# Exploring Hindu Nationalism Discourse on Reddit through Text Analysis



Author: Ananya Pujary

Affiliation:

Data Analytics and Computational Social Science Program, University of Massachusetts Amherst

This project uses sentiment analysis and topic modeling methods to visualize similarities and differences in sentiments and topics between two subreddit communities differing by size and level of extremism.

## Motivation

- Increasing Hindu nationalistic sentiments in India around the 70s (Siyech, 2021)
- Growing extremism on social media translating into the real world as hate towards religious minorities (Saraswat, 2022)
- Lack of literature on the Reddit, the third most used SNS platform in India (Saraswat, 2022)

## Research Questions

- What are the dominant themes
   discussed in prominent Indian subreddits,
   and as related to the Hindutva ideology?
- What are the dominant sentiments of users on Indian subreddits, and as related to the Hindutva ideology?
- Do some Indian subreddits engage in more extreme discussions of the Hindutva ideology than others?

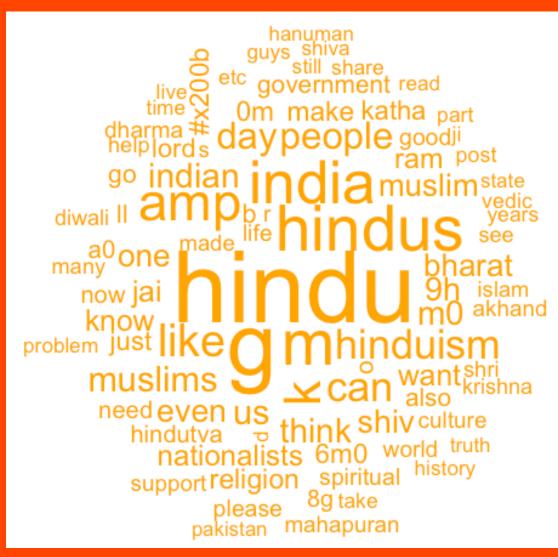
## Data and Methodology

## Data

- r/India (2M members; 747 rows) & r/HindutvaRises
   (6.4K members; 924 rows)
- Collected on December 8th, 2023 using RedditExtractoR
- Time period of 1 month, sorted by newest posts
- 8 columns (6 character, 2 numeric) and each row is a post
- Columns: date published, title, text, title\_text, timestamp, number of comments, subreddit name, url
- Data cleaning: Tokenization (removed punctuation, number, symbols, urls, English and Hindi stopwords), corpus creation

## Methodology

- Sentiment Analysis: with the NRC dictionary including emotions: anger, anticipation, disgust, fear, joy, negative, positive, sadness, surprise, trust. Used to understand user emotions related to topics; polarity score from -1 to 1 calculated by subtracting negative from positive sentiment scores
- Structural Topic Modeling: to extract four overarching topics from the subreddit datasets split by polarity scores (<0: negative polarity, = 0: neutral polarity, >0: positive polarity), generating three models for each subreddit



r/HindutvaRises

r/India

# Analysis

## **Sentiment Analysis**

#### Polarity Score Distribution

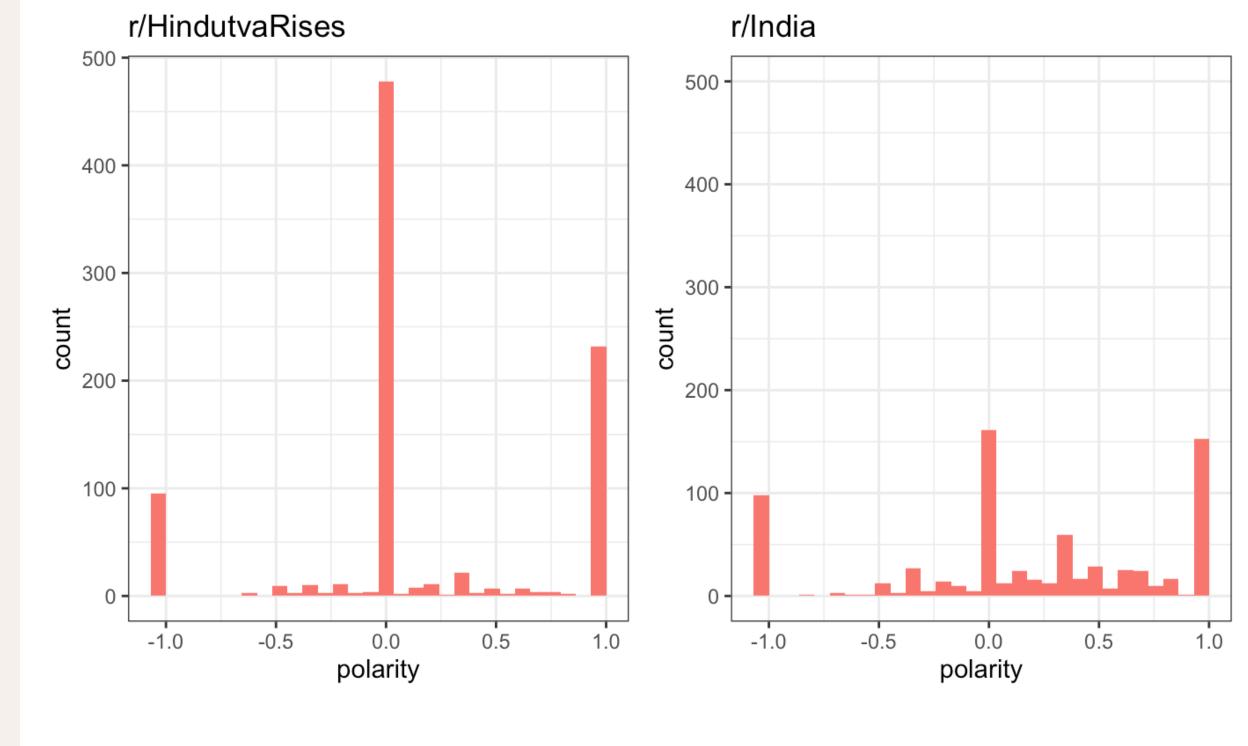


Figure I. Distribution of polarity scores in each subreddit

## Structural Topic Modeling

## r/HindutvaRises Top Topics (Polarity < 0) pic I Top Words:

FREX: schools, minority, private, rte, guru, pay, quota
Lift: \_u, #bansnapchat, #kashmirihindus, 1700s, 1830s, 19th, 20as
Score: schools, private, rte, minority, quota, pay, hindus
Topic 2 Top Words:

Highest Prob: amp, people, m, hinduism, india, theory, g FREX: theory, g, r, ancient, k, aryans, dravidians Lift: -jai, @Om, #art, #ayodhya, #bhaktione, #hindu, #mahabharat Score: theory, r, g, k, m, aryans, dravidians

Topic 3 Top Words:

Highest Prob: hindu, nationalists, moon, india, think, page, can FREX: moon, fought, communities, nepal, daksha, solah, somvar Lift: \_army, \_during, \_ii, \_war, \_world, @bengalsultanate, 00ee584382e692alfa390a4666b62a8f864cla3d Score: moon, nationalists, communities, nepal, daksha, solah, fought

Topic 4 Top Words:

Highest Prob: ram, hindus, jai, muslims, bjp, shree, mandir
FREX: ram, bjp, mandir, ban, remove, movies, shri
Lift: accused, akin, ambit, anger, assertive, attend, awareness
Score: ram, blasphemy, bjp, hindus, muslims, mandir, allies

#### r/HindutvaRises Top Topics (Polarity = 0)

Topic I Top Words:

Highest Prob: hindu, muslim, india, think, place, know, like
FREX: r, support, secularism, bose, chandra, growth, subhas
Lift: accounts, admired, bose, chandra, democracy, growth, law
Score: growth, hindu, bose, chandra, subhas, secularism, think
Topic 2 Top Words:

Highest Prob: g, m, k, m0, 9h, day, shiv FREX: g, m, k, m0, 9h, day, shiv Lift: @5, @hinducafe, #shiv, 0g, 2@8, 2g, 2gd

Score: g, m, k, m0, 9h, shiv, 0m

Topic 3 Top Words:

Highest Prob: amp, hindu, india, hinduism, people, us, hindus

FREX: amp, us, bollywood, lanka, s, sri, akhand
Lift: amp, arabia, battle, boi, chennakeshava, claiming, committed
Score: amp, us, people, lanka, sri, nationalists, bollywood
Topic 4 Top Words:
Highest Prob: like, even, dharma, santana, vedic, kali, hindus

FREX: dharma, santana, kali, asur, well, become, evil
Lift: 25k, 600k, abrahamic, achieve, afghanistan, agnostic, allah
Score: santana, kali, asur, become, evil, karma, becoming

## r/HindutvaRises Top Topics (Polarity > 0)

Topic I Top Words:

Highest Prob: hindu, amp, #x200b, hinduism, want, god, hindus
FREX: #x200b, police, army, eating, beef, pride, reading
Lift: #x200b, 1940s, 25th, 295a, aastik, ability, abstain

Score: #x200b, army, police, terrorists, god, eating, rise

Score: trans, lie, sc, fit, agni, census, tfr

Topic 4 Top Words:

Topic 2 Top Words:

Highest Prob: hindu, india, muslims, hindus, akhand, bharat, trans
FREX: trans, lie, sc, agni, census, fit, so-called
Lift: abdul, agatsya, al, ardhnarishwar, average, avtars, bahuchar

Topic 3 Top Words:

Highest Prob: india, indian, hindus, us, hindu, can, calendar

FREX: calendar, born, jai, abroad, government, hindi, break

Lift: IO-I5, I3th, 2nd, 80k, a, aamir, abcindianhindus

Score: calendar, government, jai, born, indian, success, retention

Highest Prob: diwali, lord, spiritual, mukesh, vats, part, m FREX: diwali, mukesh, vats, g, puja, avatar, online Lift: @cubicalyogi, @office\_sansad, @sammyvada, #hindu, #mahabharat, #mahabharata, #ramanavami Score: diwali, mukesh, vats, g, psychic, part, decor

#### arity < 0) r/India Top Topics (Polarity < 0)

Highest Prob: india, s, card, us, people, things, case FREX: livpure, congress, force, mp, bharat, cancel, research Lift: IOOrs, I6,I8I-crore, I8Ors, 4-hour, 5th, academic, add Score: livpure, cancel, 20th, rameswaram, card, add, deceptive

Topic 2 Top Words:

Highest Prob: india, like, just, news, caste, indian, people
FREX: caste, class, sister, agoda, brahmin, gm, man
Lift: caste, @penpencildraw, 18m, 2019–21, 210m, 4k, 50–75ish
Score: brahmin, agoda, attention, pakistan, class, caste, number
Topic 3 Top Words:

Highest Prob: can, india, like, just, never, get, even FREX: hard, non, vegan, indians, veg, diet, image Lift: hard, image, \_id, #hand, 0570d5348c7cb93e4b7ef69c5da4c82df0704bea, I-2, I0-I2 Score: vegan, non, veg, hard, diet, die, try

Topic 4 Top Words:

Highest Prob: can, parents, back, one, call, t, dad

FREX: m, asleep, clean, hai, dream, mum, prescribed

Lift: m, --, l2pm, 2lyr, 26m, 29m, 90k

Score: asleep, clean, prescribed, dream, mum, m, coach

#### r/India Top Topics (Polarity = 0)

Topic I Top Words:

Highest Prob: amp, india, indian, report, m, s, like

FREX: report, entry, amp, format, mobile, money, two

Lift: report, #x200b, I5k, I7mn, 25k, 97a083a86f3a9I08a5ealab4bII849ccca670022, access

Score: amp, sultans, tourist, visa, price, value, viewpoints

Topic 2 Top Words:

Highest Prob: india, order, since, times, inside, use, trapped

FREX: order, inside, use, back, decathlon, delivered, rescued

Lift: bribes, buddha, business, currency, glovebox, songs, spray

Score: delhi, order, decathlon, delivered, glovebox, spray, tejas

Topic 3 Top Words:

Highest Prob: s, day, roads, people, data, just, new
FREX: roads, data, every, many, took, avoid, go

Lift: \_names, \_star, #melodi, I2th, I4-jan-I969, I925-2023, Ist
Score: roads, many, data, name, avoid, took, says

Topic 4 Top Words:

Highest Prob: india, can, congress, news, s, state, old FREX: congress, news, state, coins, vigilance, book, city
Lift: 32728b62cdcb3d9488e7ce24c5ab9566, 9k, accepting, accidents, adapter, bhopal, bold Score: congress, vigilance, commission, noise, navel, saree, wearing

### r/India Top Topics (Polarity > 0)

Topic I Top Words:

Highest Prob: amp, india, people, one, like, can, just
FREX: ama, auto, lk, named, nani, webp, indigo
Lift: \_a, \_abhinandan, \_abijit, \_am, \_dhruv, \_educator, \_from
Score: ama, lk, nani, named, subsidies, cr, finals

Topic 2 Top Words:

Highest Prob: people, india, like, just, know, can, get
FREX: passport, coins, u, problems, nice, ve, non
Lift: I-2, 5th, activities, adopted, adoption, agra, aim
Score: coins, u, wife, feelings, room, godown, representation

Topic 3 Top Words:

Highest Prob: can, india, bank, online, help, get, account FREX: project, lakh, refund, zomato, gol, payments, seeking Lift: achieve, agent, agreement, astronaut, baby, bare, born Score: zomato, payments, lakh, gol, tu, academic, tashan

Topic 4 Top Words:

Highest Prob: please, india, help, like, can, university, know

FREX: certificate, protein, therapy, tickets, broadcasting, songs, university

Lift: 13-place, bajpayee, bunny, censorship, clinic, didnt, dishwasher

Score: protein, broadcasting, university, flights, forensics, universities, pet

Figure 2. Topic models split by polarity score for both subreddits

# Findings and Conclusion

- Partial support to the hypotheses.
- Smaller subreddit not as extreme as expected since most posts classified as neutral. Positively coded posts talk about Hindu rituals and extremist ideas, while negative ones discuss other religions.
- More general topics and positive polarity in r/India.
- Limitations
  - Cross-sectional data may not be very representative, and was not collected at a particularly relevant time period (e.g. elections season).
  - Not very generalizable since inferences are drawn from two subreddits only.
  - Tokenization not very effective.

### Future Directions

- Compare more subreddits over a relevant time period (e.g. during election season).
- Experiment with more K-values to see how topics in STM change.

# References

Chakraborty, S., Pal, J., Chandra, P., & Romero, D. M. (2018a). Political Tweets and Mainstream News Impact in India: A Mixed Methods Investigation into Political Outreach. Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies. https://doi.org/10.1145/3209811.3209825.

Manuvie, R., & Chatterjee, S. (2023). Automated Sentiment and Hate Speech Analysis of Facebook Data by Employing Multilingual Transformer Models. Computation and Language, 1–11.

https://doi.org/https://doi.org/10.48550/arXiv.2301.13668.

M. N., P. (2023, February 28). India's YouTube Vigilante is Wanted for Murder. Wired. https://www.wired.com/story/indias-youtube-vigilante-monu-manesar-

Saraswat, R. (2022, January 25). Out of Sight: How Reddit became a "safe space" for hate in India. TheQuint. https://www.thequint.com/neon/web-culture/out-of-sight-how-reddit-became-a-safe-space-for-hate.

Siyech, M. S. (2021). An Introduction to Right-Wing Extremism in India. New England Journal of Public Policy, 33(2).

https://scholarworks.umb.edu/cgi/viewcontent.cgi?article=1809&context=nejpp.