Online Appendix

Non-random Tweet Mortality and Data Access Restrictions: Compromising the Replication of Sensitive Twitter Studies

Andreas Küpfer[1]

[1]Technical University of Darmstadt

February 13, 2024

Appendix 1. Paper Study: Detailed Information

To determine whether a paper analyzes Twitter data, I first looked for relevant methods or dataset descriptions in the paper itself or its appendix. For the remaining portion of the papers, I automatically downloaded all publicly available official replication repositories. This lets me identify if someone shared their dataset as tweet IDs by searching for the following pattern:

```
(^|,|;|"|'| )(\d{8,})($|,|;|"|'| )
```

This regular expression looks for numeric sequences of at least eight numbers embedded in typical delimiters such as spaces, quotes, or semicolons and is applied on all files which are either text-files or readable by the R package `rio` (Chan et al. 2021). It turned out that this approach acts as a good heuristic to solve this task. I then manually checked all downloaded replication archives again, looking for false positives. After that, another identifier is added, which represents whether the replication archive contains datasets with raw textual tweet content data.

Table A.1 lists the papers that systematically analyze the content of tweets. The columns Tweet IDs and Tweet Content contain a ✓ if they share data for the respective categories and their replication archives. Sensitive content is marked with a ✓ for papers that explicitly analyze the content of tweets in datasets belonging to at least one of the following three areas (inspired by Elmas (2023)):

- Fake News/Disinformation
- Hate Speech/Violence/Terrorism
- Bots

In many cases, these categories are strongly connected with each other (e.g., bots often share fake news, or terrorism is strongly related to disinformation). My definition of sensitive datasets does not mark a dataset of tweets by Donald Trump as being sensitive, even if one would expect a few tweets containing fake news and disinformation, as most of these tweets is general political content.

Table A.1. Papers from seven political science journals mentioning the keyword "Twitter". The columns `Tweet IDs` and `Tweet content` show whether the authors shared the respective data in their replication archive. `Is sensitive content?` marks if a paper analyzes sensitive Twitter datasets.

Paper	Journal	Tweet IDs	Tweet content	Is sensitive content?
Beauchamp (2017)	AJPS	-	-	-
Benton and Philips (2020)	AJPS	-	-	-
Fong and Grimmer (2021)	AJPS	-	-	-
King, Lam, and Roberts (2017)	AJPS	✓	✓	-
Nielsen (2020)	AJPS	✓	-	-
Alrababah et al. (2019)	APSR	✓	-	✓
Barberá et al. (2019)	APSR	-	-	-
Mitts (2019)	APSR	-	-	✓
Osmundsen et al. (2021)	APSR	-	-	✓
Pan and Siegel (2020)	APSR	-	-	✓
Silva and Proksch (2021)	APSR	✓	-	-
Sobolev et al. (2020)	APSR	-	-	-
Stukal et al. (2022)	APSR	✓	-	✓
Brie and Dufresne (2020)	BJPS	✓	✓	-
Clarke and Kocak (2020)	BJPS	✓	-	-
Jones and Mattiacci (2019)	BJPS	-	-	-
Munger et al. (2022)	BJPS	-	✓	-
Bisbee and Lee (2022)	JOP	-	-	-
Boucher and Thies (2019)	JOP	-	-	-
Das et al. (2022)	JOP	-	-	-
Mitts, Phillips, and Walter (2022)	JOP	-	-	✓
Skytte (2022)	JOP	-	-	-
Bestvater and Monroe (2022)	PA	-	✓	-

Kubinec and Owen (2021)	PA	-	✓	-
Miller, Linder, and Mebane (2020)	PA	✓	✓	-
Temporão et al. (2018)	PA	✓	-	-
Castanho Silva, Proksch, et al. (2022)	PSRM	✓	-	-
Cirone and Hobbs (2023)	PSRM	✓	✓	✓
Kim (2023)	PSRM	✓	-	✓
Muchlinski et al. (2021)	PSRM	✓	-	✓
Munger et al. (2019)	PSRM	-	-	-
Settle et al. (2016)	PSRM	-	-	-
Bradshaw et al. (2020)	PolComm	-	-	✓
Cassell (2021)	PolComm	-	-	-
DiResta, Grossman, and Siegel (2022)	PolComm	-	-	✓
Gilardi et al. (2022)	PolComm	✓	-	-
Guess et al. (2019)	PolComm	-	-	-
Kang et al. (2018)	PolComm	-	-	-
Keller and Klinger (2019)	PolComm	-	-	✓
Keller et al. (2020)	PolComm	-	-	✓
Ketelaars and Sevenans (2021)	PolComm	✓	✓	-
Kligler-Vilenchik et al. (2021)	PolComm	-	-	-
Kobayashi and Ichifuji (2015)	PolComm	-	-	-
Konitzer et al. (2019)	PolComm	-	-	-
Linvill and Warren (2020)	PolComm	-	-	✓
Margolin, Hannak, and Weber (2018)	PolComm	✓	✓	✓
Muddiman, McGregor, and Stroud (2019)	PolComm	-	-	-
Popa et al. (2020)	PolComm	-	-	-
Stier et al. (2018)	PolComm	-	-	-
Yarchi, Baden, and Kligler-Vilenchik (2021)	PolComm	✓	✓	-

Appendix 1.1 Results per Journal

Figure A.1 depicts the proportions of each Twitter-related replication category grouped by the respective journal. It highlights that there are a few journals where no Twitter replication data exists for more than half of their Twitter-related papers. Authors of Twitter research in AJPS and PA either share their Tweet IDs, or the content of the analyzed tweets, or even both. On the contrary, Twitter papers published in JOP do not contain any Twitter-related replication data.

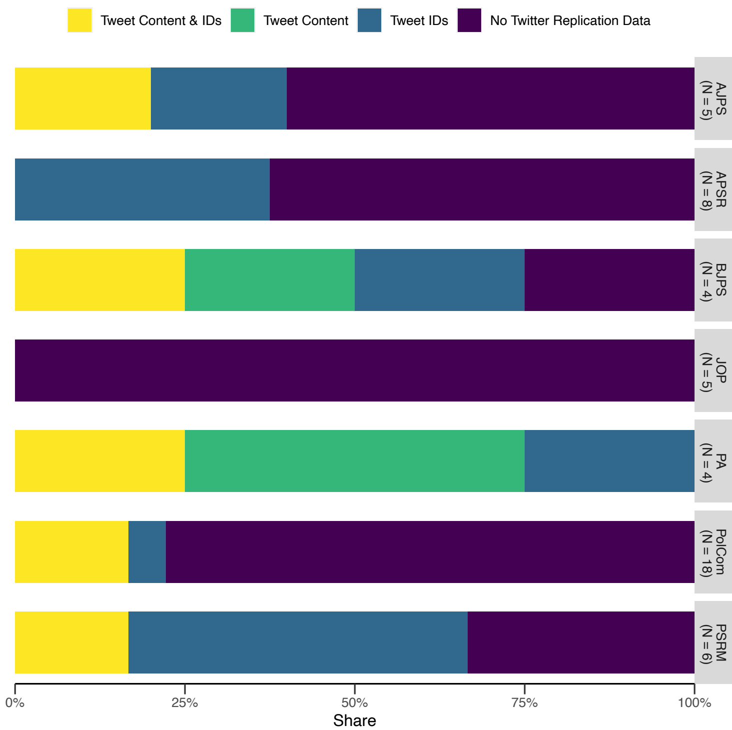

Figure A.1. Different methods of publishing Twitter datasets in papers, categorized by the journal.

Appendix 2. Kim (2023) Data Leveraging Pipeline and Replication Study

Choosing to replicate Kim (2023) in favor of other potential replication candidates is based on valid grounds. While Alrababah et al. (2019) and Muchlinski et al. (2021) do not provide much further replication data related to tweet content beyond code and tweet IDs, Stukal et al. (2022) and Cirone and Hobbs (2023) do not focus on the comparison of sensitive and non-sensitive studies. Finally, Margolin, Hannak, and Weber (2018) do not study longitudinal aspects of their dataset.

Kim's study examines violent political rhetoric on social media and its relationship with offline political violence, focusing on the Capitol Riot. The author introduces a new automated method to identify violent rhetoric on Twitter and finds that users who engage in such rhetoric are ideologically extreme and located on the fringe of the communication network. The tweets are more frequently targeted at women and Republican politicians and are often shared across the ideological divide, creating the potential for co-radicalization.

The database for these findings grounds a random proportion of 1% of all tweets in real-time by taking advantage of the Streaming API by Twitter. These tweets are then processed in a pipeline containing several keyword-based filtering approaches and finally classified as holding political violence or not, using a transformer model.

Appendix 2.1 Reasons for Unavailable Tweets

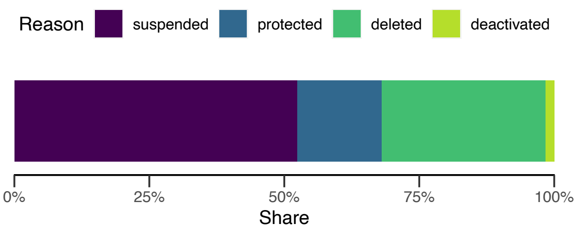

Figure A.2. Reasons for unavailable users in Kim (2023) that posted violent tweets according to Twitters' Compliance API endpoint. Results are weighted with the number of tweets each user posted according to aggregated information in the replication files of Kim (2023). The purple area shows the share of suspended users due to platform actions, while the other sections (visualized in brighter colors) highlight reasons for unavailable users due to explicit user actions.

While A.2 depicts the reasons for removed tweets based on removed user accounts weighted by their total number of tweets, there are also tweet removals without the complete user becoming unavailable. However, according to the Compliance API, only 10,425 tweets were removed explicitly, which is 5.78% of all unavailable tweets. This means that 169,946 (94.22%) tweets were removed due to account suspensions, deletions, protections, or deactivations.

Appendix 2.2 Representativity of Recrawled Content

Table A.2. Results of Welch's t-test comparing different features between a sample of 5000 violent political rhetoric tweets of the original population (published by the author along with the replication files) and the recrawled dataset. A 95% confidence interval excluding zero is an indicator that a feature is different in both datasets.

Type	95% confidence interval
Textual Content	[−3.94; −1.31]
Hashtag	[−1.72; −1.17]
User Mentions	[−5.44; 0.96]

Appendix 2.3 Hashtag Frequency

Table A.3. Hashtag frequency comparison of the original dataset and recrawl. Arrows indicate the direction of rank change, dashes show no difference, and question marks reflect that a hashtag is not available anymore. The different colors depict the intensity (red = considerable change; green = slight change).

	Count		Rank		
Hashtag	Original	Recrawl	Original		Recrawl
#wethepeople	1511	8	1	⇓	24
#1	1398	-	2	?	
#pencecard	1341	3	3	⇓	29
#maga	881	55	4	–	4
#fightback	702	1	5	⇓	31
#1776again	672	-	6	?	
#antifaarefascists	607	2	7	⇓	30
#blmareracists	607	1	7	⇓	31
#covid19	606	83	8	⇑	1
#treason	555	18	9	⇓	14
#vote	498	13	10	⇓	19
#trump	452	26	11	⇑	9
#trump2020	434	42	12	⇑	5
#walterreed	428	62	13	⇑	3
#savebrandonbernard	421	78	14	⇑	2
#pardonsnowden	365	1	15	⇓	31
#traitortrump	358	15	16	⇓	17
#freeassange	356	-	17	?	
#punkaf	354	-	18	?	
#godwins	244	1	19	⇓	31

Appendix 2.4 Fightin' Words Algorithm

Fightin' Words (Monroe, Colaresi, and Quinn 2008) is a lexical feature selection algorithm that helps in determining which terms are most distinctively characteristic of a particular textual group's (sensitive versus non-sensitive tweets) language usage. The calculation of the word importance yielded by the algorithm is implemented as follows:

$$\hat{w} = normalize(normalize(log(g\_dtm))')' \tag{1a}$$

In 1a, g_dtm describes the document-term frequency per group. Log transformation and normalization[19] of these frequencies leads to $\hat{w}$, which builds the basis for the final group-wise word importance.

$$w\_se = \sqrt{\frac{1}{g\_dtm} + \frac{1}{g\_dtm\_w} + \frac{1}{g\_dtm\_k} + \frac{1}{g\_dtm\_kw}} \tag{1b}$$

Equation 1b calculates the standard error for each word w per group. The suffix $\_w$ is about the usage of other terms in the same group k, whereas the suffix $\_k$ describes the frequency of the

19. *normalization* is applied two times to normalize both within-group and across-groups.

current word w spoken by groups other than k. Finally, $\_kw$ is the total number of words not spoken by group k other than the specific word w.

$$\hat{w}_{zeta} = \frac{\hat{w}}{w\_se} \tag{1c}$$

Finally, $\hat{w}$ is divided by the corresponding standard error. The resulting zeta scores in $\hat{w}_{zeta}$ represent how distinctive a term is for a particular group. Figure 5 leverages these scores per group and word. It contains very dense information about the group-dependent relative frequency of each keyword on the x-axis. At the same time, the y-axis (and the size of a specific word) displays the extent to which a keyword is associated with a group.

Appendix 2.5 Regression Models

Table A.4. Negative binomial regression models (original dataset)

	Model 1	Model 2	Model 3	Model 4	Model 5
(Intercept)	4.02	5.00	4.47	−8.79	−9.44
	(0.10)	(0.10)	(0.12)	(0.52)	(0.59)
Position: Governors	1.08				0.51
	(0.30)				(0.22)
Position: Senators	2.15				0.18
	(0.23)				(0.18)
Female		−0.38			0.97
		(0.21)			(0.15)
Republican			0.78		0.99
			(0.18)		(0.13)
Follower Count (log)				2.57	2.52
				(0.11)	(0.13)
AIC	5255.01	5364.26	5348.76	4689.54	4636.13
Num. obs.	585	585	585	562	562

Table A.5. Negative binomial regression models (recrawled dataset)

	Model 1	Model 2	Model 3	Model 4	Model 5
(Intercept)	2.47	2.87	2.82	−7.67	−8.69
	(0.11)	(0.09)	(0.12)	(0.49)	(0.53)
Position: Governors	0.56				0.27
	(0.25)				(0.16)
Position: Senators	0.96				0.02
	(0.19)				(0.14)
Female		0.10			0.02
		(0.20)			(0.14)
Republican			0.14		0.90
			(0.17)		(0.12)
Follower Count (log)				1.96	2.05
				(0.10)	(0.10)
AIC	2380.92	2406.76	2406.32	2017.15	1966.65
Num. obs.	328	328	328	322	322

References

Alrababah, A., W. Marble, S. Mousa, and A. Siegel. 2019. Can exposure to celebrities reduce prejudice? the effect of mohamed salah on islamophobic behaviors and attitudes. *American Political Science Review.*

Barberá, P., A. Casas, J. Nagler, P. J. Egan, R. Bonneau, J. T. Jost, and J. A. Tucker. 2019. Who leads? who follows? measuring issue attention and agenda setting by legislators and the mass public using social media data. *American Political Science Review* 113 (4): 883–901.

Beauchamp, N. 2017. Predicting and interpolating state-level polls using twitter textual data. *American Journal of Political Science* 61 (2): 490–503.

Benton, A. L., and A. Q. Philips. 2020. Does the realdonaldtrump really matter to financial markets? *American Journal of Political Science* 64 (1): 169–190.

Bestvater, S. E., and B. L. Monroe. 2022. Sentiment is not stance: target-aware opinion classification for political text analysis. *Political Analysis,* 1–22.

Bisbee, J., and D. D. I. Lee. 2022. Objective facts and elite cues: partisan responses to covid-19. *The Journal of Politics* 84 (3): 1278–1291.

Boucher, J.-C., and C. G. Thies. 2019. "i am a tariff man": the power of populist foreign policy rhetoric under president trump. *The Journal of Politics* 81 (2): 712–722.

Bradshaw, S., P. N. Howard, B. Kollanyi, and L.-M. Neudert. 2020. Sourcing and automation of political news and information over social media in the united states, 2016-2018. *Political Communication* 37 (2): 173–193.

Brie, E., and Y. Dufresne. 2020. Tones from a narrowing race: polling and online political communication during the 2014 scottish referendum campaign. *British Journal of Political Science* 50 (2): 497–509. https://doi.org/10.1017/S0007123417000606.

Cassell, K. J. 2021. When "following" the leader inspires action: individuals' receptivity to discursive frame elements on social media. *Political Communication* 38 (5): 581–603.

Castanho Silva, B., S.-O. Proksch, et al. 2022. Politicians unleashed? political communication on twitter and in parliament in western europe. *Political Science Research and Methods* 10 (4): 776–792.

Chan, C.-h., G. C. Chan, T. J. Leeper, and J. Becker. 2021. *Rio: a swiss-army knife for data file i/o.* R package version 0.5.29.

Cirone, A., and W. Hobbs. 2023. Asymmetric flooding as a tool for foreign influence on social media. *Political Science Research and Methods* 11 (1): 160–171.

Clarke, K., and K. Kocak. 2020. Launching revolution: social media and the egyptian uprising's first movers. *British Journal of Political Science* 50 (3): 1025–1045.

Das, S., B. Sinclair, S. W. Webster, and H. Yan. 2022. All (mayoral) politics is local? *The Journal of Politics* 84 (2): 1021–1034.

DiResta, R., S. Grossman, and A. Siegel. 2022. In-house vs. outsourced trolls: how digital mercenaries shape state influence strategies. *Political Communication* 39 (2): 222–253.

Elmas, T. 2023. The impact of data persistence bias on social media studies. In *Proceedings of the 15th acm web science conference 2023,* 196–207. WebSci '23. Austin, TX, USA: Association for Computing Machinery. https://doi.org/10.1145/3578503.3583630.

Fong, C., and J. Grimmer. 2021. Causal inference with latent treatments. *American Journal of Political Science.*

Gilardi, F., T. Gessler, M. Kubli, and S. Müller. 2022. Social media and political agenda setting. *Political Communication* 39 (1): 39–60.

Guess, A., K. Munger, J. Nagler, and J. Tucker. 2019. How accurate are survey responses on social media and politics? *Political Communication* 36 (2): 241–258.

Jones, B. T., and E. Mattiacci. 2019. A manifesto, in 140 characters or fewer: social media as a tool of rebel diplomacy. *British Journal of Political Science* 49 (2): 739–761.

Kang, T., E. F. Fowler, M. M. Franz, and T. N. Ridout. 2018. Issue consistency? comparing television advertising, tweets, and e-mail in the 2014 senate campaigns. *Political Communication* 35 (1): 32–49.

Keller, F. B., D. Schoch, S. Stier, and J. Yang. 2020. Political astroturfing on twitter: how to coordinate a disinformation campaign. *Political communication* 37 (2): 256–280. https://doi.org/10.1080/10584609.2019.1661888.

Keller, T. R., and U. Klinger. 2019. Social bots in election campaigns: theoretical, empirical, and methodological implications. *Political Communication* 36 (1): 171–189.

Ketelaars, P., and J. Sevenans. 2021. It'sa matter of timing. how the timing of politicians' information subsidies affects what becomes news. *Political Communication* 38 (3): 260–280.

Kim, T. 2023. Violent political rhetoric on twitter. *Political Science Research and Methods* 11 (4): 673–695. https://doi.org/10.1017/psrm.2022.12.

King, G., P. Lam, and M. E. Roberts. 2017. Computer-assisted keyword and document set discovery from unstructured text. *American Journal of Political Science* 61 (4): 971–988.

Kligler-Vilenchik, N., M. de Vries Kedem, D. Maier, and D. Stoltenberg. 2021. Mobilization vs. demobilization discourses on social media. *Political communication* 38 (5): 561–580.

Kobayashi, T., and Y. Ichifuji. 2015. Tweets that matter: evidence from a randomized field experiment in japan. *Political Communication* 32 (4): 574–593.

Konitzer, T., D. Rothschild, S. Hill, and K. C. Wilbur. 2019. Using big data and algorithms to determine the effect of geographically targeted advertising on vote intention: evidence from the 2012 us presidential election. *Political Communication* 36 (1): 1–16.

Kubinec, R., and J. Owen. 2021. When groups fall apart: identifying transnational polarization during the arab uprisings. *Political Analysis* 29 (4): 522–540.

Linvill, D. L., and P. L. Warren. 2020. Troll factories: manufacturing specialized disinformation on twitter. *Political Communication* 37 (4): 447–467.

Margolin, D. B., A. Hannak, and I. Weber. 2018. Political fact-checking on twitter: when do corrections have an effect? *Political Communication* 35 (2): 196–219.

Miller, B., F. Linder, and W. R. Mebane. 2020. Active learning approaches for labeling text: review and assessment of the performance of active learning approaches. *Political Analysis* 28 (4): 532–551.

Mitts, T. 2019. From isolation to radicalization: anti-muslim hostility and support for isis in the west. *American Political Science Review* 113 (1): 173–194. https://doi.org/10.1017/S0003055418000618.

Mitts, T., G. Phillips, and B. F. Walter. 2022. Studying the impact of isis propaganda campaigns. *The Journal of Politics* 84 (2): 1220–1225.

Monroe, B., M. Colaresi, and K. Quinn. 2008. Fightin' words: lexical feature selection and evaluation for identifying the content of political conflict. Funding Information: National Science Foundation (grant BCS 05-27513 and BCS 07-14688). *Political Analysis* 16 (4 SPEC. ISS.): 372–403. ISSN: 1047-1987. https://doi.org/10.1093/pan/mpn018.

Muchlinski, D., X. Yang, S. Birch, C. Macdonald, and I. Ounis. 2021. We need to go deeper: measuring electoral violence using convolutional neural networks and social media. *Political Science Research and Methods* 9 (1): 122–139.

Muddiman, A., S. C. McGregor, and N. J. Stroud. 2019. (re) claiming our expertise: parsing large text corpora with manually validated and organic dictionaries. *Political Communication* 36 (2): 214–226.

Munger, K., R. Bonneau, J. Nagler, and J. A. Tucker. 2019. Elites tweet to get feet off the streets: measuring regime social media strategies during protest. *Political Science Research and Methods* 7 (4): 815–834.

Munger, K., P. J. Egan, J. Nagler, J. Ronen, and J. Tucker. 2022. Political knowledge and misinformation in the era of social media: evidence from the 2015 uk election. *British Journal of Political Science* 52 (1): 107–127.

Nielsen, R. A. 2020. Women's authority in patriarchal social movements: the case of female salafi preachers. *American Journal of Political Science* 64 (1): 52–66.

Osmundsen, M., A. Bor, P. B. Vahlstrup, A. Bechmann, and M. B. Petersen. 2021. Partisan polarization is the primary psychological motivation behind political fake news sharing on twitter. *American Political Science Review* 115 (3): 999–1015.

Pan, J., and A. A. Siegel. 2020. How saudi crackdowns fail to silence online dissent. *American Political Science Review* 114 (1): 109–125.

Popa, S. A., Z. Fazekas, D. Braun, and M.-M. Leidecker-Sandmann. 2020. Informing the public: how party communication builds opportunity structures. *Political Communication* 37 (3): 329–349.

Settle, J. E., R. M. Bond, L. Coviello, C. J. Fariss, J. H. Fowler, and J. J. Jones. 2016. From posting to voting: the effects of political competition on online political engagement. *Political Science Research and Methods* 4 (2): 361–378.

Silva, B. C., and S.-O. Proksch. 2021. Fake it 'til you make it: a natural experiment to identify european politicians' benefit from twitter bots. *American Political Science Review* 115 (1): 316–322.

Skytte, R. 2022. Degrees of disrespect: how only extreme and rare incivility alienates the base. *The Journal of Politics* 84 (3): 1746–1759.

Sobolev, A., M. K. Chen, J. Joo, and Z. C. Steinert-Threlkeld. 2020. News and geolocated social media accurately measure protest size variation. *American Political Science Review* 114 (4): 1343–1351.

Stier, S., A. Bleier, H. Lietz, and M. Strohmaier. 2018. Election campaigning on social media: politicians, audiences, and the mediation of political communication on facebook and twitter. *Political communication* 35 (1): 50–74.

Stukal, D., S. Sanovich, R. Bonneau, and J. A. Tucker. 2022. Why botter: how pro-government bots fight opposition in russia. *American political science review* 116 (3): 843–857.

Temporão, M., C. V. Kerckhove, C. van Der Linden, Y. Dufresne, and J. M. Hendrickx. 2018. Ideological scaling of social media users: a dynamic lexicon approach. *Political Analysis* 26 (4): 457–473.

Yarchi, M., C. Baden, and N. Kligler-Vilenchik. 2021. Political polarization on the digital sphere: a cross-platform, over-time analysis of interactional, positional, and affective polarization on social media. *Political Communication* 38 (1-2): 98–139.