

The Effect of Mobile Platforms on Twitter Content Generation

Mathieu Perreault and Derek Ruths
School of Computer Science, McGill University
Montreal, Quebec

Abstract

The increased popularity of feature-rich mobile devices in recent years has enabled widespread consumption and production of social media content via mobile devices. Because mobile devices and mobile applications change context within which an individual generates and consumes microblog content, we might expect microblogging behavior to differ depending on whether the user is using a mobile device. To our knowledge, little has been established about what, if any, effects such mobile interfaces have on microblogging.

In this paper, we investigate this question within the context of Twitter, among the most popular microblogging platforms. This work makes three specific contributions. First, we quantify the ways in which user profiles are effected by the mobile context: (1) the extent to which users tend to be either fully non-mobile or mobile and (2) the relative activity of the mobile Twitter community. Second, we assess the differences in content between mobile and non-mobile tweets (posts to the Twitter platform). Our results show that mobile platforms produce very different patterns of Twitter usage.

As part of our analysis, we propose and apply a classification system for tweets. We consider this to be the third contribution of this work. While other classification systems have been proposed, ours is the first to permit the independent encoding of a tweet's form, content, and intended audience. In this paper we apply this system to show how tweets differ between mobile and non-mobile contexts. However, because of its flexibility and breadth, the schema may be useful to researchers studying Twitter content in other contexts as well.

Introduction

In recent years microblogging has emerged in many communities as a widespread and frequently used means of communication (Webster 2010). The governing principle of microblogging is that the user frequently posts very short, primarily textual updates that are viewable to either a restricted or fully public audience. Because this form of social interaction is new and relatively uncharacterized, recent work in the area of social media has investigated the question of how microblog content is generated and consumed by users (e.g., (Boyd, Golder, and Lotan 2010; Honey and Herring 2009; Java et al. 2009; Naaman, Boase, and Lai 2010; Krish-

namurthy, Gill, and Arlitt 2008; Zhao and Rosson 2009; Yardi and Boyd 2010)).

In this paper, we contribute to this growing literature by characterizing the way in which mobile technologies are shaping microblog users and content. We conduct this investigation within Twitter, one of the most popular microblogging platforms. This work is motivated by the fact that the widespread adoption of feature-rich mobile devices, including smartphones and tablets, has simultaneously created a class of users capable of microblogging while away from traditional computer interfaces: the desktop web-browser or application. These mobile platforms change the context within which microblogging is done in at least two important ways.

First, mobile applications are quite different from (and typically have fewer features than) their desktop equivalents: user input is typically provided via a combination of direct taps on the interface and the use of a small on-screen keyboard; compared to a monitor, screen real estate is severely limited on mobile phones and tablets; and moving content from one application to another is much less common. Combined, these differences create a very different tool by which the user may generate and access microblog content.

The second way that mobility changes the microblogging context relates to the fact that the user is no longer bound to his or her computer when generating or consuming microblogging content. This means that, with few exceptions, users may employ their mobile devices to generate and consume content wherever they happen to be. Given that the emphasis of microblogging is on making small, frequent posts, it stands to reason that allowing users to create new content, quite literally, on the go should lead users to explore new ways of integrating the phenomenon into their daily life.

Our hypothesis is that these two contextual changes have a quantifiable effect on the way that users are generating and consuming microblogging content. Some existing work has touched on this idea. One project looked at how users interact with Twitter to produce and get information about local events (Yardi and Boyd 2010). A different study looked at data from 2008 and found that Twitter users are more likely to use their cellphones to connect to the Internet (40% of Twitter users do so versus 24% of non-Twitter users) and are more likely to consume news and information on these devices (Lenhart and Fox 2009). Finally, another report found

that Twitter users are more comfortable accessing social networking services via their mobile phone, with 63% of Twitter users claiming so in contrast to 34% of all social networking platform users (Webster 2010).

While we are not the first to advance the notion that microblogging changes in the mobile setting, to our knowledge, this paper is the first to quantitatively assess what changes actually occur.

In order to study this, we collected a large data set of tweets and classified them as originating from either mobile or non-mobile applications/web interfaces (hereafter, *agents*). We were then able to contrast mobile vs. non-mobile Twitter usage in terms of both how mobile user profiles differ from the norm and how the content of mobile tweets differ from those generated on non-mobile agents.

In the user profile analysis, we were interested in understanding (1) how mobile agent usage was distributed among Twitter users and (2) the extent to which being a heavy user of mobile agents correlated with different patterns of status updates and connectivity to other users. Our analysis shows that individuals tend to be polarized toward either mobile or non-mobile platforms: very few users will regularly use both. Furthermore, this polarization does reveal that mobile communities tend to have more active and better connected users.

Our analysis of content quantified various features of the content of Twitter posts (hereafter, *tweets*) that originated from mobile and non-mobile agents. Feature included simple idiomatic structures like emoticons and web links as well as more subtle properties such as type of content and intended audience. Comparing these features, we identified a number of trends, usage patterns, and types of content that were significantly different between mobile and non-mobile tweets.

As part of this study we developed a generalized classification schema for microblog content which will be of interest to the broader community of social media researchers. Unlike previously proposed classification schemes, ours articulates and quantifies the different aspects of a microblog post: its intended audience, structure, and type of content (Java et al. 2009; Naaman, Boase, and Lai 2010). We found this schema useful for identifying and quantifying the similarities and/or differences between different data sets of tweets. While we employ it in this study to evaluate the effects of mobility, we expect that it could be fruitfully used by others to investigate a host of other variables which affect the way in which microblogging platforms are used.

Overall, our findings in this paper support the conclusion that mobile agents profoundly influence the way that users engage with microblogging content. In the following sections we discuss our methodology and results in greater detail and then conclude with a discussion of the broader implications of our results.

Data and Methods

Data collection

As the basis for most analyses in this paper, we collected a dataset of 2,000,000 tweets from January 6 to January

7th, 2011 using the Twitter Streaming API. Our access to Twitter's streaming service, called the "spritzer", provided us a random sample of 1% of all public tweets, according to Twitter's developer website. It is worth noting that the spritzer sampling can be problematic for some studies which need near-complete coverage of tweets generated over a time period. In the case of this study, because tweets are selected via a truly random process, the collected data set represents a valid, unbiased sampling of mobile and non-mobile content.

In order to study the effect of mobility on user-level features, for each user in the dataset her followers (the users who receive her posts), followees (the users whose posts she receives), and 100 most recently posted tweets were obtained using the Twitter User API. These data allowed us to identify trends in numbers of followers and followees as well as each user's preferred Twitter agents (e.g., applications and web pages). In total, we obtained this profile information for 154,311 distinct Twitter users.

English filtering Because some manual coding was used to classify tweets, we restricted our data set to contain only tweets written in English (the language shared by all our human coders). In order to retain only English tweets, a filter based on the MySpell dictionary¹, notably used in the OpenOffice suite, was created and applied on the entire dataset. Each tweet was converted into a set of words which was then filtered to contain only words related to the actual content of the tweet (as opposed to words or constructs used in the Twitter syntax such as "RT" or "@user", where "user" is a Twitter username). Using the MySpell English dictionary, only tweets that contained at least 50% of recognized English words were kept in the dataset. While other ratios were considered and tested, the 50% ratio was found to work best particularly in accommodating the grammatical mistakes (e.g., use of slang) and proper nouns that Twitter posts often contain. After filtering the original data set, we were left with 197,183 tweets.

Mobile and non-mobile data set generation

Since a major focus of this study was on understanding differences between content generated in mobile and non-mobile contexts, we needed to determine the context in which each tweet in our data set was generated. To do this, we made use of the metadata which is provided by the Streaming API for each tweet delivered. Specifically, the *source* field contains the identifier for the software agent which was used to generate the tweet. We extracted all distinct identifiers from our tweet data set and manually classified each as either *MOBILE* or *NON-MOBILE* based on any information we could find about specific platforms it supported. For the purposes of this research, mobile devices are defined as devices running mobile-oriented operating systems. These include:

- **Non-smartphones:** Often referred to as "feature phones", devices in this category lack the advanced features found in smartphones, but can access the Twitter platform

¹<http://wiki.services.openoffice.org/wiki/Dictionaries>

through the service’s SMS feature. A recent market analysis has found that 83% of all phones sold in the US in 2009 belonged in this category (Radwanick 2010). However, analysis of our dataset indicates that only 3.3% of tweets are created through SMS and 3.3% are created through a scaled-down version of the Twitter website for use on non-smartphones.

- **Smartphones:** Devices in this category are able to access a significantly higher quality version of the web than non-smartphones and for the most part are able to run mobile applications. Collected Twitter posts that belong in this category come from smartphones that connect to the Internet (either through cellular modems or Wi-Fi networks). Market leaders in this category are Apple’s iPhone and devices running on the Android operating system (Radwanick 2010).
- **Tablets:** Internet-connected tablets share some features with smartphones and in some cases offer a bigger form factor. They access Twitter mostly through mobile applications.

Some agents, such as “Tweetdeck” and “Echofon”, were found to run on both mobile and non-mobile platforms. With the metadata available, it was impossible to determine when these agents were actually mobile or non-mobile for a given tweet. Therefore, these were labeled *MIXED* and were omitted from the analysis entirely. According to agent counts in our data set, there were 215 agents, 73 mobile, 127 non-mobile, and 15 mixed. The most popular non-mobile agents in our data set were “Twitter web page”, “twitterfeed”, “Tumblr”, “HootSuite”, and “Twitter for Mac”; the most popular mobile agents were “Twitter for iPhone”, “UberTwitter”, “Twitter for Blackberry”, “SMS”, and “Mobile Web.”

The resulting mapping of agent identifiers to a mobile or non-mobile context covers more than 90% of all possible agents found in the tweet data set.

Content classification

In order to compare the differences between mobile and non-mobile tweets, it was necessary to characterize each tweet’s content using a formal classification scheme. Because tweets are relatively unstructured and content can be quite nuanced, we employed several human coders to manually perform this characterization.

The classification schema Despite the relative informal nature of most tweets, they exhibit rich variability in content and construction. We sought a classification schema that would encode the different ways in which tweets could differ. We considered several categorization schemas proposed and applied in prior work: one study broke tweets into four broad categories, daily chatter, information sharing, conversations and news reporting (Java et al. 2009); other papers have used machine learning techniques and mathematical models to derive topical classes for tweet content (Ramage et al. 2010; Ritter, Cherry, and Dolan 2010); perhaps most similar to our work is a study which used a schema to code the different message types of tweets (Naaman, Boase, and Lai 2010).

While these classification schemas have suited the goals of their individual studies, we found them too specialized to articulate the many different components that comprise tweet construction. As a result, we devised a new multi-level classification system which codes a tweet in terms of its (1) intended audience, (2) type of content, and (3) form.

1. **Audience** - a tweet should be assigned exactly one of these labels.
 - **Directed:** The tweet is directed at one or many Twitter users, through the use of the “@” character. Note that this is different from *direct messages*, which we did not consider in this study.
 - **Broadcast:** The tweet has no specifically intended recipient, implying that it is intended as a broadcast to a user’s followers.
2. **Content** - a tweet should be assigned one or more of these labels.
 - **Conversation Initiator:** a tweet that is meant to engage in a conversation and incite replies by other users. It may contain broad or specific questions or opinions, depending on the audience (as defined above).
 - **Conversation Response:** a tweet that either explicitly or implicitly responds to a conversation initiator tweet.
 - **Subjective Assertion:** a tweet expressing a personal opinion on a certain topic or event, or a person’s view on their own condition.
 - **Status Update:** a tweet reporting the user’s current condition (e.g. what they are doing or where they are going).
 - **News sharing:** a tweet reporting news (with news being defined as a recent event that is of broad public interest).
 - **Other resource:** a tweet reporting other types of information that can be shared, such as personal anecdotes, videos, photos, and websites.
 - **Spam:** a tweet with content identified as spam, either with many incoherent keywords and/or links that are qualified as spam.
 - **Unknown:** a tweet that literally makes no sense. Such tweets are often single words that give virtually no information about the tweet’s content or intended use (e.g. we have encountered tweets that simply read “triple”).
3. **Form** - a tweet should be assigned exactly one of these labels
 - **Tweet:** a simple post that contains no retweeted content.
 - **Retweet:** a tweet that does no work besides promoting someone else’s tweet to a user’s own list of followers.
 - **Mixed:** a tweet that contains a retweet with content added by the user. When employed, this form usually is used by a user to add an opinion or response to the post being retweeted.

In extensive manual coding exercises, we found this classification schema to be flexible enough to accommodate all

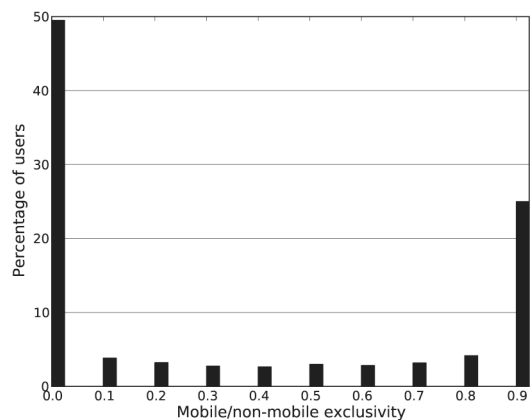


Figure 1: The distribution of users according to their mobile/non-mobile exclusivity, the fraction of statuses a user has posted from a mobile agent. Users are binned by intervals of 0.1, meaning that the bin at 0.1, for example, includes all users with exclusivity values, $0.1 \leq v < 0.2$. The distribution shows strong bimodal characteristics which indicate that the Twitter community primarily consists of two groups of users: those who nearly always use non-mobile agents and those who nearly always use mobile agents.

tweets in our data set (e.g., there were few instances where a tweet was encountered and its assignment was ambiguous).

Manual coding of tweets We used the TweetCoder platform as the manual coding application that implemented the classification schema described (Ruths and Perreault 2011). We employed three human coders to manually code 5027 tweets (2532 mobile tweets, 2495 non-mobile tweets). This number of tweets is on par with the number coded in comparable studies (Jansen, Zhang, and Sobel 2009; Thelwall and Wilkinson 2010; Naaman, Boase, and Lai 2010).

Results

User profile results

Our objective in this section was to study how mobile microblogging affected user profiles. As described in the *Data and Methods* section, we obtained the profile, follower/followee lists, and total number of statuses posted for each user present in the large tweet data set. The resulting data set contained 154,311 distinct users. This was the basis for the analysis described next.

We assessed the effect of the mobile context on user profiles in three ways.

Mobile/non-mobile exclusivity. We quantified the extent to which the usage of mobile Twitter agents precluded the use of non-mobile agents (and vice versa). Here we were interested in understanding the extent to which individual users accessed Twitter through mobile and non-mobile agents. To do this, for each user in our data set we computed the fraction of the user’s last 100 statuses that was made from a mobile agent. We called this the user’s *mobile/non-mobile exclusivity ratio*. An exclusivity value of 0 indicates that the user is an exclusive user of non-mobile agents; a value of 1 indicates that the individual never used a non-mobile agent

(and only used mobile agents); values in between indicate that the individual used both types of agents to varying degrees. We assembled the exclusivity values for all users in our data set and plotted the distribution, shown in Figure 1.

The most striking feature of this distribution is its strongly bimodal shape. The figure reveals that the mobile/non-mobile agent distinction does, indeed, segment the population into two groups: those who use non-mobile agents and those who use mobile agents nearly to the exclusion of the other. This suggests that the Twitter community consists of two groups of users: those who strongly favor mobile agents and those who strongly favor non-mobile agents. Less than 25% of all users lie in between. We refer to users with exclusivity < 0.1 as *exclusively non-mobile* and to users with exclusivity ≥ 0.9 as *exclusively mobile*. Because of the strong segmentation of the community into these two populations, the remainder of our user profile analysis focuses on the profile differences between these two groups.

Number of statuses. For each exclusively mobile and non-mobile user (exclusivity values ≥ 0.9 and < 0.1 , respectively), we determined the total number of statuses she had posted since creating her account. From this, we computed the distribution of number of statuses for the exclusively mobile and non-mobile user groups. A truncated version of this distribution is shown in Figure 2(a). As can be seen towards the right edge of the figure, the two statistics converge to a similar distribution, which continues out into the long tail (not shown).

Number of followers/followees. For each exclusively mobile and non-mobile user, we recorded the number of followers (individuals following that user) and followees (individuals whom that user follows). In much the same way as with the number-of-statuses statistic, we constructed the distributions for follower number of followee number for the populations of exclusively mobile and exclusively non-mobile users. Truncated versions of these distributions are shown in Figures 2(b) and 2(c). As with the number-of-statuses statistics, these distributions also had long tails which were similar between the mobility classes.

The very interesting feature of all three distributions is the discrepancy between the mobile and non-mobile users that occurs for small values. Furthermore, while this discrepancy is very large for small values, the two distributions eventually converge into a highly similar and overlapping long tail. These differences for small numbers effectively correspond to differences in the abundance of inactive or uninvolved users in the two groups: users who have few total statuses are not regular content contributors, users who have few followers have little chance of their tweets being read or circulated, and users who follow few other individuals receive little content to consume and propagate.

These marked and localized differences in the distributions highlights differences between the mobile and non-mobile communities. These will be expanded upon in the *Discussion* section.

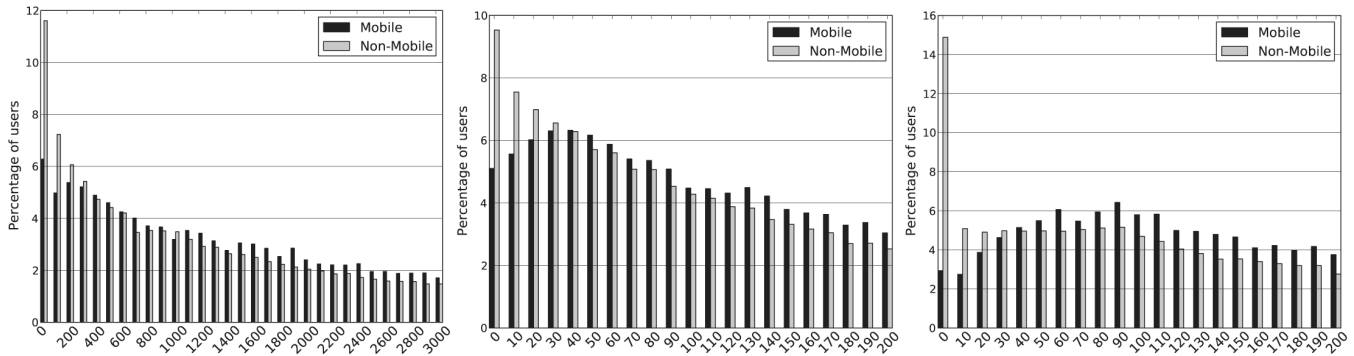


Figure 2: The truncated distributions of (a) number of user statuses, (b) number of followers, and (c) number of followees for exclusively mobile (black) vs. non-mobile (grey) users in our data set. All three distributions have very long tails that exhibit similar behavior regardless of user mobility. Of note is the behavior of the distributions for small values, which strongly distinguishes the mobile and non-mobile users.

Tweet content results

In this section we use our large tweet data set to identify and quantify differences produced by mobile Twitter microblogging environments. We investigate this in two ways. First we look at the relative frequencies of simple idiomatic structures that are common in tweets. We then look at a more comprehensive characterization of mobile and non-mobile tweet structure using the classification system discussed in the *Data and Methods* section. This second approach allows the comparison of a wide array of features a Twitter user may employ when composing a tweet.

Idiomatic structures in tweets A number of tweeting practices and conventions have become widespread among Twitter users. We use these as a starting point for detecting and understanding how mobile agents change the microblogging experience.

Directed tweets. With the exception of direct messages which we do not consider in this study, a user’s tweet is broadcasted to all of her followers. Nonetheless, Twitter-based conversations have become common in which a tweet is “directed” to a specific user by placing that user’s Twitter username at the front of the tweet: e.g., “@iman23 want to grab some dinner?”

Hyperlinks. Users will often embed links in tweets in order to share news stories, photos, videos, and a variety of other content. In effect, users are circumventing the traditional 140-character limit that is imposed on single tweets by using some of these characters in providing a URL, which contents extend or support what is being expressed. In addition, this traditional “linking” behavior has been extended by some companies to provide interfaces to a variety of third-party, Twitter-specific services such as Foursquare² and the ability to send longer tweets in Tweetdeck through “http://deck.ly” links.

Foursquare, which was founded in 2009, has been using the Twitter platform heavily to promote Foursquare’s users’

²<http://www.foursquare.com>

location updates to their existing Twitter network of followers. Upon updating their location on the Foursquare service, users are prompted to make their action known to their Twitter followers by posting a link and a short description.

Emoticons. Emoticons have become commonplace in a variety of digital (and even non-digital) textual mediums. Emoticons are used in tweets, as elsewhere, to convey a specific emotion or feeling to the reader.

Retweets. A retweet is a Twitter-specific mechanism by which a user can quote and propagate to her followers the content of another Twitter user’s tweet. Typically retweets properly attribute the quoted material to the original author by including the authors Twitter username somewhere in the retweeted text.

Hashtags. Hashtags are words in a tweet that begin with a “#” character. They have evolved to be topical labels that can be applied to a tweet. They can be used in two ways. A hashtag can be external to the tweet, not meant to be read as part of the tweet text, but rather to orient the reader to the intended point being made (e.g., “i want a dog for christmas #christmaswish”). A hashtag may also participate in the content of the tweet as a quasi-word, in which case it is both syntactically part of the tweet’s core content and a topical labeling (e.g., “saw the coolest #laptop at the store today”).

An advantage of this analysis is that these structures can be readily detected in tweets by automated means using regular expression tools. In the following analysis, we defined each of these structures as follows:

- **Directed tweets:** Tweets starting with “@” are often thought of as replies to other users, often indicating a conversation is taking place.
- **Links:** Tweets that contain hyperlinks, as indicated by the presence of “http://”.
- **Emoticons:** As proposed by (Pak and Paroubek 2010), a set of all popular emoticons and their variations was built and used to classify tweets in this category.
- **Retweets:** Tweets that are starting with “RT @” or that contain “(via @...)” are classified as retweets.

- **Hashtags:** Tweets that contain at least one word starting with a hashtag (“#”) were put in this category.
- **Foursquare updates:** Tweets that contain special links starting with “http://4sq.com” are classified as Foursquare updates.

We computed the occurrence frequency for each of these structures among mobile and non-mobile tweets. The results, shown in Table 1, were evaluated to be similar for a different day of the week. Several trends are noteworthy.

We find that retweets, hashtags, and emoticons all have nearly identical percent occurrences in mobile and non-mobile tweets. This might be expected: since most mobile and non-mobile agents provide a button that retweets the current post, mobile agents don’t introduce any unique challenges or increased ease with which content can be retweeted; because hashtags are synonymous with topic labels, it is likely that users feel compelled to add them where appropriate, regardless of platform; and emoticons in many settings can be as important as words in communicating the intent of the tweet, making them similarly indispensable to hashtags.

The three remaining constructions which do differ between the mobile and non-mobile tweets each identify a particular way in which mobility may produce different microblogging behaviors. The checkin feature underlying Foursquare underscores its dependence on mobile microbloggers, thus it is not surprising that all Foursquare links are found in mobile tweets.

The difference in occurrences of general web links, however, is quite striking and indicates that mobile tweets are predominantly used to post purely textual content. Since features such as copy and paste are not yet readily available on mobile devices, information may be harder to copy from one application (e.g., web browser or email client) into the Twitter agent. These practical constraints can explain why links are dramatically less common in mobile tweets. Additionally, the mild increase in directed tweets suggests another reason for decreased link usage in mobile tweets. Since direct tweets are often used when holding a conversation over Twitter, the absence of links combined with the increase in direct tweets may indicate that more (purely textual) conversations are conducted by mobile users.

Detailed structures in tweet content In this section we narrow our focus to consider the 5027 tweets (2532 mobile, 2495 non-mobile) that were manually coded using the

Table 1: Percent occurrence of various idiomatic constructions in mobile and non-mobile tweets.

<i>Construction</i>	<i>Mobile</i>	<i>Non-Mobile</i>
Directed tweets	35.7%	32.5%
Retweets	15.6%	15.4%
Hashtags	14.2%	13.3%
Containing emoticons	9.6%	9.6%
Containing links	7.7%	30.6%
Foursquare updates	1.8%	0.0%

classification system described in the *Data and Methods* section. The goal of this exercise was to obtain finer characterizations of the instances where mobile devices and mobility produce different microblogging behaviors. Figures 3(a-c) show the human-coded data set broken down first into mobile (black) and non-mobile (grey) tweets and then further divided into the different categories each tweet was assigned to. The figures display the percent of mobile and non-mobile tweets that were assigned to each category of the schema.

Validating human-coded classifications. In order to validate the human-coded classifications, we can compare the percentages to those obtained in the previous section to determine whether general trends agree. We believe that the trends presented below, observed to be similar both in the human and the idiomatic analyses, as well as the calculated average inter-coder reliability of 72% establish a certain degree of confidence in the human-coding. First consider retweets, which was an idiomatic construction found to be nearly identical between mobile and non-mobile tweets. In Figure 3(c), we observe that the manually coded data set reproduces this finding both in terms of the qualitative trend (i.e., nearly identical percent retweets among mobile and non-mobile data) and the quantitative percentages (in the manually coded data the percentage of posts that were retweets fell relatively close to 15% which was reported in the idiomatic analysis).

The significant difference between mobile and non-mobile usage of links is also present in the human-coded data. Notice that in Figure 3(b), the “news” and “other content” are much more common in non-mobile tweets. In the coding exercise, news tweets had to report a news story either by containing a brief textual summary text or, more often, a link to the story on a news site. The “other content” category held personal anecdotes and, much more frequently, links to photos and videos. The fact that both of these categories are more common in non-mobile tweets suggests that links, too, are much more common in non-mobile tweets.

Finally, we observe similar results for directed tweets as well. In the large-scale analysis, we found that directed tweets were slightly more common in mobile tweets. We find the same, but stronger trend in the manually coded data. This, however, is likely due to the fact that there are multiple ways to direct a tweet to a specific user, not all of which involve starting the tweet with the recipient’s Twitter username. The human coders were directed to flag every intentional directed tweet as directed whereas the regular expression-based approach only registered tweets satisfying the stricter criterion.

Discussion

Our results strongly support the hypothesis that the mobile context can induce a number of substantial changes in a microblogging community as well as in the content that they generate. In this section we highlight the differences we detected and propose possible mechanisms and processes that explain why these difference might arise in mobile contexts.

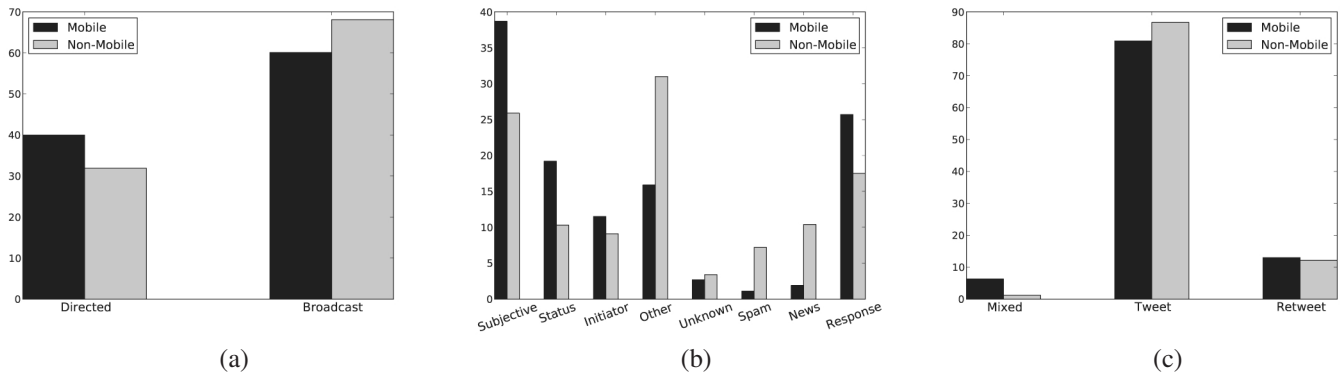


Figure 3: The percentage of tweets that were assigned to different categories in the classification schema. Results are broken up by the type of structure being analyzed: (a) audience, (b) content type, and (c) form. Mobile tweet results are in black, non-mobile tweet results are in grey. In the audience and form areas (subplots (a) and (c)), each tweet received exactly one labeling, so percentages add to 100%. In the content type area, (b), a tweet could receive one or more assignments, so these percentages do not necessarily sum to 100%.

It should be noted, however, that thoroughly evaluating the validity of these proposed mechanisms and processes is a direction for future work.

Twitter users are either mobile or non-mobile. This observation is well supported by the strong bimodal distribution of mobile/non-mobile exclusivity values in Figure 1. While initially surprising, this division makes sense when two factors are considered.

First, while many individuals own cell phones (a common mobile device), ownership of a cell phone does not translate into using the device for non-voice related activities. In fact, the barriers to microblogging on a mobile device are non-trivial: the user must either own a device capable of running a mobile agent or configure their account to permit tweet submission via text messages.

Second, once the investment has been made in setting up one’s mobile device to send and receive tweets, it may well be the most convenient way of interacting with Twitter. Typically, mobile agents remain logged in and remember passwords, eliminating the process of manually logging into Twitter via the web interface.

Thus, it is possible that the two distinct communities observed are simultaneously the result of (1) initial barriers to using mobile agents keeping much of the existing non-mobile population using computer-based interfaces and (2) the convenience and “stickiness” of mobile agents once incorporated into daily life keeping mobile users from reverting to non-mobile interfaces.

The mobile Twitter community is more active. The distributions in Figure 2 reveal that the exclusively non-mobile community has a distinctly larger contingent of inactive users. This can be observed in the discrepancies between the distributions for small values on the x-axis. Users with small number of statuses, followers, and followees are the least active members of a microblogging community: they contribute little content, what content they do produce is distributed to a relatively small community, and they follow too few other individuals to receive much information to consume. The non-mobile community when compared to the

mobile community has a much larger proportion of users who produce virtually no content (12% vs. 6% of users), maintain almost no followers (10% vs. 5%), and follow few other individuals (15% vs. 3%).

The relative scarcity of such inactive users in the mobile community may be due to two self-reinforcing factors. First, as mentioned earlier, the initial barrier for configuring and mastering a mobile Twitter agent may select for individuals who have more interest in the microblogging platform to begin with. Then, once the agent is setup, a mobile device lowers the barrier for continuing to participate since content can be generated and consumed throughout the day while in transit, during downtime, or while recreating. Thus, mobile agents may both select for individuals more likely to be active microbloggers and also make it easier for those individuals to be regular participants in the microblogging community.

Mobile content is more conversational. As can be seen in Figure 3(b), tweets that act as conversation initiators and responses are significantly more common in mobile tweets. Furthermore, mobile tweets are more often directed as well, which can be an important feature of starting or maintaining conversations. A third finding, that mixed mobile tweets are much more common than their non-mobile counterparts, adds additional support to this observation. Mixed tweets consist of a retweet with content added onto the beginning or the end. Very often we observed mixed tweets functioning as conversation responses—the retweet was the initiating tweet and the added content was the user’s response.

Overall, the finding that mobile microblogging is more conversational makes sense, particularly in light of the fact that mobile phones are now commonly used for text messaging, which is an inherently conversational form. It may be that Twitter on mobile devices is being used as a more flexible form of text messaging which can be used to reach many people at once.

On a side note, it is noteworthy that in both mobile and non-mobile data sets initiators are much less common than responses. This is likely due to the fact that tweets are al-

ways broadcasted. Thus, any initiator tweet, whether intended for one person or many, is delivered to all of the author's followers. This dramatically increases the number of individuals who may respond to the initiator. Thus, it may not be that there are few initiator tweets, but rather that there are many responses on average for every initiator.

Mobile content contains relatively few links. Table 1 indicates that mobile tweets contain on the order of 4 times fewer links than non-mobile tweets. This may be related to the fact that, according to Figure 3(b), mobile tweets contain dramatically less news and other content than non-mobile tweets, which are the main categories that contain links. As mentioned earlier, this lack of links is likely due to limitations in the functionality of mobile devices. In order to post tweets with links, one must obtain the link. This is typically done by copy and paste, which is currently a relatively unrefined or altogether unsupported feature on many mobile devices.

Mobile content is more personal. As shown in Figure 3(b) and (c), mobile tweets tend more often to be subjective and mobile users post relatively more status updates. Both of these types of content typically contain personal details either about the user and her views or about what the person is doing. Microblogging in the mobile context lends itself to more personal posts for at least two reasons which both stem from the fact that mobile devices are typically carried by individuals nearly all the time—when working, traveling, and recreating. These findings are in line with previous work by (Naaman, Boase, and Lai 2010), which indicated that personal content was more prominent on mobile platforms.

First, having a connection to Twitter available nearly all the time simply creates more opportunities for the user to make status updates. For example, often while traveling there can be downtime when the user can quickly create and post updates.

Second, having a persistent connection to Twitter means that the user can use it to capture and, potentially vent, emotions and feelings as she is having them. For example, after having locked oneself out of the house, a user can easily rattle off a post expressing exasperation, helplessness, or whatever other emotions she is experiencing.

Mobile tweets retain typical tweeting conventions. It is worth noting that, despite the differences highlighted, mobile tweets retain several of the major idiomatic features which characterize the form: retweets, emoticons, and hashtags all have near identical frequency in mobile and non-mobile tweets. It may be features such as these that unify the Twitter community and keep more distinct mobile and non-mobile communities from emerging within it.

Overall, our results show that the mobile context shapes microblogging behavior in a number of ways. Understanding how these behaviors are altered will provide significant insight into the present microblogging phenomenon and the way it will evolve over time. We identify a closer investigation into understanding these social and technological mechanisms as an important direction for future work.

References

- Boyd, D.; Golder, S.; and Lotan, G. 2010. Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In *43rd Hawaii International Conference on Systems Sciences*, 1–10.
- Honey, C., and Herring, S. C. 2009. Beyond microblogging: Conversation and collaboration via twitter. In *42nd Hawaii International Conference on System Sciences*, 1–10.
- Jansen, B.; Zhang, M.; and Sobel, K. 2009. Twitter power: Tweets as electronic word of mouth. *Journal of the . . .*
- Java, A.; Song, X.; Finin, T.; and Tseng, B. 2009. Why we twitter: An analysis of a microblogging community. In *Advances in Web Mining and Web Usage Analysis*, volume Lecture notes in artificial intelligence.
- Krishnamurthy, B.; Gill, P.; and Arlitt, M. 2008. A few chirps about twitter. In *Proceedings of the first workshop on online social networks*.
- Lenhart, A., and Fox, S. 2009. Twitter and status updating. Technical report, Pew Research Center.
- Naaman, M.; Boase, J.; and Lai, C.-H. 2010. Is it really about me? message content in social awareness streams. In *Conference on Computer Supported Cooperative Work*.
- Pak, A., and Paroubek, P. 2010. Twitter as a corpus for sentiment analysis and opinion mining. In *Conference on International Language Resources and Evaluation*.
- Radwanick, S. 2010. The 2009 u.s. digital year in review. Technical report, comScore.
- Ramage, D.; Hall, D.; Nallapati, R.; and Manning, C. D. 2010. Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora. *Proceedings of the Conference on Empirical Methods in Natural Language Processing* 248–256.
- Ritter, A.; Cherry, C.; and Dolan, B. 2010. Unsupervised modeling of twitter conversations. In *Conference of the North American Chapter of the Association for Computational Linguistics*.
- Ruths, D., and Perreault, M. 2011. Tweetcoder: a customizable application for manually coding twitter content. In *Conference on Weblogs and Social Media*, Submitted.
- Thelwall, M., and Wilkinson, D. 2010. Data mining emotion in social network communication: Gender differences in myspace. *Journal of the American Society for Information Science and Technology* 61(1):190–199.
- Webster, T. 2010. Twitter usage in america: 2010. Technical report, Edison Research.
- Yardi, S., and Boyd, D. 2010. Tweeting from the town square: Measuring geographic local networks. In *International Conference on Weblogs and Social Media*.
- Zhao, D., and Rosson, M. B. 2009. How and why people twitter: the role that micro-blogging plays in informal communication at work. In *Proceedings of the ACM 2009 international conference on Supporting group work*.