Digital Threads in the Social Fabric: An Examination of WhatsApp's Role Across Diverse Communities in the Global South

ANONYMOUS AUTHOR(S)

This research examines the nature and spread of viral WhatsApp content among everyday users in two distinct contexts: a rural Indian village and Indonesian university students. By analyzing data collected with participants' consent from hundreds of private WhatsApp groups —164 groups from India and 640 from Indonesia— we provide the first such categorization in these regions. Despite the disparate settings, our findings reveal surprising similarities in the types of groups users engage with and the viral content they receive, particularly in the prevalence of misinformation and emotionally charged messages.

Our comparative analysis shows that viral content often includes political and religious narratives, with misinformation frequently recirculated despite prior debunking by fact-checking organizations. These parallels suggest that closed messaging platforms like WhatsApp facilitate similar patterns of information dissemination across different cultural contexts. This study offers valuable insights into the challenges posed by end-to-end encrypted platforms and underscores the necessity for user-centered design interventions. By highlighting these dynamics, our work is highly relevant to the HCI community, providing a crucial baseline for developing effective moderation policies and fostering responsible use of encrypted communication channels.

ACM Reference Format:

Anonymous Author(s). 2024. Digital Threads in the Social Fabric: An Examination of WhatsApp's Role Across Diverse Communities in the Global South. 1, 1 (September 2024), 19 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

WhatsApp, a cornerstone of global digital communication, profoundly impacts billions of users worldwide. Yet, its role as an information conduit, especially in diverse settings such as rural India and Indonesia, remains critically understudied. This oversight is particularly concerning given the severe real-world consequences that misinformation on this platform has instigated, such as social unrest and violence. Our research seeks to bridge this gap by analyzing the spread and nature of 'viral' content among distinct populations: villagers in central India and university students in Indonesia. These regions represent a significant portion of WhatsApp's user base and are emblematic of the challenges and opportunities that arise in the Global South's digital ecosystems.

Both communities, while disparate in their socio-economic and cultural fabrics, reveal the dual nature of WhatsApp as a tool for widespread misinformation and a medium for everyday communication. This study, through a mixed-methods approach involving tens of thousands of messages from private groups, offers unprecedented insights into the types of digital communities formed and the content that defines their information diets. Notably, our findings expose a high prevalence of misinformation intertwined with religious and political content, highlighting a common thread of digital discourse that transcends geographical and cultural boundaries.

The complexity of researching WhatsApp stems from its encrypted, private nature and the absence of public APIs, which poses significant challenges for systematic data collection and ethical concerns regarding user privacy. Despite

Manuscript submitted to ACM

these obstacles, our methodology captures a broad spectrum of 'forwarded many times' messages, allowing us to identify macro trends in viral content dissemination.

Prior studies in this domain have primarily been qualitative, focusing on small samples, which provide deep insights but can not capture macro trends [34]. These studies do not fully appreciate the textured layers of personal WhatsApp use, especially concerning how viral messages resonate within different social groups.

The results reveal a high prevalence of misinformation, constituting over 25% in India and 30% in Indonesia of the viral content examined. We found a notable convergence between religious content and false information. Specific narratives, including those that perpetuate the notion that Hindus are under threat from Muslims and hatred towards Indonesia's current central government, were systematically amplified.

While prior research has noted the Bharatiya Janata Party's (BJP) substantial use of WhatsApp for political communication [21], 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.

In Indonesia, friends and school groups were dominant, with religious groups following not far behind. Interestingly, many misinformation did not come from the major groups. Rather, they came from mainly religious, regional, and family groups. Moreover, viral content in Indonesia still have plenty of political content even when we found no groups created specifically for politics.

This paper contributes to the HCI research by detailing how localized digital practices inform broader usage patterns and affect social dynamics. Our findings underscore the urgent need for HCI researchers to focus on underrepresented regions, advancing our understanding of how global platforms like WhatsApp can both support and undermine community resilience and social cohesion. By highlighting the nuanced ways in which misinformation spreads and is contested, we advocate for more inclusive digital communication tools that respect and reinforce local cultural practices, thereby supporting more robust democratic processes and social trust.

In doing so, this research not only addresses a critical gap in the current literature but also serves as a call to action for the HCI community to prioritize studies that explore the intersection of technology, misinformation, and society in the Global South. Such efforts are essential for designing interventions that can effectively counteract the adverse effects of misinformation and enhance digital inclusion, ensuring that technological advancements benefit all segments of society equally.

2 Background

The rise of social media platforms like WhatsApp has reshaped communication landscapes globally, but nowhere is this transformation more pronounced than in the Global South. In India, WhatsApp has amassed a user base exceeding 500 million, while in Indonesia, the platform engages around 100 million users. This vast penetration highlights WhatsApp's pivotal role in daily communications, significantly influencing social, political, and cultural dynamics [7]. These countries, characterized by their high population densities and rapidly expanding digital infrastructures, offer unique insights into the digital behaviors that pervade less-studied regions of the world.

Social media's exponential growth has, however, been shadowed by the spread of digital misinformation, leading to severe outcomes including lynchings, civic unrest, and increased political polarization [1, 4]. Prior research has extensively documented the impact of misinformation in urban settings, but there remains a significant void in our understanding of how it permeates and affects the rural populace, whose information diets are predominantly shaped by mobile-first internet access [3]. This study aims to bridge this gap by digging into the nuances of information

consumption in rural India and among university students in Indonesia, thereby providing a comparative glimpse into how misinformation spreads across different community structures within the Global South.

In terms of Human-Computer Interaction (HCI), the exploration of how digital tools like WhatsApp influence community dynamics in diverse socio-economic settings is still emerging. Existing literature often focuses on technologically advanced settings, leaving a lacuna in comprehensive studies that address the integration of technology in everyday life in rural and semi-urban areas of developing countries. Studies like that by Varanasi et al. [34] have begun to explore how digital platforms are adapted to local needs and languages, revealing a complex interplay between technology adoption and cultural context. Furthermore, research by Banaji et al. [6] underscores the role of digital platforms in forming new social networks that transcend traditional boundaries such as caste and religion, which are otherwise rigid in rural settings.

This background sets the stage for our study, which examines WhatsApp's dual role as both a facilitator of connectivity and a conduit for misinformation, considering the platform's deep embedment into the local cultural and social fabrics of India and Indonesia. By mapping out the information diets and interaction patterns among users in these countries, this research contributes to the broader HCI discourse, advocating for more inclusive and contextually aware technology design and policy interventions to mitigate the adverse effects of misinformation on society.

3 Related Work

Misinformation on WhatsApp. The spread of misinformation on WhatsApp has been extensively studied in both India and Indonesia, revealing common challenges and distinct contextual nuances that shape how information is consumed, propagated, and responded to in these countries.

Prior research has highlighted the pervasive role of WhatsApp in misinformation spread, often linking it to broader socio-political dynamics. Varanasi et al. [34] examined how rural and urban communities in India interact with misinformation, highlighting the influence of social status, local norms, and community deliberations on information sharing. Similarly, Syam et al. [33] analyzed COVID-19 misinformation in Indonesian WhatsApp groups, finding that content decisions were driven more by personal beliefs and emotions than by fact-checking, contributing to public panic and diminished trust in the platform. These studies underscore a recurring theme: misinformation is often propagated within trusted social circles, where the identity of the sender outweighs the credibility of the message.

Both contexts show that misinformation is not just a byproduct of digital interaction but is deeply embedded in the socio-political fabric of the societies. For instance, Banaji et al. [6] and Arun [5] documented the violent repercussions of misinformation in India, particularly in how anti-Muslim sentiments and political propaganda are amplified within closed groups. Comparable patterns were observed in Indonesia during the 2019 presidential elections, where misinformation and 'HOAX' posts were extensively shared due to low digital literacy and emotional engagement with content, as noted by Susilo et al. [32]. These studies suggest that misinformation on WhatsApp is not merely a technical issue but reflects deeper societal schisms around class, religion, and politics.

Across both India and Indonesia, WhatsApp functions as a platform where political and religious narratives frequently dominate, often exacerbating social divides. Chakrabarti et al. [8] identified recurring themes of nationalism, cultural pride, and political superiority in Indian WhatsApp groups, paralleling findings by Neyazi et al. [19], who observed that Indonesian WhatsApp users were more prone to political misinformation compared to other platforms like Facebook and Twitter. The end-to-end encryption of WhatsApp, while protecting privacy, simultaneously creates challenges for monitoring and mitigating the spread of false information.

Despite these shared challenges, a common gap in existing research is the lack of systematic analysis of viral content within private WhatsApp groups at scale, particularly across different types of groups. Our study addresses this gap by examining viral content in hundreds of private groups in rural India and among Indonesian university students. By analyzing tens of thousands of messages, our research extends the current literature by providing empirical evidence of the types of content that gain virality and the group dynamics that facilitate this spread. Our mixed-methods approach offers a nuanced view of how misinformation is not just a consequence of digital media but is also a reflection of underlying social structures, biases, and emotional triggers.

Fact-checking in the Global South. Efforts to understand and address misinformation on digital platforms, particularly within the Global South, have highlighted significant challenges related to fact-checking and countering misinformation on private and encrypted platforms like WhatsApp. Across both India and Indonesia, the complex social dynamics within WhatsApp groups, such as familial hierarchies, community trust, and socio-political biases, present considerable barriers to the effectiveness of traditional fact-checking methods.

In India, studies like Juneja and Mitra [14] and Haque et al. [13] delve into the fact-checking ecosystem, emphasizing the resource constraints and contextual challenges faced by fact-checkers. These works underscore how misinformation circulates unchecked, especially in private groups where fact-checking information struggles to penetrate. Similarly, Seelam et al. [28] reveal that the reach and impact of fact-checking efforts are limited in rural areas, where misinformation continues to thrive within trusted social circles, often unaffected by external corrections.

In Indonesia, Widiatmika [37] and Suminar and Hadisiwi [31] explore how entrenched social hierarchies within family and community groups further hinder the spread of fact-checked information. These studies demonstrate that in familial WhatsApp groups, older members often dominate discussions, enforcing their views and stifling counter-arguments, creating environments where misinformation is less likely to be challenged. The phenomenon of "hoax busters" identified in these studies highlights that only individuals with close relationships to the misinformation spreaders feel empowered to challenge false information, often motivated by personal experiences or concerns rather than institutional support.

Technical solutions have been proposed to address these challenges, including crowd-sourced tiplines for flagging misleading content [15] and the implementation of message tags like “forwarded many times" to signal potential misinformation [12]. However, these approaches face limitations in both contexts. The tags’ ambiguous impact on user behavior and the difficulty of implementing proactive interventions on encrypted platforms reduce the overall efficacy of these solutions.

Our research builds on these insights by revealing the persistent nature of already-debunked misinformation that continues to circulate within WhatsApp groups in India and Indonesia. We identify specialized misinformation trends, such as the propagation of health misinformation and politically charged content, that are unique to specific social contexts. This highlights the need for more targeted, context-specific moderation policies and strategies that go beyond reactive fact-checking models. Our findings call for a shift toward community-driven approaches that leverage the existing social structures within groups, promoting critical engagement with content and enhancing the reach and relevance of fact-checking efforts on private digital platforms.

4 Data Collection

In our study, we employed a tool we custom built to collect data from two distinct cultural settings: a village in Jharkhand, India, and universities across Indonesia. This tool was specifically designed to ensure the privacy and anonymity of participants’ data, with a user-friendly interface that allowed participants to donate their WhatsApp data.

Table 1. Details of the datasets from India and Indonesia.

	N	Age range	Gender	Religion	Num. groups	Num. messages
India	31	21-34	Male: 100%	Hindu: 100%	164	53,389
Indonesia	74	18-24	Male: 22% Female: 78%	Islam: 70% Christian: 30%	640	496,197

Once connected, the tool filtered out groups with fewer than six members and automatically anonymized personal information, including names, phone numbers, and emails, using Google's Data Loss Prevention library. Furthermore, any images collected had faces blurred to maintain privacy.

In India, we focused on a rural village in Jharkhand, an area noted for its communal sensitivity and frequent incidents of mob violence, which underscored the importance of studying the spread of misinformation. The recruitment process was conducted on the ground through a native researcher, which facilitated a deeper understanding and trust with participants. A total of 31 male participants from the village consented to share their data, providing insights from 164 WhatsApp groups over two months in June–August 2023. The dataset comprised 53,389 messages, predominantly from large groups with a median size of 33 members, ranging from community news to religious and political discussions.

In Indonesia, we targeted a tech-savvy demographic of university students from various universities. Recruitment was done digitally via a popular Twitter account with over 100,000 followers, @collegemenfess used by Indonesian students. We recruited 74 participants, who contributed data from 640 WhatsApp groups which had a median group size of 35. The Indonesian dataset, collected during March–July 2024, included a vast array of messages totaling 496,197, reflecting the rich tapestry of student life and interactions across different social and religious groups.

The datasets from both contexts provided a unique perspective on how WhatsApp is used in varying social settings, revealing both commonalities and differences in digital communication practices. In India, the data illuminated the alignment of group interactions along caste and religious lines, while in Indonesia, the spread of content was seen across more diverse and informal groups like those formed around school projects and social activities.

This extensive collection process, facilitated by our data donation tool, was meticulously documented and received IRB approval, ensuring that all ethical considerations, especially those concerning participant privacy and data security, were rigorously adhered to. The careful curation and analysis of these datasets enables the study of digital communication in the Global South and provide valuable insights into the role of encrypted messaging apps in shaping social dynamics.

While our datasets provide unique insights into WhatsApp usage within specific communities in India and Indonesia, they are not fully representative of the broader populations in these regions. In India, the dataset primarily reflects the perspectives of Hindu males, with no participation from the Muslim minority, who make up a significant portion of the village's population. Similarly, in Indonesia, the focus on university students from Java may not encompass the diverse experiences of individuals from other islands or different demographic groups. Table 1 shows the details of our datasets, including the sample size and the demographics.

Despite these limitations, our study is a pioneering effort in the HCI field, capturing data from underrepresented communities through ethically sound methods and explicit consent from participants. This research provides valuable glimpses into the digital interactions on WhatsApp, offering a rare and valuable perspective that is seldom seen in HCI studies.

Table 2 shows the categories of the WhatsApp groups. Even though the samples are from very different contexts, it is interesting to see a convergence in the types of groups they have.

Table 2. Categories of the WhatsApp groups.

Category	India %	Indonesia %
Regional	15.2	17.1
Activism	12.2	3.1
News	12.3	4.6
Friends	9.7	32.8
Village/Town	6.7	9.3
Religious	6.7	25.0
Family	6.6	15.6
Hindutva	8.5	-
Caste	7.3	-
University classes	-	31.2
Others	6.6	6.2

5 Data Annotation

In this study, both datasets from India and Indonesia focus on content classified as 'forwarded many times' on WhatsApp, a label used to signify viral spread within the platform. This designation indicates that a message has been forwarded through a chain of five or more separate users from the original sender.[1] The ambiguity of this definition presents challenges but also serves as a useful proxy for identifying widely circulated content in the WhatsApp ecosystem.

The core datasets consist of viral messages identified within the larger collections from each country – 604 messages from India and 177 from Indonesia. To manage these, we utilized a custom dashboard designed to facilitate the annotation process, enabling reviewers to view the context of each message by displaying the adjacent messages (10 before and after the focal message). This setup helped in understanding the interactions surrounding the viral content.

Both datasets underwent a rigorous qualitative coding process conducted by native experts who were deeply familiar with the cultural and social contexts of the respective datasets. These annotators were responsible for categorizing each message based on emergent themes and refining these into organized parent categories through multiple iterations of review and discussion among the research team.

In India, the coding was carried out by a skilled fact-checker but also native to the studied village, providing an intrinsic understanding of local dynamics such as caste hierarchies and community-specific linguistics. The annotation process identified 13 significant categories, such as misinformation combined with hate speech against Muslims, with each content potentially fitting into multiple categories. Misinformation was particularly scrutinized; each piece of content was checked for the accuracy of its claims using tools like reverse image searches and standard internet searches for text verification.

Similarly, in Indonesia, the annotator was a native Indonesian and a university student, which was crucial for interpreting the content within the contemporary social and educational contexts of Indonesian youths. The annotation resulted in 15 distinct categories, including misinformation and pro-China sentiments. Fact-checking was facilitated by the Indonesian Communication Agency's online resources, which often include 'HOAX' verification articles that helped in validating the misinformation claims found in viral posts.

For both datasets, the annotation process included documenting whether the viral content was fact-checkable and if it had been fact-checked. Additionally, the team qualitatively analyzed and recorded the types of responses each viral post received within the group chats, such as corrections or expressions of support.

[1]The exact definition by WhatsApp can be found here https://faq.whatsapp.com/1053543185312573

This meticulous annotation approach not only categorized the content but also provided insights into the social reactions to misinformation, enhancing our understanding of how misinformation spreads and is received in different cultural contexts. The categories and their prevalence are detailed in Table 3. The detailed code book containing the rubric is available in the Appendix.

Table 3. Categories of viral content on WhatsApp in India and Indonesia. The percentage does not sum up to 100 since one post can be in more than one category.

Category	**India %**	**Indonesia %**
Misinformation	26	30.5
National/International News	23.2	29.8
Inspirational/Informational Videos	21.8	20.9
Religious Propaganda	21.0	4.5
Hate Speech	19.6	2.8
Political Propaganda	18.6	3.9
Regional Information	8.5	32.7
Religious Content	7.3	7.3
Humor/Sarcasm	6.2	14.1
Good Morning Messages	3.5	1.1
Health information	1.4	6.7
Exclusive to Indonesia		
Advertising	-	31.0
Chinese Content	-	6.7

6 Findings

6.1 Types of information by group

Table 2 delineates the categorization of WhatsApp groups analyzed in this study. In India, viral content is significantly skewed towards caste-based and Hindutva groups, despite these groups comprising a smaller proportion of the total messages. This disproportionality suggests that such groups are not only active but are pivotal in spreading viral messages, often imbued with misinformation and politically charged narratives. Specifically, these messages frequently promote pro-BJP and anti-Congress sentiments and are potent in fostering anti-Muslim sentiments among young adults.

Conversely, in Indonesia, while the groups are distinctly categorized, the spread of viral content such as religious devotionals and health advice is pervasive across various types of groups, including regional, family, and even local activism groups. This widespread distribution indicates a shared interest or concern across different group types, suggesting a uniform cultural engagement with certain types of content, irrespective of the group's primary focus.

In both regions, political content infiltrates non-political groups, highlighting the pervasive nature of political discourse on WhatsApp. In Indonesia, even groups not explicitly focused on politics disseminate significant political content, particularly anti-government narratives and pro-Chinese sentiments. This mirrors the trend in India, where non-politically aligned groups still end up circulating politically charged content, demonstrating the platform's role in political mobilization and opinion shaping across different demographics.

The findings underscore the prevalence of echo chambers within WhatsApp groups, where repeated exposure to unchallenged narratives, whether they be political, religious, or health-related, can reinforce existing biases and potentially lead to indoctrination. This is evident in both India and Indonesia, where digital spaces become extensions of

existing social structures, perpetuating segregation and misinformation under the veil of trusted communal networks [6, 30]. These patterns reveal the dual role of WhatsApp as a tool for community engagement and as a conduit for misinformation. The trust inherent in social and familial connections often supersedes the imperative for message verification, making these platforms powerful but also potentially harmful in shaping public opinion and social behavior.

6.2 Popular viral narratives

In the milieu of frequently forwarded messages on WhatsApp, certain narratives demonstrated 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 datasets. Though there is a significant overlap in the categories of content (see Table 3), the differences make the findings interesting. We will qualitatively discuss the prominent narratives separately in India and Indonesia in the rest of this section. We will reflect on the similarities and differences overall in Section 7.

6.2.1 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 included exaggerated accolades, overlooking any criticisms or complexities related to his tenure. These narratives served to consolidate the Prime Minister's standing and reinforce the party's core voter base and seem to be a persistent narrative being pushed by the BJP, since such narratives were also prevalent in qualitative work by Chakrabarti et al. [8] almost 5 years ago.

Fear-Mongering and Communal Tensions. A more concerning narrative revolved around instigating fear among Hindus. Messages in this category implied 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 were often specifically portrayed as victims in such messages, magnifying the perceived threat. This narrative also extended 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 were common, along with the propagation of conspiracy theories like 'Love Jihad' [26].

WhatsApp users in most of our groups frequently consumed hate speech and misinformation about minority communities within their WhatsApp networks. We identified a pervasive climate of distrust and antipathy towards Muslims, equating them to traitors who belong to Pakistan. 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 framed Muslims as existential threats to Hindus, a narrative that has chilling parallels with rhetoric that has catalyzed genocides in Myanmar and Sri Lanka [2], and fueled white supremacist actions such as the Christchurch massacre [23]. Prior work documented the wide spread use of such fear rhetoric in certain WhatsApp groups [27] or being used by Twitter influencers [10].

Discrediting the Opposition. Lastly, we saw messages with concerted effort to discredit opposition figures, particularly leaders from the Congress party. Rahul Gandhi was a frequent target, with false claims circulating about his religious identity and mental acumen. There were 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

accused the opposition of being anti-Hindu by alleging that they restrict Hindu festival celebrations and falsely labeled 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–we found 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.

6.2.2 Indonesia. **Career Advice**. This first narrative is a recurrent theme in our viral WhatsApp dataset as both the first two most common viral content categories in Table **??**, local information and advertising, have plenty of content that support this narrative. From seminars to study abroad programs, most of the viral messages in the local information and advertising categories contain *hooks* in which potential benefits such as 'being able to work and stay permanently abroad' or to 'get a job that pays millions' were almost always mentioned. These types of information and advertisements, in a way, 'reassured' Indonesian university students that as long as they participated in whatever event was being promoted, they will have a secure life post-graduation. Of course, these *hooks* could not be explicitly proven or fact-checked since most of them were self-claimed by the advertisers. However, viral content that fit this narrative seemed to desire to use college students' financial, personal insecurity and the feeling and fear of missing out to promote any available *self-development* events they were organizing.

Health (mis)information. This second narrative comes from viral content in the inspirational message, health advice, and religious categories in Table 3 in which a lot of the posts mentioned alternative ways to health that do not concern any traditional medical help. Posts that have this narrative would usually have either a self-proclaimed doctor or a set of testimonies from unidentifiable people explaining the benefits of the alternative methods, including spiritual, to help people to heal without the need of medical help.

There are two types of viral health content that show this narrative: preventive and corrective. Preventive posts mentioned alternative ways to achieve health before illnesses are experienced. An example of a preventive post was a content telling people to fog houses and offices to prevent the possibility of dengue fever. Another example advised people to forward and spread a spiritual message about Jesus to at least 13 other groups if they do not want misfortune or sickness to befall them in the next two weeks.

Corrective posts, on the other hand, mentioned ways to achieve healing once illnesses have been experienced. An example of a corrective post was an image promoting the ability for raw papaya juice to increase platelet levels instantly, healing dengue fever without the need of medical help. Another example mentioned onions as the absolute medicine to self-heal from poisonous snake bites. Although the purpose of preventive and corrective content were different, they served the same narrative: to be healthy without any medical help.

More than half of the health-related viral content, however, are also in the misinformation category. Most of them have been debunked by the Indonesian Communication Agency as 'HOAX' [24]. Some of the viral health-related content cannot be fact-checked or are entirely opinion-based. It is alarming that this narrative of 'alternative ways to achieve health' is being extensively promoted but is rarely true.

Anti-Indonesian, pro-China politics. Another interesting narrative we found in this study is that none of the viral political, partisan, and propaganda content in Table 3 are in support of Indonesia's current president, Jokowi, and his family of political elites. Multiple viral political content talked about how Gibran, Jokowi's son, should not be allowed to be a vice presidential candidate. Posts of people talking about how degrading Indonesia started to become after

Jokowi became president were also prevalent. The narrative these posts was trying to convey was obviously to go against Jokowi and his family's administration. This was highly relevant to Indonesia's 2024 presidential election where Gibran was one of the vice presidential candidates. This anti-Indonesian politics narrative wishes to discourage people from electing Gibran and strengthening Jokowi's political dynasty.

Interestingly, we also found plenty of Chinese political, propaganda, and entertainment content in our dataset. In contrast to the viral political content concerning Indonesian politics, we found that all of the viral Chinese-related content did not have even one criticism of China. These Chinese-related viral content often also have other countries for comparison. The United States was a common comparison country used, and for that fact, was antagonized against China. For example, a post was highlighting the United States' 'inappropriate' superiority complex towards China when all they have are homeless citizens instead of advanced infrastructure. It is obvious that these posts highlighted China's best traits to portray superiority against any other countries in the world, including Indonesia.

Prevalence of Advertisements. The prevalence of advertisements in the WhatsApp conversations of Indonesian university students stands at a notable 31%, reflecting a daily inundation of promotional content ranging from academic self-development programs to job listings and Chinese products. A significant portion of these advertisements, over half, promotes academic self-development opportunities such as seminars, webinars, and internship programs, often featuring eye-catching speakers with impressive but not necessarily relevant qualifications.

Moreover, the emotional appeal extends beyond academic advertisements to include products, exemplified by viral ads for Chinese-manufactured goods like water filters. These ads often feature dramatic demonstrations and inspirational messages, targeting the viewers' empathy and environmental concerns. This emotional strategy seems to be a common thread linking various forms of content, including misinformation, suggesting that emotive content significantly enhances virality among this demographic.

6.3 Misinformation

Table 3 shows that one of the most prevalent category of viral content was misinformation across both datasets (26% in India, and 30.5% in Indonesia). Notably, for both the countries, nearly half of these (78 out of 158 in India, 29 out of 53 in Indonesia) were in video format.

6.3.1 India. . The false narratives in India 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 and is notably higher than rates observed in other contexts [9] or WhatsApp datasets [11]. 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.

We also found that the misinformation was not random, but tailored to resonate with the prevailing sociopolitical sentiments. A case in point is a message circulating the 'Love Jihad' myth [26], 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 qualitatively observed a significant overlap of misinformation themes.

However, most striking is the role of mainstream media in this ecosystem. Our qualitative analysis revealed a concerning synergy between biased national media and the misinformation cycle. Television news clips favoring BJP were often shared virally in our dataset, 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.

We found that misinformation in groups was not purely cognitive, but also affective; it capitalized on emotional resonance [38]. This is in line with Garimella and Eckles [11] who show that during key events like elections or riots, the urgency and shock value of messages surge, often inflating simmering tensions into flashpoints of potential violence.

6.3.2 Indonesia. . Contrary to the Indian case, we found that in Indonesia, the false narratives were mainly in health advice, and religion and spirituality. Health advice misinformation is very dominant in our dataset, with more health-related viral content identified as false narratives (16 posts) instead of those that are not (12 posts). A lot of these health advice misinformation were also properly tailored to gear towards college students' lifestyles. For example, there was a post saying that a college student living alone ate chocolate right after having instant noodles for dinner. Instant noodles are common meals for university students in Indonesia, as they are cheap, delicious, and available everywhere. The student died instantly after bleeding nonstop from all her five senses due to the mix of instant noodles and chocolate that disrupted the stomach's chemicals. This story itself has been published as 'HOAX' by the Indonesian Communication Agency. However, the narrative was still widely forwarded and was not counter-argued by any of the group members available in our WhatsApp dataset.

Similarly, many spiritual misinformation in our dataset are also geared or posted towards college students' lifestyles or events. For example, a viral *ritual* post was telling students to send heart emojis to at least 30 chats to enable them to get a perfect score on their final exams. Another viral post contained a largely fictional story about how an old man died pitifully without anyone by his side because he chose to chase for wealth and status during his younger years. The post then slipped a Qur'an verse and said that young people nowadays are too focused on chasing success, to the point that they forget about family and God. It should be noted that almost every viral religious content in our dataset included testimonies from a set number of people whose names nor personal identifications were written-making it unverifiable and fictional.

We found that many of the misinformation content in our dataset used group chat members' emotions, either guilt, fear, or hope, to make themselves viral. This can be a potential look at Indonesian university students and their WhatsApp group members' emotional stability, as they have been found to be correlated to the increased belief in misinformation [17]. The share of misinformation among Indonesian university students is, thus, particularly noteworthy since the possibility of the root problem of fake news sharing being emotional and personal is high. The emotional nature towards misinformation of Indonesian university students, coupled with the inability for fact-checking sites to interpolate into WhatsApp groups, raises concerns on whether having 'HOAX' information online on Google Search really is enough to combat misinformation.

6.4 Prevalence of entertainment content

As Arora et al. [3] suggest in their book "Bottom of the data pyramid: Big data and the global south", 'people go online to romance, game, be entertained, consume media, view pornography, and share their personal thoughts and feelings'. We find this to be certainly the case on WhatsApp too. Our analysis revealed that entertainment is a predominant category

of content consumed. We found a diverse range of entertainment materials, encompassing humor clips, non-political commentary, educational pieces on well-being, information about government programs, sports, beauty, art, gossip, and more. Interestingly, such messages along with religious content, political satire, and “good morning" messages constituted over a third of the virally disseminated materials [22] in both contexts.

Interestingly, contrary to findings from other studies involving similar demographics [25], entertainment was not the most prevalent type of viral content. Instead, misinformation (26% in India, 30.5% in Indonesia) and news (23.2% in India, 29.8% in Indonesia) dominated the content shared in these groups, which is surprising given the demographic profiles—young men in India and students in Indonesia. One might have expected a greater prevalence of entertainment content, reflecting their age and interests.

This suggests that WhatsApp might play a different role compared to other social media platforms for these demographics, serving less as a space for leisure and more as a medium for information dissemination, including misinformation. It raises questions about the influence of WhatsApp’s private and closed nature on content dynamics, where the spread of misinformation and news may be amplified in a way that differs from public platforms. Larger, representative quantitative studies are needed to explore these dynamics further and to understand why entertainment takes a backseat to more serious content in these settings.

6.5 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 in India and only 1 in 53 misinformation in Indonesia were counteracted or rebutted. This also extends beyond just the misinformation content, to other types such as hate speech, and propaganda. We found surprisingly in both contexts that there was 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 in India (17 out of 53 in Indonesia) 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.

These results align with the study by Seelam et al. [28], which highlights significant challenges faced by fact-checking organizations in India when it comes to reaching rural audiences with their content. Apart from top-down fact checking, we also looked for instances of within group, community driven fact-checking [16]. These are instances where a member of the group corrects misinformation posts. We found that there was only one instance of such correction out of the 158 instances of misinformation posted. Previous research in this space [20, 34] shows that such corrections do not happen often due to social pressures and power dynamics attached with correcting people in front other group members, which we think could also be at play in the groups we study.

The current model for fact-checking on WhatsApp relies on tiplines, which operate on an opt-in, reactive basis [15, 36]. 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.

7 Discussion

The paper presents a mixed-method analysis of viral content spreading in 164 Indian and 640 Indonesian private WhatsApp groups. We collected the WhatsApp groups data directly from villages in rural India and university students among Indonesian universities nationwide. This is the first study to look at viral content in private WhatsApp groups at such a scale in the Global South, giving us a look at everyday consumption both in terms of the types of groups users consume information from as well as the viral content that spreads in these groups. Though the samples from the two contexts are completely different, data reveals a surprising amount of similarity in terms of the types of groups people are a part of, and the types of viral content they receive.

7.1 Content Prone to *Virality* and Misinformation

In both India and Indonesia, we found that both religious and political content were more likely to be viral, filled with misinformation, and have no alternate arguments. In India, anti-Muslim content were very prominent. 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 minorities. 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," [1] 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 Indonesia, viral religious content were indeed *tamer* compared to those in India. Most of them did not convey any hate speech towards other religions and were mainly posts that may possibly help group members to *spiritually grow*. However, these same content included many unverifiable consequences and also *'divine punishments'* to those who did not forward, often within a specific deadline and to a specific number of people and followed the doctrines that were written in these religious posts. Instead of instilling hatred towards other groups of people, like in India, viral Indonesian religious content were created to instill fear and emotions into people who consume the content. This fact itself is troubling since emotions increase the belief in fake news [18]. The concerns following this emotional approach to religious content is that it will lead to sharing due to fear and online indirect coercion instead of personal desire. Research has also found that people would feel obliged to share misinformation if it means helping others avoid the misfortune promised by the viral content [35]. It is possible, then, that Indonesian religious content that inject fear to their readers will only worsen the spread of misinformation, as readers fear the possibilities of experiencing and causing misfortune. This study once again alerts the need of new intervention and educational methods to help WhatsApp users in the Global South to be able to prioritize recognizing fake information instead of feeding the fear and emotions originating from forwarded religious content.

Instead of having more local political content, our findings showed that Indonesians tend to share more of Chinese politics. In contrast to the negativity surrounding Indonesian politics and the Jokowi administration, viral Chinese political content were all positive in nature. China was portrayed as the most powerful nation globally, often being

compared to other countries such as the United States, Japan, and even Indonesia. Once again, the lack of alternate views on Chinese politics can instill narrow and radical judgments among Indonesians, which is an urgent matter in itself. If viral content such as in our findings continue on, it is possible that there will be a decrease in nationalism and patriotism among Indonesians, replaced instead by the desire to move abroad to countries promoted to have better governmental administrations, even when in reality may not always be the case.

7.2 The Struggle to Combat Misinformation

Our research underscores a disconcerting reality: fact-checking appears to be an exercise in futility within these WhatsApp groups. In India, not a single instance of fact-checked content surfaced in our analysis. In Indonesia, misinformation was only debunked once, 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 [29]. 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 current fact-checking efforts.

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.

Overall, the study advances the field of human-computer interaction by highlighting how technology which is currently understudied in the CHI community intersects with social, cultural, and political factors. It invites the CHI community to explore new avenues for research and practice that can lead to more effective and equitable technology design, ultimately contributing to the betterment of global digital societies.

8 Limitations and Future Work

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 samples in India and Indonesia. 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, points towards 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 specific contexts — rural India or Indonesian students, — 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. The findings, particularly the observed prevalence of politically charged content and misinformation, contribute to a deeper understanding of how WhatsApp is used in such a context. This is especially relevant given the limited research in this specific area. Some of the findings, such as the prevalence of hyper partisan messages, is interesting even in this limited context because it shows (though not fully) the range and reach of the BJP and its WhatsApp information ecosystem.

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 analysis of its kind, this study lays down the groundwork for more comprehensive investigations that could build on top of our findings.

Our study, despite its non-random sampling approach, provides significant insights into WhatsApp usage trends. The analysis focused on messages identified as "forwarded many times," which, due to their widespread circulation, likely mirror broader themes and attitudes extending beyond our immediate participant network. The methodology adopted for this research, while challenging, was the most feasible for accessing private WhatsApp groups. This challenge was compounded by the sensitive nature of the study, underscoring the crucial role of trust and pre-existing relationships in participant recruitment and data collection. Furthermore, the approach we employed holds promise for scalability. Although our current analysis was limited to a few dozen users, the methodology is adaptable and can be expanded for broader studies in the future, paving the way for a deeper and more comprehensive understanding of patterns and behaviors on WhatsApp.

References

[1] Edward Anderson and Christophe Jaffrelot. 2020. Hindu nationalism and the 'saffronisation of the public sphere': An interview with Christophe Jaffrelot. In *Neo-Hindutva*. Routledge, 116–130.

[2] Afroza Anwary. 2020. Interethnic conflict and genocide in Myanmar. *Homicide studies* 24, 1 (2020), 85–102.

[3] Payal Arora. 2016. Bottom of the data pyramid: Big data and the global south. *International Journal of Communication* 10 (2016), 19.

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

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

[6] Shakuntala Banaji, Ramnath Bhat, Anushi Agarwal, Nihal Passanha, and Mukti Sadhana Pravin. 2019. WhatsApp vigilantes: An exploration of citizen reception and circulation of WhatsApp misinformation linked to mob violence in India. (2019).

[7] Emma Baulch, Ariadna Matamoros-Fernández, and Amelia Johns. 2020. Introduction: Ten years of WhatsApp: The role of chat apps in the formation and mobilization of online publics. *First Monday* 25, 1 (2020).

[8] Santanu Chakrabarti, Lucile Stengel, and Sapna Solanki. 2018. Duty, identity, credibility: Fake news and the ordinary citizen in India. *BBC World Service Audiences Research* (2018).

[9] Simon Chauchard and Kiran Garimella. 2022. What circulates on partisan WhatsApp in India? Insights from an unusual dataset. *Journal of Quantitative Description: Digital Media* 2 (2022).

[10] Saloni Dash, Rynaa Grover, Gazal Shekhawat, Sukhnidh Kaur, Dibyendu Mishra, and Joyojeet Pal. 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.

[11] Kiran Garimella and Dean Eckles. 2020. Images and misinformation in political groups: Evidence from WhatsApp in India. *Harvard Kennedy School Misinformation Review* (2020).

[12] Natalie-Anne Hall, Brendan Lawson, Cristian Vaccari, and Andrew Chadwick. 2023. Beyond quick fixes: How users make sense of misinformation warnings on personal messaging. (2023).

[13] Md Mahfuzul Haque, Mohammad Yousuf, Ahmed Shatil Alam, Pratyasha Saha, Syed Ishtiaque Ahmed, and Naeemul Hassan. 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 (2020), 1–32.

[14] Prerna Juneja and Tanushree Mitra. 2022. Human and technological infrastructures of fact-checking. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–36.

[15] Ashkan Kazemi, Kiran Garimella, Gautam Kishore Shahi, Devin Gaffney, and Scott A Hale. 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 (2022).

[16] Pranav Malhotra. 2023. Misinformation in WhatsApp Family Groups: Generational Perceptions and Correction Considerations in a Meso-News Space. *Digital Journalism* (2023), 1–19.

[17] Cameron Martel, Gordon Pennycook, and David G Rand. 2020. Reliance on emotion promotes belief in fake news. *Cognitive research: principles and implications* 5 (2020), 1–20.

[18] Cameron Martel, Gordon Pennycook, and David G Rand. 2020. Reliance on emotion promotes belief in fake news. *Cognitive research: principles and implications* 5 (2020), 1–20.

[19] Taberez Ahmed Neyazi, Aaron Yi Kai Ng, Ozan Kuru, and Burhanuddin Muhtadi. 2022. Who gets exposed to political misinformation in a hybrid media environment? The case of the 2019 Indonesian election. *Social media+ society* 8, 3 (2022), 20563051221122792.

[20] Sheryl Wei Ting Ng and Taberez Ahmed Neyazi. 2022. Self-and Social Corrections on Instant Messaging Platforms. *International Journal of Communication* 17 (2022), 21.

[21] Billy Perrigo. 2019. How WhatsApp is fueling fake news ahead of India's elections.

[22] Newley Purnell. 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].

[23] Natasha Quek. 2019. Bloodbath in Christchurch: The rise of far-right terrorism. *RSIS Commentaries* 19 (2019), 1–3.

[24] Syarifah Ema Rahmaniah, Septiaji Eko Nugroho, Nikodemus Niko, et al. 2022. The Disinfodemic Mitigation Strategy of Mafindo in Indonesia. *International Journal of Social Science* 1, 6 (2022), 879–888.

[25] Nimmi Rangaswamy and Payal Arora. 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 (2016), 611–626.

[26] Mohan Rao. 2011. Love Jihad and demographic fears. *Indian Journal of Gender Studies* 18, 3 (2011), 425–430.

[27] Punyajoy Saha, Binny Mathew, Kiran Garimella, and Animesh Mukherjee. 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.

[28] Ananya Seelam, Arnab Paul Choudhury, Connie Liu, Miyuki Goay, Kalika Bali, and Aditya Vashistha. 2024. "Fact-checks are for the Top 0.1%": Examining Reach, Awareness, and Relevance of Fact-Checking in Rural India. *Proceedings of the ACM on Human-Computer Interaction* (2024).

[29] Ananya Seelam, Arnab Paul Choudhury, Connie Liu, Miyuki Goay, Kalika Bali, and Aditya Vashistha. 2024. " Fact-checks are for the Top 0.1%": Examining Reach, Awareness, and Relevance of Fact-Checking in Rural India. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (2024), 1–34.

[30] Farhana Shahid, Shrirang Mare, and Aditya Vashistha. 2022. Examining Source Effects on Perceptions of Fake News in Rural India. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (April 2022), 89:1–89:29. https://doi.org/10.1145/3512936

[31] Jenny Ratna Suminar and Purwanti Hadisiwi. 2021. Becoming a hoax buster in WhatsApp groups as an effort to limit the dissemination of misleading health information. *Jurnal Studi Komunikasi* 5, 1 (2021), 58–73.

[32] Muhammad Edy Susilo, Subhan Afifi, and Senja Yustitia. 2020. Hoax as a reflection on the low digital literacy in Indonesia. In *Proceedings of the Second International Conference on Social, Economy, Education, and Humanity (ICoSEEH 2019)-Sustainable Development in Developing Country for Facing Industrial Revolution*, Vol. 4. 165–174.

[33] Muhammad Taufiq Syam, Zulkifli Makmur, and Askar Nur. 2020. Social Distance Into Factual Information Distance about COVID-19 in Indonesia Whatsapp Groups. *Jurnal Ilmu Komunikasi* 18, 3 (2020), 269–279.

[34] Rama Adithya Varanasi, Joyojeet Pal, and Aditya Vashistha. 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.

[35] Rama Adithya Varanasi, Joyojeet Pal, and Aditya Vashistha. 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.

Table 4. 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
Friends	Refers to a friend group
News	Primarily focused on sharing and discussing current events or updates from official, unofficial, or community sources.
Others	Can not be categorized into any of the six groups listed above e.g. fun group, hobbies, etc.

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

[37] Keyza Pratama Widiatmika. 2019. *The lost family: the stories of Indonesian family conflicts and political messages in WhatsApp groups.* Master's thesis. University of Twente.

[38] Fabiana Zollo, Petra Kralj Novak, Michela Del Vicario, Alessandro Bessi, Igor Mozetič, Antonio Scala, Guido Caldarelli, and Walter Quattrociocchi. 2015. Emotional dynamics in the age of misinformation. *PloS one* 10, 9 (2015), e0138740.

Appendix

Definitions of group categories (India)

We manually classified all the 164 groups from India into 10 categories. Table 4 shows the definitions of the group categories.

Definitions and examples of content categories (India)

(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.

A Definitions and Examples of Content Categories (Indonesia)

(1) Local information: Viral posts containing information with specific regional boundaries mentioned that are the size of a city or smaller.
(2) Advertising: Viral posts promoting events or products that are accompanied by further contact methods for readers to RSVP or text sellers.
(3) Misinformation: Incorrect or misleading information. Both posts that are verified by the Indonesian Communication Agency as 'HOAX' and those that are unverifiable due to the lack of related information outside of the posts are included here.
(4) Inspirational Message: Viral posts that promote non-health behavioral changes in order to achieve specific non-health goals, usually in the shape of advice.
(5) National Information: Viral posts containing information with specific regional boundaries mentioned that are the size of a province or bigger, but still within the scope of Indonesia
(6) Entertainment/Humor/Sarcasm: Posts that are shared to make people happy or laugh. This includes beauty of nature, cultural posts, people doing funny acts, cartoons, etc.
(7) International Information: Posts containing information from outside Indonesia
(8) Religious: Posts that explicitly mentioned the doctrine or verses of at least one of the 6 official Indonesian religions: Islam, Protestantism, Catholicism, Buddhism, Hinduism, Confucianism
(9) Health Advice: Posts that include health content
(10) Chinese Content: Posts that include content about anything Chinese: Chinese products, Chinese politics, Chinese nature, Chinese culture, Chinese technology, posts that are written or spoken in Chinese.
(11) Religious Iconography: Posts that explicitly mentioned the characters/people of at least one of the 6 official Indonesian religions: Islam, Protestantism, Catholicism, Buddhism, Hinduism, Confucianism
(12) Political: Anything explicitly concerning the government no matter the scope: Indonesia, international, the world.
(13) Propaganda: Anything political that aims to change behavior or opinion by promoting or depromoting a specific government/party/person

(14) Hateful Content: Posts that show hostility or discrimination directly towards a specific country/group/person
(15) Partisan: Political posts that show obvious support or dislike towards a specific political party/character.
(16) Good Morning Message: Messages wishing people good morning