

Research on Social Media

Ali Ünlü, PhD aliunlu@virginia.edu



Why Social Media Research Matters?

How Is Social Media Data Collected?

How Is Social Media Data Analyzed?



Youth & Social Media

- How does social media shape our understanding of human behavior and development?
- Young people social media platforms usage (<u>PEW</u>, 2024)
 - YouTube: Nearly all teens (93%) use YouTube, making it the most popular platform among this age group.
 - TikTok: Around 63% of teens use TikTok, which has seen significant growth in recent years.
 - Snapchat: Approximately 60% of teens are on Snapchat.
 - Instagram: About 62% of teens use Instagram.
- 40% of youth aged 8 -12 use social media (The U.S. Surgeon General's Advisory, 2023)
- Social media ban for youth in Australia and ongoing discussions about regulations



Social Media & Policy and Politics

Policy Making

- Platform for public discourse and mobilization.
- Tool for advocacy and direct communication with policymakers.

Public Opinion

- Real-time access to sentiment and trends.
- Tools for monitoring reactions and predicting behavior.

Agenda Setting

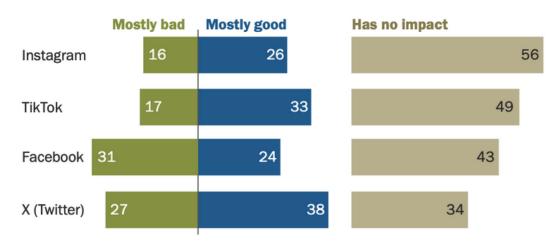
- · Amplifies issues and influences media framing.
- Highlights marginalized voices and shapes debates.

Polarization

- Creates echo chambers and filter bubbles.
- Intensifies partisan divides and spreads misinformation.

The largest shares of Instagram, TikTok and Facebook users say the platforms have no impact on democracy

% of each platform's U.S. users who say overall, they think it is ___ for American democracy



Note: Those who did not give an answer are not shown.

Source: Pew Research Center survey of U.S. adults conducted March 18-24, 2024.

"How Americans Navigate Politics on TikTok, X, Facebook and Instagram"

PEW-KNIGHT INITIATIVE

Collecting Social Media

Data

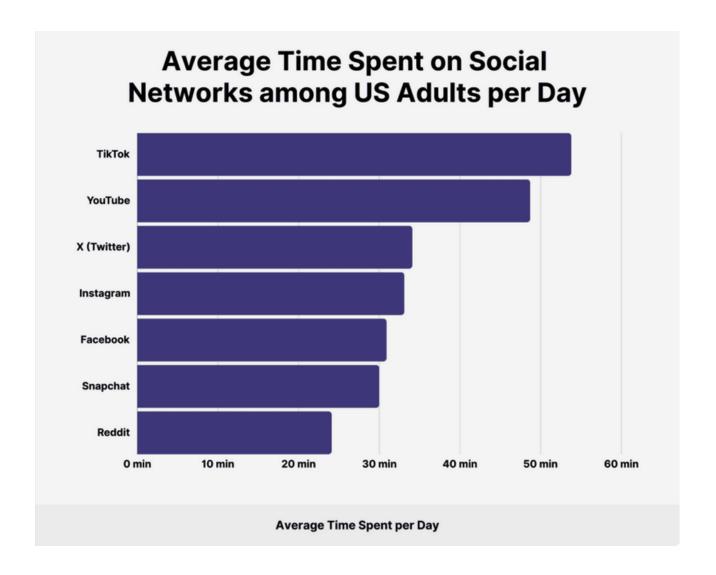
Data Sources and Types

• Platforms:

 X, Facebook, Instagram, TikTok, LinkedIn, YouTube, Reddit, etc.

Data Types:

 Text, images, videos, hashtags, and metadata.



Source: Backlinko.com

Collection Methods

- APIs and Web Scraping: Examples of tools like Twitter API, Communalytic, and NodeXL.
- Surveys and Experiments:
 Combining traditional methods
 with digital ethnography.
- Ethical Considerations:
 - Privacy, consent, and algorithmic bias.
 - Introduce frameworks such as ethical AI principles and personal data protection compliance.



Accessing Social Media Data: Tools and Methods

Platform-Wide Data Access:

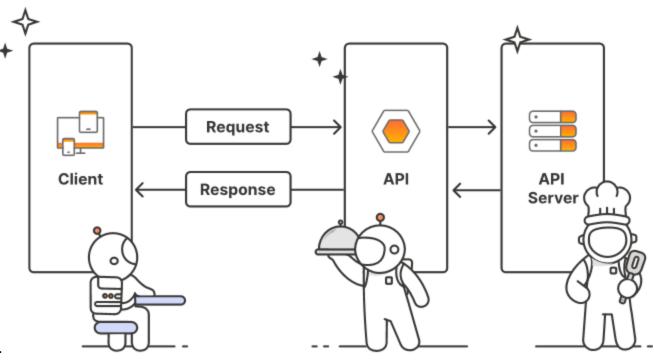
- API Access (e.g., Twitter/X, YouTube, Reddit, META):
 - Requires documentation and sometimes research approval or funding.
 - Ideal for large-scale data collection and trend analysis.

Focused Data Extraction:

- For **specific posts** (e.g., Instagram comments) or **pages** (e.g., Facebook groups):
 - Use scraping tools with third-party software.

Suggested Tools:

- Facepager Extracts data from APIs and web pages.
- Communalytic Web-based tool for text and network analysis.
- Octoparse Visual web scraping with no coding required.
- Social Media Archive (SOMAR) Inter-university Consortium for Political and Social Research (ICPSR)
- TAGS (Twitter Archiving Google Sheet) Free for Twitter/X data collection.
- NodeXL Social network analysis for Excel users.
- **Web Scraper.io** Browser-based scraping for web data collection.
- Python Libraries: BeautifulSoup, Instaloader, Selenium, Puppeteer twarch, Scrapy





How is Social Media Data Analyzed?

Approaches

Qualitative Analysis

- Content Analysis (manual coding or softwareassisted tools like NVivo).
- Thematic Analysis and Discourse Analysis—link to narratives in education and identity research.

Quantitative Approaches

- Sentiment Analysis and Network Analysis focus on group dynamics and peer influence.
- Machine Learning Techniques—classification models, topic modeling (e.g., Latent Dirichlet Allocation).

Mixed-Methods Designs

 Integrating computational tools with qualitative insights for holistic interpretations.



Source: https://www.flickr.com/photos/nicholekelly/5071174651/

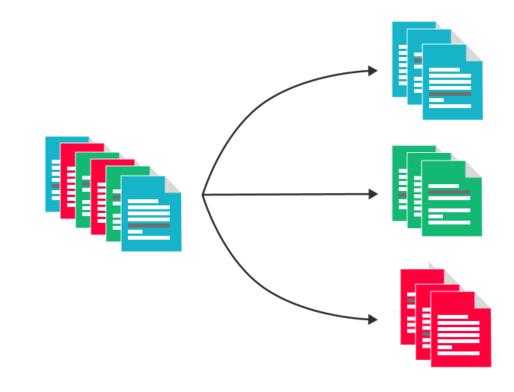
Making Sense of Social Media Data

- You have just downloaded a massive dataset from Twitter, Reddit, or Facebook. It contains millions of raw text posts.
 - What's the first step to analyzing this?



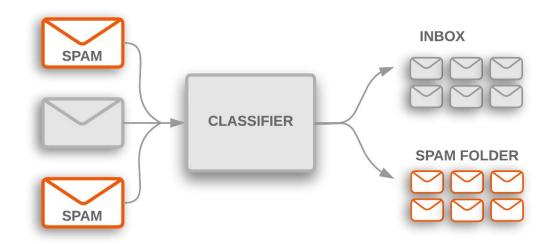
From Chaos to Clarity

- Challenges of Unstructured Text:
 - Too much data to manually process
 - No clear structure—just streams of words
 - Difficult to extract meaningful insights
- Solution:
 - Helps group similar posts together
 - Extracts key information efficiently
 - Filters noise & identifies trends



Key Steps for Text Classification

- Data Preparation Collect and preprocess corpora.
- Feature Extraction –
 Represent text (e.g., TF-IDF, embeddings).
- Model Selection & Training (e.g., Naïve Bayes, SVM, LSTM, Transformer-based models)
- Model Evaluation & Tuning



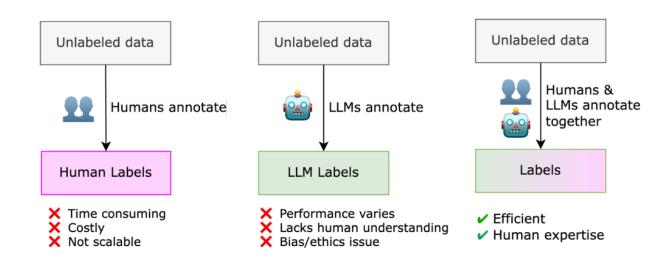
Text Classification

Data Annotation:

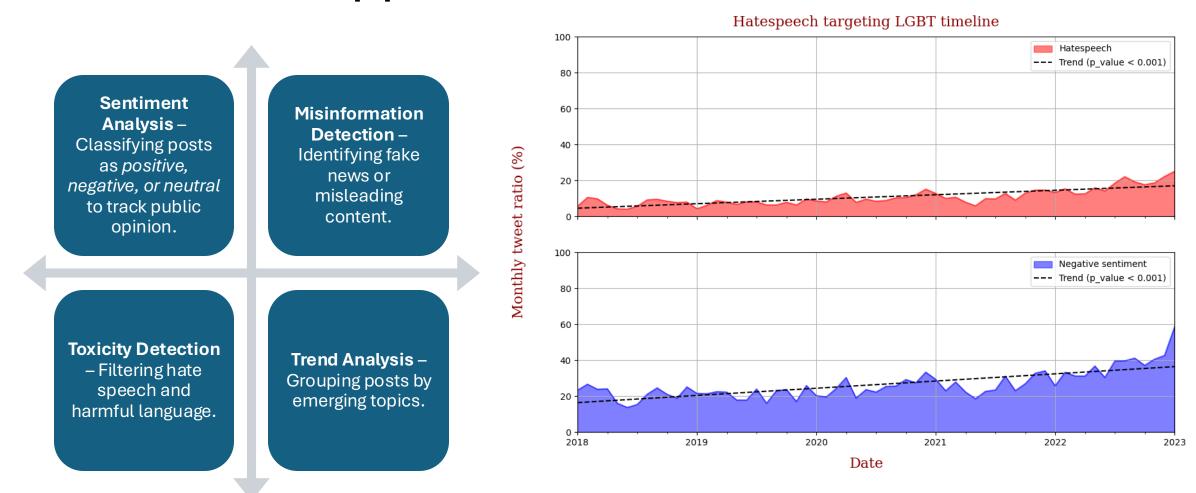
- Requires time, human effort, and financial resources.
- Needs domain knowledge and theoretical constructs for labels.
- More effective

Few Shot Learning with LLMs:

- **FastFit/Setfit:** contrastive fine-tuning of pretrained sentence transformers, then a lightweight classifier.
- Prompt-Based: Leverages LLM by designing task-specific prompts; minimal or zero finetuning required.
- Meta Learning: Optimizes for rapid adaptation across tasks with few examples, enabling quick learning.



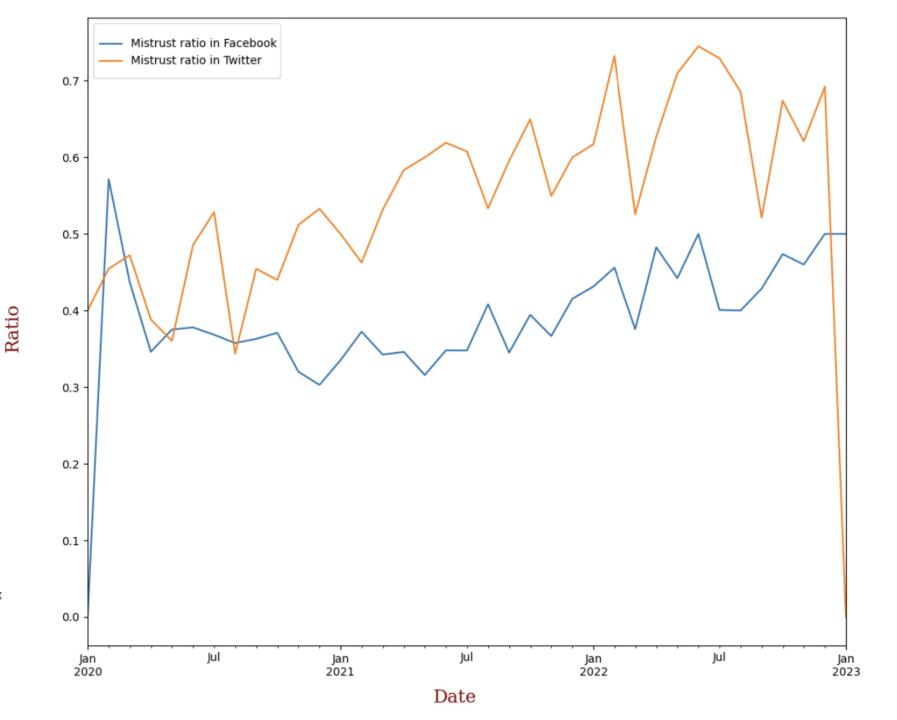
Real-World Applications



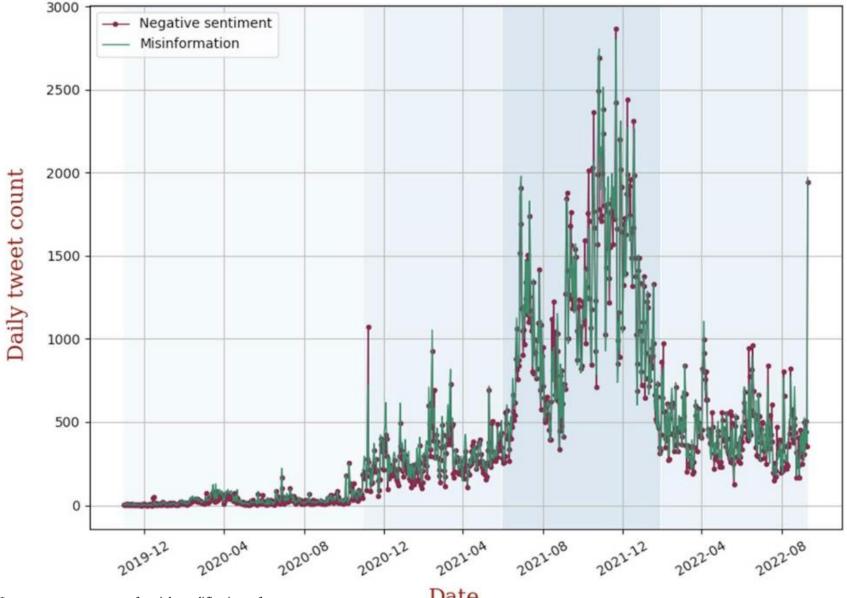
Unlu, A, Lac, T., Sawhney, N., Tammi, T. and Kotonen, T (2025) From Prejudice to Polarization: Tracing the Forms of Online Hate Speech Targeting LGBTQ+ and Muslim Communities in Finland. New Media and Society. https://doi.org/10.1177/14614448241312900

Political trust to Finnish Institute for Health and Welfare (THL) across platforms during COVID-19

Unlu, A., Lac, T., Tammi, T., and Lohiniva, A.L. (2023) Exploring Political Mistrust in Pandemic Risk Communication: A Mixed-Method Study Using Social Media Data Analysis. *Journal of Medical Internet Research – DOI: 10.2196/50199*https://www.jmir.org/2023/1/e50199/



Misinformation in COVID-19 Conversations



Beyond
Classification –
What Are People
Actually Talking
About?

Classification tells us *how many* people support the policy.

But it doesn't tell us what specific issues they are discussing.

Supporters might have different reasons—some focus on gun restrictions, others on mental health reform.

Two-Step Approach to Social Media Analysis

Classification:

- Groups text into broad categories (e.g., Support, Oppose, Neutral).
- Provides structured data for large-scale analysis.

Topic Modeling:

- Extracts **hidden themes** within each category.
- Helps policymakers, researchers, and organizations understand why people hold certain opinions.

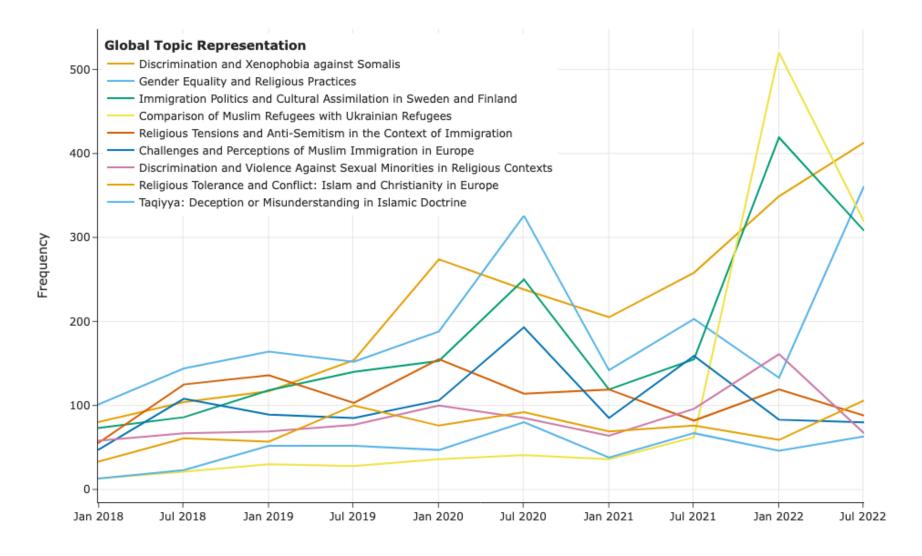
How Do We Discover Hidden Themes in Text?

What is Topic Modeling?

- A technique to automatically identify topics (hidden themes)
- Unsupervised learning no predefined labels are needed.
- Helps summarize, categorize, and interpret text data efficiently.

Common Topic Modeling Methods

- Latent Dirichlet Allocation (LDA) Probabilistic Model
- Non-Negative Matrix Factorization (NMF) Linear Algebra Approach
- BERTopic Transformer-Based Model
- Dynamic Topic Models (DTM) Topic Evolution Over Time



Unlu, A., Kotonen, T. (2024). Online polarization and identity politics: An analysis of Facebook discourse on Muslim and LGBTQ+ communities in Finland. *Scandinavian Political Studies*, 47, 199–231. https://doi.org/10.1111/1467-9477.12270

How Can We Use Topic Modeling for a Quantitative Study?

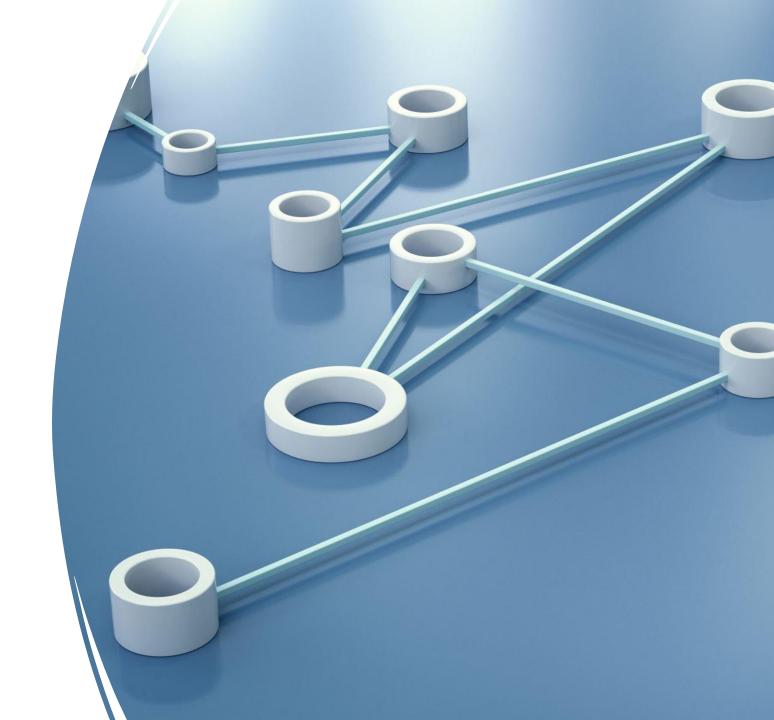
Extract Top Representative Texts for Each Topic

- After modeling, retrieve the top 20 most representative posts for each topic.
- Enables qualitative thematic analysis to interpret meaning behind topics.
 - Example: Identifying the core arguments in pro-gun control vs. pro-mental health policies.

Measure Topic Distribution Across Different Groups

- Compare topic prevalence between different demographics, locations, or time periods.
 - Example: Analyzing how public opinion shifts on school shootings before vs. after a major event.

Network Analysis



Who Really Spreads Information?

Why This Matters:

- Some posts reach millions, while others fade into obscurity.
- Information can spread **organically** (interest-driven) or be **amplified** (by influencers, bots, or coordinated campaigns).
- Understanding how content moves helps us analyze **misinformation**, activism, and public opinion shifts.

Key Factors in Information Flow:

- Shares/Retweets Signals content virality.
- Likes/Favorites Measures endorsement but doesn't spread content.
- Mentions (@username) Directs attention and influences visibility.

The Mechanisms of Information Spread

Retweet & Share Networks

A **retweet/share** creates a **cascade**: one user amplifies a message, triggering further reshares.

Some tweets **go viral quickly**, while others **build momentum over time**.

Key Metrics: Retweet trees, chain lengths, and central nodes.

Mention Networks (@username)

Mentions direct conversations and attention to specific users.

Influencers, journalists, or politicians can amplify discussions.

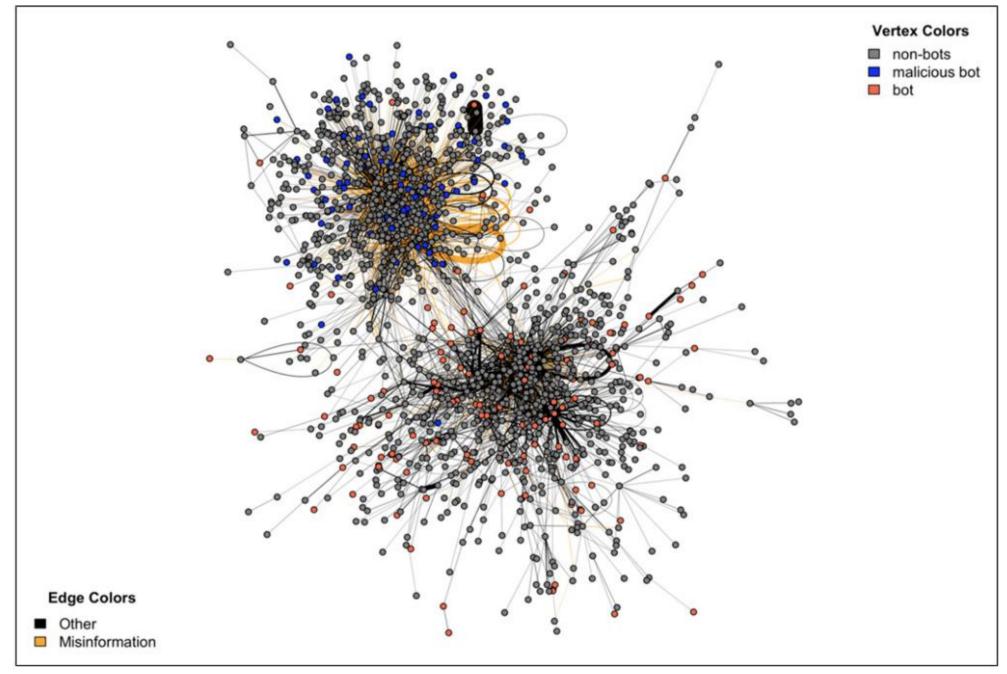
Key Metric: Centrality—who is mentioned most?

Like Networks (Engagement vs. Influence)

Likes signal **endorsement** but don't contribute to **content spread**.

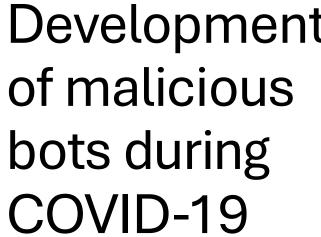
High likes + low shares = **passive agreement** (e.g., a controversial tweet people agree with but don't amplify).

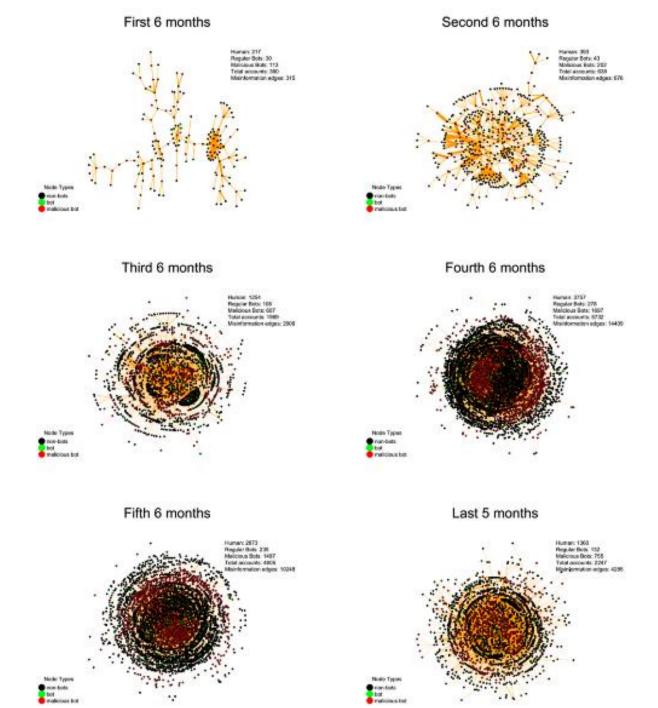
Key Metric: Engagement ratio (likes vs. shares).



Unlu, A., Truong, S., Sawhney, N., & Tammi, T. (2024). Setting the misinformation agenda: Modeling COVID-19 narratives in Twitter communities. New Media & Society, 0(0). https://doi.org/10.1177/14614448241232079

Development





Unlu, A., Lac, T., Sawhney, N., and Tammi, T. (2024) Unveiling the Veiled Threat: The Impact of Bots on COVID-19 Health Communication. Social Science Computer Review. https://doi.org/10.1177/08944393241275641

Who Shapes Conversations on Social Media?

Social media networks are **not random**—they often form **clusters** of like-minded users.

Some groups **only interact with each other** (*echo chambers*), while others are **bridges** between different perspectives.

 How can we map these structures and understand who influences discussions?

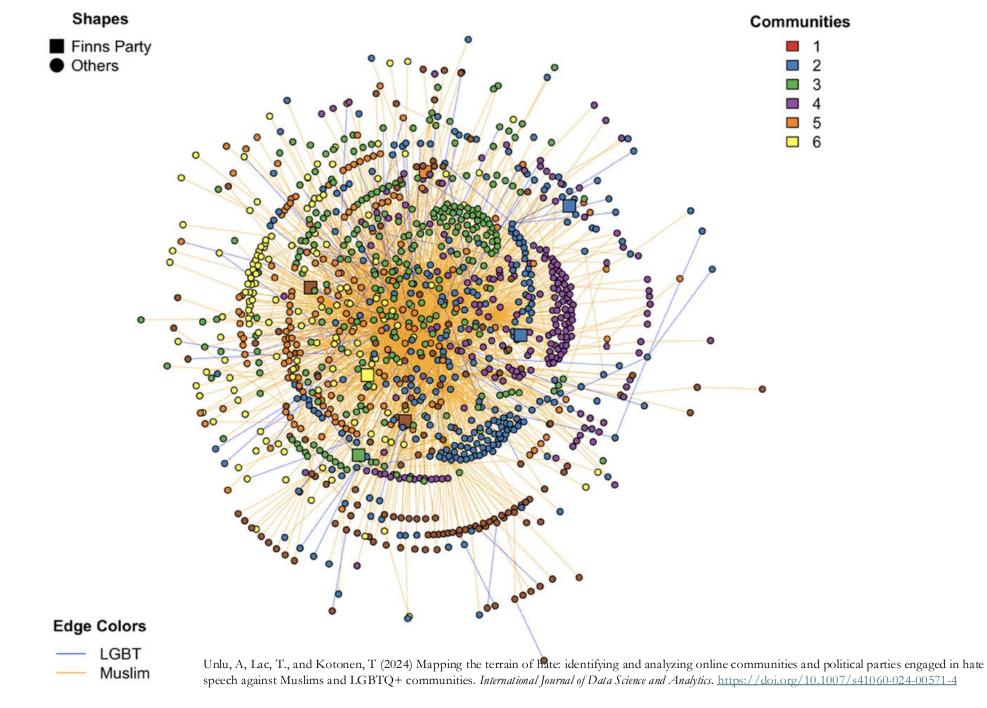
Community Detection in Social Media Networks

What is Community Detection?

- Identifies **clusters of users** that interact more with each other than with outsiders.
- Helps uncover echo chambers, interest groups, and opinion leaders.

How to Interpret Community Structures?

- Densely connected communities → Echo chambers with strong internal agreement.
- Bridges between communities → Users who connect different viewpoints.
- Influencers within communities → Key voices that drive narratives.



Which Communities Talk About What?

What is Correspondence Analysis?

- A statistical technique that maps relationships between two categorical variables.
- Helps visualize which communities are linked to which discussion topics.

How Correspondence Analysis Helps in Social Media Analysis

Maps Communities to Their Dominant Topics

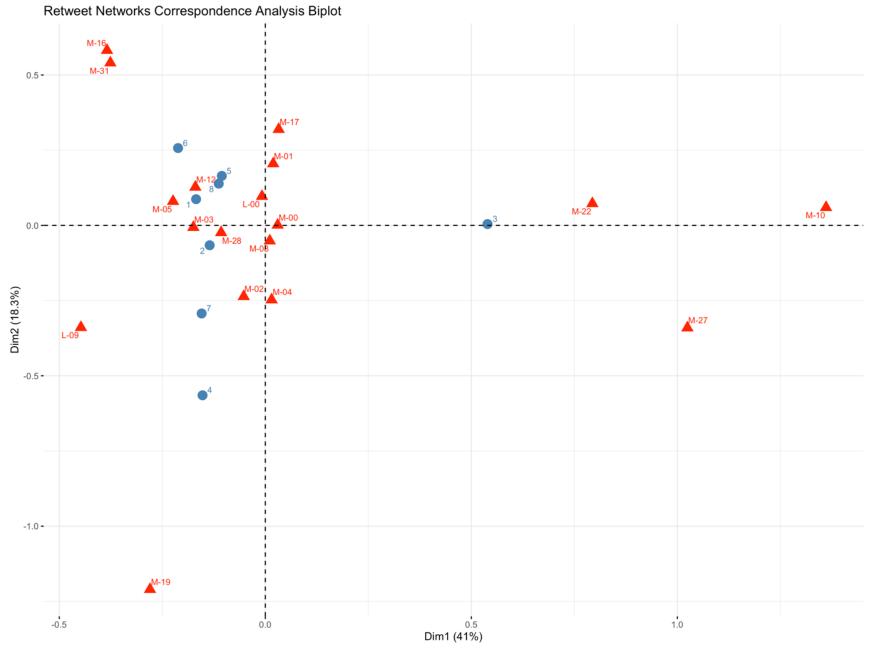
• Helps identify which groups discuss specific themes the most.

Identifies Shared & Overlapping Topics

• Shows if communities are **completely separate** or if they share **common discussion points**.

Finds Bridge Topics That Connect Communities

• Some topics act as **common ground** between polarized groups.



Unlu, A, Lac, T., and Kotonen, T (2024) Mapping the terrain of hate: identifying and analyzing online communities and political parties engaged in hate speech against Muslims and LGBTQ+ communities. *International Journal of Data Science and Analytics*. https://doi.org/10.1007/s41060-024-00571-4.

How Do Media and Social Media Shape Each Other?



Do traditional news outlets set the agenda for social media discussions?



Or does social media push certain topics into the mainstream?



Which influencers act as **bridges** between these two spaces?

Who Sets the Agenda? Social Media vs. News Media

Intermedia Agenda Setting

Traditional → Social Media (Mainstream-Led Agenda Setting)

- News coverage drives online conversations.
- Social media users react to published news articles.

Social Media → Traditional (Social-Led Agenda Setting)

- Grassroots trends influence news coverage.
- Journalists pick up viral social media discussions.

Mutual Influence (Hybrid Model)

- Social media amplifies news, while news legitimizes social media discussions.
- Key influencers act as **bridges** between the two.

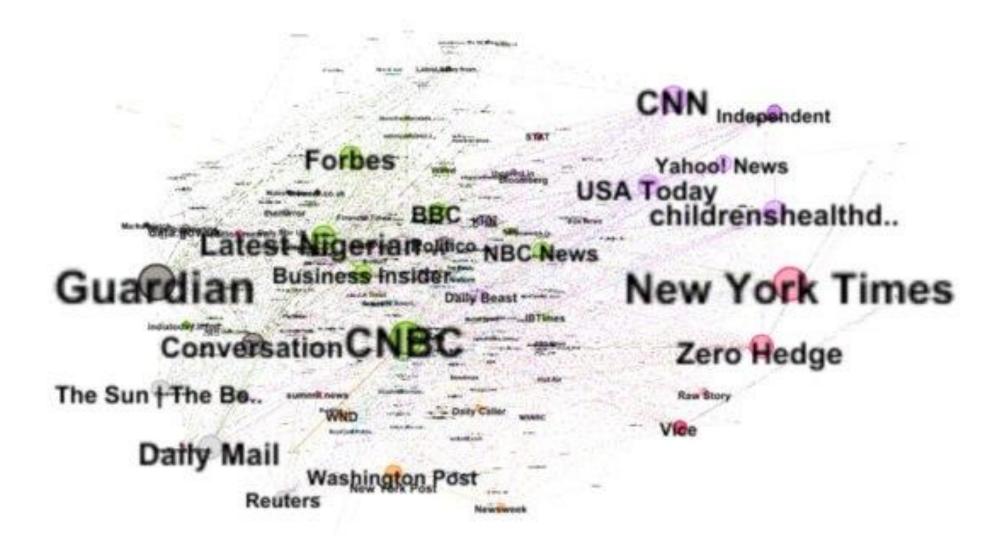


Figure 3. Intermedia Agenda Setting in COVID-19 Vaccine Digital Networks structure.



Case Study: LLM in Social Media Research

Generative Al

Analyzing Reddit Posts for CVP Strategies

Connect Young People to

Caring Adults and Activities

Intervene to Lessen Harms

and Prevent Future Risk

Mentoring programs

experiences with violence

Treatment to lessen the harms of violence
 Treatment to prevent problem behaviors and further

Hospital-based violence intervention programs

Street outreach and community norm change
 Community-justice partnerships

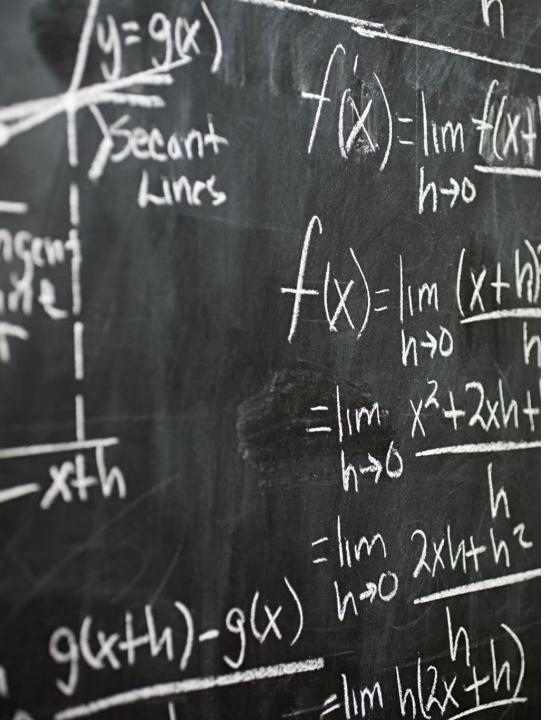
Goal: Search Reddit comments related to seven violence prevention strategies suggested by Center for Disease Control and Prevention (CDC)

- Encode each comment using sentence embedding
- To represent each strategy:
 - Ask ChatGPT to write a Reddit Comment (Cons/Pros)
 - Embedded the generated sentence with sentence transformers



Strategies

```
strategies = [
       ('economic',
        "We can talk about reducing violence without addressing the economic inequities that fuel it. Programs that provide job training, financial support, and affordable childcare can uplift entire communities.",
       "Throwing money at the problem doesn the problem
      ('education',
       "Investing in early childhood education and programs that support kids through elementary school doesn to just boost academic success—it also helps reduce the likelihood of violence as they grow.",
       "Education alone won't solve violence. Plenty of well-educated individuals still commit crimes, and these programs can only do so much without addressing larger systemic issues."),
       ('protective',
       "Parks and recreational spaces that are well-maintained and supervised provide safe places for kids and families to gather, keeping them away from harmful environments.",
       "Protective environments sound nice in theory, but modifying neighborhoods or public spaces can displace problems rather than solve them. Violence doesn t just vanish because a park gets a facelift."),
      ('family',
       "Family dynamics play such a huge role in shaping behavior. Teaching positive communication and conflict resolution within families can help prevent violence before it starts.",
       "Not every family can or wants to participate in these programs, and promoting 'healthy relationships' often ignores broader societal influences that lead to violence."),
       "Helping young people build emotional, social, and career skills is so effective. When youth feel confident and capable, they re less likely to resort to violence and more likely to succeed.",
        "Teaching skills is great, but it assumes that kids will use them instead of falling into negative influences. Without addressing those external factors, this approach feels superficial."),
      ('adults',
        "Mentorship programs are amazing for keeping kids on the right path. Even just one supportive adult can make a world of difference in how they see themselves and their future.",
       "Mentoring and after-school programs aren to a guaranteed fix. Many young people still face overwhelming pressures from their environments that even the best role model can to mitigate."),
      ('therapy',
       "Access to mental health resources and trauma-informed care is crucial for breaking cycles of violence. We need to address the root causes, not just the symptoms, to create lasting change.",
       "Focusing on intervention rather than prevention often means you re too late. By the time someone needs this kind of help, the damage is already done in many cases.")
```



Finding Posts Similar to Statements

- Find posts similar to the given statements, based on embeddings.
 - Cosine similarity
 - Pro/con similarity

Results

	body	pro_economic
708210	There is no one solution to this level of violence. Economic development, increased funding for schools/communities/youth programs, job training and job placement, criminal justice reform, increased funding for mental health and addiction treatment, all are areas that need attention to help address the issues Indianapolis is experiencing. But if you want to see this happen, you MUST be involved in your local government. Attend a neighborhood association meeting, meet your city council person, attend community meetingsmake your presence heard and known as a voting constituent. If you are tired of the violence, be a part of the solution and start as local as you can, right in your own neighborhood. We can tackle this issue if we come together as a community and say enough is enough.	0.774070
84483	You can't fix poverty like that without government grants. But education and opportunities to leave the area will probably lower violence	0.757582
1480394	Let's address another myth: that violence can be solved with better social services and economics.\n\nWhile it's true a lot of violence can happen due to family instability and bad conditions of children growing up, and a lot of that comes from lack of social safeties and economic issues, that does not explain all violence.\n\nThere has been an increase in hate motivated violence, and an increase in random unprovoked attacks due to mental issues.\n\nThere is also violence caused by violence. A family that is victimized and destabilized by violence (either by criminals or by violent cops) will breed more violence.\n\nThose all have complex issues that require deep work to address. We must not shy away from them.\n\nHowever, if wealth and access to resources prevented crimes, there would be no crimes committed by rich people.\n\nAnd when the violence is perpetrated by repeat offenders, it's very obvious how to prevent that kind of violence. And that's why it's so sad when it keeps happening.	0.749361
235113	That's not how violence works. The solutions are broad, and take a long time. Looking for immediate and impactful solutions leads people to pour money into one area, and get frustrated when there is no fast or visible change. \n\nThe programs that are making a difference are ones like READI which provides likely victims or perpetrators of violence with jobs, counseling, and therapy for 18 months. The idea is that instead of spending money on them once they're in the hospital with a gunshot wound, or in jail with a gun charge, take that same money and invest it in that person's future before it takes a turn for the worse. https://www.heartlandalliance.org/readi/\n\nOther community programs that have had successes reducing violence depend on developing deep, long lasting relationships in those neighborhoods. It's not fast, and it's not even immediately apparent. https://www.chicagotribune.com/opinion/commentary/ct-opinion-gun-violence-reaching-young-men-20200701-7jnshhzsfzaxvlcnxugcjikb7u-story.html\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n	0.732587
139512	Tackle poverty and violence will decrease.	0.729696

	body	con_education
1647941	Right, but education does not solve for the current batch of crime. And education without support at home still doesn't work.	0.761568
1619020	This all starts with better schools and an emphasis on public education/programming. It won't fix crime overnight, it'll frankly take a long fucking time, nor is it the *only* issue but if you give people tools and opportunities to do things with their life, they'll be less likely to turn to crime.\n\nThe state of SLPS in general and especially some of the schools in the areas with the highest crime areas is abysmal. Deteriorating schools, lack of resources/supplies, underfunded educators, lack of educators, zero resources for low-income families, etc. \n\nIt is a hard fight especially when the GOP has attacking public education in a decades long endeavor. Dem leaders at the local tend to toss money at quick fixes (like magnet/charter schools) which end up doing greater harm to the overall system and perpetuate inequality/crime.\n\nI want the lion's share of the Kroneke settlement money (if it ever gets used) to go into SLPS. Pay teachers more, hire more teachers, fix the buildings, reopen the closed ones, and show kids throughout the city that we give a shit about them.	0.732826
1617785	This on top of fixing education, job opportunities and social programs to help families. Middle class people aren't violent criminals.	0.713310
347959	Politicians seem to be obsessed with putting a lid on the violence but never address how you give people better alternatives. \n\nApparently high quality education which would do that is not well liked at least in this sub.	0.710728
1324283	You don't need a education or money to know that violence and crime is bad. You're just making excuses why the crime is so lopsided. \n\nSo as long as people like you make these people out to be victims they'll continue their destructive behavior. \n\n	0.710373
708210	There is no one solution to this level of violence. Economic development, increased funding for schools/communities/youth programs, job training and job placement, criminal justice reform, increased funding for mental health and addiction treatment, all are areas that need attention to help address the issues Indianapolis is experiencing. But if you want to see this happen, you MUST be involved in your local government. Attend a neighborhood association meeting, meet your city council person, attend community meetingsmake your presence heard and known as a voting constituent. If you are tired of the violence, be a part of the solution and start as local as you can, right in your own neighborhood. We can tackle this issue if we come together as a community and say enough is enough.	0.706247
403667	Investing in public housing and education for the poor will do more to stop crime than anything else.	0.698652
56767	Yeah, I have no idea, tbh. \n\nI'm fully in favor of cracking down on violent crime in the meantime, though. At the very least hopefully we can stop some of them from catching a stray bullet and/or improve the schools around them if those neighborhoods are able to improve.	0 695909

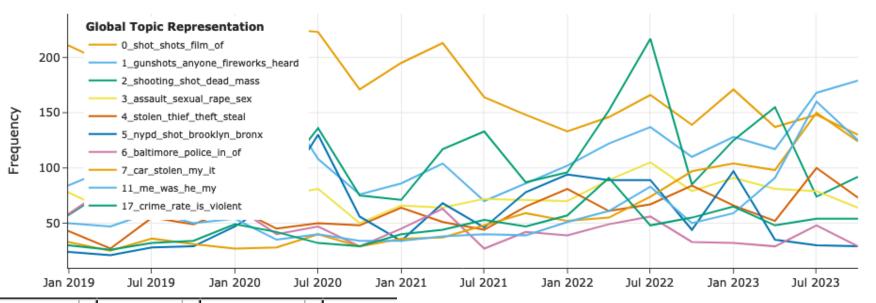
Validation layer

1	J	K	L	М	N	0	Р
pro_economic	pro_economic_gpt_agree	body					
0.771879315	YES	There is no	one solutio	n to this lev	el of violen	ce. Econom	ic develo
0.757582426	YES	You can't fi	x poverty li	ke that with	out govern	ment grants	. But ed
0.749360621	NO	Let's addre	ss another r	nyth: that v	iolence can	be solved w	ith bette
0.732586682	YES	That's not h	now violenc	e works. Th	e solutions a	are broad, a	nd take a
0.729696274	YES	Tackle pov	erty and vio	lence will d	ecrease.		
0.721635044	YES	This is why	I have issue	s with peop	le who do n	ot support f	inancial
0.71554625	YES	I think this	is only to a	ddress the ir	mmediate c	oncerns of v	iolence 1
0.71085602	NO	it's proven	that comm	unities w m	ore outreac	h programs	and acti
0.701586127	YES	I agree. It's	time to stop	the violen	ce. Let's star	t funding th	e comm
0.700724244	YES	But investi	ng in educa	tion, after s	chool progr	ams, social p	orogram
0.700717509	NO	I'd love mo	re violence	prevention	programs, b	ut where's t	he mon
0.695720613	NO	What the h	ell are you t	alking abou	ıt? If you wa	ant to cite e	conomic

- Do the following two statements express similar opinions about ways to address violence.
 - Report (declared) statement
 - Comment statement
 - Yes/No

Topics over Time

Results



Name Box name ▼	pro_economic ▼	pro_education ▼	pro_protectiv ▼	pro_family ▼	pro_skills
83_violence_gun_march_rally	0.215906226	0.063271994	0.01499874	0.053188808	0.064280
carjackings_chicago_carjacking_cwb	0.175315344	0.066262175	0.020277822	0.040076641	0.077917
_school_high_students_student	0.107422267	0.117652959	0.019759278	0.043430291	0.071915
91_violent_crime_mayor_total	0.190330008	0.033768227	0.016884114	0.023791251	0.038219
52_louis_st_cities_murder	0.175266817	0.031920102	0.026436894	0.026045236	0.03387
31_violence_gang_violent_black	0.147919877	0.027664939	0.02122146	0.033688192	0.034178
17_crime_rate_is_violent	0.158364558	0.023804276	0.025934906	0.016273602	0.029608
6_baltimore_police_in_of	0.137199198	0.03450869	0.01453877	0.024941511	0.037516
acking_carjackings_attempted_carjack	0.113397204	0.038898487	0.013035967	0.027066646	0.053190
_chicago_police_shot_weekend	0.133231435	0.028085594	0.012960132	0.029836963	0.03477
14_old_year_shot_girl	0.10599322	0.036077504	0.014742015	0.031816627	0.049482
77_adams_eric_mayor_nyc	0.162947415	0.016460451	0.012262483	0.013367212	0.022757
5_lightfoot_lori_mayor_chicago	0.137473374	0.023922661	0.012289038	0.021792561	0.028838
86_orleans_new_nola_the	0.106389391	0.034358047	0.019891501	0.022403054	0.036769
19_york_nyc_new_crime	0.141656004	0.013982818	0.012977518	0.014485469	0.023121
49_baltimore_city_and_to	0.064083458	0.02251157	0.10518472	0.007216252	0.015295
44_the_of_that_to	0.108949743	0.026712832	0.011289476	0.020459553	0.035715
5_nypd_shot_brooklyn_bronx	0.114827708	0.013218304	0.011755209	0.020483326	0.024519
10_stabbed_death_man_fatally	0.09869863	0.012226027	0.021643836	0.01619863	0.023253
6_chicago_violence_crime_gun	0.110381228	0.018932331	0.022661427	0.017125154	0.02332

Q&A or Discussion

Where are the **theoretical** or **methodological gaps**?

How do you envision future research with these tools?

