



Education Engagement on Twitter

CHEN FAN WANG

Executive Summary

- Users with most original tweets often only post **mundane personal topics**
- **Influencer** twitterers with high retweet counts post **politically charged** topics that model their political party's latest views on education
- Tweets about education are especially **popular** when **school starts** around August/September and **die down** during **summer vacation**
- Tweets predominantly originate from the U.S, with discussion of hot button topics dominated by U.S tweets
- **Influencers** have much **broader reach** than all other user groups and hence **should be focused on** for tracking emerging topics
- **Original tweets** of influencers are also **unique** and of **interesting** topics regarding education instead of repeated mundane tweets

Methodology & Data Source



Data was collected from Twitter API

- Contained ~100 million tweets mostly in English
- Tweets were collected between April 5th 2022 and November 6th 2022
- Retained ~30 million tweets after filtering out tweets not directly related to education



Analytical Method

- Data was processed and analyzed in PySpark
- Utilized GCP for data storage and analysis
- Pandas and Seaborn/Matplotlib for plots
- Minhashing for text similarity analysis

Raw Data EDA Key Insights



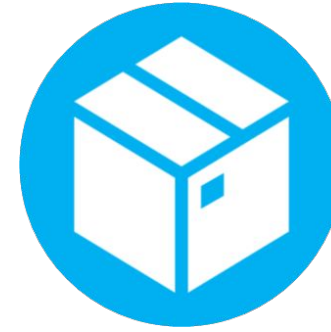
Many columns had unreliable data

- E.g user location can be populated with any text; hence many users populate it with their pronouns instead of location
- Engineered so the place object was used to populate location



Tweets were immediately recorded after posting and never updated

- Hence many tweet variables were 0 as they were not updated after being posted (e.g retweet count)
- These variables were re-engineered by looking at tweets created at later dates that contained updated information in JSON objects



Embedded JSON Objects contained more reliable info

- E.g Retweet status contained updated retweet and favorite counts for original tweet
- User, Retweet status, and Place JSON objects will be exploded to populate unreliable fields

Filtering Education Related Tweets

- Only tweets with certain keyword combinations were retained, with keywords determined through EDA
 - Tweets with words such as [primary school](#), [secondary school](#), [higher education](#) etc., were directly related to K-12 and Higher education
 - “these books helped us a lot in primary school 🤔🤔 <https://t.co/maj8bg9yow>”
 - Words such as [curriculum](#) and [literacy](#) were mostly related to K-12 schooling
 - “@subvversion media literacy needs to be part of high school curriculum so fucking bad.”
 - Subject keywords such as [math](#) and [humanities](#) were almost always about K-12 schooling
 - “hearing the news coming out of Roehampton... about mass redundancies in arts & humanities subjects makes me feel physically sick...”
 - Most [R18](#) tweets were filtered out with column [possibly sensitive](#)



Filtering Education Related Tweets

- EDA on remaining tweets showed many tweets on certain topics that should be removed even if they contained keywords on education
 - Tweets about **school sports** or **live streams** were removed
 - “northeast metro rvt vs essex tech - high school lacrosse ● live stream watch live game: [https...](#)”
 - **School shootings** though tangentially related was mainly about **gun control**
 - “...what kind of fukcing monster kills children, fucking primary school age children??”
 - Most **vaccine** tweets were mostly about wider **public health**
 - “your government overthrew you and you can't even see it. if you've masked up and vaxxed you did nothing but comply to stupidity. our child goes to school with over 1000 unmasked kids, zero cases. your healthcare stinks. your country is a prison now.”



Top 5 Most Prolific Twitterers

USERS WITH THE **MOST ORIGINAL TWEETS** ARE ALMOST ALL DAILY TWEETS REPEATING THE SAME PERSONAL TOPICS WITH LITTLE SUBSTANCE

	NJ School Jobs.com	Kae	Dennis Stemmler	KQ education group	Kevin Edwards
Unique Tweets	9887	9357	8411	6162	3113
Tweets About	Job website posting school job openings	Grad student raising money for lab equipment	Author promoting his self-help books about college	Links to their own external website about education	Teacher raising money for classroom supplies

Examples:

“intervention teacher at plainfield public schools in plainfield, nj: qualifications: nj department of education standard teacher certification ten years of classroom teaching experience in elementary or secondary grade levels...”

“hello, mima shopee 🧡✨ i am in my internship yr and a graduating student who need funds para makabili ng laboratory equipments, school supplies for me and my brother and also, home necessities para kay mama. thank you shopee, we love you”

Top 5 Most Influential Twitterers

TWITTERERS WITH **HIGHEST NUMBER OF RETWEETS** (OF THEIR OWN ORIGINAL TWEETS) ARE OFTEN POLITICALLY CHARGED INFLUENCERS POSTING INCENDIARY TOPICS IN THE EDUCATION SPACE

	Libs of Tik Tok	Occupy Democrats	Christopher F. Rufo	No Lie with Brian Tyler Cohen	Scorpiio
Times User Original Tweets are Retweeted	167290	59812	41818	38845	34063
Tweets About	Right-wing influencer account claiming education is corrupted by the left	Grassroots organization sharing left-wing news and views	Individual that often shares right-wing news and views	Individual that often shares left-wing news and views	Former university student sharing UMich medical student walkout news

Examples:

“ your daily reminder that when they accuse republicans of “banning books” they’re referring to pornographic books which are offered to kids in schools across the country.”

“ I’m working on a ten-part investigative series on radical gender theory in American schools.”

What Type of Accounts are These Users?

We categorized users into 5 main categories most likely involved with education based on their user descriptions, tweets, and usage statistics to understand who these prolific/influential twitterers are



**Official Government
Accounts,
Congressmen,
Senators etc.**



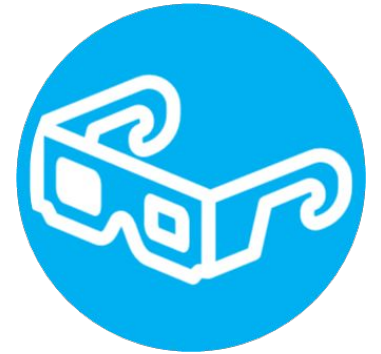
**Official News
Agency Accounts,
Journalists,
Reporters**



**Official
Non-Governmental
Agency Accounts**



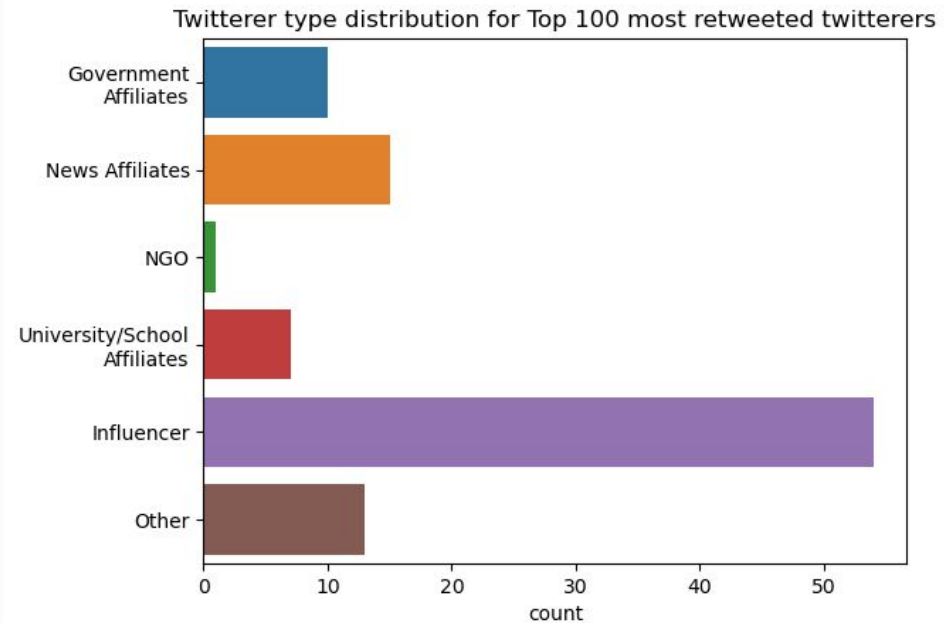
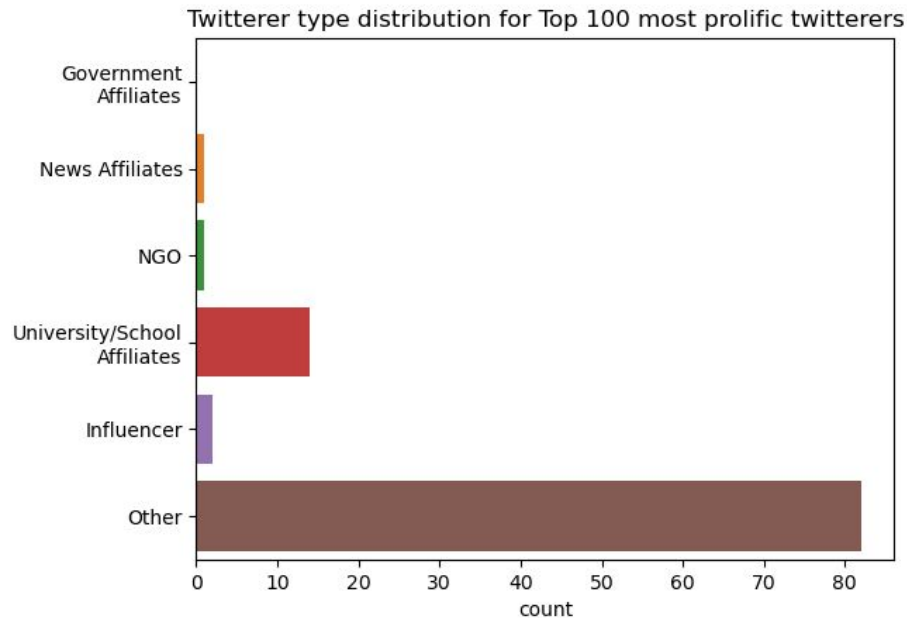
**Official
University/School
Accounts,
Professors,
Teachers**



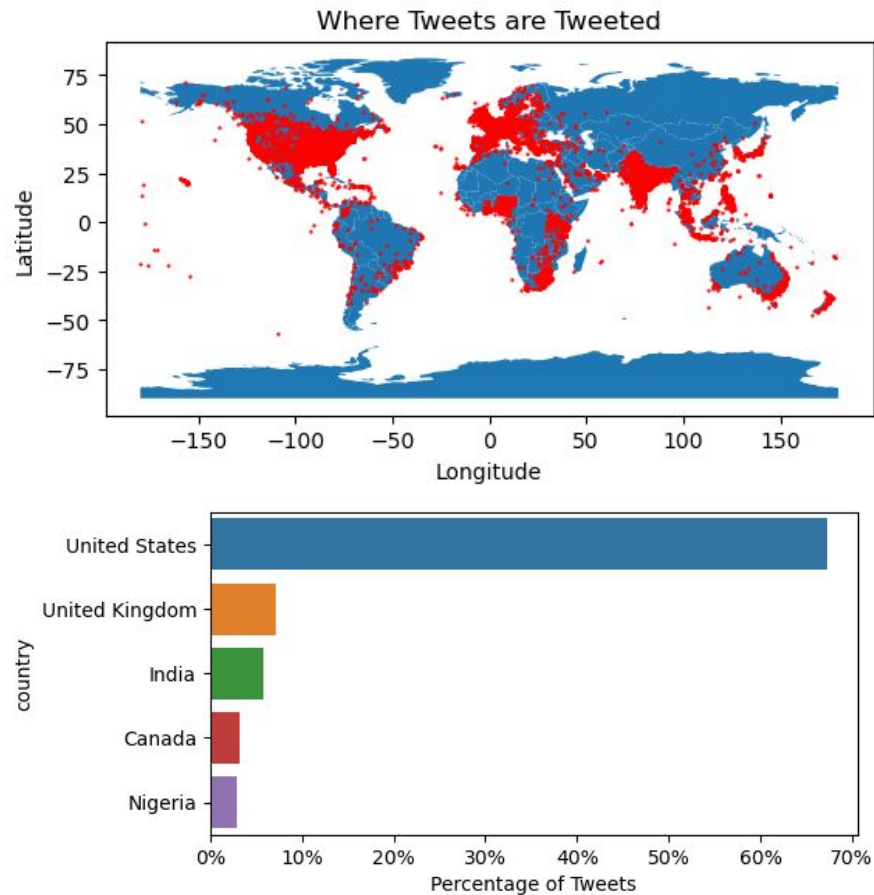
Influencers

What Type of Accounts are These Users?

- Twitterers with the most **original** (not retweeted) tweets are predominantly random individuals posting education related content of only personal relevance, hence are of **low importance**
- Twitterers with the most **retweeted** original tweets are predominantly influencers posting **politically charged** content **widely supported** and **shared** by those of the same political alignment



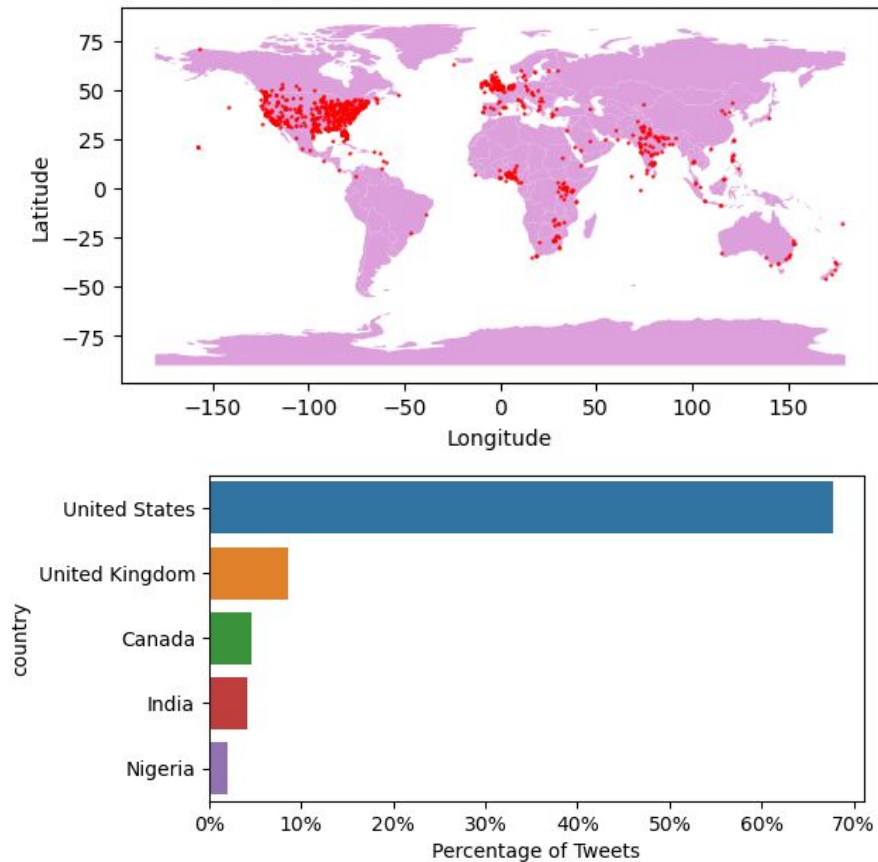
Where are Tweets Coming From?



- English Educational tweets can come from **all over the world**, not solely english speaking countries
- Most of the tweets are however coming **from the United States**

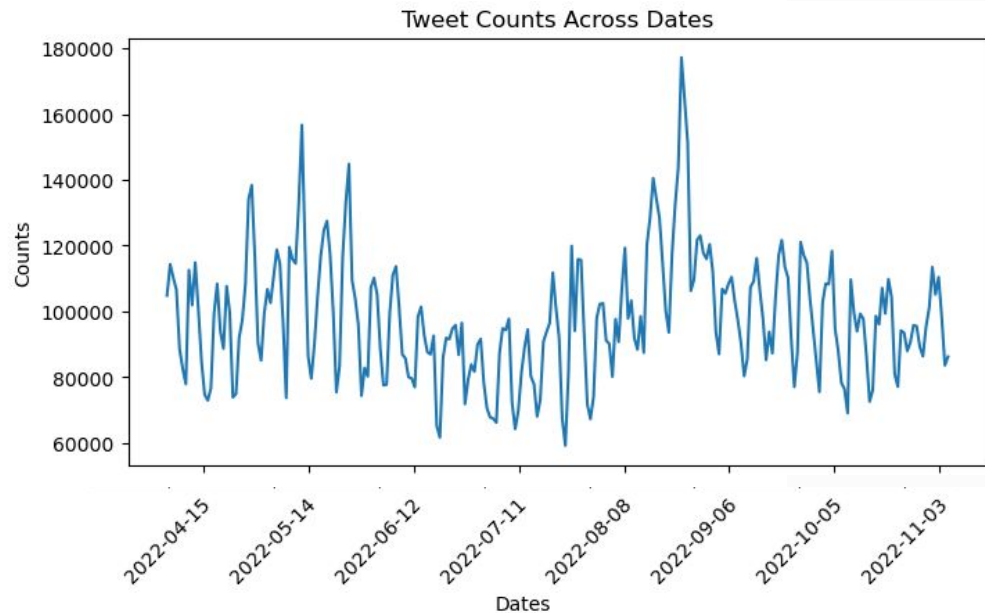
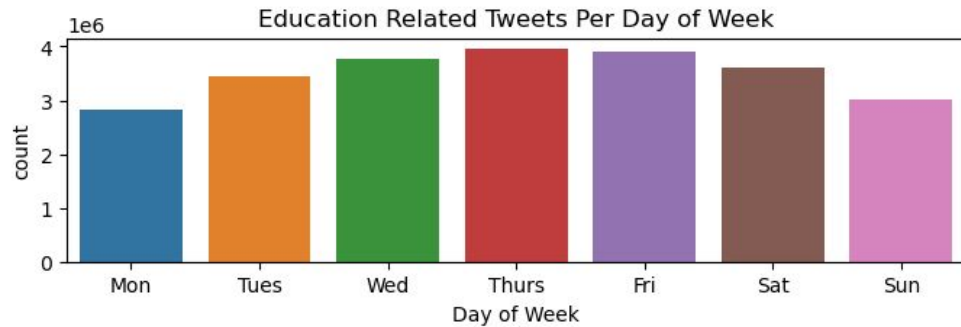
Educational Issues vs. Location

Where Tweets Regarding Gender/Sex Education In School are Tweeted



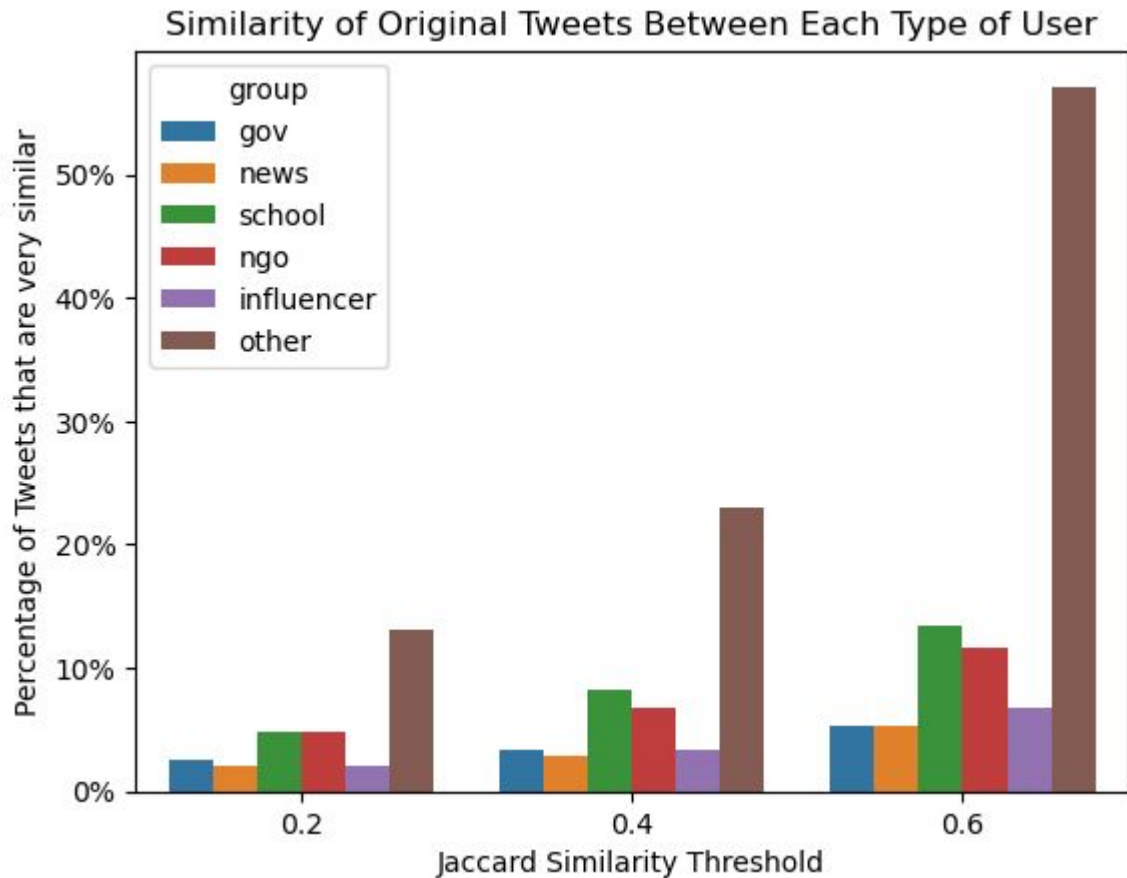
- Gender/Sex education in school was a hot topic in the educational space during the data collection period
- Top countries tweeting on the topic are the same as top countries tweeting in general
- Most tweets regarding gender/sex education is, however, further concentrated in English speaking countries
 - India and Nigeria had significant less gender/sex ed tweets, with India falling behind Canada in tweet count

When are Tweets Tweeted?



- Number of education tweets in general follow a **weekly cyclic** pattern, with most tweets occurring on **Thursday**
- Number of education **tweets dip** during **summer vacation** and **spike** as **school starts**

How Similar are the Tweets?



- For random users, even **original tweets** are often similar and not unique
 - This is because most random users normally tweet short and simple status updates such as “high school sucks bro”
- Though not as much as random users, tweets within **schools** and **NGO** are also **relatively similar**
 - This is because many of their tweets are about fundraising events that look very similar to each other
- Government, news, and influencers have relatively unique tweets
 - This is because they seek to actually unearth interesting topics to bring the public’s attention to, hence their tweets have more substantive content worth reading

Conclusion & Action Items

- It is possible to extract trends or topics on education from twitter
- Tweets of **influencers**, **government**, and news affiliates post **original tweets** of **substance**
- Tweets of **influencers**, however, have far **wider reach** and audience than government or news affiliates and therefore **better reflect public sentiment**
- To quickly get a sense of **hotly discussed** education topics of **interest**, one can view **tweets** of **top influencers** in the US with keywords related to education