What the Hashtag: Temporal Shifts in Hashtag Position in Twitter

Christopher Mascaro, Drexel University, cmascaro@gmail.com (corresponding author)

Nora McDonald, Drexel University

Alan Black, Drexel University

Sean Goggins, The University of Missouri

Abstract

Hashtags on Twitter are both contextual signals and links to relevant streams of discourse and are often used as selection criteria for identifying a domain of study. In this paper we seek to describe the semantics of hashtag position in Twitter corpora. We show that hashtag positions can vary over time, and that this variance is sometimes a signal of events in the physical world. We consider how differences in hashtag position constitute both differences in meaning and social signaling on Twitter. We also explore how hashtag position varies in combination with the usage of other syntactical features, like retweeting, mentioning and URL sharing. We describe how shifts in hashtags position signals different expressions of information. We also show the difference between hashtag use for information navigation and social hashtag use.

1. Introduction

Twitter users and Twitter itself have developed a set of syntactical features that make information exchange and short topical conversations tractable in a set of 140 character messages. The most widely used syntactical features include at-mentions, hashtags, retweets, and URLs. These syntactical features enable social and information navigation for Twitter users to locate the rivers and tributaries of content shared and linked one tweet at a time.

This discoverability of content and the resulting, dynamically constructed context of Twitter *depends on* Twitter users embedding navigational cues in the content of their message; this is referred to as tagging. Previous studies of tags in socio-technical systems have shown how tags are used not for memory but "primarily for the benefit of others" and as "a social act" (Miller & Edwards, 2007). "Tags" on Twitter have been operationalized as at-reply's at-mention's, URL's or a hashtags (Kaptelinin & Nardi, 2012). Each operationalization of "Tag" is a link to reference information, people or classifications. All of these different syntactical features help to create a different experience for each individual; they are essential markers of context and navigation on Twitter.

Hashtags are framed by prior information science research as a multivalent kind of linguistic component that is used for context development on Twitter (Naaman, Becker, & Gravano, 2011). A good deal of literature examines context on Twitter by focusing on one or several related hashtags associated with discourse or an event (Conover et al., 2011; Shamma, Kennedy, & Churchill, 2010). This literature tends to examine single domains in areas of news, politics (Hu, John, Seligmann, & Wang, 2012; Mascaro, Black, & Goggins, 2012), sports (Chakrabarti & Punera, 2011), entertainment (Doughty, Rowland, & Lawson, 2011), crisis (Starbird, Palen, Hughes, & Vieweg, 2010). These analyses tend to focus on how the hashtag is adopted (Chang, 2010) or temporal occurrences of the hashtag (Lehmann, Gonçalves, Ramasco, & Cattuto, 2012), but do not focus on how the hashtag is used within the tweet and how location varies.

Zappavigna (Zappavigna, 2011) describes hashtags as "searchable talk" and other studies have illustrated the utility of hashtags as a form of conversational tagging (Huang, Thornton, & Efthimiadis, 2010). Hashtags have not, however, been previously operationalized as linguistic markers with different, Tweet-Location-Dependent meanings. Consequently, little is understood about the semantics of hashtag location, and how this linguistic frame might reveal a new dimension for analysis in social tagging research. This linguistic frame adds to existing research that shows hashtags can be used to inject a tweet into a topically focused stream of discourse, interest or concern (Conover et al., 2011). To the person making the tweet, it is a form of audience selection (Marwick & boyd, 2011), which is conceptually similar to socio-technical community selection (Yang, Sun, Zhang, & Mei, 2012).

Hashtags make discourse discoverable. Where the user decides to put the hashtag reflects a linguistic, user choice that, once more well understood, holds promise for unlocking a greater understanding of adaptive use, and social production on Twitter.

In this paper we examine the dynamics of hashtag location within a corpora of US civic and political discourse tweets. We focus on hashtag position across several specific search phrases, draw inferences about purpose and function related to position and consider how differences in hashtag position constitute both differences in meaning and social signaling. These analyses build on the only previous work on hashtag location that we were able to identify (Zappavigna, 2011). This paper builds on prior work by examining the dynamics of hashtag position over time and by operationalizing hashtag position as a linguistic indicator in the context of other hashtags and syntactical features.

Specifically, our findings illustrate that there is a range of dynamism in hashtag positioning on Twitter, and the degree of stasis, combined with the direction of movement from beginning, to middle to end of a set of tweets reflects an external context in the Twittersphere that may not be readily apparent. Hashtags are our focal syntactic feature, but from this center we explore how hashtag position varies with other syntactical features such as URLs, at-replys, at-mentions and retweets. This research is a systematic first step towards developing a framework for further understanding "searchable talk" and how medium specific syntactical features facilitate technological interaction (Zappavigna, 2011).

2. Related Literature

While hashtags can serve a variety of functions – e.g. affiliation (Conover, Gonçalves, Ratkiewicz, Flammini, & Menczer, 2011), part of speech (in-line meta data), a meme and, thus, a kind of community process (Zappavigna, 2011) – they are ultimately a way of finding information that is not built into the technology, but rather is imputed by its users. Because hashtags serve as a technologically specific hyperlink (a "learning affordance") (Kaptelinin & Nardi, 2012) we conceptualize them here as a technological affordance. In that sense, they are a feature of the user experience. In clarifying the meaning of "affordance," Kaptelinin and Nardi (2012) adopt a mediated action perspective that insists that previously understood notions of an affordance in HCI are problematic. When considering tagging technologies (and Twitter in particular) as affordances, they are technologically constitutive and only perceived when made. One must conceive of and affix a hashtag to content for it to be viewed as a technological affordance.

2.1 Tagging Broadly Conceptualized

The practice of tagging spans multiple mediums and media and ranges from personal to social motivations. Ames and Naaman (Ames & Naaman, 2007) study the motivations behind use of tags on the popular photo-sharing service Flickr. They classify social motivations for tagging such as organization of the data as well as communicating to the individuals. In the case of social organization, tagging increases the likelihood that the public will view a user's photos.

There is still a difference in motivations when we move further from mediums in which social and public navigation and personal and public organization is a goal. In one sense, the factors contributing to decisions to tag and what to tag *with* are much more likely to be motivated and coordinated by groups. Sen et al. (Sen et al., 2006) demonstrates that community tagging does influence tag choice in the MovieLens recommender system. Other studies show that people tend to adopt the same set of hashtags based on the community (Cunha et al., 2011).

2.2 Tagging on Twitter

Hashtag use on Twitter is examined in many studies of Twitter since it was established in 2006. This research focuses primarily on how hashtags spread (Romero, Meeder, & Kleinberg, 2011) and on the use of hashtags during events; particularly crisis events (Starbird & Stamberger, 2010). Hashtags first emerged on Twitter during the San Diego Fires in 2007 as a way of tracking relevant information about a natural disaster. In crisis informatics, researchers use hashtags to identify streams of discourse for information gathering and sharing by witnesses on scene. Heverin and Zach (Heverin & Zach, 2011) used hashtags to identify tweets associated with three separate school shootings to understand how information diffused through the network and applied the framework of sense-making to understanding how individuals used this information.

Similarly, hashtags and other shared keywords serve to identify other situations so individuals can discover important information and share information around a specific topic outside the context of their network of existing relationships (Bruns & Burgess, 2011). In disaster situations, greater investment in coherent channels of communication has led to proposals that users adopt hashtags as a “syntax” (Starbird & Stamberger, 2010). The emergent strategic use of syntax hashtags is possibly well understood in the context of an event – especially when it is prescribed.

According to Huang (Huang et al., 2010), a hashtag can serve as either, “a label in the traditional sense of a tag” or a “prompt for user comment”. In the context of trends, the result of this prompt can be “massively-multi-person conversation” (Huang et al., 2010). Huang shows the way in which hashtags in Twitter are used for fundamentally different purposes than tagging in the past. On sites like del.icio.us, tagging artifacts were used for archival purposes and for easy retrospective access, whereas tagging through the use of hashtags in Twitter is for a user to identify and find topically relevant streams of discourse, though they may also be used to make a joke or coin a sentiment intended for spreading. Following certain hashtags can allow users to be made aware of a certain area of discourse that they would not be exposed to by just looking at the individuals that they follow.

Romero et al. (Romero et al., 2011) examine the adoption of hashtag idioms, hashtags that concatenate two or more words (and sometimes abbreviations and numbers) to create a mimetic structure for conversational tweets, such as #dontyouhate, and more well known, #musicmonday and #followfriday. The adoption of these hashtags illustrates adoption of a ritual behavior of prompting people to contribute to a discourse. We do not examine this type of structure in our paper as our analysis indicates little variance of the use of these hashtags from the beginning of the tweet.

In addition to emergent hashtag use we see the artificial injection of hashtags into the stream of discourse such as from marketers or political campaigns. Based on an analysis of political hashtag utilization, Conover et al. (Conover et al., 2011) found that hashtag usage and network analysis based on retweets and mentions of users were reliable predictors of political affiliation of Twitter users. In some cases, where the affiliation was found to be wrong or ambiguous, it was determined that “content injection,” the inclusion of a hashtag in an unrelated tweet, was to blame. This utilization was a way for individuals to get information into discourse streams that they would otherwise not. In that way, hashtags are susceptible to what Priedhorsky et al. (Priedhorsky et al., 2007) defines as a kind of "spam." Their role in commercial and celebrity behavior can be viewed as a marketing tactic that runs somewhat counter to emergent, organic discourse, or relevant information.

Tsur and Rappoport (Tsur & Rappoport, 2012)study of the role of hashtag content in the propagation of memes across 400 million tweets, they consider three categorical hashtag positions: prefix, infix and suffix, which denote the beginning, middle, and end of a tweet, respectively. Their descriptive study provides statistics of overall hashtag location without considering the difference between hashtags. An important next step that builds on this prior work is to examine differences in hashtag positions across different hashtags and corpora. Such contrasts illustrate the role of hashtags as an emergent and evolving affordance for social production.

3. Data Collection and Methods

As part of a larger collection effort examining the 2012 United States Election, we collected data for 11 hashtags that represent a set of sustained and general civic interest in the United States over a five-month time period from April 10, 2012 through August 12, 2012. Our data collection was conducted using the TwitterZombie architecture that queries the SEARCH API for tweets that meet specific selection criteria identified by the researcher (Black, Mascaro, Gallagher, & Goggins, 2012). We bound our data as this timeframe represents a distinct time in the United States where civic discourse shifted from the Republican Party primary season to the general election up through the time that immediately precedes the party conventions. Although our data represents political events, we do not intend this to be a politically specific paper. Instead, we use this narrow context to isolate other factors that are present in larger scale analysis of Twitter data.

We calculated the position of the hashtag by calculating the length of the tweet and then subtracting the length of the hashtag from this number and dividing the position by the length to get the position value. The subtraction of the length of the hashtag from the length of the tweet allowed for the clear identification of a hashtag in the final position. In this formula any hashtag that appeared at the end of a tweet is noted as being 100% through the tweet. This method normalizes our analysis to remove the confounding variable of hashtag length that may significantly affect the data.

Following the identification of the overall position of the hashtag in each tweet, we ran a series of analytical scripts to identify the mean and median position of each hashtag using the day as the unit of analysis to help identify longitudinal trends. The longitudinal data was then examined using both a standard linear regression function and locally weighted scatterplot smoothing (LOESS). The LOESS analysis allows for a more granular model of variation that highlights slight changes in longitudinal position in our time series analysis. We provide both the LOESS and standard regression lines on the graphs, as they are both useful in understanding the longitudinal trend. To further augment our overall and time series analysis we subset our data based on the presence or absence of common syntactical features in Twitter such as retweets, at-replys, at-mentions and URLs and compare how hashtag position varies based on these syntactical features. The identification of how hashtag position changes in these subsets further identifies causal relationships that indicate why certain hashtag positions may shift.

4. Research Questions

Our review of the literature and analysis of prior work that examined hashtags lead us to identify a gap in the analysis of hashtag position. Our dataset represents a specific context, but these research questions and our methods can be easily applied to other contexts and datasets.

1. To what extent does hashtag position in tweets vary?
2. To what extent does the position of hashtags vary longitudinally?
3. How does the presence of different syntactical features affect the position of hashtags?

5. Dataset

We choose to focus on 11 hashtags related to the United States 2012 election and political activity because they represent a significant amount of activity in Twitter and also provide a tractable scope for analysis. The maturity of these hashtags varies. The hashtags #tcot (Top Conservatives on Twitter) and #p2 (A progressive Response to #tcot) have existed for years, whereas other hashtags such as #Romney in the context of the 2012 Presidential Election are much newer as Mitt Romney was only chosen as the Republican Party Presidential nominee in summer 2012. We do not take into account geographic data or data relative to what was trending at the time of tweets as the amount of geographic data is still significantly small and trending topics vary by geographic region.

In our analysis, we choose to focus on three distinctly different classes of hashtags that were identified through a broader analysis of over 1,000 hashtags that form the basis of another stream of research. These three types of hashtags reflect broad topics that are appropriated as hashtags (#politics, #congress, #economy), individual's names (#MittRomney, #Romney, #BarackObama, #Obama) and affiliation memes that exist mostly in Twitter (#tcot, #tlot, #p2, #p21) (Table 2).

Examining these three classes of hashtags allows us to examine different positions for different types of hashtags. We exclude a commonly used type of hashtag, often identified as a micro-meme that occurs at the beginning of a tweet as a prompt because our analysis of a set of hashtags that were promoted in this way identifies limited variance of position. Of the hashtags that we examined that fit the criteria of a micro-meme, few were used beyond one week and those that did tended to become a mechanism of spam propagation.

Hashtag Type	**Hashtags**
Person	#MittRomney, #Romney, #BarackObama, #Obama
Topic	#congress, #politics, #economy
Affiliation	#p2, #p21, #tcot, #tlot

Table 2: Hashtag Type

Table 1 illustrates the syntactical feature distributions for each of the datasets in which we examine the hashtag positional data. In total, our analysis focuses on hashtags in 10,300,244 tweets. We provide the following table as reference, but also to ground our analysis of hashtag position and how syntactical features affect position, which we discuss in the third part of our findings. Our provision of this data could be a starting point for other researchers to contrast the data they are examining on Twitter. Our analysis identifies that there are certain ranges for certain syntactical features that help to define the dataset. For example, all of the at-reply distributions for the datasets are below 10% indicating that these are the least common syntactical feature we study. Further, links and mentions tend to be the most common in this dataset with significant differences in each dataset as to what occurs more often.

Our analysis of the descriptive statistics shows that there is no correlation between the distributions of any two syntactical features in our corpus. In other words, more retweets do not equal more links, although, we do demonstrate in the later findings section that there is a significant correlation between these syntactical features and the position of the hashtag among the hashtags we examine. In the context of the hashtag categories, we see that there are a higher number of at-replys with the person hashtags and a lower percentage of single contributors in the affiliation hashtags (indicating a participant community). Although these correlations do exist, they are limited so we do not believe they warrant their own findings section.

6. Findings

6.1 Hashtag Position

Our first finding identifies that hashtags have a finite position and this position correlates to the hashtag context and type. Table 3 identifies the mean and median of the collected datasets. We rank the table with the lowest median position (earlier in the tweet) at the top. We choose median as a measure for analysis as the distributions are highly skewed and not normally distributed. Additionally, our distribution is finite and bounded. Therefore, we believe the median is a more valid measure than mean. We provide the mean to illustrate the limited differences in the mean and median for robustness, but note that there is little difference.

Dataset	**Median**	**Mean**
#romney	36.42%	43.05%

#MittRomney	37.86%	45.92%
#congress	38.58%	47.00%
#obama	46.41%	49.09%
#economy	61.29%	60.67%
#BarackObama	77.53%	65.78%
#politics	82.86%	72.75%
#p21	91.01%	86.67%
#tcot	91.54%	83.09%
#tlot	92.94%	88.31%
#p2	93.98%	87.39%

Table 3: Hashtag Position Mean and Median

We see that hashtag position varies significantly in median percentage through the tweet for the hashtags that we examined. We see that there are two sets of positions, those that occur in the middle of the tweet that ranges from 36.42% - 46.41% and those that occur near the end 77.53%-93.98%.

One of the interesting cases is #economy, which has a median position of 61.29% and is an outlier in our data. #Economy exists in the middle of the two positional clusters we identify, and its two closest hashtags, #obama and #BarackObama are relatively distant: Tweet length positions of 14.88% and 16.24% on each side, respectively.

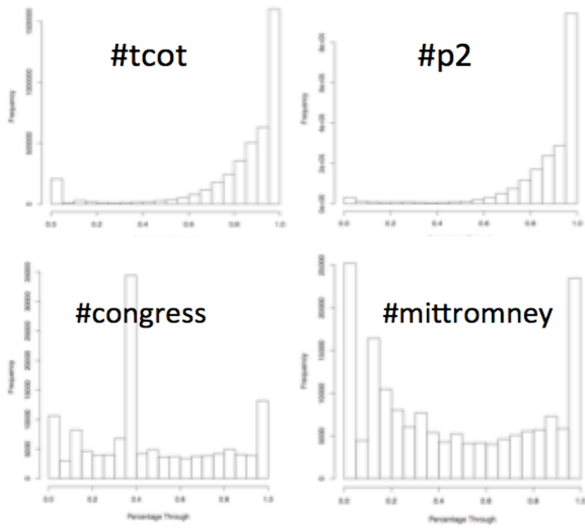

Figure 4: Affiliation Hashtag Distribution

Hashtags that occur near the end of a tweet are more likely to be affiliation hashtags as they are more metadata than message. The most noticeable affiliation hashtags are #tlot (Top Libertarians on Twitter), #tcot (Top Conservatives on Twitter), #p2 (A progressive Response to #tcot), #p21 (a response to hijacking of #p2). These four hashtags were created and have propagated throughout Twitter as a way for individuals to affiliate with certain streams of discourse. These hashtags, unlike the other hashtags we collected, have limited utility and recognition outside of the Twitter context. We see this contrast in figure 4 where we provide a sample of distributions of the hashtags we analyzed.

We would expect the hashtags we identified as names to be in similar positions in the tweet since they serve similar functional purposes. In these types of hashtags we see greater variety of position than in the other hashtags we examine. The hashtags associated with Mitt Romney (#MittRomney, #romney) appeared earlier in the tweet compared to similar hashtags for President Barack Obama (#BarackObama, #obama). This may be the result of individuals using

Mitt Romney as more of a subject than as an affiliation such as our data would indicate for Barack Obama. Our more in-depth analysis of the syntactical features of each of these datasets helps to illustrate why such differences may occur between Mitt Romney and Barack Obama.

Our last finding in this section relates to the hashtags that we classify as topics that are appropriated as hashtags. In our dataset we examined #congress, #economy and #politics. These hashtags were chosen because they are words that could exist without a hashtag in a tweet, but have been explicitly made a hashtag by the user and thus represent a different type of linguistic function compared to the affiliation hashtags and the individual names. Each of these hashtags appears in distinctly different parts of the tweet. Congress appears more towards the beginning, economy in the middle and politics near the end.

The different positions in tweets may indicate the relative usage of these words, where individuals use #economy and #congress as a subject referring abstractly to the topic of economy or congress as opposed to tagging what may be a tweet that is political in nature and tagging the tweet with #politics. In this way, #politics functions as an affiliation not in the partisan way, but in the traditional sense of metadata tagging where individuals identify their content with a label. While the affordance of a link produced by a hashtag doesn't change, its position does imply differences in it's meaning as subject *now* versus an issue or meta-topic *later* or an affiliation *established*.

6.2 Hashtags shift positions over time

Our second finding focuses on the longitudinal analysis of these hashtags. The time period that we examined represents the tail end of the 2012 Republican Primary season and the beginning of the 2012 Presidential General Election Campaign season. It is during this time that significant political discourse occurs as individuals begin looking towards the general election. In order to examine the longitudinal movement of hashtags we took the median position of the hashtag data on a daily basis and plotted those positions longitudinally. For the purpose of illustration and comparative analysis we plotted a standard regression line of best fit (red) and LOESS derived line (blue) that more closely identifies the localized movement of the medians over time.

Our longitudinal analysis identifies that hashtag median position varies on a daily basis, but that for most of the hashtags we examined, the position is within a well-defined range. These boundaries further validate our first finding that hashtag position is not randomly distributed. We see that the hashtags that we identified as affiliation hashtags (#tcot, #p2, #tlot, #p21), which tend to occur near the end of tweets, had very little variation in their placement longitudinally with a maximum daily variation in the median based on the two regression analyses of less than 1%. Lack of variance demonstrates the persistence of the hashtag in the affiliation spot of a tweet. Although there was minimal slope in the two measures in these graphs of the changing position of affiliation hashtags overtime these indicate such small overall changes that they are not indicative of a shift in position.

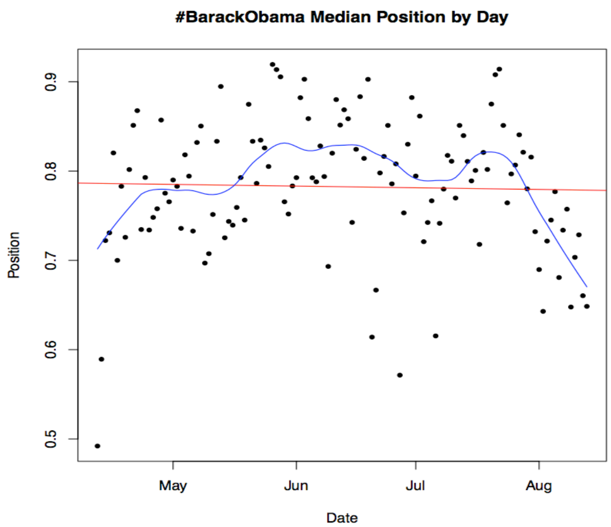

Figure 5: #barackobama median longitudinal position

In addition to the affiliation hashtags maintaining their position over time, we also identified that most of the other hashtags we examined tend to not fluctuate their median position over time. For example, #congress and #politics both had limited variations in the two regression measures we used of less than 2% over the time we studied. This illustrates that there appears to be a hashtag equilibrium state where on a daily basis hashtags may fluctuate in their position to some extent, but tend towards a stable position in the tweet.

Although we found significant differences in the overall position of the hashtags associated with individuals, our analysis of the longitudinal medians of these hashtags illustrates that these medians tended to not fluctuate over time. Based on our LOESS analysis (blue line), we see that #barackobama shifted from .72 to just over .8 in late May through June, but then returned back down (figure 5). This shift was not found with #obama as the overall median position stayed constant longitudinally.

Examining #mittromney (figure 6) we see that there are two slight shifts in the LOESS analysis that indicate a slight dip in the middle of May, but that the position increases throughout June and July before leveling off in late July and August. This shift saw a slight positive slope of the standard regression line that indicated that #mittromney had a slight shift from the beginning of the tweet to the end of the tweet.

In contrast to #obama, #romney has a slight decreasing slope in both the standard regression and LOESS analysis (figure 7). The LOESS analysis indicates that the decrease in position (moving earlier in the tweet) levels off in late June just as that the position of #mittromney levels off. We also see that the decrease that occurs in late July occurs with both #MittRomney and #Romney before a significant uptick in early August for both of the hashtags. This similar behavior (the decrease in position, followed by a significant uptick) illustrates the possibility these two hashtags are being used in a similar manner.

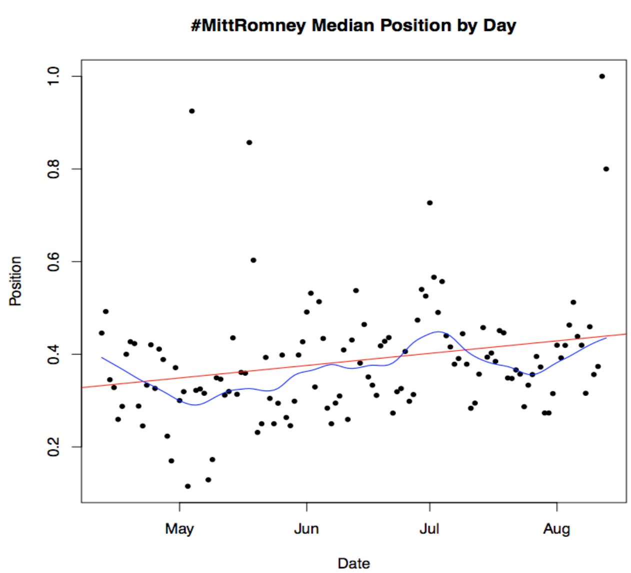

Figure 6: #mittromney median longitudinal position

6.3 Why does economy fluctuate?

The most significant fluctuations in the median of the hashtag position over time occurred in the hashtag #economy (Figure 8). This hashtag also had the highest percentage of tweets that contained URLs and the fewest percentage of retweets.

Examining the difference between this two week time period with a sustained higher hashtag position and the rest of the time period allows for an interesting case of how certain attributes of a dataset can influence hashtag position. In our analysis of the tweets from the two week time period where #economy's median position spiked significantly (n=54,277, 14.5%), we found that the difference in the syntactical feature distribution was negligible between the two datasets with the greatest variance between the two being a 2.5% difference in retweets where the subset of data had about 12.5% as retweets and the larger dataset had 15% as retweets. Further examination of the syntactical features found that the top 10 hashtags and mentions were the same with the only differences occurring in the rank order of the mentions.

Our analysis did not identify the raw count of syntactical feature distribution as the cause of the shift. Of all of the hashtags we examined, #economy had the lowest percentage of English tweets, but we also note that this difference is not significant for the two time periods we identified as the percentage of tweets that are English during the two weeks is within 1% of the larger sample to the overall time period.

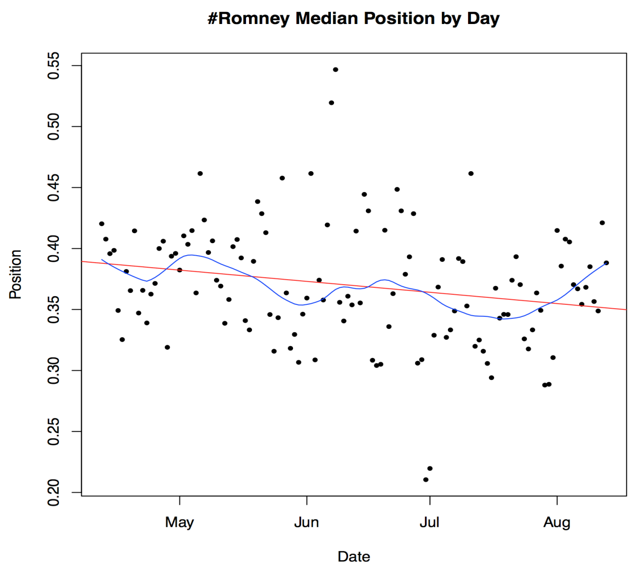

Figure 7: #mittromney median longitudinal position

Going beyond analysis of the syntactical feature distribution, we used content analysis to identify the cause of this significant shift during that time period. During that time period, four of the six most common tweets (which were retweets) had #economy at the end position of the tweeted. These tweets discussed the economic situation in Greece and another tweet that quoted Calvin Coolidge: *"The chief business of the American people is business" - Calvin Coolidge #PresidentsDay #quote #economy."* What is of interest is that this original tweet that quoted Calvin Coolidge was copied many times (as an original tweet) and also was retweeted numerous times.

In the complete dataset, the most commonly occurring tweets do not have #economy at the end of the tweets. In fact, only 1 of the top 25 tweets in the complete dataset had #economy beyond the 80% percentile in position. This more granular analysis allows for us to identify the actual reasons behind the hashtag shift and illustrates how looking beyond just syntactical feature distribution helps to identify the cause of such shifts in position. In this case it was not the raw count of the number of retweets, as there were fewer overall retweets in the two week time period, but the tweets that were retweeted.

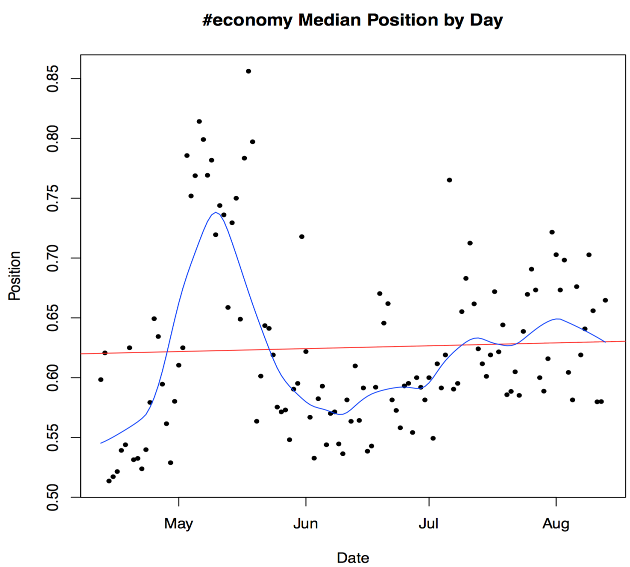

Figure 8: #economy longitudinal median position

6.4 Syntactical Feature Perspective

Our third finding builds on our analysis of #economy by focusing on the dependence that positional analysis of hashtags has to other syntactical features and how this may have an effect on position and may provide an indicator for further analysis such as was the case with retweets and #economy. Hashtag position does not occur independently of other content of the tweets and our previous analysis indicates that there is some equilibrium in the overall position of hashtags among the different hashtags. Through our analysis of the utilization of certain syntactical features in the datasets we identify that there may be other factors that influence the position of hashtags on a more granular level.

Other syntactical features such as URLs, at-mentions and the presence of an at-reply or retweet could significantly alter the position of the hashtag in the tweet. Examining the position of the three different types of hashtags in the context of other syntactical features highlights interesting dependencies between hashtag position and the utilization of other syntactical features. We note that some of the shifts in position may at first seem obvious. For example, the removal of retweets and at-replys would shift the position of a hashtag to the left. Nonetheless, we identify numerous dependencies that are counterintuitive illustrating distinct use cases of hashtags and other syntactical features.

Examining the hashtags in the categories we identify earlier in the paper we find that there is limited correlation with position and syntactical features except that we see a higher amount of at-reply tweets with the person hashtags and a lower percentage of single contributors in the affiliation hashtags. These two findings indicate distinct use cases. For example, there are more conversational messages that utilize the personal hashtags and fewer single contributors in the affiliation dataset; suggesting that affiliation hashtag users are interacting around these hashtags as a kind of community.

6.4.1 Affiliation Hashtags

The hashtags we classify as affiliation hashtags shift little when examined in the context of other syntactical features. The most drastic shift is when at-replys are extracted from the #p2 dataset. Looking at the set of hashtags more holistically, the change in position when the syntactic features of URLs and mentions are extracted are all positive except for URLs in the #p21 dataset. This illustrates that the presence of URLs and mentions causes affiliation hashtags to appear earlier in the tweet. On the other hand, this shift when examining at-replys and retweets is negative except for #tlot and at-replys where there is a negligible shift of .31% when at-replys are extracted.

This illustrates that affiliation hashtag position has similar behavior when examined in the context of other syntactical features and that shifts are negligible. We also see that URLs and mentions seem to have similar effects, as do at-replys and retweets. Overall, this limited shift in the context of other syntactical features that you would expect to shift the position is negligible as these hashtags have a stable position at the end of the tweet as a marker of affiliation.

6.4.2 Person Hashtags

The shift based on syntactic features and hashtags we classify as names is more diverse than affiliation hashtags. Eliminating at-replys shifts all of the hashtags to the left significantly with the exception of #barackobama where the shift is only 4.5% to the left. Extracting retweets from the dataset causes a slight shift to the right for all of the hashtags except for #barackobama where there is a significant shift to the left (-21.31%). The latter shift (to the left) makes intuitive sense: given that retweets text always appears on the far left, one would expect that removing retweet text ("RT @[username]") would shift the text of a tweet left.

The former shift (to the right) is puzzling, however, and suggests something about retweets. Since retweets tend to be found among content that is shared by mass media and spam, perhaps there is a tendency to place hashtags in the content or at the beginning among those handles, causing a skew left that is corrected for when we remove retweets. Eliminating mentions causes a negative shift in hashtags (to the left) in position from -7 to -13% except for #barackobama, which causes a shift of 17% to the right. Eliminating URLs causes a negative shift for all hashtags except for #obama, though this is a slight positive shift (.17%). The most drastic position is for #barackobama which has a significant shift to the left (-29%).

This data illustrates that the hashtags #obama and #barackobama have unique positional attributes compared to #romney and #mittromney. The different positions identified in the first section of findings may indicate that these hashtags, though similar in construction and purpose from a high level serve different roles. This differing role may be the result of the context of these individuals, but may also illustrate how an individual's role in an electoral discourse may lead to a difference in the way that a hashtag is utilized.

6.4.3 Topic Hashtags

Similar to the shift when examining hashtags classified as names, the hashtags we classify as nouns are much more diverse. Extracting links and mentions shifts all of the nouns from 10-20% right except for #economy which shifts -24.7% (to the left). The shift based on at-replys and retweets is split: #congress and #economy shift about -21% left when at-replys are extracted, but politics shifts 20% to the right. #congress and #economy shift 8% and 23% respectively to the right when retweets are excluded, but #politics has a -9.75% shift to the left when retweets are excluded.

This analysis helps to further illustrate the unique characteristics that each of these topical hashtags has and the different functions that they serve. As we identified earlier, the topical hashtags have different overall positions and in the case of #economy has a significant shift at certain times. Therefore, we see that topical hashtags may have more variance and less of a positional relationship with syntactical features.

7. Discussion

The rapid growth of Twitter opens the possibility that explicit syntactic features, like hashtags, also contain implicit signals to be discerned from the manner of use. In this paper, we show that the subtle, contextual cue of hashtag position is one such signal. Hashtag position is an indicator that has not been previously identified as useful for designing tools to help users filter and search tweets. The hashtags examined here have an equilibrium position, but those positions shift on specific days; the shifts signal a change in events outside Twitter, in the physical world.

We see that affiliation hashtags tend to occur at the end of a tweet. Shifting affiliation hashtags signal information sharing, and are often coupled with a URL. Most shifts are subtle, but more significant shifts in the location of an affiliation hashtag could be an indication of a campaign or other coordinated effort. Having a baseline is essential to detecting a signal in the future.

In the case of Obama and Romney, we see that this shift in position occurs with an event such as the convention or the winning of the nomination as there are noticeable shifts in the middle of August preceding the General Election and in the middle of the summer when Governor Romney became the presumptive nominee. In the context of an election or political discourse, a shift in the position may be a signal that precedes a shift in opinion or the transition of one form of discourse to something different. Campaigns and other political monitors such as the media could use this type of analysis as a way to identify how information is spreading or if a message has been adopted.

Contemplation of broader implications for our findings suggests that hashtag use during crisis situations may help identify the different phases of crisis and response, or signal changes in on the ground situations. In this and other domains, it is worthwhile to examine how the signal of hashtag position can be used for design of systems. Its also worth noting that making the use of hashtag position something explicit, and within a user's embodied interaction with a system, could lead to richer and more reliable signals.

This paper provides scholars with a starting point for the investigation of new signals in Twitter;

signals implicit in the modes of use defined by users. This could enable Twitter users to better filter their stream (Golbeck, 2012) in the future. For example, if a user was interested in only looking at tweets in which individuals were discussing a particular hashtag (or set of hashtags) as the subject, but not when it is being used to signal a reference to a related topic, they might narrow their search to hashtags appearing in the middle versus the end, respectively.

Future research should involve the development of computational linguistic algorithms to identify how hashtag positions might shift when used as a classifier (#obama supporters), a thing (#obama) or a process (Zappavigna, 2011). We suggest a computational linguistic approach in the future because positional shifts can be coupled with analysis of syntactical features.

In this paper, we show that the location of a hashtag is not random. Consequently, hashtag location, and shifts in hashtag location in a corpus can be used to identify interesting content about the economy or other domain specific topic. Our findings suggest that shifting hashtag position warrants a user alert. This proposed technique might prove even more useful on lower volume hashtags, where common heuristics like tweets per minute that are often applied in Twitter research will not be useful. This approach could be adapted for other syntactical feature analysis, which may surface other causal effects of syntactical features.

While we focused on hashtag positions in the context of political activity, this type of inquiry could be applied to any domain. Our examination and provision of this data could be a starting point for other researchers to contrast the data they are examining on Twitter with ours. This might move us forward in standardizing measures that help users

8. References

Ames, M., & Naaman, M. (2007). *Why we tag: motivations for annotation in mobile and online media.*

Black, A., Mascaro, C., Gallagher, M., & Goggins, S. (2012). *TwitterZombie: Architecture for Capturing, Socially Transforming and Analyzing the Twittersphere*. Proceedings from ACM Group, Sanibel Island, FL.Sanibel Island, FL

Bruns, A., & Burgess, J. E. (2011). #Ausvotes: how Twitter covered the 2010 Australian federal election. *Communication, Politics and Culture*, *44(2)*(2), 37-56.

Chakrabarti, D., & Punera, K. (2011). Event Summarization Using Tweets. *ICWSM*, *11*, 66-73.

Chang, H.-C. (2010). *A New Perspective on Twitter Hashtag Use: Diffusion of Innovation Theory*. Proceedings from ASIS&T, Pittsburgh, PA.Pittsburgh, PA

Conover, M. D., Gonçalves, B., Ratkiewicz, J., Flammini, A., & Menczer, F. (2011). *Predicting the political alignment of twitter users*. Proceedings from Privacy, Security, Risk and Trust (PASSAT), 2011 IEEE Third International Conference on and 2011 IEEE Third International Confernece on Social Computing (SocialCom).

Conover, M. D., Ratkiewicz, J., Francisco, M., Goncalves, B., Flammini, A., & Menczer, F. (2011). *Political Polarization on Twitter*. Proceedings from ICWSM, Barcelona, Spain.Barcelona, Spain

Cunha, E., Magno, G., Comarela, G., Almeida, V., Gonçalves, M. A., & Benevenuto, F. (2011). *Analyzing the dynamic evolution of hashtags on twitter: a language-based approach.*

Doughty, M., Rowland, D., & Lawson, S. (2011). *Co-viewing Live TV with Digital Backchannel Streams*. Proceedings from EuroITV, Lisbon, Portugal.Lisbon, Portugal

Golbeck, J. (2012). *The twitter mute button: a web filtering challenge.*

Heverin, T., & Zach, L. (2011). Use of Microblogging for Collective Sense-Making During Violent Crises: A Study of Three Campus Shootings. *Journal of American Society for Information Science,* 10.1002/asi.21685.

Hu, Y., John, A., Seligmann, D. D., & Wang, F. (2012). *What Were the Tweets About? Topical Associations between Public Events and Twitter Feeds*. Proceedings from ICWSM, Dublin, Ireland.Dublin, Ireland

Huang, J., Thornton, K. M., & Efthimiadis, E. (2010). *Conversational Tagging in Twitter*. Proceedings from HT, Toronto, Ontario.Toronto, Ontario

Kaptelinin, V., & Nardi, B. (2012). *Affordances in HCI: toward a mediated action perspective.*

Lehmann, J., Gonçalves, B., Ramasco, J. J., & Cattuto, C. (2012). *Dynamical classes of collective attention in twitter.*

Marwick, A. E., & boyd, d. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media Society*, *13*.

Mascaro, C., Black, A., & Goggins, S. (2012). *Tweet Recall: Examining Real-Time Civic Discourse on Twitter*. Proceedings from ACM GROUP, Sanibel Island, FL.Sanibel Island, FL

Miller, A. D., & Edwards, W. K. (2007). *Give and take: a study of consumer photo-sharing culture and practice.*

Naaman, M., Becker, H., & Gravano, L. (2011). Hip and Trendy: Characterizing Emerging Trends on Twitter. *Journal of American Society for Information Science*, *62(5)*(5), 902-918.

Priedhorsky, R., Chen, J., Lam, S. K., Panciera, K., Terveen, L., & Riedl, J. (2007). *Creating, Destroying and Restoring Value in Wikipedia*. Proceedings from GROUP, Sanibel Island, FL.Sanibel Island, FL

Romero, D. M., Meeder, B., & Kleinberg, J. (2011). *Differences in the Mechanics of Information Diffusion Across Topics: Idioms, Political Hashtags, and Complex Contation on Twitter*. Proceedings from WWW, Hyderabad, India.Hyderabad, India

Sen, S., Lam, S. K., Rashid, A. M., Cosley, D., Frankowski, D., Osterhouse, J. et al. (2006). *Tagging, communities, vocabulary, evolution.*

Shamma, D. A., Kennedy, L., & Churchill, E. F. (2010). *Conversational Shadows: Describing Live Media Events Using Short Messages*. Proceedings from ICWSM, Washington, DC.Washington, DC

Starbird, K., Palen, L., Hughes, A. L., & Vieweg, S. (2010). *Chatter on The Red: What Hazards Threat Reveals about the Social Life of Microblogged Information*. Proceedings from CSCW, Savannah, GA.Savannah, GA

Starbird, K., & Stamberger, J. (2010). Tweak the tweet: Leveraging microblogging proliferation with a prescriptive syntax to support citizen reporting.

Tsur, O., & Rappoport, A. (2012). *What's in a hashtag?: content based prediction of the spread of ideas in microblogging communities.*

Yang, L., Sun, T., Zhang, M., & Mei, Q. (2012). *We know what@ you# tag: does the dual role affect hashtag adoption?*

Zappavigna, M. (2011). Ambient affiliation: A linguistic perspective on Twitter. *new media & society*, *13(5)*(5), 788-806.