Filename	Description	Data Structure/Types

all_jkr_did_data_test_all.tsv	Formulation of JKR dataset for DiD analysis.	**id**: integer; row number of the account (one per account)
		X: vector of 8 0s (artifact from previous analysis).
		G2: always 0 (artifact from previous analysis).
		G: integer; indicates the account type.
		  0: non-interest actor, non-TERF.
		  1: interest actor, non-TERF.
		  2: non-interest actor TERF
		  3: interest actor, TERF.
		period: number of days since the first day of the data collection period.
		Y: following rate (interpolated) on that day. Given as the ratio of the number of new follows by users in the treatment/control group to the number of users in the treatment/control group.
		treat: 0 or 1; indicates whether the account had been retweeted by JKR at that time
		ts: days relative to retweet by JKR

all_jorts_did_data_test_all.tsv	Formulation of Jorts dataset for DiD analysis.	**id**: integer; row number of the account (one per account)
		X: vector of 8 0s (artifact from previous analysis).
		G2: always 0 (artifact from previous analysis).
		G: integer; indicates the account type.
		  0: non-union-related
		  1: union-related
		  2: non-interest actor TERF

		3: interest actor, TERF. **period**: number of days since the first day of the data collection period. **Y**: following rate (interpolated) on that day. Given as the ratio of the number of new follows by users in the treatment/control group to the number of users in the treatment/control group. **treat**: 0 or 1; indicates whether the account had been retweeted by Jorts at that time **ts**: days relative to retweet by Jorts
HASHED_jkr_followers_full_past_201806 15.json	Maps Unix cursor timestamps to chunks of JKR's followers who followed after that particular timestamp.	JSON dict. **Keys:** nanosecond Unix cursor timestamps (see this repo for more information on the Twitter V1 API cursor trick) **Values:** lists of hashed follower account IDs (as strings); these users followed JKR **after** this timestamp but **before** the next timestamp in the sorted sequence of keys.
HASHED_jkr_rts_labels_final.tsv	Indicates account labels (human-generated) for each account retweeted by JKR in the period studied.	**No name:** row ID (int) **Unnamed:0**: row ID (int) **agreement_terf:** Y or N, indicating whether coders concluded the account was a TERF-affiliated account **agreement_deleted:** Y or N indicating whether the account was deleted at any point before coding concluded. **agreement_ professional poster:** Y or N indicating whether the account's professional identity involved their Twitter account.

		agreement_non-profes: Y or N indicating whether the account was **not** on Twitter in a professional capacity. **agreement_political entity:** Y or N indicating whether the account was involved in politics (as an organization or individual). **agreement_interest actor:** Y or N indicating whether the account met the definition of an interest actor ([see L.Moses, 2023](#)). **agreement_institution/org:** Y or N indicating whether the account was an institution or organization. **hashed_user_id:** hashed handle of the retweeted account
HASHED_jkr_rts_past_20180615.tsv	Indicates the timestamps at which accounts were retweeted by JK Rowling.	**No name:** row ID (int) **Unnamed:0**: row ID (int) **created_at**: timestamp at which the retweet occurred (YYYY-mm-DD HH:MM:SS [UTC offset] **created_at_ts**: same as above **ts_created**: same as above **hashed_user_id:** hashed handle of the retweeted account.
HASHED_jorts_follower_data_by_cursor_all.json	Maps Unix cursor timestamps to chunks of Jorts' followers who followed after that particular timestamp.	JSON dict. **Keys:** nanosecond Unix cursor timestamps (see [this repo](#) for more information on the Twitter V1 API cursor trick) **Values:** lists of hashed follower account IDs (as strings); these users followed Jorts **after** this timestamp but **before** the

		next timestamp in the sorted sequence of keys.
HASHED_jorts_ground_truth_full.json	Maps hashed account handles to a string indicating whether or not the account is union-related.	JSON dict. **Keys:** hashed account handles (of retweeted accounts). **Values:** "Y" if union-related; "N" otherwise.
HASHED_jorts_rt_authors_to_ts.json	Maps hashed account handles to the timestamp at which they were retweeted.	JSON dict. **Keys:** hashed account handles (of retweeted accounts) **Values:** YYYYmmdd retweet dates.
jkr.tar.gz	tarball of a directory. Contains JSON files with filenames in the format HASHED_{hashed_account_handle}_following_data_pre_{YYYYmmdd}_all.json. **hashed_account_handle** is the hex digest of the hashed account handle. **YYYYmmdd** refers to the date that upper bounds the timestamps of the following events contained in the file. There are **3** JSON files per hashed account handle. One of them has YYYYmmdd **equal** to the date of the retweet; another has YYYYmmdd equal to **two weeks before** the retweet; and the last one has YYYYmmdd equal to **two weeks after** the retweet.	Each file is a JSON dict. **Keys:** Unix Twitter API cursors (int). The key upper bounds the timestamps at which the followers in the corresponding value followed the account whose hashed handle is in the filename. **Values:** Lists of hashed follower IDs who followed the account before the YYYYmmdd date listed in the filename. These can be cross-referenced with hashed follower IDs in HASHED_jkr_followers_full_past_201806 15.json
jorts.tar.gz	tarball of a directory.	Each file is a JSON dict.

| | | Contains JSON files with filenames in the formats HASHED_{hashed_account_handle}_following_data_2_weeks_pre_{YYYYmmdd}_all.json HASHED_{hashed_account_handle}_following_data_2_weeks_after_{YYYYmmdd}_all.json HASHED_{hashed_account_handle}_following_data_post_{YYYYmmdd}_all.json

hashed_account_handle is the hex digest of the hashed account handle. **YYYYmmdd** refers to the date that the account handle in **hashed_account_handle** was retweeted by Jorts.

There are **3** JSON files per hashed account handle. One of them contains following events upper bounded by 2 weeks after the retweet event; another contains following events upper bounded by the retweet event itself, and the final one contains following events upper bounded by 2 weeks before the retweet event. | **Keys:** Unix Twitter API cursors (int). The key upper bounds the timestamps at which the followers in the corresponding value followed the account whose hashed handle is in the filename. **Values:** Lists of hashed follower IDs who followed the account before the YYYYmmdd date listed in the filename. These can be cross-referenced with hashed follower IDs in HASHED_jorts_follower_data_by_cursor_all.json. |