"Longform" Twitter: How Twitter threads are used to share COVID-19 information

Katy Ilonka Gero, Jennifer Lee, Sarah Huang, Lydia B. Chilton
Department of Computer Science, Columbia University

CUEPID Data Challenge, April 2020

How does “longform” info spread on Twitter?

01 Twitter is an important venue for information sharing, especially during disasters, but studies typically focus on individual tweets [1, 2]

02 **We focus on Twitter threads**: linked lists of tweets from a single author

03 We collected ~1000 coronavirus-related Twitter threads and annotated the ones with over 50 likes (163 threads).

04 This project helps:

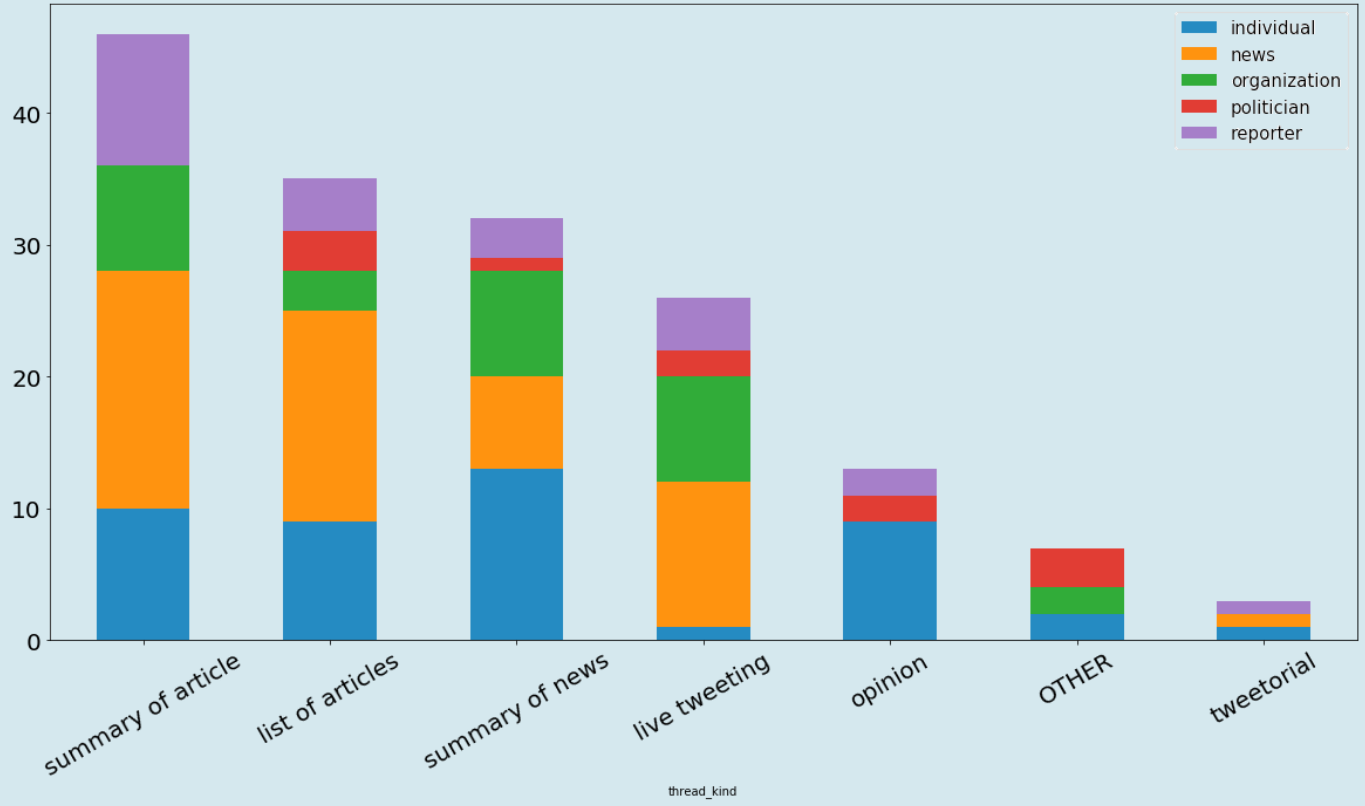
- Organizations understand how they can share info on Twitter
- Researchers understand how info is spreading on Twitter

1. Stowe, Kevin, et al. "Developing and evaluating annotation procedures for Twitter data during hazard events." *Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018)*. 2018.
2. Anderson, T. Jennings, et al. "Far Far Away in Far Rockaway: Responses to Risks and Impacts during Hurricane Sandy through First-Person Social Media Narratives." *ISCRAM*. 2016.

Data collection and annotation

- Collected tweets with 'coronavirus' or 'covid19' in the text using twarc
- On a daily basis search for threads: tweets that include 3+ replies from the same author
- **We collected ~1k Twitter threads from 4/9-4/19**
- **We annotated threads with 50+ likes (163 in total) according to**:
  - Locality (focus of content, e.g. USA, global, India)
  - Source (kind of author, e.g. reporter, politician, news organization)
  - Type (structure of content, e.g. summary of article, live tweeting)
- Additional Twitter data includes # retweets, # likes, user profile info

Most Twitter threads are about external news

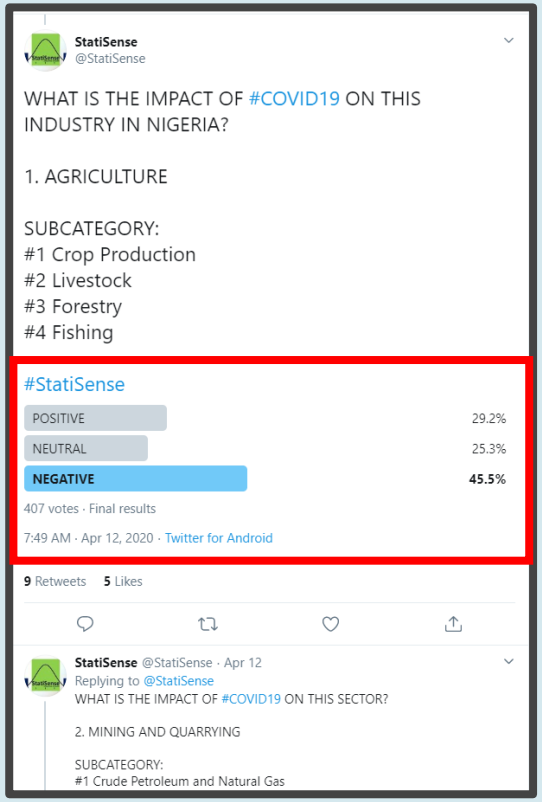


- Most threads summarize or refer to external news sources
- Most threads are tweeted by news organizations

Next steps:

- How many readers read the external sources?
- Are there predictors of likes or retweets other than the author of the thread?

Surprising uses of Twitter threads

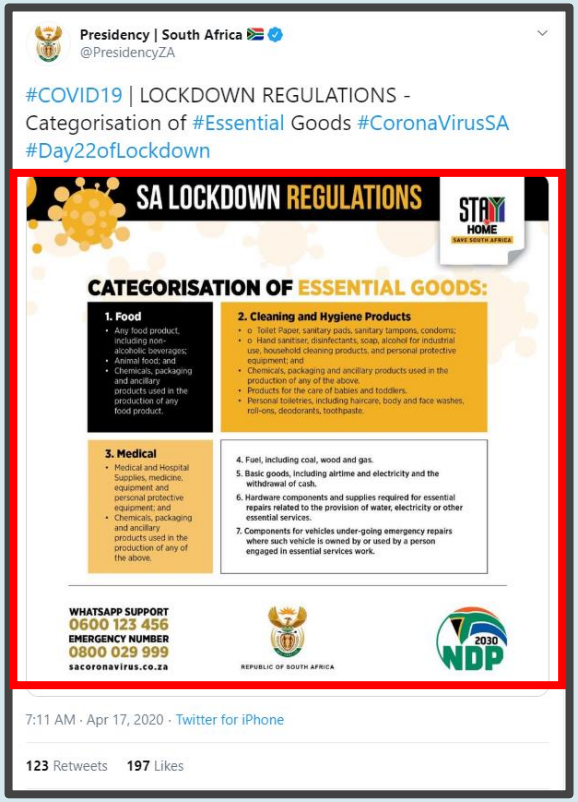


To gather data

This thread was a series of polls from a data science organization

To explain a concept

This thread explains science behind antibodies as protection



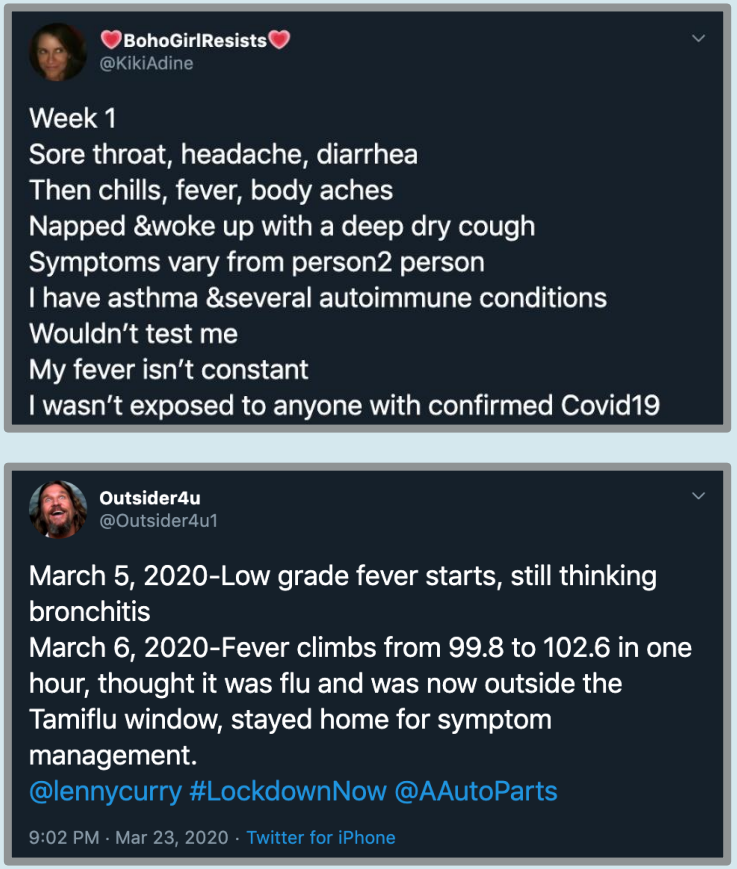
Official gov communication

This is a thread of announcements from South Africa's president

Why were so few personal stories collected?

1. Method of data collection
   - Often people will have a series of Tweets that are not explicitly a thread.
   - Individuals are less likely to have many followers leading to fewer likes.
2. Format of individual tweets
   - Character limit inhibits free-form writing.
   - People may prefer to tweet personal stories on private accounts.

To improve future data collection:

- Expand search to indirect terms (such as “pandemic”, “virus”, “quarantine”, “isolation”)
- Filter out tweets made by news organizations and focus a bigger part of search on individuals

Examples of personal stories not identified by our algorithm: