

Social Media and Political Polarisation in the 2024 Maharashtra Legislative Assembly Elections

Abstract

This study explores the relationship between social media usage by political parties and affective polarization within the Indian voting population. Focusing on the Maharashtra State Legislative Assembly election, sentiment analysis was conducted on Twitter posts and replies from BJP4Maharashtra and INCMaharashtra. Using RoBERTa, a natural language processing tool, the sentiments were classified as negative, neutral, or positive, alongside a corresponding score. Results indicate that posts containing opposition-blaming and communal rhetoric garnered more polarized responses, while posts centered on policy discussions and updates prompted more neutral responses and nuanced engagement. Interviews with a political social media manager and a political analyst further substantiated these findings, emphasizing the role of emotionally triggering content in driving engagement. To address this, the paper proposes a regulatory framework leveraging machine learning to identify and moderate hate speech and polarizing content while acknowledging technical and ethical limitations. The findings underscore the need for computationally-based solutions to regulate the digital political landscape and mitigate polarization within Indian democracy.

Section1: Introduction and Research Question

Social Media and Political Polarisation has garnered a decent amount of scholarly attention. This scholarly attention is mostly concentrated within the context of the United States and the apparently widening gap between democrats and republicans which has reached new heights in the last decade (Arceneaux, Johnson, & Cryderman, 2013; Pew Research Center, 2017). A lot of the literature on polarization holds that social media has played a role in widening this gap (Jones, 2002, Lau et al., 2017) although not all sources agree on the degree of that effect (Valenzuela et al. 2019). India has been expanding its social media frontiers exponentially within the last few years and there is a need for scholarly attention to be paid to the dynamics of Indian Politics with increasingly mediatized election campaigns. Polarisation in India is not just about political parties but about identity. Tribal and religious riots aren't uncommon to our country and it becomes necessary in that regard to scrutinize the digital space that more and more Indian voters are inhabiting everyday and how this space has affected citizen perceptions.

To approach this subject methodically, there is first a need to define our dependent variable, that is, Polarisation. Ideological and Affective Polarisation have been the two distinctions made by scholars who study polarisation (Kubin & von Sikorski, 2021). Ideological Polarisation refers to the divergence of political opinions, beliefs and stances of political adversaries (Dalton, 1987). This is based around a divergence in opinions and schools of thought that inhabit a society and is generally considered positive for the functioning of a democracy. Affective Polarisation is centered around the role of identity in politics and how one-group animosity can foster, this is generally about the emotions and attitudes of the public about political parties and entities

(Yarchi, Baden, & Kligler-Vilenchik, 2020). It is a measure of the extent to which people like their political allies and dislike their political opponents (Iyengar, Sood, & Lelkes, 2012). For this paper, we will hedge our focus on affective polarisation because its focus on identity and emotions is appropriate to study the political ethos of our country.

In its pre-poll survey for the 2019 and 2023 General Elections, Lokniti, Centre for the Study of Developing Societies (CSDS) asked its survey participants about their social media usage and their like/dislike for a political party. Within their analysis, they found higher levels of like/dislike among social media users for particular political parties (*Social Media & Political Behaviour*, n.d.). This finding forms the backdrop for this research project to further explore the relationship between social media political content and voter polarisation.

The reach garnered by political parties is based as much on their popularity as it is on how their content interacts with app algorithms. With the influx of multiple political parties on social media in India, it has become a competitive space where parties and their members need to design and post content which has virality¹ in order to reach the most people. To this extent, it is also necessary to study how political content interacts with social media algorithms. In simpler terms, what type of political posts tend to get the most engagement? This helps us establish two things: strategies used by political parties to attain viewer engagement and the consequent effect of this type of content on the voter. In this study, we hypothesize that this effect is an increase in political polarisation of voters. Thus, our central research question is: How does social media

¹ The potential for spreadability of any given content; or particular qualities which are considered to have led to content 'going viral' (Chandler & Munday, 2016).

usage by political parties affect the state of affective polarisation within the Indian voting population?

Section 2: Methodology

The dependent variable, that is, Affective polarisation, will be studied on the grounds of sentiments expressed by users responding to posts on Twitter by BJP4Maharashtra and INCHaryana during the Maharashtra State Legislative Assembly election. The tweets are assigned negative, neutral and positive based on their language along with a score between 0 and 1 for how negative, neutral or positive the sentiment expressed is. This was done through the usage of RoBERTa, an NLP (Natural language processing) tool available on Hugging Face (Loureiro et al., 2022). The Maharashtra election was chosen due to its recency and the availability of local engagement with social media posts, i.e. mostly people from Maharashtra engaged with the posts as most replies were in Marathi.

Independent Variable: Social Media Usage by political entities. To identify patterns within posts, we notice keywords that occur multiple times and say something about the content of the post. If our hypothesis is true, posts with polarising content such as opposition blaming and communal sentiments lead to more affective polarisation, i.e the replies would be more strictly negative or strictly positive in sentiment, suggesting a majority of dichotomous responses. Whereas the posts about general policy discussions and updates would have more neutral sentiments, suggesting less dichotomy and more discussion.

The posts were chosen from 16th November to 20th November. Posts with a minimum of 10 replies were chosen so as to conduct a viable sentiment analysis of the comment section. Additionally, 5 posts were chosen at random (following the above criteria) from each day to ensure equality in distribution. This gave us 25 posts from BJP4Maharashtra and INCMaharashtra each. The tweets and their replies were manually scraped.

As part of the fieldwork, two interviews were conducted, one with a Social Media manager for a political entity who asked to be anonymised (hereafter referred to as Jack) and the Second with Yogendra Yadav, former psephologist and convener of the Bharat Jodo Abhiyaan.

Section 3: Results

The results obtained indicate more neutral responses overall but this changes with the content posted. Posts that talked about general updates and policy issues attained more neutral responses than others. While posts with other party shaming obtained the most strictly positive or strictly negative responses. This follows with the hypothesized results. This implies little room for discussion within the comment section, with people instead choosing to voice their pre-formed opinions instead of engaging with the content of what other users have said. Tweets that are talking about opposition parties, i.e. keywords: ShameOnCongress and MVA for BJP4Maharashtra formed the 36% of the posts analysed for the BJP. Additionally, BJP relies on the keywords 'Vote Jihad', 'Naxals' and 'Hindutva' and communal sentiments are appealed to consistently, with the hashtag "#EkHaiTohSafeHai" being one of the most popular. These posts have subsequently garnered the most polarised comment sections as well, with other posts that

talk about specific issues such as development, women safety and farmers enjoying a more neutral discussion, with users adding insights from their experiences suggesting that certain types of content invoked more divisive responses while others did not.

A similar result was also obtained from the INCMaharashtra twitter account, but the difference was that of the 25 posts analysed, 62 percent of them were about invoking anger within the audience through calling out the opposition government for their acts. (keywords: BJP, coalition, Fednavis/Modi). Criticising the BJP was the overarchingly consistent pattern within the INC's digital playbook. Even for posts that talked about specific issues like Women's safety, Constitutional Validity and Farmer rights, most of them mentioned as a side or as the main focus, the many mistakes of the incumbent government. This resulted in comment sections for congress being more divisive, with negative sentiments being expressed in 42% of the overall responses to all the posts. People tend to react overtly negatively when their favoured party is attacked or overtly positively when the party they hate is being attacked, and both these trends were witnessed clearly within the discussions on INCMahrashtra's Twitter.

Section 4: Discussion

The discussion section will be subdivided into 2, the first section is about the the interplay between engagement and content, the second is about the Election commission and its regulatory mechanisms of election campaigns. With this discussion, we hope to understand the digital landscape of Indian politics and attempt to reach potential solutions for the depolarization of mediatized political discourse.

Section 4.1: Algorithms and Politics

Engagement forms key aspects of political advertisement strategies, as social media doesn't guarantee an audience, it has to accumulated and maintained with content that appeals the most to viewers. App algorithms are designed to deliver curated content in line with the individuals pre-existing beliefs and associations with the goal of maximizing their engagement (Vallor, 2011). This implies that even though different actors have different content strategies, the content which reaches the most people in a particular domain, say politics, can accrue similarities in content design in lieu of maximising reach. Jack said that the posts which tend to get the most engagement are usually geared with emotions and sensationalism. This is congruent with the work of Abishek Sharma and Vandita Gupta² who correlated the high usage of emotionally triggering content with higher engagement (Sharma, 2024). Despite different advertising strategies, the BJP and Congress garnered the most engagement out of specific sensationalised talking points which tended to use emotional undertones of anger, disgust, sadness and pride to rile up the viewer and make them actively engage. This is where a potential underlying principle of online political advertisement seems to be surfacing: a digital playbook employing the use of emotionally triggering content. Yogendra says that algorithms end up giving more reach to content with hate-speech as this content causes more controversy and clicks. While this has resulted in increased engagement for both the parties, the effect of such advertisement on users is unclear. Our analysis shows that this might have a polarising effect on voters which could be

² Abishek Sharma and Vandita Gupta carried out an analysis for The Hindu whereby they collated the 2,152 posts made by BJP4Haryana and INC-Haryana from the period of 12th September (beginning of campaigning period) to 5th October, 2024 (start of polling period) and analysed them for seven distinct types of emotions: Neutral, Anger, Joy, Fear, Disgust, Sadness and Surprise to see if the post appealed to any of these.

hypothesized to be a consequence of being emotionally triggered and producing strong reactions which make audiences more sure of what they believe in, strengthening their pre-existing beliefs.

The answer to our question about how social media political advertising affects voter behaviour is thus two-fold: First, Political Advertisement that focuses on opposition blaming and communal messaging correlates with higher affective polarisation of users of social media who actively engage with such content, secondly, political advertisement of the aforementioned kind tends to get more engagement and interacts positively with social media algorithms.

Rather than providing a platform for information influx and diversification of viewpoints, which is an expectation one would have of a digital landscape with a free flow of information, we instead get feedback loops of *certain kinds* of information that set the dominant agenda being discussed on social media.

Section 4.2: Election Commission of India

The Election Commission of India has a Model Code of Conduct (MCC) for social media for candidates and political parties campaigning for an election. This outlines ‘consensual guidelines’ for political parties and candidates during election campaigning. They have recently involved within the MCC the usage of social media platforms for information dissemination. Here, they have recommended candidates to “include [sic] in any activity which may aggravate existing differences or create mutual hatred or cause tension between different castes and communities, religious or linguistic” (Para I.1). These guidelines are vague and are often difficult to enforce within election time. We already saw that within our analysis the prevalence of words

and hashtags like “vote jihad” and “#EkHaiTohSafeHai”, the EC has taken notice of the usage of the phrase “Vote Jihad” and have said that they are conducting thorough and comprehensive investigation about its usage. This statement about the investigation was issued on December 11th, 18 days after the votes for the Maharashtra election had been counted and the winner of the election was announced (PTI, 2024). The vague nature of the MCC’s guidelines is on display here and it highlights the key issues with the body: lack of enforcement of its guidelines and a lack of technical capacity to deal with the scale of information on the Internet.

Section 5: Towards a Solution

The digital realm of politics has been expanding its reach in the country. CSDS data shows that the number of Indians on Social Media and engaging with political content on it has been increasing leaps and bounds (*Social Media & Political Behaviour*, n.d.). As we have established, the interplay between algorithms which maximise engagement and political content is one of boosting posts with emotionally geared and polarising rhetoric, mostly rife with hate speech and fake news. A lot of this content ends up being in violation of MCC of the Election Commission of India but managing these violations has proven to be an arduous task. Whether because of scale or incompetence, such content is everywhere on digital platforms. This has been steadily polarising voters as seen within the comment sections and given the spread of social media platforms, this trend is not looking to slow down. Both Jack and Yogendra said that any potential solution to polarising content on social media could only come through computational means as the scale of the problem far exceeds the technical capacity of just the Election Commission and its members.

If an algorithm can learn to promote certain content, it is also capable of demoting a class of content. Counter Algorithms if I may. There already exist algorithms on these platforms which block and slow down reach for certain content. A common form of this practice is shadow banning, where a profile's reach is severely limited upon multiple scenarios. This is done for violating Terms of service of a given platform, or sometimes, by using political power for shadow banning a political adversary. The latter use case has been employed often with Yogendra yadav and his social media accounts. Thus, the technology required to auto assess violations and hand reach blocks to faltering entities exists. This forms the backbone of the solution I propose. Note that it is one of precarious nature, with the same potential set of problems that we began with, hence it must be handled with caution and subtlety. The proposed solution would operate in 4 steps, identification, action determination, implementation and enforcement and adjudication of disputes. Before I breakdown these steps in detail, it is important to list the assumptions I have made about the Indian democracy when compiling this solution:

1. The Election Commision is politically neutral and wants to implement the Model Code of Conduct and laws of the IPC (Indian Penal Code), ITA (Information Technology Act) and RPA (Representation of People Act) in the right spirit.
2. The Judiciary is politically neutral.
3. Bureaucracy and Government are not politically aligned.
4. The Machine Learning Model to be used is transparent and human interpretable.
5. Technological literacy is equally distributed.

While the assumptions above are unrealistic and not good ones to make, they represent some of the most challenging meta-problems³ our country faces and are thus, beyond the scope of this paper.

Identification: The first step is the preparation of a comprehensive lexicon of violations: words and phrases that promote hate speech and actions which would violate the MCC or relevant sections of the IPC (Indian Penal Code), ITA (Information Technology Act) and RPA (Representation of People Act). These need to be defined with precision. The current vagueness of the MCC leaves room for arbitrary usage and non-usage. This list should be non-exhaustive in nature and its contents should be public knowledge. It should be prepared by the Election Commission and the Supreme Court in tandem. Phrases here are important as hate speech rhetoric is formed contextually, not just through usage of certain terms. The purpose here is for a machine learning model to be trained to identify and pick up usage of these phrases on a scale that humans cannot.

Action Determination: The next step would be to associate usage of phrases with appropriate retort in tandem with different levels of offences. These could range from temporary shadow banning/inability to post to banning a profile entirely. The EC has already displayed its access to power granted to it by Article 324 of the Constitution at various points, including disbarring candidates from running, calling an election null and void, etc. Such measures as listed above would fall comfortably under the purview of powers available to the Election Commission.

³ An issue underlying several other problems. Essential to resolving these problems is understanding, and possibly resolving, this underlying issue (The Law Dictionary)

Implementation and Enforcement: This step involves the legislative and judicial action that demands corporations that run social media platforms to comply with the above definitions and mete out the prescribed actions to the offenders. This would mean that the social media platform develops the algorithm which detects the codified violations and enforces the prescribed action. This model has to be human-interpretable and available to the public as a means to ensure checks and balances on the model to prevent potential biases to creep in. This fact assumes technological literacy is distributed equally across the population as the publicly available nature of the model would imply that everyone can inspect and understand the working of the model.

Adjudication of Disputes: Any disputes arising from the above is taken first to the platform, appeals go to the election commission and then the Supreme Court. Any disputes about the lexicon and what it should and should not contain are taken up by the Supreme Court.

Section 6: Limitations, Ethical Implications and Further Research

The research done for this project was done by one person with no funding. This resulted in some corners being cut in terms of methodology and scale which should be reworked given more time, people and funding. The biggest of this includes translation. All of the tweets analysed for this research were in Marathi. The only available source for translation with no monetary costs was Google translate. The accuracy rates for google translate for regional Indian Languages is about 74% (IIT Patna) and this rate falls more when considering tweets with noisy text, hashtags and slang. Thus for further research, either a human translator or machine translation with higher

levels of accuracy for direct translation and context identification would be ideal. The second limitation has to do with scale. As twitter API has become wildly expensive, scraping had to be done manually. This is manageable for scraping tweets but for scraping replies to tweets, the scale of the problem becomes about 30 times larger in case of our analysis. Further research would involve the use of automated scrapers for this purpose. Next, the interviews and fieldwork should be conducted with a larger variety of experts. For this project, I interviewed two people with political inclinations which are not fully representative of the media landscape.

In regards to the solution proposed, with more time and effort, obtaining a solution of a similar vein which doesn't rely on an ambitious set of assumptions is possible. Ethically, the solution is a precarious one. It involves using AI models which have the potential of unintended biases creeping in. To possibly combat this, the model has to be as humanly interpretable and democratic as possible.

For further research, firstly expanding the scale of the work done here as laid out in the limitations is essential. Secondly one could also look at TV Media and the affective polarisation which follows from there, as TV news channels enjoy a much larger audience than Social Media platforms currently do, and thus hold a larger sway on public opinion and discourse.

Section 7: Conclusion

This research aimed to explore the relationship between social media usage by political parties and affective polarisation within the Indian voting population, with a focus on the Maharashtra State Legislative Assembly elections. By analyzing Twitter content posted by BJP4Maharashtra and INCMaharashtra, this study identified clear trends in engagement and sentiment polarisation. Posts containing divisive rhetoric, such as opposition blaming and communal messaging, were

shown to generate more emotionally charged, dichotomous responses, whereas posts focused on policy updates and developmental issues fostered relatively neutral and constructive discussions. This supports the hypothesis that political content leveraging emotionally triggering narratives contributes to heightened affective polarisation among social media users.

The interplay between social media algorithms and political content was also brought to light, demonstrating how platforms prioritize emotionally engaging posts to maximize user interaction. This practice inadvertently amplifies polarising content, creating feedback loops that entrench pre-existing beliefs and reduce space for deliberative discourse. Interviews with key stakeholders further underscored how sensationalism drives both content strategy and user engagement, while regulatory mechanisms like the Election Commission of India's Model Code of Conduct remain limited in their enforcement capabilities.

To address this issue, the proposed solution envisions the development of counter-algorithms designed to identify, moderate, and demote polarising content on social media platforms. While technologically feasible, this approach is contingent upon critical assumptions about the neutrality of governing institutions, transparency of AI models, and equitable technological literacy across the population. The solution's ethical and practical implications highlight the need for careful implementation, with robust oversight to ensure fairness and prevent misuse.

The study acknowledges its limitations, including the challenges of translation accuracy, limited data collection due to resource constraints, and a relatively small pool of expert interviews. Future research should focus on expanding the scale of analysis, incorporating automated tools for data scraping, and examining other forms of media, such as television news, which continues to exert significant influence over public opinion.

In conclusion, the growing role of social media in Indian politics has introduced new dynamics that both reflect and exacerbate affective polarisation. Addressing this challenge requires a multi-pronged approach that combines technological solutions, institutional oversight, and public awareness to foster a more balanced and inclusive digital political discourse. While the proposed measures offer a starting point, further interdisciplinary research is essential to develop sustainable solutions that uphold the democratic ideals of informed, diverse, and constructive political engagement.

References

- Arceneaux, K., Johnson, M., & Cryderman, J. (2013). Communication, Persuasion, and the Conditioning Value of Selective Exposure: Like Minds May Unite and Divide but They Mostly Tune Out. *Political Communication*, 30(2), 213–231. <https://doi.org/10.1080/10584609.2012.737424>
- Author, N. (2015, April 21). Americans' Views on Open Government Data. Retrieved December 17, 2024, from Pew Research Center website: [https://www.people-press.org/2017/10/05/the-partisan-divide-on-political-values-grows-even-wider/\(open](https://www.people-press.org/2017/10/05/the-partisan-divide-on-political-values-grows-even-wider/(open)
- Chandler, D., & Munday, R. (2016). *A Dictionary of Social Media*. Oxford University Press. <https://doi.org/10.1093/acref/9780191803093.001.0001>
- Dalton, R. J. (1987). Generational Change in Elite Political Beliefs: The Growth of Ideological Polarization. *The Journal of Politics*, 49(4), 976–997. <https://doi.org/10.2307/2130780>
- Iyengar, S., Sood, G., & Lelkes, Y. (2012). Affect, Not Ideology: A Social Identity Perspective on Polarization. *Public Opinion Quarterly*, 76(3), 405–431. <https://doi.org/10.1093/poq/nfs038>
- Jones, D. R. (2001). Party Polarization and Legislative Gridlock. *Political Research Quarterly*, 54(1), 125–141. <https://doi.org/10.1177/106591290105400107>
- Kubin, E., & von Sikorski, C. (2021). The Role of (social) Media in Political polarization: a Systematic Review. *Annals of the International Communication Association*, 45(3), 188–206. Retrieved from <https://www.tandfonline.com/doi/full/10.1080/23808985.2021.1976070#abstract>

- Lau, R. R., Andersen, D. J., Ditonto, T. M., Kleinberg, M. S., & Redlawsk, D. P. (2016). Effect of Media Environment Diversity and Advertising Tone on Information Search, Selective Exposure, and Affective Polarization. *Political Behavior*, 39(1), 231–255. <https://doi.org/10.1007/s11109-016-9354-8>
- PTI. (2024, December 11). ECI analysing impact of terms like “vote jihad” during Maharashtra polls: Electoral Officer. Retrieved December 17, 2024, from The New Indian Express website: <https://www.newindianexpress.com/nation/2024/Dec/11/eci-analysing-impact-of-terms-like-vote-jihad-during-maharashtra-polls-electoral-officer>
- Sharma, A. (2024, December 7). Joy, fear and anger — the BJP and INC’s digital playbook. Retrieved December 17, 2024, from The Hindu website: <https://www.thehindu.com/data/joy-fear-and-anger-the-bjp-and-incs-digital-playbook/article68954635.ece>
- Social Media & Political Behaviour*. (n.d.). Retrieved from https://www.lokniti.org/media/upload_files/Social%20media%20and%20Political%20Behaviour%20report%20new.pdf
- Valenzuela, S., Bachmann, I., & Bargsted, M. (2019). The Personal Is the Political? What Do WhatsApp Users Share and How It Matters for News Knowledge, Polarization and Participation in Chile. *Digital Journalism*, 1–21. <https://doi.org/10.1080/21670811.2019.1693904>
- Vallor, S. (2011). Flourishing on facebook: virtue friendship & new social media. *Ethics and Information Technology*, 14(3), 185–199. <https://doi.org/10.1007/s10676-010-9262-2>

Yarchi, M., Baden, C., & Kligler-Vilenchik, N. (2020). Political Polarization on the Digital Sphere: a Cross-platform, Over-time Analysis of Interactional, Positional, and Affective Polarization on Social Media. *Political Communication*, 38(1-2), 1–42.