

Upvoting Hurricane Sandy: Event-Based News Production Processes on a Social News Site

Alex Leavitt

Annenberg School for
Communication & Journalism,
University of Southern California
aleavitt@usc.edu

Joshua Clark

Annenberg School for
Communication & Journalism,
University of Southern California
joshuaac@usc.edu

ABSTRACT

This paper uses the case of Hurricane Sandy and reddit's topical community (subreddit) /r/sandy to examine the production and curation of news content around events on a social news site. Through qualitative analysis, we provide a coded topology of produced content and describe how types of networked gatekeeping impact the framing of a crisis situation. This study also examines, through quantitative modeling, what kind of information becomes negotiated and voted as relevant. We suggest that highly scored content shared in a social news setting focused more on human-interest media and perspective-based citizen journalism than professional news reports. We conclude by discussing how the mechanisms of social news sites conflict with the social norms and culture of reddit to produce differing expectations around news.

Author Keywords

Social news site; news production; reddit; networked gatekeeping; news framing; crisis communication; mixed methods

ACM Classification Keywords

H.5.3. Group and Organization Interfaces; Asynchronous interaction; Web-based interaction.

INTRODUCTION

Individuals use social media for a variety of information and media sharing purposes. Personal reactions to events, sometimes called citizen journalism, have been a crucial emergent behavior on networked platforms, particularly for their role in moments of crisis or disaster (e.g., [31]). Keegan, Gergle, & Contractor [12] argue that “[s]tudies of peer production in sociotechnical systems... often presuppose that membership in and motivations to contribute to these communities are relatively stable and constant, yet such communities often self-organize and coordinate their activities under conditions of unstable and sudden collective action” (3). However, social networking platforms and microblogs (namely Twitter) have dominated

the research of these situations. Another genre of social media, “social news sites,” provides a unique way for participants to curate and consume information through collaborative voting processes. Not only have social news sites been rarely studied from mixed methods perspectives, but reddit and /r/sandy present a unique example of a social news site adapted for crisis communication. As sites like reddit become more popular and mainstream, it is important that we understand how technical and social features of social news site communities impact information flows. This paper examines how one social news platform, reddit.com, collected and organized information in a crisis situation. We explore the production and filtering of this information through the lens of networked gatekeeping, where a mix of average users and elite moderators work together to curate and make visible the most important content. Specifically, we examine one “subreddit” (topical community message board), /r/sandy, that was adopted in late October 2012 to curate information related to Hurricane Sandy. We then ask how a social news site can produce news around event-based topics:

RQ1: *How do technical and social processes work together in a social news site to determine the production, moderation, and curation of information?*

RQ2: *What types of content emerge around an event from a community-driven curation platform, and who contributes to its framing?*

PLACING THE “NEWS” IN SOCIAL NEWS SITES

To explain news production in /r/sandy, we first contextualize this subreddit with other social news sites. Social news sites, like Slashdot, Digg, and reddit, allow users to submit hyperlinks to a feed of content, which other users can vote up or down (dubbed “upvotes” and “downvotes,” respectively). As posts gain more upvotes, this popular content eventually reaches the top of the site’s page based on a total score relative to up- and downvotes. Lerman & Ghosh [17] find that social news sites like Digg are able to disseminate information quickly. Through a combination of the platform infrastructure and underlying social networks, users continue to vote on content that had become popular through earlier voting-based moderation. Additionally, on reddit, when a user posts a hyperlink or comment, the total score it receives is added to the user’s “karma” points, a measure of reputation. Users are also able

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org. CHI 2014, April 26 - May 01 2014, Toronto, ON, Canada. Copyright is held by the owner/author(s). Publication rights licensed to ACM. ACM 978-1-4503-2473-1/14/04 \$15.00. <http://dx.doi.org/10.1145/2556288.2557140>

to submit “self” posts: text posts whose total score is not added to the poster’s total karma.

Rather than a single main page where all users contribute, reddit is composed of subreddits: community message boards curated around a particular topic, like /r/pics for sharing pictures or /r/funny for posting jokes. Reddit as a whole is thereby founded on a network of subcommunities, where users will post across a variety of subreddits, each with their own norms and methods of evaluation. Reddit’s front page displays the user-submitted hyperlinks from twenty “default” subreddits (which every viewer automatically sees), and users can create new subreddits, which other users then subscribe to in order for respective popular content to be interspersed on their personal front pages. It is important to note that content submitted to any subreddit goes through an automated detection process to see if the particular hyperlink already exists; if so, the user is forwarded to the existing page (though users are allowed to cross-post content to different subreddits). While millions of users have registered with reddit, many more merely browse the front page as viewers, resembling the structure of a mass media platform and validating reddit’s slogan as “the front page of the internet.”¹

While news is shared daily on large, general subreddits like /r/news or /r/worldnews, users might try to situate information about particular events by sharing it to other topic-oriented subreddits (for example, /r/olympics becomes primarily active around its related events). Of course, users may improvise ways, outside of a particular topical subreddit, to contribute news on reddit, such as continually updating a single thread (see [6] covering an example during the Aurora, CO shootings).

Still, one of the barriers for new users to participating in reddit’s extended network is discovering the variety of subreddits that exist. Subreddits that are not included in the default set must recruit users through advertising or word of mouth or rely on serendipity; therefore, newer subreddits may take some time to grow a committed collective of contributors.

RELATED WORK

Event-Based News Sharing in Social Media

Most work on social news sites has demonstrated the effects of their unique technical system on the potential for a large collective of users to curate relevant content in a democratic manner. Such democratic processes allow for various public spheres to form [9, 25]. However, little research has examined what kinds of content are shared in social news settings and how those types of content frame the whole of what is contributed.

The occurrence of a news event especially impacts how information sharing plays out [1, 24]. Prior research on uses of social media in crisis settings has shown at a high level that particular types of content dominate what is produced and shared, as well as what is seen as important, about breaking events. For example, Hughes & Palen [10] demonstrated that information broadcasting and brokerage were important facets of content shared during emergency situations and events with popular attendance. As far as the validity of content, Schultz, Utz, & Goritz [29] demonstrated that newspaper reports are deemed as more reputable sources in crisis communication than online media like blogs. As an exception, Chew & Eysenbach [7] showed in detail that on Twitter, in reaction to the spread of the H1N1 virus, news websites and resource-related information were the most shared type of content, and over time content like personal experiences increased while others like humor decreased. Bruns, et al. [4] also found that Twitter conversation, by looking at one coordinated hashtag, #qldfloods, during a flood situation, primarily produced information and media sharing, with fewer reactions, help, and personal experiences. Given the size and importance of news in other platforms, we then suggest that:

H1: *Professional news reports are more likely to be scored higher than other content.*

In addition to what kinds of content are curated around events in social news sites, we expect that more proficient users will know how to produce important and relevant information. For instance, Suh et al. [32] found that Twitter users’ account ages positively affected retweetability. Burke, Marlow, & Lento [5] found that newcomers to Facebook participate less than experienced users, and increased participation on the part of newcomers depends on their experiences becoming acculturated to the platform and learning social norms early on. Similar differences in experience exist on social news sites as well [15]. While Shklovski, Palen, & Sutton [31] note that individuals in crisis situations are prone to adopt new communication technologies in order to share information, we still anticipate more experienced users to be able to produce relevant information given their experience with the community. We therefore expect that:

H2a: *Compared with other content, professional news reports are more likely to be contributed by more-experienced users.*

H2b: *Compared with other content, professional news reports are more likely to be contributed by more-active posting users.*

H2c: *Compared with other content, professional news reports are more likely to be contributed by more-reputed posting users.*

¹ For statistics describing the popularity and widespread use of reddit, see [21].

News Framing & Networked Gatekeeping

Framing focuses on the way that information is presented and comprehended around a specific topic [11]. The kinds of information shared around an event frame it in a particular way as they spread inside and around the social news site. However, various gatekeepers will also affect the framing of that content by limiting its production, visibility, and access. These actors shift traditional news framing from journalism [28] into “networked gatekeeping,” a distributed process where information is vetted and obstructed not by hierarchical editorship but a mix of distributed processes [2, 12, 16, 26]. On a social news site like reddit, the users negotiate information generated by the community through consensus with upvotes and downvotes. However, powerful figures that act as gatekeepers filter content too: frequently, these are elite members of online communities that take up moderation roles within the platform, through either technical or socially-mediated roles. On reddit, volunteer moderators are granted special powers to remove content (or even users) they feel conflicts with the subcommunity’s rules, norms, or goals. Still, there are opportunities for more distributed moderation: for instance, Shaw [30] outlines how relational boundaries between participants in a participatory blog setting lead to a mix of centralized gatekeeping by elite users with decentralized moderation through processes like content filtering, all supported by an egalitarian ethos. In this mixed setting, moderation pushes relevant and factual information higher. Information flows directed through both moderators’ decisions and the voting behaviors of users collectively mix to determine information relevancy and to produce a “networked framing” of the event [24]. On reddit, framing occurs both as an effect of the content contributed by users, what content is allowed to stay (as determined by moderators), and what content becomes visible (as voted upon by users).

Our study contributes to research on framing news events because the platform and purposes of a social news site remain distinct from microblogs and social networking sites, which have dominated the literature. The latter represents a mix of egocentric information sharing with differing audience dynamics (see [14] for the evolution of these dynamics on Facebook; for Twitter, see [22]), while social news sites particularly attract users to broadcast and share to an unknown mass audience. Just as with breaking news stories on Wikipedia [13], for a social news site like reddit, the contributions of submitters are moderated by other participants; therefore, only particular content becomes popular and visible, setting the frame and tone of how the crisis becomes transmitted in this particular space.

METHOD & DATA

In order to analyze the impact of networked gatekeeping on the framing of news in /r/sandy, we combine qualitative observation with statistical analysis of content-coded posts. Through these observations, we identify examples of networked gatekeeping processes at work on reddit. We then use a statistical model to observe how community

voting as a gatekeeping process impacts the types of content that become the most popular.

During the duration of Hurricane Sandy and its aftermath, one author conducted two months of qualitative observation, maintaining an active reddit.com account, viewing live activity in the subreddit, messaging contributors and moderators, and using a combination of field notes with archival search to structure the analysis. Both authors are also active users of reddit, having participated in the broader community for the past two years, which provides additional context to understanding behavioral norms across the site.

We later collected, using Python, MySQL, and the reddit API, a dataset of 2458 threads posted to /r/sandy during the subreddit’s entire lifespan (beginning in late-October 2012). These threads consist of a population sample of all posts contributed to the subreddit up until the collection period (4 February 2013), including the 1000 highest-voted posts (with total vote scores ranging from 22 to 2724). During the hurricane, though, moderators deleted hundreds of posts, and these posts were not collected.

We also scraped user data for the 1895 distinct users that contributed each post, which includes each user’s post and comment karma scores (sum of votes users receive for their own submissions and comments) and account age. Since posting, some users deleted their accounts or removed the association between their account and the post; in these cases, we removed them from the model.

From these data, we coded the top 1000 highest-voted threads (for which, along with thread authors, we provide summary statistics below in Table 1) according to content type and content format. We use the top 1000 posts because we are interested in patterns among different types of popular content (rather than differences between successful and failed content). For content type, we distinguished among four categories. We derived them using a grounded theory approach, where we took a sample of posts and coded them based on their content, using categories in line with those employed by [4] and [7]:

- Professional news reports (PROF)
- Users’ perspective-centric content, commonly referred to as “citizen journalism” (PERSP)
- Content that users connected to on an emotional level or that would be of human interest related to Hurricane Sandy (INTER), and
- Conversational posts related to meta-discussions, as well as asking questions, sharing tips, and issuing corrections (CONV).²

² We further subcategorized the content into humor, tips, questions, corrections, and meta posts, though these were only used for qualitative analysis.

The content format was split among four categories:

- External textual websites (SITE)
- Internal subreddit textual self posts (REDDIT)
- Video and audio multimedia (VIDEO), and
- Images, infographics, charts, and maps (IMAGE).

Examples are shown in Table 2. Between two coders, we each labeled every post; for those we disagreed on initially, we reached consensus through discussion. The coder inter-reliability rate was calculated at Cohen's kappa = 0.72 ($p < 0.001$).

The variables we use in the statistical model are as follows:

- Score: *H1*. A measure of upvotes minus downvotes (calculated by reddit's platform).
- Comments: *Control*. The number of comments a post has garnered.
- Approximate Views: *Control*. A rounded number of logged in users viewing (and able to vote on) posts in the subreddit at the time of a post's submission.³
- User Experience: *H2a*. A measure of poster's account age in days (normalized to years).
- User Activity: The number of submissions (*H2b*) or comments (*Control*) of the poster (rescaled to every 100 posts).
- User Reputation: The number of submission (*H2c*) or comment (*Control*) karma points the poster has earned (rescaled to every 1000 karma points).
- Content Code: *Dependent variable*. Categorical variable of PROF, PERSP, INTER, or CONV.

RESULTS

Networked Gatekeeping in Social News Sites

A social news site, through its technical architecture, allows the community to vote on what content is important and relevant to those users. The algorithmic implementation of upvoting and downvoting only goes so far, though: social norms and processes structure those behaviors too.

/r/sandy was repurposed on October 26, 2012. All subreddits have a sidebar where the moderators can post information and rules for the subcommunity, and from the start /r/sandy's read "Information and discussion around Hurricane Sandy and the possible Frankenstorm/ Snowicane/ Noreastercane/ Floodmageddon. Post your personal experiences, photos, and video of Hurricane Sandy." Initially, users driven to the subreddit did not have much guidance regarding what would be "proper" to post, and the first few submissions included a cocktail recipe list, a joke image, and a note to stock up your fridge. As the subreddit attracted more and more viewers (in fact, viewership eclipsed the number of subscribers early on; a rare occurrence on reddit), hundreds of posts were

Variable	N	Min	Max	Median	Mean	Std. Dev.
Score ⁴	999	22	2724	63	176.7	332.6
Comments	999	0	514	9	21.63	38.5
Approximate Views	999	10	6860	1300	2516	2480.9
User Experience (Days)	966	46.9	2607	440.9	610.7	548.6
User Submission Activity	966	0	1000*	39.5	131.2	232.8
User Comment Activity	966	0	1000*	269	422.9	387.1
User Submission Reputation	966	1	50000**	1031	5000	10795.2
User Comment Reputation	966	0	50000**	921	3824	7628.2

* Capped at 1000. ** Capped at 50000.

Table 1. Descriptive statistics of top 1000 /r/sandy posts and their contributors (excluding rows with NA values).

Type	N	Example
PROF	162	Some subway tunnels flood - CBS
PERSP	331	My moms backyard in Hatteras. That dock is usually about 3 feet above water (<i>sic</i>)
INTER	282	Sandy from space
CONV	224	Please, if you're currently without power please read
Subtype	N	Example
HUMOR	81	Horse Man Caught On Film During Live Sandy Coverage
TIP	128	Tips from a Louisiana native who lived through Katrina
QUEST	76	Is there a list of places I can volunteer in NYC?
CORR	7	9 Viral Photos That AREN'T Hurricane Sandy
META	13	Please consider including location in the title of images you post.
Format	N	Example
SITE	208	Al Jazeera gives /r/sandy a shoutout... Thought you guys might like to know...
REDDIT	211	Update your situation here
IMAGE	495	Photo taken at 9am EST in downtown NYC
VIDEO	85	The Weather Channel Live Stream

Table 2. Content types, subtypes, and formats of highest scored /r/sandy submissions.

³ These numbers were scraped by stattit.com.

⁴ Reddit employs a vote fuzzing algorithm to prevent spam, but ultimately the score reflects the total contributions of users' voting behaviors.

contributed, and eventually users took to self posts to help shape norms around submitting, voting, and moderation.

For example, one user posted `Mods, can you please stop all the jokes? This is a serious storm and I am using /r/sandy for knowledge and safety.` This post marked one of the earliest points of conflict for the subreddit: between its use as a news and information hub versus a space for any related media, including jokes. A second user commented:

I think we're okay making jokes and humor so long as they're clearly distinguishable from the actionable information. ... I support the humor. I support the useful information. If we don't confuse the two, we can stay safe and happy without feeling too alone.

Less than an hour later, another user contributed a post entitled `I love you all, but can we please try and keep the jokes at a minimum?` The top voted comment in response was “You shouldn't rely on reddit for life critical information. Use your local news.” Debates like these continued throughout the first few days of the hurricane, with users debating back and forth about what should and should not be allowed.

In addition to the content validity debate, other users were concerned about participatory behaviors of the community, namely that more users needed to vote so that important and relevant contributions could become visible on the top of /r/sandy's front page. These calls to action are directly in line with Gilbert's observations of underprovision on reddit [10] and the consequences this lack of participation has for users in ignoring content at the time of its first introduction to the platform. Two of the highest voted conversational meta posts were `We need to keep this subreddit fresh! View the new! and If you're not in the storm, one thing you can do is stroll through New to upvote life-saving posts and downvote jokes.` These related to the fact that not enough users were visiting the “new” page, which aggregates all the newest submissions, and voting up the important information and downvoting less relevant content. In the comments of these posts, we see sustained negotiation of proper content – “And please downvote the blatant karma-whoring. Idiot Redditors are trying to capitalize on an emergency situation with memes and screen caps of novelty Twitter accounts.” – as well as continued suspicion around the value and purpose of the subreddit – “I see what you're trying to accomplish but I also find it hard to believe someone's gonna be in a life threatening situation and be like ‘Oh lets check /r/sandy for advice on what I should do at this precise moment.’”

Of course, as different types of content were contributed over the few days, debates arose over the value of /r/sandy for posting and reading critical news content. One user complained, “You shouldn't rely on reddit for life critical information. Use your local news.” Other users countered

this point, given that many professional news articles had been shared, making the subreddit a helpful source of information. When a user responded “I'm pretty sure there are better sources for knowledge and safety on the internet than Reddit,” another replied “[C]ould you please post them then? [I]’ve been following here and NOAA and trying to stay away from the news.”

Along with these community-level debates, elite-level moderators actively helped structure much of what would be allowed in the subreddit. In an email interview with three of /r/sandy's moderators, one commented, “I was watching the news and moderating for several days with little sleep.” The mods engaged in similar discussions as users about content. One responded to us:

The goal from the beginning was to keep information flowing and accurate... to remove hearsay and things we just knew to be lies or posts from previous hurricanes just recirculating. ... We wanted this sub to be an emergency lookup location, so we wanted our information to be solid and more serious content. At one point... we were getting about 15 posts every 20 seconds posted. It was just a massive influx of data that we had to just constantly hit refresh on and delete content not deemed to be following what our mindset was.

The moderators deleted hundreds of contributions as they were posted to the subreddit, though they were relatively permissive of what content went through. In our interview, one moderator wrote, “The only things deleted were crass, rude humor that was deemed insensitive and memes...” Everything else was allowed to be voted upon by the users. In addition, the moderators actually removed *another* moderator who they felt was conflicting with the goals of the subreddit. As one moderator recounted:

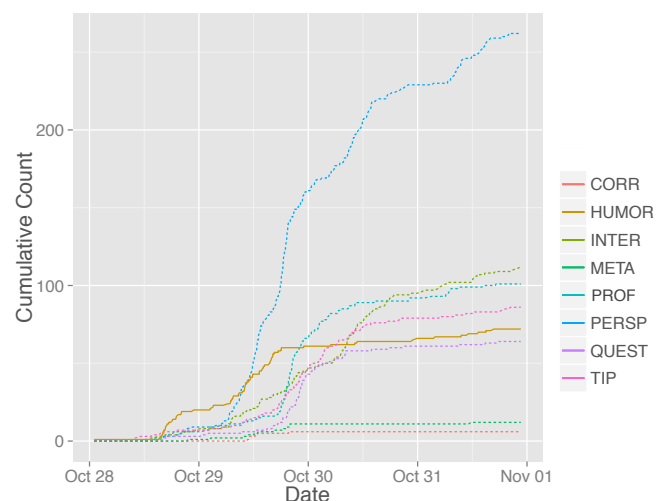


Figure 1. Truncated timeline of content types and subtypes (subtracted from types) during first few days of activity in /r/sandy, highlighting the rise and fall of humor (solid orange line) compared to other content (dotted lines).

	Score	H1		Comments			Approximate Views			User Experience	H2a	
	$Exp(\beta)$	p	z	$Exp(\beta)$	p	z	$Exp(\beta)$	p	z	$Exp(\beta)$	p	z
INTER	1.0033	***	3.4169	1.0018		0.2605	0.9997	***	-6.2783	0.9483		-0.8293
PERSP	1.0026	***	2.6703	1.0034		0.5092	0.9999	*	-1.8345	0.9310		-1.0907
CONV	1.0006		0.5574	1.0194	***	2.9328	1.0000		0.7192	0.8413	**	-2.4612
	User Submission Activity	H2b		User Comment Activity			User Submission Reputation	H2c		User Comment Reputation		
	$Exp(\beta)$	p	z	$Exp(\beta)$	p	z	$Exp(\beta)$	p	z	$Exp(\beta)$	p	z
INTER	0.9189	*	-1.5635	0.9755		-0.7353	1.0072		0.5802	0.9904		-0.6303
PERSP	0.7474	***	-4.9703	0.9919		-0.2399	1.0010		0.0697	0.9992		-0.0488
CONV	0.8644	**	-2.4617	1.0449		1.2734	0.9870		-0.8619	1.0068		0.4317

Baseline: PROF

McFadden's Pseudo R^2 : 0.1119* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

N: 967

AIC: 2455.4

Table 3. Multinomial logistic regression for H1, H2a, H2b, and H2c (in blue; controls in grey).

We had a problem with one mod and I had to contact [the reddit admins] to get rid of him. I talked with three other mods who I thought also had a problem with him so we had a little vote and I wrote a letter [to the admins] with my concerns... [they] told him his services were no longer needed... [H]e was going around and undeleting things we deleted (and sent explanatory replies to), just generally undermining. Really hard to deal with...

These top-down actions, along with the distributed moderation-through-voting of the users, represent the ecosystem of networked gatekeeping that structures what becomes (and can become) visible as important or relevant in a social news setting. In a visual example, Figure 1 demonstrates how humor, in relation to other categories, dominated the top 1000 submissions in the first few days until the social norms and actions of networked gatekeeping put a halt to its continued growth.

Event-Based Content Types in Social News Sites

Given that networked gatekeeping determines what can and does become visible on a social news site, we next use a predictive model to examine how voting behaviors and users in the community impact the emergence of certain types of content related to the crisis. Hypothesis 1 and 2 examine how types of content become marked as important through the distributed gatekeeping of a social news site (upvoting and downvoting, as demarcated by a post's score) as well as what kind of users produce those types of content. To assess both hypotheses, we carried out a multinomial logistic regression with content categories as the dependent variables (using PROF as a baseline) and post, poster, and subreddit attributes as the independent variables. The results are laid out in Table 3; the exponentialized beta ($Exp(\beta)$) represents the relative risk ratio, or an increase or decrease in the relative likelihood (compared to PROF) of a post being a specific type of content per a one unit increase in the independent variable.

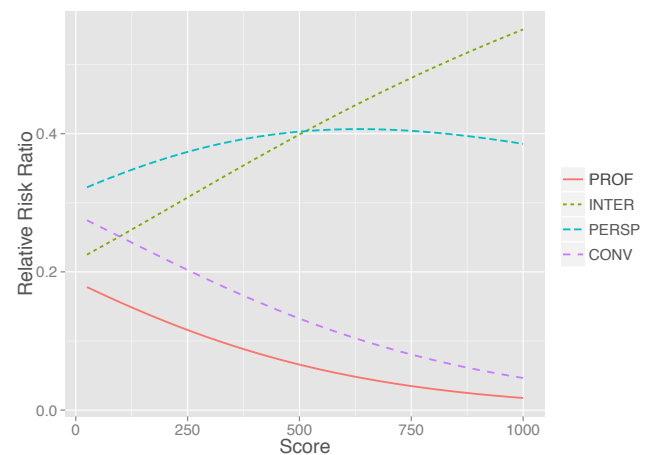


Figure 2. Predicted relative probabilities of content types from multinomial logistic regression using post's score (H1) as x-axis (25-1000) and holding other variables at their means.

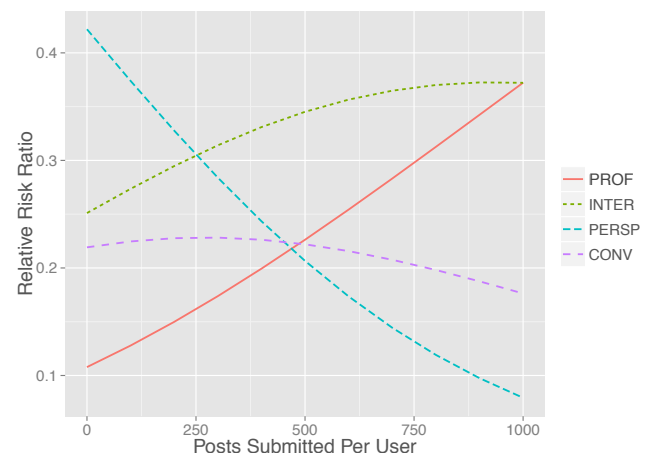


Figure 3. Predicted relative probabilities of content types using user's submission activity (H2b) as x-axis (0-1000) and holding other variables at their means.

Hypothesis 1 stated that professional news reports are more likely to be scored higher by the community than other content posted to /r/sandy. Conversely, the model predicted that all other types of content are relatively more likely than news content as a score increases for any particular post (the predicted relative risk ratio of each content type is visualized in Figure 2). Of these types, only human-interest posts and perspective-based citizen journalism were significant. Thus, Hypothesis 1 was not supported.

Hypothesis 2 stated that professional news reports are more likely to be contributed by a user with higher experience (2a), activity (2b), and reputation (2c). In relation to a user's experience, more experienced users were less relatively likely of posting other content compared to professional reports but only for conversational posts. Therefore, Hypothesis 2a was partially supported. Users' activity for submissions led to non-professional content being relatively less likely than news, and the effects were significant across all types of content, supporting Hypothesis 2b (like H1, summarized in Figure 3). The effects based on users' submission reputation moved in varying directions, and none were significant, suggesting Hypothesis 2c was not supported.

DISCUSSION

This study analyzes the dynamics of event-based news production around framing and gatekeeping in a social news setting. We show that professional news reports are less likely to receive higher scores compared to other types of content. Also, we demonstrate that more experienced users are more likely to contribute professional reports than conversational posts, and users that contribute more submissions are more likely to contribute professional news reports than all other types of content.

Who participates in “news” on a social news site?

As a massively-popular social news site, reddit occupies a particularly interesting space in terms of viewing and sharing content. While reddit boasts tens of millions of viewers on its front page, many fewer participate on the site's default subreddits. Fewer still collaborate to post and vote on content in smaller topical subcommunities. A niche, event-based subreddit like /r/sandy epitomizes a unique case of reddit subcommunities, in that it was created for a specific purpose at a very particular time period. Because of the subreddit's relevancy as a potential news source, it gained an abnormal amount of subscribers and viewers in a short duration. To attract contributors to /r/sandy, reddit administrators posted an advertisement (Figure 4) for the /r/sandy subreddit on the right-side column of the platform, making users aware of the new subreddit's existence. As stated by one /r/sandy moderator (via email interview, quoting statistics only available to mods), the subreddit received 401,675 unique visitors and 4,429,554 pageviews in October (just from the last four days of the month, with the peak occurring just after the storm), 172,102 unique

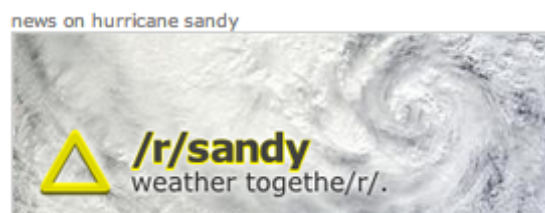


Figure 4. Screenshot of the advertisement on reddit's front page to introduce /r/sandy to the wider community.

viewers and 1,148,667 views in November, and 15,568 unique viewers and 28,435 views in December.

We establish in Hypothesis 2 that certain types of users are relatively more likely to post professional news reports than other content: namely, users that have posted many submissions before (and older, experienced users, when comparing news and conversational contributions). However, the insignificance of most of the user attributes suggests that both inexperienced and experienced users, users of all levels of reputation, and users who post any number of comments are all relatively likely to post any type of content. This falls in line with Shklovski, Palen, & Sutton [31] who show that, in crisis situations, newer and less experienced users actively contribute alongside experienced participants in platforms oriented toward disaster communication.

From a methodological perspective, the varying utility of user attributes employed in this paper point to the necessity of understanding the platform's cultural context in its use. For example, reddit's karma points are awarded by individuals in the community but across a variety of subcommunities with differing purposes. Therefore, karma may be insignificantly related to the posting of types of content because, when the community upvotes posts and comments with karma, users will have varying reasons for awarding reputation points (e.g., upvoting for funny posts in one subreddit versus for relevant or erudite posts in another subreddit).

We do see, though, that if a user has posted more, they are more likely to post professional news content in /r/sandy than other types of content. We surmise that this effect could stem from highly-active posters being able to identify uses of subreddits well, since they have a lot of experience posting in reddit's subcommunities. For /r/sandy, then, active posters may have identified its core purpose as sharing professional news.

Conversely, less active users (who have posted fewer times in general on reddit) may have come to post on /r/sandy to contribute a one-off submission without much awareness of the point of that post. Still, users who are lurkers (who have high experience but low activity) might have come to the subreddit based on exogenous factors (like its relevancy to their local situation). With high experience, these users understand reddit's culture and context well, contributing

content that they know reddit's audience appreciates (humor, interesting media, etc.), or they possibly found that they could participate in some way, such as posting a picture from their backyard – as the model suggests, posting something other than professionally-reported news content. Further qualitative research on posting behaviors could illuminate the discrepancies behind this conduct.

How does reddit frame “news”?

The nature of reddit's technical platform mixed with its social norms allow for a unique context of networked gatekeeping to impact the framing of event-based news. In Hypothesis 1, we expected to find that, amongst the popular submissions, professional news content would likely be scored higher relative to other content. However, our model suggests that human-interest and perspective-created content would more likely appear in higher-voted submissions.

The results of Hypothesis 1 are intriguing in that they conflict with what the literature has suggested would be seen as more important by social media users (ie., professional news reports). Instead, we see that social news sites operate differently from other social media sites. We must also keep in mind that attention drives much of what is produced and mediated in social news sites: as Muchnik, Aral, and Taylor show [23], even one initial upvote in a social news platform can lead to a strong and consequential social influence bias.

But why do we see /r/sandy score other content higher more than professional news reporting? As we explored in our qualitative analysis, it is clear that some members had an idea of the subreddit's ideal use, but the technical platform and the moderators did not strictly limit what types of content could be shared. While some information was considered useful, a presumption that “news” would be produced mischaracterizes the purpose of a social news site, where instead of limiting types of information, the community makes visible through its own means content they find relevant. The /r/sandy collective still had expectations about proper content, but what resulted was in a sense engineered through networked gatekeeping: moderator actions, discursive negotiation, and bottom-up voting.

The expectations around proper content also derive from reddit's participatory culture. As Bergstrom [3] notes, participation on reddit is accompanied by a strong subculture and peculiar sets of community expectations, depending on the particular subreddit collective. Content like humor and processes like bandwagon effects permeate the behaviors of reddit's community at large. Thus, social norms become particularly important in structuring what kind of submissions and comment-driven conversations are denoted as valuable to each subreddit community. Even though we saw a drop-off in humor and a significant attempt by both moderators and users to push for helpful information, the highest voted submissions were shown to

relatively less likely be traditional news reports. While news was not necessarily downvoted or removed from the subreddit, the community relatively preferred other content. Keep in mind though that some users *did* downvote this content, but more people upvoted it, resulting in positive scores.

Moderation of these social norms especially becomes important when considering the technical implementation of social news sites' core feature, voting. Reddit's “reddiquette” emphasizes upvoting if content contributes to conversation, and downvoting if it is off-topic and does not contribute to the subreddit.⁵ Still, some users might upvote if they like the content and downvote if they dislike it (actions that we observed users talking about when discussing their general reddit use, albeit in subreddits other than /r/sandy). Preferential voting might then lead to the outcomes we observe in /r/sandy, where things that people like on an emotional level (like human-interest content) win out over objective (and possibly, given the context, more relevant) information.

Also, /r/sandy was not the only space on reddit for news sharing, so it is important to highlight /r/sandy's status as a small subreddit within reddit's larger culture of posting behaviors. Not all users post content only to their respective topical subreddits; usually, more general content can be found interspersed across the default subreddits. To illustrate this point, we can look at the top 50 *all-time* submissions that mention Hurricane Sandy across reddit, which were posted to /r/pics (14), /r/politics (10), /r/sandy (7), /r/funny (5), /r/gifs (3), /r/AdviceAnimals (2), and each of nine other subreddits. Therefore, some users might not even consider /r/sandy as a possible place to post content about the hurricane. But, as our analysis shows, users framed the important contributions in /r/sandy – again, through voting – as interesting tidbits about the hurricane or content that reflected the personal positions of those affected locally (as compared to news content).

LIMITATIONS & FUTURE RESEARCH

Additional data analysis could help illuminate the dynamics underlying the production and framing of news on social news sites. For example, we did not include coded content of the population of submissions posted to /r/sandy (instead focusing solely on more-popular content), but investigation of the spectrum of posts to an event-based topical community could further illustrate the differences between content types (ie., looking at the least popular content in addition to the most popular). We also attempted, in the exploratory phase, a temporal analysis of the data to control for types of content fluctuating at different periods. However, after slicing the data into pre-, during-, and post-event periods, the samples in each set spread across our

⁵ <http://www.reddit.com/wiki/reddiquette>

four coded categories were not sufficient for multinomial logistic regression.

Further qualitative research could also examine why social news settings – and reddit in particular – generate particular types of content and how these types flow across different subcommunities (e.g., /r/sandy appears to be filled with more personal content than the more-officially-reported /r/olympics). In particular, studying the dynamics of networked gatekeeping – especially voting behaviors – can help us understand how certain social values are translated to gatekeeping norms, as we note above.

This study contributes knowledge that has implications for the design of community-moderated platforms, as well as the design of such platforms for event-based situations like communication during crises. We especially emphasize the importance of architecting tools for determining strong social norms in a social news site context, as it becomes clear from our qualitative observation that these rules strongly influenced content production and negotiation in the subreddit. We recognize that /r/sandy represents a unique example of event-based news production processes in a social news setting, particularly in its use for crisis communication. During the writing over this paper, there were no comparable examples of a subreddit used for this purpose, though studying /r/sandy provides insight into the potential design of such platforms. However, we do believe that the social negotiations and cultural context of reddit permeate most subcommunities on the site, and thus these observations could be seen as generalizable.

Overall, this study shows in detail what kinds of content emerge in critical information moments under the moderation of networked gatekeeping. We demonstrate that information sharing is not a normative practice and differs in social news sites as compared to other social media based on the literature. Future research should consider approaches that provide nuanced details about the data that is produced and shared to further understand the dynamics behind such practices. The study also illustrates that various types of moderation crucially impact what content becomes relevant (and indeed what content can be curated to be voted on at all). Unlike other social media platforms, reddit's integration of networked gatekeeping into the norms of information production within its subcommunities ultimately changes the way that we can think about the emergence of relevant information in sociotechnical systems.

ACKNOWLEDGMENTS

We would like to thank Kate Starbird, Brian Keegan, Peter Monge, Dennis Wixon, Dmitri Williams, fellow Annenbergers, and our anonymous reviewers from ICWSM, CSCW, and CHI for their helpful feedback. We also extend our gratitude to the /r/sandy moderators and reddit users that provided data and comments.

REFERENCES

1. Acar, Adam, and Yuya Muraki. 2011. "Twitter for Crisis Communication: Lessons Learned from Japan's Tsunami Disaster." *International Journal of Web Based Communities* 7 (3): 392–402.
2. Barzilai-Nahon, Karine. 2008. "Toward a Theory of Network Gatekeeping: A Framework for Exploring Information Control." *J. Am. Soc. Inf. Sci. Technol.* 59 (9): 1493–1512.
3. Bergstrom, Kelly. 2011. "Don't Feed the Troll": Shutting down Debate About Community Expectations on reddit.com." *First Monday*; 8(1).
4. Bruns, Axel, Jean E. Burgess, Kate Crawford, and Frances Shaw. 2012. "#qldfloods and @QPSMedia: Crisis Communication on Twitter in the 2011 South East Queensland Floods." In *Brisbane QLD Australia: ARC Centre of Excellence for Creative Industries and Innovation*, Queensland University of Technology.
5. Burke, Moira, Cameron Marlow, and Thomas Lento. 2009. "Feed Me: Motivating Newcomer Contribution in Social Network Sites." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 945–954. CHI '09. New York, NY, USA: ACM.
6. Chen, Brian. 2012. "How Reddit Scooped the Press on the Aurora Shootings." *The New York Times*. Retrieved from <http://bits.blogs.nytimes.com/2012/07/23/reddit-aurora-shooter-aff/>
7. Chew, Cynthia, and Gunther Eysenbach. 2010. "Pandemics in the Age of Twitter: Content Analysis of Tweets During the 2009 H1N1 Outbreak." *PLoS ONE* 5 (11).
8. Gilbert, Eric. 2013. "Widespread Underprovision on Reddit." In *Proceedings of the 2013 ACM Conference on Computer Supported Cooperative Work*, 803–808. CSCW '10. San Antonio, TX, USA: ACM.
9. Goode, Luke. 2009. "Social News, Citizen Journalism and Democracy." *New Media & Society* 11 (8) (December 1): 1287–1305.
10. Hughes, Amanda Lee, and Leysia Palen. 2009. "Twitter Adoption and Use in Mass Convergence and Emergency Events." *International Journal of Emergency Management* 6 (3): 248–260.
11. Iyengar, Shanto. 1994. *Is Anyone Responsible?: How Television Frames Political Issues*. Reprint. University Of Chicago Press.
12. Keegan, Brian, and Darren Gergle. 2010. "Egalitarians at the Gate: Onesided Gatekeeping Practices in Social Media." In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*, 131–134. CSCW '10. New York, NY, USA: ACM.

13. Keegan, Brian, Darren Gergle, and Noshir Contractor. 2013. "Hot Off the Wiki: Structures and Dynamics of Wikipedia's Coverage of Breaking News Events." American Behavioral Scientist Forthcoming.
14. Lampe, Cliff, Nicole B. Ellison, and Charles Steinfield. 2008. "Changes in Use and Perception of Facebook." In Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work, 721–730. CSCW '08. New York, NY, USA: ACM.
15. Lampe, Cliff, and Erik Johnston. 2005. "Follow the (slash) Dot: Effects of Feedback on New Members in an Online Community." In Proceedings of the 2005 International ACM SIGGROUP Conference on Supporting Group Work, 11–20. GROUP '05. New York, NY, USA: ACM.
16. Lampe, Cliff, and Paul Resnick. 2004. "Slash(dot) and Burn: Distributed Moderation in a Large Online Conversation Space." In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 543–550. CHI '04. New York, NY, USA: ACM.
17. Lerman, Kristina, and Rumi Ghosh. 2010. "Information Contagion: An Empirical Study of the Spread of News on Digg and Twitter Social Networks." Proceedings of 4th International Conference on Weblogs and Social Media (March 12).
18. Lewin, Kurt. 1943. "The Forces Behind Food Habits and Methods of Change." In The Problem of Changing Food Habits Report of the Committee on Food Habits 1941-1943. The National Academies Press.
19. Long, J. Scott, and Jeremy Freese. 2005. *Regression Models for Categorical Dependent Variables Using Stata, Second Edition*. 2nd ed. Stata Press.
20. Lotan, Gilad, Erhardt Graeff, Mike Ananny, Devin Gaffney, Ian Pearce, and danah boyd. 2011. "The Revolutions Were Tweeted: Information Flows During the 2011 Tunisian and Egyptian Revolutions." International Journal of Communication 5.
21. Martin, Erik. 2012. "2 Billion & Beyond." Blog.redd.it. <http://blog.redd.it.com/2012/01/2-billion-beyond.html>.
22. Marwick, Alice E., and danah boyd. 2011. "I Tweet Honestly, I Tweet Passionately: Twitter Users, Context Collapse, and the Imagined Audience." New Media & Society 13 (1): 114–133.
23. Muchnik, Lev, Sinan Aral, and Sean J. Taylor. 2013. "Social Influence Bias: A Randomized Experiment." Science 341 (6146): 647–651.
24. Palen, Leysia, Sarah Vieweg, Sophia B. Liu, and Amanda Lee Hughes. 2009. "Crisis in a Networked World Features of Computer-Mediated Communication in the April 16, 2007, Virginia Tech Event." Social Science Computer Review 27 (4) (November 1): 467–480.
25. Poor, Nathaniel. 2005. "Mechanisms of an Online Public Sphere: The Website Slashdot." Journal of Computer-Mediated Communication 10 (2).
26. Roberts, Chris. 2005. "Gatekeeping Theory: An Evolution." Paper presented at the Annual Meeting of the Association for Education in Journalism and Mass Communication, San Antonio, TX.
27. Scellato, Salvatore, and Cecilia Mascolo. 2011. "Measuring User Activity on an Online Location-based Social Network." In 2011 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), 918–923.
28. Scheufele, Dietram A. 1999. "Framing as a Theory of Media Effects." Journal of Communication 49 (1): 103–122.
29. Schultz, Friederike, Sonja Utz, and Anja Göritz. 2011. "Is the Medium the Message? Perceptions of and Reactions to Crisis Communication via Twitter, Blogs and Traditional Media." Public Relations Review 37 (1): 20–27.
30. Shaw, Aaron. 2012. "Centralized and Decentralized Gatekeeping in an Open Online Collective." Politics & Society 40 (3): 349–388.
31. Shklovski, Irina, Leysia Palen, and Jeannette Sutton. 2008. "Finding Community Through Information and Communication Technology in Disaster Response." In Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work, 127–136. CSCW '08. New York, NY, USA: ACM.
32. Suh, Bongwon, Lichan Hong, P. Pirolli, and E.H. Chi. 2010. "Want to Be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network." In 2010 IEEE Second International Conference on Social Computing (SocialCom), 177–184.