"THIS TWEET IS UNAVAILABLE": #BLACKLIVESMATTER TWEETS DECAY

Yiran Duan
Syracuse University

Jeff Hemsley
Syracuse University

Alexander Owen Smith
Syracuse University

Introduction

By virtue of its user base, which skews towards those who disproportionately affects news coverage (Chadwick, 2017; Himelboim et al., 2013), Twitter plays an important role in setting the public agenda. However, Twitter is a highly dynamic space that evolves over time, where users' norms impact Twitter's environment and functionality (Halavais, 2013). Another way that Twitter is dynamic results from information decay (Walker, 2017), which happens when users delete their tweets, close or make their accounts private, or Twitter suspends or deletes an account. While a few studies have looked at Twitter information decay (Almuhimedi et al., 2013; Bastos, 2021; Bhattacharya & Ganguly, 2016; McCammon, 2022; Zubiaga, 2018) most have tended to focus on a limited set of topic spaces, or were before Musk took over Twitter, which has dramatically altered the distribution of users (Safak, 2022). As researchers grapple with figuring out how to do research in this evolving space, understanding how information decay may impact research seems key.

This preliminary work adds to the existing literature in two keyways. First, we broaden the kinds of discussion spaces where information decay has been studied by focusing on #BlackLivesMatter (hereafter, BLM), which in the U.S. has become a highly politicized movement (Stewart et al., 2018). Second, to our best knowledge, this is the first study looking at social movement tweets since Musk took over and it may highlight how the landscape on Twitter has shifted.

Background

Previous studies show that random tweet collections that include more than one hashtag had relatively low rate of unavailable tweets while political related datasets have a higher rate of tweet decay (Almuhimedi et al., 2013; Bastos, 2021; Bhattacharya & Ganguly, 2016; McCammon, 2022; Zubiaga, 2018). For example, Zubiaga's (2018) study indicates that 81.4% of their 30 randomly selected real-world events datasets remained available after 4 years, while Bastos' (2021) study shows that only 67% of Brexit debate related tweets were available after the same amount of time. With Musk's takeover, and reinstatement of many accounts previously shut down for harassment (Lorenz, 2022), some of these numbers may look different today.

We take up this speculation by examining a large collection of #BlackLivesMatter data at three different time points. Pew polling data shows that roughly 42% of adults in the U.S. held a negative view of the BLM movement (Hurst, 2022). Research has also suggested that the Twitter discussion space around BLM is highly polarized (Stewart et al., 2018), therefore we expect to find a low rate of decay since some accounts that were negative towards BLM could have been reinstated on the grounds of free speech after Musk's takeover. We also speculate that highly publicized events such like BLM might have a different rate of decay since they might have a larger audience.

Given above, we propose the following two research questions:
RQ1: *how does the percentage of unavailable BLM tweets compare to the numbers published by previous research*?
RQ2: *is the rate of decay substantially different for a highly publicized, BLM related event, to a more typical rate for BLM?*

The key event we chose is Derek Chauvin's trial, which took place on April 20th, 2021. Chauvin, a former police officer, murdered Black American George Floyd on May 25th, 2020, which caused world-wide protests for police brutality and unfair treatments of Black Americans.

Methods

Using Twitter's academic API, we have been collecting tweets with the #BlackLivesMatter hashtag since July 2020. This API collects tweets as soon as they are posted, so even if someone deletes their tweet right after posting it, we have a copy in our dataset. To answer our research questions, we pulled tweets from our collection for three time periods: February 17th to February 23rd of 2021, April 17th - April 23rd, 2021, and February 17th to February 23rd, 2022. Thus, we have tweets from three, one week block of time.

The April block captures tweets about three days before and after Derek Chauvin's trial. The other two blocks offer data before and after the trial within Black History Month. Then, we used the rehydratoR (Coakley & Steinert-Threlkeld, 2019) package to check the availability of our tweets and generated a list of unavailable ones. We then ran the academicR package (Barrie et al., 2022) on the unavailable tweets. This package provides error messages for unavailable tweets that helps us understand if the tweet

was deleted or the account has been marked private or was suspended. Note that no package that we know of provides more detail on what a tweet might be unavailable.

Having the results, we answer the first question by finding the average missing tweet rates across the three date ranges and comparing it to the published rates. With the results we provide some observations given Musk's take over and some of the sparse literature since then. To answer the second question, we compare the results during Derek Chauvin's trial with the results from before and after.

Preliminary Findings

Table 1 shows the information decay at the three time periods.

Timeline	Available tweets	Unavailable tweets		Total
		deleted	suspended/private account	
February 17th -23rd, 2021	166,926 (58.54%)	67,268 (23.59%)	50,940 (17.87%)	285,134 (100%)
April 17th - 23rd, 2021 (Chauvin trial)	747,087 (60.10%)	303,417 (24.41%)	192,656 (15.50%)	1,243,160 (100%)
February 17th -23rd, 2022	142,937 (70.79%)	29,165 (14.44%)	29,820 (14.77%)	201,922 (100%)
Average	63.14%	20.81%	16.05%	

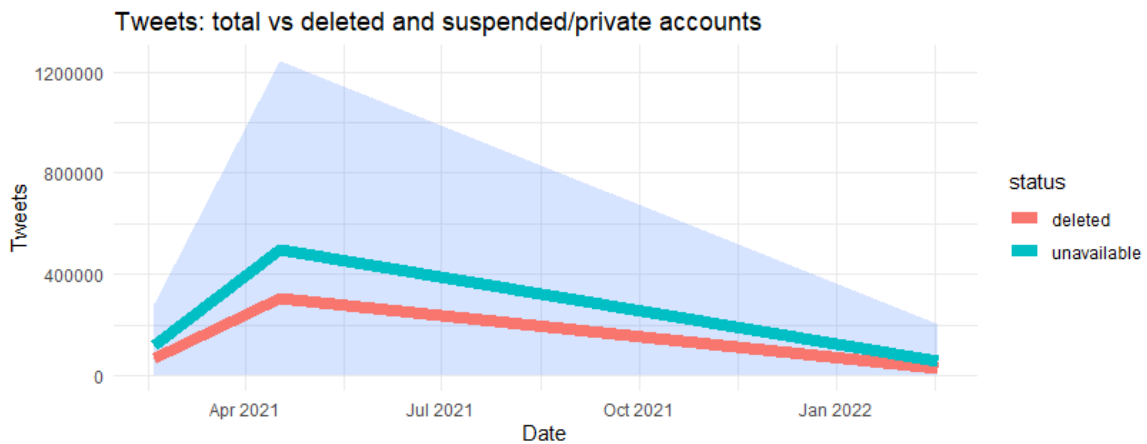Table 1. summary of three time periods

Figure 1. Visualization of three time periods

As table 1 shows, on average, 37% of the tweets in our data collection became unavailable within 2 years. Compared to previous studies, such as Zubiaga (2018)'s randomly selected real-world events datasets (19% information decay after 4 years) and Bastos (2021) Brexit debate (33% information decay after 4 years), the BLM information decay is 4 to 18% higher. This appears to be a substantially higher rate of decay than previous literature, which answers RQ1. We also note that while the most recent dataset has the lowest decay rate, the two older datasets have similar rates, suggesting

that an event like Chauvin's trial seems not to influence decay rate; the unavailable tweets rate increases as more time passes, which answers RQ2. The noticeably higher decay rates suggest that either more decay happens in the BLM discussion space than others, or that since Twitter was taken over by Musk, more information decay has taken place. Future work could compare the BLM discussion space with other social movements to conduct further analysis.

One limitation of this study is that we do not know when exactly a tweet or an account became unavailable, but this limitation does not impact us comparing the percentage of unavailable tweets in our dataset with previous studies'. Another limitation is that although some studies show that users delete their tweets or close their accounts due to regrets (McCammon, 2022; Murillo et al., 2018; Petrovic et al., 2013), we cannot say for sure why some tweets became unavailable in our dataset. However, this limitation does not affect our analysis since this study focuses on the percentage of unavailable tweets. We plan on running sentiment analysis on the unavailable tweets for future studies since the original dataset captured the original message of a tweet.

Conclusion

This preliminary study shows that tweets from the #BlackLivesMatter discussion space have decayed at a higher rate than other discussion spaces from previous studies. More work needs to be done to understand if this a byproduct of Musk taking over Twitter, or if something is inherently different about the BLM discussion space that results in higher rates of decay.

References
Almuhimedi, H., Wilson, S., Liu, B., Sadeh, N., & Acquisti, A. (2013). Tweets are forever: A large-scale quantitative analysis of deleted tweets. *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, 897–908. https://doi.org/10.1145/2441776.2441878
Barrie, C., Ho, J. C., Chan, C., Rico, N., König, T., & Davidson, T. (2022). *academictwitteR: Access the Twitter Academic Research Product Track V2 API Endpoint* (0.3.1). https://CRAN.R-project.org/package=academictwitteR
Bastos, M. (2021). This Account Doesn't Exist: Tweet Decay and the Politics of Deletion in the Brexit Debate. *American Behavioral Scientist*, 65(5), 757–773. https://doi.org/10.1177/0002764221989772
Bhattacharya, P., & Ganguly, N. (2016). Characterizing Deleted Tweets and Their Authors. *Proceedings of the International AAAI Conference on Web and Social Media*, 10(1), Article 1. https://doi.org/10.1609/icwsm.v10i1.14803
Chadwick, A. (2017). *The Hybrid Media System: Politics and Power*. Oxford University Press.
Coakley, K., & Steinert-Threlkeld, Z. (2019). *rehydratoR: Downloads Tweets from a List of Tweet IDs* (0.5.2). https://CRAN.R-project.org/package=rehydratoR
Halavais, A. (2013). Structure of Twitter: Social and Technical. In *Twitter and society* (1th ed., Vol. 89, pp. 29–42). Peter Lang.
Himelboim, I., Hansen, D., & Bowser, A. (2013). Playing in the same twitter network. *Information, Communication & Society*, 16(9), 1373–1396. https://doi.org/10.1080/1369118X.2012.706316

Hurst, K. (2022). U.S. teens are more likely than adults to support the Black Lives Matter movement. *Pew Research Center*. https://www.pewresearch.org/fact-tank/2022/06/15/u-s-teens-are-more-likely-than-adults-to-support-the-black-lives-matter-movement/

Lorenz, T. (2022, November 30). 'Opening the gates of hell': Musk says he will revive banned accounts. *Washington Post*. https://www.washingtonpost.com/technology/2022/11/24/twitter-musk-reverses-suspensions/

McCammon, M. (2022). Tweeted, deleted: An exploratory study of the US government's digital memory holes. *New Media & Society*, *24*(3), 741–759. https://doi.org/10.1177/1461444820934034

Murillo, A., Kramm, A., Schnorf, S., & De Luca, A. (2018). "If I press delete, it's gone": User understanding of online data deletion and expiration. *Proceedings of the Fourteenth USENIX Conference on Usable Privacy and Security*, 329–339.

Petrovic, S., Osborne, M., & Lavrenko, V. (2013). *I Wish I Didn't Say That! Analyzing and Predicting Deleted Messages in Twitter* (arXiv:1305.3107). arXiv. http://arxiv.org/abs/1305.3107

Safak, V. (2022). *Elon Musk's Twitter Takeover: Politician Accounts*. https://doi.org/10.48550/arXiv.2205.08491

Stewart, L. G., Arif, A., & Starbird, K. (2018). *Examining Trolls and Polarization with a Retweet Network*. 6.

Walker, S. (2017). *The Complexity of Collecting Digital and Social Media Data in Ephemeral Contexts* [Thesis]. https://digital.lib.washington.edu:443/researchworks/handle/1773/40612

Zubiaga, A. (2018). A longitudinal assessment of the persistence of twitter datasets. *Journal of the Association for Information Science and Technology*, *69*(8), 974–984. https://doi.org/10.1002/asi.24026