# Chatter on *The Red*: What Hazards Threat Reveals about the Social Life of Microblogged Information

**Kate Starbird, Leysia Palen, Amanda L. Hughes & Sarah Vieweg**
connectivIT Lab
Computer Science & ATLAS Institute
University of Colorado at Boulder
*{Catharine.Starbird, Leysia.Palen, Amanda.Hughes, Sarah.Vieweg}@colorado.edu*

**ABSTRACT**

This paper considers a subset of the computer-mediated communication (CMC) that took place during the flooding of the Red River Valley in the US and Canada in March and April 2009. Focusing on the use of Twitter, a microblogging service, we identified mechanisms of information production, distribution, and organization. The Red River event resulted in a rapid generation of Twitter communications by numerous sources using a variety of communications forms, including autobiographical and mainstream media reporting, among other types. We examine the social life of microblogged information, identifying generative, synthetic, derivative and innovative properties that sustain the broader system of interaction. The landscape of Twitter is such that the production of new information is supported through derivative activities of directing, relaying, synthesizing, and redistributing, and is additionally complemented by socio-technical innovation. These activities comprise self-organization of information.

**Author Keywords**

Computer-mediated communication, crisis informatics, disaster, emergency, risk communication, microblogging

**ACM Classification Keywords**

H.5.3 Groups & Organization Interfaces—collaborative computing, computer-supported cooperative work, organizational design, K.4.1 Public Policy Issues, K.4.2 Social Issues

**General Terms**

Human Factors

## INTRODUCTION: A FLOOD OF DATA

Microblogging is one of the most recent incarnations of computer-mediated chat. Chat applications have been available since the dawn of the Internet, and have provided casual, rapid and synchronous means for communication. As chat applications have migrated to multiple platforms and morphed to include different speaker-audience relationships (one-to-one; one-to-many; many-to-many; known-to-known; known-to-unknown; unknown-to-unknown), they continue to figure centrally in our evolving computer-mediated interactions. As more people adopt and maintain a digital presence, these ever-advancing forms of chat-based environments draw attention not only because of the synchronous and lightweight interactions they support, but also for the new information relationships they produce and the manner in which the media is adapted to suit technological constraints and social conditions [7, 15].

This paper reports on a study of the use of Twitter, a popular microblogging service, during the 2009 seasonal flood threat period to the Red River Valley, a valley whose river separates the US states of North Dakota and Minnesota in a region that extends across the US-Canadian border into the province of Manitoba. The research enumerates and describes the nature of computer-mediated communication (CMC) chatting around a significant, safety-critical event that affects a broad, populated region. Goals of the research are two-fold: First, we aim to theorize about CMC-based chat in the new age of "social media" or Web 2.0 applications and services. Such applications and services receive a great deal of popular attention and therefore are given incomplete and often zealous descriptions of their role and potency. The second aim of the research is to consider and describe features of the relationship that chat has to mass emergency events so that we might more accurately predict its potential in a future where the hope is that information and communication technology (ICT) can mitigate damage incurred by hazards.

### CMC in the Age of "Social Media"

Computer mediated communication is receiving new attention with the progression of cross-platform applications and services collectively called Web 2.0. Just as with previous big end-user advancements, Web 2.0 and "social media" have produced a huge spike of interest and technology adoption. For example, Forrester Research reports that in 2008, 75% of the US adults online used "social tools" compared to 56% in 2007 [1]. Social media


*CSCW 2010*, February 6–10, 2010, Savannah, Georgia, USA.

applications and services include social networking sites (eg., Facebook, MySpace, Friendster, Linkedin, Orkut, BlackPlanet), map and other data mashup services (eg., Google Maps, AlertMap, FlickrVision, Unfluence*)*, and microblogging services and applications, among others.

Microblogging is a form of lightweight chat that allows users to send short messages to people subscribed to their streams. Microblogging services include Jaiku, Plurk, me2DAY, as well as the popular Twitter and several others. We scoped this investigation specifically to the use of Twitter.

Twitter allows its users to send short messages (140 characters or less) to others. These messages— tweets—can be sent and retrieved through a variety of means and front-end clients, including text messaging, e-mail, the web, and other third-party applications, which are enabled through Twitter's public API. Over time some aspects of Twitter behavior have normalized, and even incorporated into the feature set of some end-user client interfaces (an observable instance of Orlikowski's technology and structuration conceptualizations [17]). A notable instance is the use of the '@' symbol, followed by a username (ie. @johndoe) in the text of a message to direct the message to a specific user (even though the message is still public for others to see).

Each Twitter user has a profile, designated as private or public. Private profiles and the tweets sent from these accounts can only be viewed by those who have permission. Twitterers can choose to "follow" other Twitterers, which means subscribing to their tweet streams. Consequently, Twitterers have both "followers" who read them, and those they are "following" themselves. Though this follower-following configuration sets up a form of narrowcasting, all publicly available tweets are also sent to a much broader public stream and remain searchable and accessible by anyone until space caps out on Twitter servers. As one data point, about 70% of tweets sent in the August-September 2008 time frame were public [9].

### Collective Behavior: Mutual Reinforcement Between Threat Conditions & CMC

The implications of social media are significant for mass emergency events. The reasons for this go beyond sometimes popular presumptions that all social media interaction leads to bigger and better forms of information. Rather, social media has made CMC so popular to the point that ubiquity seems inevitable; a realistic understanding of what role social media could or should play in human coordination, especially in cases of emergency is critical to design, practice and policy.

In cases of mass emergency, particular socio-behaviors known as collective behavior phenomena are apparent [6]. These include intensified information search, social convergence in physical space, and information contagion. Collective behavior ideas are powerful in the space of widescale CMC-based interaction during emergency events because the activities of the distributed, decentralized, digital world and those on physical display during disaster events are mutually reinforcing. In other words, the tendency to search for and provide information during a mass emergency event complements the immediacy and breadth of CMC, particularly with today's social media capabilities [18].

The information produced under such pressing and impoverished conditions, however, is heterogeneous and scattered. It is differentially helpful, depending on timeliness and actor relation to the event. Information that was once accurate might later become inaccurate as time goes on; spatio-temporal context for accuracy matters significantly. We do not yet know how much is deliberately harmful, though the presumption that much of it is misleading (deliberately or otherwise) is incorrect [3, 13].

The recent flooding events in the Red River Valley provided conditions for examining closely just what microblogging-based interaction might mean in a disaster event. This hazard possesses seasonal, latent and extended threat: residents are on alert for a long period of time every spring. People in the region have accumulated knowledge about the signs, dangers and mitigation of floods. Furthermore, the several townships along the river have a relationship to each other, as there is some correlation (though not always direct) between upstream and downstream conditions and dangers. And, unfortunately, in this 2009 event, damage was extensive in some areas.

In 1997, the region also experienced devastating floods, which have been the subject of much sociological research. Wachtendorf [28] examined how Canadian and American organizations responded to the disaster transnationally. Buckland and Rahman [4] conducted a study of how the 1997 floods affected three communities in rural Manitoba, finding that differences in "physical, human and social capital" explained the degree of community-level resilience in response. Burn [5] reported on flood risk perception, finding that prior flood experience influenced future flood response.

## BACKGROUND: RED RIVER VALLEY FLOODING

### Geography

For nearly 3000 years, the Red River Valley sat at the bottom of the enormous Lake Agassiz, before it drained about 9200 years ago [21]. That lake carved out the fertile but shallow valley that exists today (see Fig. 1). The Red River flows from just south of Fargo to the north along the North-Dakota (ND)-Minnesota (MN) state border in the US, and into Lake Winnipeg, just north of the city of Winnipeg in Manitoba, Canada. These topographical features make it prone to flooding. With its northerly directional flow, rising waters from southern run-off can pool behind still-frozen northern channels. Because the valley and river channel are flat, rising waters have nowhere to run off but outward onto the flood plain [21].


**Figure 1. Red River Drainage Map:** The Red flows from south to north in a shallow plain. (Credit: Natural Resources Canada)

The valley is therefore plagued by spring flooding, affecting outlying farm areas as well as riverside towns and cities. Larger townships routinely inundated include Fargo/Moorhead; Grand Forks/East Grand Forks; and Winnipeg. Recent decades have experienced significant flooding every five to ten years. The flood of 1997 was the worst modern day flood for the region, with floodwaters in Fargo reaching a height (or crest) of 39.6 feet—21.6 feet above the flood stage of 18 feet.

The *crest* of a flood is the highest level that the water reaches before receding. It is measured in a variety of ways, including maximum height from the base of the riverbed and, during a flood, height above flood stage. Timing and relative crest heights vary across different cities along the river. Typically, cresting will move with the flow of the river (south-to-north), so Winnipeg's seasonal crest will occur weeks later than Fargo's. Other factors, like temperature changes, precipitation and ice jams, lead to variance in crest height and timing along the river.

Over time, flood prevention efforts have helped to mitigate damage. Fargo and Grand Forks have both raised their dikes or levees above their 1997 flood levels. In response to the 1950 flood and then after 1997, Winnipeg constructed and continues to expand its Red River Floodway, which channels excess water during a flood around the city.

### 2009 Flooding

Residents of the Red River Valley were first warned of potential flooding in late February 2009, when the National Oceanic and Atmospheric Administration's (NOAA) National Weather Service released a flood forecast for a spring crest in Fargo at the mid-30 foot range [8]. The official flood stage at Fargo is 18 feet; without the preventative measures of dikes, the predicted crest of 35 feet would flood significant sections of the city [20]. NOAA would continue to revise its preliminary flood forecast on March 19 and 26, raising the predicted height of the Fargo crest to heights of up to 43 feet [10, 19] .

The Red River crested in Fargo on March 28 (see Fig. 2) at a new all-time record height. Fortunately dikes were high enough to avoid catastrophic damage and a cold front had arrived earlier that week, helping flood waters abate by freezing upstream run-off sources. Though temporarily relieved, residents of Fargo were soon warned by the National Weather Service that a second crest, likely higher than the first, would occur mid to late April [16].

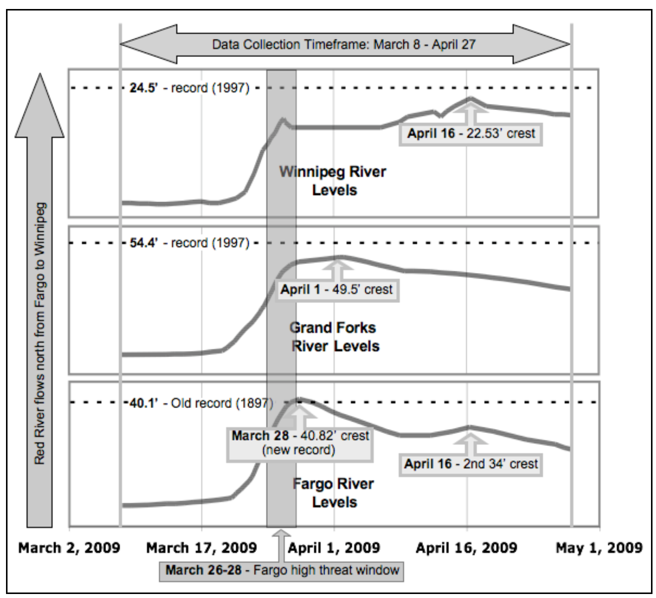


**Figure 2. 2009 Red River Flood Timeline** [24-26]

Meanwhile, downstream towns and cities to the north monitored the conditions in Fargo as well as their own rising waters. In Grand Forks, the river crested on April 1. Ice jams complicated the situation further downstream in the Canadian province of Manitoba. Residents there suffered through a prolonged flood threat accompanied by waves of flash flooding and evacuations in several areas

[22]. In Winnipeg, waters rose and fell and rose again over the course of many weeks. Ice-jams prevented officials from opening the Floodway until April 8. Flooding began at the end of March, but the river did not crest until April 16. The City declared a state of local emergency that morning. Parts of the city and surrounding areas would remain flooded for many weeks.

## METHOD & DATA DESCRIPTION

This research focuses on Twitter communications that took place over a 51-day period surrounding the flooding event. The data collection window began on March 8, 2009, when the Fargo area was operating under predictions of flood, but before threat concerns were raised on March 19. We continued to collect tweets until April 27, when most of the apparent flood danger had passed.

Data collection for this event occurred soon after the onset of the flood threat period. The challenge in studying an emergency event as well as short-lived chat-based information is that one has to make rapid decisions about an emergent event before being sure about what the scope and surrounding communications of the event are. In this section, we outline our data collection and analysis steps as a "virtual ethnographic" method under these conditions.

### Data Collection Steps

Data collection occurred in two parts. In the first phase, we used the Twitter Search API to pull publicly available tweets containing the case-insensitive search terms `red river` and `redriver`. These terms returned relevant data with relatively little noise. Any choice of terms is automatically a constraint so the choices must be carefully made, though there is little opportunity to dwell on the choice because the window on retrievable Twitter data is short (and getting shorter). After investigating the public stream, we settled on these terms as producing a good first cut sample. This initial search activity resulted in 13,153 tweets and 4983 unique tweet authors. In the second part of our data collection, we developed a script to collect the entire Twitter stream for each user in the sample. The result was a data set of 4,592,466 tweets.

The data necessary for the analysis described below exist in two different, but related, data sets: the *Tweet Overview Sample*, a keyword search-generated dataset which allows the examination of a large number of tweet authors ("Twitterers") and their tweets, and the *Local-Individual User Streams*, which provided insight into how (emergency) event-driven tweets are incorporated within the whole of a single Twitterer's tweet stream. We identify these here to support the reporting of findings in the remainder of the paper.

### Qualitative Data Coding

For data analysis, we qualitatively examined and coded individual tweets and user tweet streams to enable information visualization of the entire data corpora.

#### E-Data Viewer

We visualized and qualitatively coded the data sets using the E-Data Viewer, an in-house software application designed for analyzing large CMC-based data sets [23]. The E-Data Viewer allows researchers to see thousands of data points in time, to visualize the interaction of multiple variables, and to quickly test hypotheses (see Figs. 5 & 6 for print versions of these visualizations). Data navigation and visualization tools are combined with features that allow researchers to code the data based on a coding scheme tailored to the investigation at hand.

#### Analytical Iterations and Data Visualizations

Using E-Data Viewer (which we iteratively customized to support these data sets as the analysis evolved), we immersed ourselves in the data, reading through hundreds of user streams and thousands of messages. For each tweet author, we navigated to profile pages to read bios and current update streams. We also traversed links specified within tweets to locate the original source of information for each tweet. The traversal to these out-of-data set sources was built into the E-Data Viewer. The Viewer environment also permitted fluid movement between macro-visualization (of the entire data set or a subsection) and micro-analysis of individual tweets. It enabled us to share insights and develop a common understanding of the two large data sets. Additionally, we could work simultaneously on the data sets (MySQL was the backend).

In a first analytical pass, the keyword-generated data set was trimmed through a process of coding individual tweets and tweet authors as on- or off-topic. A second round of trimming confined the data set to tweets from users with more than three on-topic Red River keyword tweets to allow a focus on the more active Twitterers. This removed the low-level pervasive chatter and made ethnographic investigation of 10,000+ data points more focused and tractable. The resulting data set, the *Tweet Overview Sample*, consists of 358 Twitterers and 7183 tweets.

The *Tweet Overview Sample* was qualitatively coded by multiple researchers working across sections to assure that each tweet was analyzed at least twice. The coding scheme evolved through an iterative and ground-up process that combined insights gained from the first pass with considerable revision and refinement during frequent all-hand data analysis sessions. We coded individual tweets as well as tweet author characteristics. Each tweet within the sample was coded for the apparent original source (or sources) of the information. Information was coded to be *original* to the Twitterer; *secondarily synthesized* by the Twitterer from multiple sources; *re-sourced,* meaning that other on-line sources were reused and passed on; or *retweeted,* where tweets were forwarded wholesale. We also coded for instances of *providing* or *seeking information*. Additionally, tweets that were marked as containing original or synthesized information were additionally coded for other themes and functions that are known in the disaster literature on social convergence [12] or otherwise emerged from this particular data set (i.e. *spiritual support, humor, fear, celebrating, hopeful, educational, exploiting*, etc.).

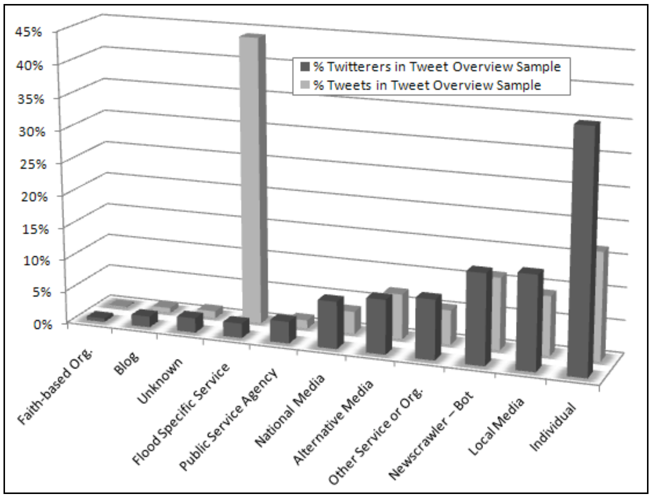


**Figure 3. Distribution of Twitters & Their Produced Tweets by Affiliation** (Note: Graph is descriptive of the Tweet Overview data set and not necessarily representative of all Twitter activity)

For tweet authors, key distinctions that emerged as important descriptors included affiliation, geographical location, and relative distance from the event. Though many Twitterers are private individuals without a stated affiliation (see Fig. 3), others act as representatives for organizations. The distance category—local, peripheral (within 6 hours driving distance), personally connected (has direct personal connection to physical area, but is otherwise remote), and non-local—captures physical distance as well as other connections to the affected geography (see Fig. 4).

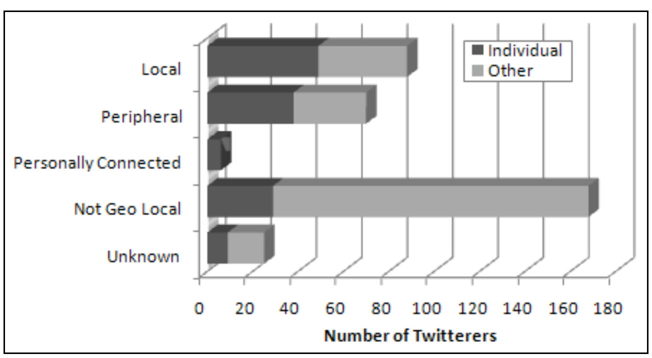


**Figure 4. Number of Twitterers (Individual/Other) by Distance**

From this initial analysis, we could better understand the data set. Unaffiliated individuals, the largest group, comprised over a third (37%) of Twitterers. Though Flood Specific Services were a small portion of the author pool, they were responsible for nearly 44% of all tweets, skewing other analyses towards features of their auto-generated tweets. Removing them, we found that individuals comprise 30% of all tweets. By *distance*, we found that non-locals are the largest group, but locals make up a much larger portion of *individual* Twitterers (Fig. 4).

Distribution of tweets varies over time by both distance and geographical locations along the river. For those outside the affected regions, Red River keyword tweets are focused in a tight window, between March 26 and March 28, leading up to the Fargo crest when predictions were dire. Tweets from U.S. located peripherals are concentrated around that same window, but trail off at a slightly lower rate than tweets from non-locals, while tweets from Fargo locals have a higher intensity over a much broader window. Cities and towns further up the river show local tweet patterns that are less focused than non-locals and locals to Fargo. In all cases, relevant tweets begin to show up during the first major flood predictions and rise in intensity during the Fargo threat window, but do not fall off as rapidly due to later crest times of northern cities.

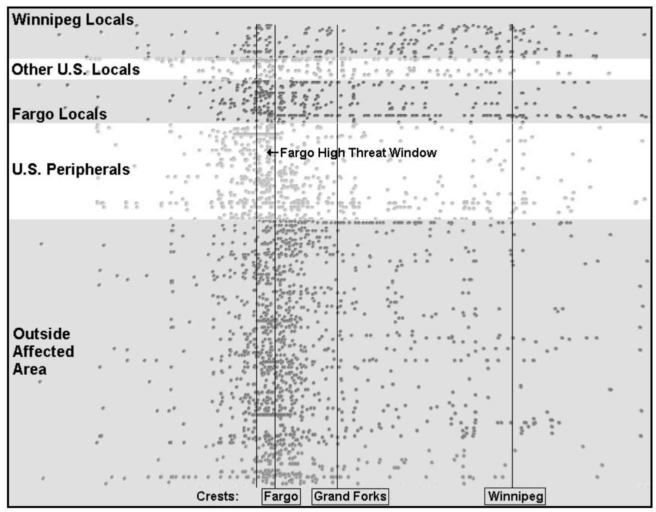


**Figure 5. Temporal Distribution of Overview Tweet Set Sorted by Location from March 8 to April 28**

This descriptive analysis set the stage for additional qualitative analysis of tweet streams for all local-individual Twitterers. Though these users were generating the most original data in our initial sample, we theorize that relevant tweets were still underrepresented. Because individual tweet streams have some level of conversational context and local users can presume context of their followers, these Twitterers would be least likely to incorporate the keywords and hashtags we used to generate our sample. To examine their tweet behavior more closely and better understand generative information production, we compiled all tweets from these authors during the March 8 to April 28 timeframe into a second data set, the *Local-Individual User Streams* data set. This set contains 49 Twitterers and their 19,162 tweets. We coded tweets within this set only as on- or off-topic (14% were coded as on-topic). The remaining qualitative analysis was done at the user stream level.

Finally, we examined other on-line sources to situate Twitter activity relative to other CMC-based opportunities for interaction, but did so without deep content analysis.

## ANALYSIS

Collection and analysis of large data sets generated from CMC during newsworthy events first reveals an utterly unsurprising observation: that publicly available CMC is heterogeneous and unwieldy. However, with extended micro- and macro- examination of those data, we suggest that information production and re-use activities reveal bottom-up ordering activities that support the means of self-organization (in the collective behavior sense) of such information spaces. Our aim in the analysis is to describe features of the "social life" of microblogged information (with apologies to Brown and Duguid [2]).

Complex and even unbounded information spaces begin with *generative information activity*. Generative activity creates the body of material that then requires organization through highly distributed, decentralized and diffuse social cognition processes. The information generated through these narratives becomes part of the public, searchable account of the event. This material then acts as primary source material for downstream *derivative* and *synthetic* processes. *Innovative* activity is another form of synthesis and derivation that reinterprets information and its representation with inclusion of cross-domain expertise and interpretation.

### Generative Information Production

Generative information is at the core of the information production cycle, providing the raw material that later production behavior works to shape into a meaningful informational resource. Generative tweets in our *Tweet Overview Sample* are those coded as *original*. Original source tweets result from two distinct types of Twitter behavior that reflect different orientations towards the information. The first type is autobiographical narrative, which includes first-person observations and status updates. The second type of behavior consists of introducing common knowledge or adapting information from other sources to the discussion space. This latter behavior often takes the form of commentary, as in the example below:[1]

```
@plruark (Mar 28 12:22): Thinking that the Red
River is not cresting, it's more of a temporary
shrinking affect due to the cold weather
```

Original tweets make up less than 10% of our *Tweet Overview Sample*. Locals and peripherals produce over 80% of these. When tweets generated automatically by the Flood Specific Services are removed, local and peripheral authors' tweets are three times as likely to be original.

[1] The @ is a Twitter convention to mark usernames. Individual usernames are anonymized. Usernames of mainstream media sources and public entities remain unchanged.

Though analysis of the *Tweet Overview Sample* helps us frame generative activities within the entire Twitter communication space, analysis of the *Local-Individual User Streams* provides insight into the kinds of information being produced, as well as the ways that individual, local users presented themselves and the information they were conveying. Our tweet-by-tweet analysis of the *Local-Individual Users Streams* indicates that most are broadcasting autobiographical information in narrative form, though many contain elements of commentary and the sharing of higher-level information as well. Even as some Twitterers shift focus to the flood, most continue tweeting within their established Twitter persona.

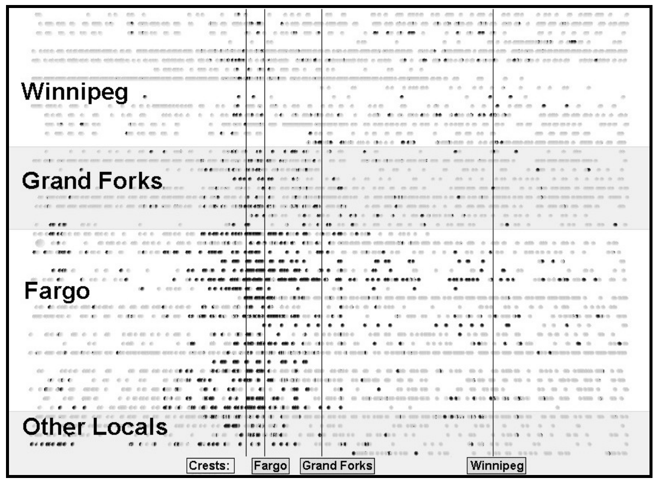


**Figure 6. Local-Individual User Streams. Dark spheres represent on-topic tweets. Light spheres are off-topic.**

When news of flood predictions and warnings appears, local individuals who are already Twitterers begin to tweet more about flood-related issues. For example, in our visualization (Fig. 6) we see spikes in flood-relevant activity among most Fargo users leading up to the first crest on March 28. Some Twitterers who are regular users begin to tweet almost exclusively on flood-related matters during the most critical times, mentioning sandbagging, evacuation information and other related subjects. During the floods, everyday updates are no longer the focus of their Twitter activity. However, once the river level begins to subside, they return to tweeting about their everyday lives. In Fig. 6, these horizontal tweet streams show long, uninterrupted strings of dark spheres representing tweets about the flooding. Authors from Fargo are the most likely local authors to demonstrate intense, on-topic tweet patterns.

One of these authors, @kathy123, had lived through the 1997 flood and decides to use Twitter to narrate her experience during the current flood. On March 19, nine days before the Fargo crest, she begins asking for volunteers to come to the Red River Valley to sandbag. She continues to seek help until March 26, when she decides to evacuate. Three days later, on March 29, she returns home, and her subsequent tweets are celebratory in nature.

Another user, @jordi in Fargo/Moorhead, starts tweeting flood warnings on March 21, the same day he begins to volunteer. He continues to tweet regularly, and almost exclusively about the floods until April 3, when other subjects begin to reappear in his stream. Unlike @kathy123, @jordi chose to stay in a flood-affected city, and documented his experiences sandbagging, helping others and sending hopeful messages to the Twitter audience. In addition to personal updates on flood activities, @jordi broadcasted river levels and official municipal updates, combining generative information production with other types of production activities.

Followers of these Twitterers are able to share in the stories of flood-affected locals through short, direct messages that conveyed the worry, fear, uncertainty and joy of the experience. The information generated through these narratives becomes part of the public, searchable account of the event. This material then acts as primary source material for downstream derivative processes.

### Synthetic Information Production

A second key piece of the information production cycle within our Twitter data is synthesis. Over a quarter of all tweets in the *Tweet Overview Sample* are original tweets created by synthesizing outside knowledge, including other tweets, web and news sources, common knowledge of geographical or historical facts, etc. Synthesis tweets shape the information space by digesting, filtering, relaying, and adapting information to the 140 character maximum format.

Within our *Tweet Overview Sample*, national and local media use synthesis at the highest rates (90% and 71%). Many of these tweets are generative, because their authors are acting (within the role of traditional media) to bring information to the web space for the first time, synthesizing from non-web, non-Twitter sources.

```
@CBSRadioNews (Mar 27 11:50): Coming up at 1pmET: The Red River continues to rise, reaching a record high. The mayor calls for more evacuations and National Guard troops
```

Individuals, blogs, alternative media, and faith-based organizations also use synthesis at above average rates. Much of this synthesis activity functions as an informal, end-user driven filter for the massive flood of information surrounding the emergency event. In these cases, Twitterers process and digest other informational media, then redistribute it on the Twitter platform.

```
@markp (Mar 22 22:20): WDAZ says the predicted crest of the Red River is now 52 feet. Follow @egffloodstage to get hourly updates of the river level.
```

Often, as in the previous example, synthesis activity accompanies other, derivative production behavior, such as directing and/or re-sourcing (defined below).

Synthesis activity is also illustrated in cases when a Twitterer lends expertise to add information value. This didn't happen often but is noteworthy activity that may predict future behavior. For example, in an attempt to contribute through his domain knowledge, a person identified through his username as a pilot says:

```
@i_fly (Apr 6 08:42): The lowest runway at Fargo Hector airport is 896 Ft. above sea level. I wonder how that elevation compares with the Red River at flood.
```

Along a similar vein, 2.5% of tweets and 8% of tweets from individual authors in the *Tweet Overview Sample* are coded as *educational*. These tweets synthesized common knowledge and current information from multiple sources to educate a broader audience.

```
@speakup (Mar 27 06:06): Red River serpasses record highs set in the 1800s, expected to rise more after snowstorm. http://www.fakeURL.com
```

### Derivative Information Production

Our data indicate that Twitter activity cannot be defined completely in terms of generative and synthetic information production. Twitter is not simply a platform for broadcasting information, but one of informational interaction. Activities classified as derivative information production occur in over three-quarters of tweets in our *Overview Tweet Sample* and a smaller, but still substantial portion of our *Local-Individual User Streams*. Though information generated in Twitter and across any number of on-line sources may be valuable, navigation of this unwieldy space is difficult. Many of these conventions have evolved to aid this navigation, directing other users to valuable information, placing virtual signposts within a complex information space. Other behaviors function as filters and informal, user-driven recommendation systems.

Through these activities, Twitterers both self-organize and create the need for more self-organization, as they generate even more noise that gives rise to the need for more directing and focusing behaviors. Derivative information production is therefore a user-driven cycle of shaping and re-shaping a shared interaction and information space.

#### *Retweets*

Retweeting is a user-driven convention that emerged in Twitter communication. In one of our user streams, an experienced Twitterer instructs a new user on its utility:

```
@ozacko: @SuzyQ RT is re-tweeting. Passing along another message. So you go: "RT @ [username] [message]. Bumps it along to more people! =)
```

By allowing Twitterers to pass on information that they deem interesting, important, entertaining, and so on, retweets function as an informal recommendation system within a platform that lacks a formal mechanism. Retweets act to both recommend the information and the original author, and we often see retweeted authors thanking others for passing along their tweets.

Retweets have linguistical features that can be readily measured across a large data set. In the *Tweet Overview Sample,* 5.7% of all tweets and 23% of tweets by local, individuals contain the conventional language for retweeting. Most of these retweets were from tweets that originated in the local media or Flood Specific Services.

When individuals were the original source for the retweets, two-thirds of the time they were local or peripheral. The interpretation of this is that locals and peripheral Twitterers (individuals, media or flood-specific services) are the locus of retweeted information. Additionally, we know from deeper investigation of the *Local-Individual User Streams* that local-individuals retweeted more often about flood-related matters than they do across the rest of their Twitter streams. This supports the assumption that people spread information that they feel or know to be newsworthy through retweeting.

#### *Follow @ Tweets*

Another convention within Twitter is to explicitly recommend authors by telling others to follow them.

```
@markp (Mar 26 10:20): @art2 follow @fargofloodstage for current height. follow @homer cause he is funny.
```

Almost all of the explicit follow recommendations within our *Overview Tweet Set* direct Twitterers to sources that are local or peripheral to the flooding event. Most point to Flood Specific Service accounts that post flood level data at regular intervals. Twitterers use this convention to guide other users to sources they deem trustworthy. Interestingly, the authors who compose "Follow @" tweets are all local or peripheral. These directors are Twitterers who have earned or presumed the credentials to tell other people whom to follow [27].

#### *Re-sourcing*

Another widespread directing behavior we identified in the data is *re-sourcing*, the act of pointing to other sources or copying information from elsewhere into a tweet. Re-sourcing tweets can act to organize the information space by directing user attention to specific information, but their pervasiveness within Twitter also results in more confusion, more clutter, and the need for more organization.

69% of tweets in our *Overview Tweet Sample* were interpretatively coded as incorporating re-source behavior. When Flood Specific Services are removed from the sample, that percentage drops by half (to 34%) and we find non-locals to be much more likely to use their tweets to re-source than locals and peripherals. This is not surprising, as few have first-hand knowledge of the event, so much of their Red River conversation has to rely on information from others. In many cases, these Twitterers use their accounts to pass on information and links to sources with more credentials, mainly local and national media.

One affiliation group, the Newscrawlers-Bots, tweeted re-sourced information only. These are computer-operated accounts that grab information from other CMC sources, filter it in various ways, and redistribute it through their tweet streams. A similar affiliation type was the Flood Specific Services, a group of Twitterers who initiated their accounts during the flooding event and tweeted flood-related information exclusively. Though three of these authors aggregated then re-sourced or synthesized information from multiple sources, the remaining five had a single purpose: to distribute at even intervals automated flood level data from their various locations.

We also discovered automatic-feeds incorporated into accounts of authors with affiliations other than Newscrawler-Bots and Flood Specific Services, including individuals. Services like Friendfeed (a social media aggregator) and Twitterfeed (an automatic blog updater) enable users to pull web and Twitter-based information from other sources and add it to their update streams. We found several users within our data sets using these plug-ins to auto-generate tweets. Filters chosen by these Twitterers allow them to auto-re-source and redistribute information they deem important, an action that again works to both populate as well as shape the overall information space.

#### *URLS*

The presence of URLs within tweets is an indicator of another organizational activity, often tied to re-sourcing: providing a direct link to external, web-based information. These URLs can be used as an economizing strategy to overcome the 140 character limit by pointing to more information created by the original tweet author (e.g., a blog). In most cases within our set, we see these URLs directing readers to information created by another author, a media outlet, or an unaffiliated website.

URLs are present in more than half of the tweets in the *Overview Tweet Sample* (56%) when Flood Specific Services and Newscrawler-Bots are removed. In that sample, locals and individuals use URLs within their tweets at much lower rates than other distances and affiliations. Those Twitterers are more likely to rely on first-hand information and synthesis than external re-sourcing.

### Innovation Through Tweeting

In another notable part of the information production cycle, Twitterers use personal skills and expertise to contribute to the information space through innovation.

Five of the eight Flood Specific Service accounts sent precise flood stage measurements at regular intervals. The regularity of tweet posting intervals and text for these streams indicated that they were auto-generated "bots." Their tweets were often retweeted and re-sourced. Within days of account creation and leading up to the Fargo crest, flood information from their tweets was spreading throughout the broad Twitter network, allowing users to follow water level changes in almost real-time.

```
@egffloodstage (Mar 29, 19:00): Red River at East Grand Forks is 48.70 feet, +20.7 feet of flood stage, -5.65 feet of 1997 crest. #flood09
```

Investigation in our *Local-Individual User Streams* reveals that each flood stage service was created by a single local individual Twitterer. Though the original sources of the water data were official agencies (including the US Geological Survey and NOAA), none of the distributing Twitter accounts were affiliated with a public service agency or media outlet. Each author maintained a separate, personal Twitter account which references the new service he created. Not insignificantly, most of these authors self-identified as a "geek" or a "nerd" in their bios. Evidence within their streams points to transfer of both innovative ideas and the techniques required to produce them.

```
@markp (Mar 23 15:13): Looks like I have started a twitter mini-meme @egffloodstage @fargofloodstage @redinwinnipeg and @redriveratfargo #flood09.
```

In response to this tweet, the author of @fargofloodstage initiated a discussion with the author of @egffloodservice, about the scripts each used to scrape web-available water resource data and populate their bots. The original author even offered to send the copycat a version of his code.

Motivations for creating these streams may have varied, from increasing Twitter status to showing off technical skills, but some authors explicitly indicated that they felt good about being able to help through their expertise.

```
@homer (Mar 23 20:58) : It is really amazing watching the followers grow on @fargofloodstage I feel special being able to make something that helps
```

These instances of innovation follow the sociology of disaster research, which has repeatedly shown that *improvisation* is a feature of self-organizing activity across emergency response [11, 14].

Others saw the floods as a reason to log onto Twitter for the first time, particularly those in the media. @weather_guy and @janecleary (a local news anchor) both started accounts during the floods, and updated them regularly with news of river levels, traffic reports and similar information. In these cases, the floods served as an impetus for both innovation and adoption of Twitter.

Some Twitter users adapted their understanding of how to best use the application during this critical time. @macsmth maintains a personal Twitter account, but when the flood threat became serious, he created a new account, @risingredriver, another *Flood Specific Service* (though not a script-controlled account). He populated the account with flood-related information only, most of it derivative. Throughout the event window, he continued to update his personal account with mostly autobiographical, generative tweets, but saved flood-related tweets for the @risingredriver account.

During the height of the Fargo threat, another local Twitterer attempted to leverage his understanding of Twitter search mechanisms to manipulate public attention.

```
@ozacko (Mar 27 03:05): Fargo is now a hot topic on Twitter Search. Keep tweeting to keep the attention, guys! #fargo #redriver #flood09
```

## CONCLUSION

Twitter, a new incarnation of computer mediated chat, is a platform without formal curation mechanisms for the massive amount of information generated by its (burgeoning) user base. There is no rating or recommendation system support—key features of commerce sites like Amazon and information aggregators like Digg. Nor is there a complex system of validation that, for example, Wikipedia has implemented. Also unlike Wikipedia, content passed through Twitter is short-lived, and therefore cannot be discussed, verified and edited. While most social media have "places" for interaction, interaction in Twitter occurs in and on the data itself, through its distribution, manipulation, and redistribution. Without regular retransmission, communications quickly get lost in the noise and eventually die off.

Adapting to these unique characteristics of interaction, Twitterers have evolved their own curation mechanisms, a form of bottom-up self-organizing. Users determine what is valuable and what is not. Information is part of a lifecycle of generation, derivation, synthesis, and innovation that marries skills with information production to shape the information space.

The conditions of the 2009 Red River Floods provided a window through which to examine Twitter activity over a concentrated period, where stable elements of geography and features of the hazards threat may be connected to Twitter communications. The result is the description and consideration of an information lifecycle offered here.

However, the results of this work also have bearing on practical, societal level matters of emergency management. One of the challenges for emergency management today is to know "what to do" with social media applications. The new digital world provides both an opportunity but also a real and understandable dilemma for emergency management: How can they make sure that the information that is "out there" is accurate during an emergency event?

Though we dwell on the details of these non-trivial concerns in greater depth elsewhere [27], this research suggests a reassuring new framing of the relationship between emergency response and communications by members of the public: That official information remains important and is complemented, not usurped, by information generated by the public. People use and rely on official sources and other believable eyewitness accounts from which to source their information.

In this flooding event, we see this idea manifest in the derivative and innovative information activities concentrated on distributing water level data that was originally published and made available online by government agencies. Though popular literature places high value on eyewitness accounts as provided through social media—and we know from ongoing research that they can indeed be helpful—this finding demonstrates that official and objective data are still actively sought and important. Though the data itself are valued, they are made useful and locally relevant through active manipulations by interactive members of the information space who add context to it, support it, refute it, and, in this case, create new representations of and new distributions for it.

The lesson here is not that agencies need to be everywhere online but rather to know that people have the capacity through social cognition—and individual enterprise motivated by a perceived audience—to re-use data for their local needs. Understanding this relationship between the provision of quality data for the purposes of user-driven redistribution and innovation is where confidence in the release of control of information needs to reside.

**AKNOWLEDGEMENTS**

This research has been supported by the US National Science Foundation through grants IIS-0546315 and IIS 0910586, as well as an NSF Graduate Fellowship awarded to the first author. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

**REFERENCES**

1. Bernoff, J., C.N. Pflaum, & E. Bowen, *The Growth of Social Technology Adoption*. 2008, Forrester Research.
2. Brown, J.S. & P. Duguid. *The Social Life of Information*. Harvard Business School Press, 2000.
3. Bryant, S., A. Forte, & A. Bruckman. Becoming Wikipedian: Transformation of Participation in a Collaborative Online Encyclopedia. In *Proc. GROUP 2005*.
4. Buckland, J. & M. Rahman. Community-based Disaster Management During the 1997 Red River Flood in Canada. *Disasters* 23, 2 (1999), 174-191.
5. Burn, D.H. Percetions of Flood Risk: A Case Study of the Red River Flood of 1997. *Water Resources Research* 35, 11 (1999), 3451-3458.
6. Dynes, R.R. & E.L. Quarantelli. *Group Behavior Under Stress: A Required Convergence of Organizational and Collective Behavior Perspectives*. Columbus, OH. Disaster Research Center at OSU, 1968.
7. Farnham, S. & P. Keyani. Swarm: Hyper Awareness, Micro Coordination, and Smart Convergence through Mobile Group Text Messaging. In *Proc. HICSS 2006*, IEEE.
8. Forecaster: Red River Flood Crest May Top 30 Feet in Fargo. Feb 24, 2009. *USA Today.*
9. Hughes, A.L. & L. Palen. Twitter Adoption and Use in Mass Convergence and Emergency Events. In *Proc. ISCRAM 2009*.
10. Jenkins, N. & D. Kolpack. Red River Could Crest at 43 Feet. Mar 27, 2009. *The Boston Globe.*
11. Kendra, J. & T. Wachtendorf. Creativity in Emergency Response to the World Trade Center Disaster. *Beyond September 11th: An Account of Post-Disaster Research* (2003), 121-146.
12. Kendra, J.M. & T. Wachtendorf. Reconsidering Convergence and Converger Legitimacy in Response to the World Trade Center Disaster. *Research in Social Problems and Public Policy* 11, (2003), 97-122.
13. Kriplean, T., I. Beschastnikh, & D.W. McDonald. Articulations of Wikiwork: Uncovering Valued Work in Wikipedia Through Barnstars. In *Proc. CSCW 2008*, ACM Press (2008), 47-56.
14. Mendonça, D. & W.A. Wallace. Studying Organizationally-situated Improvisation in Response to Extreme Events. *Int'l Journal of Mass Emergencies & Disasters* 22, 2 (2004), 5-29.
15. Nardi, B.A., S. Whittaker, & E. Bradner. Interaction and Outeraction: Instant Messaging in Action. In *Proc. CSCW 2000*, ACM Press, 79-88.
16. NOAA Forecasters: Red River Will Crest Again in Fargo-Moorhead in Late April. Apr 3, 2009. *NOAA.*
17. Orlikowski, W.J. The Duality of Technology: Rethinking the Concept of Technology in Organizations. *Organization Science* 3, 3 (1992), 398-427.
18. Palen, L. & S. Vieweg. The Emergence of Online Widescale Interaction in Unexpected Events: Assistance, Alliance & Retreat. In *Proc. CSCW 2008*, ACM Press, 117-126.
19. Rampton, R. & D. Gregorio. High Risk for Flooding in US Red River Valley: NOAA. Mar 19, 2009. *Reuters.*
20. Red River: Flood Stages Versus Geographic Locations. http://www.ndsu.edu/fargo_geology/floodgeography.htm.
21. Schwert, D.P. A Geologist's Persepective on the Red River of the North: History, Geography, and Planning/Management Issues. In *Proc. Int'l Water Conference 2003.*
22. Skerritt, J. & B. Redekop. Manitoba Coping with Third-worst Flood of Century. Apr 14, 2009. *Winnipeg Free Press.*
23. Starbird, C. E-Data Visualization for Qualitative Research. Interactive Poster Presented at *CSCW 2008*.
24. This Year's Winnipeg River Levels. http://www.winnipeg.ca/waterandwaste/drainageFlooding/riverlevels/thisYear.stm.
25. USGS Water Data for Red River at Fargo, ND. http://waterdata.usgs.gov/nd/nwis/uv/%3fsite_no=?station=05054000.
26. USGS Water Data for Red River at Grand Forks, ND. http://nwis.waterdata.usgs.gov/nwis/uv/?site_no=05082500.
27. Vieweg, S., L. Palen, & K.M. Anderson. Supporting the "Everyday Analyst" in Time- and Safety-Critical Situations. Unpublished Manuscript, University of Colorado at Boulder.
28. Wachtendorf, T. Interaction Between Canadian and American Governmetal and Non-Governmental Organizations During the Red River Flood of 1997. In *Proc. Int'l Joint Commission Ottawa/Washington 2000.*