

Newsworthiness and Network Gatekeeping on Twitter: The Role of Social Deviance

Nicholas Diakopoulos

Columbia University, School of Journalism
nicholas.diakopoulos@gmail.com

Arkaitz Zubiaga

Dublin Institute of Tech., Applied Intelligence Research Centre
arkaitz@zubiaga.org

Abstract

Publishers of news information are keen to amplify the reach of their content by making it as re-sharable as possible on social media. In this work we study the relationship between the concept of social deviance and the re-sharing of news headlines by network gatekeepers on Twitter. Do network gatekeepers have the same predilection for selecting socially deviant news items as professionals? Through a study of 8,000 news items across 8 major news outlets in the U.S. we predominately find that network gatekeepers re-share news items more often when they reference socially deviant events. At the same time we find and discuss exceptions for two outlets, suggesting a more complex picture where newsworthiness for networked gatekeepers may be moderated by other effects such as topicality or varying motivations and relationships with their audience.

Introduction

The analytics purveyor Chartbeat recently reported that 26% of the traffic they measure to news sites is from social sources.¹ Now more than ever news publishers need to understand how to harness social platforms like Twitter and Facebook to disseminate information and reach larger audiences, both for breaking news as well as for headlines pointing users back to the publisher's content (Kwak et al. 2010; Messner, Linke, and Eford 2011). A strong social presence allows news publishers not only to engage their community of readers with the latest news, but also to implicate those readers as network gatekeepers (Barzilai-Nahon 2008) who can further share that news.

Twitter offers an unprecedented opportunity to quantify and analyze how news arouses interest by observing the number of users who re-share a news story. Motivated by findings proffered in previous research on news coverage, here we take up the relationship between the concept of *social deviance* and the re-sharing of news headlines by

network gatekeepers. Previous literature has found that events with high deviance were more likely to gain coverage in the main stream media (MSM) (P. Shoemaker, Danielian, and Brendlinger 1991), “the role of news media is not to mirror the world as it is, but rather to spotlight and draw public attention to problems and situations that need solutions and repair,” (Pamela Shoemaker 2006). Shoemaker's theory posits that MSM will select for and favor socially deviant stories and events.

But do network gatekeepers share a similar predilection for selecting socially deviant news items? Is social deviance a professionally imbued newsworthiness criterion, or something that more generally explains interest, attention, and sharing of news? Networked gatekeeping theory explores the idea that every user on Twitter is a gatekeeper, with the discretion to share or not share a news item with their audience (Barzilai-Nahon 2008). Each user can have their own criteria for what becomes “news”—what's worthy of sharing. In this paper we explore how social deviance relates to the re-sharing of news headlines by network gatekeepers on Twitter.

In particular, we study 8,000 news stories posted on Twitter by 8 major U.S. news outlets and examine the distribution of socially deviant tweets and the relation to number of retweets. Our results show that network gatekeepers do tend to re-share news stories at a higher rate when they reference socially deviant events, particularly for tabloid news outlets' content. At the same time we find and discuss exceptions for two outlets, suggesting a more complex picture where newsworthiness for networked gatekeepers may be moderated by other effects such as varying motivations or relationships with the audience. In addition, our study complements and broadens previous research that has looked at how network characteristics (Bhattacharya 2012), Twitter-specific features such as hashtags and URLs (Suh et al. 2010), context (Nahon and Hemsley 2013) and sentiment and emotion (Berger and Milkman 2012) impact the social spread of information.

Analysis of Social Deviance

In this section we describe our data collection process, define and operationalize social deviance, and present results for how social deviance varies across outlets and plays a role in re-sharing of headlines.

Data Collection

We identified the top 25 U.S. daily newspapers by circulation from the 2012 AAM circulation report.² To ensure sufficient variance in the number of retweets for each outlet we focused on the eight newspapers with more than 100k Twitter followers: the Wall Street Journal (@WSJ), USA Today (@USAToday), New York Times (@nytimes), Los Angeles Times (@latimes), NY Daily News (@nydailynews), New York Post (@nypost), Washington Post (@washingtonpost), and Chicago Tribune (@chicagotribune).

For each outlet's account we collected all tweets from Nov. 1st, 2011 to Oct. 31st, 2012. Across the eight outlets this resulted in 119,498 tweets, of which 107,066 had news headlines with content links to the site of the form: <story headline> + <link>. Since this is the most prevalent pattern of use of Twitter by news organizations (Messner et al. 2011) we focus our analysis on these tweets. We used the Topsy API to collect historical tweets due to Twitter API limitations.³ However, retweet information was collected directly from the Twitter site.

Defining Social Deviance

Different societies, cultures, and sub-cultures often have their own social norms of what constitutes acceptable behavior within that group. More formally, a social norm can be defined as a "stable, shared conception of the behavior appropriate or inappropriate to a given social context, that dictates expectancies of others' behavior, and provides 'rules' for one's own behavior" (McKirnan 1980). When a person violates a social norm, it is considered to be a socially deviant action. Social norms often but not always overlap with legal norms of behavior: for instance, murder is a violation of both a legal code and a social norm, but plagiarism, while not illegal, is a violation of a social norm. In this work we consider a news event to be socially deviant when it violates a social or legal norm. (P. Shoemaker et al. 1991).

Here we study the re-sharing behavior of network gatekeepers on Twitter depending on whether the news headline references a socially deviant event. Examples of news stories involving social deviance include robbery, homicide, or violence, while examples with no social deviance include a sports game, a political event, someone's passing, or a natural disaster where no human is violating any norm.

² <http://www.auditedmedia.com/>

³ <http://code.google.com/p/otterapi/>

Measuring Social Deviance

With the aim of producing a subset of tweets coded for whether they are socially deviant or not, we carried out an iterative process to create a groundtruth and to define sound coding guidelines for Amazon Mechanical Turk (AMT) workers. In the first iteration, 100 tweets were randomly selected from our dataset and each coded by the two researchers with a binary value for social deviance. From an initial agreement of 92% between the two researchers, most coding disagreements were resolved through iteration; discussion resulted in an agreement of 98% and a Fleiss' Kappa inter-coder agreement of 0.96. Through this initial process we were able to articulate a more precise definition of social deviance to be provided to the AMT workers:

"Does the headline refer to an event or topic that is socially deviant, such as a violation of social norms or challenge to the status quo, as understood from a contemporary U.S. cultural point of view?"

Each of the 100 tweets was coded by five workers, with the restriction that workers had to be from the U.S. Using majority voting, we labeled the tweets deemed deviant by workers and compared them to the researchers' groundtruth. This led to an agreement rate of 96% and a Fleiss' Kappa of 0.93 with respect to the ground-truth, indicating that the workers were largely able to understand and apply the definition of social deviance that we supplied.

We then selected a larger subset of tweets to be coded according to deviance. We sampled 1,000 tweets from each of the eight news outlets under study. These were selected such that 200 came from each of five quintiles as defined by number of retweets. This provided a variety of tweets across a range of levels of retweeting and allowed us to see differences in deviance patterns across quintiles. The final set of tweets was coded by AMT workers according to the guidelines above: each tweet was coded by five workers and the final decision was made using majority voting. For example, "Principal busted for spying on students through Facebook quits" was rated as a deviant tweet in our corpus, whereas "Oil jumps to 9-month high after Iran cuts supply" was rated as non-deviant.

Results

The coding process resulted in a balance of 1,512 deviant (18.9%) and 6,488 non-deviant (81.1%) tweets. Three outlets stand out with deviance rates above 25%: NY Daily News (30.6%), New York Post (27.4%) and Chicago Tribune (26.5%). The least deviant outlet was the Wall Street Journal (7.7%). The other outlets had moderate amounts of deviance: New York Times (12.9%), Washington Post (14.0%), USA Today (14.8%), and LA Times (17.3%). Outlets tweeted 121 links that were repeated across tweets with different headlines e.g. to a homepage, news app, or evolving blog, however there was

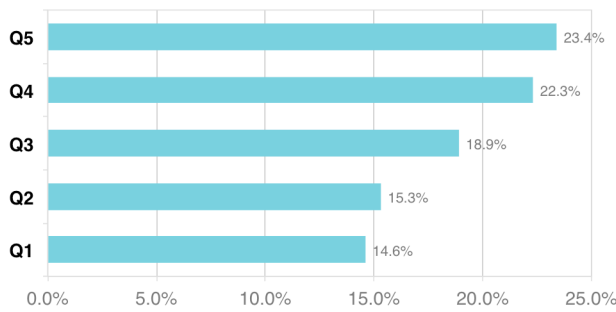


Figure 1. Percentage of deviant tweets across quintiles defined by number of retweets (Q5 is top, Q1 is bottom)

no difference in the rate of deviant vs. non-deviant events in those tweets with repeated links. In line with related work (P. Shoemaker et al. 1991), these results underscore deviance as a salient newsworthiness criterion for many of these outlets, though with considerable variance between them.

Figure 1 shows the distribution of deviant tweet rates across quintiles defined by the number of retweets and aggregated for all outlets. The two top quintiles (Q4, Q5) have the highest rate of deviant tweets and the rate decreases in lower quintiles, the bottom quintile being the lowest ($\chi^2 = 35.298$, $DOF = 4$, $p < 0.0001$). This supports the idea that deviant news items do evoke additional attention and re-sharing from network gatekeepers: among the tweets that are more often retweeted (i.e. Q5) deviant topics are more prevalent and comprise a larger portion of the news.

We also looked at the distribution of deviance for each outlet to understand any variation in network gatekeeping behavior across outlets. With the exceptions of the WSJ and Chicago Tribune (which we discuss further below), the other six outlets had substantial differences between their bottom and top quintiles in terms of proportion of deviant tweets (See Figure 2). Papers known to have reputations as more balanced and objective sources (i.e. NY Times, LA Times, and Washington Post) tend to have smaller disparities among bottom and top quintiles, whereas papers with tabloid reputations (i.e. NY Post and NY Daily News) ⁴ have much larger disparities (e.g. NY Post, at 24%, was largest). So not only do the tabloids have higher rates of socially deviant headlines (30.6% and 27.4%), but those headlines also comprise a larger fraction of the top retweeted tweets, suggesting that the audiences for those outlets may be particularly interested in re-sharing socially deviant news; for those network gatekeepers social deviance is an especially salient newsworthiness criterion.

To provide stronger statistical evidence of the relationship between deviance and retweets, both overall and for each outlet, we perform Mann-Whitney-Wilcoxon (MWW) tests. We normalize retweet counts by computing z-scores from the averages and standard deviations for each outlet, which allows us to compare outlets as well as

⁴ http://en.wikipedia.org/wiki/New_York_Post

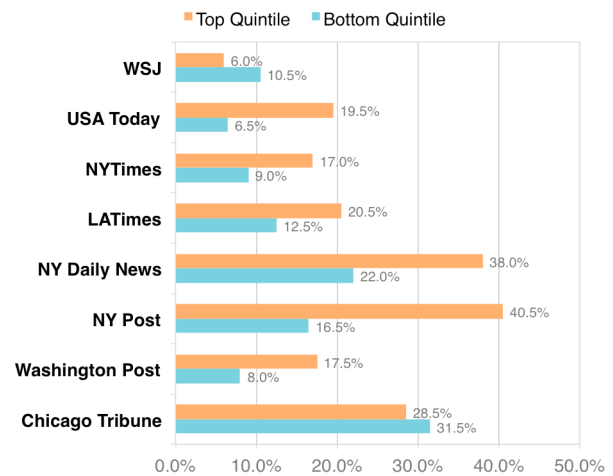


Figure 2. Percentage of deviant tweets across outlets and between top and bottom quintiles.

compute an aggregate statistical test across all outlets. Table 1 shows the Z statistic and significance values of each test. When we consider all outlets together we find a statistically significant difference, with deviant tweets garnering more retweets (mean z-score = 0.29) than non-deviant tweets (mean z-score = 0.13). When we look at each individual outlet we find that for six of the outlets deviant news tends to be retweeted significantly more. The exceptions include the Wall Street Journal (WSJ), which is significant in the opposite direction, and Chicago Tribune (CT), whose difference is not statistically significant.

To better understand the two outlets where network gatekeepers did not seem to find social deviance as important a newsworthiness criterion, we carefully read through all of the deviant headlines for those outlets. The deviant stories covered by WSJ included few breaking news items, opting more for investigative pieces about criminal issues or for international stories. Moreover, the WSJ has a specific focus on business, market, and economic news, the audience for which may be less interested in reading about and sharing socially deviant news, especially in comparison to some of the tabloid outlets we examined. At the CT the majority of top re-shared deviant stories included many shootings and crime-reports, with headlines often explicitly including numbers

Table 1. Wilcoxon Rank Sum tests showing statistical differences in # of retweets for deviant vs. non-deviant news.

News outlet	Z	p
Wall Street Journal	2.120	0.034
USA Today	-3.584	3.385x10 ⁻⁴
New York Times	-3.020	2.524x10 ⁻⁴
LA Times	-3.086	2.030x10 ⁻³
NY Daily News	-3.777	1.586x10 ⁻⁴
NY Post	-6.097	1.077x10 ⁻⁹
Washington Post	-4.144	3.410x10 ⁻⁵
Chicago Tribune	1.531	.126
Overall	-7.535	4.887x10 ⁻¹⁴

of dead or wounded. At the same time, many of the *least* re-shared deviant stories also included crime, but of lower salience, such as local robberies or abuses.

To help further explain the disparity in results for WSJ and CT, we computed the average z-score value for all words used in any deviant tweet. This allowed us to rank words used in deviant tweets by the extent to which they are re-shared and to see where there was a divergence between the topics that MSM was publishing and the topics that the network gatekeepers were most interested in re-sharing. We carefully read through that list looking both at the words in deviant tweets that were least, as well as most, retweeted. In the least retweeted list we found many words relating to court processes such as “accuser,” “lawyer,” “guilty,” “sues,” “charge,” “allegation,” “sentence,” and “jury” as well as some related to fatality such as “suicide,” “death,” and “slaying.” In the most-retweeted list were words relating to crime such as “police,” “cops,” “arrest,” “killed,” “shooting,” “attack,” as well as titillations such as “porn” and “sex”. We computed the fraction of deviant tweets that used any of the 25 different least retweeted words related to court activity or fatality and found that both WSJ and CT used those words at a higher rate than the other outlets ($\chi^2=98.79$, $DOF=1$, $p < 0.0001$). By emphasizing these topics in their coverage, WSJ and CT may not be getting as much re-sharing activity on their deviant news headlines.

Discussion and Conclusions

In this work we have studied the connection between social deviance and news headline re-sharing on Twitter. Our study contributes evidence that there is indeed a statistically significant inclination of network gatekeepers toward re-sharing news involving socially deviant events, just as for their professional counterparts. Our results indicate a solid overall relationship between social deviance and what network gatekeepers select for re-sharing: social deviance is generally correlated to more retweeting. Outlets particularly known for tabloid journalism exhibited an exaggerated effect, but results for WSJ and CT indicate other moderating factors at play.

Upon deeper investigation we found that for the WSJ and CT there was a focus on topics that, although deviant, were not highly re-shareable on Twitter. These results reinforce previous research, which has shown that professionals are more interested in public affairs topics, such as international stories or court proceedings, than audiences (Boczkowski and Mitchelstein 2013). By emphasizing these topics in their deviant tweets, it appears to have dampened the impact of deviance on re-sharing and helps explain the lack of statistical significance we found for those outlets.

In a broad analysis and synthesis of modern professional newsworthiness values, Harcup and O’Neill (Harcup and O’Neill 2001) identify the following list of ten factors:

reference to the power elite, reference to celebrities, entertainment, surprise, bad news, good news, magnitude (i.e. significance to a large number of people), cultural relevance to audience, follow-up, and newspaper agenda. In the same way that there are other dimensions of professional newsworthiness decisions besides deviance, there are other factors that also help explain what makes something newsworthy for a network gatekeeper. For instance, self-presentation effects and concern for your audience (boyd, Golder, and Lotan 2010), as well as motivations to share content such as humor which are unrelated to the newsworthiness of the event (André, Bernstein, and Luther 2012) could modulate whether a network gatekeeper chooses to re-share a headline.

Future work will strive to better understand the entire matrix of newsworthiness effects and the nuance of how and why network gatekeepers respond and may respond differently than professionals. While our study has focused on U.S. news outlets that release daily print copies, future work might also examine outlets from other countries, and other kinds of news media, such as digital native sites, news wires, or broadcast news outlets.

References

- André, P., Bernstein, M., & Luther, K. (2012). Who Gives A Tweet? Evaluating Microblog Content Value. In *Proc. CSCW*.
- Barzilai-Nahon, K. (2008). Towards a theory of network gatekeeping: A framework for exploring information control. *JASIST*, 59(9), 1493–1512.
- Berger, J., & Milkman, K. (2012). What makes online content viral? *Journal of Marketing Research*, 49(2), 192–205.
- Bhattacharya, D. (2012). Sharing News Articles Using 140 Characters: A Diffusion Analysis on Twitter. In *Proc. ASOAM* (pp. 966–971).
- Boczkowski, P., & Mitchelstein, E. (2013). *The news gap: When the information preferences of the media and the public diverge*. MIT Press.
- boyd, danah, Golder, S., & Lotan, G. (2010). Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter. In *Proc. HICSS-43*.
- Harcup, T., & O’Neill, D. (2001). What is News? Galtung and Ruge Revisited. *Journalism Studies*, 2(2), 261–280.
- Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? In *Proc. WWW*.
- McKirman, D. (1980). The identification of deviance: A conceptualization and initial test of a model of social norms. *European Journal of Social Psychology*, 10(1), 75–93.
- Messner, M., Linke, M., & Eford, A. (2011). Shoveling tweets: An analysis of the microblogging engagement of traditional news organizations. In *Proc. ISOJ*.
- Nahon, K., & Hemsley, J. (2013). *Going Viral*. Polity Press.
- Shoemaker, P. (2006). News and newsworthiness: A commentary. *Communications*, 31(1), 105–111.
- Shoemaker, P., Danielian, L., & Brendlinger, N. (1991). Deviant acts, risky business and US interests: The newsworthiness of world events. *Journalism & Mass Communication Quarterly*, 68(4), 781–795.
- Suh, B., Hong, L., Piroli, P., & Chi, E. (2010). Want to be retweeted? Large scale analytics on factors impacting retweet in Twitter network. In *Proc. of SocialCom* (pp. 177–184).