International Journal of Communication 5 (2011), Feature 1375-1405 1932-8036/2011FEA1375

The Arab Spring
The Role of ICTs

The Revolutions Were Tweeted: Information Flows During the 2011 Tunisian and Egyptian Revolutions

GILAD LOTAN
Social Flow

ERHARDT GRAEFF
Web Ecology Project

MIKE ANANNY
Microsoft Research

DEVIN GAFFNEY
Web Ecology Project

IAN PEARCE
Web Ecology Project

DANAH BOYD
Microsoft Research

This article details the networked production and dissemination of news on Twitter during snapshots of the 2011 Tunisian and Egyptian Revolutions as seen through information flows—sets of near-duplicate tweets—across activists, bloggers, journalists, mainstream media outlets, and other engaged participants. We differentiate between these user types and analyze patterns of sourcing and routing information among them. We describe the symbiotic relationship between media outlets and individuals and the distinct roles particular user types appear to play. Using this analysis, we discuss how Twitter plays a key role in amplifying and spreading timely information across the globe.

Introduction

The shift from an era of broadcast mass media to one of networked digital media has altered both information flows and the nature of news work. Mainstream media (MSM) outlets have adopted Twitter as

Please note the postscript that follows this article.

a means of engaging with and enlarging audiences, strengthening their reach and influence while also changing how they rely on and republish sources. During unplanned or critical world events such as the Tunisian and Egyptian uprisings, MSM turn to Twitter, both to learn from on-the-ground sources, and to rapidly distribute updates.

In this paper, we analyze Twitter information flows during the 2011 Tunisian and Egyptian uprisings. Our information flows are drawn from two datasets of public tweets, each shared during a period of approximately one week. The first covers the Tunisian demonstrations from January 12–19, 2011; the second covers the Egyptian demonstrations from January 24–29, 2011. We analyzed both data sets to identify different types of users who posted to Twitter regularly, sorting them into what we call "key actor types," e.g., MSM organizations, individual journalists, influential regional and global actors, and other participants who actively posted to Twitter on these two revolutions. We look at how each actor produced and passed information over the networks of Twitter users. In each case—Tunisia and Egypt—we describe how information flowed across different actor types and discuss why we see certain patterns. We conclude by discussing the symbiotic relationship between news media and information sources.

Context

Tunisia and Egypt

The Tunisian Revolution, which successfully ousted longtime President Zine El Abidine Ben Ali, consisted of a series of street demonstrations in January 2011 following the self-immolation of Mohamed Bouazizi on December 17, 2010 ("Timeline: Tunisia's uprising," 2011). The demonstrations were an expression of citizens' frustration over economic issues like food inflation and high unemployment, as well as a lack of political freedoms like rights to free speech (Mohyeldin, 2011; Sadiki, 2010). The phrase "Sidi Bouzid" (Bouazizi's home city) became shorthand for the revolt. On Twitter, participants began labeling messages discussing the uprisings with #*sidibouzid*, effectively indexing the Tunisian Revolution through a hashtag.[1] Despite President Ben Ali's attempts to quell the demonstrations through violence and last minute reforms, the Tunisian military intervened against loyal security forces, leading to the January 14 resignation of Ben Ali.

Following the success of the Tunisian protesters, opposition groups and activists in Egypt organized a demonstration in Cairo for January 25, 2011—National Police Day—to protest abuse by police ("Timeline: Egypt's revolution," 2011). These protests also emerged from similar frustrations with unemployment, corruption, and the lack of political freedoms, with #*Jan25* becoming the common Twitter hashtag used to mark messages relevant to the Egyptian Revolution. The Egyptian protests were well-organized through both old and new media, with veteran new media activists, such as the "April 6 Youth Movement," using social media, blogging, and video sharing to encourage people to protest (Kirkpatrick & Sanger, 2011). A series of protests involving civil resistance—which were illegal in Egypt—ensued over

[1] When a Twitter user places a "#" before a string of text, that string can then be clicked as a link to a global search of tweets using that string, a platform feature meant to facilitate a global discussion on a topic beyond a user's follower network. These are called hashtags.

several weeks, spreading to other major cities in the country, resulting in violence as protesters clashed with police forces loyal to longtime President Hosni Mubarak ("Security forces to deal," 2011). The military refused to fire upon the protesters, most notably in Tahrir Square, where protesters camped out in civil resistance. Mubarak resigned on February 11, 2011.

Both revolutions featured prominent use of social media, both by activists organizing the demonstrations, and by those disseminating or discussing news of the events locally and globally. Twitter emerged as a key source for real-time logistical coordination, information, and discussion among people, both within the Middle East and North Africa (MENA) and across the globe. This was especially true in Tunisia, where, prior to the uprisings, few mainstream media organizations had a formal presence or staff. By the time the Egyptian protests began, MSM news outlets began using both old and new media to document the uprisings. Al Jazeera covered the Egyptian Revolution with streaming video starting with the "Friday of Anger" on January 28, 2011, while journalists from Western media organizations started reporting from Egypt at a later stage. Most starkly, the Internet's role as a key networked information infrastructure was seen in the Egyptian government's decision to deny citizens access to it from January 26–February 2 ("Timeline: Egypt's revolution," 2011).

As these events unfolded, Twitter served both as a common medium for professional journalism and citizen journalism, and as a site of global information flow. People from around the world tuned in to Twitter feeds to learn about the revolutions and share what they learned. Although heavily critiqued (Nelson, 2011; Taylor, 2011), countless TV stations, commentators, and government officials noted the key roles that Facebook and Twitter had in facilitating the revolutions, raising critical questions about the role of social media in the production and dissemination of news about the "Arab Spring" uprisings.

Organizational and Networked Production of News

Studies of mainstream news production can be understood in terms of two broad phases of research. The first focuses on how journalists work within formal news organizations, while the second—a newer body of literature—investigates how news emerges from networked actors who span different professional and organizational identities and contexts.

Organizational studies of news production find that mainstream, professional journalists tend to structure their work according to a set of unspoken heuristics. Excluding longer-term reporting like investigative and feature journalism, daily news was largely understood in terms of physical beats and an expectation that news occurs at particular places in predictable moments (Gans, 1979; Molotch & Lester, 1974). Such reporting norms exist both within and among news organizations. For example, reporters, editors, and publishers of smaller-sized papers historically looked to well-established organizations like *The New York Times* for cues about how they should behave and what kind of news they should report—creating a kind of "arterial effect" (Breed, 1955) in which editorial staff tried to mimic journalists who were thought to produce high-quality news. Often unconsciously, journalists tend to produce news designed for their publishers and editors,[2] seeking informal feedback on their work from friendly non-journalists—a "known audience" of friends, neighbors, and family members (Gans, 1979, p. 236). Mainstream

[2] See Breed (1955) for a foundational study and Chomsky (2006) for an update.

professional journalists often write for a personal reference group (Darnton, 1975, pp. 182–188), have a "fear" of large, undifferentiated audiences, (Gans, 1979, p. 235), and see those who initiate contact with a newspaper (e.g., those who write letters to the editor) under an "idiom of insanity" (Wahl-Jorgensen, 2002). These and other strategies are part of well-established and strategic "rituals of objectivity" (Tuchman, 1972) that journalists use to distance themselves from audiences, relying instead on professionally sanctioned cues about what constitutes "good" editorial judgment.

In contrast, studies of the networked press tend to see news production in terms of connected actors who span different organizational contexts, personal and professional identities, geographic locations, and normative models of journalism. To be sure, the old organizationally situated dynamics still exist online, but with some important differences:

1) Mainstream news organizations still mimic each other's coverage in leader-follower relationships, albeit within a single day and faster news cycles (Boczkowski, 2010).

2) Mainstream news organizations and blogs exhibit predictable patterns in which the professional press still lead bloggers, albeit very quickly, to particular stories (Leskovec, 2009).

3) Although there are certainly more ways for people to participate in or comment on online news—and ways of commodifying and deriving revenue from readers who participate (Alexander, 2010; Vujnovic et al., 2010)—journalists tend to be deeply skeptical about how valuable or relevant user involvement is to their work, worrying that low-quality content may displace their professionally produced work and result in degraded overall news environments (Hermida & Thurman, 2008).

What is not well understood is exactly how mainstream news organizations—who have traditionally prided themselves on a kind of professionalism not easily accessible to general publics—understand and negotiate their identities and unique functions in networked news environments. This paper is one step in understanding how mainstream news organizations relate to, rely upon, and distinguish themselves from "non-professionals" within the context of Twitter information flows during fast-breaking newsworthy events.

Social Media and the News

Launched in 2006, Twitter is a microblogging service that was built by much of the same team that created the popular blogging service Blogger. Twitter was designed to let participants post short 140-character textual updates that could easily be disseminated via text messaging. Although Twitter was designed with the mobile phone in mind, many of Twitter's users consume content through the Web service or third-party software applications. Drawing on the features of social network sites, Twitter lets participants "follow" other users, although it does not require reciprocity. Thus, there are a handful of users with millions of followers, while the majority of users only have dozens of followers. Some Twitter users are quite influential, broadcasting messages that are widely received, while others have smaller spheres of influence (Cha, Haddadi, Benevenuto, & Gummadi, 2010).

As of October 2010, Twitter had more than 175 million registered users (Rao, 2010), and in March 2011, Twitter reported that a billion tweets are sent per week (Penner, 2011). What people do on Twitter varies tremendously. Some use the service to communicate with close friends, attract attention, or seek the attention of celebrities like Justin Bieber (Marwick & boyd, 2011). Others use the service to track news, share information, and offer links. Twitter has been used for a host of news-related events, including serving as a backchannel to American TV political debates (Shamma, Kennedy, & Churchill, 2009), a site for coordination in emergency events (Hughes & Palen, 2009; Sutton, Palen, & Shklovski, 2008), and a space for making sense of emergent news events (Yardi & boyd, 2010). While the practices on Twitter vary, they are not segregated; individual participants may post personal notes intended for friends alongside links to important news topics.

The relationship between social media and the press has become increasingly complex, as self-described non-professional journalists, using tools like Twitter, begin to influence and co-construct the kind of news traditionally produced by mainstream broadcasters. This has prompted scholars to question whether Twitter is a social media service or a news medium (Kwak, Lee, Park, & Moon, 2010). At an institutional level, traditional mainstream media journalists express concern about how their long-standing business models, professional standards, and relationships with the public are changing as information has begun to flow among multiple, networked actors with no stated journalistic affiliations or ethics (Carlson, 2007; Lowrey, 2006; Overholser, 2009; Singer, 2007; Singer, 2010). When these emergent, hybrid news ecosystems are analyzed, it is often unclear how networked information actors influence each other, who they look to as authorities, what kind of diversity exists among them, how professionals insert themselves into such networks, and how professionals use social media tools and sources in their own reporting (Braun & Gillespie, 2010; Hindman, 2008; Kelly, 2010; Wallsten, 2007).

Essentially, there is an evolving and dynamic relationship between the traditional, mainstream press organizations that have historically broadcasted news to audiences, and the emerging networked actors who consume mainstream news stories, remix and interpret them, and sometimes conduct original, high-quality reporting that stands alongside professionally produced content.

Twitter and Information Flow

Social media, and Twitter in particular, is a new venue for studying how people communicate and how information flows. Unlike many genres of social media, blogging and Twitter do not encourage reciprocal sharing; however, they have been shown to enable rapid information flow. Marlow (2005, p. 37) argues that information flows on blogs are a curious combination of broadcast diffusion and media “contagion,” emphasizing the person-to-person dissemination of information. Likewise, Kwak et al. (2010) conclude that the non-reciprocal nature of information sharing on Twitter means that it operates more like an information-sharing network than a social network, complete with well-positioned influencers who can shape how information flows.

This information-sharing behavior has been studied for decades using the two-step flow theory of communication first developed by Katz and Lazarsfeld (1955). They determined that mass media had very little direct effect on how citizens voted, and that the disproportionately greater influence came from other people with whom they regularly associated. These individuals were termed “opinion leaders.” Wu et al. (2011) tested the two-step flow theory on Twitter for normal traffic and found strong similarities in their information diffusion models.

One way of conceptualizing information flow on Twitter is through the frame of “information cascades,” or situations where “it is optimal for an individual, having observed the actions of others ahead of him, to follow the behavior of the preceding individual without regard to his own information” (Bikhchandani, Hirshleifer & Welch, 1992, p. 944). On Twitter, information cascades are easily amplified through the common practice of “retweeting” content, or reposting the content while referencing either the source of the content or the last person who shared it (boyd, Golder, & Lotan, 2010), and the use of hashtags, which make it easier for participants to follow content on a particular topic (Romero, Meeder, & Kleinberg, 2011). Finally, Twitter’s “trending topic” feature highlights content that its algorithm determines to be collectively related and statistically outstanding across the system at a point in time. Thus, if several users suddenly start talking about Egypt, “Egypt” becomes visible to all users through the trending topic feature. Twitter’s features and people’s tweeting practices thus make it easy for information cascades to occur.

Twitter Revolution?

Given that Twitter and other social media tools can be leveraged to spread information, Shirky (2009) has argued that social media may have the potential to provoke and sustain political uprisings by amplifying particular news and information.[3] After the 2009 Iranian election, many Twitter users altered their profile images so that they were tinted green (the color of the revolution) and switched their location to Tehran in a sign of solidarity with the movement. In addition, a remarkable spike in user account creation was seen during the event, further indicating the close relationship between Twitter and critical world events (Gaffney, 2010).

The aim of this article is to investigate the role of different types of social media actors in spreading information on Twitter during critical, time-sensitive world events. In situations like these, it is often difficult to distinguish between truthful information and rumor, or even to understand where information originates and how it changes over time. To help us make these distinctions and determinations, we need to understand the dynamics of *how* information spreads among networked actors.

This study of these dynamics is bounded by the context of information flows among a series of very similar tweets that are posted and reposted by users within a particular timeframe. By looking at these flows, we can identify key characteristics: who starts an information flow, what types of actors are

[3] For different popular perspectives on the role of social media in these contexts, compare the utopian views of Shirky (2009), the more critical views of Gladwell (2010) and Morozov (2010), and Zuckerman’s (2011) analysis of limitations such contexts place on the mainstream media’s ability to report news.

involved in the flow, how many users participate, and which actor types appear to be more successful in spreading information. We then divide each flow into sub-flows and analyze recurring patterns among actor types. With this data set and methodology, we are unable to comment on how information or understandings change over time, make broad statements about all social media or other uses of Twitter, or even say exactly why information flowed in the way it did on Twitter. Rather, our aim here is to focus on the flow of communication on Twitter, and to describe how communication begins from and disseminates among actors that, together, constitute a new kind of global audience.

Methodology

Our data collection and analysis involved three steps: data collection, information flow identification, and actor type classification. Each step is described in detail below. We conclude with a discussion of the study's methodological limitations.

Data Collection

Our analysis is based on two datasets acquired during the height of the Tunisian and Egyptian uprisings. We used the Twitter application programming interface (API) to query for the most recent Twitter posts every 5 minutes, requesting the last 100 publicly posted tweets (i.e., those from accounts that are not "protected") containing specific chosen keywords.

The first dataset includes 168,663 tweets posted January 12–19, 2011, containing the keywords "#sidibouzid" or "tunisia." The second includes 230,270 tweets posted January 24–29, 2011, containing the keywords "egypt" or "#jan25." Although not every relevant post included such keywords, their use was widespread among Twitter users, giving us a fairly representative sample of Twitter posts within the two time segments. We identified 39,696 distinct users in the Tunisia dataset, and 62,612 in the Egypt dataset.

From February 12–13, 2011, we queried the Twitter API for all publicly available profile information for each user appearing in either dataset, gathering self-reported location and time zone. Although we collected this information after the two events occurred, we estimate that it was not significantly skewed from the observed time segments.

Information Flow Identification

We define an information flow as an ordered set of near-duplicate tweets. We identify these flows by finding very similar tweets in our datasets using the shingling method for string comparison[4] (Manning, Raghavan, & Schütze, 2008), which converts a string of text (such as a tweet) into a fingerprint summary of the words it comprises. This fingerprint can then be efficiently compared against other strings (other tweets) to find near-duplicates. This methodology parallels the one used in Lotan's (2009) visual analysis of tweets surrounding the 2009 Iranian election protests.

[4] Twitter has a built-in function for retweeting, which produces metadata available via the API; however, we chose to use string comparison to find all retweets, since not all users use the built-in function.

Using the shingling method, we identified a total of 20,848 Tunisian and 29,403 Egyptian flows with size greater than two (involving at least 2 near-duplicate tweets that were posted by different users). For each dataset, we sorted these sets of flows by total number of tweets, thus creating a rank-ordered set of retweeted posts. Since our goal was to characterize the most common information flows and assess users' roles in dissemination, we wanted to make sure that our sample set included the longer flows, and not flows that consisted of small numbers of users. For that reason, we selected the top 10% of this rank-order. We recognize that this is not a representative sample of tweets but, rather, a selection of the most prominent information flows. In the same manner, we could have defined a flow as any group of retweets that included 19 or more posts in the Egypt dataset, and 16 or more in the Tunisia dataset.

We then randomly chose one-sixth of this top 10%, which resulted in a sample of 500 Egypt and 350 Tunisia flows. Out of these chosen flows, we extracted a list of users, whom we then classified into actor types (discussed in the next section).

To summarize how we arrived at the chosen flows, we did the following: 1) classified tweets that were very similar into bins, 2) sorted bins by size (number of tweets included), 3) chose the top 10%, and then 4) randomly chose one-sixth of them to identify a total of 850 flows which we would analyze in more detail.

Classifying Actor Types

There is a growing body of literature investigating how to quantify influence on Twitter, most prominently via in-degree (number of followers), retweets, mentions (Cha et al., 2010), and TunkRank (takes into account followers of followers) (Tunkelang, 2009). In order to determine how information flows between actor types, we had to cut down the number of users that would be hand coded. We selected 963 users total, from both the Egypt and Tunisia datasets, who either were first to post in a flow, or were retweeted or mentioned at least 15 times. Of these 963, 774 were part of our Tunisia dataset, and 888 were present in our Egyptian dataset; 699 (or 73%) of the actors we coded were involved to some extent in both datasets. We developed a classification schema based on the following types of actors, which was refined through several phases of coding:

- Mainstream media organizations ("MSM"): news and media organizations that have both digital and non-digital outlets (e.g., @AJEnglish, @nytimes).

- Mainstream new media organizations ("Web news orgs"): blogs, news portals, or journalistic entities that exist solely online (e.g., @HuffingtonPost).

- Non-media organizations ("non-media orgs"): groups, companies, or organizations that are not primarily news-oriented (e.g., @Vodafone, @Wikileaks).

- Mainstream media employees ("journalists"): individuals employed by MSM organizations, or who regularly work as freelancers for MSM organizations (e.g., @AndersonCooper).

- Bloggers: individuals who post regularly to an established blog, and who appear to identify as a blogger on Twitter (e.g., @gr33ndata).

- Activists: individuals who self-identify as an activist, who work at an activist organization, or who appear to be tweeting purely about activist topics to capture the attention of others (e.g., @Ghonim).

- Digerati: individuals who have worldwide influence in social media circles and are, thus, widely followed on Twitter (e.g., @TimOReilly).

- Political actors: individuals who are known primarily for their relationship to government (e.g., @Diego_Arria, @JeanMarcAyrault).

- Celebrities ("celebs"): individuals who are famous for reasons unrelated to technology, politics, or activism (e.g., @Alyssa_Milano).

- Researchers: an individual who is affiliated with a university or think-tank and whose expertise seems to be focused on Middle East issues (e.g., @JRICole).

- Bots: accounts that appears to be an automated service tweeting consistent content, usually in extraordinary volumes (e.g., @toptweets).

- Other: accounts that do not clearly fit into any category.

To allow multiple coders to hand-sort the 963 users into one of the above actor types, we built a browser-based coding tool that displayed the stored user profile data. Coders determined each user's actor type by looking at their stored profile data, current Twitter profile and latest tweets, and any websites they linked to in their profile. Coders could also search the Internet for a user's given name or handle to find personal websites, LinkedIn profiles, or bylines on news websites to help determine actor type. The first round of coding involved two different coders classifying each Twitter user. When coders disagreed on a user's categorization, that user went through a second round of classification that required a third coder to choose. Finally, we were left with 42 users (4%) that still had three different actor types (i.e., each of the three coders selected a different classification). These were coded through in-person consensus building. Four of the authors contributed to all rounds of the coding process.

Methodological Limitations

Limitations of our methodology fall into three categories: data representativeness, actor type edge cases, and selection bias.

First, the raw datasets do not contain all relevant Twitter messages: they lack tweets outside of the dates we studied, tweets that did not contain popularly used keywords, and those unavailable through the public Twitter API, including (but not limited to) non-public or "protected" tweets. Furthermore, Twitter and its API have limitations of their own: 1) Tweets contain no curated topic-based metadata, so it is difficult to know whether the search terms used to collect the tweets are, in fact, related to a tweet's content; 2) the API only returns the 1,500 most recent tweets, so when we queried Twitter every 5 minutes, we missed any tweets beyond the 1,500 most recent tweets from the preceding 5 minutes; and 3) in situations where Twitter is being used heavily, the platform's own internal latency results in some tweets simply being missed without any indication by the API.

Second, actor types were generally difficult to classify using the available and highly dynamic information; and in many cases, users seemed to span actor type categories. One example is Jillian C. York (@jilliancyork) who is an active blogger, studies technology use in developing countries, and is a committed activist involved deeply in the MENA region, having previously lived in Morocco. Another example is Xeni Jardin (@xenijardin) who is an active blogger, a freelance journalist, and a member of the digerati. For the purpose of this paper, we decided on a best fit for each of these edge cases while acknowledging that they warrant further study in and of themselves.

Finally, by sampling the largest 10% of flows, we may have introduced a selection bias correlated to individual actors' numbers of followers and/or network centrality. Choosing to cut off the bottom 90% means that our sample only includes flows containing at least 16 posts in Tunisia, or 19 posts in Egypt. Overall, our sample sizes are 850 flows out of a total of 5,024, and 963 coded users out of a total of 30,949 participating in the chosen flows.

Findings

In this section, we analyze the role of different actor types in information dissemination on Twitter within our datasets of flows. We first examine the distribution of coded actor types. Next, we analyze sections of flows, or sub-flows, between distinct actor types, pointing to recurring interactions between actor types. In aggregate, the recurring relationships between actor types shed light on how content spreads—effectively, how information flows on Twitter. We conclude by reproducing a few exemplar information flows to provide context and depth.

Actor Types

As described in the methodology section, we categorized actors from each dataset into 12 distinct types (see Figure 1). (As a reminder, we classified 963 users: 774 were from the Tunisia dataset, 884 were from the Egypt dataset, and 699 appeared in both.) In both datasets, bloggers, journalists, and activists are prominently represented, and at similar frequencies in both datasets.

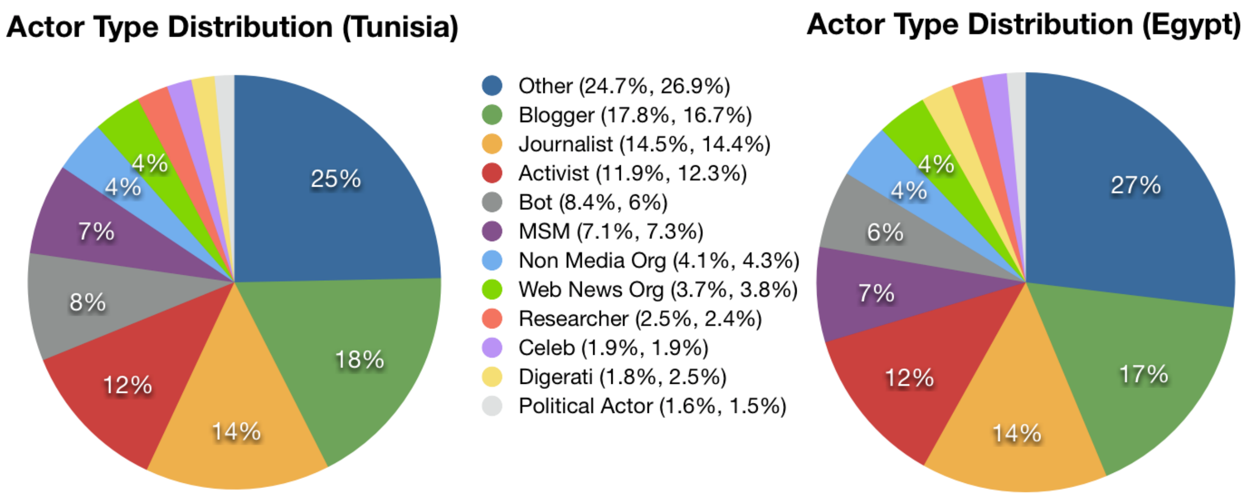

Figure 1. Actor Type Distributions for Tunisia (left) and Egypt (right).

We assumed that an organization's Twitter account plays a different role than an individual account, often serving as the official voice of a group, company, or organization. We define organization accounts as the following: MSM, non-media org, Web news org, and bots (which, in many cases, are controlled by automated programs representing non-individual interests). All other actor types are considered individual accounts. In comparing organization accounts to individual accounts in our datasets (see Figure 2), we found that roughly 70% of the actors in each dataset are individuals.

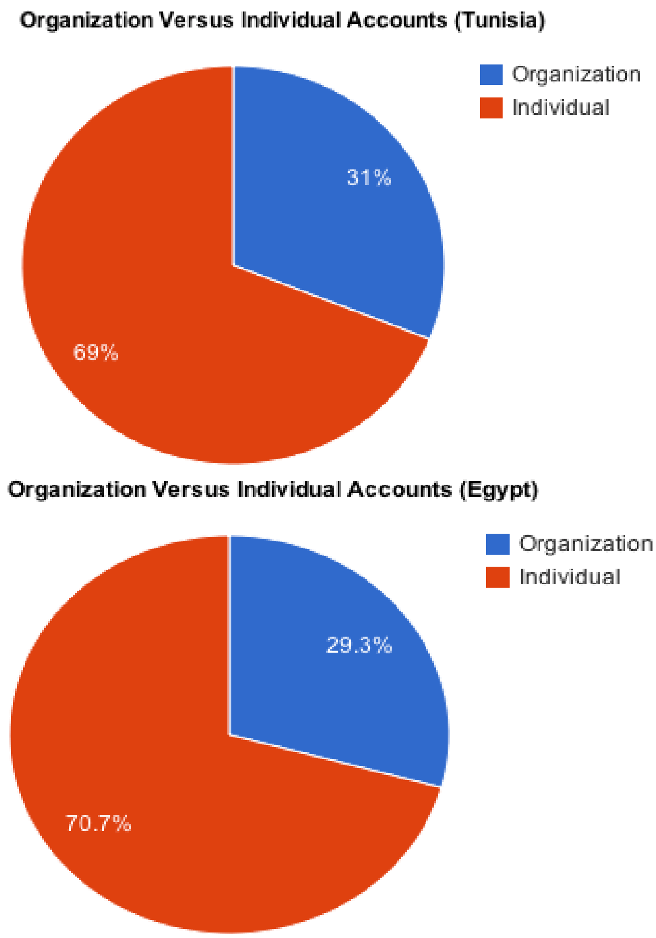

Figure 2. Organization vs. Individual Accounts for Tunisia (top) and Egypt (bottom).

Organization accounts are often managed strategically, and their tweets tend to be more polished and grammatically correct. Their follower counts tend to be higher, and they tend to tweet more frequently (see Table 1). Among individuals, there is also variation in tweeting frequency and follower counts; the actor types that we coded explicitly tend to post more frequently and have more followers than those who are categorized as “other.”

Table 1. Twitter User Behavior: Number of Followers and Level of Activity per Type.

	Median Tweets/Day	Median # of Followers
Organizations	15.98	4004
Individuals (excluding Others)	11.45	2340
Others	9.35	340

To understand further how different actor types behaved, we looked at their tweet to retweet ratio (see Tables 2 and 3). This is an indication of how often different actors' tweets are retweeted by their followers. We take this to be a measure of how well actors engage their audiences. At the low end of this metric are "other" users, who are able to elicit retweets approximately 30% of the time, compared to 88% for MSM accounts. Additionally, Twitter accounts of organizations (MSM, Web news org, and non-media org) have substantially higher retweet rates (i.e., flow sizes) than do individual accounts.

Table 2. Chart of Flow Dynamics by Single Actor Types, as well as by Full Paths: Tunisia Dataset.

Actor Type	Average # Responses per Tweet	Average # Responses per Tweet (for tweets with responses)	Total # Tweets across threads	Total # times tweets had responses across threads	Tweets/Tweet responses	Total number of actors
MSM	14.32	16.15	53	47	0.8868	13
Web News Org	12.06	17.65	79	54	0.6835	14
Non Media Org	12.86	28.30	66	30	0.4545	13
Bot	0.23	2.36	223	22	0.0987	60
Journalist	11.49	22.73	182	92	0.5055	51
Blogger	8.23	21.89	303	114	0.3762	90
Activist	4.20	11.04	271	103	0.3801	58
Digerati	9.21	10.75	14	12	0.8571	6
Political Actor	39.71	55.60	7	5	0.7143	4
Researcher	6.29	15.10	24	10	0.4167	8
Celeb	21.28	63.83	18	6	0.3333	3
Other	6.91	22.75	349	106	0.3037	116

Table 3. Chart of Flow Dynamics by Single Actor Types, as well as by Full Paths: Egypt Dataset.

Actor Type	Average # Responses per Tweet	Average # responses per tweet (for tweets with responses)	Total Tweets across threads	Total times tweets had responses across threads	Tweets/Tweet responses	Total number of actors
MSM	28.75	32.86	64	56	0.8750	15
Web News Org	13.89	21.90	82	52	0.6341	12
Non Media Org	26.82	43.04	90	54	0.6000	16
Bot	1.68	4.06	114	47	0.4123	36
Journalist	12.58	20.86	423	255	0.6028	71
Blogger	3.72	8.54	693	302	0.4358	111
Activist	7.14	17.94	706	281	0.3980	89
Digerati	4.47	5.63	34	27	0.7941	8
Political Actor	21.54	35.00	13	8	0.6154	5
Researcher	8.93	24.00	43	16	0.3721	15
Celeb	14.45	30.42	40	19	0.4750	7
Other	9.05	27.58	789	259	0.3283	157

To understand the impact of actor types on the information flows, we look at two important attributes: *source* and *size*. An information flow's *source* refers to the user who first posted the content. If we look at the distribution of information flows across source types, the differences in dynamics between the Tunisia and Egypt datasets are prominent (see Figure 3).

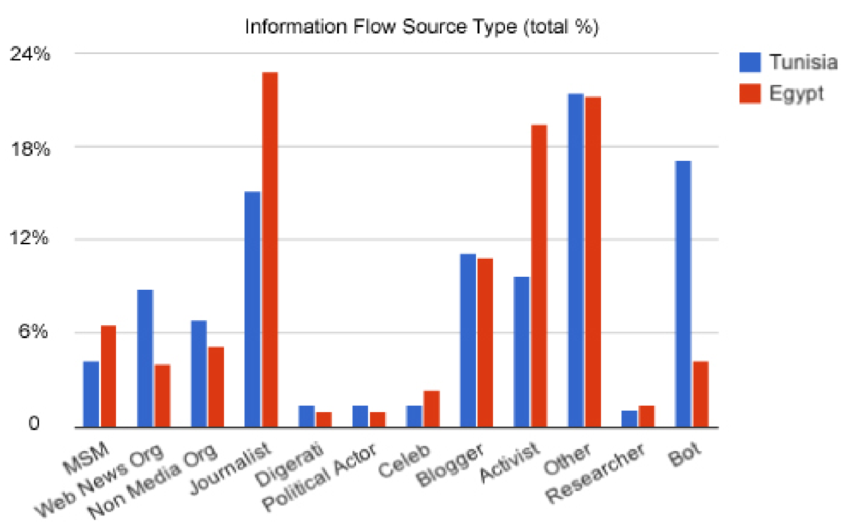

Figure 3. Distribution of Information flows by Source Type for Tunisia and Egypt.

Note: Bars represent the number of threads (as a % of total threads) in each dataset that were seeded by an actor of the given type.

We define an information flow's *size* as the total number of participatory tweets, namely, tweets that are close copies or retweets of the information flow source (see Figure 4).

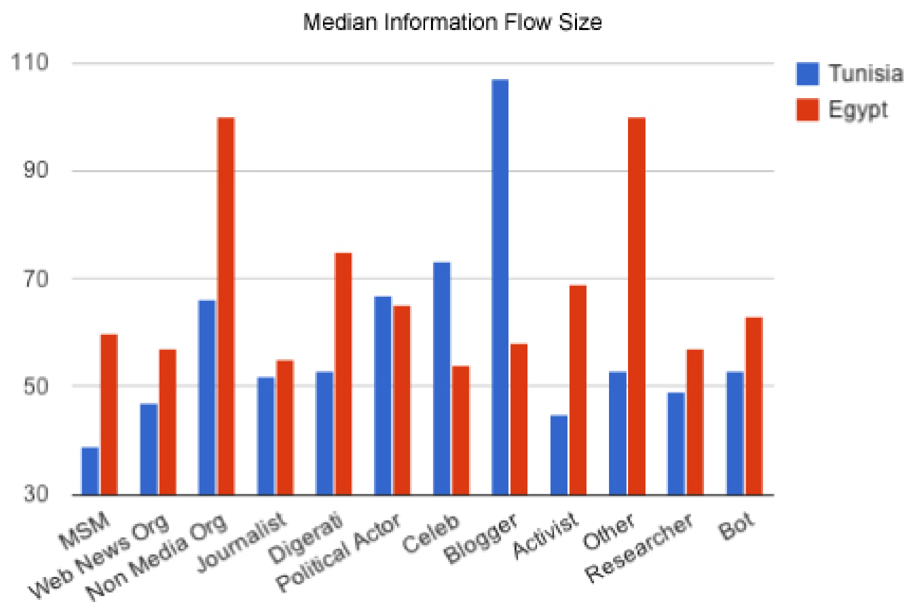

Figure 4: Information Flow Sizes for Tunisia and Egypt.

Note: Bars represent the median number of tweets in threads that were originated by an actor of the given type.

When considering the Tunisia dataset, Figures 3 and 4 suggest that, while more journalists than bloggers served as sources for information flows in Tunisia, those flows started by bloggers were substantially larger in size. This suggests that bloggers played an important role in surfacing and disseminating news from Tunisia, as they had a substantially higher likelihood to engage their audience to participate, compared with any other actor type. Additionally, the Tunisia dataset showed less engagement from MSM, journalists, or activists, compared to Egypt.

When looking at the Egypt data, there are very clear distinctions: MSM, journalists, and activists were much more engaged in information flows, serving as the main sources of flows much more than in the Tunisia dataset. Additionally, they drew larger participation from their audience, as measured through flow size. Meanwhile, although non-media orgs account for being the source of 5% of all flows (26 out of 500), they had the largest average size, most notably a flow started by the official WikiLeaks account, which read: *"WikiLeaks did "more 4 Arab democracy than decades of backstage U.S. diplomacy." http://bit.ly/iitGiF #egypt #tunisia."*

Sub-Flows

In order to gain another dimension of understanding of the flow of information on Twitter and the relationship between actor types in our data, we examined what we call sub-flows. Each information flow is made up of multiple sub-flows. A sub-flow between user A and B (A→B) exists if user B retweeted text that user A had previously posted.

By collapsing every sub-flow within all chosen information flows, we see recurring patterns of retweet behavior among actor types. In the ten most common sub-flow paths between coded actors across both datasets, journalists, activists, bloggers, and "other" actor types are the most prominent (see Table 4). This reinforces the claim that, while organizational actors have larger followings on average, individual actors are much more likely to play an active role in information dissemination.

Table 4. Ten most Common Sub-flows for each Dataset (Tunisia left, Egypt right).

Sub-flows (Tunisia)	Count		Sub-flows (Egypt)	Count
Activist → Activist	49		Journalist → Activist	111
Journalist → Other	48		Journalist → Other	109
Journalist → Blogger	41		Journalist → Blogger	102
Activist → Blogger	38		Activist → Other	102
Other → Blogger	37		Activist → Activist	100
Journalist → Activist	34		Other → Other	97
Blogger → Blogger	31		Activist → Blogger	85
Blogger → Other	31		Blogger → Blogger	78
Journalist → Journalist	30		Journalist → Journalist	70
Activist → Journalist	29		Blogger → Activist	69

In both datasets, journalists and activists serve primarily as key information sources, while bloggers and activists are more likely to retweet content and, thus, serve as key information routers. While there are substantially more journalists actively posting and reposting content about Egypt, the general retweet behavior between the two datasets is similar. In both datasets, journalist content tends to be re-posted frequently by bloggers, activists, and other journalists.

Table 5. Breakdown of Sub-flows from Journalists to Other Actor Types in both Tunisia and Egypt (i.e., Who Reposts Content Coming from Journalists?).

Journalist → ...			
Tunisia		Egypt	
Actor Type	Frequency	Actor Type	Frequency
Other	48	Activist	111
Blogger	41	Other	109
Activist	34	Blogger	102
Journalist	30	Journalist	70
Non Media Org	6	Celeb	10
Researcher	4	Web News Org	9
Digerati	2	Non Media Org	7
MSM	1	Researcher	5
Bot	1	Bot	4
Web News Org	1	Digerati	3
Celeb	1	Political Actor	2
		MSM	1

Journalists appear to have a strong preference for retweeting other journalists' content over content from other actor types. Journalists covering Egypt retweeted other journalists at a substantially higher rate than any other actor types (see Table 6), while in Tunisia, journalists also heavily retweeted activists. Blogger content was retweeted substantially less often in the Tunisia dataset than in the Egypt dataset, suggesting an important and distinct role played by bloggers in disseminating information to journalists during the Egyptian demonstrations.

Table 6. Sub-flows to Journalist Actor Type in both Tunisia and Egypt (i.e., Whose Content do Journalists Tend to Retweet?).

… → Journalist			
Tunisia		Egypt	
Actor Type	Frequency	Actor Type	Frequency
Journalist	30	Journalist	70
Activist	29	Activist	29
Blogger	9	Other	28
Other	8	Blogger	22
MSM	6	Non Media Org	9
Web News Org	5	MSM	5
Non Media Org	2	Web News Org	3
Digerati	1	Digerati	3
Celeb	1	Political Actor	3
		Researcher	2
		Bot	1
		Celeb	1

Bloggers' sub-flows show different characteristics than those from journalist accounts. However, just as journalists prefer to retweet other journalists, bloggers tend to retweet other bloggers. Activists are also retweeted quite heavily by bloggers (see Table 7).

Table 7. Who Bloggers Retweet (left) and Who Retweets Bloggers (right) Across both Tunisia and Egypt Datasets.

Blogger → …				… → Blogger			
Tunisia		Egypt		Tunisia		Egypt	
Actor Type	Frequency	Actor Type	Frequency	Actor Type	Frequency	Actor Type	Frequency
Blogger	31	Blogger	78	Journalist	41	Journalist	102
Other	31	Activist	69	Activist	38	Activist	85
Activist	22	Other	55	Other	37	Blogger	78
Journalist	9	Journalist	22	Blogger	31	Other	58
Researcher	3	Researcher	3	Web News Org	9	Non Media Org	15
Non Media Org	2	Non Media Org	3	Researcher	8	MSM	10
Web News Org	2	Web News Org	2	Non Media Org	5	Web News Org	7
Celeb	1	Digerati	2	MSM	5	Researcher	6
Digerati	1	Bot	1	Digerati	3	Celeb	5
		Celeb	1	Political Actor	1	Bot	4
						Political Actor	3

By looking at retweet behavior among actor types, we are able to identify a significant difference between individual and group accounts. We also see clear preferences for certain actor types to retweet content from the same actor type.

Example Flows

In order to better situate the sub-flow paths, we now consider some exemplar flows that provide depth to the patterns we found in the data.

1. Journalist→…

On January 25, 2011, @adamakary (Adam Makary), an Al Jazeera producer, writes: *"Police guard in tahrir tells me, I'm just following orders, doing my job. Otherwise, I'd be with the protesters #jan25 #egypt.*" Within a minute, @evanchill (Evan Hill), another Al Jazeera producer, retweets Adam's original post. Within 10 minutes, it is retweeted by @exiledsurfer (activist) and @ashrafkhalil (journalist). After a few hours, it reaches @octavianasr (journalist), and is then retweeted widely again (see spike on right side of graph in Figure 5). This is an example of an information flow that started with a journalist and was heavily picked up by other journalists.

Figure 5. Screenshot of Interactive Visualization Showing Volume of Tweets over Time and Participation of User Types. Link to site: ***<u>http://giladlotan.com/ijoc/viz/thread.php?country=egypt&thread=7109</u>***

2. Journalist→…

On January 15, 2011, @bencnn, a CNN reporter on the ground in Tunisia, posts: *"No one I spoke to in Tunis today mentioned twitter, facebook or wikileaks. It's all about unemployment, corruption, oppression. #Tunisia.*" The flow structure is that of a typical broadcast, where, following Ben's initial post, there's an immediate spike of retweets, which then subsides within several hours. Among the identified actor types who retweeted this content, we found: @LaurenBohn (researcher), @digiphile (digerati), @HalaGorani (journalist), @Dream23fb (activist), and @AnonOC (other). These varied actors pick up the original journalist content and spread it to their own audience. Notably, the connection between journalist (@bencnn) and digerati (@digiphile) is found here; @digiphile's tweet prompted a new wave of retweets.

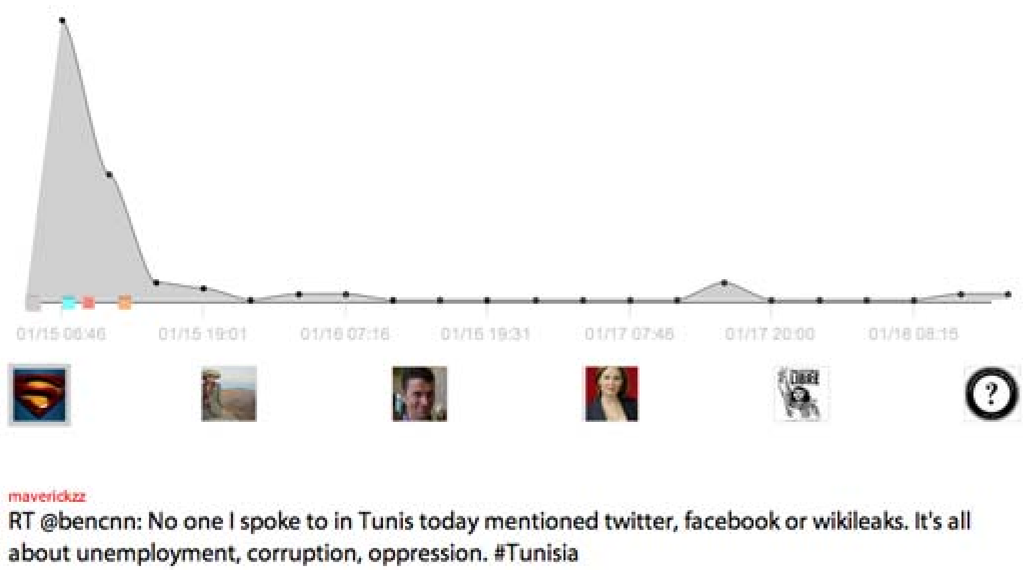

Figure 6. Screenshot of Interactive Visualization Showing Volume of Tweets over Time and Participation of User Types. Link to site:
http://giladlotan.com/ijoc/viz/thread.php?country=tunisia&thread=12358

3. Activist→...

On January 13, 2011, Mauritanian activist @weddady posted: "*I have been an activist for 20 years of my life. What is happening in #Tunisia is unprecedented in Arab World #sidibouzid.*" The flow that follows is short-lived (3 hours and 40 seconds). However, in that period, a variety of bloggers (@s_a_cosgrove, @ByLasKo, @ibnkafka, @Zeinobia) and journalists (@NatashaTynes, @Dima_Khatib) amplify his message.

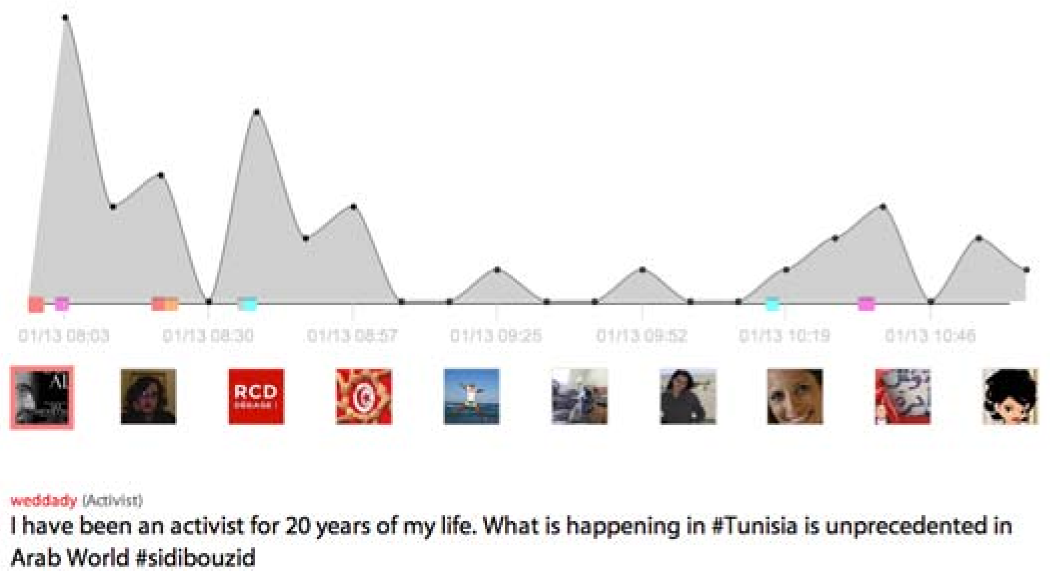

Figure 7: Screenshot of Interactive Visualization Showing Volume of Tweets Over Time and Participation of Coded User Types. Link to site: http://giladlotan.com/ijoc/viz/thread.php?country=tunisia&thread=7875

4. MSM-org→…

On January 15, 2011, @guardiannews (official *Guardian* account) posts *"Tunisia gets third leader in 24 hours http://gu.com/p/2mev2/tf*" with a link to an article from its Tunisia coverage. Within a few hours, a number of journalists (@acarvin, @monaeltahawy, @Brian_Whit, @mfatta7) and activists (@Sonja_jo, @exiledsurfer) repost the article; this generates a number of retweets. This is representative of a typical MSM information flow, where there is little to no interaction with their audience.

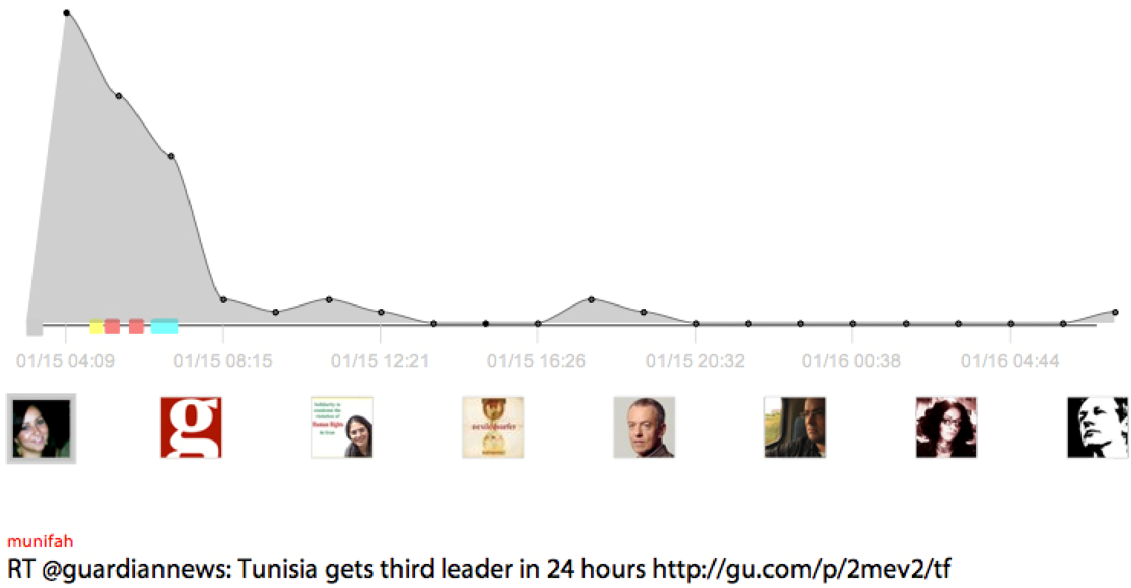

Figure 8. Screenshot of Interactive Visualization Showing Volume of Tweets over Time and Participation of Coded User Types. Link to site: ***http://giladlotan.com/ijoc/viz/thread.php?country=tunisia&thread=9874***

These are just a few examples of the types of information flows present in our data. More examples are available on Lotan's website: http://giladlotan.com/ijoc/viz/

Discussion

Our findings about the distribution of actor types across the Tunisia and Egypt datasets and the trends within information flows and sub-flows give us an indication as to how news might be co-constructed on Twitter among MSM and other actors.

The Distribution of Actor Types

The high degree of overlap between users in the Egypt and Tunisia datasets may suggest that patterns of Twitter usage simply highlight pre-existing relationships among people with similar interests. That is, there is a set of people interested in events like the Egyptian and Tunisian revolutions that Twitter makes visible. Alternatively, Twitter may serve as a convening site wherein people without previously shared interests or existing relationships gather around a particular topic. Twitter is less of a permanent site of conversation among users who know each other, and more of an ad-hoc place where people gather to discover others with similar interests. Additionally, there might have been a learning effect: During the Tunisian uprising, Twitter may have been a place where users honed a set of practices and established relationships that were then further developed during the Egyptian revolution. This learning effect may even stem back as far as the Iranian election of 2009 and the ensuing discussions of “Twitter Revolutions” in various media outlets, possibly priming the Twitter user population to engage in news events like the Tunisian and Egyptian Revolutions. Twitter could be evidence of order effects, in which networks and expertise develop around similar topics that occur in sequence. Or, similarly, these events may suggest special conditions for information cascades over the network.

It is also worth noting how actors are differently able to engage audiences. For instance, MSM accounts in both cases were able to command the highest response rates, measured by their audience’s level of engagement. In both cases, however, journalists—perhaps by virtue of expertise in media dissemination—were able to generate response levels comparable to bloggers, bots, activists, researchers, and others. Digerati were capable of generating the highest response rates—other than MSM—despite their raw number of responses being fairly low. This may imply a certain consistency in being able to command an audience. Whereas many other actors may have a “hit or miss” situation, digerati, perhaps by virtue of a dedicated audience and a personal dedication to the platform itself, are able to routinely generate buzz with their posts.

Additionally, our datasets involved a significant number of users who were difficult to classify—others—suggesting that many influential Twitter users do not easily fit into traditional categorization schemes that attempt to distinguish among actors. That is, it may be that unaffiliated people who do not easily fit into traditional categories of media actors can play a significant role in global news events like the Tunisian and Egyptian Revolutions. Future work focusing on such users, and adopting case study methodologies, might help to explicate these new practices.

Trends in Information Flows

The information flow data graphed in Figure 3 reveal a relatively low number of flows started by organizations versus individuals, most distinctively seen in the Egypt dataset. We see a more balanced distribution across organizations and individuals regarding flow size (Figure 4). Considering that, on average, organizations tend to have more followers than individuals, this finding suggests that influencing audiences to participate on Twitter might be, in part, derived from individual personality, balancing out raw follower count in the flow size data.

If individuals are generally more successful than organizations in seeding prominent information flows, it may be that they are perceived as more trustworthy than organizations—even when they work for organizations, as in the case of some individual journalists. It could also be that there are simply more individual Twitter accounts, giving them an influential advantage over organizational accounts. Or, it could be that, during politically volatile events, individuals are more willing to spread information than organizations. More normatively, it is important to note that this does not mean that individuals are necessarily better at spreading quality information, or that their information is more trustworthy than organizations’ information. Indeed, another interpretation is that individuals are more likely to seed information flows because they share information more liberally than organizations, spreading information before it has been vetted or verified. It could also be that individuals more casually share information of uncertain value because they lack the resources required to evaluate information themselves, assuming (perhaps incorrectly) that their network of Twitter followers will determine a tweet’s veracity.

A different study examining why individuals seem to occupy these positions within information flows could look at the content and motivations of tweets to determine what kind of information individuals were spreading, and whether this information proved to be helpful, trustworthy, or true.

The findings related to an information flow's size—effectively, the number of participants engaged by a particular actor type—suggest that there is, indeed, a difference between how individuals' and organizations' tweets are perceived. In Tunisia, tweets from bloggers had the highest number of retweets, while in Egypt, those from non-media organizations were the most likely to spread.

The differences in information flows between Egypt and Tunisia suggest that Twitter reveals differences in how each country behaves as a media system. Since our study was bounded within the use of Twitter, we cannot make broader claims about the two countries, but we can note that the following things appear to be true on Twitter: mainstream media and individual activist tweets appeared to generate many more responses in Egypt than in Tunisia; non-media orgs appeared to generate many more responses in Egypt than in Tunisia; journalists appeared to have equally large information flows in both countries; and bloggers in Tunisia had greater information spread than those in Egypt. We also cannot say whether these patterns are related to how media systems behave within Egypt and Tunisia—we did not cross-index these patterns with an actor's geographic location—but we can observe that Twitter makes such differences visible. Recalling the high overlap between users in the Egypt and Tunisia datasets, these findings suggest that Twitter itself—without knowing where its users may be located—is a platform on which a similar set of users behaves differently depending on the topic they are using it to discuss.

Trends in Sub-Flows

The sub-flow data give us a more atomic look at the patterns of interaction among actor types, shedding light on the notion that some interactions among actors generate more retweets than others. We clearly see that the most prominent retweet interactions happen between journalists and activists in both the Egypt and Tunisia datasets. Table 4 clearly shows that journalists and activists were the main sources of information on Twitter, engaging their audiences to retweet much more, on average, than other actor types. We can speculate that journalists and activists are similar, in that they are often based in the region at the center of the news event, i.e., within Tunisia, Egypt, or MENA more generally. Within this context, proximity to the event may lend credibility to the source, thus increasing the user's likelihood of being retweeted. The numbers shift slightly from Tunisia to Egypt, in which journalists overtake activists as the top sources. This could reflect the fact that, in Tunisia compared to Egypt, MSM were highly censored, and Western media were generally not very welcome to work ("Reporters without borders," 2011). In both of these cases, we would expect to see a larger role for journalists on the ground in Egypt, and activists and bloggers filling that news void in Tunisia.

Table 4 also indicates that bloggers and activists were the most frequent retweeters in both cases. Each of these actor types may have a more personal agenda for getting the latest news out to their regionally-based followers. In both Egypt and Tunisia, journalists preferred to retweet other journalists, suggesting that Twitter reveals institutional dynamics *within* the mainstream media that may be similar to or different from historical patterns in how news organizations organize work among themselves. And in the case of Tunisia, journalists also often retweeted activists—reinforcing the argument that activists provide on-the-ground news sources during an event that are perceived to be valuable to journalists. These findings offer the possibility of a unique two-step flow phenomenon occurring on Twitter, one in which there may be a "boomerang" effect from on-the-ground reportage to MSM and back to regional sources—an emerging symbiosis between professionals and non-professionals sharing news on Twitter.

Conclusion

Our findings suggest that news on Twitter is being co-constructed by bloggers and activists alongside journalists. This confirms the notion that Twitter supports distributed conversation among participants and that journalism, in this era of social media, has become a conversation (Gillmor, 2004). Specifically, in the context of a major news event like a natural disaster (Sutton, Palen, & Shklovski, 2008) or the Tunisian and Egyptian Revolutions, these conversations involve a host of interested parties. These interested parties fall into roughly three categories:

1. People directly connected to an incident, either as residents or expatriates that want to know about dangerous conditions and the state of their homes and families, or who are experiencing a crisis event firsthand;

2. MSM who want to learn about developments on the ground so that they can provide up-to-date coverage across media channels and hold audience attention; and

3. General interest readers who want to know about events as they happen.

Understanding how news organizations use Twitter can offer insights into the situated and embedded natures of contemporary journalistic practices. That is, MSM's use of Twitter suggests that news emerges not from a single set of stable sources, but from a hybrid and dynamic information network whose structures and influences change depending upon how a variety of actors behave. As demonstrated in this analysis, these actors, working together, can constitute a particular kind of online press.

For news organizations, our research raises questions about how they should use Twitter, understanding how their reporting may be disseminated through both formal organizational channels and the quasi-official accounts of staff. For example, it may be more effective to let journalists control their individual Twitter accounts and build audiences through them, than to disseminate information through official accounts with organizational identities. Most broadly, our observations raise questions about the meaning of objectivity in contemporary journalism. If, historically, objectivity has represented an ideal that a story or piece of information stands on its own regardless of the reporter, our data suggest that, within these Twitter networks, individual journalists were sometimes more effective disseminators of information than organizations. Of course, such a finding should also be viewed critically. Who controls the Twitter accounts of individual journalists? Are such accounts, in fact, simply differently branded organizational accounts with little connection to a particular journalist? Are they strategic instruments used by news organizations to convey an impression of personalization? And if such accounts often attempt to link readers and retweeters back to organizational news sites, are they simply tools for driving traffic, as opposed to means of providing the kind of individual interpretation that has long existed in journalism, but is rarely openly acknowledged?

More work is needed to better understand how information flows among sources. How does information cross linguistic barriers? What are the relationships between regional and global actors? To what degree are journalists or news agencies consuming tweets and incorporating that knowledge into articles without retweeting the messages? Which tweets are actually read by followers, or seen as most valuable? How are different actors viewed in terms of their trustworthiness and accountability?

While there is plenty of future work to do, this article highlights that information is, indeed, flowing among different actor types during events like the Tunisian and Egyptian uprisings, and that the revolutions were, indeed, tweeted.

References

Alexander, A. (2010, April 4). Online readers need a chance to comment, but not to abuse. *The Washington Post*. Retrieved from http://www.washingtonpost.com/wp-dyn/content/article/2010/04/02/AR2010040202324.html

Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A theory of fads, fashion, custom, and cultural change in informational cascades. *Journal of Political Economy, 100*(5), 992–1026.

Boczkowski, P. (2010). *News at work: Imitation in an age of information abundance*. Chicago, IL: University of Chicago Press.

boyd, d., Golder, S., & Lotan, G. (2010). Tweet tweet retweet: Conversational aspects of retweeting on Twitter. *Proceedings of HICSS-42*.

Braun, J. A., & Gillespie, T. (2010). Hosting the public discourse: News organizations, digital intermediaries, and the politics of making newsmedia social. Paper presented at the 11th International Symposium on Online Journalism, Austin, TX, April 1–2, 2010.

Breed, W. (1955). Social control in the newsroom: A functional analysis. *Social Forces, 33*, 326–355.

Carlson, M. (2007). Blogs and journalistic authority: The role of blogs in U.S. election day 2004 coverage. *Journalism Studies, 8*(2), 264–279.

Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, K. P. (2010). Measuring user influence in Twitter: The million follower fallacy. *Proceedings of ICWSM'10*.

Chomsky, D. (2006). "An interested reader": Measuring ownership control at *The New York Times*. *Critical Studies in Mass Communication, 23*(1), 1–18.

Darnton, R. (1975). Writing news and telling stories. *Daedalus, 104*(2), 175–194.

Gaffney, D. (2010). #iranElection: Quantifying online activism. *Proceedings of WebSci10*.

Gans, H. (1979). *Deciding what's news*. New York: Vintage.

Gillmor, D. (2004). *We the media*. Sebastapol, CA: O'Reilly Media.

Gladwell, M. (2010, October 4). Small change: Why the revolution will not be tweeted. *New Yorker*, 42–49.

Hermida, A., & Thurman, N. (2008). A clash of cultures: The integration of user-generated content within professional journalistic frameworks at British newspaper websites. *Journalism Practice, 2*(3), 343–356.

Hindman, M. (2008). *The myth of digital democracy*. Princeton, NJ: Princeton University Press.

Hughes, A. L., & Palen, L. (2009). Twitter adoption and use in mass convergence and emergency events. *Proceedings of ISCRAM.*

Katz, E., & Lazarsfeld, P. F. (1955). *Personal Influence*. Glencoe, IL: Free Press.

Kelly, J. (2010). Parsing the online ecosystem: Journalism, media, and the blogosphere. In G. Einav (Ed.), *Transitioned media: A turning point into the digital realm* (pp. 93–108). New York: Springer.

Kirkpatrick, D. D., & Sanger, D. E. (2011, February 13). A Tunisian-Egyptian link that shook Arab history. *The New York Times*, A1.

Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? *Proceedings of WWW'10.*

Leskovec, J., Backstrom, L., & Kleinberg, J. (2009). Meme-tracking and the dynamics of the news cycle. *Proceedings of KDD '09*.

Lowrey, W. (2006). Mapping the journalism-blogging relationship. *Journalism, 7*(4), 477–500.

Lotan, G. (2009). Retweet revolution methodology. Retrieved from http://giladlotan.org/viz/iranelection/methodology.html

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval.* Cambridge, UK: Cambridge University Press.

Marlow, C. (2005). *The structural determinants of media contagion.* (Unpublished PhD thesis). MIT Media Lab, United States. Retrieved from http://cameronmarlow.com/papers/phd-thesis

Marwick, A., & boyd, d. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media and Society*, *13*, 96–113.

Mohyeldin, A. (2011, January 20). Suicide sparked Tunisia revolution. *Al Jazeera English*, Opinion. Retrieved from http://english.aljazeera.net/indepth/opinion/2010/12/20101231161958792947.html

Molotch, H., & Lester, M. (1974). News as purposive behavior: On the strategic use of routine events, accidents, and scandals. *American Sociological Review, 39*(1), 101–112.

Morozov, E. (2010). *The net delusion: The dark side of Internet freedom*. New York: PublicAffairs.

Nelson, A. (2011, February 24). The limits of the "Twitter revolution." *The Guardian.* Retrieved from http://www.guardian.co.uk/commentisfree/cifamerica/2011/feb/24/digital-media-egypt

Overholser, G. (2009, Fall). What is journalism's place in social media? *Nieman Reports*. Retrieved from http://www.nieman.harvard.edu/reportsitem.aspx?id=101882

Penner, C. (2011, March 14). #numbers. *Twitter Blog*. Retrieved from http://blog.twitter.com/2011/03/numbers.html

Rao, L. (2010, October 31). Twitter added 30 million users in the past two months. *Tech Crunch.* Retrieved from http://techcrunch.com/2010/10/31/twitter-users

Reporters without borders in Tunisia: A new freedom that needs protecting. (2011, February 10). *Reporters Without Borders.* Retrieved from http://en.rsf.org/tunisie-reporters-without-borders-in-10-02-2011,39519.html

Romero, D., Meeder, B., & Kleinberg, J. (2011). Differences in the mechanics of information diffusion across topics: Idioms, political hashtags, and complex contagion on Twitter. *Proceedings of WWW'11.*

Sadiki, L. (2010, December 27). Tunisia: The battle of Sidi Bouzid. *Al Jazeera English*, Opinion. Retrieved from http://english.aljazeera.net/indepth/opinion/2010/12/20101227142811755739.html

Security forces to deal strictly with "illegal" 25 January protest. (2011, January 23). *Al Masry Al Youm.* Retrieved from http://www.almasryalyoum.com/en/node/303405

Shamma, D. A., Kennedy, L., & Churchill, E. F. (2009). Tweet the debates: Understanding community annotation of uncollected sources. *Proceedings of ACM Multimedia.*

Shirky, C. (2009, December 11). The net advantage. *Prospect Magazine, 165*. Retrieved from http://www.prospectmagazine.co.uk/2009/12/the-net-advantage

Singer, J. B. (2007). Contested autonomy: Professional and popular claims on journalistic norms. *Journalism Studies, 8*(1), 79–95.

Singer, J. B. (2010). Quality control: Perceived effects of user-generated content on newsroom norms, values and routines. *Journalism Practice, 4*(2), 127–142.

Sutton, J., Palen, L., & Shklovski, I. (2008). Back-channels on the front lines: Emerging uses of social media in the 2007 Southern California wildfires. *Proceedings of ISCRAM.*

Taylor, C. (2011, February 24). Why not call it a Facebook revolution? *CNN Tech.* Retrieved from http://articles.cnn.com/2011-02-24/tech/facebook.revolution_1_facebook-wael-ghonim-social-media

Timeline: Egypt's revolution. (2011, February 14). *Al Jazeera English*, Middle East. Retrieved from http://english.aljazeera.net/news/middleeast/2011/01/201112515334871490.html

Timeline: Tunisia's uprising. (2011, January 23). *Al Jazeera English*, Africa. Retrieved from http://english.aljazeera.net/indepth/spotlight/tunisia/2011/01/201114142223827361.html

Tuchman, G. (1972). Objectivity as strategic ritual: An examination of newsmen's notions of objectivity. *American Journal of Sociology, 77*, 660–679.

Tunkelang, D. (2009, January 13). A twitter analog to PageRank. *The Noisy Channel*. Retrieved from http://thenoisychannel.com/2009/01/13/a-twitter-analog-to-pagerank/

Vujnovic, M., Singer, J. B., Paulussen, S., Heinonen, A., Reich, Z., Quandt, T. et al. (2010). Exploring the political-economic factors of participatory journalism. *Journalism Practice, 4*(3), 285–296.

Wahl-Jorgensen, K. (2002). The construction of the public in letters to the editor: Deliberative democracy and the idiom of insanity. *Journalism, 3*(2), 183–204.

Wallsten, K. (2007). Agenda setting and the blogosphere: An analysis of the relationship between mainstream media and political blogs. *Review of Policy Research, 24*(6), 567–587.

Wu, S., Hofman, J. M., Mason, W. A., & Watts, D. J. (2011). Who says what to whom on Twitter. *Proceedings of WWW'11*.

Yardi, S., & boyd, d. (2010). Tweeting from the town square: Measuring geographic local networks. *Proceedings of ICWSM'10*.

Zuckerman, E. (2011, January 14). The first Twitter revolution? *Foreign Policy*. Retrieved from http://www.foreignpolicy.com/articles/2011/01/14/the_first_twitter_revolution

PUBLICATION CONTEXT AND CONFLICT OF INTEREST STATEMENT

The data for this paper was first captured when Lotan, a Middle East author for Global Voices, began personally obsessing over the events unfolding in the Middle East. At the time, he was an employee of Microsoft and his interests in the Middle East were unrelated to his job. He captured the Twitter data for personal use using white-listed Twitter accounts and a series of scripts that he had developed to collaborate with boyd. The idea to analyze the network of relevant actors emerged during a Web Ecology Project weekend where a group of self-identified geeks came together to ask questions of Internet culture, analyze data, and otherwise engage with Web 2.0 phenomenon, even though most of them have jobs that are unrelated to the industry. Lotan offered his dataset to the group and this started the project that resulted in the IJoC paper.

At the time of writing and analysis, Lotan, Ananny, and boyd all technically had access to the Twitter data "firehose" through their employment with Microsoft, but this was not used both because of the limitations on how that data can be used and also because this project began as a personal project for Lotan. Ananny and boyd were both affiliated with Harvard University's Berkman Center for Internet and Society during the period in which this project was created and much of the inspiration for this project was rooted in conversations at Berkman and on Global Voices. Both the Berkman Center and Global Voices (which began as a Berkman Center project) are known for engaging on issues related to news, politics, and civil society. Numerous projects related to the Arab Spring have emerged from people who have connections to the Berkman Center.

This article was submitted for consideration by IJoC after Lotan had left Microsoft and before Lotan had joined SocialFlow, where he is currently the Vice President of Research and Development. SocialFlow is a for-profit company that analyzes Twitter data for corporate clients. While Lotan has access to the "firehose" through SocialFlow, he continues to also collaborate with academic researchers using data he collects with his personal scripts. The analysis he did for this paper predates his employment at SocialFlow; he joined during the period between submission and publication and, thus, used his new affiliation on this paper. Likewise, Microsoft uses Twitter data as part of their business. Yet, as part of Microsoft Research, boyd and Ananny have little insight into the corporate relationship between Microsoft and Twitter. Both SocialFlow and Microsoft learned of this paper after it was submitted to IJoC.

There is little doubt that all six authors are shaped, in part, by their institutions and affiliations (including both professional and personal connections), and that multiple, cross-sector identities can raise questions about the nature of any collaborative work. Many scholars—both in academia and industry—are tied to funding processes that bias their outlook and must grapple with the ethical issues they face when they get access to corporate data. That said, this paper was not written for, reviewed or approved by, nor is it necessarily beneficial for, any corporation or organization with which any author is or has been affiliated. All authors are committed to engaging openly and professionally with all critiques of this and any other collaborative work. We believe that analyzing large corporate datasets presents new ethical challenges and that the economic structure of public scholarship is both complex and intertwined with corporate funding. We look forward to the rise of public conversations about what it means to be a responsible scholar in this networked age.