Deciphering Viral Trends in Everyday WhatsApp Use in Rural India

Kiran Garimella, Bharat Nayak, Aditya Vashistha

kiran.garimella@rutgers.edu, bharatnayak34@gmail.com, adityav@cornell.edu

Abstract

This research studies the nature and spread of WhatsApp content among everyday users in a rural Indian village. Leveraging a dataset of hundreds of private WhatsApp groups collected with consent from participants, our study uncovers the kinds of WhatsApp groups users are part of, marking the first such categorization. The dataset comprises tens of thousands of messages, out of which we manually classified 600 pieces of content designated as 'forwarded many times'—indicating their viral status.

Our key findings indicate a high prevalence of content focused on national politics, with the viral messages overwhelmingly supporting a specific political party and disparaging the opposition. Significantly, these messages were fraught with misinformation, engendering hate against Muslims and promoting a narrative of Hindus being under threat. This trend was particularly noticeable within caste-based groups, which were dominated by misinformation, pro-BJP rhetoric, anti-Congress content, and Hindutva propaganda. Remarkably, much of the misinformation circulating had previously been discredited by established fact-checking organizations. This suggests not only a recurring cycle of debunked information reappearing but also that fact-checks are failing to penetrate these specific groups.

As the first quantitative analysis of everyday WhatsApp use in a rural context, this research has far-reaching implications for understanding the unique challenges posed by end-to-end encrypted platforms. It serves as a crucial baseline for designing more effective moderation policies aimed at combating misinformation and fostering a more responsible use of encrypted communication channels.

1 Introduction

While WhatsApp is a key communication platform for billions globally, its role in disseminating information, particularly in rural settings like those in India, remains understudied academically. This lack of scrutiny becomes problematic in light of the severe offline consequences that rumors on the platform have incited, such as lynchings (Arun 2019). The research void is particularly acute in understanding the nature and spread of 'viral' content among everyday users in these rural communities, especially in India where hundreds of millions of first time internet users depend on the platform for news and social discourse. Our study, focusing on a village in central India, aims to fill this gap. It captures the intricate dynamics of caste, religion, and social ideology and highlights the importance of understanding how misinformation and political propaganda proliferate, potentially affecting democratic processes.

The challenges of studying this issue are multi-fold. The absence of public APIs coupled with WhatsApp's end-to-end encryption and inherently private nature significantly hinders systematic data collection. The sampling process of identifying rural participants itself is intricate, as reaching out to users for such sensitive and private information poses ethical concerns and practical hurdles, further complicating the research process.

Prior studies in this domain have primarily been qualitative, focusing on small samples, which provide deep insights but can not capture macro trends. These studies do not fully appreciate the textured layers of rural WhatsApp use, especially concerning how viral messages resonate within different social groups.

In contrast, our study makes use of a mixed-methods approach, starting with a large dataset of over 50,000 WhatsApp messages from private groups collected with consent from the participants. Using this, for the first time, we describe the types of WhatsApp groups present in this rural community and thus capture information diets of these users. Next, we delve into rich qualitative analysis using a manual classification of 600 messages 'forwarded many times' in the dataset. These 'forwarded many times' messages indicate viral content on the platform, and help us get macro trends on content being widely circulated on WhatsApp.

The results reveal a high prevalence of misinformation, constituting over 25% of the viral content examined, and there's a notable convergence between anti-Muslim sentiment and false information. Specific narratives, including those that perpetuate the notion that Hindus are under threat from Muslims, appear to be systematically amplified.

While prior research has noted the Bharatiya Janata Party's (BJP) substantial use of WhatsApp for political communication (Perrigo 2019), it was uncertain whether this messaging spread beyond party-controlled groups. Our analysis now offers a clearer understanding, showing that about 20% of viral content in the observed WhatsApp groups overtly supported the BJP, even though we did not make any

explicit effort to sample BJP supporters.

One of the most striking findings of our study is the distinct nature of WhatsApp groups within the village. These groups are primarily organized around caste lines, reinforcing existing social structures. An overwhelming majority of the viral content were concentrated in caste-based groups. The content in these groups showed a strong bias, often disseminating misinformation and politically charged messages favoring the BJP while opposing the Congress party. These groups also included materials aligned with Hindutva ideology, further inflaming sectarian divisions.

No evidence was found of any attempt to debunk or challenge the false information circulating within these groups. This lack of counter-narratives signifies an unchecked propagation of misinformation, posing serious implications for social cohesion and democratic processes. Further exacerbating the problem is the resurfacing of misinformation that had previously been fact-checked by established agencies. Among the 600 pieces of content we examined, not a single item contained any form of fact-checking, whether it be an image, video, text, or link. While fact-checking links were occasionally shared, none gained traction, thereby failing to mitigate the spread of misinformation.

These findings bring to light the nuanced ways in which viral content contribute to the social and ideological landscape in a rural community. It uncovers the pressing need for interventions aimed at stemming the unchecked flow of misinformation and bias, as well as fostering a more balanced information ecosystem.

2 Background and Related Work

Social media's exponential growth in the Global South has coincided with a surge in digital misinformation. This has resulted in severe outcomes, including lynchings, civic unrest, and increased political polarization (Anderson and Jaffrelot 2018; Arun 2019). While existing research provides valuable perspectives on how residents of the Global South interact with misinformation, there's a notable gap in understanding the information diets of rural users. This study aims to fill the gap.

WhatsApp and Information Propagation Varanasi, Pal, and Vashistha (2022) take an interview-based approach to examine the differences and commonalities in the way rural and urban communities in India interact with misinformation on WhatsApp. Their study brings out a nuanced view, pointing to localized influences on perception, the role of social status, and community deliberations as vital factors in misinformation spread. They delve into how both rural and urban communities discover, verify, propagate, and counter misinformation within closed WhatsApp groups.

Banaji et al. (2019) employ focus groups and interviews to explore WhatsApp usage across demographics in four Indian states. The study is significant for its focus on violent incidents resulting from misinformation and the role of pre-existing biases, particularly against Muslims. Banaji et al. also discuss the role of mainstream media in perpetuating misinformation.

Arun (2019) study the role of WhatsApp in amplifying rumors that led to violent lynching incidents across India in 2018. Their work focuses on the immediate social impact of misinformation, detailing how false information spirals into real-world consequences.

Chakrabarti, Stengel, and Solanki (2018) conducted an in-depth qualitative study that allowed for a nuanced understanding of WhatsApp usage in India. Participants provided the researchers with week-long access to their phones, followed by detailed interviews. The study identified four dominant narratives: Hindu power and superiority, preservation and revival of culture, progress and national pride, and the prowess of political leaders. Though the study was limited to a small sample size of 40 users, its findings were rich in qualitative insights.

Despite variations in location, time of study, population, and research methodology across different papers, our research echoes existing work in identifying similar content consumption and narratives. We further substantiate the role of WhatsApp in disseminating misinformation and its impact on rural communities in India. Our study finds common ground in themes of nationalism, social and political biases, and the mechanisms through which misinformation circulates, affirming that these issues persist across time, geography and population.

Our work also extends prior research in several key ways. Our study distinguishes itself through a mixed-methods approach, allowing for a robust analysis of tens of thousands of messages, particularly focusing on content that is viral on WhatsApp. We uniquely focus on consumption patterns in rural Indian settings and add a temporal dimension by assessing the persistence of identified issues over time. This extensive dataset enables us to provide a more nuanced understanding of the complexities involved in information consumption and sharing behaviors, particularly related to national and hyperlocal content.

Politics on Twitter in India There is a lot of work exploring the various dimensions of Indian electoral politics, with a focus on discourses, electoral strategies, and the role of social media, particularly Twitter. We provide a brief overview of this landscape here. Martelli and Jaffrelot (2023) argues that populist leaders in India use implicit rhetorical strategies like simplicity and intimacy to connect with the masses. Jaffrelot and Verniers (2020) gives an exhaustive account of the 2019 Indian elections, touching on multiple facets from the BJP's robust electoral machine to the impact on women's representation. It emphasizes how the BJP's control over (social) media and resources led to its decisive victory and the weakening of regional parties. Rajadesingan, Panda, and Pal (2020) quantifies the extent to which candidates make their campaigns leader-centric rather than party-centric, finding that BJP candidates strategically run Modi-centric campaigns. Finally, Dash et al. (2022) examines how influential Twitter accounts in India disseminate dangerous speech against vulnerable groups, contributing to societal polarization.

Our study distinguishes itself from this line of work focusing on the micro-level prevalence of content (including misinformation) spread through personal messaging platforms like WhatsApp, rather than examining the macro-level strategies or outcomes related to electoral politics on public platforms like Twitter. While these works collectively pro-

vide a multifaceted understanding of social media use in Indian politics, they don't specifically examine how political messaging on personal messaging platforms impact everyday users.

Political Propaganda on WhatsApp Political organizations have constructed an expansive network of WhatsApp groups to tap into the burgeoning user base, especially targeting people who have recently joined the digital sphere (Perrigo 2019). Numerous journalistic pieces corroborate the existence of such large-scale networks, citing hundreds of thousands of politically-oriented WhatsApp groups (Arnimesh 2020).

In a focused study, Garimella et al. analyzed images shared within 5,000 publicly accessible WhatsApp groups during the 2019 elections. Their results revealed that approximately 10% of these shared images could be categorized as misinformation (Garimella and Eckles 2020). Studies conducted in other geopolitical contexts, such as Brazil (Resende et al. 2019) and Pakistan (Javed et al. 2022), further validate these findings, showing a consistent pattern of misinformation prevalence.

Taking a different angle, Chauchard and Garimella (2022) scrutinized private WhatsApp groups managed directly by political parties. With permission from group administrators, they joined over 500 such groups and annotated a corpus of more than 40,000 images. Contrary to public groups, they found surprisingly low rates of misinformation and hate speech within these party-controlled settings (Chauchard and Garimella 2022). The bulk of the content appeared to focus on routine activities and updates from party workers.

These studies have limitations in that they do not address the concept of virality, which is the focal point of our research. Furthermore, the studies do not explore the WhatsApp groups as standalone entities. Although extensive research exists that sheds light on the nature of the content within these networks (Garimella and Eckles 2020), current public reports largely rely on self-disclosures from political parties regarding the scope and magnitude of their digital operations. Absent from the scholarly discourse is any measure of the impact these groups have on the daily life of typical WhatsApp users, as opposed to just party workers and ardent supporters who are most often the core members of public groups.

Rural Internet Usage and Media Consumption Mobile adoption is soaring in India, especially in rural regions, fueled by affordable smartphones and data. A Nielsen survey indicates that over 300 million rural Indians are online, a figure expected to exceed 500 million soon (Nielsen 2019). The purposes for which these users access the internet remain largely unexplored. Rangaswamy and Arora (2016) explores how the unregulated expansion of mobile internet in India's urban slums has led to unconventional uses, particularly among the youth. The paper focuses on how activities often deemed as 'leisure' or 'non-productive,' such as socializing on mobile Facebook, actually serve as powerful avenues for learning and cultural participation. In contrast, Arora (2016) scrutinizes the often-celebrated impact of big data in the Global South, arguing that its framing as an empowerment tool masks underlying biases and neoliberal agendas. It calls for a more nuanced understanding of how big data can genuinely serve as a social good in emerging economies, rather than merely treating impoverished populations as new consumer bases. Collectively, these studies contribute to an intricate understanding of technology's role and societal impact in India. They urge for a comprehensive analysis that acknowledges the socio-cultural intricacies of technology usage.

Fact-checking in the Global South Efforts to understand and address misinformation on digital platforms have predominantly centered around the mechanisms of fact-checking and the infrastructures that support it. Juneja and Mitra (2022) dissects the fact-checking ecosystem, interviewing participants from diverse stakeholder groups, such as editors, external fact-checkers, and social media managers, among others, to understand the nuances of real-world fact-checking processes. Haque et al. (2020) shifts the discourse from the western context to the Global South, specifically focusing on the fact-checking landscape in Bangladesh. The authors highlight the lack of infrastructural support for voluntary fact-checkers and explore the role of the audience and journalists in combating misinformation. Both studies underline the collaborative nature of the fact-checking process and its contextual complexities, particularly in a landscape where information spreads on private platforms like WhatsApp, and how resource-constrained fact-checkers play catch-up with viral misinformation spreading on an encrypted platform where no data is available.

There have been technical solutions designed for the unique challenges posed by end-to-end encrypted platforms like WhatsApp, recent work has explored the utility of crowd-sourced tiplines for flagging potentially misleading content (Kazemi et al. 2022). This study found that tiplines serve as valuable lenses into 'viral' conversations on WhatsApp, thus acting as a critical source for discovering false or misleading information.

Hall et al. (2023) conducted a qualitative analysis to assess the effectiveness of WhatsApp's 'forwarded' and 'forwarded many times' tags, which are designed to make users consider the authenticity of the content they receive. They find that these tags are subject to varied interpretations, reducing their intended efficacy in combating misinformation. While the tags were designed to minimize negative platform-user interactions, their research suggests that the link between 'forwarding' and misinformation dissemination should be made more explicit.

Our research not only complements the insights from fact-checkers in identifying and countering misinformation but also reveals the persistent cycle of already-debunked false information. We find specialized misinformation trends within WhatsApp groups, underscoring the need for more targeted moderation policies. Our results question the efficacy of current reactive fact-checking models and opt-in technical solutions for combating misinformation. Moreover, by serving as a baseline that includes the cycle of resurfacing debunked information, our study lays the groundwork for designing more effective, context-specific strategies to tackle misinformation on platforms like WhatsApp.

3 Data Collection

We collected data from a village in Jharkhand state which is located in Central India.[1] The state is marked by its communal sensitivity and frequent incidents of mob violence, making it a critical location for studying the spread of misinformation (Times 2017).

The WhatsApp data for the study was collected through an opt-in tool built for this research. The tool allows users to scan the QR code for web WhatsApp, thus providing the researchers access to the users WhatsApp account. Once the user is authenticated, the tool provides a list of the groups which the user is a part of. The tool automatically removes any groups with less than 6 members (to protect privacy and avoid any re-identification issues) and groups with no activity. The users can then select the groups they were comfortable donating. The tool was designed with user privacy in mind. No personal identifying information (like phone numbers) are ever stored — the data is anonymized before being stored. In addition, names, emails and phone numbers are automatically removed using Google's Data Loss Prevention library[2] and all faces in images are blurred before storing. The tool has been approved by our institution IRB. The entire process of donating WhatsApp groups takes less than 5 minutes of a user's time. The users were not compensated for their data donation.

Recruitment for the data collection tool was done by one of the authors who hails from the village through an iterative, convenience sampling approach. Potential participants were selected based on prior acquaintance, age, and openness to technological understanding. Conversations for recruitment typically commenced in public gathering places, most notably a central gathering place in the village. If the users consented, they were explained the process, including the privacy measures put in place to protect their data.

The study eventually onboarded 31 participants, who contributed data from a total of 164 distinct WhatsApp groups. These participants were all male,[3] primarily Hindu, and belonged to different castes. Their ages largely ranged from 20 to 30 years, although there were a few older participants.

The 164 WhatsApp groups we collected were then manually categorized into various categories such as village groups, caste groups, religious groups, Hindutva groups,[4] activism groups, and others. Table 1 shows the distribution of the types of groups, and their total messages. A majority of the groups fall under the 'Other' category including a long tail of groups about family, friends, job search, education help, hobbies, etc. A detailed description of the labels given to the groups is shown in Table 3 in the Appendix.

We collected data from these groups for around 2 months spanning mid June–August, which gave us a total of 53,389 messages of which 68.8% were text messages, 28.2% images and 3% were videos.

While the dataset is robust in its coverage of the Hindu majority narratives, it does have limitations that must be acknowledged. The data predominantly represents the Hindu viewpoint, as no Muslims were onboarded for the study. Moreover, the Muslim minority, constituting approximately 30% of the village's population, might possess distinct narratives that this dataset does not capture. Lastly, the dataset is skewed towards younger, primarily jobless, males, presenting a potential bias in understanding the full scope of narratives prevalent in the community. That said, this research marks a pioneering effort in gathering private WhatsApp data with the explicit consent of users. The dataset, along with the methodologies employed for its collection, serves as a significant contribution that holds applicability beyond the current study's context.

4 Data Annotation

The dataset contains information on content that is 'forwarded many times' on WhatsApp. A content is classified as 'forwarded many times,' a term defined by WhatsApp as any content forwarded a minimum of five times.[5] The specificity of this classification remains ambiguous; it is uncertain whether a 'forwarded many times' message has been shared just five times or exponentially more (say a million times). Despite this lack of clarity, we utilize this classification as a proxy for viral spread within the WhatsApp ecosystem, implying that the content has traveled through at least five different nodes from its origin. Throughout this study, the terms 'forwarded many times', 'frequently forwarded' and 'viral content' are used synonymously.

The core dataset used in the rest of the paper is a subset of the 50k messages we collected, consisting of 600 frequently forwarded messages in the two month period between June 21, 2023 and August 20, 2023. The data annotation process was performed on a custom dashboard designed to track messages that have been forwarded multiple times in our dataset. For each message, the annotators could see the chat context in which the content was shared, by browsing through the previous/next 10 messages sent along with the message.

The coding was done by one of the authors of the paper, who was also from the village and knew the context the data was collected. We followed an inductive coding method where we iteratively improved the coding as we annotated the data. In case where the codes were not clear, the annotator discussed with the other authors. We annotated using an iterative coding process. First, we decided on the categories for which the content should be coded. Then, we began qualitatively coding for emergent themes. These codes were refined over multiple iterations until we finally grouped

[1] The village has a population of around 9,000 people. To protect the privacy of the participants, we are not publishing the name of the village.

[2] https://cloud.google.com/dlp

[3] Despite concerted efforts, we were unable to include female participants in our dataset. This is reflective of broader gender disparities in terms of smartphone access, privacy, media literacy, and financial capacity for device and data usage, as indicated by (Banaji et al. 2019). Women either lacked personal phones or expressed discomfort in sharing their data, leading to an all-male sample from the village.

[4] According to the Oxford dictionary, Hindutva is an ideology advocating, or movement seeking to establish, the hegemony of Hindus and Hinduism within India.

[5] https://faq.whatsapp.com/1053543185312573

Table 1: Statistics of the groups. %Groups shows the percentage of the total number of groups (164) we obtained in each category. %Total messages shows the percentage of the total messages we collected (53,389). %Viral messages shows the percentage of them which are viral. %Viral annotated shows the percentage of viral content in each category of the 600 messages annotated. The column does not sum to 100 because the content can be viral in more than one category.

Category	%Groups	%Total messages	%Viral messages	%Viral annotated
Village	6.7	12.9	1.3	15.1
Caste	7.3	7.7	6.6	45.1
Religious	6.7	6.1	2.2	12.1
Hindutva	8.5	2.5	10.0	22.3
Activism	12.2	19.5	0.56	9.8
Regional	15.2	13.4	2.3	27.3
Others	43.3	37.5	0.31	10.5

Table 2: Categories of the viral content shared on WhatsApp.

Category	Percentage
Misinformation	26%
Information/Inspirational videos/ Commentary/Amusement videos/ Religious harmony/Educational messages	21.8%
Religious propaganda to influence Hindus	21.0%
Hate against Muslims	19.6%
Pro-BJP political propaganda	18.6%
Anti-Congress political propaganda	13.8%
Regional information	8.5%
Religious	7.3%
Good morning messages	3.5%
Political or religious sarcasm/satire	2%
Political opinion not to benefit any political party	2%
Health misinformation	1.4%
Anti-BJP propaganda	0.5%

similar codes together to create organized parent codes. We ended up with 13 significant categories. Content can belong to multiple categories, e.g. a content can be misinformation and also contain hate against Muslims. The categories and their prevalence are shown in Table 2. Note that since a piece of content can belong to multiple categories, the numbers do not sum to 100. A detailed description of the categories can be found in the Appendix.

To code for misinformation, one of the authors who is a trained fact-checker and has worked in the fact-checking industry for almost a decade went through each piece of content to ascertain both the source of the information and the accuracy of any claims made. For images and videos, reverse image searches were conducted across multiple search engines including Google, Reveye, Bing, and Yandex, to fact-check the content. Text-based messages were verified through standard Google searches.

For each piece of content, we also annotated whether it is fact-checkable, and if so, if it was fact-checked and when. Since we were looking at the chat for coding the content, we also qualitatively coded responses to the misinformation post, specifically documenting whether the content received any responses (e.g. corrections, support, etc).

5 Findings

We examined 600 viral messages from a 50,000-message dataset of WhatsApp groups. More than half of these viral messages were videos, even though they make up only 3% of total shared content. The rest were images (32%), text (12%), and links (2%). Videos were usually forwarded with original captions, though some users added their own comments. The multimedia nature of videos makes them more emotionally impactful than images or text.

5.1 Types of information by group

Table 1 delineates the categorization of WhatsApp groups analyzed in this study. A conspicuous finding is the disproportionate rate at which viral messages are present in caste-based and Hindutva groups. While messages in caste-based groups constitute only 8% of the total dataset, they account for an alarming 45% of viral messages. Similarly, Hindutva-focused groups, which contribute a scant 2.5% of the overall messages, attract over 22% of the viral content. The ratio of viral content (as depicted in the fourth column of Table 1) provides further evidence of this skewed distribution.

These figures corroborate the role of social trust in information sharing. Drawing on findings from Banaji et al. (2019), trust within social and communal networks tends to outweigh considerations of message credibility. The identity of the person or group forwarding a message critically influences the decision to re-forward it. When the sender is perceived as a trusted in-group member, even implausible messages often gain credence and circulation.

Significantly, caste-based groups were rife with misinformation and political content, largely skewed towards pro-BJP and anti-Congress narratives. These groups primarily served rural populations in Jharkhand and mainly aligned with specific castes or religious communities. Though most of our participants had no direct affiliation with the BJP or other right-wing organizations, they were often members of groups that propagated such views. A notable demographic observation is that our sample predominantly comprised youth aged 20-30, who are exposed to messages fostering anti-Muslim sentiments; nearly 20% of the viral content aimed to incite hatred against Muslims. This raises concerns about the potential for long-term indoctrination among youth in such closed communities. Such restricted exposure exacerbates the creation of echo chambers, significantly narrowing the spectrum of narratives to which these individuals are exposed.

The persistent segregation based on caste–still a glaring reality despite being legally prohibited–extends its tentacles into digital spaces as well. These online caste-specific groups serve not merely as digital replicas of existing social structures, but also as fertile grounds for the perpetuation of

hate speech and misinformation. These findings are particularly crucial given the current debates on the role digital platforms play in perpetuating social hierarchies and forming echo chambers that serve as hotbeds for misinformation and hate speech.

Contrary to expectations, village-based WhatsApp groups, which have multi-caste and multi-religious memberships, did not showcase any substantial counter-narratives or factual corrections against the prevalent misinformation, even though certain castes and religions of the members present in the groups were frequently targeted with misinformation and hate speech. This raises questions about the efficacy of such 'diverse' digital spaces in combatting misinformation and the role of class hierarchies in shaping information flows on WhatsApp.

5.2 Popular narratives of content

In the milieu of frequently forwarded messages on WhatsApp, certain narratives demonstrate a recurring pattern. These narratives, often entrenched in political and social discourse, not only hold sway over public opinion but also serve to perpetuate existing divides. This section delves into the predominant narratives that surfaced in our dataset, which notably align with broader ideological currents sweeping through India.

PM Modi as a Global and National Leader One persistent narrative positions Prime Minister Narendra Modi as an unparalleled leader, respected both nationally and globally. In this narrative, Modi is often depicted as the 'savior' of Hindus—a figure capable of restoring Hindu pride and protecting the community from external threats. Messages with such content often include exaggerated accolades, overlooking any criticisms or complexities related to his tenure. These narratives serve to consolidate the Prime Minister's standing and reinforce the party's core voter base and seems to be persistent narrative being pushed by the BJP, since this narrative was also observed to be the most prevalent in qualitative work by Chakrabarti, Stengel, and Solanki (2018) almost 5 years ago.

Fear-Mongering and Communal Tensions A more concerning narrative revolves around instigating fear among Hindus. Messages under this category imply that failing to vote for the BJP would lead to an impending demographic shift where Muslims become the majority, thereby posing a threat to Hindu safety. Hindu women are often specifically portrayed as victims in such messages, magnifying the perceived threat. This narrative also extends to the vilification of particular minority groups, such as Rohingya Muslims and Bangladeshi immigrants, as disruptive elements. Calls to boycott Muslim shopkeepers and their products are not uncommon, along with the propagation of conspiracy theories like 'Love Jihad' (Rao 2011).

We identified a pervasive climate of distrust and antipathy towards Muslims, equating them to traitors who belong in Pakistan. WhatsApp users in most of our groups seem to frequently consume hate speech and misinformation about minority communities within their WhatsApp networks. The credence given to such (mis)information appears to stem from its alignment with the users' own set of ideologically prejudiced and discriminatory beliefs, irrespective of the accuracy of the sources or content.

Our analysis revealed disturbingly consistent messaging that frames Muslims as existential threats to Hindus, a narrative that has chilling parallels with rhetoric that has catalyzed genocides in Myanmar and Sri Lanka (Anwary 2020), and fueled white supremacist actions such as the Christchurch massacre (Quek 2019). Prior work documented the wide spread use of such fear rhetoric in certain WhatsApp groups (Saha et al. 2021) or being used by Twitter influencers (Dash et al. 2022).

Discrediting the Opposition Lastly, there is a concerted effort to discredit opposition figures, particularly leaders from the Congress party. Rahul Gandhi is a frequent target, with false claims circulating about his religious identity and mental acumen. There are also derogatory allegations against historical figures like former Prime Minister Jawaharlal Nehru, portraying him as a womanizer or questioning his religious affiliations. These messages further extend to accuse the opposition of being anti-Hindu by alleging that they restrict Hindu festival celebrations and falsely label the Congress as a 'Muslim party.'

In light of the three predominant narratives identified–elevating PM Modi as an unparalleled leader, inciting fear among Hindus, and discrediting the opposition–there is a concerning pattern of endorsement and amplification. Prominent figures within the BJP frequently echo these narratives during interviews and public rallies, thereby granting them a level of credibility. This is then amplified by the mainstream media, covering it as 'news', which then makes its way to social media.

5.3 Misinformation

Table 2 shows that the most prevalent category of viral content in our dataset is misinformation. Out of the 600 messages analyzed, 26% (158) were identified as misinformation. Notably, nearly half of these (78 out of 158) were in video format. The false narratives primarily centered on three themes: Pro-BJP propaganda, Anti-Congress content, and religious propaganda specifically targeting Hindus. Within these, a significant subset (44% of the misinformation) was engineered to incite hatred against Muslims. This intersection of hate speech and misinformation is extremely problematic.

This is notably higher than rates observed in other contexts (Chauchard and Garimella 2022) or WhatsApp datasets (Garimella and Eckles 2020). The virality indicates not just the prevalence but also the impact and reach of these messages. Misinformation is not just being generated; it is also being massively disseminated and consumed. This significant difference in virality patterns demands targeted attention, for it implicates that whatever mechanisms are causing this have more potency in this context.

The misinformation was not random but tailored to resonate with prevailing sociopolitical sentiments. A case in point is a message circulating the 'Love Jihad' myth (Rao 2011), falsely claiming Hindu women were being lured by

Muslim men. This example is indicative of the dual nature of misinformation: the capability to serve as both a mirror and a molder of public opinion.

Our data not only underscores the resilience of misinformation but also its nimble adaptability, existing and thriving across different platforms. Context-dependent misinformation, such as narratives blaming Muslims for the BJP's electoral loss, denotes a reactive and highly coordinated system in operation. This system is not siloed within the confines of WhatsApp but extends to a broader ecosystem, including other social networks like Twitter and Facebook, where we observed a significant overlap of misinformation themes.

However, most striking is the role of mainstream media in this ecosystem. Our qualitative analysis reveals a concerning synergy between biased national media and the misinformation cycle. Television news clips favoring BJP are often shared virally, reinforcing and legitimizing the narratives initially propagated as misinformation. This legitimacy gains further weight because these media segments are likely seeded by the BJP's organized machinery. In such a scenario, misinformation gains a veneer of credibility, amplifying its social impact and making it much more challenging to counter.

Misinformation on WhatsApp is not purely cognitive but also affective; it capitalizes on emotional resonance. During key events like elections or riots, the urgency and shock value of messages surge, often inflating simmering tensions into flashpoints of potential violence. This raises concerns about the platform being used to escalate long-term discriminatory attitudes into immediate acts of aggression.

5.4 Prevalence of pro-BJP content

As seen in Table 1, even without explicitly sampling pro-BJP users or content, 8.5% of the groups we collected were related to the Hindu right wing. Expanding on the topic, it is essential to contextualize our findings within the broader socio-political landscape. The preeminence of pro-BJP narratives within viral WhatsApp content suggests an effective strategy at the grassroots level, where the party's messaging seems to resonate strongly. The absence of anti-BJP content further underscores the potency of their information dissemination mechanisms. Importantly, the high virality rates for pro-BJP and anti-Congress messages (18.6% and 13.8% respectively from Table 2) indicate not just efficient dissemination strategies, but also genuine public engagement. This suggests that the messages are being actively forwarded by everyday users, revealing an underlying level of popular support.

Moreover, the robust presence of pro-BJP content in our data resonates with the party's documented efforts at various organizational tiers, including booth-level groups, to saturate the information ecosystem (Perrigo 2019). While these efforts had been previously documented, their actual reach and influence beyond party lines remained unclear until this study. Our work illuminates this aspect, affirming that the party's narratives have infiltrated general discourse to a significant extent.

The conspicuous absence of regional content, making up only 8% of the viral messages, also deserves special attention. The focus on national politics at the expense of regional issues could be reflective of a broader trend toward the 'nationalization' of the political discourse. This parallels findings in other contexts, like the United States, where local news and issues are increasingly sidelined (Martin and McCrain 2019). Furthermore, the trend aligns with existing literature discussing the centralization of politics in India (Jaffrelot 2022).

Collectively, these findings extend far beyond mere party politics. They raise critical questions about the evolving dynamics of information dissemination and political engagement in India, set against a backdrop of global trends. With increasing nationalization and centralization of politics, we must question what is lost when local issues are obscured. Our data thus not only quantifies but also qualitatively enriches our understanding of how political narratives are constructed, consumed, and perpetuated in the age of social media.

5.5 Prevalence of entertainment content

In line with previous qualitative studies on first-time internet users, our dataset substantiates that entertainment is a predominant category of content consumed (Rangaswamy and Arora 2016). Our sample revealed a diverse range of entertainment materials, encompassing humor clips, non-political commentary, educational pieces on well-being, information about government programs, and more. Interestingly, such messages along with religious content, political satire, and "good morning" messages constituted over a third of the virally disseminated materials (Purnell 2018).

Our findings also demonstrate the presence of nationalist undertones within these entertainment segments. For instance, the dataset included videos lamenting the perceived erosion of Indian culture or criticizing Western food habits. Additionally, our data incorporated misinformation, such as the debunked claim that UNESCO awarded the Indian national anthem as the best anthem in the world.

It is surprising that our analysis revealed that viral messages contain more misinformation (26%) than entertainment content (21.8%, see Table 2), a concern that merits further investigation. This is particularly noteworthy because, according to prior research, the primary usage of WhatsApp in similar demographics is geared towards entertainment (Rangaswamy and Arora 2016).

The ubiquity of "good morning" messages in our dataset echoes findings from other studies (Purnell 2018), indicating that these messages are not merely cultural quirks but represent a wider phenomenon in digital communication in rural India.

The implications of these findings are multifold. First, they signal the need for a deeper analysis of entertainment as a vector for various social, cultural, and political messages. Second, the surprisingly high prevalence of misinformation amidst entertainment content necessitates targeted interventions for effective public awareness and information literacy. Finally, understanding these prevalent communication patterns can offer insights into the unique digital behaviors of rural Indian WhatsApp users, thereby informing more culturally and contextually relevant tech policies and initiatives.

5.6 Fact-checking and its impact

Our analysis underscores a startling yet critical observation: not a single one of the 158 pieces of misinformation across various types of groups was counteracted or rebutted. This extends to the entire 600 pieces of content we examined, where no form of fact-checking—be it images, videos, text, or links—was present. This scenario is paradoxical, especially since much of the misinformation circulating on WhatsApp had already been debunked by mainstream fact-checking agencies.

This implies one of two scenarios: either the fact-checks are failing to penetrate the circles in which misinformation circulates, or they are ineffectual in deterring users from sharing previously debunked information. Our data leans towards the former hypothesis. The resilience of debunked content is astonishing—60 of the 158 misinformation posts were already fact-checked. We noted instances where identical pieces of misinformation related to health have been systematically fact-checked annually by reputable agencies since 2017. Yet, these debunked claims continue to disseminate widely on WhatsApp, reflecting the lack of effectiveness of current fact-checking strategies in reaching the target audiences.

The current model for fact-checking on WhatsApp relies on tiplines, which operate on an opt-in, reactive basis (WhatsApp 2021; Kazemi et al. 2022). This is fundamentally flawed for several reasons. First, mass adoption is lacking. Second, and perhaps more crucially, trust in the WhatsApp ecosystem typically arises through personal networks, creating a barrier to external information, including fact-checks.

The staggering ineffectiveness of current fact-checking mechanisms calls for a radical rethinking of our approach. Given the trust dynamics and the persistence of debunked content, a more integrated and proactive fact-checking mechanism involving platform operators, community leaders, and policy stakeholders seems imperative.

6 Discussion

The paper presents a mixed-methods analysis of viral content spreading in 164 private WhatsApp groups collected from participants in rural India. This is the first study looking at viral content on WhatsApp, in a rural setting, giving us a look at every day consumption.

The data reveals a disconcerting proliferation of misinformation, constituting over 25% of viral content, much of which is also tinged with anti-Muslim sentiment. These narratives appear to be orchestrated to amplify specific, often troubling themes–such as hate speech targeting Muslims–thus warranting serious consideration of their societal and national ramifications. Such a constant deluge of divisive content raises concerns about its long-term psychological impact, particularly in fostering a climate of hostility towards Muslims. While the specific media consumption habits of Muslim individuals within this context remain unknown, the overall trend is troubling. Invoking Anderson et al.'s concept of the "saffronization of the public sphere," (Anderson and Jaffrelot 2018) the findings highlight an unsettling trajectory potentially leading to offline violence, as observed in situations like Sri Lanka and Myanmar. This study offers empirically grounded alerts, demanding immediate policy interventions to curb this escalating issue before large scale violent clashes between communities threaten the fabric of the country.

In the digital sphere, we observe that the entrenched caste-based hierarchies that mar Indian villages are faithfully replicated. Specifically, WhatsApp groups organized around caste and religious identities serve as major repositories for misinformation. The high frequency of hate speech and misinformation in these segmented communities may be indicative of the deep-seated trust and comfort that members share with each other. This trust, in turn, mirrors offline ideologies steeped in hate and a flagrant disregard for verifiable information. Far from being mere platforms for conversation, these groups function as insular echo chambers that resist external corrections, effectively eliminating any opportunities for fact-checking or dissemination of accurate information.

While existing studies have documented the BJP's extensive utilization of WhatsApp for political messaging (Perrigo 2019), it was previously unclear if this content effectively proliferated beyond groups directly operated by the party. Our data clarifies this picture, revealing that approximately one in five pieces of viral content in the monitored WhatsApp groups was unmistakably in favor of the BJP. When we include messages specifically targeted at Hindus, as well as those containing misinformation, the proportion increases to nearly one-third. The conspicuous absence of alternative political perspectives is a matter for concern. In a functioning democracy, political parties should vie for public support through open debate and diversified messaging. While it may not be inherently problematic for the BJP to gain such extensive reach in this medium, questions arise: When does this become an undue advantage? Is it even possible for other parties to compete under these skewed conditions? The implications for democratic discourse and electoral competition in this asymmetric landscape of information distribution remain ambiguous.

Our research underscores a disconcerting reality: fact-checking appears to be an exercise in futility within these WhatsApp groups. Not a single instance of fact-checked content surfaced in our analysis, highlighting a deep-seated issue with information veracity in these settings. Moreover, there is a pervasive lack of awareness—or perhaps interest—about the concept of fact-checking among group members. The reactive nature of current fact-checking methods is exacerbated by WhatsApp's end-to-end encryption, making it impossible to proactively counter misinformation at the source. Even narratives that have been discredited nearly a decade ago continue to circulate virally, undeterred by the efforts of mainstream fact-checking agencies.

The real challenge, then, is not just writing or updating fact-checks, but ensuring they reach the appropriate audiences—a feat that current methods are failing at. In this context, a more grassroots approach may be essential. Leveraging community-driven fact-checking models, which are designed to be participatory and decentralized, might be one

of the few sustainable ways to combat this issue. Such a system could better capitalize on the inherent trust and social ties within WhatsApp groups, creating a more resilient and adaptable mechanism for information verification.

This situation raises critical questions about the effectiveness of current methods and the urgent need for innovative solutions that can adapt to the unique challenges posed by encrypted platforms. It implies that the battle against misinformation may need to shift from simply debunking false narratives to fundamentally altering the ecosystem that allows these narratives to thrive unchecked.

Our research comes with several limitations that warrant discussion while interpreting the results:

Sample size and convenience sampling: While our dataset presents a disturbing prevalence of politically charged and misleading narratives, it is important to remember that the data originates from a relatively small, convenience sample of rural WhatsApp groups in India. Additionally, to protect user privacy, we only chose groups with a certain size and activity, thus missing out personal conversations where misinformation could be shared. Consequently, the scope for generalization remains limited. However, the presence of such troubling content in even a small sample raises urgent questions about the extent and depth of the issue in larger, more diverse populations.

Observational nature: The study is essentially observational, limiting our ability to ascertain the intent behind the forwarding of misinformation or the belief in the misinformation. While our analysis focuses on content exposure rather than spread dynamics, the data doesn't reveal whether individuals forwarded messages knowingly or unknowingly. This limits our understanding of user motivations and the psychological underpinnings of the dissemination process.

Prevalence vs. Exposure: In our analysis, it is vital to clarify that the proportion of misinformation among viral messages does not directly translate into the same proportion of misinformation in the total messages each user encounters. However, this distinction should not minimize the severity of our findings. While misinformation may not dominate an individual's daily message feed, the fact that debunked misinformation is still reaching a viral status–especially when laden with hate speech–is deeply concerning. This recurring cycle of misinformation, even after public debunking, underscores a systemic issue that extends beyond the scope of individual exposure and speaks to a larger, more pervasive problem.

Geographic and cultural context: While the study is focused on a rural Indian context, one might argue that its findings have limited applicability beyond this particular demographic. However, we posit that our methodology and findings have broader implications. The mechanics of misinformation and political propaganda can be similar across different contexts, making our research a valuable case study for understanding the dynamics on end-to-end encrypted platforms globally.

These limitations, far from undermining the study, serve as important qualifiers that provide direction for future research. They raise compelling questions about the true extent of misinformation spread and political propaganda on encrypted platforms, while also pushing for nuanced approaches that account for local context and individual behavior. As the first quantitative analysis of its kind, this study lays down the groundwork for more comprehensive investigations that could build on top of our findings.

References

Anderson, E.; and Jaffrelot, C. 2018. Hindu nationalism and the 'saffronisation of the public sphere': An interview with Christophe Jaffrelot. *Contemporary South Asia*, 26(4): 468–482.

Anwary, A. 2020. Interethnic conflict and genocide in Myanmar. *Homicide studies*, 24(1): 85–102.

Arnimesh, S. 2020. 9,500 IT cell heads, 72,000 WhatsApp groups — how BJP is preparing for Bihar poll battle — theprint.in. https://theprint.in/politics/9500-it-cell-heads-72000-whatsapp-groups-how-bjp-is-preparing-for-bihar-poll-battle/451740/. [Accessed 16-09-2023].

Arora, P. 2016. Bottom of the data pyramid: Big data and the global south. *International Journal of Communication*, 10: 19.

Arun, C. 2019. On WhatsApp, rumours, lynchings, and the Indian Government. *Economic & Political Weekly*, 54(6).

Banaji, S.; Bhat, R.; Agarwal, A.; Passanha, N.; and Sadhana Pravin, M. 2019. WhatsApp vigilantes: An exploration of citizen reception and circulation of WhatsApp misinformation linked to mob violence in India.

Chakrabarti, S.; Stengel, L.; and Solanki, S. 2018. Duty, identity, credibility: Fake news and the ordinary citizen in India. *BBC World Service Audiences Research*.

Chauchard, S.; and Garimella, K. 2022. What circulates on partisan WhatsApp in India? Insights from an unusual dataset. *Journal of Quantitative Description: Digital Media*, 2.

Dash, S.; Grover, R.; Shekhawat, G.; Kaur, S.; Mishra, D.; and Pal, J. 2022. Insights Into Incitement: A Computational Perspective on Dangerous Speech on Twitter in India. In *ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies (COMPASS)*, 103–121.

Garimella, K.; and Eckles, D. 2020. Images and misinformation in political groups: Evidence from WhatsApp in India. *Harvard Kennedy School Misinformation Review*.

Hall, N.-A.; Lawson, B.; Vaccari, C.; and Chadwick, A. 2023. Beyond quick fixes: How users make sense of misinformation warnings on personal messaging.

Haque, M. M.; Yousuf, M.; Alam, A. S.; Saha, P.; Ahmed, S. I.; and Hassan, N. 2020. Combating Misinformation in Bangladesh: roles and responsibilities as perceived by journalists, fact-checkers, and users. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2): 1–32.

Jaffrelot, C. 2022. Populism Against Democracy or People Against Democracy? In *Contemporary Populists in Power*, 35–53. Springer.

Jaffrelot, C.; and Verniers, G. 2020. Assessing the 2019 Indian Elections. *Contemporary South Asia*, 28(2 (June 2020)): 113.

Javed, R. T.; Usama, M.; Iqbal, W.; Qadir, J.; Tyson, G.; Castro, I.; and Garimella, K. 2022. A deep dive into COVID-19-related messages on WhatsApp in Pakistan. *Social Network Analysis and Mining*, 12: 1–16.

Juneja, P.; and Mitra, T. 2022. Human and technological infrastructures of fact-checking. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2): 1–36.

Kazemi, A.; Garimella, K.; Shahi, G. K.; Gaffney, D.; and Hale, S. A. 2022. Research note: Tiplines to uncover misinformation on encrypted platforms: A case study of the 2019 Indian general election on WhatsApp. *Harvard Kennedy School Misinformation Review*, 3(1).

Martelli, J.-t.; and Jaffrelot, C. 2023. Do populist leaders mimic the language of ordinary citizens? Evidence from India. *Political Psychology*.

Martin, G. J.; and McCrain, J. 2019. Local news and national politics. *American Political Science Review*, 113(2): 372–384.

Nielsen. 2019. Nielsen's Bharat 2.0 Study reveals a 45Internet Users in rural India since 2019 — Nielsen — nielsen.com. https://www.nielsen.com/news-center/2022/nielsens-bharat-2-0-study-reveals-a-45-growth-in-active-internet-users-in-rural-india-since-2019/. [Accessed 16-09-2023].

Perrigo, B. 2019. How Whatsapp Is Fueling Fake News Ahead of India's Elections — time.com. https://time.com/5512032/whatsapp-india-election-2019/. [Accessed 14-09-2023].

Purnell, N. 2018. The Internet Is Filling Up Because Indians Are Sending Millions of 'Good Morning!' Texts — wsj.com. https://www.wsj.com/articles/the-internet-is-filling-up-because-indians-are-sending-millions-of-good-morning-texts-1516640068. [Accessed 15-09-2023].

Quek, N. 2019. Bloodbath in Christchurch: The rise of far-right terrorism. *RSIS Commentaries*, 19: 1–3.

Rajadesingan, A.; Panda, A.; and Pal, J. 2020. Leader or party? Personalization in twitter political campaigns during the 2019 Indian Elections. In *International conference on social media and society*, 174–183.

Rangaswamy, N.; and Arora, P. 2016. The mobile internet in the wild and every day: Digital leisure in the slums of urban India. *International Journal of Cultural Studies*, 19(6): 611–626.

Rao, M. 2011. Love Jihad and demographic fears. *Indian Journal of Gender Studies*, 18(3): 425–430.

Resende, G.; Melo, P.; Sousa, H.; Messias, J.; Vasconcelos, M.; Almeida, J.; and Benevenuto, F. 2019. (Mis) information dissemination in WhatsApp: Gathering, analyzing and countermeasures. In *The World Wide Web Conference*, 818–828.

Saha, P.; Mathew, B.; Garimella, K.; and Mukherjee, A. 2021. "Short is the Road that Leads from Fear to Hate": Fear Speech in Indian WhatsApp Groups. In *Proceedings of the Web conference 2021*, 1110–1121.

Table 3: Descriptions of the group categories.

Category	Description
Village	Has the village name in it and is apolitical
Caste	Has the name of any caste
Religious	Has the name of any religion, a god or symbol of a religion
Hindutva	Has the BJP, a hindutva idealogue, or a hindutva group
Activism	Is non-political and demands any rights
Regional	Has the name of any village, town, city, district of Jharkhand and not any of the five categories listed above
Others	Can not be categorized into any of the six groups listed above e.g. fun group, friends group, hobbies, etc.

Times, H. 2017. Jharkhand lynching: When a WhatsApp message turned tribals into killer mobs — hindustantimes.com. https://www.hindustantimes.com/india-news/a-whatsapp-message-claimed-nine-lives-in-jharkhand-in-a-week/story-xZsIlwFawf82o5WTs8nhVL.html. [Accessed 16-09-2023].

Varanasi, R. A.; Pal, J.; and Vashistha, A. 2022. Accost, accede, or amplify: attitudes towards COVID-19 misinformation on WhatsApp in India. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 1–17.

WhatsApp. 2021. IFCN fact-checking organizations on WhatsApp — WhatsApp Help Center — faq.whatsapp.com. https://faq.whatsapp.com/5059120540855664. [Accessed 16-09-2023].

Appendix

Definitions of group categories

We manually classified all the 164 groups into 7 categories. Table 3 shows the definitions of the group categories.

Definitions and examples of content categories

1. Misinformation: Incorrect or misleading information. Since the annotator was a trained fact-checker, any claim which we found to be false by fact-checking them was tagged as misinformation. There were contents which could not be fact-checked, e.g. opinions. These are not included in this category.
2. Information/Inspirational videos/Commentary/Amusement videos/Religious harmony/Educational messages: A catch-all category containing videos providing information, inspirational videos, providing commentary/opinion about certain (non political) issues, educating people, or promoting religious harmony.
3. Religious propaganda to influence Hindus: Messages targeted to make Hindus feel that their rights are being taken, they are discriminated, deprived of their religion, fear mongering about the extinction of Hindus/Hinduism, etc.

4. Hate against Muslims: Characterized by or expressing hostility or discrimination toward Muslims or the Islamic faith, including content that clearly and deliberately incites hatred against Muslims. Examples include: asking Hindus to unite, Muslims rulers being looters and thus justifying attrocities against Muslims, Muslims being traitors/Pakistan supporters, etc.
5. Pro-BJP political propaganda: Information–facts, arguments, rumours, half-truths, or lies–to influence public opinion in favor of the BJP. Includes any post that endorses BJP, Uttar Pradesh CM Yogi Adityanath, Prime Minister Modi or defends BJP, or other BJP leaders.
6. Anti-Congress political propaganda: Content targeted towards the Congress party or any of its leaders Nehru, Sonia Gandhi, Rajiv Gandhi or Rahul Gandhi, etc.
7. Regional information: Regional news or are related to local demand of Jharkhand. The messages in the regional news were mostly related to demands for jobs.
8. Religious: Any content which is not political and is not right-wing or left wing but contains mentions of god.
9. Good morning messages: Messages wishing people good morning.
10. Political or religious sarcasm/satire: Spoof, Humor, Sarcasm to advance political or religious propaganda.
11. Political opinion not to benefit any political party: Political opinion not to endorse or malign any particular party.
12. Health misinformation: Any health claim that lacks evidence or a claim that goes against current evidence.
13. Anti-BJP propaganda: Content targeted against BJP and its leaders.