nation). These are much stronger results than for Bush, Christie, and Rubio, the candidates regarded as Kasich's main rivals on the insider track of the race for the nomination.

During the GOP debate, Ted Cruz's castigation of the media wound up fourth in terms of retweets. A photo of Bush's personalized cowboy boots finished fifth, in an effort to brandish the candidate's macho side akin to his remarks about playing fantasy football. Kasich fared poorly on the retweet during debate metric. As for the Democrats, Clinton's comments topped Sanders during the GOP debate.

We conclude this fourth report with our overall rankings.