



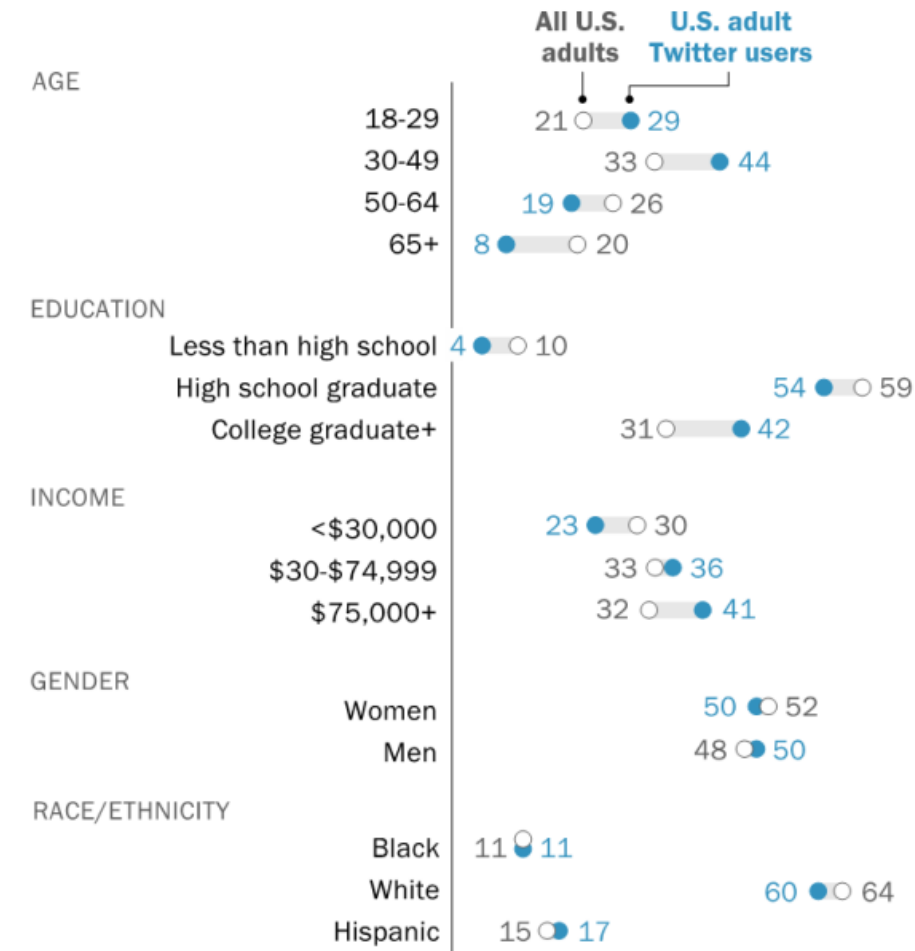
# Notes from the **API**calypse: The **Twitter** V2 API

---

Ed Summers  
edsu@umd.edu  
mith.umd.edu

## Twitter users are younger, more highly educated and wealthier than general public

% of \_\_\_\_\_ who are ...



Pew Research Center, April, 2019, "Sizing Up Twitter Users".



## Google Scholar

"Twitter" in title

2018: 61,300 papers

2021: 108,000 papers

"Facebook" in title

2021: 103,000 papers



2.8 Billion > 330 Million

# API

Application  
Programming  
Interface

# URLS FOR DATA

**<https://api.twitter.com/2/users/813286/tweets>**



Information, Communication &amp; Society &gt;

Volume 22, 2019 - Issue 11: Locked Out

Submit an article

Journal homepage

Enter keywords, authors, DOI, ORCID etc

This Journal



Advanced search

3,753

Views

43

CrossRef citations  
to date

110

Altmetric

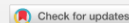
Articles

# After the 'APIcalypse': social media platforms and their fight against critical scholarly research

Axel Bruns

Pages 1544-1566 | Received 24 Nov 2018, Accepted 06 Jun 2019, Published online: 11 Jul 2019

Download citation

<https://doi.org/10.1080/1369118X.2019.1637447>

Full Article

Figures &amp; data

References

Citations

Metrics

Reprints &amp; Permissions

Get access

## ABSTRACT

In the aftermath of the Cambridge Analytica controversy, social media platform providers such as *Facebook* and *Twitter* have severely restricted access to platform data via their Application Programming Interfaces (APIs). This has had a particularly critical effect on the ability of social media researchers to investigate phenomena such as abuse, hate speech, trolling, and

## Related research

People also  
readRecommended  
articlesCited by  
43

An end to the wild west of social media research: a response to Axel Bruns >

Cornelius Puschmann

Support the Guardian

Available for everyone, funded by readers

Contribute →

Subscribe →

Search jobs

Sign in

Search

US edition ▾

The  
Guardian  
For 200 years

News

Opinion

Sport

Culture

Lifestyle

More ▾



**The Cambridge  
Analytica Files**  
Cambridge Analytica

**Carole Cadwalladr and  
Emma Graham-Harrison**

Sat 17 Mar 2018 18.03 EDT



48,162



484

🕒 This article is more than 3 years old

## Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach

**Whistleblower describes how firm linked to former Trump adviser Steve Bannon compiled user data to target American voters**

- **'I made Steve Bannon's psychological warfare tool': meet the data war whistleblower**
- **Mark Zuckerberg breaks silence on Cambridge Analytica**







**Adam Tornes**

[@atornes](#)

Staff Product Manager,  
Developer & Enterprise  
Solutions



**Leanne Trujillo**

[@leanne\\_tru](#)

Sr. Program Manager,  
Developer & Enterprise

Tools

# Enabling the future of academic research with the Twitter API

By [Adam Tornes](#) and [Leanne Trujillo](#)

Tuesday, 26 January 2021



When we introduced the [next generation of the Twitter API](#) in July 2020, we also shared our plans to invest in the success of the academic research community with tailored solutions that better serve their goals. Today, we're excited to launch the Academic Research product track on the new Twitter API.

## Why we're launching this & how we got here

Since the Twitter API was first introduced in 2006, academic researchers have used data from the public conversation to study topics as diverse as the conversation on Twitter itself - from [state-backed efforts to disrupt the public conversation](#) to [floods and climate](#)



# IT'S ACADEMIC



- 
- Research-focused employee at an academic institution
  - Clearly defined research objective
  - Non-commercial



# THE GOOD



- 
- No more contests for data
  - Search the full archive (not just last 7 days)
  - No more stream volume throttling
  - Adjusting stream rules on the fly
  - Collecting conversation threads
  - Tweeting entities: people, places, organizations

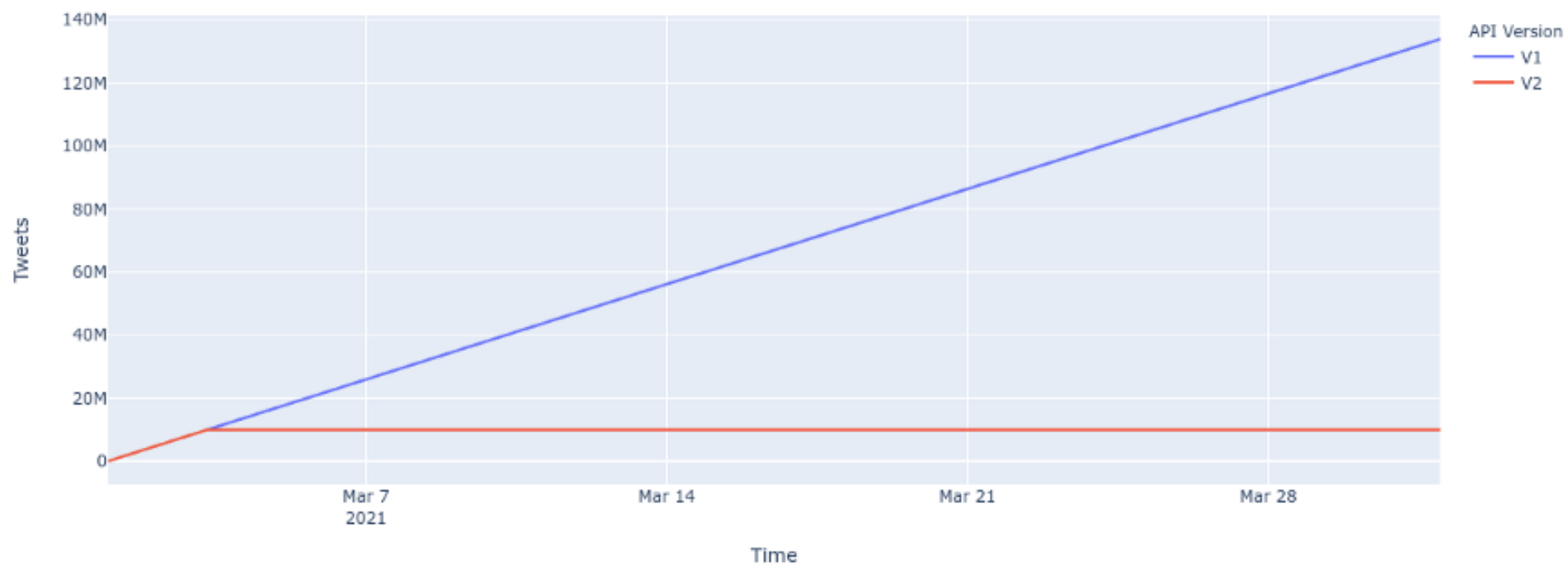


# THE BAD

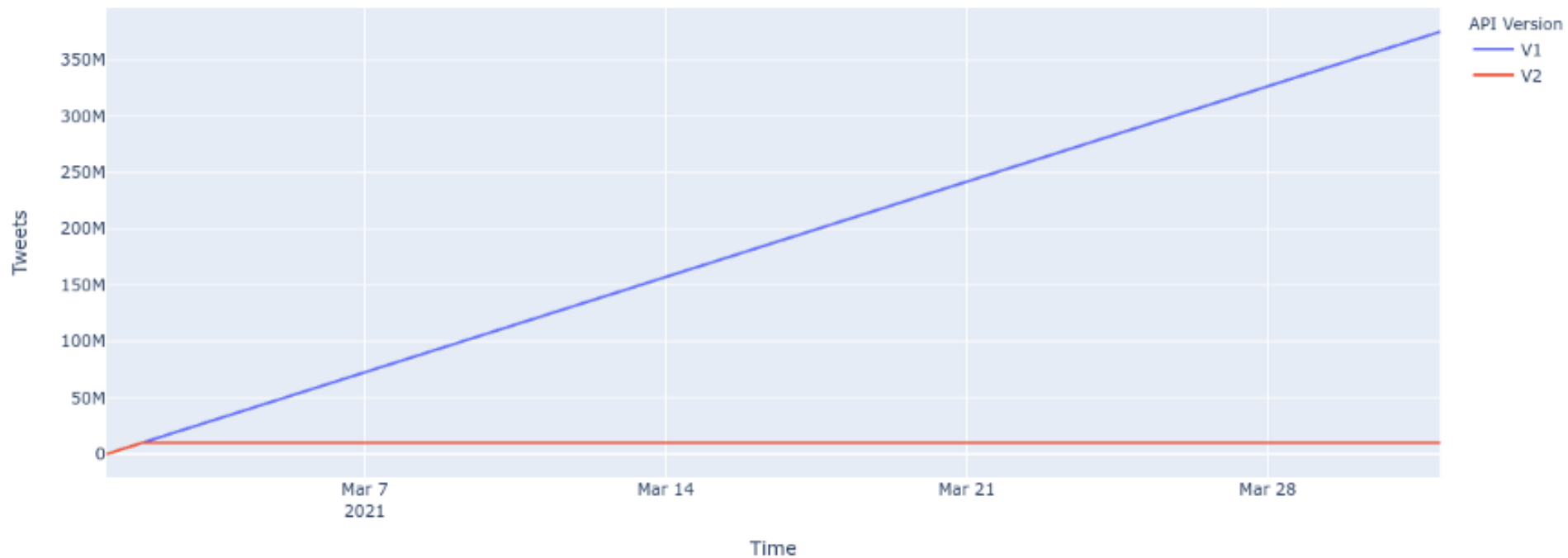


- 
- Premium product offerings going away
  - Twitter as gate keeper of what is research
  - No more canonical tweet representation
  - *Much* lower quotas (slightly better for academic researchers)

## Monthly Data Collection from Application Twitter Search API



Monthly Data Collection from Application Filter Stream API



# WHAT'S CANONICAL?

**expansions:** author\_id in\_reply\_to\_user\_id referenced\_tweets.id  
referenced\_tweets.id.author\_id entities.mentions.username  
attachments.poll\_ids attachments.media\_keys geo.place\_id

**users:** created\_at description entities id location name pinned\_tweet\_id  
profile\_image\_url protected public\_metrics url username verified  
withheld

**tweets:** attachments author\_id context\_annotations conversation\_id  
created\_at entities geo id in\_reply\_to\_user\_id lang public\_metrics text  
possibly\_sensitive referenced\_tweets reply\_settings source withheld

**media:** duration\_ms height media\_key preview\_image\_url type url width  
public\_metrics

**polls:** duration\_minutes end\_datetime id options voting\_status]



**<https://github.com/docnow/twarc>**





<https://www.docnow.io>

# TOOLS AND PRACTICES

---

1. **twarc**: tweet archiving utility
2. **docnow**: appraisal application
3. **catalog**: clearinghouse for twitter id datasets
4. **hydrator**: desktop app for tweet id datasets