
Using Historical Twitter Data for Research: Ethical Challenges of Tweet Deletions

Jim Maddock

University of Washington
HCDE, iSchool, DUB,
maddock@uw.edu

Kate Starbird

University of Washington
HCDE, DUB
kstarbi@uw.edu

Robert Mason

University of Washington
iSchool, DUB
rmason@uw.edu

Paste the appropriate copyright/license statement here. ACM now supports three different publication options:

- **ACM copyright:** ACM holds the copyright on the work. This is the historical approach.
- **License:** The author(s) retain copyright, but ACM receives an exclusive publication license.
- **Open Access:** The author(s) wish to pay for the work to be open access. The additional fee must be paid to ACM.

This text field is large enough to hold the appropriate release statement assuming it is single-spaced in Verdana 7 point font. Please do not change the size of this text box.

Every submission will be assigned their own unique DOI string to be included here.

Abstract

This paper surfaces ethical concerns for social media researchers related to content deletion. We first provide a case study from our own work, which highlights a tradeoff between user rights and methodological consistency. We then offer partial anonymization as a possible answer, but acknowledge that better solutions may exist. Ultimately, we hope to engage the academic community to develop accepted practice for future work, and to contribute to the larger ethical debate surrounding social media research.

Author Keywords

Twitter; social media; ethical concerns; privacy; deletion

ACM Classification Keywords

H.5.3 [Information Interfaces & Presentation]: Groups & Organization Interfaces - Collaborative computing, Computer-supported cooperative work; K.4.2 Social Issues

Introduction

Twitter is becoming an increasingly powerful tool for social science research. Records of interactions left by its users provide researchers with large quantities of

trace data—in this cases tweets and their meta-data—which can be both qualitatively and quantitatively analyzed long after their creation. These data allow for the study of informational and social phenomena in new, potentially powerful ways.

Collecting and analyzing Twitter data, however, raises unique ethical questions. The Association of Internet Researchers (AoIR) outlines three “considerations” that encapsulate the debate: whether this research constitutes study of human subjects, how definitions of public and private information apply to internet data, and whether an avatar or profile is a person [4]. Previous work raises these ethical concerns through specific case studies—such as the T3 study’s leak of thousands of Facebook users’ profile data—and urges caution in future research [2][6][8].

Here we address a specific component of the social media ethics debate. Tweets are not necessarily permanent; like other social media platforms, Twitter allows users to delete their content. Previous work shows that many users believe their tweets to be inherently ephemeral [5], and complications therefore arise when researchers collect, store, and analyze these data more permanently [6]. These ethical considerations apply to both passive archiving—where the researcher treats both deleted and undeleted content equally—and active study of deleted content [1].

Through the following case study we intend to explore the broader ethical implications of tweet deletion within social media research. At the crux of this account is a trade-off we confronted between methodological validity and a user’s right to be forgotten. We offer our

current approach of selective anonymization as one possible answer; however, we acknowledge that this may not be the best solution. Ultimately we hope to engage the research community and motivate further discussion in a collective effort to establish best practices for future work.

Case Study

This case study emerges from ongoing research that attempts to interpret and quantify online rumoring behavior during crisis and disaster events. The larger research project encompasses dual goals: to better understand, describe, and model information propagation within a crisis context; and to automatically detect false rumors, first retroactively and later in real time [3][7].

Contemporaneous and Historical Datasets

Our analysis utilizes two Twitter datasets from the 2013 Boston Marathon Bombings. We collected the first—or *contemporaneous*—dataset with the Twitter Streaming API, filtering on the key words “boston”, “bomb”, “marathon”, “explosion”, and “blast”. The collection ran from 5:25pm EDT on April 15, 2013 until April 22 at 3:05pm EDT and produced a corpus of 10,621,415 tweets. Notably, the dataset shows periods of inconsistent collection, due to both rate limiting (from Twitter) and technical issues (from our collection scripts).

We purchased the second dataset from GNIP, a subsidiary of Twitter, in an effort to mitigate methodological issues created by data gaps in our collection. GNIP sells complete *historical* collections, which are not subject to rate limiting or technical problems. While these collections are often

unobtainable to researchers due to cost, they theoretically eliminate biases present in a similar Streaming API collection. Our historical collection consisted of tweets from the same time period, filtered on the same key words in order to maximize consistency between the two datasets, which resulted in a corpus of 23,701,467 tweets.

Missing Tweets

Calculating the overlap between the two datasets using unique tweet IDs revealed that our historical collection contains 14,432,153 tweets—or 61% of its total volume—that do not exist in its contemporaneous counterpart. This makes sense; rate limiting and our own technical limitations would explain the discrepancy.

More surprising, however, is the absence of tweets from the historical collection that exist within the contemporaneous collection. The contemporaneous collection contains 1,351,643 tweets—or roughly 13% of its total volume—that do not appear in the historical collection. In other words, the collection we purchased through GNIP was missing approximately 13% of the tweets that we originally captured in our real-time, *contemporaneous* collection.

While the percentage of missing tweets in this direction is substantially smaller, the historical collection theoretically suffers none of the limitations of its contemporaneous counterpart, and therefore completely captures *all* tweets over the given time interval using a given set of key words. That *any* tweets are missing reveals unexpected limitations of the dataset.

We extrapolate that these missing tweets likely stem from three different, though related, sources: tweets that have been deleted by their author, timelines that have been made private, and accounts that have been deleted or suspended. These deletions intersect in meaningful ways with rumoring behavior—tweets which passed along false rumors were more likely to be missing in the *historical* set (and likely deleted) than other tweets we collected. For one of the rumors we identified and coded in the *contemporaneous* set, more than 50% of the tweets were missing in corresponding data from the *historical* set. Additionally, tweets that corrected false rumors were less likely to be missing tweets than those that affirmed false rumors. Upon reflection, it would not be surprising that users who passed along misinformation were deleting those tweets at a higher rate than typical tweet content.

Our conclusion that this missing data was likely due to tweet deletions and account suspensions aligns with Twitter policy; in its terms of service (TOS), Twitter requires those collecting data to remove all tweets that have been deleted by their owner—i.e. tweets that fall under any of these three categories. GNIP must follow developer guidelines before reselling Twitter data, and would therefore scrub these deleted or private tweets from its historical collections. Ultimately, the historical collection would be “updated” to reflect these removals at the time of sale.

Emergent Legal and Ethical Issues

Removal of deleted tweets creates methodological concerns—for example, it may be hard to accurately characterize rumoring behavior by examining datasets where deleted tweets are missing. However, in this paper we surface both the *legal* and *ethical* concerns

that became evident during our study. We raise these concerns because we believe they extend beyond our work, relevant not only to analysis of historical collections, but to *all* Twitter research. They are as follows:

Legal Obligation to Remove Deletions

Twitter requires developers to remove deleted tweets from any collection, should those deletions become apparent. In describing status deletion notices that are sent through the Streaming API for some types of searches, the API documentation states specifically:

Status Deletion Notices (Delete)

These messages indicate that a given Tweet has been deleted. Client code must honor these messages by clearing the referenced Tweet from memory and any storage or archive, even in the rare case where a deletion message arrives earlier in the stream than the Tweet it references.

Account holders can remove their content at any time. Legally, individuals and organizations who collect Twitter data are therefore required to remove deleted tweets *when they become aware of the deletions* (although there is an important loophole, which we will address later). Though these rules were likely established for the client applications that use Twitter's APIs, they apply to researchers as well. In our case study, awareness of deletion through comparison of our two datasets obligates us to remove these tweets from our contemporaneous collection.

The Ethics of Researching Deleted Tweets

Beyond our legal accountability, there exists some ethical ambiguity regarding obligation to remove

deleted content from our collections. Twitter research does not require human subjects approval at many universities, including the University of Washington where this research took place. This is because social media data is published, public content—though researchers debate these characterizations—and consent is therefore implied.

Yet the related acts of deleting or making tweets private show a deliberate intent to remove content from public view. In traditional human subjects research, a participant's withdrawal requires researchers to remove this material—for example, an interview or a survey response—from the study. Though in our case study we were not explicitly contacted by any Twitter account holders, it would be difficult to argue that deletion does not imply a request to withdraw. We must therefore consider whether awareness of deliberate content removal thereby obligates us to scrub these tweets from our collection for ethical reasons.

Practical Barriers of Collection Curation

Though there are legal and ethical factors suggesting that researchers *should* remove deleted content from research collections, whether researchers *can* identify and remove deleted content—due to technical and methodological constraints—complicates the issue.

Technical

Identifying deletions within large datasets is technically complex and resource-heavy. Manual analysis is not scalable to any degree, and is therefore largely irrelevant here. Similarly, harnessing the API to look up individual tweet IDs on large datasets requires

considerable computational resources and is likely to run up against rate limits imposed by Twitter.

Overlap calculations between contemporaneous and historical datasets require two datasets with identical search parameters, and obtaining historical collections is prohibitively expensive for many researchers. Additionally, this calculation is only valid for the time at which the dataset is purchased; subsequent deletions would still exist within the historical data.

Twitter produces Status Deletion Notices through specific User Streams within the Streaming API, which could provide a computational alternative, however these deletion notices are difficult to collect. Collections based on key words—such as ours—do not receive Status Deletion Notices, and the computational overhead of following millions of users is prohibitively large. Notably, Twitter offers only a partial solution. A comment from a Twitter Product Manager on a Q&A site for Streaming API users explains the obligations for those collecting and analyzing data under these conditions:

You are responsible for removing and respecting deleted tweets that you are made aware of. When not using a streaming API like user or site streams, you're not streamed the events when a user deletes a tweet and it's more difficult to be aware of them. If you become aware of a deleted tweet through some other means (whether programmatic, socially, in an article, recrawling, etc.) then you'd want to honor that decision.

In other words, researchers are required to remove tweets from their own collections only if they become

aware of the deletion. Our collection methods—and we suspect those of other researchers—follow only keyword streams, which do not provide deletion notices. We cannot know which tweets have been subsequently deleted, and we therefore remain technically within the TOS. In other words, we are not obligated to delete removed tweets only because we do not confirm that they remain public.

Methodological

Curation in accordance with the TOS also creates methodological limitations, which could hinder future research. Deletion patterns are likely inconsistent, and therefore produce an unpredictably biased collection. Furthermore, the data—and the resulting bias—would constantly change over time as users delete their tweets and accounts weeks, months, or years after the collection period. Even if researchers could honor deletion requests, doing so would, in some cases, invalidate the collection as an accurate representation of the information space at the time.

Selective Anonymization and Call for Discussion

We recognize that our case study illustrates a specific but rare instance in which researchers became aware of *all* deleted tweets. Many researchers do not have the capacity—either financially or computationally—to identify deletions within their collections. We hope, however, to use this as a starting point for conversation surrounding issues of tweet deletion. The current strategy—either plausibly deniability or blissful ignorance—should be addressed. Though Twitter seems to passively encourage this approach, we hope researchers will actively engage the issue to develop a

better solution that both facilitates future research and respects users' deliberate intent to remove content.

Selective anonymization is one such approach that we employ. We do not actively remove deleted tweets, and we use these tweets indiscriminately in our quantitative analysis so as not to introduce unpredictable bias. Because deleted tweets are never explicitly cited and only contribute to statistical metrics, we can maintain a high level of anonymity. Alternatively, before using specific tweets in qualitative analysis we check to make sure they are still publically available. This subset is limited, and we can therefore check their status manually.

Our mixed approach, though temporarily functional in the absence of other standards, has clear limitations. It does not completely honor our subjects' intention to remove data, it does not address legal issues, and it arguably favors quantitative research techniques over their qualitative counterparts. We therefore open this discussion to the wider research community in hopes of developing a better solution.

Acknowledgements

This research is a collaboration between the emCOMP Lab and SoMe Lab at UW and was partially supported by NSF Grants 1342252 and 1243170. The authors thank Shawn Walker for his efforts to transform the original Twitter data into datasets that were more easily accessed and analyzed.

References

- [1] Almuhimedi H., Wilson S., Liu B., Sadeh N., Acquisti A. (2013). Tweets are forever: A large-scale quantitative analysis of deleted tweets. *Proc. ACM Conf. Comput. Support. Coop. Work CSCW Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW, (2013)*, 897–907.
- [2] boyd, d., and Crawford, K. (2011). Six Provocations for Big Data. *A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society*, September 2011.
- [3] Maddock, J., Starbird, K., Al-Hassani, H., Sandoval, D., Orand, M., & Mason, R.M. (Accepted). Characterizing Online Rumoring Behavior Using Multi-Dimensional Signatures. *To appear in CSCW 2015*.
- [4] Markham, A. and Buchanan, E. *Ethical decision-making and Internet research: Recommendations from the AoIR ethics working committee (version 2.0)*. Association of Internet Researchers, 2012.
- [5] Proferes, N. (2014). What Happens to Tweets? Descriptions of Temporality in Twitter's Organizational Rhetoric. *In iConference 2014 Proceedings (p. 76 - 87)*.
- [6] Small, H., Kasianovitz, K., Blanford, R., and Celaya, I. What Your Tweets Tell Us About You: Identity, Ownership and Privacy of Twitter Data. *IJDC International Journal of Digital Curation 7, 1 (2012)*, 174–197.
- [7] Starbird, K., Maddock, J., Orand, M., Achterman, P., & Mason, R.M. (2014). Rumors, False Flags, and Digital Vigilantes: Misinformation on Twitter after the 2013 Boston Marathon Bombings. Short paper. *iConference 2014*.
- [8] Zimmer, M. But the data is already public: on the ethics of research in Facebook. *Ethics Inf Technol Ethics and Information Technology 12, 4 (2010)*, 313–325.