

# Noisy Candidates and Informative Politicians: Analyzing Changes in Tweet Behavior using Tweet Quality Assessment Framework

Sikana Tanupabrunsun<sup>1</sup>, Jeff Hemsley<sup>1</sup>, Bryan Semaan<sup>1</sup>, Jennifer Stromer-Galley<sup>1</sup>  
<sup>1</sup>Syracuse University

## Abstract

Politicians use Twitter as a strategic tool for campaigning and posting messages once elected. Our work focuses on the ways U.S. state governors' use twitter differently when they were campaigning vs. after they have taken office. Our data consists of tweets posted by winning gubernatorial candidates during and six months after the 2014 elections. Using regression analysis, we find that post-election tweet volume is related to factors such as pre-election tweet volume, incumbency status, and if they were a third party candidate. We also develop and utilize a novel Tweet Quality Assessment Framework (TQAF) to show that during elections politicians try to engage topically with their audience more than once elected, but tend to produce higher quality information once in office. Our work contributes to the understanding of politician's use of Twitter. We also believe our TQAF will be useful for researchers wishing to compare differences in tweet behavior across time or groups.

**Keywords:** Elections; Twitter; Retweets; Politicians; Media Richness

**doi:** 10.9776/16235

**Copyright:** Copyright is held by the authors.

**Contact:** stanupab@syr.edu, jjhemsle@syr.edu, bsemaan@syr.edu, jstromer@syr.edu

## 1 Introduction

While Barack Obama is considered to be the first politician to successfully harness social media as a strategic tool for his presidential campaign in 2008 (Carr, 2008), other politicians have quickly followed suite<sup>1</sup>. Twitter, for example, has gained a great deal of traction among politicians as a platform to inform and interact with their audience. For example, in 2014, 36 states in the U.S. held elections for state governors. Between September 14th and November 11th, 2014, 78 U.S. gubernatorial candidates posted a total of 35,639 tweets. Certainly some candidates posted many more tweets than others. Additionally, some candidates appeared to create richer content by taking better advantage of the affordances of the Twitter platform (e.g. @mentions, hashtags, embedded URLs, sentence construction).

In this work we are interested in understanding how candidate's use of Twitter differs after a successful bid for office. That is, once in office, in what ways are U.S. state governors' tweeting differently? According to Coleman's direct representation concept (Coleman, 2005), interaction between voters and representatives should be of "on-going and permanent nature" throughout their services, not only the election period. As one would expect, we find evidence that politicians tend to tweet significantly less after being elected, but we are also interested in identifying factors related to post-election tweet volume, as well as identifying how the richness of their tweets may change once in office.

To the best of our knowledge, no current literature exists that focuses on comparing pre and post-election Twitter behavior of elected officials. To answer our questions, we collected tweets posted by gubernatorial candidates before the election and then again six months after the successful candidates took office. By employing regression analysis to identify factors related to post-election tweet volume, our work contributes to the growing body of literature concerned with politician's use of Twitter by identifying factors related to the ways their behavior differs after a successful bid for public office. We also develop, utilize and evaluate a novel *Tweet Quality Assessment Framework* (TQAF) that we show is useful in measuring tweet richness and use the framework to dig deeper into how our successful candidates tweet differently after becoming elected officials. We expect this framework to be useful for comparing changes in tweet quality over time for a specific user, and for comparing users or groups of users in large corpuses of tweets. For example, in future work we hope to examine how tweet quality differs between political actors, scholars, media personalities and pop celebrities.

---

<sup>1</sup> We note that in 2004 Howard Dean was the first candidate for presidency who raised money for his campaign using the technique of micro payments via the World Wide Web, which could be considered 'social media'. Our intent here is just to highlight the growing use of platforms like Twitter among politicians.

## 2 Background Study

### 2.1 Political engagement with Twitter

The use of twitter by politicians is now commonplace. At its best, it provides public with an opportunity to directly interact with, and engage in, political discourse with politicians and elected officials (Ausserhofer & Maireder, 2013). A common, recent research theme has been to study the tweet behavior of specific sets or types of politicians over a single finite period of time. Conway, Kenski, & Wang (2013) studied Twitter use by the presidential candidates of the 2012 primary election in the United States. They found that the most active Twitter users in their data were not from the major Democratic and Republican parties, but alternate parties, such as the Green Party or the Libertarian Party. Christensen (2013) suggested that while candidates from alternate parties suffered from limited support and resources, social media platforms like Twitter offered them opportunities to gain attention and move towards the political front. Compared to the mainstream candidates, these “third party” candidates tended to create the highest Tweet volumes during the debates. Of course, a large number of tweets is not a measure of audience engagement. Christensen (2013) found that candidates that employed hashtags in creative ways tended to have higher rates of audience engagement, as measured by how often they were retweeted.

Of course, use of Twitter for political reasons is not exclusive to U.S. politicians seeking election. Candidates of the minor Liberal Party in the U.K. utilized Twitter to promote themselves during the 2010 general election more than their major party counterparts did (Graham, Broersma, Hazelhoff, & Haar, 2013). In Australia, politicians are generally “noisier” than the public in that they tend to broadcast more than convers (Grant, Moon, & Grant, 2010). They also tend to cluster by party, forming relatively small, ‘small-world’ networks. Likewise, recent scholarships exploring the tweets of members of congress (Golbeck, Grimes, & Rogers, 2010; Hemphill, Otterbacher, & Shapiro, 2013) find that they use social media as a broadcast mechanism, rather than as a mechanism for interaction with constituents. However, Ausserhofer & Maireder (2013) examined the network formed by Austria’s most engaged political Tweepsters and found that while the network is dominated by political elites, significant activity does come from the public.

Collectively these studies suggest that 1) While politicians are generally quite active on Twitter, those with few options to engage with the public tweet more, 2) Interaction with the public, via Twitter, is the exception, not the rule and 3) Tweets that are richer in content (hashtags) tend to gain more attention from the public, as measured by retweets. However, all of the above studies present a snapshot in time of political actors’ tweet behavior. None provide insight specifically into the differences of campaigning vs elected politicians, nor do they show how a political actor’s tweet behavior changes over time. Coleman (2005) suggests that in an ideal conversational Democracy, politicians should maintain a consistent level and quality of interaction with the public throughout their services. Thus, we are interested in whether or not their Twitter use is consistent throughout their transition from campaigning to political public service.

More specifically, our first set of research questions is:

- 1) Are user’s pre-election tweet volume related to their post-election volume?
- 2) Do post-election tweet volumes differ by party?
- 3) Do post-election tweet volumes differ by a candidate’s election incumbency status?

### 2.2 Content Richness and the Quality of a Tweet

In general we tend to think of the *richness* of a tweet being a reflection of the presence of *content features* (Suh, Hong, Pirolli, & Chi, 2010). Content features are characteristics of a message. For tweets, these content features could be hashtags, URLs or @mentions. Of course each of these content features have a different purpose in a tweet. Hashtags are one key way in which Twitter users engage with specific current topics and signal the context within which the tweet occurs (Boyd, Golder, & Lotan, 2010; Huang, Thornton, & Efthimiadis, 2010). An embedded URL can be thought of as an information resource (Bennett, Segerberg, & Walker, 2014). And @mentioning someone else can be seen as a kind of interaction with other users (Boyd et al., 2010). Retweeting can also be thought of as interacting with others (Boyd et al., 2010) as well as a form of information diffusion (Kwak, Lee, Park, & Moon, 2010; Petrovic, Osborne, & Lavrenko, 2011).

Content features can also be textual elements within the text of the tweet, such as the presence of question marks and explanation points. Such features are what Naveed, Gottron, Kunegis, & Alhadi (2011) refer to as *low-level content-based features*. For them, *low-level features* also include the sentiment of the text, as measured by the presence of specific keywords. In their work looking at which content features were more likely to be retweeted, they found that a tweet is more likely to get retweeted

when it includes hashtags, usernames, URLs and question marks. Moreover, a tweet with strong sentiment, i.e. a tweet with positive or negative strong words, is more likely to be retweeted. On the other hand, they found that a tweet is less likely to get retweeted if it is a direct message (@mentioning others) and includes exclamation marks. *High-level features* are associated with topics of broader public interest and can be identified using topic modeling (Naveed et al., 2011). The more a tweet is associated with a current topic, the more likely it is to be retweeted.

Others (Purohit, Ruan, Joshi, Parthasarathy, & Sheth, 2011; Suh et al., 2010) have also found that the presence of content features such as hashtags and @mentions are related to the number of times a tweet is retweeted. Purohit et al. (2011) also found that how relevant an embedded URL is also impacts retweetability. Thus, there is a relationship between the presence of content features and the likelihood that a tweet will be retweeted. Coletti (2013) suggests that *high quality tweets* are those being rich in the content features mentioned above. But there are certainly many other ways in which we could define quality in a tweet. For example, we could measure the quality of tweet text by how well the message can be understood by a human. Becker, Naaman, & Gravano, (2011) define a high-quality tweet as a “crisp, clear, and effective text that is easy to understand” and a low-quality tweet an “incomprehensible text with heavy use of short-hand notation, spelling and grammatical errors, and typos”. The kinds of textual problems present in low quality tweets can be measured using the out-of-vocabulary (OOV) score developed by Luo, Osborne, Petrović, & Wang (2012). Becker et al., (2011) found that, like the content features above, high-quality tweets, using their definition of quality, were retweeted more than lower quality tweets.

We suggest that a useful metric for tweet quality combines both the understandability of the message as well as the various ways tweets can be rich. Looking at the different kinds of content features we identify three dimensions of quality. We think of hashtags as topic markers that reflect *Contextual richness*, and @mentions, @replies and retweeting someone else as *Interaction richness*. *Information richness* captures the understandability of the text as well as the presence of URLs. In the methods section we provide more detail about how these are combined into our TQAF, but we note that any useful combination of these different dimensions should provide a straightforward way to make comparisons of tweet quality and should be testable. That is, we should be able to evaluate whether or not a high score from the TQAF reflects some measure of quality. Since tweets with content features and tweets that facilitate human understanding have been found to be more likely to be retweeted, we believe that a correlation between TQAF score and number of retweets should validate the framework.

Note that the vast majority of tweets are never retweeted (Boyd et al., 2010; Kwak et al., 2010). In the Twitter context, a retweet can be thought of as a measure of tweet’s success and is a reflection of what community is interested in or find worth talking about (Graham et al., 2013). To retweet is to make a decision to forward a particular message into a user’s own network. Alternately, it can be conceived of as a measure of tweet’s *interestingness* on a global scale (Purohit et al., 2011). In studying retweeting, Purohit, Ruan, Joshi, Parthasarathy, & Sheth (2011) suggested that “content is engaging by its quality and nature”. Thus, we believe that verification of TQAF using retweet rates is a robust form of validation.

### 2.3 Application of TQAF

As outlined in section 2.1, one of goals in this work is to understand how tweet behavior may change once a candidate becomes an elected official. Using our TQAF we can explore the ways in which different groups, or actors over different timeframes, tweet differently. By either taking the average score of tweets for a given a group or actor, we can get a single number indicating tweet quality, or we can look at the different dimensions that make up the TQFA as a way to compare the ways usage is different among groups or has changed over time. Thus, our final set of questions is:

- 4) Do elected officials create higher quality tweets than they did as candidates?
- 5) Along what dimensions of TQAF do the pre- and post-election tweet contents differ?

## 3 Analysis

We examine how governors use Twitter differently before and after election to office with 2 comparative analyses. Both are conducted on 2 sets of tweets collected a) during the election cycle when they were candidates; and b) six months into their tenure as an elected governor. In the first analysis we construct a regression model to predict their post-election tweet frequency using their pre-election frequency and a number of control variables. Our intent with this model is to answer research questions 1, 2 and 3, noted above. The second analysis employs our TQAF that measures the quality of tweets. This work is intended to answer our remaining research questions. The results from two analyses help us compare the tweeting behaviors of elected officials.

### 3.1 Regression Analysis

#### 3.1.1 Dataset

To make our before and after collections we collected two sets of tweets. We employed an open-source toolkit (Hemsley, Ceskavich, & Tanupabrungsun, 2014), that collected tweets from Twitter's streaming API. We used the "follow" parameter of the Streaming API ("The Streaming APIs | Twitter Developers," n.d.), which allows developers to collect real-time tweets from a specific user or set of users.

The first collection contains tweets from 72 candidates running in the 2014 U.S gubernatorial elections. We retrieve the list of candidate names, their party, incumbency status and if the race was competitive (toss-up) or not from [www.RealClearPolitics.com](http://www.RealClearPolitics.com). For each of these we manually identified their Twitter accounts. We found that some candidates were using more than one Twitter account: a personal and a campaign accounts. Campaign accounts typically noted in their profile description that were an official account, and that tweets were posted by campaign staff. For each candidate we merged both person and campaign accounts under the assumption that candidates and elected officials may both maintain social media management personnel. This collection spans 51 days from September 15th 2014 to November 4th 2014 and contains 34,021 tweets created by our candidates. The collection was terminated at the midnight of the election date.

For the second collection, we identified the list of winning candidates. Of these now elected officials, 33 were still active on Twitter six months after taking office. Again we combined related accounts and note that some governors abandoned campaign accounts and created new accounts. We initiated our collection on June 5<sup>th</sup>, 2014 and collected until July 25<sup>th</sup>, 2014: a period of 51 days, the same duration as collection 1. This second collection contains 3,259 tweets. Note that for our analysis we filtered tweets out of collection 1 from the candidates who did not win the election, leaving us 11,617 tweets for our final count in collection 1. Table 1 provides frequencies for different categories of the data sets.

	Republican			Democrat			Third Party		
	Users	Before	After	Users	Before	After	Users	Before	After
Incumbent	16	4,775	1,143	7	2,507	534	0	0	0
Challenger	6	3,476	881	3	833	627	1	26	74

Table 1. Tweets distribution of 2 collections by incumbency status and party

#### 3.1.2 Analysis

We start our analysis by initially operationalizing the 'tweet behavior' for each elected governor as tweet frequency and construct a multivariate regression model to help us identify some of the related factors. Regression analysis examines the relationship between a dependent variable and the explanatory variables by estimating how much of the variance in the dependent variable can be explained by each independent variable in the model (Faraway, 2004; Kahane, 2001). Thus, regressions are said to estimate the relationship of each independent variable to the dependent variable while holding the others constant. Thus, this type of model (assumptions for validity discussed below) allows us to simultaneously examine the relationship between all of our variables of interest and the dependent variable. For our model, the number of tweets for each governor in the second dataset is our dependent variable (*Gov\_Freq*). The model includes one independent variable for each of the first three research questions as well as control and interaction variables as described below.

##### *Model Variable Selection*

We identify 8 variables potentially related to *Gov\_Freq* and perform bi-direction stepwise regression (Kahane, 2001) on variables to select the optimal subset. Stepwise regression starts with an initial model and iteratively adds and removes one variable at a time in an attempt to choose the subset that maximizes the adjusted R-Squared value. Specifically, if variables are not related to *Gov\_Freq* they will lower the adjusted R-Squared value and thus, as suggested by the stepwise technique, should be removed. However, its performance is affected by the order of adding and removing variables. As such, we mitigate this effect by using bi-directional approach (Faraway, 2004), meaning removed variables are retested in different orderings. In the stepwise process we allow and test for the interactions of up to 2-variables, resulting in a final model that includes both main and interaction effects. All interaction are multiplicative, meaning we multiply two main effects variables to create an interaction variable. Due to space limitations we only report on the final model.

Note that regression assumes variables are normally distributed. To satisfy this assumption we transform our continuous variables using a logarithm function and perform Shapiro-Wilk normality test (Shapiro & Francia, 1972) to validate that the transformed variables are now normally distributed. We

provide information about our initial set of variables below. Independent variables addressing research questions are noted as such, others are control variables.

Candidate Freq (RQ 1): To answer RQ1, we include frequency of pre-election tweets to analyze its relation to post-election tweet volumes. The relationship shows whether or not Twitter use is consistent throughout their transition from campaigning to political public service. Logarithm transformed continuous variable ranging from 2.71 – 7.68

Gov Party Dem, and Gov Party Third (RQ 2): For RQ2, we create 2 dummy variables for party to analyze if Twitter uses differ by party. As Williams & Gulati, (2012) suggest, party is the main driver leading to an adoption and extensive usage of Twitter in political sphere. Democrat (Gov\_Party\_Dem = 1), Third party candidate (Gov\_Party\_Third = 1) or Republican (both are 0).

Gov Inc (RQ 3): We also ask whether or not post-election tweet volumes differ by their incumbency status during the election. We hypothesize that challenger governors feel more need to reach out to their constituents on Twitter. Incumbent (1) or Challenger (0)

State Twitter User: We assume that in states with larger numbers of Twitter users, governors will tweet more frequently. Data obtained on July 3<sup>rd</sup> 2015 from Twellowhood ([www.twellow.com/twellowhood](http://www.twellow.com/twellowhood)), which is a directory of Twitter users based on locations listed in their profile. Logarithm transformed continuous variable ranging from 8.49 – 12.51.

State Pop: Like above, we assume governors in states with higher populations may be more active on Twitter. Data obtained from [www.census.gov](http://www.census.gov) on July 3<sup>rd</sup> 2015. Logarithm transformed continuous variable ranging from 13.35 – 17.47

Gov Follower: We assume that governors with higher numbers of followers are more active on Twitter. Logarithm transformed continuous variable ranging from 0.69 – 3.95

Gov Account: Newer accounts may have lower numbers of followers than older accounts. Boolean indicates the governor created a new post-election account. New (1) or False (0)

Gov Race: We assume that governors from more competitive races will need to interact with the public more (work harder), and thus tweet more. Tossup (1), not competitive (0).

### Linear Regression

The optimal model, shown in Table 2 below, includes 8 main effect and 5 interaction effect variables. Bold variables are significant beyond the 0.05 level. The multiple R-Squared value of 0.79 (adjusted R-Squared of 0.64) indicates the data fit the model reasonably well. Note that based on our development of an optimal model, we kept several non-significant variables in the model.

One of the assumptions for regression analysis is that residuals are normally distributed. We perform the Shapiro-Wilk normality test on our model's residuals with the null hypothesis claiming that data is normally distributed. The p-value of 0.6 indicates no evidence to reject the null hypothesis, in other words, the residuals are normally distributed. We also looked at the variance inflation factors to check if there is multicollinearity among variables; we found that none of the significant effects has a VIF of greater than 4, indicating we need not be concerned about multicollinearity in the model.

Gov_Freq =	Coef.	p-val.	SE.	Gov_Freq =	Coef.	p-val.	SE.
<i>intercept</i>	-2.01	0.38	2.24	<i>Gov_Account</i>	-5.00	0.15	3.31
<b>Candidate_Freq</b>	<b>+0.85</b>	<b>&lt;0.00</b>	<b>0.16</b>	<i>State_Twitter_User</i>	+0.10	0.60	0.19
<i>Gov_Party_Dem</i>	-0.11	0.77	0.38	<b>Gov_Inc * Gov_Account</b>	<b>+2.20</b>	<b>&lt;0.00</b>	<b>0.73</b>
<b>Gov_Party_Third</b>	<b>3.59</b>	<b>&lt;0.00</b>	<b>1.19</b>	<i>Gov_Party_Dem * Gov_Account</i>	+1.36	0.06	0.69
<b>Gov_Inc</b>	<b>-1.65</b>	<b>&lt;0.00</b>	<b>0.53</b>	<i>Gov_Race * Gov_Follower</i>	-0.77	0.06	0.39
<b>Gov_Follower</b>	<b>+0.55</b>	<b>0.02</b>	<b>0.22</b>	<i>Gov_Race * Gov_Account</i>	-1.14	0.09	0.65
<i>Gov_Race</i>	+2.29	0.08	1.24	<i>State_Twitter_User * Gov_Account</i>	+0.41	0.21	0.32

Table 2. Regression model results

### 3.1.3 Findings

The result suggests the following findings:

- RQ 1: Governors tweet frequency during the election is positively related to their post-election tweet behavior. Note that this holds true even accounting for the other factors we control for in the model.
- RQ 2: Note that both Gov\_Party\_Dem and Gov\_Third\_Party are compared to the base state of Republican. So the finding that Gov\_Party\_Dem was not significant indicates that our model did not find a difference in post-election tweet volume between Republicans and Democrats. However, we did find a difference positive between the Third party candidates and Republicans, indicating our third party candidates, when controlling other factors, tended to tweet more. This finding aligns with other

research (Conway et al., 2013; Graham et al., 2013). Note that our variable construction does not test for a difference between Democrats and third party candidates.

- RQ 3: Our findings suggest that governors who were an incumbent during the election tended to create fewer tweets once elected than governors who were challengers. More work needs to be done to understand this relationship, but it is possible that incumbents feel less need to reach out to their constituents on Twitter. Perhaps incumbents also tweeted less during the election than challengers. Our current model does not provide insight here, but does open the door to more questions.

Interpretations of control variables:

We note that of the main effects control variables, only Gov\_Follower is significant. It is also positive, indicating that elected officials with more followers, or a larger Twitter audience, tend to tweet more. Note that even though State\_Twitter\_User is not significant, its presence in the model helps to control for its effect, such that we can say that governors with more followers tend to be more active on Twitter, even when controlling for the size of the Twitter population. However, a regression gives us no indication of causal direction. It is reasonable to assume that governors who are more active on Twitter will tend to gain more followers.

Interestingly, the fact that State\_Twitter\_User is not significant implies that the size of the active Twitter population within a state does not push elected officials into being more active on Twitter. It is also interesting that our variable for tight election races (Gov\_Race) was not significant. This implies that governors who won tight election races, controlling for other factors, are not necessarily inclined to tweet more. One could make the argument that when a race is tight, and an elected official does not have a clear mandate from their constituents, they might work harder to maintain public support. Of course, since we are only looking at Twitter we can't know what they are doing on other platforms or through the mainstream media.

Of the interaction variables only Gov\_Inc \* Gov\_Account is significant. The implication is that if Governor, who was an incumbent during the election, created a new account after the election, they tended to be an active tweeter. We can easily imagine a case where an incumbent who had not used twitter before this last election, found it useful enough during the election to continue using it after the election.

## 3.2 Quantitative Analysis on Content Richness

### 3.2.1 Dataset

For the second analysis, we turn our focus to the quality aspects of their Twitter use. We use both of the datasets described in the previous analysis but process the data differently. While in the first analysis we built a model to understand factors related to post-election tweet behavior, now we are interested in understanding how the quality of their tweets might be different. As noted above, we measure tweet quality along three different dimensions: contextual richness, interaction richness and information richness. This work requires mining the text of tweets and quantifying their content as richness scores for each governor using techniques we describe below.

### 3.2.2 Analysis: Quality Assessment Framework

The quality dimensions of our framework are contextual, interaction and information richness. We express each dimension of richness as a score, as described below:

*Contextual richness* looks at how governors attempt to bind themselves to different conversations/discussions with the greater public. To find this score we count the number of hashtags in their tweets ( $Count_{ht}$ ) and then normalize to a range of 0 to 1 across the set of tweets.

$$Score_{cont} = \frac{\max(Count_{ht}) - Count_{ht}}{\max(Count_{ht}) - \min(Count_{ht})}$$

*Interaction richness* considers how governors interact with the public on Twitter. We thus count the occurrence of @username in a tweet and the number of times they retweet others. For @username, we count how many Twitter handles were tagged in a tweet,  $Count_{un}$ , but exclude those in retweets and cases where they tagged themselves. Note that by counting @username, we capture both @mentions as a means to mention someone and @replies as a means to reply to another tweet. For retweets RT is 1 if it is a retweet, otherwise RT is 0. Before adding these two variables, we normalized  $Count_{un}$  to a range of 0 to 1 otherwise it overwhelms the RT effect in the formula.

$$Score_{int} = \sqrt{\frac{\max(Count_{un}) - Count_{un}^2}{\max(Count_{un}) - \min(Count_{un})}} + RT$$

*Information richness* considers how informative their tweets are to the public. Note that we do not only look at how much information they provide but also the quality of the text, as measured by the not-out-of-vocabulary (NOOV) ratio discussed in section 2.2. We express information richness with a normalized URL count and the NOOV ratio. We calculate the NOOV of a tweet's text by using qDap (Goodrich, Kurkiewicz, & Rinker, 2015), an R package used to calculate the ratio of misspelling words, subtracting by the tweet length and divide by tweet length. The two scores are then combined with the Euclidean distance. Euclidean distance function is a common approach for projecting multi-dimensional data to a single dimension. In this work, our Euclidean distance represents a combination of URLs and NOOV as a one-dimensional score, which allows us to compare across each candidate.

$$Score_{inf} = \sqrt{\frac{\max(Count_{url}) - Count_{url}^2}{\max(Count_{url}) - \min(Count_{url})}} + NOOV^2$$

For each dataset (pre and post-election tweets), we calculate 3 scores for each tweet then calculate average scores for each governor. For each of the 33 governors, we have 3 richness scores for pre-election and 3 richness scores for post-election. Next, we project the three richness scores for each governor to one dimension with the Euclidean distance function expressed below.

$$Score = \sqrt{Score_{cont}^2 + Score_{int}^2 + Score_{inf}^2}$$

Recall that our 4th research question is: do elected officials create higher quality tweets than they did as candidates? We answer this question with a simple paired t-test on the before and after overall scores. The result of the test is not significant ( $t=-0.53$  and  $p=0.6$ ), indicating that we cannot detect a difference in the mean quality scores before and after election for our governors. However, looking at the individual scores provides more insight.

Research question 5 asks: along what dimensions of TQAF do the pre and post-election tweet contents differ? We answer this question with a series of 3 paired t-tests comparing the averages of  $Score_{cont}$ ,  $Score_{int}$  and  $Score_{inf}$  pre and post-election. The results lead us to several interesting findings. First, our governors created tweets with higher contextual richness when they were candidates ( $t=-2.66$  and  $p=0.01$ ). That is, they worked harder to engage with public conversations or discussions when they were candidates but less so when they are in the office. Second, their tweets when they are in office are more informative than when they were candidates ( $t=6.16$  and  $p<0.01$ ). We further investigate their URL use vs. NOOV ratio, and find that the difference is mainly contributed by the inclusion of URLs; they provide URLs more frequently after the elections. Third, we do not find any statistically significant difference in  $Score_{int}$  of pre and post-election tweets ( $t=0.7$  and  $p=0.49$ ). This means that they maintain roughly the same interaction level with the public regardless of their political status.

### 3.3 Framework Evaluation

This section presents the evaluation of our TQAF on its capability to assess tweets from the audiences' perspective. Our hypothesis is that higher quality tweets are more likely to get retweeted, but as we saw in the last section, it is necessary to look at all three dimensions. Thus, we report the evaluation in 2 levels: How well our three richness scores work together, and how well each score works individually, as measured by how many times a tweet got retweeted.

First, we present an analysis showing that our three richness scores together are good overall measurements of tweet's quality. We do this by breaking the scores into 7 groups: 3<sup>rd</sup>, 10<sup>th</sup>, 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup> and 90<sup>th</sup> percentiles, reporting the average retweets for different statuses of our governors. Figure 1a presents the plot of retweets and projected richness scores grouped by incumbency status. Table 3 presents the average retweets counts of each group of scores. They show that,

regardless of incumbency status, tweets in the third percentile (a low richness score below 1.22) were never retweeted. Beyond this point, the average number of retweets tends to increase by score with a slight decline for tweets in 75<sup>th</sup> - 90<sup>th</sup> percentile. We also notice that when incumbent governors created

%	Inc.	Cha.	Rep.	Dem.	Ind.
0-3	0.00	0.00	0.00	0.00	0.00
3-10	20.35	0.32	15.86	0.65	0.00
10-25	19.25	6.80	17.72	8.24	3.33
25-50	29.44	9.56	28.59	9.38	3.19
50-75	32.38	8.17	28.21	7.81	2.29
75-90	24.75	5.57	18.81	8.21	3.35

Table 3. Average RTs of each score group

high quality tweets (score > 1.22), they are more likely to get retweeted than those created by challenger governors. Specifically, while up to 77.83% of high quality tweets (1,633 tweets) from incumbent governors got at least one retweet, there was only 62.52% of high quality tweets (1,521 tweets) from challenger governors that got at least one retweet. Certainly a limitation of this analysis is that we don't control for how many followers a governor has, but in aggregate the analysis is still suggestive. Moreover, when incumbent governors' high quality tweets did get retweeted, they got higher number of retweets. As shown in Table 3, the average numbers of retweets for incumbent governors are clearly higher than challenger governors' in every group. On average, while incumbent governors got 27.50 retweets per tweet, challenger governors got only 7.25 retweets per tweet.

Interestingly, we found similar patterns for the party attribute as illustrated in Figure 1b. That is, high quality tweets from Republican governors are more likely to get retweeted than those from Democrat and Independent governors. Specifically, up to 71.47% of high quality tweets (1,963 tweets) from Republican governors got retweeted at least once, there were only 67.62% and 66.30% for Democrat and Independent tweets. And for those retweeted tweets, Republican tweets got an average of 23.76 retweets while it was 8.09 and 3.29 retweets per tweet for Democrat and Independent governors.

We conclude that our framework is capable of assess tweets to some extent but there exists a limitation. That is, public figures like politicians are special cases because of their identity and reputation. Thus, one should also take author attributes, i.e. Twitter account attributes like follower count into consideration to predict the retweetability (Purohit et al., 2011). However, we would like to emphasize that our goal is not to predict the retweetability but only to demonstrate the usefulness of the TQAF.

The second level of our evaluation is to look at the performance of each score individually. Figure 2a, 2b and 2c present the plots of the number of retweets by  $Score_{cont}$ ,  $Score_{int}$  and  $Score_{inf}$  respectively. All figures clearly show that high interaction, contextual and information richness tweets are more likely to get retweeted. Again, this does not mean that high quality tweets always get retweeted but it shows that the higher score a tweet has, the higher chance of being retweeted. Additionally, we find an interesting pattern of  $Score_{inf}$ . Most of retweeted tweets are clustered around the score of 1.0 and 1.25 – 1.41. The first cluster is a collection of tweets with either URL or NOOV ratio of 1 (no misspelling). This shows that audiences take the misspelling issue seriously. The second cluster is a collection of tweets with high NOOV ratio. Since all was embedded with URLs, scores are varied by NOOV ratio only. This finding conforms to the previous discussion about the misspelling issue. We conclude that our scores are good reflection of how tweets affect the audiences even when they are used individually. But then again, they are not the only factors of retweetability still they are sufficiently good for assessing tweet contents.

#### 4 What is this *Quality* thing?

Our use of the word "quality" is multilayered and intended to be a loose interpretation of the word. First, we think we are measuring specific qualities or characteristics that can be used to distinguish one tweet

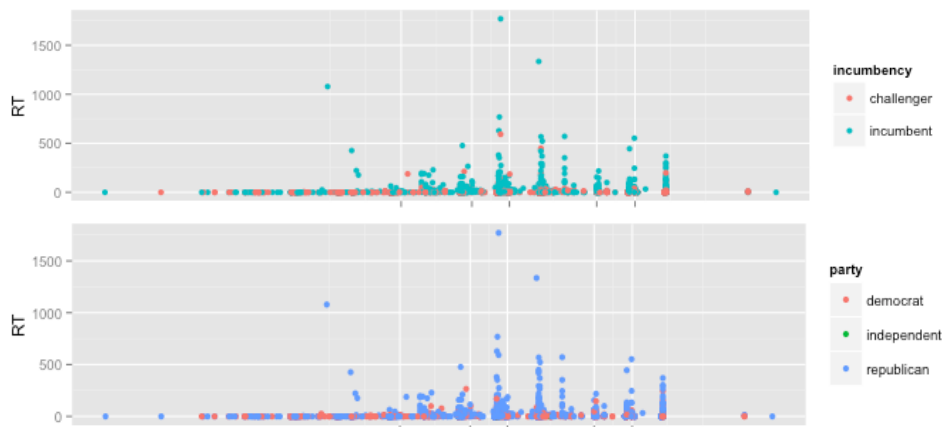


Figure 1. Plots of richness scores and RTs by incumbency status (top) and party (bottom)

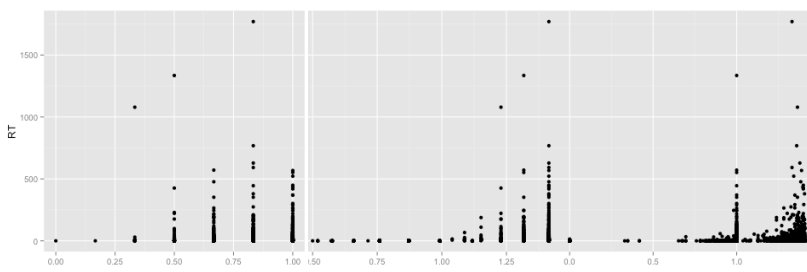


Figure 2. Number of RTs by a. contextual richness (left), b. interaction richness (middle) and c. information richness (right)



from another (e.g. *Contextual richness* and *interactional richness*) of tweets. In this context we think of the qualities as neither good nor bad, but things we can point to that are present. For example, a car may have specific qualities that distinguish it from other cars such as having an air conditioner or sensors in the tire that can warn us of an impending flat tire. The second layer of our use of the word quality is intended to a subjective, though quantitatively derived, measure of how *good*, or rich, a tweet is. Quality in this case reflects how well a message is constructed along a number of different dimensions (or qualities). As an example, consider two movies: one a comedy and one a high production science fiction film. There are many ways we could measure these films and if you prefer comedies you might be inclined to rate the former as being of higher quality than the later. Our purpose in developing this framework is to provide a reasonably objective way to quantify and compare the quality and richness of large sets of tweets, while preserving the different dimensions along which we might measure quality.

For our governors we compared the quality of their tweets at two different time points. Thus it makes sense that we did not find a difference in overall quality scores from before to after the election. But campaigning politicians must certainly create and propagate different kinds of messages than elected officials, and thus we found differences in how they prioritized different qualitative elements of their messages.

We recognize that by using retweets to validate our framework we are possibly falling into the same problem that one faces when picking a movie: do you trust the expert movie reviewer, or number of stars that the crowd assigns? We have chosen the latter under the assumption that Twitter is a crowd environment and individuals in the crowd constantly promotes content, and thus expands the reach of messages, whenever they retweet a message. In other words, retweeting is a kind of voting done by the crowd that indicates what they think is worth talking about: messages the crowd deems as having the qualities worthy of sharing into their own networks. We take heart in the finding that the crowd thinks spelling is important.

Certainly the framework has limitations. There may be other dimensions worth including such as the presence of images; there are other ways we could standardize or weigh the dimensions; we have not yet controlled for the number of followers in our assessment; our work is limited to the Twittersphere. And yet, we think this framework could be quite useful in comparing different sets of actors and different actors over time. For example, how do the tweets of academics differ from politicians, media personalities, business leaders or pop stars? We believe that understanding how actors prioritize the content in their messages can give us clues about what they think is important, how they present themselves or how they hope to be perceived.

## 5 Conclusion

In this study, we are interested in the consistency of governors' Twitter use before and after the elections. That is, we explore the ways in which their post-election tweet behavior is related to and different than their activity while campaigning. Using Twitter data collected during the elections and six months into their terms as elected officials, we examined 5 research questions concerning both their general behavior and the quality of the tweets they posted. We employed a regression model to examine factors related to the post-election tweet volumes and then used the Tweet Quality Assessment Framework to look at how the quality and richness of their tweets changed over time.

Our findings indicate that while candidates who were active tweeters before the election tended to also be active as governors, overall our elected officials tended to tweet less once elected. Of more interest is that we found that third party governor tended to be more active than his counterparts and that politicians who were challengers during the election tended to post more tweets than their incumbent counterparts once elected. We also note that governors with more followers tend to also be more active tweeters. While the nature of our analysis does not suggest a causal relationship, real world experience suggests that people who are more active tend to gain more followers.

Once we apply our TQAF, a more nuanced picture emerges about the changes in their tweeting activity. We found that during elections politicians tended to have higher contextual richness scores than once they were elected officials. Our contextual richness score is related to their hashtag use and so higher scores suggest that they are either attempting to join into ongoing discussions or they are attempting to initiate new discussions. We note that the scores for interaction richness remained unchanged pre and post-election. This is interesting because it suggests that governors tend to maintain the same level, in terms of a percentage of their overall tweet activity, of interacting with others on Twitter in terms of @mentions, @replies and retweeting others. So while they tend to engage in group discussions less, they interact with individuals about the same amount. Once in office, our elected governors do tend to have higher information richness scores than pre-election. This suggests that their

tweets are more informative when they are elected officials. We note that all of these scores reflect a percentage change, not a nominal change. As we stated before, governors tweet far less than candidates.

When we aggregate all three richness scores into a single quality score we find that the tweet quality scores are unchanged. Our work shows that while the overall quality of the messages is unchanged, the nature of the messages are different. Perhaps these differences reflect the nature of politics: networking is far more important for getting elected than informing the public, but once elected, influencing the flow of information, or propagating one's own version of a story, is more important than engaging with the crowd.

## 6 References

- Ausserhofer, J., & Maireder, A. (2013). National Politics on Twitter. *Information, Communication & Society*, 16(3), 291–314. <http://doi.org/10.1080/1369118X.2012.756050>
- Becker, H., Naaman, M., & Gravano, L. (2011). Selecting Quality Twitter Content for Events. In *Fifth International AAAI Conference on Weblogs and Social Media*.
- Bennett, W. L., Segerberg, A., & Walker, S. (2014). Organization in the crowd: peer production in large-scale networked protests. *Information, Communication & Society*, 17(2), 232–260.
- Boyd, D., Golder, S., & Lotan, G. (2010). Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In *System Sciences (HICSS), 2010 43rd Hawaii International Conference on* (pp. 1–10).
- Carr, D. (2008, November 10). The Media Equation: How Obama Tapped Into Social Networks' Power. *New York Times*, p. 1.
- Christensen, C. (2013). Wave-Riding and Hashtag-Jumping. *Information, Communication & Society*, 16(5), 646–666. <http://doi.org/10.1080/1369118X.2013.783609>
- Coleman, S. (2005). *Direct Representation Towards a conversational democracy*. Retrieved from <http://www.ippr.org/publications/direct-representation-towards-a-conversational-democracy>
- Coletti, C. (2013). *Blueprint for a Literate Nation How You Can Help*. Xlibris Corporation.
- Conway, B. A., Kenski, K., & Wang, D. (2013). Twitter Use by Presidential Primary Candidates During the 2012 Campaign. *American Behavioral Scientist*, 57(11), 1596–1610.
- Faraway, J. J. (2004). *Linear models with R* (Vol. 63). Chapman and Hall/CRC.
- Golbeck, J., Grimes, J. M., & Rogers, A. (2010). Twitter use by the U.S. Congress. *Journal of the American Society for Information Science and Technology*, 61(8), 1612–1621.
- Goodrich, B., Kurkiewicz, D., & Rinker, T. (2015). qdap: Bridging the Gap Between Qualitative Data and Quantitative Analysis (Version 2.2.2).
- Graham, T., Broersma, M., Hazelhoff, K., & Haar, G. van 't. (2013). Between Broadcasting Political Messages and Interacting with Voters. *Information, Communication & Society*, 16(5), 692–716.
- Grant, W. J., Moon, B., & Grant, J. B. (2010). Digital Dialogue? Australian Politicians' use of the Social Network Tool Twitter. *Australian Journal of Political Science*, 45(4), 579–604.
- Hemphill, L., Otterbacher, J., & Shapiro, M. (2013). What's congress doing on twitter? In *Proceedings of the 2013 conference on Computer supported cooperative work* (pp. 877–886). ACM.
- Hemsley, J., Ceskavich, B., & Tanupabrungsun, S. (2014). Syracuse Social Media Collection Toolkit (Version 0.1). Syracuse University, School of Information Studies.
- Huang, J., Thornton, K. M., & Efthimiadis, E. N. (2010). Conversational tagging in twitter. In *Proceedings of the 21st ACM conference on Hypertext and hypermedia* (pp. 173–178). ACM.
- Kahane, L. H. (2001). *Regression basics*. Thousand Oaks, [Calif.]: Sage Publications.
- Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web* (pp. 591–600). ACM.
- Luo, Z., Osborne, M., Petrović, S., & Wang, T. (2012). Improving twitter retrieval by exploiting structural information. In *In AAAI*.
- Naveed, N., Gottron, T., Kunegis, J., & Alhadi, A. C. (2011). Bad News Travel Fast: A Content-based Analysis of Interestingness on Twitter. Presented at the ACM WebSci'11, Koblenz, Germany.
- Petrovic, S., Osborne, M., & Lavrenko, V. (2011). RT to Win! Predicting Message Propagation in Twitter. In *Fifth International AAAI Conference on Weblogs and Social Media*.
- Purohit, H., Ruan, Y., Joshi, A., Parthasarathy, S., & Sheth, A. P. (2011). Understanding User-Community Engagement by Multi-faceted Features: A Case Study on Twitter.
- Shapiro, S. S., & Francia, R. S. (1972). An Approximate Analysis of Variance Test for Normality. *Journal of the American Statistical Association*, 67(337), 215–216.

- Suh, B., Hong, L., Pirolli, P., & Chi, E. H. (2010). Want to be retweeted? Large scale analytics on factors impacting retweet in Twitter network. In *Social Computing (SocialCom), 2010 IEEE Second International Conference on* (pp. 177–184).
- The Streaming APIs | Twitter Developers. (n.d.). Retrieved April 23, 2013, from <https://dev.twitter.com/docs/streaming-apis>
- Williams, C. B., & Gulati, G. J. "Jeff." (2012). Social networks in political campaigns: Facebook and the congressional elections of 2006 and 2008. *New Media & Society*, 1461444812457332.