(How) Will the Revolution be Retweeted?
Information Diffusion and the 2011 Egyptian Uprising

Kate Starbird
ATLAS Institute, Technology, Media & Society
University of Colorado Boulder
Boulder, CO 80309-0430
starbird@colorado.edu

Leysia Palen
Department of Computer Science
University of Colorado Boulder
Boulder, CO 80309-0430
palen@colorado.edu

ABSTRACT

This paper examines microblogging information diffusion activity during the 2011 Egyptian political uprisings. Specifically, we examine the use of the retweet mechanism on Twitter, using empirical evidence of information propagation to reveal aspects of *work* that the crowd conducts. Analysis of the widespread contagion of a popular meme reveals interaction between those who were "on the ground" in Cairo and those who were not. However, differences between information that appeals to the larger crowd and those who were doing on-the-ground work reveal important interplay between the two realms. Through both qualitative and statistical description, we show how the crowd expresses solidarity and does the work of information processing through recommendation and filtering. We discuss how these aspects of work mutually sustain crowd interaction in a politically sensitive context. In addition, we show how features of this retweet-recommendation behavior could be used in combination with other indicators to identify information that is new and likely coming from the ground.

Author Keywords

Crisis informatics, information contagion, information diffusion, microblogging, political disruption, Twitter

ACM Classification Keywords

H.5.3 Groups & Organization Interfaces—collaborative computing, computer-supported cooperative work, K.4.2 Social Issues

General Terms

Design; Human Factors

INTRODUCTION

Microblogging tools have been appropriated for a wide range of applications, including, but certainly not limited to, the networking and socializing we have come to associate with social media. Several research studies have documented the use of platforms like Twitter (the most popular microblogging tool, which has global reach) and Sino-Weibo (primarily used in China) during mass emergencies and large-scale crises [12,18,20,22,27]. Messages about the use of microblogging, specifically Twitter, during political protests conflict. Though celebrated by the Western media as a force for good during the Iran Election protests in June 2009 [7], follow-up research suggests the role of Twitter had been over-stated, noting that the social media revolution failed to affect the change it intended, and that the service may have been used by the government to crush opposition protests and identify protesters, whose lives were put in immediate danger [4].

During the "Arab Spring"—a string of political uprisings that took place in the winter and spring of 2011 across the Arab world—attention was again paid to the emerging role of social media. Following Tunisia's successful demonstrations, protesters took to the streets of Egypt to demand reform [15]. As had happened during the Iranian Election protests, people across the world tuned into the event via social media; there again mainstream media described a significant role being played by platforms like Facebook and Twitter [16].

In his attention-grabbing editorial about the difficulties of using social media to affect revolutionary change—change he claimed required high-risk activism—Gladwell wrote that the "revolution will not be tweeted" [6]. Our empirical consideration of the events in Egypt calls this claim into question.

Here, we report on this research, finding—to foreshadow a small bit of it—that 30% of the 1000 mostly highly retweeted Twitterers who were using popular hashtags related to the protests *were in Cairo* during the event, and that many of these were "on the ground" in street protests. Tweets from these users contained information about meeting times, injuries, violence, supplies needed, etc. Revolutionaries were clearly using social media services to coordinate their actions and garner support. While we will not confront head-on Gladwell's argument that social media fail as organizing tools for high-risk activism, this paper will address a piece of that argument, asserting that the low-risk "activism" enabled by social media (what others have

termed "slacktivism" [13]) may indeed have been a productive component of this recent revolution.

In this paper, we examine the use of one Twitter feature—the retweet, a mechanism by which Twitter users pass on or forward information to other users. We use the retweet mechanism as a means to understand broader Twitter behavior around these protests, and to demonstrate how remote individuals participated in Egypt's 2011 revolution through low-risk, social media-enabled activities. To be clear about aims, we neither present a full description of Twitter activity during the protests, nor an analysis of the role that social media played (or didn't play) in *orchestrating* this event. Rather we show how consideration of the retweet mechanism reveals a good deal about information contagion across a large number of people and how this behavior figures into social movements.

To that end, we theorize that "the crowd" participated in at least two ways. By uncovering interaction via information propagation between those on the ground and those in the broader Twitterverse, we first demonstrate how the crowd, through a show of interest, expressed solidarity with the cause. We then show how some protesters embraced this show of interest and aligned with it, giving credibility to the idea of solidarity through social media. Second, we show how the crowd did the "work" of information processing (through retweeting), and discuss how we might be able to leverage that recommendation process to increase situational awareness during events.

BACKGROUND

Twitter and Mass Disruption
Twitter is a popular microblogging service with global reach, boasting well over 100 million registered users worldwide [25]. Although considerable attention has been paid by the media to the use of Twitter during political protests after the Iran Election and again during the Egyptian protests, few empirical studies have been done on the use of microblogging during political disruption. There is some literature that addresses this issue from a political perspective. Mungiu-Pippidi and Munteanu give an accounting of the role of microblogging during the failed "Twitter Revolution" in Moldova in 2009 [14]. Burns and Eltham warn that the hype around the role of Twitter during the Iran Election protests has been overblown and blinded to the negative consequences of the service's use [4].

A growing body of research in the area of crisis informatics, including several studies on the use of microblogging during mass emergencies and disasters, helps elaborate the topic of civil disruption and ICT. Though obvious disparities exist between these types of social convergence events—not least of which are the special brand of concerns about privacy and security that arise during planned internal protests—the behavior of the attendant audience as mediated by social media likely has some similarities by virtue of the disruptive nature of either kind of event.

Microblogging has been used during crisis events all over the world, including the 2007 San Diego Wildfires [12], the Oklahoma wildfires and Red River floods of 2009 [20,27], the 2010 Haiti earthquake [22], the 2010 Yushu earthquake in China [18], and other more recent earthquakes in Christchurch New Zealand [11] and Japan [17]. Research indicates that Twitter has been employed to seek information [27], solicit donations [18], organize volunteers [22], publicize the names of missing persons [11], and broadcast immediate needs [18], among other uses.

Information Propagation and Twitter
This paper focuses specifically on how information spreads through Twitter during political protests by way of the retweet mechanism. The retweet is a user-driven convention—now formally supported by one-click functionality on the Twitter platform—that acts to forward tweets, giving attribution to the original (or another upstream) author. Boyd et al. [3] found that retweets are used both for information diffusion and for engaging others. That study also noted the difficulty of identifying retweets and traces due to different syntaxes, user truncation, and added commentary across the tweet propagation. In the data collected for this study, we found four forms of the retweet convention in use: the most popular is `RT @username`. Also used at high volumes are `via @username` and `R @username` followed by the message, as well as `"@username:,` where the entire tweet is quoted.

Several studies have described dynamics of the retweet across large, random samples of tweets. Examining features that lead to increased retweetability, Suh et al. [24] found that tweets with URLs and hashtags were more likely to be retweeted. They also noted a strong linear relationship between a user's number of followers and the likelihood that that user's tweet would be retweeted. They suggested that "social context"—information about the Twitter author including followers, following, account inception date, etc.—can help identify the "value" of information. Kwak et al. [8] observed that retweets spread quickly and broadly, noting that the speed of diffusion could be an indication of strength of influence of the Twitterer, and suggesting that the number of followers a user has and the number of times that user is retweeted are different measures of popularity. Comparing Twitter and Digg, Lerman and Ghosh [9] found that information diffuses faster on Twitter, and that initial number of followers is not as closely correlated with retweetability as it is on Digg. Van Liere [26] used retweets to study the distance that a tweet travels and identified a type of Twitterer—the "information broker"—who connects to others according to shared interest and who spreads information across geographic distance.

Others have addressed information diffusion through microblogging sites during crises. Looking specifically at the retweet mechanism during crises, Starbird and Palen [21] found that retweets with topical keywords were more likely to be on-topic related to the disaster than non-retweets, suggesting a role of information recommendation

performed by the crowd. They also noted different retweet patterns between those local to a disaster—who preferred to retweet messages written by people who were also locals—over those who were remote—who retweeted almost exclusively messages that possessed "broad appeal." Vieweg et al. [27] noted that retweeted information is more likely to include information that contributes to situational awareness than non-retweeted information, and Qu et al. [18] also found that retweets were more likely to contribute to situational awareness as well as to contain "action-related" information.

Complex contagion, a concept from sociology related to information contagion [5], may be an important dynamic of social media use during political protest. Romera et al. [19] researched how socially sensitive topics, including political ones, propagated through Twitter. They identified two different properties of diffusion: *stickiness*, the likelihood of information being spread after one contact; and *persistence*, the likelihood of information being spread after repeated contacts. They found that complex contagion to be at work on Twitter, reporting that political hashtags were more persistent than other types of tags, meaning that they were more likely to be spread after multiple exposures. Twitterers who are initially unlikely to join a conversation on a sensitive topic become more likely to join as they see increasingly more people becoming involved.

METHOD

Event Description: The 2011 Political Uprising in Egypt

This paper investigates Twitter activity during the 2011 political uprising in Egypt, a "mass disruption" event with mass social and informational convergence properties. Mass protests of Egyptian autocratic governance began on 25 January 2011, and continued for eighteen days until Egyptian president Mubarak resigned on 11 February [15]. Early on, social media appeared to have an active presence. In reaction to the initial, mostly peaceful protests, the Egyptian authorities cut internet access to major providers on 25 January at 12:20am EET[1], with service largely resumed by 2 February. The government also moved to disband protesters using a security police force notorious for torture. By 28 January, hundreds of thousands of people had gathered in Tahrir Square [15]. An ad-hoc medical facility was set up near the square to tend those wounded in skirmishes that had broken out between security forces and the protesters [1]. 2 February marked a significant shift from relatively peaceful protests to violent clashes between pro-Mubarak groups and anti-Mubarak protesters. Several non-Egyptian reporters on the ground in Tahrir Square, a central location for the protests, and other parts of Cairo were reportedly attacked by pro-Mubarak "thugs," as they came to be called [15,23]. Protests continued and tensions increased over the following week, finally culminating on 11 February when Mubarak stepped down.

---

[1] All times in this paper in local Cairo time, GMT +02:00

Data Collection

This research is based on Twitter data collected between 2 February, hours before Internet access was restored in Egypt, and 15 February, four days after Mubarak stepped down. Using sophisticated search architecture that takes advantage of several Twitter APIs [2], we collected both tweet and Twitterer data. For tweet collection, we relied on Twitter's Streaming API to collect tweets in real-time, filtering on the following terms: `egypt`, `#egypt`, and `#jan25`.[2] This captured tweets in English, as well as a range of other languages, including tweets written in Arabic, which constitute 15% of the data. For each unique Twitterer who contributed a tweet to this collection, we also captured all of their Twitter profile information including follower count at the time of the first tweet grabbed by our collection. These data—2,229,129 tweets and 338,895 unique Twitterers—comprise the *Egypt Twitterverse Proxy* set. Due to technical issues with Twitter data collection, we were unable to collect data during the early stages of the event (between 25 January and 2 February) and lost data during three short windows within the collection period. However, the type of analysis conducted here does not depend on a full accounting of all possible tweets. Additionally, the delayed capture of profile data, which resulted in elevated initial follower counts for local Twitterers, likely weakened some of the effects we report on later in this paper.

Highly Retweeted Twitterers

Accepting the retweet feature to be a measure of popularity [8] and possibly a recommendation feature during mass disruption events [20], we then attempted to identify and examine the most highly retweeted Twitterers. Retweets (using `RT @` and `via @` conventions) make up 58.5% of the *Egypt Twitterverse Proxy* set (all languages combined). Furthermore, 56.3% of all tweets with non-English characters (which include Arabic language tweets) in that set are retweets, which suggests that the retweet conventions have been at least partially adopted by Arabic language tweeters, and indicates that retweeting continues to be an important behavior during mass disruptions.

Searching against the two retweet conventions mentioned above, we calculated how many times each Twitterer was retweeted within the *Egypt Twitterverse Proxy* set, and then identified the top 1000 most-retweeted accounts. We then cross-referenced these usernames against the usernames of the Twitterers whose profile information we had collected, and located the profile and follower data of 956 of the 1000 highly retweeted usernames. The 44 missing usernames are likely absent because users protected (or restricted access to) their accounts or because they untraceably changed their

---

[2] Collecting Twitter data during mass disruption events requires rapid selection of search terms in an evolving information space [19]. Arabic speakers in our research lab selected these terms as the most popular during the early days of this event, resulting in a large, low-noise sample.

account names within the data collection window. The 956 remaining users and all of their keyword tweets comprise the *Egypt Highly RTed Twitterers* set.

For further analysis, we randomly selected a one-quarter sample from these 956 Twitterers. This resulted in 254 initial Twitterers. We then removed six accounts: three were removed due to a zero following count, which indicates a problem with the account or an account suspension; two were removed because their account names had changed during the event, altering their retweet numbers; and a final account was removed because it was a bot with more than 10,000 automatically sent tweets during the time period. The remaining 248 Twitterers comprise the *Egypt Highly RTed Sample* set.

Measuring diffusion of individual tweets
To further understand information diffusion, we traced the propagation of all retweeted tweets that were originally authored or attributed to a Twitterer in the *Egypt Highly RTed Sample*. We were able to trace the history of 313,662 retweets back to 34,605 original tweets authored by a Twitterer in this sample. We calculated both how many times an author was retweeted, as well as how many different tweets of theirs were retweeted. We also captured how many times each tweet was retweeted.

DATA SET	# Twitterers	# Tweets
Egypt Twitterverse Proxy Tweets sent during Egyptian protests Identified using keyword search: #jan25 #egypt Egypt *Used to identify highly RTed & track tweet/meme diffusion.*	338,895	2,299,129
Egypt Highly RTed Subset of *Egypt Twitterverse Proxy* Top 1000 most RTed accounts, where profiles could be found.	956	282,010
Egypt Highly RTed Sample 25% sample from *Egypt Highly RTed* *Coded for Location of Author.*	248	69,461

Table 1. Egyptian Protest Data Set Statistics

Qualitative Coding
To understand the types of accounts that are most retweeted, we categorized the Twitterers in the *Egypt Highly RTed Sample* along several criteria, including their account affiliation and location during the event. Because users occasionally include locations in their profile that are not their actual locations, or because they move around during events, we do not accept the location character string in the user profile as the true location. Instead, we manually "code" users for location by reading through each user's entire collected twitter stream (which ranged from tens to thousands of tweets for each Twitterer) looking for assertions of or references to being in Cairo during the period of the protests. Because this study centers around the activities in Cairo's Tahrir square, we coded each account's

location relative to Cairo, using the following categorizations: *In Cairo, In Egypt but not in Cairo* and *Not in Egypt*. Twitterers who were in Cairo at any time during the event were coded as being *In Cairo*.

Many of the top 1000 most retweeted accounts in the Egypt data set had Arabic language tweets in their user streams. To make determinations of location and affiliation for Arabic language Twitterers, we first analyzed the profiles and user streams using an English translation generated by Google Translate software. When coding determinations could not be made in this way, we asked a native Arabic speaker to translate the tweet and profile data. Ultimately, we could not confidently determine the locations for three of 248 Twitterers in the *Egypt Highly RTed Sample*.

FINDINGS
Using both qualitative and quantitative analyses, we examine how the retweet was used to diffuse information during the Egyptian protests. We begin at the tweet level, investigating the propagation and re-appropriation of a single tweet meme that spread widely in the set. We then use a measure of the number of times a tweet is retweeted to uncover distinctions in retweetability among different types of information. Next, we shift back to look at the retweet as a method of recommending Twitterers as well as tweets, and examine the features of Twitterers who were most highly retweeted in our data.

Propagation of a Metaphor for the Revolution
The most popular tweet in the *Egypt Twitterverse Proxy* set is a proclamation of support for the removal of Mubarak, represented by an ascii-created "progress bar" graphic:

```
@adelshehadeh (10 Feb, 18:37): Uninstalling
dictator ... 99% complete ██████████████████
████████▒ #egypt #jan25 #tahrir #mubarak
```

Variations of the "Uninstalling dictator" with progress bar tweet appear 19,836 times in the data set. With this finding, we further investigated the morphology of the progress bar text and graphic, discovering that its use was widespread, cutting across Twitterverse, and began much earlier, progressing to this most popular form. It appears more frequently among those who are not local, though those who were on the ground also propagated it—a point to which we return at the conclusion of this section.

Over 1% of our *Egypt Twitterverse Proxy* set (23,012 tweets) reference the progress bar theme, as determined by inclusion of the block characters as well as occurrence of the % character. In total, of 338,895 Twitterers in the entire set, we found 20,727 who tweeted at least one progress bar tweet. We examined these tweets by hand to identify the major forms of the progress bar tweet and how they shifted over time, as well as the smaller changes that Twitterers made as evidence of the "work" that was put to sustaining and adapting the *meme* over time.

Throughout the time period of the protests, variations on the progress bar meme appear, morphing into new forms, but in most cases, keeping to and elaborating on a common

"computing" metaphor. The first occurrence happens the morning of 2 February, suggesting that the progress bar was already propagating, though in smaller numbers, prior to data collection start. The tweet is retweeted twice:

```
@NotNaif (2 Feb, 11:49): Hosni #Moubarak Escape
loading... [███████████████] 99% #alarabiya
#egypt #cairo #jan25 #FREEEGYPT #Internet #25jan
#Arab
```

The next form extends the computing metaphor that is implied by the progress bar. It is only retweeted once in this form, but introduces both the ideas of "installing" and "freedom" into the information space:

```
@AntiMubarak (3 Feb, 01:41): RT @thameralzaidy
Installing Freedom in Egypt [██████████████] 99%
#Jan25
```

The first Arabic language tweet with a progress bar appears almost two days later. Adjusting the sentiment of the tweet to reflect a slightly less positive outlook after the violent clashes that took place on 2 February in Tahrir Square, @almuraisy suggests that "some more patience" is needed. This tweet was retweeted 6 times.

```
@almuraisy (4 Feb, 14:26): [████████████████████]
99% جماعة يا هانت# #Egypt #Jan25 #Tahrir
```

Of note, only 52 instances of the progress bar tweet (much less than 1%) had Arabic characters in the text.

The next form expands upon the "freedom" idea, and makes a strong connection to the progress bar with "loading." This is the first progress bar to diffuse broadly in our set, appearing 5977 times in a similar format:

```
@fakroona (4 Feb, 14:39): FREEDOM LOADING
[████████████████] 99% #Egypt #Jan25 #Tahrir
#Cairo #mubarak #sidibouzid #fridayofdeparture
#yemen #syria #jordan
```

Interestingly, even though the "FREEDOM LOADING" and progress bar text, as well as the attribution to @fakroona, diffuses widely, the #fridayofdeparture hashtag is dropped in about half of the retweets, like the one below.

```
@StephanJourdan (4 Feb, 15:00): RT @fakroona:
FREEDOM LOADING [█████████████████] 99%
#Tahrir #egypt #jan25
```

All these major progress bar tweet forms include variations in hashtag listings, as well as attributions to the original and downstream Twitterers. Boyd et al. [3] report that text and attribution often change over the course of a retweet's life, an effect they attribute partly to character length limits of tweets and to added commentary by downstream Twitterers. We see evidence in this study of people in different locales trying to make the tweet their own by slightly adjusting the meaning to take into account the event progression, or contributing new tags or short comments.

In a move similar to @almuraisy's Arabic tweet above, @RuwaydaMustafah also adjusts the popular meme to acknowledge renewed uncertainly in the success of the protests, introducing an "error message" on 4 February:

```
@RuwaydaMustafah (4 Feb, 15:25): FREEDOM LOADING
[████████████████] 99% [ Error : Please remove
Mubarak and try again!] #Jan25 #Egypt
```

This tweet results in 324 instances, but is distributed fairly steadily until Mubarak steps down on the 10th. Notice that this tweet drops attribution to the original "freedom loading" author. Other variants of the metaphor appear, but are not as frequently propagated. Their presence nevertheless suggests attention to a "collective riff," and includes "downloading freedom" and the "blue screen of death" which appear 3 and 9 times respectively in our set:

```
@JameedKaraky (4 Feb, 17:21): #Jan25 Downloading
Freedom in #Egypt [████████████████] 99%
```

```
@BanglarPain (4 Feb, 14:50): @fakroona #Mubarak
Overload [███████████████] 99% Blue screen of
death! #Tahrir #jan25 #Egypt
```

The latter tweet is a direct reply to @fakroona, explicitly playing off of @fakroona's popular progress bar tweet. The investment into the collective riff continues with some attention to incrementing the percentage value associated with the progress bar by 0.1%:

```
@Ahmoooos (10 Feb, 17:57): FREEDOM LOADING
[███████████████] 99.8% [ removing mubarak in
processing right now please wait ! ] #Jan25 #Egypt
#Mubarak #tahrir #cairo
```

New forms continue with the addition of more, and shortened progress bars, here with attention to leaders. Another variation on this style appears in the form of other nation states in the Arab world where recent protests have taken place or future protests have been proposed:

```
@alihabibi1 (10 Feb, 18:21):
Ben Ali [██████] 100% Mubarak [██████] 99%
Qaddafi [██__] 42% Boutaflika [█__] 
19% #Jan25 #SidiBouzid
```

It was @adelshehadeh's tweet (at the beginning of this section), however, that generated the most propagation, and it didn't appear until 10 February, when the resignation of Mubarak was appearing imminent. That tweet accounts for 78% of the "progress bar" tweets in the set, and introduces the morphological change of "uninstalling."

Still more extensions of the computing metaphor continue to appear after the first appearance of the most diffused form of the meme, with references to "Egypt 2.0," "Fatal error," "kernel error," and "install Democracy, Disk #1 into drive E: gypt."

In these extractions of the major forms of the meme morphology identified here, we see the evolution of a tweet meme. The progress bar ascii graphic is probably not something that everyone can easily create. Once someone introduced it to the public sphere, it got picked up and modified. Though these different forms comprised a fairly high absolute number, their totals paled in comparison to one particular form that did not appear until 10 February. In the time leading up to that date, however, variations on the first form transformed it into a meme, a widely understood idea. Centola and Macy [5] call this "complex contagion," where multiple exposures from multiple "long-tie" sources

ignite and support diffusion. The remixing of elements, all keeping within a computing metaphor, shows some degree of shared understanding of its purpose, and kept the meme propagating. The variations show collective "work" happening on the meme to keep it relevant and interesting, until it reached the point of widespread coverage and possibly saturation with the 10th of February "uninstalling dictator" tweet.

We note that most progress bar tweets appeared in English, though a small number also appeared in Arabic, Japanese, Korean, Dutch, German and Spanish. We suspect that the number of tweets in Arabic was lower than expected, and that the overall distribution tended strongly toward English because of the desire for contributions to the collective meme to be widely understood. To this point, many of the authors who originated new versions of the progress bar were tweeting from other locations in the Middle East.

Propagation by Others v Locals
To the question of whether tweet propagations like this effectively contribute to the basic work of the revolution, we look to how much this meme appeared in the tweet streams of those who were on the ground in Cairo versus those who were not.

In the *Egypt Highly RTed Sample*, where we coded for location, Twitterers who were not in Cairo were considerably more likely (27.6%) to send a progress bar tweet than those who were in Cairo during the protesting (17.6%). This suggests that this "sticky" meme had broad appeal in the wider Twitterverse, offering a short abstract of the event [21]. Indeed, another interpretation of this finding might be that non-local Twitterers were boosted into highly retweeted status by the popularity of the progress bar meme. We will explore this further in the latter half of the Findings section.

Perhaps more interesting than the discrepancy here is that 13 of our sample's 74 highly retweeted Twitterers who were tweeting from Cairo *were* taking part in propagating the progress bar tweet, including @sandmonkey, one of the ten most retweeted accounts in our set. Tracing the origin of several of the most popular variations, we find little evidence that any began in the streams of Cairo Twitterers. This suggests that Twitterers on the ground were, to some extent, embracing external declarations of solidarity as legitimate messages and voices within their cause.

In sum, the propagation of the progress bar meme by both those on the "inside" as well as the large attendant audience suggests interplay between both realms. Though such a meme could be perceived as "silly" and incidental to the event—and indeed, some people publicly asked for its distribution to stop—such phenomena need attention to fully understand their impact on newsworthy and risky events like a political revolution. It could be that wide propagation of a tweet-idea helps cast a light on events happening in other parts of the world. Participation by those in the midst of the geographical event suggests active

endorsement of the meme. In other words, if those on the ground find value in the propagation of a tweet, then—in contrast to Gladwell's point [6]—perhaps that is one of the real measures by which one decides how much a role social technology plays in revolutions.

Broad Appeal v Local Utility
By tracing the most popular (highly retweeted) tweets originated by authors in the *Egypt Highly RTed Sample,* we see a similar bias towards broad appeal tweets, messages of solidarity with the Egyptian cause, as well high-level news.

@JoeUnfiltered: The people of #Egypt have shown the world that youth activism can change ANYTHING. 18 days has ended 30 years of oppression *[RT 3989x³]*

@DominicKavakeb: Its so vital that world understands this is not a divided #Egypt. This is the state attacking the people. #jan25 *[RT 2660x]*

@HuffingtonPost: BREAKING: Mubarak will step down, hand power to military tonight – Reuters http://huff.to/egyptnews #Egypt #jan25 *[RT 1233x]*

These were the three most highly retweeted tweets originated by any author in our sample. The top tweet was retweeted 3989 times, the second 2660 times and the final tweet 1233 times. As above, none of these tweets were originally authored by individuals who were in Cairo at the time of the protests.

Highly propagated tweets originated by locals often follow the broad appeal pattern as well. Tweets about violence, detained friends or colleagues, requests for solidarity, and humorous tweets are often the most retweeted messages authored by the group. The following tweet, retweeted 998 times, is the most popular tweet authored by a local.

@sharifkouddous: A couple just got married in Tahrir in front of army tanks. A revolution wedding. #Egypt #Jan25 *[RT 998x]*

While witty and humorous tweets experienced "sticky" retweetability, serious tweets remarking on violence, especially violence against the media, and asking for support were also among the most popular. These two tweets asking for support for the release of Ayman Mohyeldin, a reporter for Al Jazeera, were retweeted a total of 1386 times.

@nolanjazeera: the hashtag is #freeayman please retweet as much as u can to get this trending. Attacks on media in #Egypt MUST end now *[RT 985x]*

@AymanM: Ayman has been detained by #Egypt military. Will keep everyone posted on his status as things develop (tweeted by friend) *[RT 401x]*

Popular tweets from locals also included first hand reports of violence or tactical information. The following tweet, a real-time accounting of violence on the ground, written by an NBC news reporter, was retweeted 591 times. The two subsequent tweets, which received considerably fewer retweets relayed tactical information about the evolving protests.

---

[3] This notation indicates the number of retweets.

@richardengelnbc: #egypt.. Just saw a protester hit by a molotov .. Saw him catch on fire.. Other protesters managed to put him out.. *[RT 519x]*

@alaa: Army gave up and let us control flow of ppl at tv building #Jan25 *[RT 336x]*

@3arabawy: The army has allowed a group of 100 Mubaraks thugs into Qasr el-Nil bridge now. #Jan25 *[RT 175x]*

Less "popular" tweets by locals, as measured by the breadth of their diffusion within the larger Twitterverse, often contained more detailed information from the ground. The following tweets were retweeted fewer than 20 times each, but offer very specific details about the situation in Cairo.

@wilyawil: Im hearing from friends inside Tahrir Sq that best entrance is from Falaki st. #jan25 #egypt *[RT 18x]*

@mosaaberizing: Weve taken over Talaat Harb entrance as well. Just the musuems left now. #Jan25 *[RT 14x]*

The popularity of individual tweets shows a bias towards English language tweets. Though 34% of tweets in the *Egypt Highly RTed Sample* contain Arabic characters, only 20% of tweets that were retweeted from these Twitterers had Arabic characters. Arabic language tweets that were retweeted at least once received an average of five retweets, while non-Arabic tweets were retweeted an average of eight times. The first Arabic tweet from the Twitterers in our highly retweeted sample, a tweet from the @nytimes requesting an interview with someone on the ground, appears as the 83rd most popular tweet, retweeted 242 times.

@nytimes : في القاهرة او في اسكندرية او
كا Nadia8@ في انجليزي او عربي. في سويز و تحب ان تكم عن ال ضع؟
#Egypt #Jan25 *[RT 242x]*

Analyzing the diffusion of popular tweets through retweeting suggests that actionable or action-related information [18], information coming from the ground and, perhaps, appearing in the information space for the first time, can be found in the streams of local Twitterers, but not necessarily in their most popular tweets. Tracing the most popular retweets, even those sent by locals, may result in broad appeal messages. But tweets like the following may be more valuable to those acting on the ground, and perhaps to others who are remote, trying to better understand and support the situation on the ground:

@occupiedcairo: Medical supplies needed at temp. hospital at tahrir: neck supports and stitching thread #jan25 *[RT 39x]*

How might the noisy retweeting by the well-intentioned and easily entertained crowd be leveraged to home in on actionable information being tweeted from the ground?

Who Were the Most Retweeted Accounts

Possibly a more effective strategy than investigating the most popular tweets during the Egyptian protests is identifying the most popular Twitterers. The number of times that an account is retweeted can be used as a measure of popularity [8] or, during a specific, mass disruption event, an implicit recommendation by the crowd that an account is a useful source of information [21]. In this

section, we will describe features of the most highly retweeted accounts, and investigate the utility of using the retweet mechanism, in concert with other aspects of an account's "social context" [24] to home in on accounts with "local authority" [21].

Research has shown that many features of information propagation through social media have power law or other heavy-tailed distributions [9,10]. For our *Egypt Highly RTed* set, the number of times that an account was retweeted has a heavy-tailed distribution. Figure 1 shows this distribution, the number of accounts with x number of tweets, in linear-linear and log-linear scale. Note that the first bar in this distribution is at 170 tweets, the cut-off point for inclusion in the top-1000 Twitterer sample. The mean number of times retweeted is 1089, and the most highly retweeted Twitterer, @ghonim, was retweeted over 35,000 times.
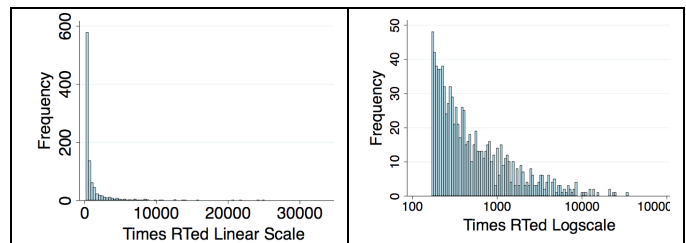

Figure 1. Times Retweeted, Linear-Linear & Log-Linear Scale

Top Ten Most Retweeted Accounts

Examining the top ten most retweeted accounts in the *Egypt Twitterverse Proxy* set (Table 2), we see the retweet acting to recommend Twitterers who, for the most part, were actively tweeting from Cairo. Many of these highly retweeted accounts were operated by journalists, or mainstream media with journalists on the ground. Others were popular Egyptian bloggers, and all have above average follower counts (see Figure 3).

Name	Times RTed	Followers	Affiliation	Location
@ghonim	35265	12491	individual	Cairo
@dima_khatib	25062	11320	journalist	Latin America
@bencnn	24066	21147	journalist	Cairo
@3arabawy	21607	9022	journalist	Cairo
@sandmonkey	20714	6943	blogger	Cairo
@alarabiya_ar	15681	14011	mainstream media	Arab World*
@monaeltahawy	14150	16048	journalist	Cairo
@ajenglish	13791	145246	mainstream media	Arab World*
@ajarabic	12687	30244	mainstream media	Arab World*
@monasosh	12609	2535	blogger	Alexandria

Table 2. Top 10 Most Retweeted Accounts (* denotes organizations with reporters on the ground.)

Examining the location of the Twitterers in the *Egypt Highly RTed Sample* (Table 3), we see a relatively high number of local Twitterers. 89 of 248 Twitterers in the sample (36%) have evidence within their tweet streams of being in Egypt at some point between 2 February to 15 February. Of these, 74 (30% of the sample) were in Cairo during that time. 43 others were located elsewhere in the Middle East and the rest were spread across the globe, with the highest concentrations in Europe (27) and North America (54). Among the 248 there were three for whom we could not determine if they were in Egypt or not.

Location	# Twitterers	% of Sample
Egypt – Cairo	74	29.8%
Egypt – Other	15	6.0%
Unknown, possibly in Egypt	3	1.2%
Arab World	43	17.3%
Outside Arab World	113	45.6%

Table 3. Location for the *Egypt Highly RTed Sample*

Though we do not have the locations of the remaining 800,000 plus Twitterers who sent one or more tweets hashtagged with `#egypt` or `#jan25`, it is probable that nowhere near a third of them were on the ground in the protests in Egypt. Highly retweeted accounts appear to be more likely than other Twitterers to be local in this event. This finding suggests that during mass-participation events of global import, the retweet may serve as a mechanism through which the larger Twitterverse identifies and recommends local Twitterers.
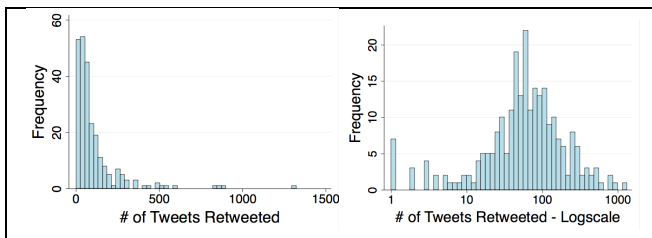

Figure 2. Number of Popular Tweets by Twitterer

Top 1000 Most Retweeted: Number of Retweets

As mentioned in the above section on Broad Appeal versus Local Utility, tracing the propagation of an individual tweet offers some insight into the value of that tweet measured by its contribution to situational awareness and newness in the information space. Similarly, the number of unique tweets, sent by the same author and propagated by the crowd, may be a measure of overall value of that user's tweet stream, in relation to a single event. In other words, though their overall number of times retweeted is the same, a Twitterer retweeted 500 times for a single tweet may be more likely to be a broad appeal Twitterer who is not on the ground, than a Twitterer retweeted 50 times for 10 different tweets.

Figure 2 shows the distribution of the numbers of individual popular tweets for each user in the *Egypt Highly RTed*

Sample in linear-linear and log-linear scale. The number of popular tweets for highly retweeted Twitterers has a log-normal distribution. The mean is 108 retweeted tweets, with seven Twitterers in our sample having tweeted only one Egyptian protest tweet that was retweeted, albeit very broadly. Twitterers in Cairo were retweeted, on average, for a higher number of tweets (123), than those not in Cairo (102). Though this comparison of the number of retweeted tweets between locals and non-locals shows some difference, things become more interesting when we break down this effect in relation to different variables, below.

Top 1000 Most Retweeted: Followers

For Twitterers in the *Egypt Highly RTed Sample*, the number of followers has a log-normal distribution with a long tail. The mean number of followers is 116,318, with a high standard deviation of 500,000. The median is 1730. Comparing the count across Twitterer location shows a stark difference between the median number of followers for a highly retweeted local (790) and a highly retweeted non-local (2609).
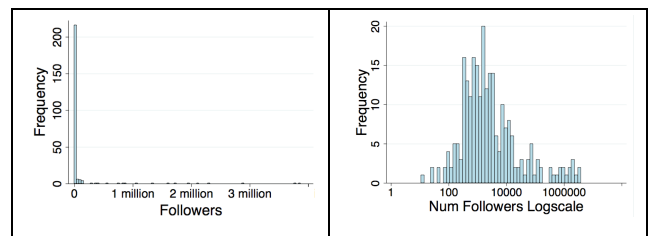

Figure 3. Number of Followers, *Highly RTed Sample*

Accounts with a higher number of followers get more exposure for their tweets, and therefore, each tweet they send has a proportionally higher chance of "infecting" the recipient or being passed on to others [18,24]. In our sample, the number of followers is positively correlated with the number of times a Twitterer is retweeted. However, the Pearson coefficient between the two variables is only 0.1306, indicating that follower number does not fully account for retweet popularity. While recognizing that retweets are one measure of popularity, Kwak at al. [8] noted this gap between the popularity measures of follower count and retweet count. The stark difference between average number of followers for highly retweeted locals and highly retweeted non-locals may suggest that during mass disruption events the follower-retweet gap is related to whether or not the Twitterer is in or near the affected area.

Using retweets and "social context" to identify locals

In the final section of this analysis, we will address the possibility of leveraging what we know about the retweet-follower popularity gap, as well as other measures of retweet behavior and social context to identify Twitterers and tweets from "the ground."

The preceding analysis suggests a relationship between different measures of tweet behavior, social context, and location. To evaluate these relationships we did a multinomial logistic regression, using location (*In Cairo, In*

Egypt, Not in Egypt, Unknown) as a polytomous (non-ordered) dependent variable. For our independent variables, we used follower count, number of times an individual is retweeted, number of different tweets authored by a Twitterer that are retweeted by the crowd, and the total number of total keyword tweets that each user sent during the collection window. To normalize the heavy-tailed distributions, we did natural log transformations on all of the independent variables.

Variable	Coef.	Std. Err	Z	P>|z|
log_followers	-.691792	.1286891	-5.38	0.000
log_timesRTed	.5020533	.1802369	2.79	0.005
log_numtweets	-1.08295	.3713641	-2.92	0.004
log_uniques	1.18107	.4061834	2.91	0.004
_cons	1.716295	1.44004	1.19	0.233

Table 4. Multinomial Logistic Regression:
***In Cairo* case vs *Not in Egypt* base case**

Treating *Not in Egypt* as the base case for the analysis, we find statistically significant effects from all four independent variables, but only for those who were in Cairo during the event.[4]

Log number of followers and log number of tweets both have significant negative effects on a Twitterer's odds of being in Cairo during the event. The log number of times retweeted and the log number of different tweets that were retweeted were significantly, positively related to being in Cairo. These descriptive statistics suggest that a *highly retweeted* Twitterer is only more likely to be local when their initial follower count is low. Number of followers and total number of keyword tweets contributed raise a Twitterer's chance of being retweeted, but do not make them more likely to be local. These results also support the idea that a Twitterer with only one or only a few popular tweets related to an event is less likely to be local than a Twitterer with many different popular tweets.

CONCLUSION

Twitter content during mass emergencies consists of both "original" information—that coming into the information space for the first time—and "derivative" information which can be found elsewhere, as our previous research indicates [19]. Retweets, which constitute nearly 60% of tweets sent with tags referencing the Egyptian protests, are one form of derivative information. From this point of view, one could consider retweets as merely noisy output and members of the crowd only cheerleaders.

Another view that incorporates the ideas put forth in this paper is that retweets serve as a crowd-powered recommendation system. Plenty of Twitterers put thought

---

[4] One other statistically significant effect was found for log number of followers on the Unknown case, which had a slightly negative effect, with P = 0.025 significance.

into what they retweet [3]. This constitutes a form of work—and sometimes a form of collective work, as we saw in the propagation and morphing of a metaphor-driven tweet across the duration of the Egyptian protests.

We also demonstrate how considering the crowd from this angle of a working collective can help identify, among a vast number the tweeters, locals who may be introducing "original" information into the space. However, this process is not so straightforward such that finding the most popular tweets or most popular Twitterers will solve the "noisy crowd" problem. Instead, it is important to understand how and why people retweet what they do, including the distinction between broad appeal and local utility tweets, and leverage this understanding to home in on new, locally relevant information. This can be useful as state of the art moves forward toward more automatic forms of curation and collation of vast computer-mediated communication.

When we consider these tweet activities of the locals—who, in this case, were closest to the activist cause and performed some of the work of activism as represented by their tweets—in relation to the activities of the on-lookers, the interaction between them suggests a more integrated relationship than the "slacktavism/activism" [6,13] construct suggests. Here we see that parts of the crowd activity do more than create noise, and that the crowd's zealous but safe vicarious participation has perhaps a more meaningful and new connection to the hard work of the social movement than is readily visible.

This research on the use of the retweet mechanism shows the crowd doing work in two ways that make it a *functional*, if not necessarily central (and that debate needs to continue), aspect of social movements. First, expression of social solidarity is found through the collective, observable creation of a tweet-based, visual and metaphorical meme. This is a deliberate form of collective work. It is critical to recognize that the meme creation was also visibly acknowledged by local activists. These expressions of social solidarity, we propose, are the kinds of activities that draw and sustain attention on a cause, which in turn may sustain the cause itself. The second form of work happens individually but collectively has an effect—retweeting as information filtering and recommendation.

The behaviors of people both on the perceived "inside" and "outside" of the social movement show that responsibilities and weights of responsibility differ across the collective. However, the behaviors are interrelated, and as such, have elements of mutual benefit. Social movements indeed happened prior to social media, but we would be remiss to imagine social movements of the future not leveraging and adapting these tools to gather more to a cause.

REFERENCES

1. Al-Ghazawy, O. Tahrir field hospital saves hundreds of lives, Nature Middle East, 2011. Retrieved 15 March, 2011. http://www.nature.com/nmiddleeast/2011/110214/full/ nmiddleeast.2011.18.html

2. Anderson, K. & Schram, A. Design and Implementation of a Data Analytics Infrastructure In Support of Crisis Informatics Research. In the *33rd International Conference on Software Engineering*, 21-28. (Honolulu, Hawaii, 2011)

3. boyd, d., Golder, S. & Lotan, G. Tweet, tweet, retweet: Conversational aspects of retweeting on Twitter, *Proc. of HCSS'10*, (Koloa, Hawaii, 2010), 1-10.

4. Burns, A. & Eltham, B. Twitter free Iran: An evaluation of Twitter's role in public diplomacy and information operations in Iran's 2009 Election Crisis. In Communications Policy & Research Forum, (2009).

5. Centola, D. & Macy, M. Complex contagion and the weakness of long ties. *American Journal of Sociology*, 113:702-734, 2007.

6. Gladwell, M. Why the revolution will not be tweeted. The New Yorker, 2010. Retrieved May 22, 2011. http://www.newyorker.com/reporting/2010/10/04/101004fa_fact_gladwell

7. Grossman, L. Iran's protests: Why Twitter is the medium of the movement. *Time,* 2009. Retrieved May 22, 2011. http:// www.time.com/time/world/article/0,8599,1905125,00.html

8. Kwak, H., Lee, C., Park, H. & Moon, S. What is Twitter, a social network or a news media? *Intl. WWW Conference,* (Raleigh, NC, 2010), ACM, 591-600.

9. Lerman, K. & Ghosh, R. Information contagion: An empirical study of the spread of news on Digg and Twitter social networks, in *4th Intl. AAAI Conference on Weblogs and Social Media*, (2010), AAAI, 90-97.

10. Leskovec, J., McGlohon, M., Faloutsos, C., Glance, N., & Hurst, M. Cascading behavior in large blog graphs, in *SIAM Intl Conference on Data Mining*, (2007).

11. Manhire, T. & Tran, M. Christchurch Earthquake: Hunt for survivors begins. *The Guardian*, 2011. Retrieved May 31, 2011. http://www.guardian.co.uk/world/2011/feb/22/christchurch-earthquake-hunt-for-survivors

12. Messina, C. Twitter hashtags for emergency coordination and disaster relief. *FactoryCity*, 2007. Retrieved May 31, 2011. http://factoryjoe.com/blog/2007/10/22/twitter-hashtags-for-emergency-coordination-and-disaster-relief/

13. Morozov, E. The brave new world of slacktivism. Net.Effect Foreign Policy, 2009. Retrieved: May 31, 2011. http://neteffect.foreignpolicy.com/posts/2009/05/19/the_brave_new_world_of_slacktivism

14. Mungiu-Pippidi, A., & Munteanu, I. Moldova's "Twitter Revolution" *Journal of Democracy 20*,3 (2009) 136-142

15. New York Times. Egypt News—Revolution and Aftermath. *New York Times*, 2011. Retrieved 31 May, 2011. http://topics.nytimes.com/top/news/international/countriesandterritories/egypt/index.html

16. New York Times. Times Topics – Wael Ghonim. *New York Times,* 2011. Retrieved 15 March 2011. URL: http://topics.nytimes.com/top/reference/timestopics/people/g/wael_ghonim/index.html

17. Neubig, G., Matsubayashi, Y., Hagiwara, M., & Murakami, K. Safety Information Mining – What Can NLP Do in a Disaster. *Proc of Joiru Conf. on Natural Language Processing*, (Chiang Mai, Thailand, 2011), 965-973.

18. Qu, Y., Huang, C., Zhang, P. & Zhang, J. Microblogging after a major disaster in China: A case study of the 2010 Yushu Earthquake*, Proc. of CSCW'11*, (Hangzhou, China, 2011), 25-34.

19. Romero, D., Meeder, B. & Kleinberg, J. Differences in the mechanics of information diffusion across topics: Idioms, political hashtags, and complex contagion on Twitter, in *Intl. WWW Conference*, (Hyderabad, India, 2011), ACM, 695-704.

20. Starbird, K, Palen, L, Hughes, A & Vieweg, S. Chatter on *The Red*: What hazards threat reveals about the social life of microblogged information, *Proc of CSCW* (2010), 241-250.

21. Starbird, K. & Palen, L. Pass it on? Retweeting in Mass Emerency, *Proc. of ISCRAM Conference*, (Seattle, WA, 2010).

22. Starbird, K. & Palen, L. "Voluntweeters": Self-organizing by digital volunteers in times of crisis, *Proc of CHI'11,* (Vancouver, CA, 2011), ACM.

23. Sweney, M. Al-Jazeera office attacked in Egypt protests. *guardian.co.uk,* 2011. Retrieved 15 March, 2011. http://www.guardian.co.uk/media/2011/feb/04/al-jazeera-office-attacked-egypt-protests

24. Suh, B., Hong, L., Pirolli, P. & Chi, E.H. Want to be retweeted? Large scale analytics on factors impacting retweet in Twitter network, in *IEEE Intl Conference on Social Computing*, (2010), IEEE, 177-184.

25. Twitter Blog. The evolving ecosystem. 2010. Retrieved May 30, 2011. http://blog.twitter.com/2010/09/evolving-ecosystem.html

26. van Liere, D. How far does a tweet travel? Information brokers in the Twitterverse, in *MSM'10*, (Toronto, 2010)

27. Vieweg, S., Hughes, A., Starbird, K., & Palen, L. Micro-blogging during two natural hazards events: What Twitter may contribute to situational awareness, *Proc of CHI'10*, (Atlanta, GA, 2010), ACM, 1079-88.