academictwitteR: an R package to access the Twitter Academic Research Product Track v2 API endpoint

Christopher Barrie[1] and Justin Chun-ting Ho[2]

1 School of Social and Political Sciences, University of Edinburgh, Scotland, UK. **2** Centre for European Studies and Comparative Politics, Sciences Po, France.

Statement of need

In January, 2021, Twitter announced the "Academic Research Product Track." This provides academic researchers with greatly expanded access to Twitter data. Existing R packages for querying the Twitter API, such as the popular `rtweet` package (Kearney, 2019), are yet to introduce functionality to allow users to connect to the new v2 API endpoints with Academic Research Product Track credentials. The `academictwitteR` package (Barrie & Ho, 2021) is built with academic research in mind. It encourages efficient and responsible storage of data, given the likely large amounts of data being collected, as well as a number of shortcut and query building functions to access new v2 API endpoints.

Summary

The Twitter Application Programming Interface, or API, was first introduced in 2006. It was designed principally with commercial objectives in mind. Over time, however, researchers began to repurpose the Twitter API for academic ends. In January, 2021, Twitter announced the "Academic Research Product Track", noting that "[t]oday, academic researchers are one of the largest groups of people using the Twitter API."

Authorization for the Academic Research Product Track provides access to the Twitter v2 API endpoints, introduced in 2020, as well as much improved data access. In summary the Academic Research product track allows the authorized user:

1. Access to the full archive of (as-yet-undeleted) tweets published on Twitter;
2. A higher monthly tweet cap (10m—or 20x what was previously possible with the standard v1.1 API);
3. Ability to access these data with more precise filters permitted by the v2 API.

The `academictwitteR` package was designed: 1) to make the Academic Research Product Track easily accessible for R users by providing dedicated functions to query the the v2 API endpoints; 2) to encourage academic researchers efficiently and safely to store their data.

The functions allow the user to collect tweets from (or to) specified users and to collect tweets containing specified words or sets of words. In particular, queries that include so-called "conjunction-required" operators can also be accessed via a set of shortcut functions for accessing e.g. tweets containing media content, tweets containing geographic location information, or tweets containing urls. Additionally, separate query builder functions allow the user to specify complex queries to incorporate into the API call.

Data is stored in serialized form as RDS files or as separate JSON files. The former represents the most efficient storage solution for native R data-file formats; the latter helps mitigate loss by storing data as separate JSONs for each pagination token (or up to 500 tweets). Convenience functions are also included to bind tweet- and user-level information stored as JSON files, and to pick up data collection where it left off in the case of unplanned interruption.

References

Barrie, C., & Ho, J. C. (2021). *academictwitteR: An r package to access the twitter academic research product track v2 API endpoint.* https://doi.org/10.5281/zenodo.4714637

Kearney, M. W. (2019). Rtweet: Collecting and analyzing twitter data. *Journal of Open Source Software*, *4*(42), 1829. https://doi.org/10.21105/joss.01829