University at Buffalo
College of Arts and Sciences

A Tale of Vaccination Debates & Public Responses: *Data-Driven Insights from a Multi-medium Exploration*

Presenter: Qingqing Chen
Advisor: Dr. Andrew Crooks
Department of Geography
Jul 20, 2023 @ PIPP Modeling Workshop

PART I: Vaccine Sentiments

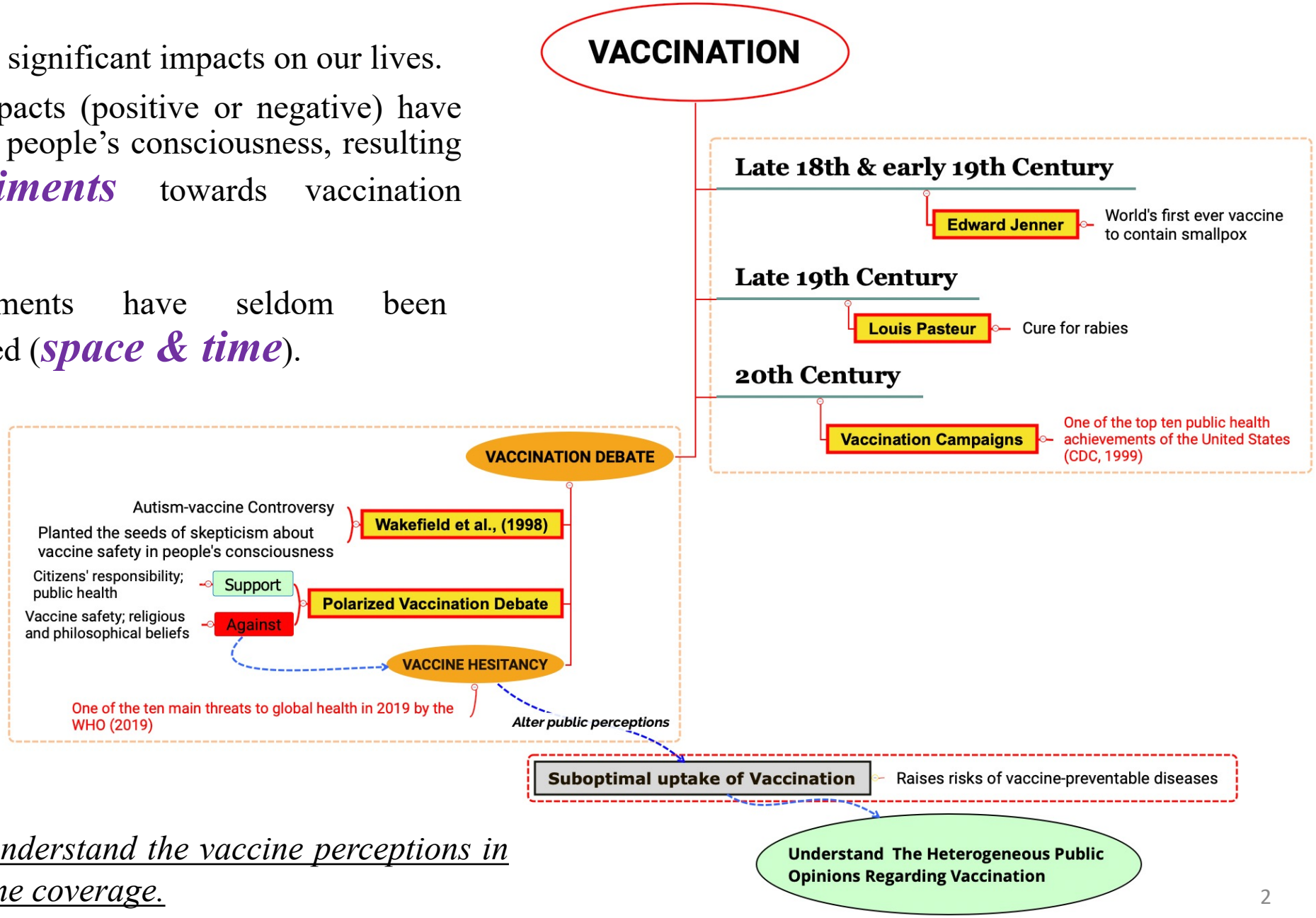
(Social Media)

Problem Statement

Disease outbreaks have significant impacts on our lives.

The accompanying impacts (positive or negative) have gradually taken root in people's consciousness, resulting in ***diverse sentiments*** towards vaccination campaigns.

The diverse sentiments have seldom been systematically quantified (***space & time***).



It becomes crucial to understand the vaccine perceptions in order to enhance vaccine coverage.

Analyze ***dynamic vaccine sentiments*** in ***SPACE*** and ***TIME*** based on Social Media Data over a long-term period by leveraging Machine Learning techniques.

Research questions:

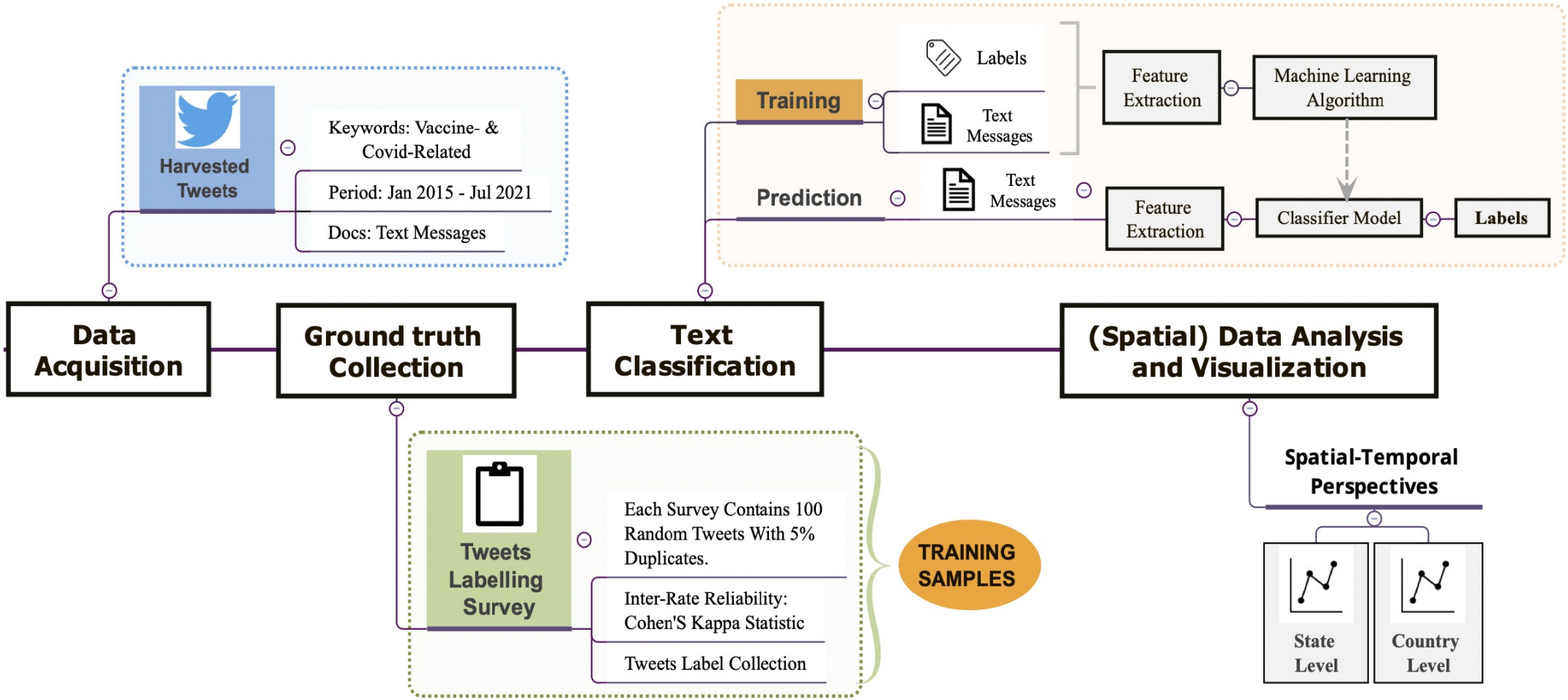
- What is the dominant vaccine sentiment before and after the outbreak?
- Did vaccine sentiment change over time and where did such changes take place?
- What are the relationships between different vaccine sentiments and the actual vaccination rates?

Direct policy implications:

- The effectiveness of strategies for enhancing vaccine uptake and immunization coverage;
- The psychological, social, and political factors that sustain public trust in vaccines.

Analyze ***dynamic vaccine sentiments*** in ***SPACE*** and ***TIME*** based on Social Media Data over a long-term period by leveraging Machine Learning techniques.

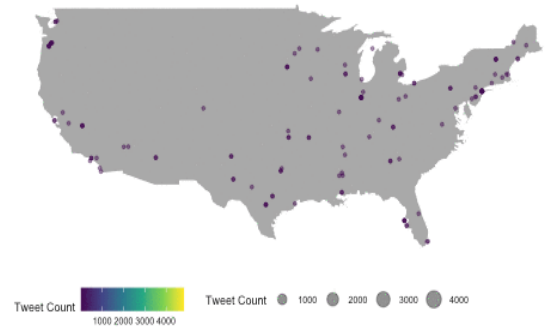
Research Outline:



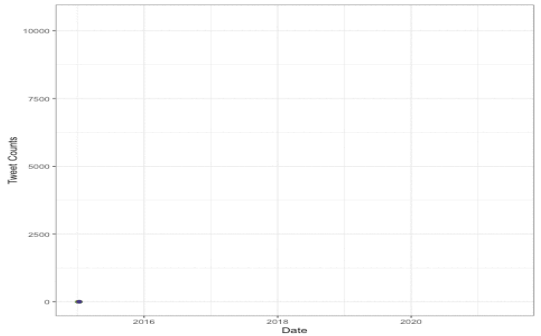
Twitter Data

- Keywords: Vaccination and their derivatives (e.g., “vaccine”, “vaxx”, “vaxine”)
- Jan 2015 – July 2021 (~ 12 million tweets sent by 2 million users)

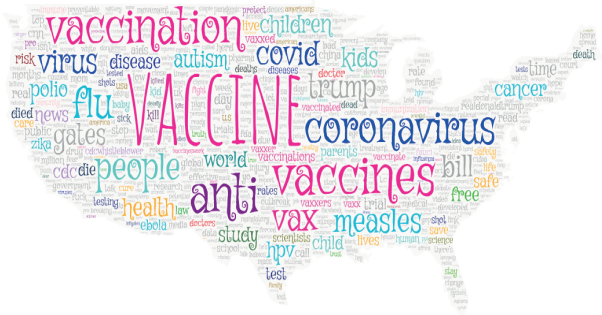
Spatial



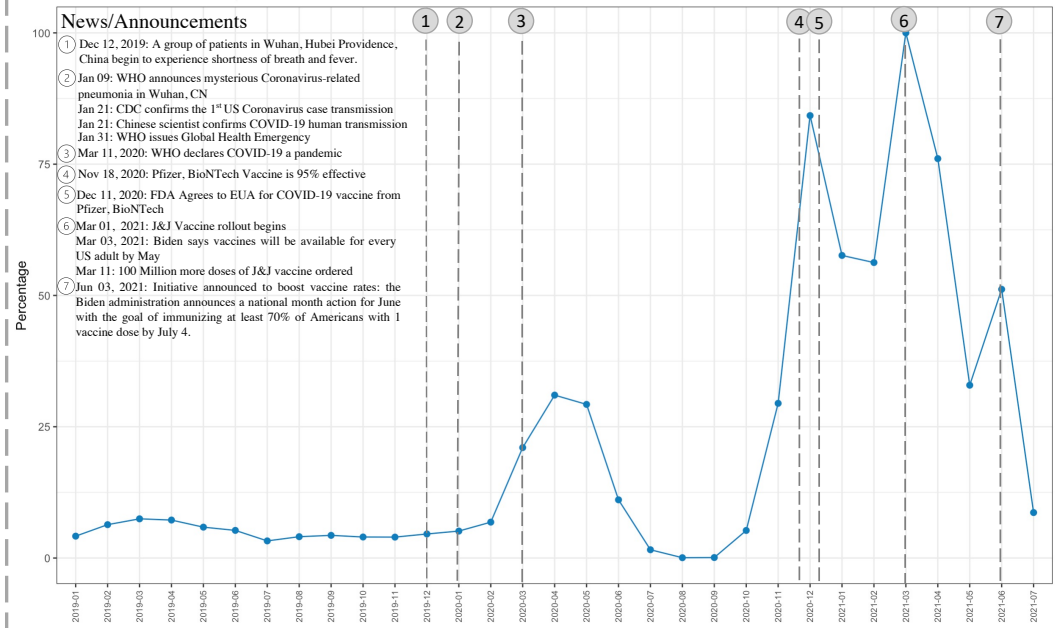
Temporal

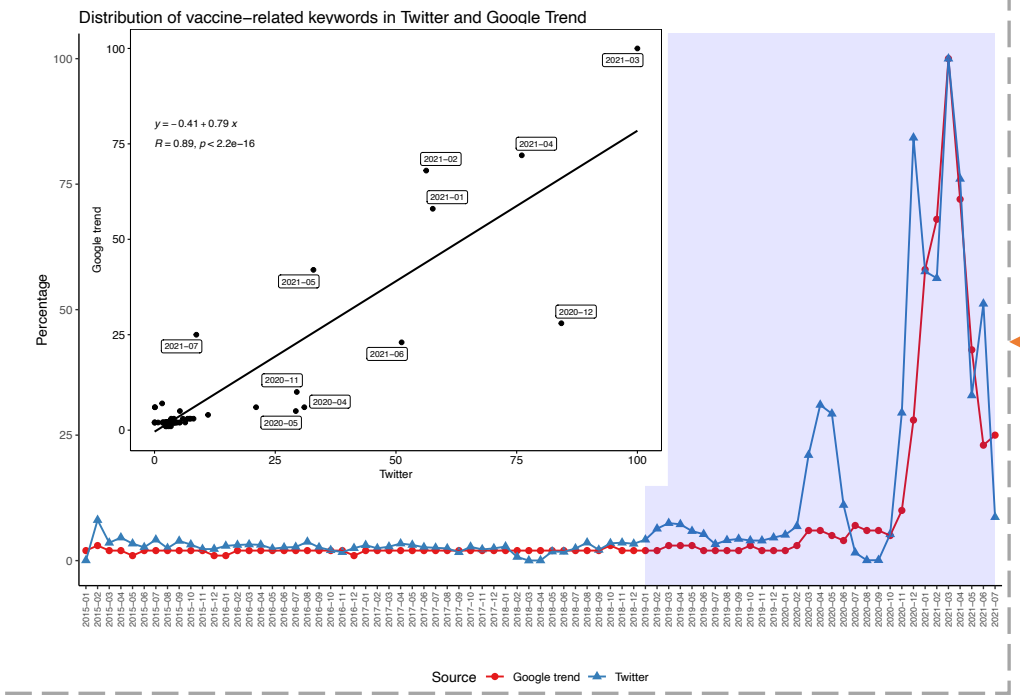


Textual



Sanity Check

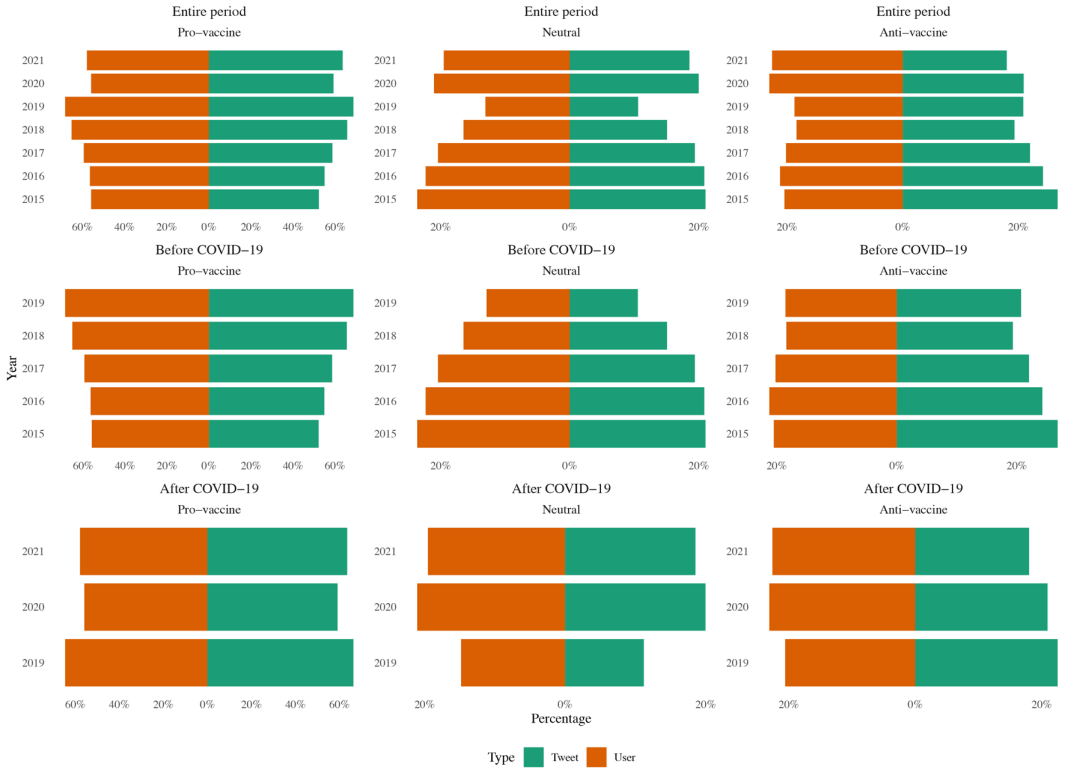




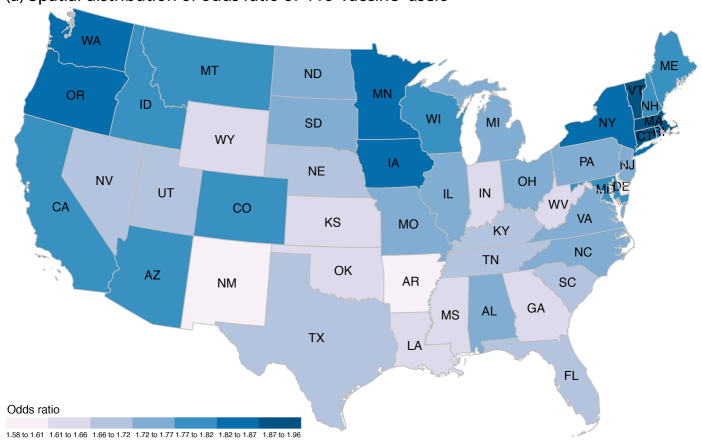
Sentiments	Number of Users	Number of Tweets
Entire period		
Pro-vaccine	2,055,959 (56.97%)	7,190,846 (61.58%)
Neutral	729,868 (20.22%)	2,158,271 (18.48%)
Anti-vaccine	822,926 (22.8%)	2,327,495 (19.93%)
Before COVID-19		
Pro-vaccine	544,365 (61.56%)	1,655,642 (60.56%)
Neutral	161,609 (18.28%)	457,925 (16.75%)
Anti-vaccine	178,339 (20.17%)	620,103 (22.68%)
After COVID-19		
Pro-vaccine	1,631,444 (56.2%)	5,535,204 (61.89%)
Neutral	595,655 (20.52%)	1,700,346 (19.01%)
Anti-vaccine	675,706 (23.28%)	1,707,392 (19.09%)

- The positive vaccine sentiment was the dominant opinion
- The rate of "Pro-vaccine" users decreased after the outbreak (61.56% → 56.20%)
- The percentage of "Anti-vaccine" users revealed a modest increment after the outbreak (20.17% → 23.28%).
- The outbreak indeed moderately shifted public attitudes towards vaccination.

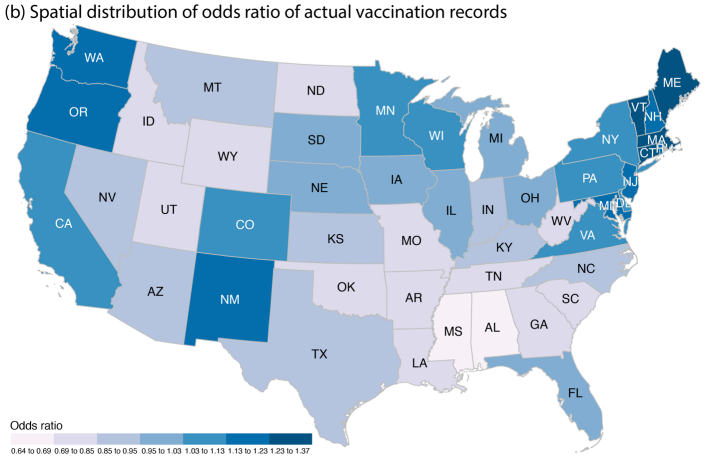
- The rate of "Anti-vaccine" users approached the highest point in 2020, then slightly shrank in 2021;
- The uptake of the coronavirus vaccine(s) in some cases is accompanied by various side effects;
- Coincided with Yousefinaghani et al.'s (2021) results



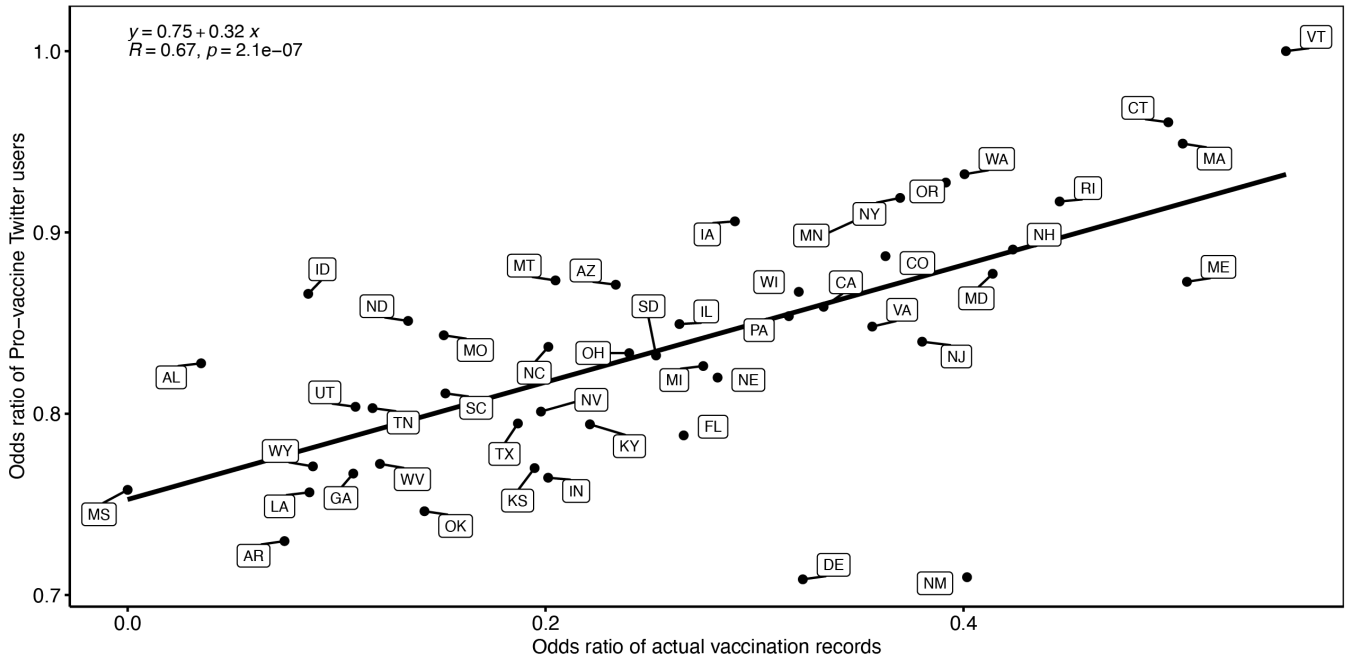
(a) Spatial distribution of odds ratio of "Pro-vaccine" users



(b) Spatial distribution of odds ratio of actual vaccination records



(c) Correlation between "Por-vaccine" users and the actual vaccination records



Online v.s. Offline

Online: *"Pro-vaccine" users online*
Offline: *Actual vaccination rate [Our World in Data (Ritchie et al., 2020)]*

- Odds ratio: to alleviate size-related issues;
- Geographic difference in Pro-vaccine sentiment on Twitter: MA, CT, VT, CO, WA, NY had relatively higher Pro-vaccine odds than other states → relatively complete health system;
- Follow a similar trend to that of the actual vaccination rate;
- A positive correlation (R = 0.67)

The proposed approach for identifying positive vaccine sentiments online can be used as an indicator for evaluating offline vaccination rates.

PART II: Vaccine Comparison

(Social Media)

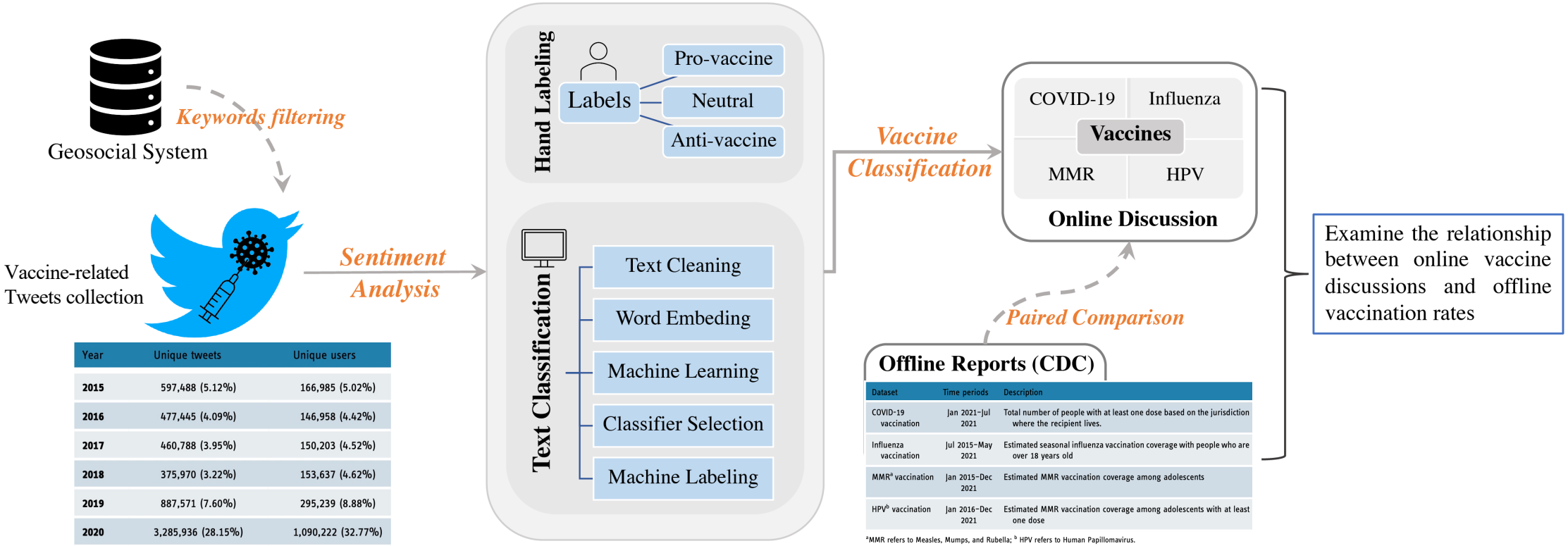
Basic premise: social media can impact human behavior.

- This use of online platforms not only impacts the vaccination discussion itself;
- but also more broadly the way by which vaccine related information is consumed and produced.

Understanding the dynamics of public attention among competing themes for attention in the context of vaccination.

- How public attention be reshaped/allocated when different public health issues are prominent on social media?
- How public attention compares to actual vaccination rates?

- **Four vaccines**: COVID-19, Influenza, Measles, Mumps, and Rubella (MMR), Human Papillomavirus (HPV)
  - COVID-19: impacts on the current world order
  - The other three: the robust nature of the debate around them in recent years before the emergence of COVID-19
  - Public availability of authoritative data on their respective vaccination rates.



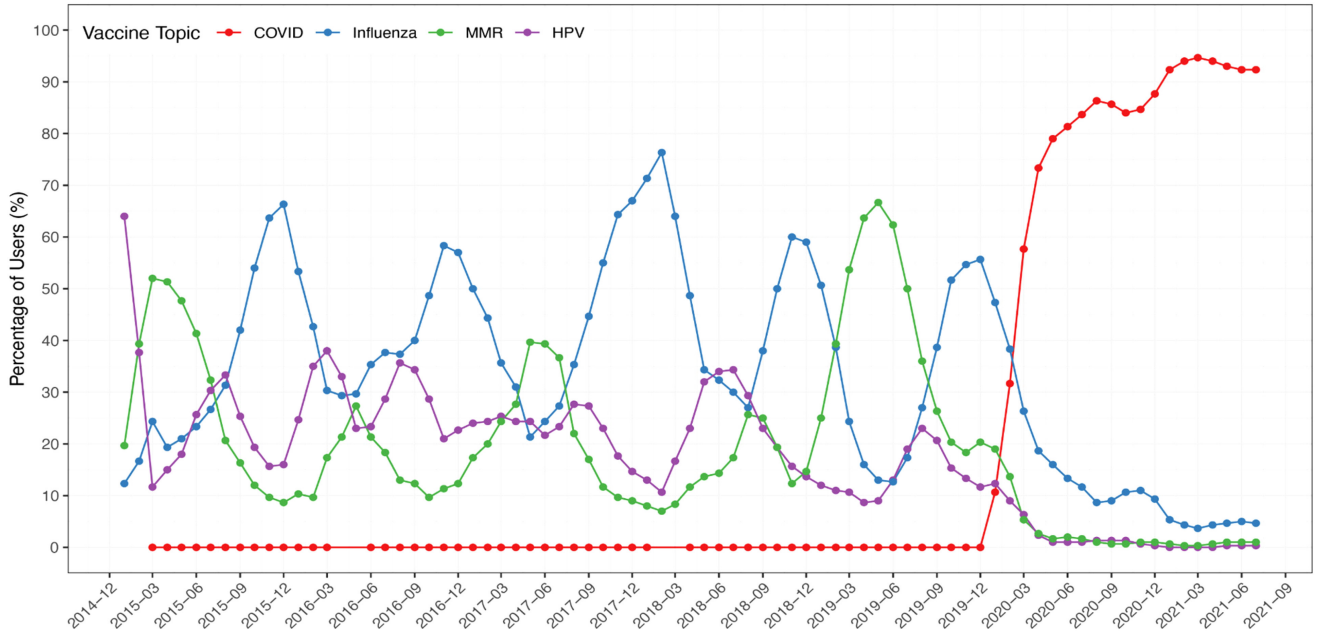
Year	Unique tweets	Unique users
2015	597,488 (5.12%)	166,985 (5.02%)
2016	477,445 (4.09%)	146,958 (4.42%)
2017	460,788 (3.95%)	150,203 (4.52%)
2018	375,970 (3.22%)	153,637 (4.62%)
2019	887,571 (7.60%)	295,239 (8.88%)
2020	3,285,936 (28.15%)	1,090,222 (32.77%)
2021	5,589,713 (47.88%)	1,323,331 (39.78%)

Dataset	Time periods	Description
COVID-19 vaccination	Jan 2021–Jul 2021	Total number of people with at least one dose based on the jurisdiction where the recipient lives.
Influenza vaccination	Jul 2015–May 2021	Estimated seasonal influenza vaccination coverage with people who are over 18 years old
MMR[a] vaccination	Jan 2015–Dec 2021	Estimated MMR vaccination coverage among adolescents
HPV[b] vaccination	Jan 2016–Dec 2021	Estimated MMR vaccination coverage among adolescents with at least one dose

[a]MMR refers to Measles, Mumps, and Rubella; [b] HPV refers to Human Papillomavirus.

The quarterly distribution of percentage of users by different vaccine discussion (2015 to 2021)

Year Quarter	COVID-19	Influenza (Flu)	MMR	HPV
2015 Q1	0.0% (0)	15.7% (5413)	69.6% (24,035)	14.8% (5102)
2015 Q2	0.0% (0)	25.5% (5926)	47.7% (11,098)	26.8% (6237)
2015 Q3	0.0% (0)	53.2% (11,859)	18.4% (4109)	28.4% (6342)
2015 Q4	0.0% (0)	74.2% (22,667)	8.7% (2656)	17.1% (5232)
2016 Q1	0.0% (0)	34.9% (7643)	16.5% (3608)	48.6% (10,652)
2016 Q2	0.0% (0)	42.0% (6142)	26.8% (3929)	31.2% (4566)
2016 Q3	0.0% (0)	45.7% (9188)	12.4% (2492)	41.9% (8416)
2016 Q4	0.0% (0)	63.8% (13,486)	11.4% (2418)	24.7% (5226)
2017 Q1	0.0% (0)	41.8% (7355)	27.1% (4770)	31.1% (5468)
2017 Q2	0.0% (0)	26.8% (5677)	45.5% (9636)	27.7% (5879)
2017 Q3	0.0% (0)	50.3% (8590)	18.2% (3111)	31.5% (5387)
2017 Q4	0.0% (0)	73.4% (20,787)	10.0% (2840)	16.6% (4701)
2018 Q1	0.0% (0)	81.3% (14,690)	5.7% (1035)	12.9% (2338)
2018 Q2	0.0% (0)	39.0% (3057)	14.9% (1167)	46.1% (3611)
2018 Q3	0.0% (0)	44.9% (10,671)	26.2% (6233)	28.8% (6851)
2018 Q4	0.0% (0)	66.8% (28,262)	15.3% (6460)	18.0% (7598)
2019 Q1	0.0% (0)	25.3% (14,833)	62.1% (36,438)	12.6% (7375)
2019 Q2	0.0% (0)	12.1% (7260)	74.8% (45,050)	13.1% (7892)
2019 Q3	0.0% (0)	43.1% (13,461)	32.7% (10,202)	24.2% (7537)
2019 Q4	0.0% (0)	59.6% (28,851)	26.6% (12,884)	13.8% (6661)
2020 Q1	73.5% (135,566)	19.1% (35,224)	3.5% (6480)	3.8% (7061)
2020 Q2	86.5% (301,884)	11.5% (40,283)	1.3% (4556)	0.7% (2432)
2020 Q3	88.1% (8346)	10.1% (961)	1.0% (96)	0.8% (75)
2020 Q4	95.1% (681,480)	4.4% (31,297)	0.4% (2690)	0.2% (1398)
2021 Q1	97.4% (1,061,667)	2.1% (23,030)	0.2% (2551)	0.2% (2651)
2021 Q2	96.3% (644,512)	3.0% (19,862)	0.5% (3045)	0.5% (2022)
2021 Q3	96.1% (36,797)	3.2% (1211)	0.4% (161)	0.3% (127)



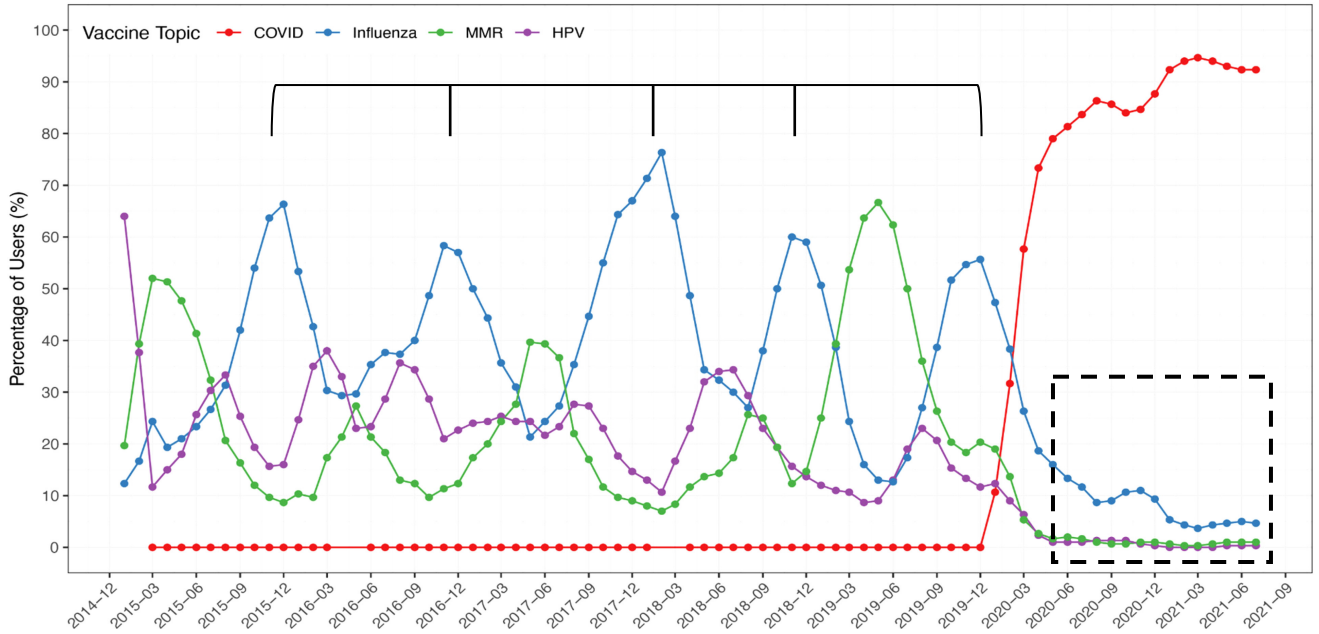
Shows how the public's attention, while being finite due to the zero-sum theory, switches from one vaccination to another over time.

- 1st half of 2015: MMR dominated (measles outbreaks in CA)
- 2nd half of 2015: Influenza during the winter period
- End of 2019: COVID-19 vaccine & maintained dominance until 2021

- The media has the capability in shaping people's agenda/priority of issues, that public attention is finite.
- The public is uncomfortable in new settings until they achieve some degree of orientation to their new surroundings.

The quarterly distribution of percentage of users by different vaccine discussion (2015 to 2021)

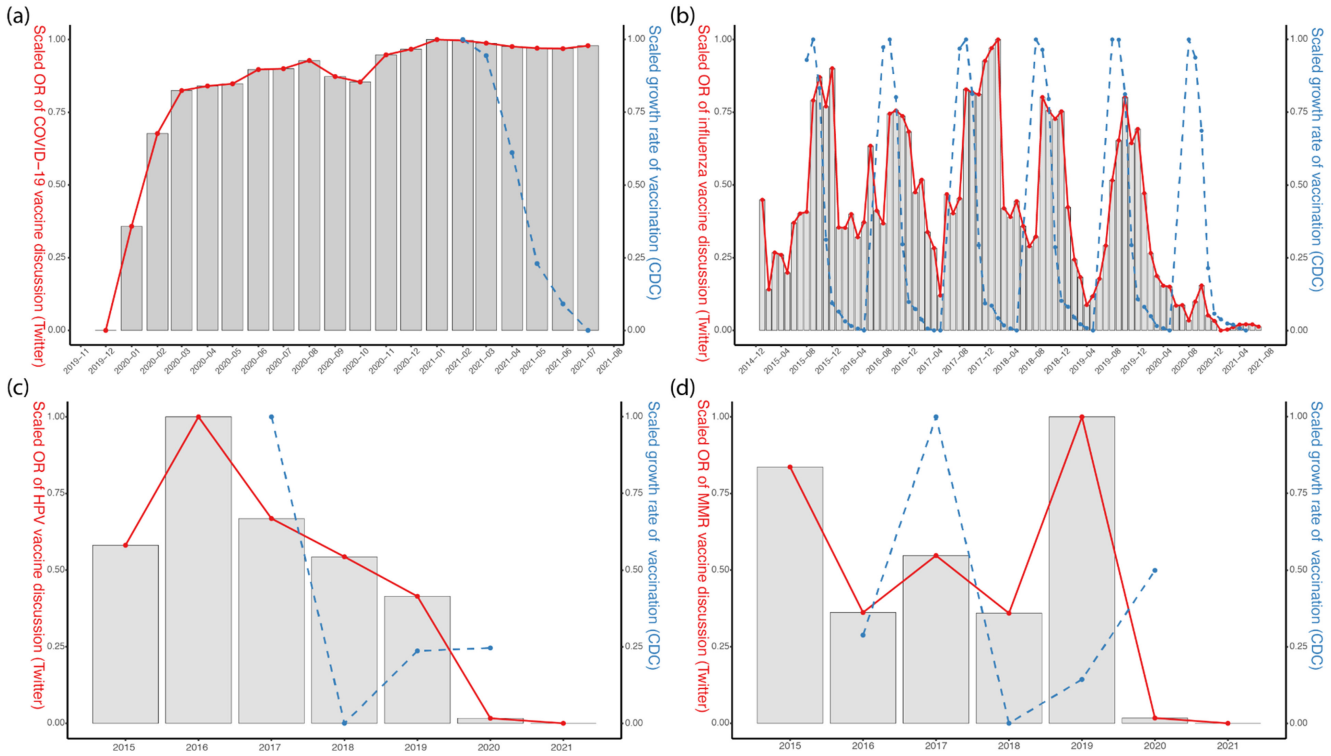
Year Quarter	COVID-19	Influenza (Flu)	MMR	HPV
2015 Q1	0.0% (0)	15.7% (5413)	69.6% (24,035)	14.8% (5102)
2015 Q2	0.0% (0)	25.5% (5926)	47.7% (11,098)	26.8% (6237)
2015 Q3	0.0% (0)	53.2% (11,859)	18.4% (4109)	28.4% (6342)
2015 Q4	0.0% (0)	74.2% (22,667)	8.7% (2656)	17.1% (5232)
2016 Q1	0.0% (0)	34.9% (7643)	16.5% (3608)	48.6% (10,652)
2016 Q2	0.0% (0)	42.0% (6142)	26.8% (3929)	31.2% (4566)
2016 Q3	0.0% (0)	45.7% (9188)	12.4% (2492)	41.9% (8416)
2016 Q4	0.0% (0)	63.8% (13,486)	11.4% (2418)	24.7% (5226)
2017 Q1	0.0% (0)	41.8% (7355)	27.1% (4770)	31.1% (5468)
2017 Q2	0.0% (0)	26.8% (5677)	45.5% (9636)	27.7% (5879)
2017 Q3	0.0% (0)	50.3% (8590)	18.2% (3111)	31.5% (5387)
2017 Q4	0.0% (0)	73.4% (20,787)	10.0% (2840)	16.6% (4701)
2018 Q1	0.0% (0)	81.3% (14,690)	5.7% (1035)	12.9% (2338)
2018 Q2	0.0% (0)	39.0% (3057)	14.9% (1167)	46.1% (3611)
2018 Q3	0.0% (0)	44.9% (10,671)	26.2% (6233)	28.8% (6851)
2018 Q4	0.0% (0)	66.8% (28,262)	15.3% (6460)	18.0% (7598)
2019 Q1	0.0% (0)	25.3% (14,833)	62.1% (36,438)	12.6% (7375)
2019 Q2	0.0% (0)	12.1% (7260)	74.8% (45,050)	13.1% (7892)
2019 Q3	0.0% (0)	43.1% (13,461)	32.7% (10,202)	24.2% (7537)
2019 Q4	0.0% (0)	59.6% (28,851)	26.6% (12,884)	13.8% (6661)
2020 Q1	73.5% (135,566)	19.1% (35,224)	3.5% (6480)	3.8% (7061)
2020 Q2	86.5% (301,884)	11.5% (40,283)	1.3% (4556)	0.7% (2432)
2020 Q3	88.1% (8346)	10.1% (961)	1.0% (96)	0.8% (75)
2020 Q4	95.1% (681,480)	4.4% (31,297)	0.4% (2690)	0.2% (1398)
2021 Q1	97.4% (1,061,667)	2.1% (23,030)	0.2% (2551)	0.2% (2651)
2021 Q2	96.3% (644,512)	3.0% (19,862)	0.5% (3045)	0.5% (2022)
2021 Q3	96.1% (36,797)	3.2% (1211)	0.4% (161)	0.3% (127)

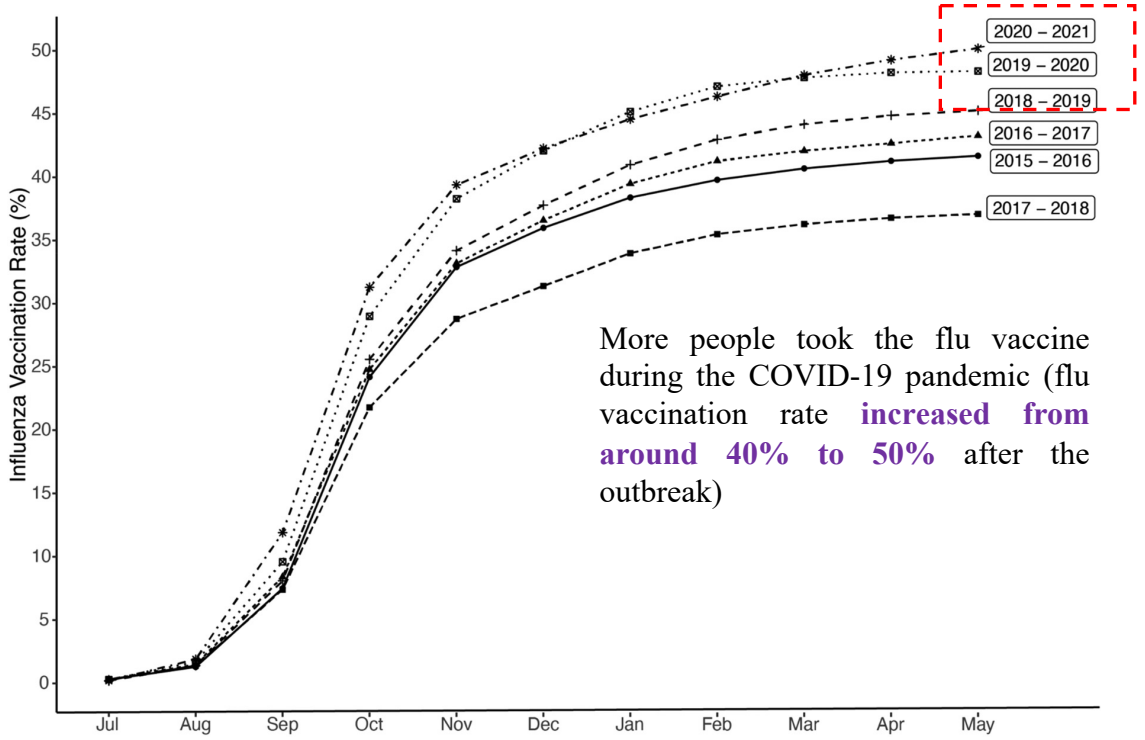


- Demonstrates a cyclical pattern, with peaks generally occurring during the winter flu seasons before COVID-19 outbreak.
- A smell peak during winter flu season in 2020 under the dominance of COVID-19 vaccination debate.

A potential association between COVID-19 and flu vaccines that may result from the perceived similarity between the two illnesses (e.g., similar symptoms)

Vaccine discussions on Twitter & growth rate of the actual vaccination rate from CDC (a) COVID-19; (b) Influenza; (c) HPV; (d) MMR.

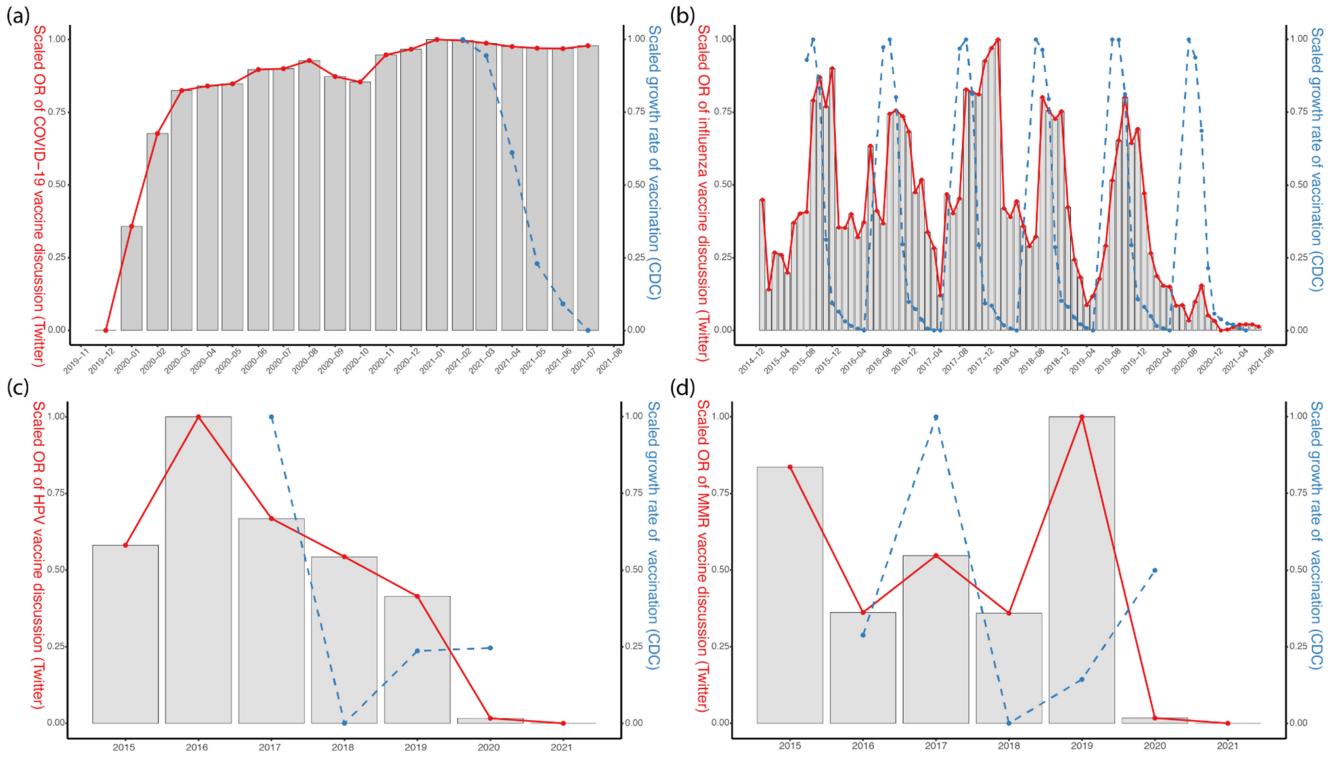


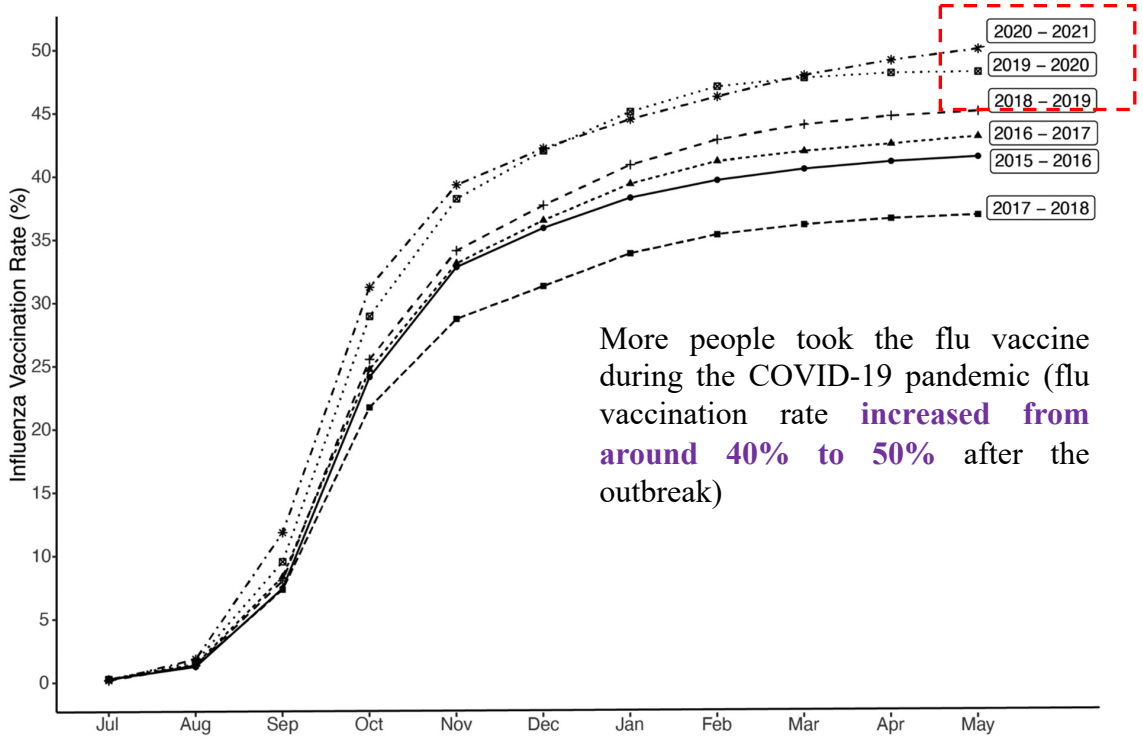


- The *HPV* and *MMR* vaccine rates are rather *volatile*, without any apparent patterns in their trends in Twitter and actual vaccination rates
- Do observe a *periodic change* in the *influenza* vaccine, peak rate of flu vaccinations emerges close to the peak of the flu vaccine discussion on Twitter

- The prominence of an issue (e.g., *COVID-19 vaccine* discussion) on social media has the potential to affect the public's behavior on another issue (e.g., *flu vaccine uptake*)
- Network Agenda Setting: the role of *cognitive components* in the process of representing reality
  - Information describes the symptoms or the vaccine

Vaccine discussions on Twitter & growth rate of the actual vaccination rate from CDC (a) COVID-19; (b) Influenza; (c) HPV; (d) MMR.





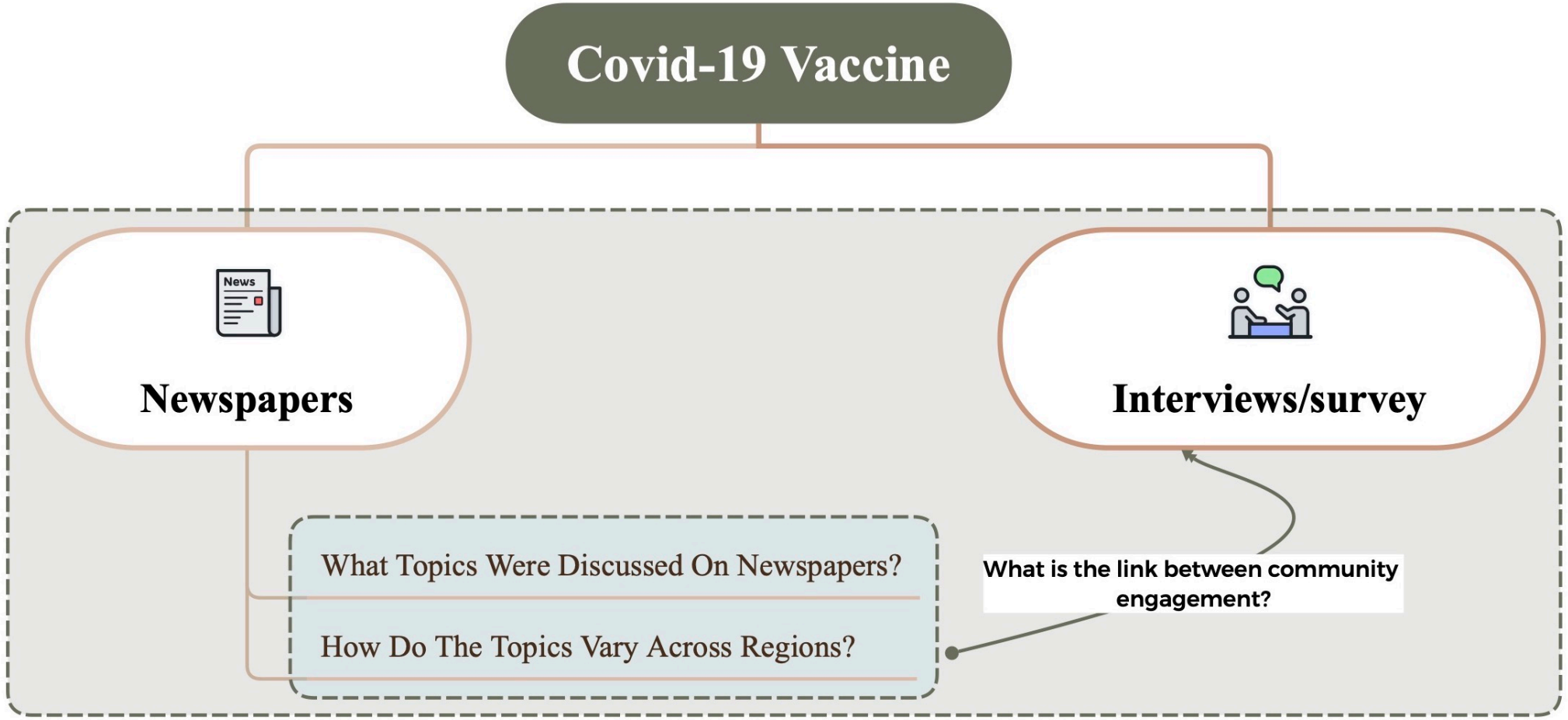
More people took the flu vaccine during the COVID-19 pandemic (flu vaccination rate **increased from around 40% to 50%** after the outbreak)

The more frequently two issues are associated within the media (e.g., COVID-19 & influenza share certain similarities), the more likely they are be perceived as interdependent on the public agenda

PART III: Vaccine Community Engagement

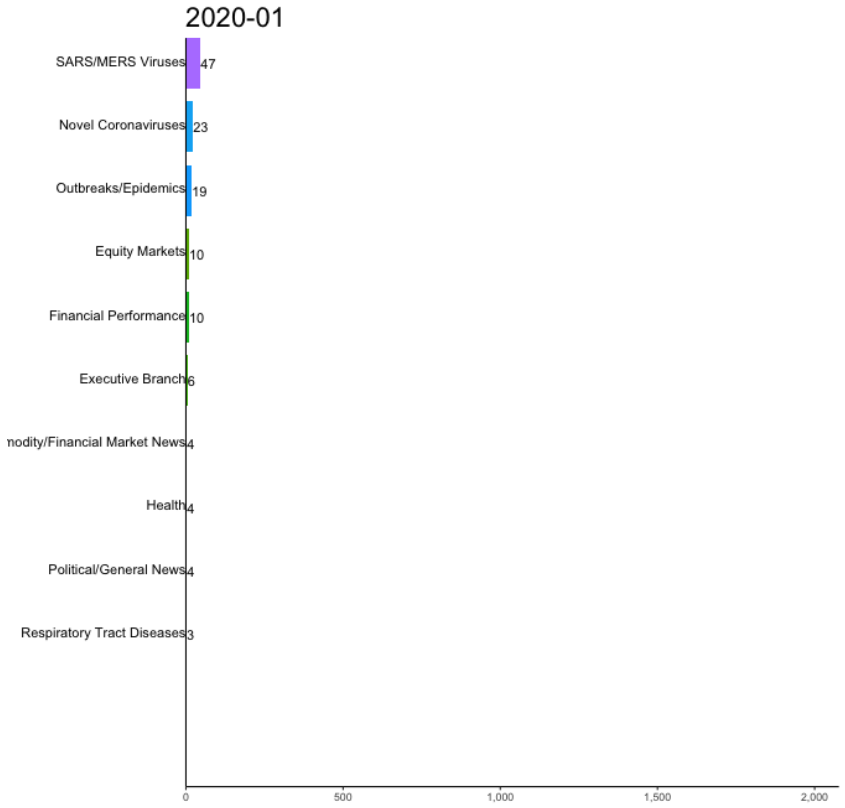
(Mass Media)

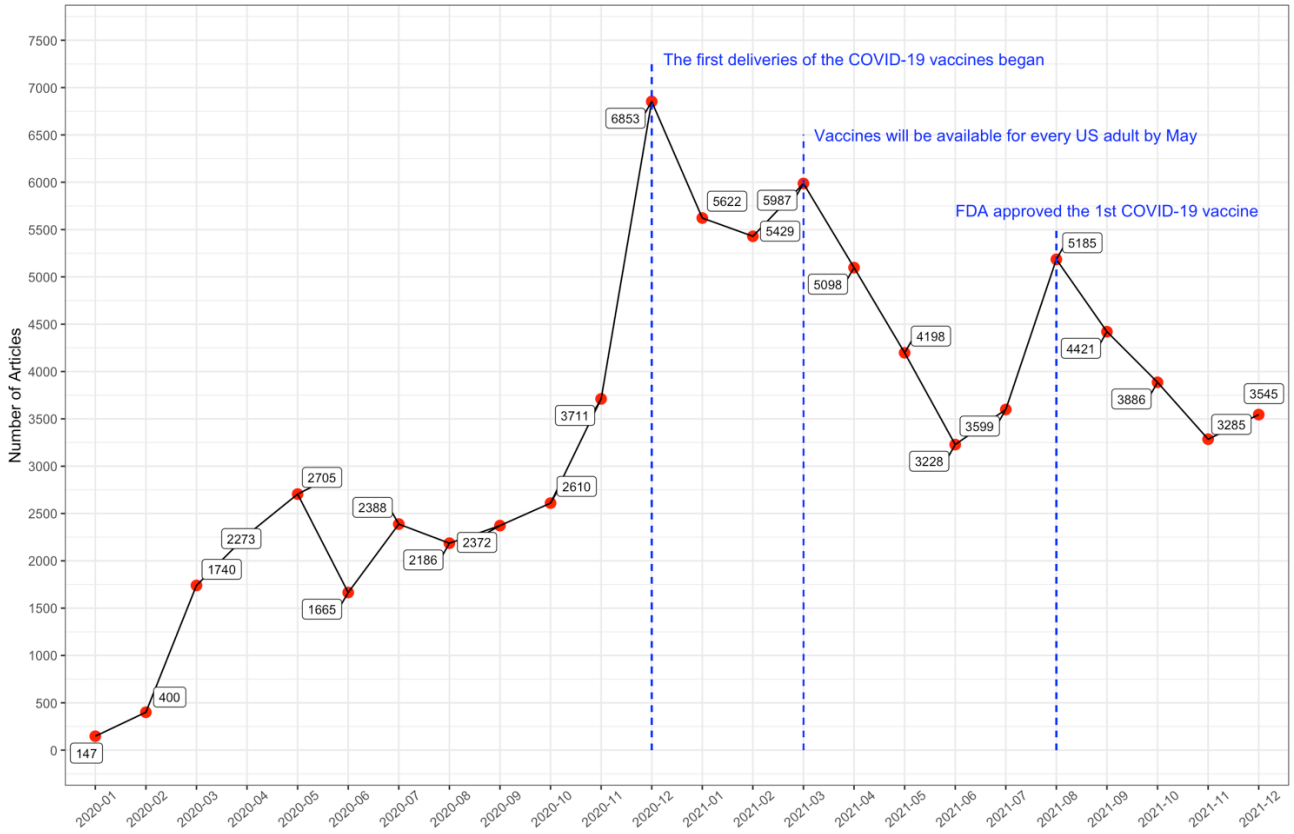
(Ongoing project)



How the study could be leveraged to better understand pandemics?

- Data: all newspapers in the United States collected from Factiva[1]
  - ~83K news articles
- Time period: Jan 2020 – Dec 2021
- Keywords: (vaccine, vaccination, Pfizer, Moderna, Novavax, Janssen/J&J) AND (covid, covid-19, covid19, coronavirus)

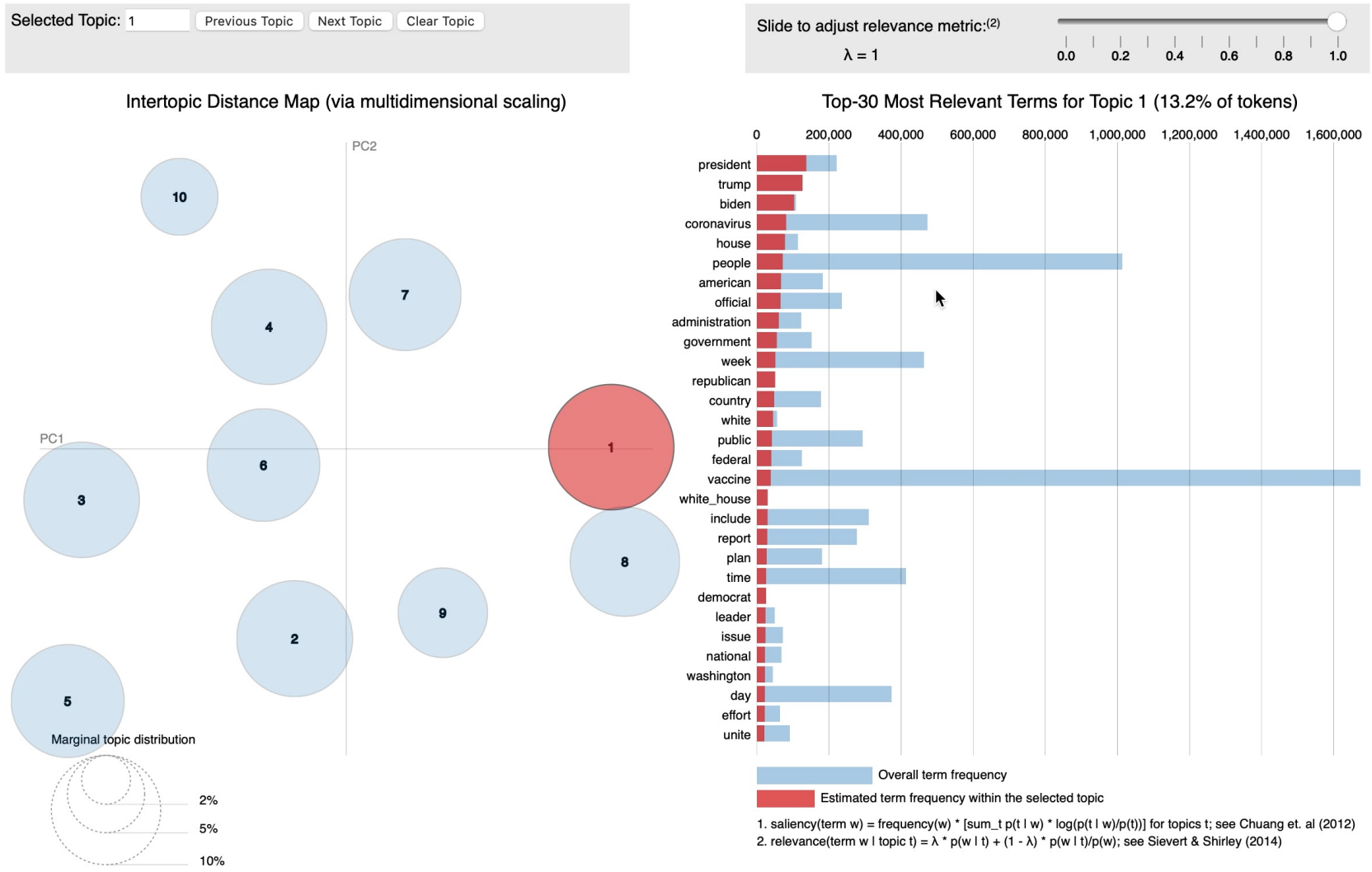


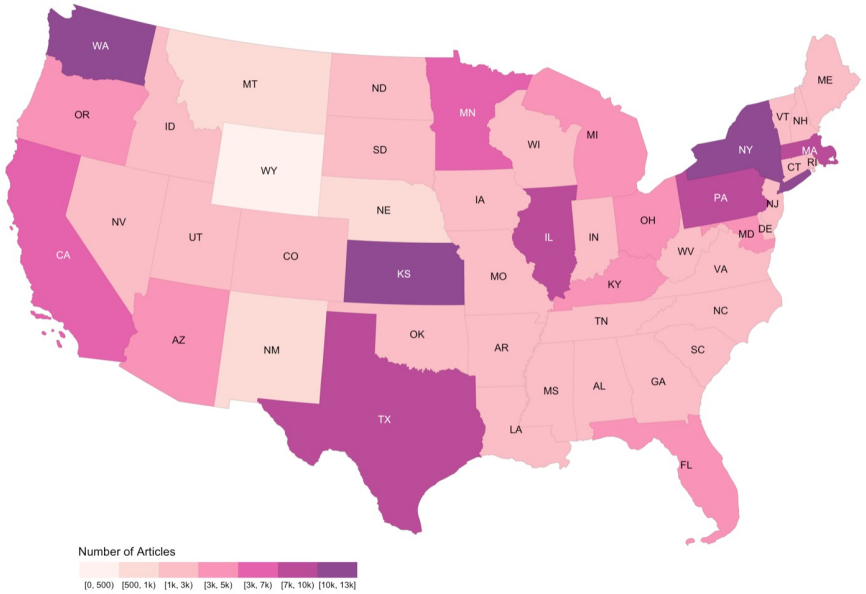


[1] Factiva is a powerful database system which provides current and retrospective news stories, periodical articles, and financial data from thousands of sources worldwide, covering virtually every subject category.

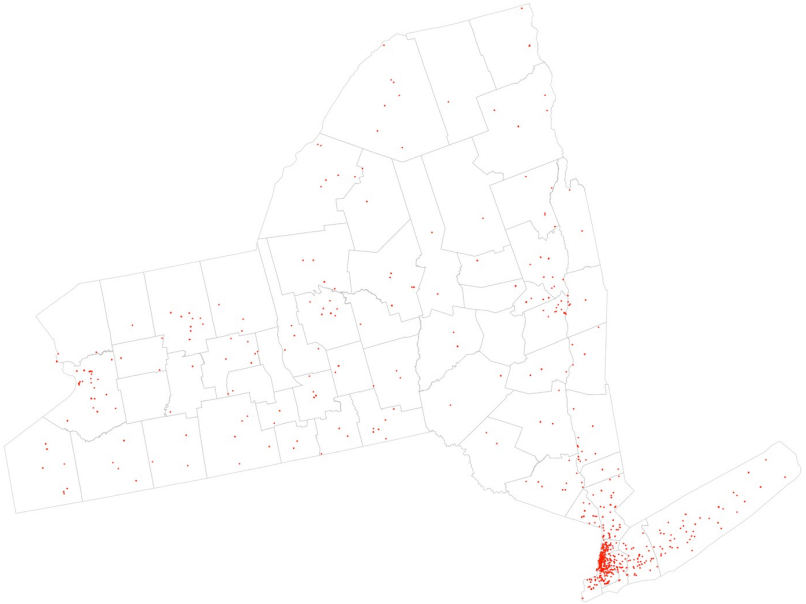
Topic Modeling:

- Used for discovering the abstract topics that occur in a collection of documents
- LDA (Latent Dirichlet Allocation): one of the most popular algorithm (efficient, highly interpretable topics)

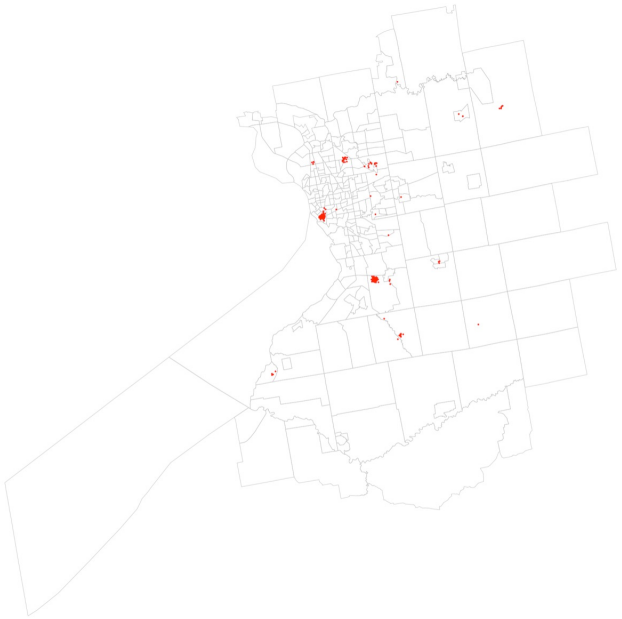




Country level



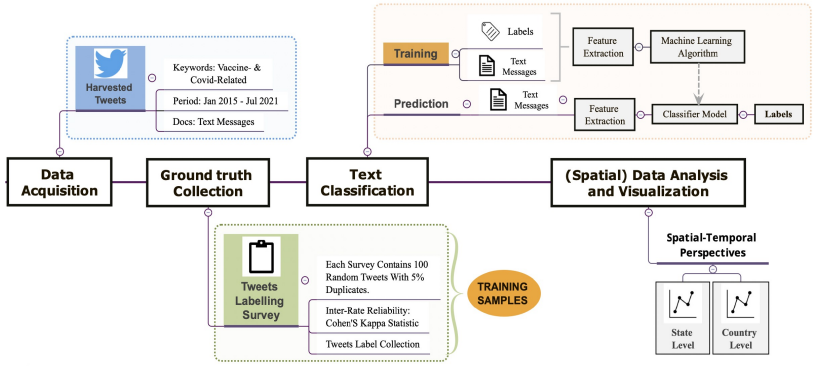
State level



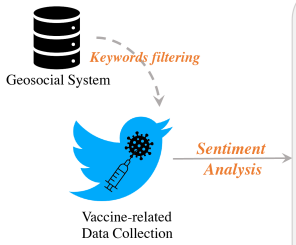
City level

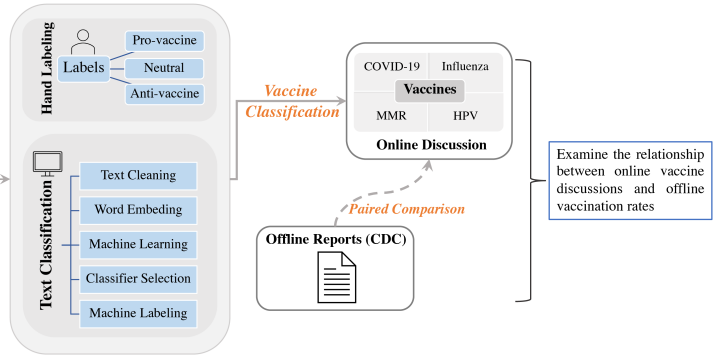
Next step: compare identified topics across different regions and interviews.

- These studies show how can we use machine learning and NLP techniques to understand vaccination debates and public responses, especially the combination of analytical latitude offered by multi-media data;
- Our findings emphasizes that we can not only identify people's sentiments towards diverse sets of public health issues, but also link such analysis to places over time;
- We hope these studies could provide insights into emerging topics in public health.

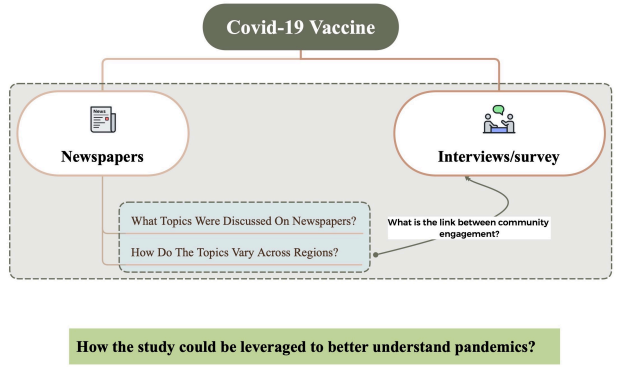


Vaccine Sentiment





Vaccine Comparison



Vaccine Community Engagement

THANK YOU FOR YOUR ATTENTION

qchen47@buffalo.edu https://qingqingchen.info @qingqingchen77