Tattle

Crowdsourcing Aid: A Case Study of Information Chaos

During India's Second Covid-19 Wave

Crowdsourcing Aid: A Case Study of Information Chaos during India's Second Covid-19 Wave

25 July 2021

Contributors

Tarunima Prabhakar
Denny George
Swair Shah

Acknowledgements

We'd like to thank Connie Moon Sehat for reading an early draft of this report and pushing us to slow down and reframe the analysis; to CoViD Action Initiative: cov.social for responding to a cold email on a short notice and sharing their crowdsourced database with us; to Scott Rogowski for his persistent work on the WhatsApp scraper; and Tanima Saha for the editorial scrubbing of this report. We'd ultimately like to thank all the open source contributors who have contributed to Tattle's code base. The analysis presented here builds on all their work.

Despite an incredible support crew, this report might contain errors. Such errors are the responsibility of the authors alone.

Layout and Design: Saakshita Prabhakar

Any Questions?
Details or questions about this report please email at:
tarunima@tattle.co.in or denny@tattle.co.in

CONTENTS

EXECUTIVE SUMMARY

From April to June 2021, India was ravaged by the second wave of the Covid-19 pandemic. With a steep increase in the number of infections, cities ran out of drugs, medical oxygen, hospitals and other necessary medical supplies. The second wave also ravaged rural India that had weaker public health infrastructure than urban India.

The second wave was a public health crisis, but also an information crisis. As cities ran out of medical aid, people turned to social media to request for resources outside their geographies and immediate networks. The circulation of *leads* for medical aid led to a concomitant increase in circulation of inaccurate leads. Fraudulent leads—of scamsters duping people of money in promise of medical supplies—were mixed in the pool of information leads on social media platforms. In addition, the status of medical leads changed rapidly—any available hospital beds or drugs were taken up within minutes of being available. Obsolete and fraudulent leads resulted in loss of critical time in medical care, money and ultimately lives. Moreover, phone numbers of individuals from marginalized identities, circulated for requesting or providing aid, were used for targeted harassment.

The crisis thus resulted in spontaneous volunteering—individuals and groups played the role of intermediaries, connecting those in need to *verified* leads for medical aid. WhatsApp, the most widely used social media platform in India, emerged as a natural choice to organize volunteering energies. WhatsApp could be used to source requests for help, coordinate with other volunteers and connect people to adequate aid. But how was WhatsApp used to surface actionable information and push back on inaccurate leads when the platform eludes centralized takedowns and moderation? How did individuals filter credible leads from a glut of information leads?

This report contends that the second wave of the pandemic in India showed a new facet of the Information Disorder. It was driven by a specific typology of information—of hyperlocal information leads shared during a crisis. While this category of information shares some features with the prototypical political and medical misinformation, it also merits unique attention.

We present preliminary analysis from 8 weeks of conversations in 21 Covid-19 relief WhatsApp groups that were operational during the second wave of the crisis, to shed light on the information chaos that ensued during that period. We use vector embedding based machine learning to aggregate images based on visual and semantic similarity, and language processing techniques to make sense of the multi-lingual content shared on these groups. Our preliminary analysis suggests that:

- Even on WhatsApp, people relied heavily on other social media platforms such as Twitter and Instagram to find verified leads. Sharing screenshots of posts from these platforms was the commonly used method to cross-post information. Twitter has fewer than 20 million users in India. People act as 'go-betweens' and connect WhatsApp users to information on Twitter and Instagram, giving content on these platforms greater reach.
- We compared the phone leads shared in the 21 WhatsApp groups with a national level database of verified leads maintained by a fact-checking group and with a crowdsourced database of 'scam' numbers. We found that less than 17% of the leads were common between the WhatsApp groups we were tracking and the databases. This indicates the scale of information that was circulated and challenge of verifying content during the second wave of the pandemic.
- The frequency of conversations declined in these groups over time. Some groups were repurposed to share information unrelated to Covid-19 such as chartered accountancy related webinars reflecting the use of WhatsApp in digital marketing.

- Volunteering groups asked for patient information in specific templates to make relief work more efficient. People shared doctor prescriptions, medical receipts and sensitive personal information (including the Biometric ID Aadhar) in these groups when requesting for help. Private information was circulated in groups of unknown persons who had come together for public oriented service. Public and private boundaries are more blurred in emergencies which demands greater attention to data deletion protocols by group admins.

Taking note of the citizen and crowdsourced verification that sprung about during the second wave of the pandemic, we posit that the spontaneous increase in hyperlocal information demanded distributed but coordinated verification. While we have seen coordinated fact-checking operations around elections, we propose the possibility of similar coordination during sudden events such as natural disasters and wars.

We note that the information chaos during the second wave in India eluded responses conceived around political and medical misinformation. 'Leads for medical aid' as a typology of information was hyperlocal and not created or propagated in coordination. This information did not have to be viral to be harmful. Leads about medical resources are concise units of information that don't rely on or trigger cultural, social or political beliefs. People had strong incentives to seek out accurate information. Despite deliberate reasoning, the truth status of such information was not easy to discern. How platforms could have best intervened to reduce the circulation of such content is also unclear.

A case study of relief work coordinated on chat apps during the second wave of Covid-19, this analysis highlights a facet of the information disorder that could emerge in any situation where the need for reliable actionable information is high but trusted and expected information channels fail. Accounting for such situations in emerging agendas for research and action could lead to more robust toolkits for dis/misinformation response.

THE NEED FOR RELIEF WORK

DURING INDIA'S SECOND WAVE

When the second Covid-19 wave hit India in April 2021, the demand for medical oxygen, medicines, hospital beds and plasma donors exceeded the capacity of public and private health service providers. The crisis first hit the metropolitan cities of New Delhi and Mumbai. But soon enough, unlike the first, the second wave also severely affected smaller cities and rural India.[1]

A year into the pandemic, it was clear that timely care was critical in Covid-19 treatment. But as cases rose exponentially, health facilities faced an unanticipated and incomparable burden. In the capital city, New Delhi, any available hospital bed had many takers. The status of availability of beds in hospitals changed rapidly and government-provided dashboards were not updated frequently enough. People who arrived at hospitals based on some information, were turned away because the vacant beds would be occupied by the time they reached. As the city started facing a shortage of medical oxygen, hospitals stopped the intake of patients to optimize for their reserve oxygen. Consequently, leads for hospital beds in the city or neighboring cities, or for medical oxygen became critical information. When the usual channels such as government helplines, neighborhood hospitals and pharmacies became unhelpful, patients or their family and friends began to reach out to their networks for help.

1 The Financial Express. (2021, June 6). 'Rural India ravaged by Covid-19 second wave'. The Financial Express. https://www.financialexpress.com/lifestyle/health/rural-india-ravaged-by-covid-19-second-wave/2266111/

Those who were on social media platforms, broadcast requests on platforms such as Twitter, Instagram and WhatsApp. On Twitter, requests were noticed and amplified by political representatives and celebrities. The heightened demand also created black markets for medical resources and proliferation of scam operations. Several advertisements for oxygen cylinders and concentrators turned out to be fraudulent. Since patients and their families were in isolation at home, or geographically distant from the medical service providers, they relied on providers to deliver medical supplies. By requesting for a payment ahead of delivery, scamsters duped patients by taking the money but not delivering the medical supplies. People lost money and critical time in Covid-19 care.

The proliferation of scam operations and the pace at which leads became irrelevant resulted in a need to verify leads. Motivated individuals, neighborhood communities, religious organizations and corporate social responsibility groups took the role of intermediaries, connecting patients and their families to 'verified leads' for hospitals, oxygen, medicines and plasma donors. While some volunteering groups were operational since the first wave of the pandemic in March 2020, the second wave led to spontaneous organizing for Covid-19 relief. Most of these groups focused on medical needs, but some also focused on providing food to patients, frontline works and to those who had lost their livelihood in lockdowns. An Indian news article captured the role of social media with the headline: 'How volunteers have made social media the national Covid "helpline" for beds, oxygen, plasma.'[2]

Twitter and Instagram became platforms to broadcast requests and leads on and WhatsApp emerged as a crucial platform to organize volunteering energy.[3] The group chat feature on

2 Basu, M. (2021, April 19). 'How volunteers have made social media the national Covid 'helpline' for beds, oxygen, plasma'. Accessed on 6 July 2021 from The Print. https://theprint.in/health/how-volunteers-have-made-social-media-the-national-covid-helpline-for-beds-oxygen-plasma/642010/

3 Bose, M. (2021, April 30). 'These Covid volunteers were helping save lives, but now they're really scared'. India Today. Accessed on 6 July 2021 from https://www.indiatoday.in/coronavirus-outbreak/story/corona-wave-covid-volunteers-sos-shortage-oxygen-remdesivir-hospital-bed-help-1796702-2021-04-30

WhatsApp allowed for creation of functional boundaries within which similarly intentioned individuals could coordinate efforts and triage information. The WhatsApp group delineated a space with informally specified protocols, enforced by loose social and technical hierarchies—even if the feature was not used, group chats allowed for exclusions of norm-breakers by group admins.

With over 400 million users, WhatsApp is the most widely used app and social media platform in India.[4] As per one estimate, over 90% of internet users in India aged 16 to 64 use chat apps.[5] As per Lokniti's 2019 survey on social media and political behavior, even though WhatsApp's user base is skewed towards younger demographics, it is the most widely used social media platform across age groups. In 2019, 19% of respondents in 46–55 age group claimed to be daily or weekly users of WhatsApp. This is a stark contrast to Twitter, where only 7% of respondents across all age groups reported using Twitter regularly.[6]

Even though WhatsApp allows groups of only 256 individuals, the high adoption makes it the a natural choice for spontaneous orga nizing. One feature of Covid-19 relief work is that it is regional. Patients need resources in the cities/towns that they are living in. Lockdowns and restrictions are implemented locally based on the severity of the outbreak in a district or state. Covid-19 facilities are largely managed by states. Finally, verifying leads and connecting patients to medical resources could require familiarity with regional languages.

4 Estimated by monthly active users.

5 Kemp, S. (2021, February 11). 'Digital in India: All the Statistics You Need in 2021 - DataReportal – Global Digital Insights'. DataReportal (Slide 65). Accessed on 6 July 2021 from *https://datareportal.com/reports/digital-2021-india*

6 Centre for the Study of Developing Societies, Lokniti, and Konrad Adenaur Stiftung. (2019). Social Media & Political Behaviour. 'Social Media & Political Behaviour – Study'. Accessed on 6 July 2021 from *https://www.lokniti.org/otherstudies/social-media-political-behaviour-study-208*

Volunteering groups might be seeded by a group of individuals who are geographically co-located and know each other in person, They converge on WhatsApp, a platform accessible to anyone with a mobln. To sidestep the group size limitation, a volunteering community may run multiple WhatsApp groups, with some admins common across the groups. While WhatsApp isn't designed as a platform for broadcasting information, volunteering groups providing assistance or in need of volunteers, have to broadcast information about their existence to be discoverable by broader public. For this purpose, group admins may advertise the links to join these groups on more open platforms such as Twitter or on a volunteering website.

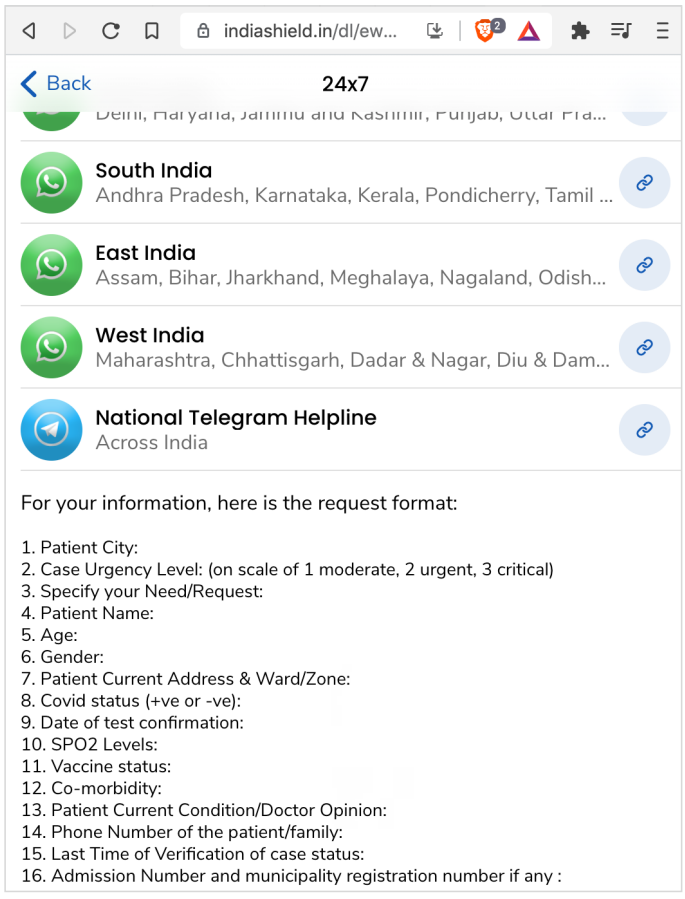


Figure 1: Request for Templatized Message on WhatsApp Groups Run by a Relief Group

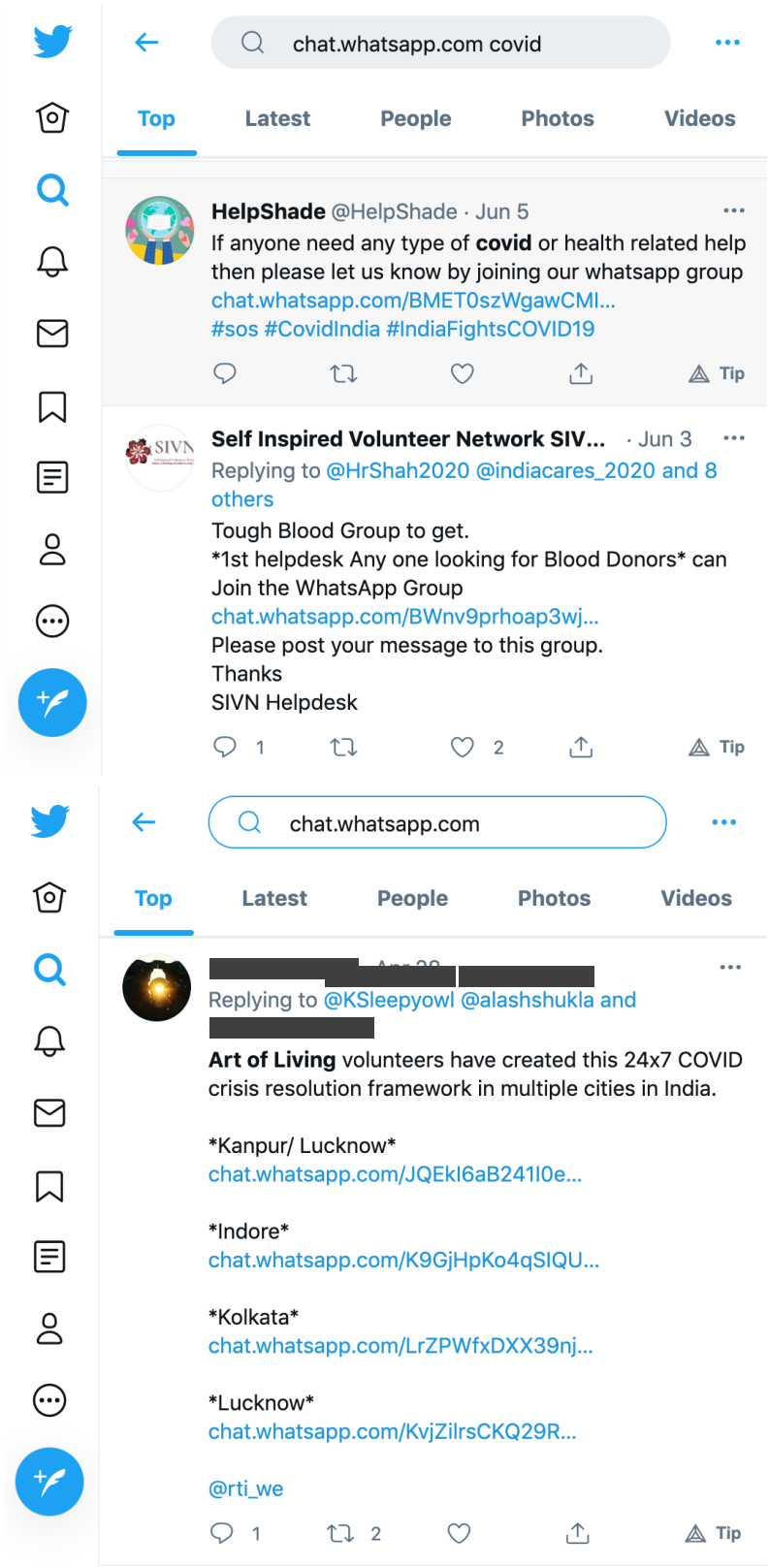


Figure 2: Links to WhatsApp Relief Groups Shared on Twitter

Figure 1 and 2 show several such WhatsApp group links shared by volunteering groups on Twitter or the group websites. Figure 1 also reflects processes such as templatized request formats that groups converged on to make relief work more efficient. These processes emerged since the nature of aid requested by different people were similar, such as when requesting for plasma it was important to know what blood group was being sought; the level of medical support needed by a patient depended on their SPO2 levels and co-morbidities. However, the format in which information was requested changed depending on the volunteering group and the medical need for example, when requesting oxygen support, people often shared their home address.

While the patient or the immediate family members of a patient might not be in these WhatsApp groups, people in their extended networks might post their requests in these groups. These requests were also shared on more open platforms such as Twitter, Facebook and Instagram. The qualitative difference between messaging apps such as WhatsApp and open platforms such as Twitter is that information amplification on Twitter is primarily created through retweets and can be controlled to some extent by deletion of posts by the original creator or by the platform. Messaging apps might allow a grace period (one hour for WhatsApp) during which the sender can delete the message for all the receivers and senders. But on the whole, WhatsApp doesn't allow for retrospective editing or deletion of messages. It only allows for additional context to be added to a message through the reply-to-message option. A message once sent stays in the phones of all those in the group who have received the message.

WHY STUDY COVID-19 RELIEF GROUPS

‘ What we really need to do to design is look at the extremes... Because if we understand what the extremes are, the middle will take care of itself. ’[7]

Accounting for a New Typology of (Mis)information

Rise in popularity of messaging apps for individual and group communication has gone hand in hand with an increase in inaccurate, harmful and low credibility information on these platforms.[8] In 2018, lynchings in India were connected to rumors circulating on WhatsApp. Understanding how chat apps are used for coordinated messaging and how harmful ‘viral’ content can be addressed on platforms without centralized moderation, has assumed urgency over the last few years. Prior research on messaging apps tended to focus on circulation of content related to contemporary political themes, especially around elections.[9][10][11]

7 Product vs digital design: the Extreme vs the Majority. Studio frankly. (2019, July 24). Accessed on 6 July 2021 from https://frankly.studio/product-vs-digital-design-the-importance-of-power-users/

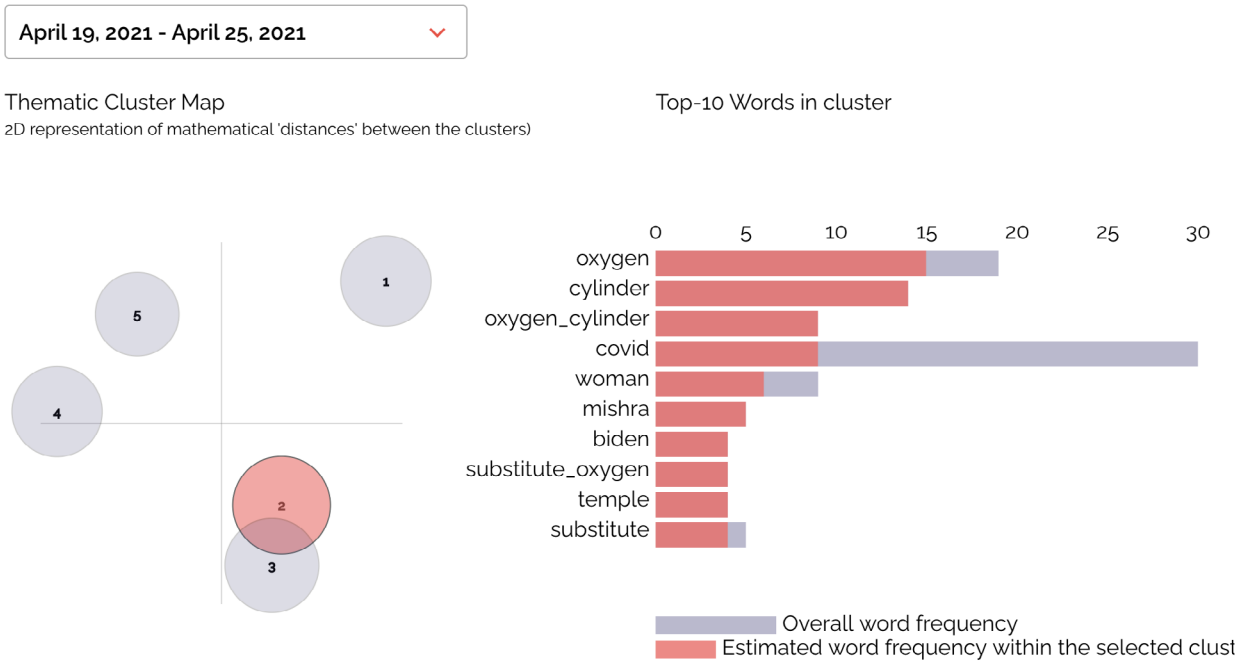
8 Also called misinformation/disinformation.

9 Banaji, S., Bhat, R., Agarwal, A., Passanha, N., & Pravin, M. S. (2019). ‘WhatsApp Vigilantes: An exploration of citizen reception and construction of WhatsApp messages’ triggering mob violence in India’. London School of Economics. Accessed on 6 July 2021 from https://www.lse.ac.uk/media-and-communications/research/research-projects/whatsapp-vigilantes

10 Garimella, K., & Eckles, D. (2021, January 26). ‘Images and misinformation in political groups: Evidence from WhatsApp in India: HKS Misinformation Review’. Misinformation Review. Accessed on 6 July 2021 from https://misinforeview.hks.harvard.edu/article/images-and-misinformation-in-political-groups-evidence-from-whatsapp-in-india/

11 Bengani, P. (2019, October 16). ‘India had its first “WhatsApp election.” We have a million messages from it’.. Columbia Journalism Review. Accessed on 6 July 2021 from https://www.cjr.org/tow_center/india-whatsapp-analysis-election-security.php

The pandemic has led to renewed interest in countering health related misinformation. In health emergencies, uncertainty is inevitable.[12] The anxiety accompanying uncertainty and information voids[13] make health emergencies especially fertile events for mis/disinformation. Figure 3 shows a snapshot of the kinds of misinformation shared during the peak of the second wave in India.[14]



Viral list of oxygen cylinder suppliers is partially true
https://newsmeter.in/fact-check/fact-check/viral-list-of-oxygen-cylinder-suppliers-is-partially-true-677083

Fact Check: Photo of woman sitting outside hospital with oxygen cylinder not related to pandemic
https://newsmeter.in/fact-check/fact-check/fact-check-photo-of-woman-sitting-outside-hospital-with-oxygen-cylinder-not-related-to-covid-19-pandemic-677190

Vadodara's BAPS temple converted into COVID center, not Mumbai's Swaminarayan temple
https://newsmeter.in/fact-check/fact-check/vadodaras-baps-temple-converted-into-covid-center-not-mumbais-swaminarayan-temple-677226

Delhi school converted into COVID ward, not Lucknow
https://newsmeter.in/fact-check/fact-check/delhi-school-converted-into-covid-ward-not-lucknow-677404

Figure 3: Tattle Fact Checking Sites Dashboard Summarizing Fact Checks from Third Week of April 2021.[14]

12 Hyland-Wood, B., Gardner, J., Leask, J., & Ecker, U. K. H. (2021, January 27). 'Toward effective government communication strategies in the era of COVID-19'. Nature. Accessed on 6 July 2021 from https://www.nature.com/articles/s41599-020-00701-w

13 Shane, T. (2020, August 10). People are using Facebook and Instagram as search engines. During a pandemic, that's dangerous. Neiman Lab. Accessed on 6 July 2021 from https://www.niemanlab.org/2020/08/people-are-using-facebook-and-instagram-as-search-engines-during-a-pandemic-thats-dangerous/

14 Image created from Tattle's Fact Checking Sites Dashboard: https://services.tattle.co.in/khoj/dashboard

The second wave in India became an emergency within a public health emergency. The death toll rose rapidly and reports of deaths of fully vaccinated individuals, health practitioners and younger demographics spread panic about the double mutant, later named the Delta variant.[15] [16] As more people tested positive, the course of treatment became less clear. Protocols for prevention and recovery developed in the preceding months seemed insufficient. Medicines were in short supply. Eventually, in cities like New Delhi, medical oxygen too ran short. Even as patients were being prescribed convalescent plasma therapy, doctors on social media dissuaded this course of treatment, except in specific cases.[17] Misinformation around Covid-19 cures has existed throughout the pandemic; but during the second wave, such misinformation latched onto prominent anxieties around the second wave. In one viral video, a medical practitioner claimed that nebulizers could be used in absence of medical oxygen.[18] Doctors on televised news and social media rushed to debunk the claim.

The second wave, however, also introduced another unprecedented form of information crisis, by creating demand for a specific kind of information. In the crisis, a lot of people were simultaneously looking for information, such as leads for medical oxygen suppliers, hospital beds and plasma donors, that was immediately actionable. This is in contrast to the usual typologies of misinformation that aim to shape opinions or affect a long-term decision. Another contrast is that this information sought

15 Ghosh, B. (2021, May 24). 'Coronavirus: 43 doctors die in Bengal in second wave'. The Hindu. Accessed on 6 July 2021 from https://www.thehindu.com/news/national/other-states/coronavirus-43-doctors-die-in-bengal-in-second-wave/article34635081.ece

16 Qureshi, S. (2021, June 6). 'Doctors explain why young people are more affected in second wave of Covid-19'. India Today. Accessed on 6 July 2021 from https://www.indiatoday.in/coronavirus-outbreak/story/doctors-explain-young-people-more-affected-second-wave-1811619-2021-06-06

17 Correspondent, S. (2021, May 17). 'ICMR drops plasma therapy from COVID-19 treatment guidelines'. The Hindu. Accessed on 6 July 2021 from https://www.thehindu.com/sci-tech/health/icmr-drops-plasma-therapy-from-covid-19-treatment-guidelines/article34582184.ece

18 BBC. (2021, May 1). 'India Covid-19: Fact-checking misleading claims on oxygen treatments'. BBC News. Accessed on 6 July 2021 from https://www.bbc.com/news/world-asia-india-56925650

was local or hyperlocal—people needed leads in their vicinity. There was no reliable centralized source for such immediately actionable information.

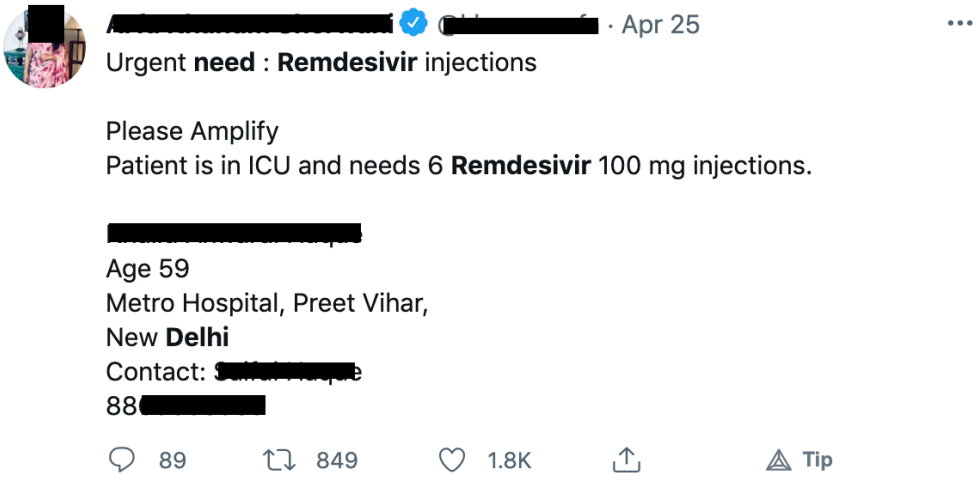


Figure 4: A Tweet Requesting for Remdesivir Injections for a Patient

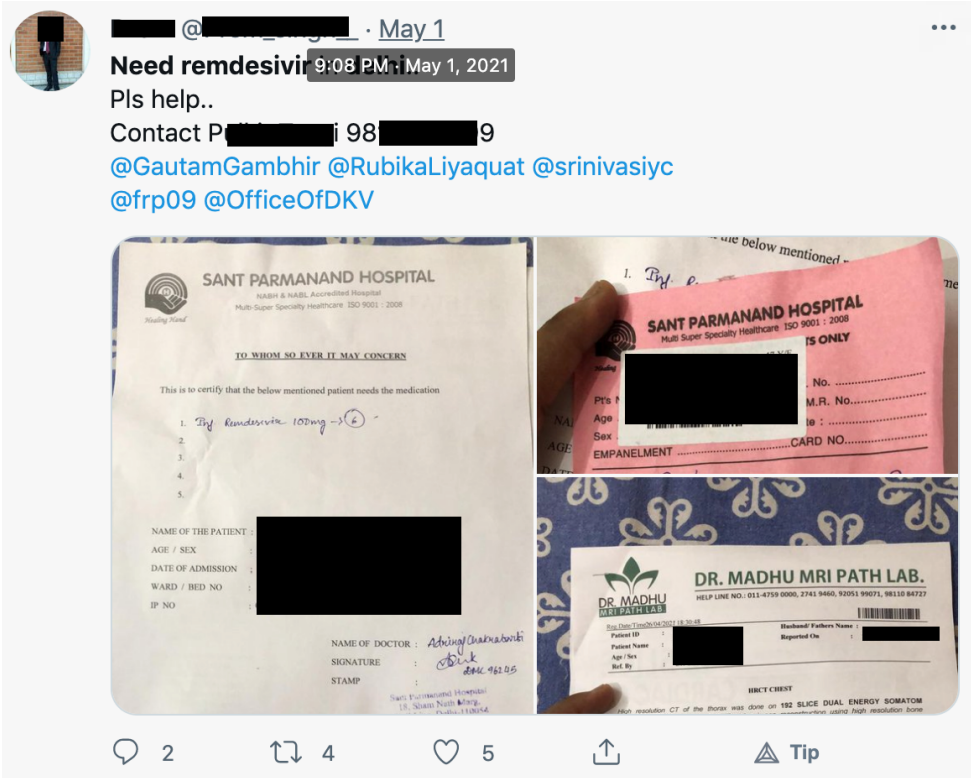


Figure 5: A Tweet Requesting for Remdesivir Injections. Doctor's prescription for Remdesivir as well as other patient medical records are shared in the tweet.

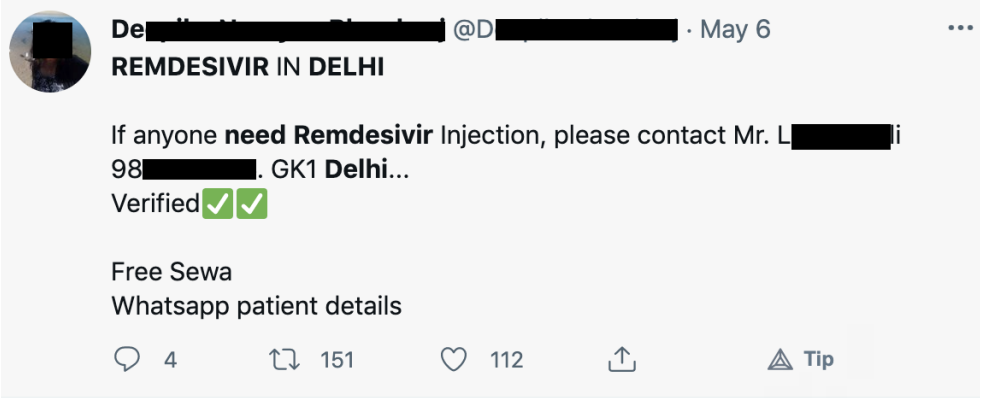


Figure 6: A Tweet advertising a WhatsApp number for Remdesivir Injections

The deficit in usable information about medical aid was filled with leads that were old or fraudulent. Information leads might have been inaccurate on the date of creation of message (with the intent to scam), or might be inaccurate on the date of receipt (information is obsolete and not actionable).[19] In both scenarios, regardless of whether a message was viral or not, inaccurate messages could cause imminent harm to any recipient who acted on it. These harms can be separated intp three overlapping categories:

Loss of time

For every inaccurate or obsolete lead, patients lost more time in accessing critical care. Victims of fraudulent operations lost hours, and sometimes days, waiting for oxygen supply, which could be the difference between life and death.[20]

Loss of money

Scarcity of resources resulted in price hike of medical resources. The price of medical oxygen increased by 700%

19 Digital Methods Initiative. (2020). 'Disinformation in anti-Eu Facebook groups'. Winter School 2020 Disinformation EU. Accessed on 6 July 2021 from: https://wiki.digitalmethods.net/Dmi/WinterSchool2020DisinformationEU#A_3._Research_Questions

20 Peng, Y., Xu, B., Sun, B., Han, G., & Zhou, Y.-H. (2020, May 11). 'Importance of timely management of patients in reducing fatality rate of coronavirus disease 2019'. Journal of Infection and Public Health. https://www.sciencedirect.com/science/article/pii/S1876034120304652

during the two months of the second Covid wave in India.[21] In black markets, the drug Remdesivir could cost five times its actual price.[22] The desperation of patients and surge of unregulated black markets created room for scamsters. People who paid for medical resources on the promise of future delivery, lost money equivalent to months' worth of earnings or their entire life savings.[23]

Figure 7: A Tweet expressing frustration on being scammed

21 Shanker, K.S. (2021, June 13). 'Drop in rates of oxygen concentrators, cylinder refill'. The Hindu. Accessed on 6 July 2021 from https://www.thehindu.com/news/cities/Hyderabad/drop-in-rates-of-oxygen-concentrators-cylinder-refill/article34806996.ece

22 Gill, P. (2021, May 10). 'The utter chaos and confusion over Remdesivir in India is making the COVID-19 second wave worse'. Business Insider. Accessed on 6 July 2021 from https://www.businessinsider.in/science/health/news/remdesivir-injection-price-jumps-five-times-as-covid-19-cases-rise/articleshow/82083414.cms

23 Jha, A. (2021, January 16). 'Number Theory: How much does an average Indian earn?' Hindustan Times. https://www.hindustantimes.com/india-news/number-theory-how-much-does-an-average-indian-earn-101610760612856.html.Accessed on 6 July 2021 from https://www.hindustantimes.com/india-news/number-theory-how-much-does-an-average-indian-earn-101610760612856.html

Loss of life

Since the beginning of the pandemic, the WHO had warned that misinformation could be life threatening.[24] As per some estimates, at least 800 people may have died globally due to Covid-19 related misinformation and thousands more may have been hospitalized.[25] In addition to medical misinformation,[26] the second wave also led to proliferation of inaccurate and obsolete leads for medical resources, delaying access to timely care. This ultimately contributed to higher mortality from the disease.

The definition of the terms disinformation and misinformation are still emerging. The terms, while broad and flexible in definition, have primarily been used to describe content aimed at changing people's perceptions.[27] [28] The typology of information shared in Covid-19 relief groups and considered in this analysis falls outside this prototypical definition. The messages circulated in relief groups are short and contain precise information such as a phone number or address. Online scams, typically absorbed under cybersecurity, have not been systematically analyzed in disinformation research, possibly due to their relatively small volume in day-to-day social media discourse. In Covid-19 relief work, however, these messages constituted the dominant disinformation typology that needed to be managed. In the

24 World Health Organization. (2020, August 25). 'Immunizing the public against misinformation'. World Health Organization. Accessed on 6 July 2021 from https://www.who.int/news-room/feature-stories/detail/immunizing-the-public-against-misinformation

25 Islam, M. S, et al. (2020, October 7). 'COVID-19–Related Infodemic and Its Impact on Public Health: A Global Social Media Analysis'. The American Journal of Tropical Medicine and Hygiene. Accessed on 6 July 2021 from https://www.ajtmh.org/view/journals/tpmd/103/4/article-p1621.xml

26 Deokar, M. (2021, April 25). 'No, Nebulizer Should Not Be Used As a Substitute for an Oxygen Cylinder'. FactCrescendo. Accessed on 6 July 2021 from https://english.factcrescendo.com/2021/04/25/no-nebulizers-cannot-be-used-as-a-substitute-for-an-oxygen-cylinder/

27 Southwell, B.G., Thorson, E. A., Sheble, L., (2017, January 27). 'The Persistence and Peril of Misinformation'. American Scientist. Vol. 105, No. 6. Accessed on 6 July 2021 from https://www.americanscientist.org/article/the-persistence-and-peril-of-misinformation

28 Wardle, C. (2018, December 10). Fake news. It's complicated. First Draft Footnotes. Accessed on 6 July 2021 from https://medium.com/1st-draft/fake-news-its-complicated-d0f773766c79

past, fact checkers have verified phone numbers circulated on social media and provided them with a 'Fact-Check' status label such as False or Misleading, similar to more news-worthy social media content.[29] [30] Such verification work became urgent during the second wave.[31] These messages demanded similar kinds of intervention from platforms, governments and citizens for suppression of low-quality information, as more news-like dis/misinformation.

The category of messages describing local resources for medical aid serves as microcosm within the information ecosystem, where several aspects of the misinformation challenge are heightened. People needed actionable information during the crisis. Anxiety and uncertainty further drove a demand for such information- in fear of not getting medical care when needed, people preemptively stocked on drugs, oxygen concentrators and cylinders. Inaccurate information was indistinguishable from relevant and actionable information. People had to find credible information despite accurate and actionable information becoming obsolete in minutes.

Despite these overlaps with previously considered typologies of misinformation such as political content or medical misinformation, the type of information circulated during the relief work in the second wave was unique and merits attention. The rise in user-generated content has pushed the scope of misinformation beyond fake 'news' that can be typified by a source, website or renowned media channel. Some attributes of traditional 'fake news' such as linguistic features and network of

29 Goldhamer, M. (2020, April 17). 'Posts list phone sex numbers as helplines to track US stimulus payments'. AFP. Accessed on 6 July 2021 from: https://factcheck.afp.com/posts-list-phone-sex-numbers-helplines-track-us-stimulus-payments

30 Sutaria, S. (2019, December 6). 'Hyderabad Vet Rape: Inactive Numbers Shared As Emergency Helpline For Women'. BOOM. Accessed on 6 July 2021 from: https://www.boomlive.in/fake-news/hyderabad-vet-rape-inactive-numbers-shared-as-emergency-helpline-for-women-6221

31 FactChecker.in (2021, June 22). 'FactChecker Called Up All COVID-19 Helplines'. FactChecker.in. Accessed on 6 July 2021 from: https://www.factchecker.in/fact-check/factchecker-verified-covid-19-helplines-remdesivir-hospital-beds-oxygen-743175

dissemination have been used to rank quality of user generated content such as tweets.[32] But such markers don't easily extend to information leads shared in Covid-19 relief work. The only purpose of these leads is to provide contact details of suppliers, who are (reasonably) assumed to be the source of the information. Message attributes such as grammatical mistakes don't help identify whether the information is actionable or not. For one, many of these information messages may be too short to capture any meaningful grammatical mistakes. Second, grammatical mistakes such as spelling errors are not unsurprising if suppliers don't speak or type well in English. Network of dissemination is another marker used by researchers as well as social media consumers to assess the credibility of messages online.[33] The primary purpose of sharing requests for help on social media and messaging apps is to look beyond one's immediate networks—one is proactively seeking information from strangers. Future research can clarify how this marker of credibility evolved in deliberate interaction with strangers during relief work.

A message on WhatsApp comes with no information about when it was created or who created it. Both these data points, the source and the time of creation, serve as important context when trying to gauge the relevance of specific information for relief work. An update about the status of availability of hospitals in a city from two days ago is likely to be obsolete and unactionable. In absence of the source or time-stamp, what markers and processes did people seek or fall back on to gauge credibility? How were rapid updates about specific messages communicated through messaging networks? How quickly, if at all, did inaccurate content fade from these networks? Understanding the answers to these question in this specific context could shine a light on the broader phenomenon of misinformation on chat apps.

32 de Beer D., Matthee M. (2021) Approaches to Identify Fake News: A Systematic Literature Review. In: Antipova T. (eds) Integrated Science in Digital Age 2020. ICIS 2020. Lecture Notes in Networks and Systems, vol 136. Springer, Cham. https://doi.org/10.1007/978-3-030-49264-9_2

33 Banaji et al. (2019). 'WhatsApp Vigilantes: An exploration of citizen reception and construction of WhatsApp messages'. Pg 14.

This analysis, carried out in the month after data collection, relies primarily on automated techniques. Far from providing conclusive answers to the questions raised above, it aims to provide an overview of the information chaos that ensued during the second wave of the pandemic in India; and propose another direction in which research on disinformation, especially pertaining to chat apps, can be broadened. This broader research agenda could not only build resiliency for future global crises such as natural disasters, wars and cyber-attacks. but could also enhance our understanding of the broader phenomenon of mis/dis-information.

Understanding the Public/Private Boundary on Chat Apps

In focusing on public-oriented action coordinated on closed messaging platforms, we also hope to shed more light on emerging online publics. As highlighted in the previous section, these WhatsApp groups became a means for community organizing and consequently public engagement during Covid-19 relief work. Messaging apps exemplify the blurring of distinctions between public and private spaces online. These apps are designed to be private yet social communication platforms. The simultaneous existence of private and viral content on the same platform not only presents challenging questions about the form of communication, but also about how to understand what is happening in these spaces. One model of research is to identify and join groups that can be assumed to be reasonably public, and analyze the conversations in these spaces. We worked with the assumption that if an invitation to join a group is broadcast on Twitter[34], it is reasonable to assume that the group admins aimed for it to be discovered by (some) strangers. Specifically, messaging groups created exclusively to coordinate information and aid for those

34 By sharing group invitation links on Twitter.

affected during the pandemic serve a public purpose. Yet, the nature of relief work during the second wave meant that private information about individuals such as their phone numbers and medical histories were shared on these WhatsApp groups. Such information was also shared on more open platforms such as Twitter, Facebook and Instagram. People were broadcasting such information to reach outside one's immediate networks. But content on these platforms could be taken down by the user or the platform. Many people deleted tweets once their request for aid was met. On WhatsApp, on the other hand, a user lost control over the message once it has been sent.

While a lot of good came out of coordination of information on social media and WhatsApp, tangible harms could also be linked to identification of individuals through and on these platforms. Several women reported receiving sexually explicit messages and calls after sharing their number for seeking Covid-19 relief resources for family and friends, on social media.[35] Volunteers who shared their contact number were also similarly harassed.[36] In Bengaluru, after a local political leader singled out sixteen Muslim volunteers in a 'scam' at the city's Covid-19 response helpline, the list of 205 volunteers working in that specific zone was circulated on WhatsApp.[37] Volunteers on the list were inundated with Islamophobic calls and messages.[38] Not only did this hamper their ability to respond to calls on the helpline effectively, the harassment also led to psychological stress. The heavy use of digital tools in accessing and coordinating relief work implied that access

35 TNM Staff. (2021, April 28). 'Women shared their contact for COVID help – they got obscene pics, harassment in return'. The News Minute. Accessed on 6 July 2021 from: https://www.thenewsminute.com/article/women-shared-their-contact-covid-help-they-got-obscene-pics-harassment-return-147972

36 Saleem, S., & Richariya, J. (2021, May 31). 'Women Talk About Receiving Obscene Messages During Covid-Relief Work'. Live Wire. Accessed on 6 July 2021 from: https://livewire.thewire.in/gender-and-sexuality/women-talk-about-receiving-obscene-messages-during-covid-relief-work/

37 Bengaluru Bureau. (2021, May 7). 'Bed allotment scam: Emergency response hit by attrition from war rooms'. The Hindu. Accessed on 6 July 2021 from: https://www.thehindu.com/news/cities/bangalore/bed-allotment-scam-emergency-response-hit-by-attrition-from-war-rooms/article34500444.ece

38 Rao, M. (2021, May 7). 'Flood Of Islamophobia After BJP MP Publicly Names, Shames Govt Workers'. Article 14. Accessed on 6 July 2021 from: https://www.article-14.com/post/flood-of-islamophobia-after-bjp-mp-publicly-names-shames-govt-workers

to critical medical care was skewed towards those with digital and functional literacy.[39] Over 50% of India still lacks access to the internet. The targeted harassment, further restricted the group of people who could request or offer help. Women reported relying on male members of the family for any social media outreach. The harassment placed additional burden on friends and family of patients or volunteers, stealing critical time away from care-giving responsibilities, and further affecting their mental health in a stressful situation.

A less egregious concern was the secondary use of the private information such as phone numbers for secondary uses such as telemarketing during and after the crisis.

The harms from identification of individuals in relief work are not unique to chat apps. It is, however, possible that the blurred boundary of public and private communication on chat apps made people comfortable in sharing more personal information in these groups, even if these were groups of strangers. Understanding people's motivations and usage patterns of different social media platforms demands more dedicated research than the bird's-eye view analysis presented here. We did, however, hope to outline broad information sharing patterns that speak to the public/private distinction on these platforms and identify any emergent practices to respond to harms from identification of individuals on chat apps.

39 Barik, S. (2021, April 27). '"What is Twitter?": COVID-19 exposes India's digital divide'. Entrackr. Accessed on 6 July 2021 from: https://entrackr.com/2021/04/what-is-twitter-covid-19-exposes-indias-digital-divide/

ANALYSIS OF 21 WHATSAPP 'COVID RELIEF' GROUPS

Data Collection

We joined eighteen WhatsApp groups on 29th April 2021. By this time, India was deep into the second wave of the pandemic. It was still 10 days away from the peak, which hit on 8th May 2021. Cases in New Delhi, a city which saw acute shortage or medical oxygen, had started declining. Cases in Maharashtra, another badly hit state, were also on a downward trend.[40] As the pandemic spread to other parts of the country, volunteering energy spread beyond the main metropolitan centers.

We searched for WhatsApp group links shared on Twitter. Several of these groups had already reached the 256 limit, and we were thus unable to join them. Two groups had the 'disappearing messages' setting turned on. We took this to be an indication of intention to keep the communication on the group private and thus exited the groups immediately.

In the second phase of the analysis, a month into data collection we searched for more Covid-19 relief groups on Twitter. The number of such groups had declined. We identified and joined another

40 Hannah Ritchie, Esteban Ortiz-Ospina, Diana Beltekian, Edouard Mathieu, Joe Hasell, Bobbie Macdonald, Charlie Giattino, Cameron Appel, Lucas Rodés-Guirao and Max Roser (2020) - "Coronavirus Pandemic (COVID-19)". Published online at OurWorldInData.org. Accessed on 11 July 2021 from: *https://ourworldindata.org/coronavirus' [Online Resource]*

five groups on 9th June 2021. By this time the total number of Covid cases had declined to numbers at the peak of the first wave. The central and state health institutions had been able to streamline isolation and treatment protocols. We were motivated to join these groups to understand if the nature of relief groups had changed in the month since the peak of the second wave.

We have confined this analysis to an 8-week period from 29th April 2021 to 23rd June 2021. From preliminary analysis, we realized that two groups in the first batch of groups we had joined (on 29th April 2021), were not Covid-19 relief groups but were run by media organizations to exclusively share content from their organization. We exited these groups and excluded them from the analysis. Thus, the final data that we analyzed came from a total of 21 groups, 16 of which we joined on 29th April 2021 and 5 of which we joined on 9th June.

We joined these groups using phone numbers that were registered to Tattle Civic Tech. The username on the WhatsApp account was 'Tattle Civic Tech'. The 'about' section on the profile states: 'Tattle is a civic tech project that aims to archive content from Indian chat apps and social media. More details: www.tattle.co.in.' The account did not have a profile photograph.[41]

We used two parallel approaches to move data from the mobile device to a cloud service through which data could be analyzed. We used WhatsApp's export chat feature to frequently back up the history of a chat on the cloud service. WhatsApp's export chat feature results in a text file which contains the sender's name/ phone number, the time stamp of receipt and message content. WhatsApp claims that the export chat feature allows users to export 10,000 messages with media items and 40,000 messages

41 More details about our data collection protocol can be found on the Tattle website: https://tattle.co.in/products/whatsapp-archiver

with without media items.[42] In practice, the export chat feature fails to export media items reliably. Many media items downloaded on the device are not exported, and the total number of messages exported is also inconsistent. In total, we used the 'export chat with media items' feature 61 times. The maximum lines in any file exported were 4060. Exporting chat without media items was more reliable and captured most of the text messages. Excluding media items, however, is to exclude all multimedia related content, which is a crucial part of WhatsApp conversations.

In order to make up for this discrepancy in media retrieved from exported chats, we used a complementary approach of saving the media items that were downloaded in the 'WhatsApp Media Downloads' folder. Although WhatsApp is supposed to automatically download media items it receives if the 'Auto Download Media' setting is enabled, we were not able to reproduce this behavior on the Samsung Galaxy M10 we used. We resorted to manually clicking the media items to download them. We then backed up the media items download folder to a secure cloud service.

Since we were using one phone exclusively for the purpose of data collection from the Covid-19 relief groups, the media items in the media downloads folder from this duration came only from these groups. Unlike the export chat feature, the media downloads folder does not provide information about the group or the time at which the media item was received. It, however, captured over a thousand more media items than the export chat feature. To avoid repetition in media items, we discarded media items collected from the export chat feature for any media analysis and used media items only from the media downloads feature.

42 https://faq.whatsapp.com/android/chats/how-to-save-your-chat-history/?lang=en

Data Reporting

A consequence of the blurred lines between public and private communication on WhatsApp is that people may share personal and personally identifiable information on a WhatsApp group, even if it corrals strangers. Furthermore, while our goal was to analyze how information quality was managed on these groups, we did not want this report to become the means of discovering any of the leads shared during the second wave, whether they were actionable or not.

All the analysis presented in this report was conducted on deidentified text messages and media items. We did not access original phone numbers of senders. During the second stage of joining groups (on 9th June 2021), we did scan through the original groups we had joined, on the mobile phone, to understand which ones to exit.

All personal details in images used in this report and visualizations were obfuscated after the analysis. For images used in this report, we manually blacked out phone numbers and personally identifiable details. For images used in the visualizations, we followed four steps for obfuscation: we blurred the images; resized them; used a machine learning algorithm to detect faces and conceal them with black boxes; and finally, manually scanned all the images and concealed phone numbers and other personal information still visible after the first three steps. The Appendix contains more details on the technical aspects of anonymization.

Prior to the publication of this report, we removed all non-anonymized data as well as original media items from any cloud service we had used.

Despite our intentions and effort to scrub out phone numbers or personal information from the data, prior to reporting some media items may have escaped our attention. If any such media items

or security vulnerabilities in our data management practices are discovered, please email us at tarunima@tattle.co.in. We will act on it immediately.

The Analysis

Over the 8-week period we collected 16,694 text messages, 2,415 images (2,296 unique). We also collected over 200 videos during this time, but we did not undertake any video analysis in this study.

At the time of joining, we were primarily targeting groups in English, Hindi and Marathi, since these were the languages understood by the team members. In addition, we also joined one group in Telugu. In the final sample of posts, we also found messages in Tamil, Telugu and Gujarati.

Total Number of Groups	21
Number of Text Messages	16,694
Total Number of Images	2,415
Number of Unique Images	2,296
Number of Unique Senders[43]	1,192
Duration of Analysis	29 April 2021 – 24 June 2021 (for 16 groups) 9 June 2021 – 24 June 2021 (for 5 groups)

Figure 8: Snapshot of data from WhatsApp Conversations

43 Two separate phone numbers (even if used by the same person) would be counted as two unique senders.

In this analysis, we have focused on the content rather than the networks through which information was circulating. We didn't try to understand individual sharing activity in these groups or overlap of individuals across groups. Our goal was to broadly understand **what** was shared in these groups. To that end, we relied primarily on automated techniques for broad insights about the data. Specifically for analyzing images, we use vector embeddings extracted from the images using a pretrained ResNet model. [44]For textual analysis, we relied on word-frequency based techniques. All non-English text was translated to English using a Google Translate API.[45] Since most images contained textual content, we also applied textual analysis to text in images. To extract text from images, we used Google's Cloud Vision API. Figure 9, shows the results of text extracted from images using the Cloud Vision API.

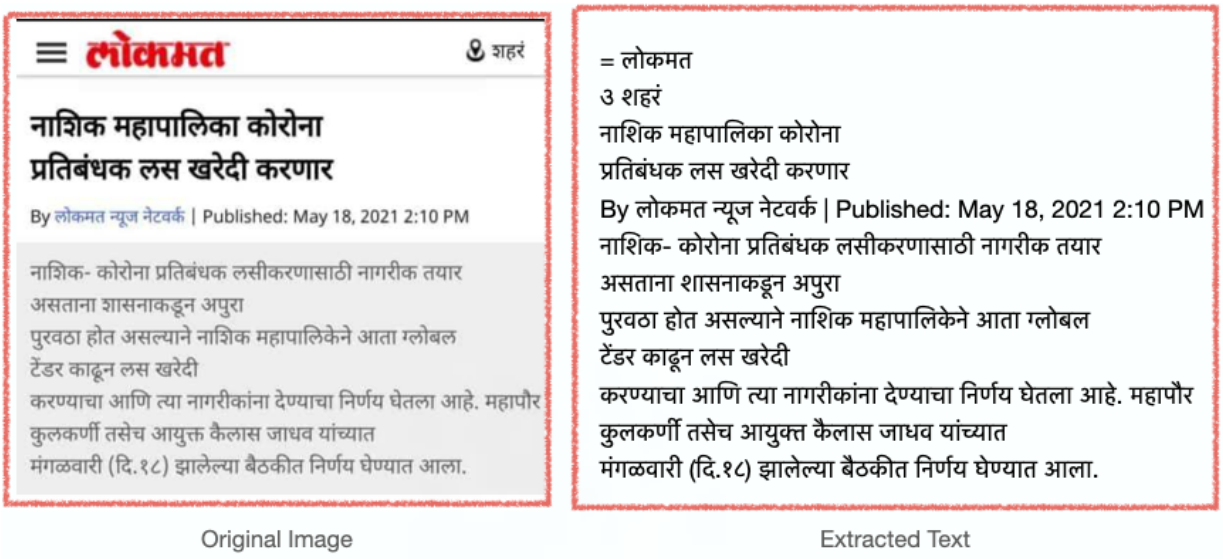


Figure 9: Result of Text Extraction from Images Using Google Cloud Vision API

44 The interactive T-Sne visualization can be viewed on the Tattle Website

45 See Appendix for methodology.

Each of these techniques is described in detail in the Appendix. While these techniques allow for at-scale analysis of social media content, these automated techniques also come with inherent limitations. These are described in the Limitations section of the Appendix. Machine learning based techniques, be it vector-embeddings, language translation or computer vision based text extraction, are error-prone. While we attempted some correction of known errors, a lot more can be done. The analysis presented here should be treated as indication of trends that merit further investigation, and not conclusive assertions.

Trend 1: Heavy Use of Information from Other Social Media Platforms

The biggest cluster in the vector embeddings based image grouping is of screenshots of posts from Twitter and Instagram. A scan of this cluster shows that majority of these screenshots are leads for medical oxygen suppliers, drugs, hospital and ICU facilities and other medical supplies, indicating heavy reliance on other social media platforms for finding medical aid.

The cluster of web and mobile screenshot images is also similarly large. This cluster contains a greater diversity in the content of images—some are screenshots of WhatsApp and Facebook posts and some of apps and websites with resources of Covid-19 related information.

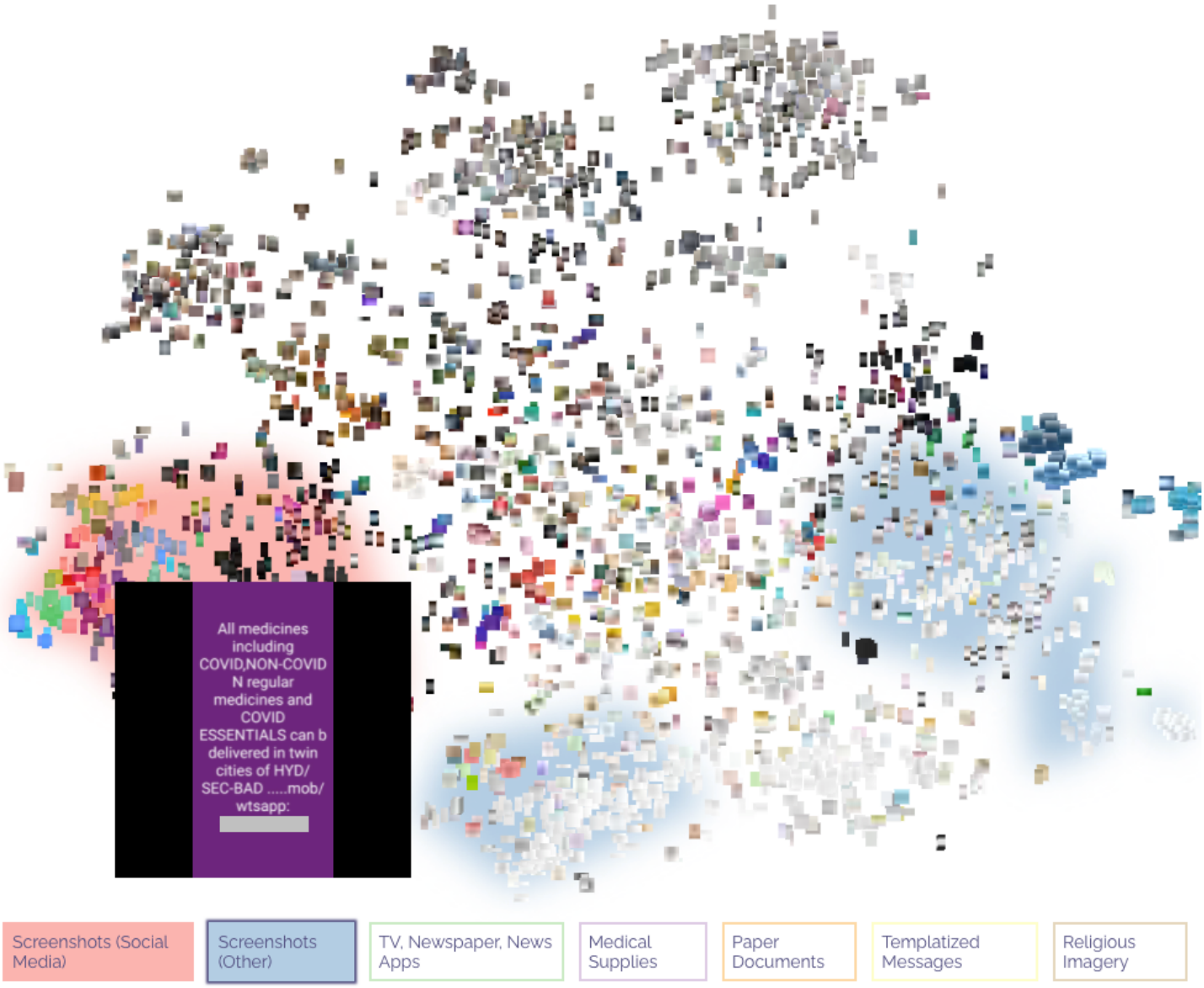


Figure 10: Image Clustering based on Visual and Semantic Meaning
https://tattle.co.in/articles/covid-whatsapp-public-groups/t-sne/

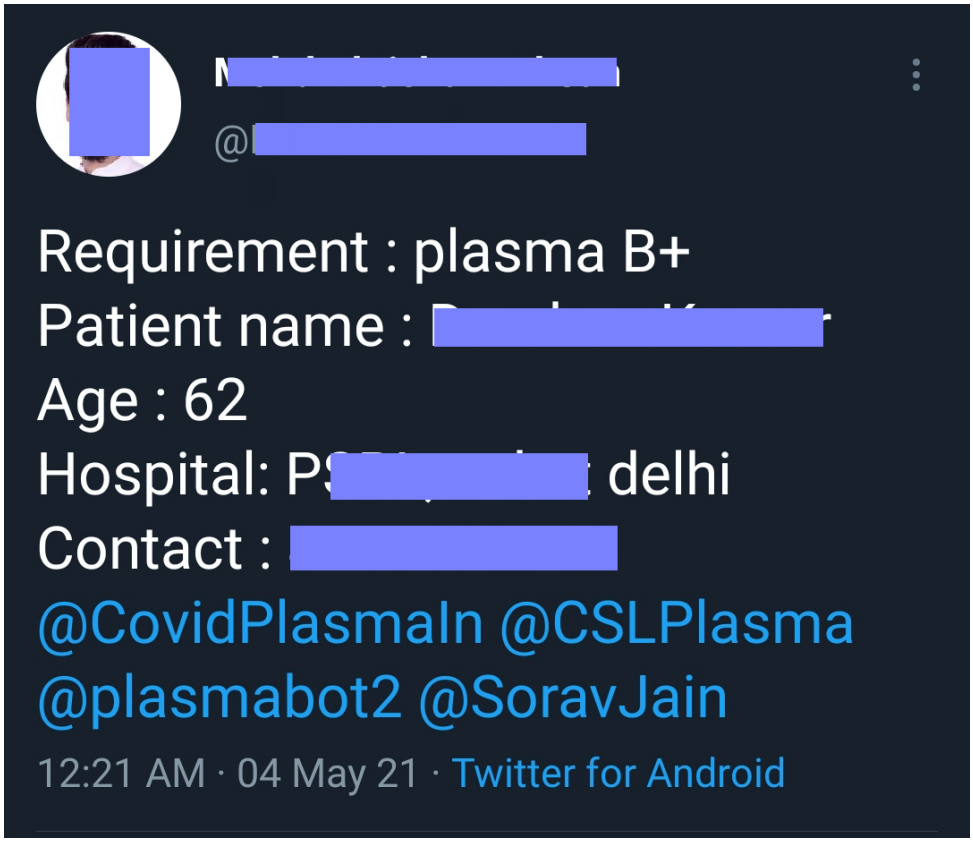


Figure 11: Screenshot of a Tweet Shared in the WhatsApp Groups

At least 9% of the text messages (1,551 messages) contained links to other websites. Twitter was the most popular social media in these WhatsApp groups. We found that 21% of all external links (330 messages) contained links to tweets. There were 168 messages with YouTube links; 49 with Instagram links; 190 messages contained links to other WhatsApp chat groups; and 30 contained links to Telegram groups.

The Instagram, Twitter and Facebook posts shared on WhatsApp as images were primarily text posts on their native platforms. Images emerge as a modality to conveniently share content across platforms. Content from government apps, online dashboards and other social media platforms is given broader reach by sharing screenshots on WhatsApp. This circumvents the need for people

to create accounts or access specialized websites. To understand if the extent of reliance on external sources was unique to Covid-19 relief work on chat apps, would require comparison with other datasets collected from WhatsApp. To the best of our knowledge, this analysis is the first example of similarity based grouping of WhatsApp content, allowing for visualization of broad trends in images collected from the platform. Similar aggregation of other WhatsApp datasets could allow for a more direct comparison and contrast of media typologies shared on WhatApp around different events.

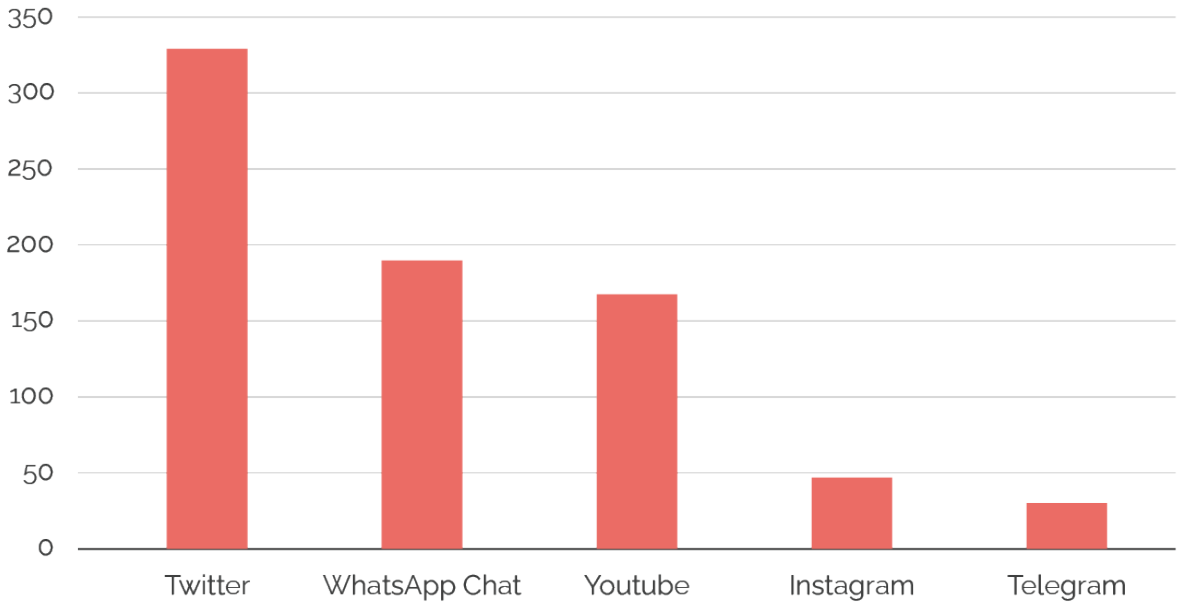


Figure 12: Number of messages with links to specific platforms

Trend 2: Images of Prescriptions, Medicines and Receipts

Another big cluster in the image similarity grouping is of medicines, concentrators, medical prescriptions, receipts and other paper documentation. Sharing medical prescriptions and receipts could be attributed in part, to the need to assert legitimacy and urgency of need for drugs, hospital beds and blood donations in a situation where many were requesting for aid. Images also circumvent the effort needed in typing out information into a text message format. These prescriptions and receipts, however, contain several personal details such as patient name, hospital of admission, blood group and co-morbidities.

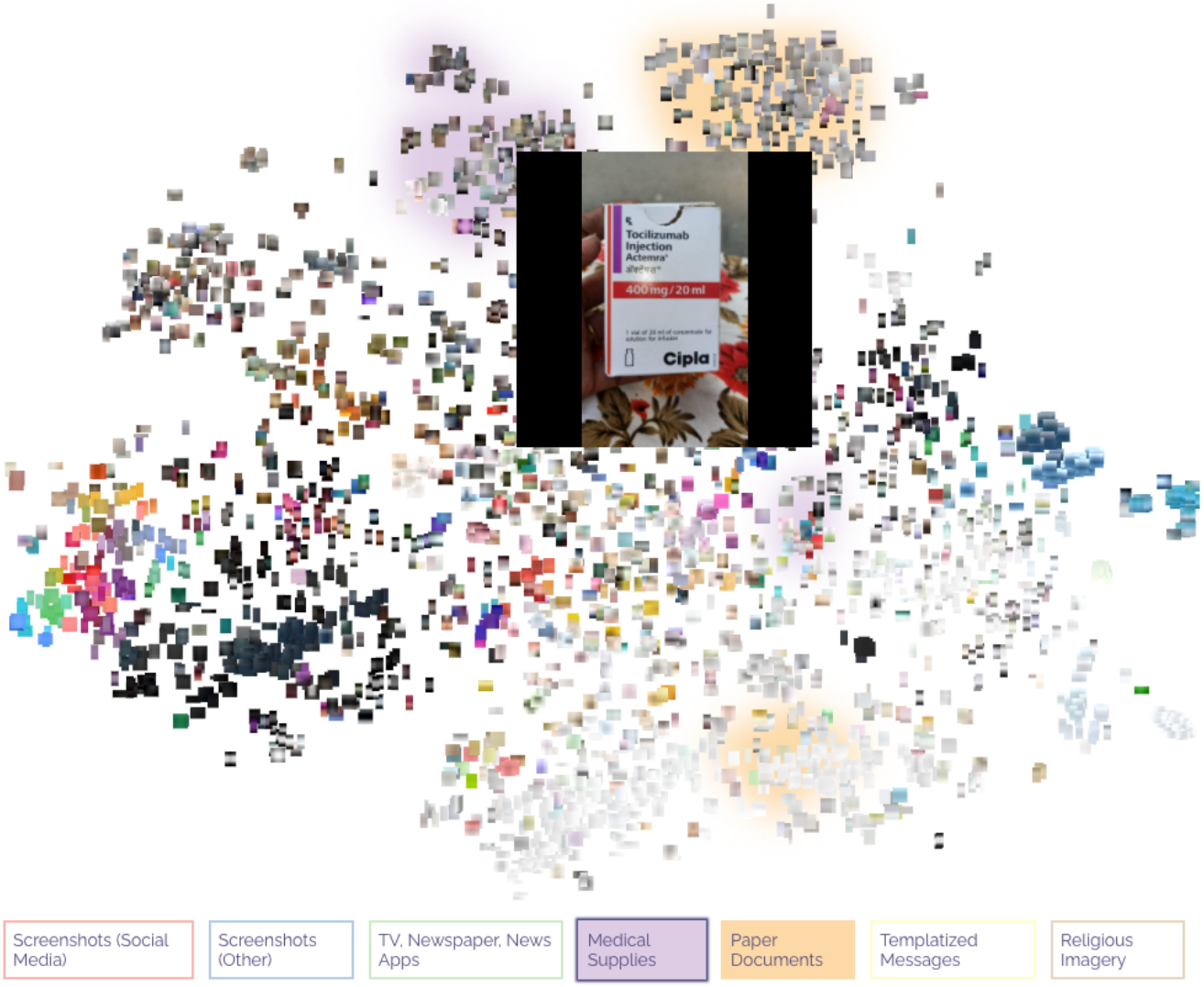


Figure 13: Group of Images of Medical Supplies and Paper Documents
https://tattle.co.in/articles/covid-whatsapp-public-groups/t-sne/

The images of medical supplies, in particular, parallel the demand (gaps) observed for these during the crisis. For example, there are a number of images of the drug Amphotericin. Amphotericin is used to treat fatal fungal infections such as Mucormycosis (Black Fungus). The data collection period coincided with the rise of Mucormycosis tied to Covid-19 treatment. The multiple instances of images of Amphotericin and concentrators indicate an interest in availing or correcting information about these specific medical resources.

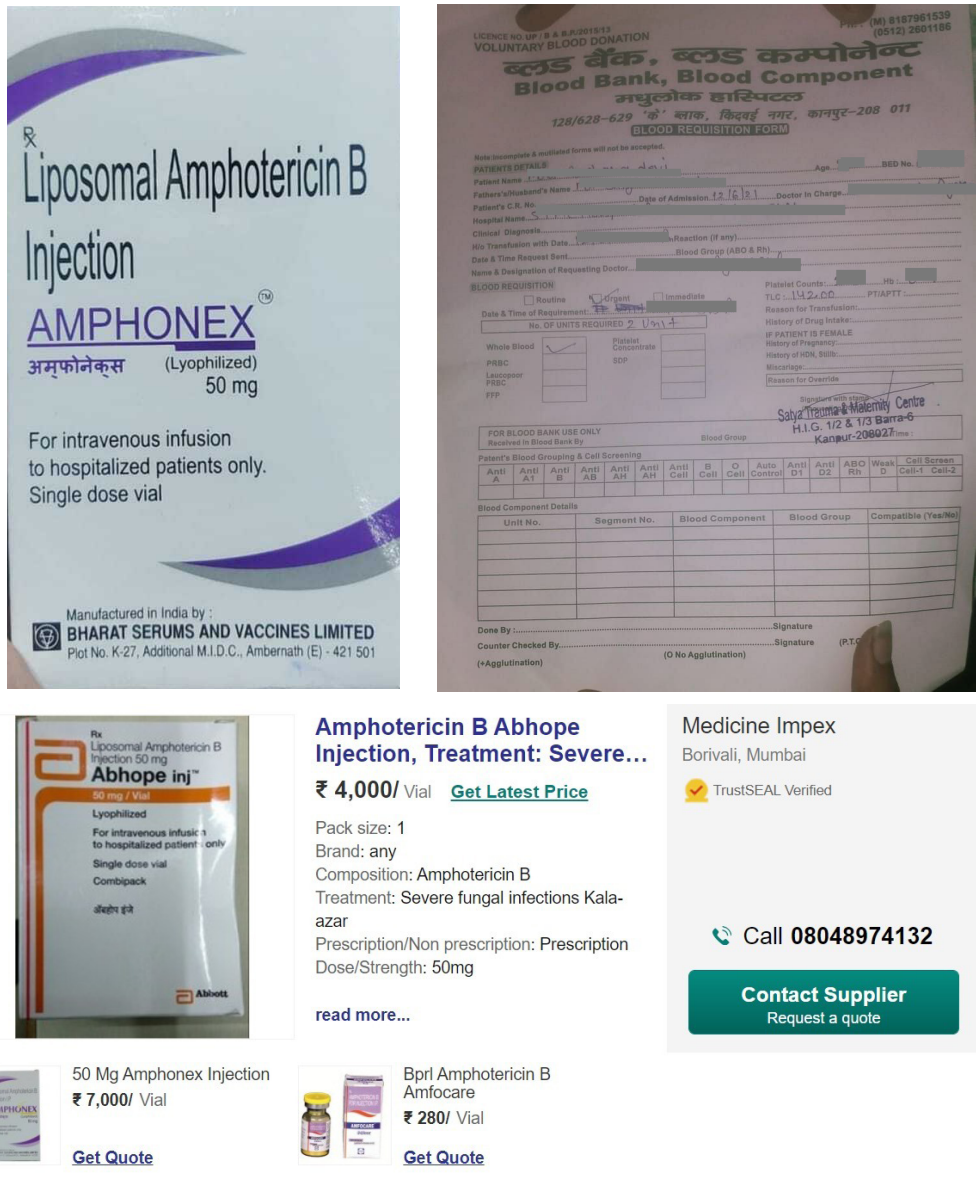


Figure 14: Images of Amphotericin Shared in the Groups

Trend 3: Healing Does Not Imply Only Medical Treatment

In the image grouping, we also found two unexpected clusters of images of gods and of close-up of people's faces. We tracked the images of people's faces to a specific spiritual WhatsApp group which had the terms 'Covid' and 'healing' in the group name. The group aimed to provide healing to Covid patients remotely. Similarly, we tracked images of gods to a specific group which was sometimes used for sharing resources for Covid-19 relief, but was predominantly used for sharing images of different Indians gods. In a time of desperation, when medical advice or medicines were not available, people looked beyond medical science for healing of loved ones.

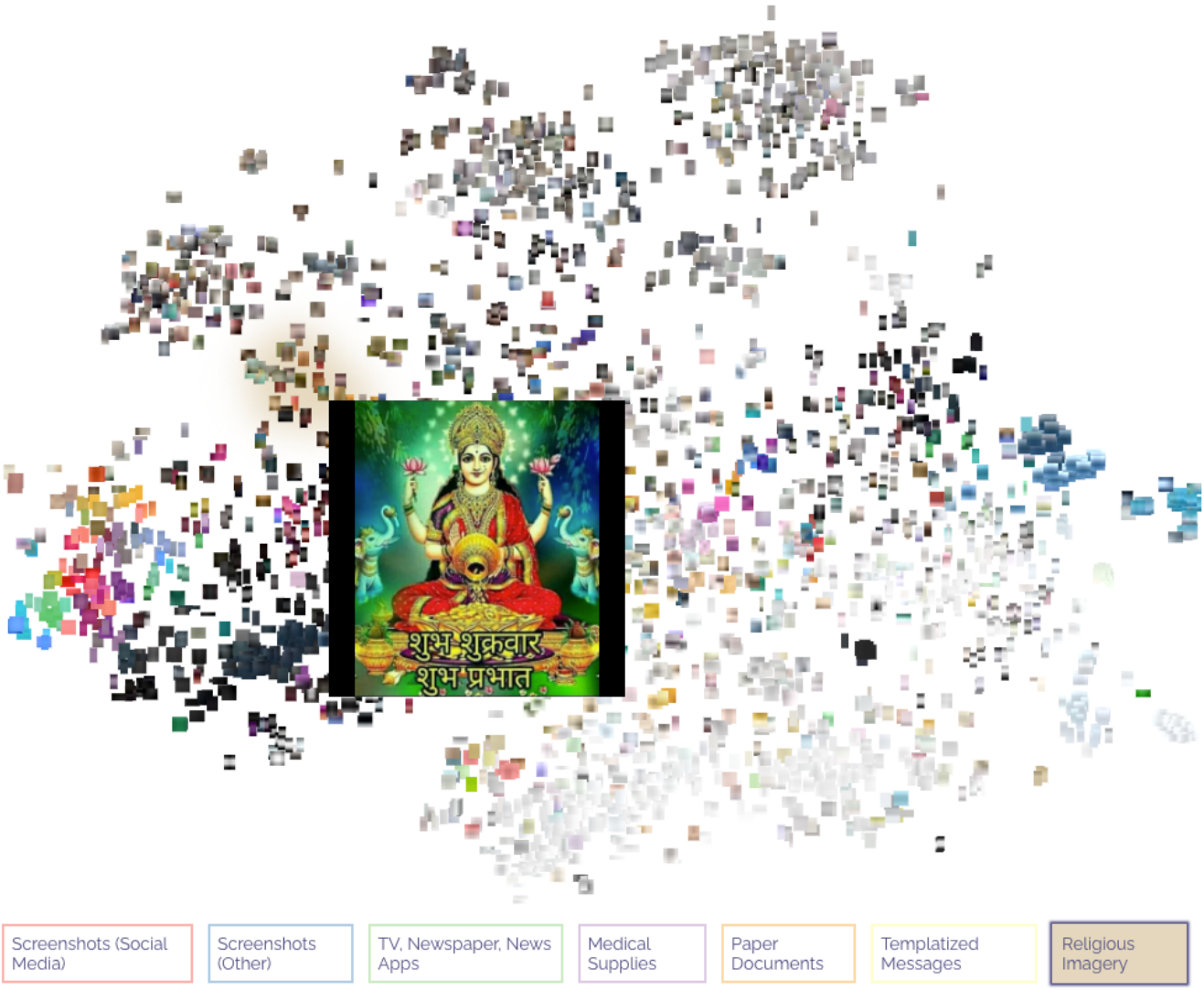


Figure 15: Cluster of Images of Indian Gods
https://tattle.co.in/articles/covid-whatsapp-public-groups/t-sne/

Trend 4: Low Overlap with Database(s) of Verified Leads/ Scams

With the proliferation of leads for medical resources during the second wave, multiple volunteering groups started maintaining databases of 'verified leads'. In preliminary manual annotation of the text messages on the 21 WhatsApp groups, we discovered at least 257 unique phone numbers shared as leads. We compared this list to a database of Covid-19 Helpline numbers verified by FactChecker.in, the oldest fact-checking group in India. The fact-checking group has been verifying Covid-19 helpline numbers sourced through a tip line as well as through social media monitoring since the beginning of the second wave.[46] As of July 4, 2021 the list had 510 'verified' phone numbers.[47]

We found that less than 17% of the leads shared in the WhatsApp text messages (37 of 257 leads) were captured by the FactChecker.in database. Accounting for the leads shared in images in these groups takes the tally of overlapping unique leads to 42. There were five leads in the images that were not contained in the text messages.[48]

Possibly to prevent amplification of inaccurate information, the FactChecker.in database does not list the leads that were checked and were found to be inaccurate or inactive. To estimate the leads shared on the WhatsApp group that were known to be fraudulent, we also compared the phone leads in the text messages against a crowdsourced database of scam numbers called The CoVid Scam Directory. The database is maintained by the volunteering group: CoViD Action Initiative.[49] Any individual could add an entry for a 'scam' number. As of July 4, 2021, the database had 812 phone

46 FactChecker.in (2021, June 22). 'FactChecker Called Up All COVID-19 Helplines'. Pg 20.

47 Analysis based on the status of the database on July 4, 2021.

48 There could be some error here due to imperfection in computer vision techniques for extracting text from images.

49 https://cov.social/#/

records of which 647 were unique. In addition to an API, the group's website provides a simple search functionality to search for specific numbers in the database. The entire database of numbers is not open access, but the volunteer group shared the database with us on request.

16 phone numbers reported as 'scams' on the CoViD Scam Directory were found in the text messages. But only 2 of the numbers from the CoViD Scam Directory were shared in the WhatsApp groups as warnings. The remaining 14 numbers were shared as genuine leads.

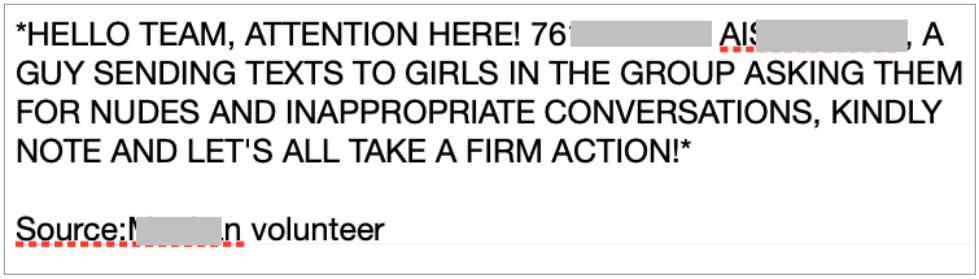


Figure 16: A Text Message in the Groups Warning Others about a Specific User

Since the CoViD Scam Directory is crowdsourced, all numbers listed in the directory shouldn't be assumed to be scam numbers. Some numbers might have been added because they were inactive at the time a person looking for aid called them. Some might have also been added by people frustrated by the inability of the helpline to aid them. What is salient, however, is that even in this database the overlap of numbers is low—less than 6% of the leads shared in the 21 WhatsApp groups were captured in the CoViD Scam Directory.

Despite the large volume of phone leads collected by both the Factchecker.in and the CoViD Scam Directory, majority of phone leads shared in the 21 WhatsApp groups were not found in either of these databases. This speaks to the volume of information that was shared and needed to be verified during the second wave of the pandemic in India.

Trend 5: Differences between Text Contained in Images and Text Contained in Text Messages

The visualization of image groups in the WhatsApp conversations revealed a common theme of images being used as convenient mechanism to share information from other platforms. We wanted to understand if the images contained similar information as text messages or if the information shared varied with the modality. We thus compared the textual content in the images with that of text messages.

Figure 17, 18 and 19 provide a high level summary of the words contained in images and texts. The word clouds hint that while the words used are common across text messages and images, they vary in their relative frequency. Words such as 'hospital', 'patient', 'available' were common to both text messages and images. But words such as 'help', 'need', 'contact', which are amongst the five most frequently used words in text messages, were not amongst even the ten most frequently used words in images.

Figure 17: (Left) Words Contained in Images (Right) Words Contained in Text Messages

FIVE MOST FREQUENT WORDS IN TEXT MESSAGES	
Term	*Number of Occurences*
Hospital	1,335
Need	957
Patient	835
Contact	783
Lead	763

Figure 18

FIVE MOST FREQUENT WORDS IN TEXT IN IMAGES	
Term	*Number of Occurences*
Patient	1,332
Hospital	1,290
Available	1,270
Day	682
Report	648

Figure 19

Absolute numbers of occurrence of terms across images and text can't be directly compared since the volume of text messages is significantly higher than media messages. We thus compare the proportion or percentage of occurrence of terms (number of times a term is used divided by total number of words) in both these datasets.

Figure 18. shows the percentage frequency of terms in images against the percentage frequency of terms in text. The graph is restricted to the 30 most frequent words in images and text messages (a total of 38 words). Please see the Appendix for more details.

This comparison shows that while words such as 'available', 'blood', 'patient', 'oxygen', had nearly equal representation in images and text messages (close to the y=x line), words such as 'need' and 'help' were significantly more common in text messages. This

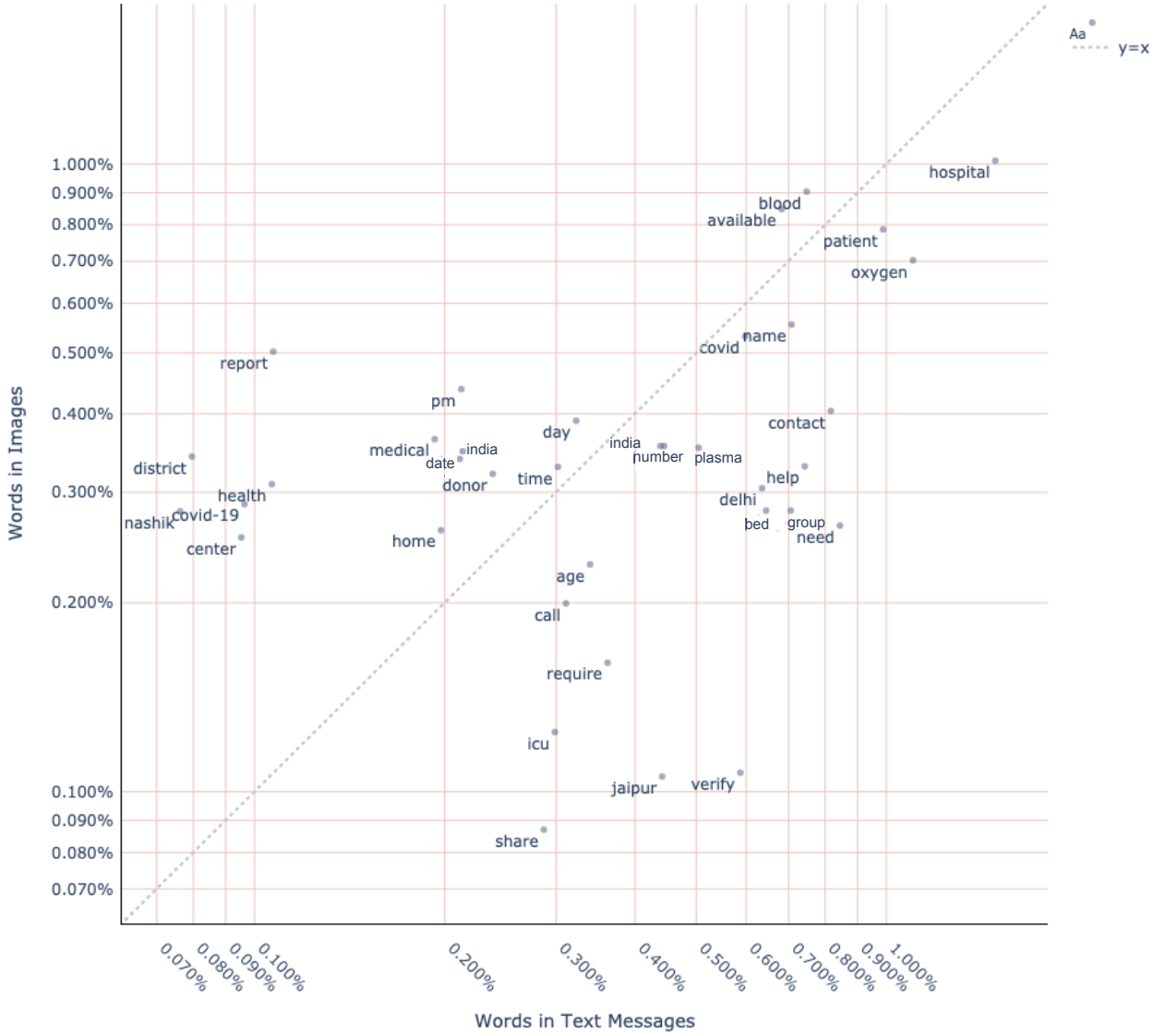


Figure 20: Comparison of Words in Images and Text Messages

preliminary analysis suggests that while both text messages and images (which are screenshots of information from other platforms) were used to advertise for availability of oxygen or blood donors, the request for medical aid was more frequently circulated as text messages native to WhatsApp. This analysis relies on text extracted from images using Cloud vision techniques and automated language translation, both of which are prone to error. The claim presented here merits more in-depth research, with more manual scanning of individual messages.

Trend 6: Variation in Conversation Over Time and Across Groups

The analysis, so far, summarized conversations in 21 groups across the 8-week period. The situation during the second Covid-19 wave in India, however, was changing rapidly. We tried to analyze if the conversations in these groups reflected the changing status of the pandemic.

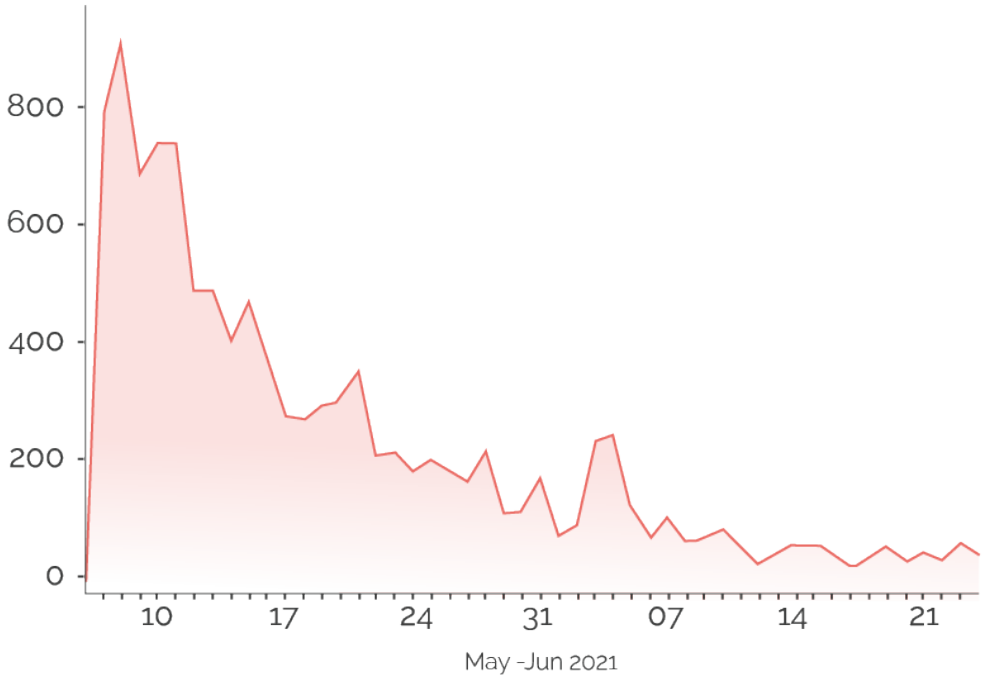


Figure 21: Frequency of Messages by Week

Since we had tracked only sixteen groups for the entire 8-week duration, we limited the temporal analysis to the sixteen groups. Furthermore, we discovered that text messages were missing for a few days of the first week. We thus discarded content from the first week and carried out temporal analysis over a 7-week period starting from 6th May 2021 and ending on 25th June 2021.

To carry out the temporal analysis, we analyzed the prominent words uses in each of the 7 weeks. The aggregate analysis of word frequencies in text messages showed that words such as 'hospitals', 'patient', 'oxygen', 'blood' were prominent in text messages (see previous section). Even in the list of 30 most frequently used, the frequency of usage of these terms is significantly higher than the frequency of other terms on the list.

While recognizing that these words were important in the text messages during the 7 weeks, we wanted to capture the unique themes in a conversation in any week. For that, we used a technique called Term Frequency - Inverse Document Frequency (TF-IDF) which gives prominence to words in a week that are more salient compared to terms in other 6 weeks.[50] Words such as 'hospital' and 'oxygen' may still feature in a specific week, which would imply that their usage in that week was notably higher than the other weeks.

Figure 22 ranks 20 words as per their relative importance in any week. The word 'oxygen' stops appearing in the frequent words list after 20th May 2021 indicating that the need for medical oxygen had declined by then. The prominence of the words 'hospital' and 'plasma' also decline over the 7 weeks. Instead, we see words associated with hyperlinks like 'https', 'com' become more prominent. In the last week in particular, it seems that WhatsApp chat links were the prominent theme. Unexpectedly, we also see words such as 'CA', 'income' and 'tax' feature in the last two weeks.

We tracked these terms to a specific group that had started as a Covid-19 relief group but towards the end turned into a group for

50 Please see the Appendix for detailed methodology.

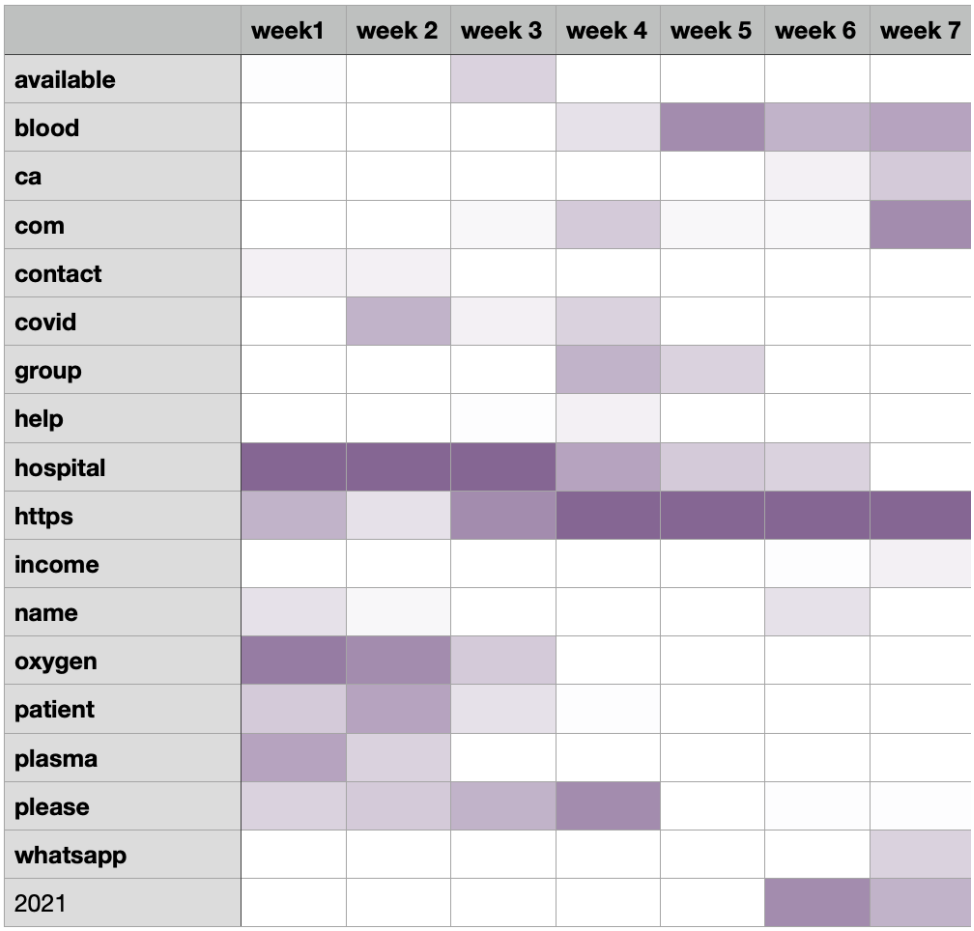


Figure 22: Top Sixteen Words Across Weeks (TF-IDF), dark means more important, light means less important

primarily sharing material related to chartered accountancy. A few links on chartered accountancy related webinars were shared on this group even in the last week of May. Towards late June, information about webinars on lung recovery and other health related topics was still shared on these groups, but its proportion relative to information about chartered accountancy declined.

From the conversations alone, it is not possible to ascertain the intention of the administrators in creating this group—did they create the group around a topical issue to find newer audiences for marketing their other interests (related to chartered accountancy), or was the decision to use the channel for marketing of other topics a post-facto one? Regardless of the intention, this group highlights a less explored use of WhatsApp groups.

DISCUSSION

Public and Private Boundaries Are More Blurred in Emergencies

The pandemic has raised an important debate on whether individual privacy is negotiable during an emergency.[51] This debate has tended to focus on demands for personal data such as mobile location or biometric ID placed on citizens by states. This analysis from a small sample of WhatsApp groups shows that in a state of crisis people have willingly shared not only personally identifiable data but also sensitive personal data with a group of strangers.[52] In addition to prescriptions and medical records, we also found images of Aadhar cards on these groups.

Experiments in behavioral economics reveal that people's estimation of the value of privacy is highly sensitive to their circumstances and immediate situation.[53] With Covid-19, patients and their families were reaching out to these relief groups because they were desperate for life-saving aid. It is thus unsurprising that people were willing to share sensitive and personal data in the hopes of receiving the urgent care needed during treatment of Covid-19.

As documented earlier in this report, there are tangible harms from identification of people from the information shared in these

51 Zwitter, A., Gstrein, O.J. Big data, privacy and COVID-19 – learning from humanitarian expertise in data protection. Journal of International Humanitarian Action 5, 4 (2020). https://doi.org/10.1186/s41018-020-00072-6

52 As defined in the Personal Data Protection Bill (2019). Bill No. 373 of 2019 Introduced in The Indian Lok Sabha. Accessed on 12th July 2021 from http://164.100.47.4/BillsTexts/LSBillTexts/Asintroduced/373_2019_LS_Eng.pdf

53 Alessandro, A. John, L.K, Loewenstein, G. 'What Is Privacy Worth?'. The Journal of Legal Studies 42, No. 2 (June 2013): Pg. 249–74. https://doi.org/10.1086/671754

groups. Templatized information makes it easier to find resources for people, but it also produces data ripe for secondary uses. While on Twitter and Instagram, the person asking for aid has the agency to pull down the request if they choose to, on WhatsApp, the sender cannot affect the circulation of information once they have sent it out.

When we were first joining groups, two groups had the 'disappearing messages' setting enabled (which is why we did not collect data from these groups and exited them). This setting implied that all the messages from the group would be removed from it after a duration decided by the admin of the group. These messages would not only disappear from the sender's phone, but also from the phones of all the participants of the group. As the second wave revealed, the lifespan of information is short during a crisis. Despite the groups being discoverable by strangers through a more public forum, people shared sensitive data on these groups. Enabling messages to disappear after a certain duration can minimize the duration for which an individual's personal information is available to ill-intentioned actors on these groups and minimize harm from identification of individuals.

Admins might also consider deleting the groups altogether, if the group is no longer serving the purpose it was created for. In many of the groups, the frequency of conversation declined towards the end of the analysis period. In some cases, participants started posting information unrelated to the pandemic. Destruction of data is the final step in data life cycle management and one that WhatApp group admins could heed more carefully. [54]

54 John R. Talburt, Yinle Zhou, Chapter 2 - Entity Identity Information and the CSRUD Life Cycle Model, Editor(s):John R. Talburt, Yinle Zhou. Entity Information Life Cycle for Big Data. Morgan Kaufmann. 2015, Pp. 17-29, ISBN 9780128005378, https://doi.org/10.1016/B978-0-12-800537-8.00002-8

The Social Media Mix-and-Match

This analysis reveals a specific way in which different social media platforms were used in conjunction, during relief work. WhatsApp, with 400 million users, is the most popular platform in India. Twitter has fewer than 20 million users. The media and text analysis indicated that despite the low user base, relief work, even on WhatsApp, relied heavily on Twitter. In the specific groups we were tracking, it appears that WhatsApp was the primary channel to collect requests for aid, but when it came to advertising availability of resources, people also sought information on Twitter and Instagram. These platforms could emerge as centralized, constantly updating repositories of information in a way that WhatsApp could not.[55] While information flow on WhatsApp depends on users pushing information, Twitter and Instagram also allow for pulling information. This ability to pull information, on demand, is important when the goal is to connect those in need with useful information.

This analysis also reveals an important role for the 'go-betweens' who connect WhatsApp users to information on Twitter and Instagram, giving content on these platforms greater reach than that reflected by the engagement metrics on the platforms. People shared screenshots as well as direct links to tweets in the groups we were tracking. Possibly because Instagram provides restricted functionalities for unregistered users, we see a big cluster of screenshots of Instagram stories but very few direct links to Instagram content in these conversations.

Credibility Indicators for WhatsApp

Even though people relied on other social media platforms such as Twitter and Instagram for information about medical aid, leads for the same were also shared as native WhatsApp messages. On social media platforms such as Twitter, the source of the message (the person or account sharing it) can be used to evaluate its

55 Several dashboards aggregated information resources shared on Twitter. For example, see: https://external.sprinklr.com/insights/explorer/dashboard/601b9e214c7a6b689d76f493/tab/4?id=-DASHBOARD_601b9e214c7a6b689d76f493

credibility.[56] Similarly, the time stamp of creation of a tweet or a web page, can help understand whether the content is out of context or references old/inaccurate information.

Such indicators are absent on any messaging platform. By definition, messaging apps are meant to be used for direct messaging between users. For a recipient of a message, the source of the message is the person sending it. Origination information, be it of the person creating a message or the time stamp of creation, can't be technically hard-coded in a message within the ambit of private and secure messaging.

The technical design of messaging apps notwithstanding, some credibility markers to assess whether a lead shared could be trusted were needed. In the groups we were tracking, we saw several messages being 'signed' with a time stamp and/or information about the originator of the information. While sharing the lead, people added a line about who had verified it and when the message was last verified. The information about who had verified the lead, is relevant context for people in the group who know the individual. But the time of verification would be relevant context, even if the messages were circulated outside the groups.

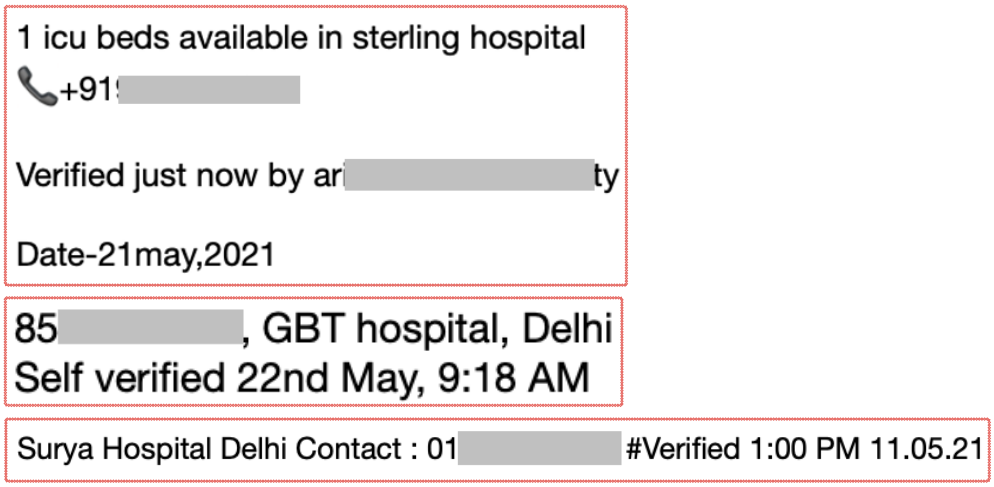


Figure 23: WhatsApp text messages with information about the time of verification and the person who verified it.

56 Castillo, C., Mendoza, M., Poblete, B. Information credibility on Twitter. In: Proc. WWW, pp. 675–84. ACM (2011), http://doi.acm.org/10.1145/1963405.1963500

Since these credibility markers are user created, they can be spoofed. Scamsters could just as easily add such a line to their messages. But within a group, where a baseline level of trust between members can be assumed, the additional information is useful context. While recent regulatory, academic and media attention has focused on technical mechanisms for suppressing misinformation on chat apps,[57,58] these examples reveal social practices for information quality management, which also deserve more attention.

The Need for Distributed but Coordinated Verification

As the previous analysis shows, the phone numbers shared in the 21 Covid-19 relief groups we were tracking, had low overlap not only with a database of verified leads, but also with a crowdsourced database of scam phone numbers—the majority of numbers shared in this small sample of 21 groups could not be or had not been verified.

This reflects the volume of information that needed to be verified during the crises. During the second wave, contact details of pharmacies, hospitals, medical oxygen suppliers, ambulance services in every neighborhood became objects of verification. The goal of verification was not only to weed out scammers, but to also check on the status of availability of resources with different suppliers. Verification work was directed to aggregate the leads that were still active and actionable. Every location had its own list of relief channels that could be verified.

Fact-checking and verification of political and health related misinformation has primarily been concerned with debunking

57 Agarwal, S. (2021, March 23). 'India proposes alpha-numeric hash to track WhatsApp chat'. The Economic Times. https://economictimes.indiatimes.com/tech/technology/govt-proposes-alpha-numeric-hash-to-track-whatsapp-chat/articleshow/81638939.cms?utm_source=contentofinterest&utm_medium=text&utm_campaign=cppst

58 Reis, J. C. S., Melo, P., Garimella, K., & Benevenuto, F. (2020). 'Can WhatsApp benefit from debunked fact-checked stories to reduce misinformation?'. Harvard Kennedy School (HKS) Misinformation Review. https://doi.org/10.37016/mr-2020-035

specific posts that have wide geographical reach. The more than 50 incidents of lynching in India in 2018 were linked to specific rumors on WhatsApp around child-lifting.[59] The rumors were debunked early in the year. The bigger challenge was in disseminating the results of fact-checking broadly to the Indian population. Specific posts or narratives become viral. Recent academic research has emphasized that the difficulty of misinformation response is not in debunking, but rather in matching and countering the virality of the original posts.[60] [61]

The second wave of the pandemic in India resulted in an unprecedented situation that challenged the verification process itself. Every location had unique leads that needed to be verified. Leads for medical resources did not have to 'go viral' to cause harm. Even one person acting on an inaccurate or fraudulent lead could lose critical time or money. The second wave resulted in spontaneous generation of hyperlocal information across India, simultaneously. All of this (mis)information was a worthy object of verification.

The analysis reflects on the volume of content that needed to be verified, and the challenge associated with it. Only 37 phone numbers shared in the WhatsApp conversations across eight weeks were captured in a national-level database of over 500 verified leads. There are other databases maintained by volunteering groups that we could also compare the WhatsApp conversations to. The existence of these multiple databases, however, also speaks to the challenge of coordinating and triaging

59 IndiaSpend. (2018, July 9). 'Child-lifting rumours caused 69 mob attacks, 33 deaths in last 18 months'. Business Standard. Accessed on 6 July 2021 from https://www.business-standard.com/article/current-affairs/69-mob-attacks-on-child-lifting-rumours-since-jan-17-only-one-before-that-118070900081_1.html

60 C. Geeng, T. Francisco, J. West, and F. Roesner. (2020, Dec) 'Social media covid-19 misinformation interventions viewed positively, but have limited impact'. Accessed on 8th July 2021 from https://arxiv.org/abs/2012.11055

61 National Academies of Sciences, Engineering, and Medicine; Health and Medicine Division; Board on Population Health and Public Health Practice; Roundtable on Health Literacy; Wojtowicz A, editor. 'Addressing Health Misinformation with Health Literacy Strategies: Proceedings of a Workshop—in Brief'. Washington (DC): National Academies Press (US); 2020 Dec 17. Accessed on 8th July 2021 from: https://www.ncbi.nlm.nih.gov/books/NBK565935/ doi: 10.17226/26021

actionable information during a crisis. A patient or their family member looking for actionable information in the emergency could have relied on multiple volunteer run channels. Not all volunteer run channels are equally effective or devoted to focused or high quality information (as reflected by chartered accountancy related content in one of the groups we were tracking).

The nature of information needed during the second wave (phone numbers, status of medical facilities, etc.) demanded localized verification efforts. But it also demonstrated the need for coordination of these efforts.

The righteousness and importance of truth-seeking inherent in fact-checking and verification work can sometimes frame the entities being fact-checked as immutable objects that are unaffected by the *process* of such work. Fact-checking and verification focuses on changing perceptions about that which is being investigated, at the risk of making invisible the effects of investigation on those subject to it.

Many of the phone numbers and leads shared during relief work were genuine. Verification work during the second wave also entailed calling these genuine leads—hospitals and medical suppliers who were over-burdened in responding to and helping patients or their families. From their perspective, calls verifying if they were functional or had specific medical resources could be a distraction from responding to patients with an urgent need. Furthermore, since multiple volunteers and groups were involved in verification, hospitals and medical suppliers received multiple calls or messages verifying their status. The verification work during the second wave of Covid-19 in India, thus, also humanized those subject to fact-checking.

The volunteering energy during the second wave of the pandemic in India helped save lives. The localized nature of information that needed to be verified also speaks to the importance of localized fact-checking. However, volunteering action in the second wave

showed that the decentralized verification by multiple groups simultaneously, is not always efficient. Specifically for the kinds of information shared in Covid-19 relief, coordination of the distributed efforts could have helped. This raises the question of who can play the coordinating role. There can be several candidates for this role, but the conditions that enable such candidates to appear are not obvious. Multiple groups have to agree to trust the coordinating entity to maintain quality control, resolve overlaps and disagreements between the groups and continually convince the participating groups that the transaction costs of coordinating with other groups are worth it. Can such conditions appear in an emergency? Or can they only be created in more 'normal' times, but leveraged in crises?

The implications of crowd-sourced verification work during the second wave also bear on professional fact-checking work. Over the last four years, there has been a steep rise in fact-checking operations. In India, there are at present 15 fact-checking groups certified by the IFCN.[62] There is redundancy in the stories covered by these fact-checking groups. Redundancy in fact-checking can be helpful when it serves as a self-correction and standard setting mechanism within the fact-checking community. At the same time, as the verification work during the second wave made obvious, it can also be inefficient and possibly interfere in the work of well-intentioned actors.. There is a balance to be struck to make redundancy in fact-checking and verification productive. In situations such as disasters, pandemics or elections, where new information may emerge or information may change rapidly, how could professional fact-checking groups better coordinate efforts? Such coordination has been attempted around elections in several countries.[63] [64] Elections, however, are a planned, foreseeable event. Is it possible to enable coordinated fact-checking efforts to kick-in around sudden and unexpected events?

62 https://www.ifcncodeofprinciples.poynter.org/signatories

63 Flueckiger, S. (2020, July 14). Verificado 2018: Fighting misinformation collaboratively. World Association of News Publishers. Accessed on 12th July 2021 from https://wan-ifra.org/2019/11/verificado-2018-fighting-misinformation-collaboratively/

64 EKTA News Coalition. EKTA. Accessed on 12th July 2021 from https://ekta-facts.com/

A New Facet of The Information Disorder

Disinformation and 'fake news' assumed the status of a global crisis in 2016 after the Brexit vote and the US presidential elections. The ability of false information to influence public opinion has countries legitimately concerned about narratives that undermine elections, drive political polarization and fuel ethnic, racial or communal tensions. Prior to the pandemic, civic action focused on responding to 'political' misinformation—misinformation that pertained to elections, political candidates or topical political issues.[65] The pandemic drew attention to health related misinformation. Some health related topics such as vaccines can become politicized, but many health related issues are less prone to motivated reasoning than outright political news such as action by elected representatives.[66] Responding to health related misinformation came with its own set of challenges, such as communicating information under uncertainty while honoring existing sociocultural norms. But as a less politicized topic, it also allows for alignment of incentives and more streamlined communication across different stakeholders such as governments, multilateral organizations, domain experts and community based organizations.

Emerging conceptions of misinformation response can be loosely divided into demand side and supply side interventions. Supply side interventions have focused on detecting and suppressing low quality information and coordinated inauthentic behavior on online platforms. On the demand side, misinformation response has focused on de-bunking or pre-bunking for social media consumers, possibly led by ideologically aligned sources.[67]

65 Swire-Thompson, B., & Lazer, D. (2020). Public Health and Online Misinformation: Challenges and Recommendations. Annual Review of Public Health, 41(1), 433–451. Accessed on 12th July 2021 from https://doi.org/10.1146/annurev-publhealth-040119-094127

66 Schaffner, B. F. Roche, C..(1 March 2017). 'Misinformation and Motivated Reasoning: Responses to Economic News in a Politicized Environment.' Public Opinion Quarterly. Vol. 81, Issue 1, Pg. 86- 110. https://doi.org/10.1093/poq/nfw043

67 Lazer, D., Baum. M, Grinberg, N., Friedland, L., Joseph, K., Hobbs, W. and Mattsson, C. 99 (2017) 'Combating Fake News: An Agenda for Research and Action'. The Shorenstein Center, Harvard University. Accessed on 9th July 2021 from https://shorensteincenter.org/combating-fake-news-agenda-for-research/

Enhancing truth discernment through human reasoning underscores existing conceptions of misinformation resiliency. [68]

The relief work during the second wave resulted in a new typology of (mis)information that eludes these emerging conceptions of misinformation response. During the second wave, people had strong intentions and incentives to seek out accurate information. Leads about medical resources are concise units of information that don't rely on or trigger cultural, social or political beliefs. Despite deliberate reasoning, the 'truth status' of such information was not easy to discern. On the supply side, the information was hyperlocal, not created or propagated in coordination and high in volume. This information did not have to be viral to be harmful. Identifying inaccurate information was difficult, as was reducing its circulation—could platforms have relied on crowdsourced verification by volunteering groups to make decisions about takedowns of harmful leads? When any usable lead for medical aid could save lives, should platforms and volunteers err towards suppressing bad leads at the risk of reducing useful ones, or the other way round? Reducing circulation of harmful information on chat apps remains an open question, but one that becomes more challenging in presence of small-scale, but widespread, criminal activity on the platform.

This analysis presents a case study of a different kind of dis/mis-information than political and medical content that has received mainstream attention so far. This phenomenon observed in India during the second wave of Covid-19 in India, is one that could occur in any location or situation where the need for reliable actionable information is high but trusted and expected information channels fail. This could be in subsequent waves of the pandemic, natural disasters, cyberattacks or wars. In such contexts, the harm from disinformation is immediate and tangible. Accounting for these less frequent, but extreme situations can strengthen our conceptions and agendas for misinformation response.

68 Pennycook, G. Rand, D.G. 'The Psychology of Fake News'. Trends in Cognitive Sciences. Vol. 25, Issue 5, 2021, Pp. 388-402, ISSN 1364-6613. https://doi.org/10.1016/j.tics.2021.02.007

POSSIBLE FUTURE DIRECTIONS OF WORK

This analysis only scratches the surface, both in terms of the technical methods adopted as well as the thematic contributions that such a crisis makes to our understanding of misinformation.

In terms of methodology, we found vector embedding based image clustering to be a useful summarizing technique for image content. This technique is versatile and can also be used in real-time misinformation response. The centroid of different image clusters can help determine the category of a new image with some confidence. This can help prioritize a specific post for verification. For example, in the images collected from WhatsApp groups, all images that have religious imagery could be ignored but images that fall in the cluster of medical supplies can be prioritized for verification. A real-time visualization of incoming images can also help identify groups that are irrelevant to the analysis early on, so that researchers may leave the group and minimize personal data collected during the research.

A deeper analysis with greater manual intervention could provide more definitive evidence on some of the trends shared in the report. For example, manually annotating the text messages and media items could uncover more qualitative differences in the use of these modalities for communicating information. This analysis also excluded temporal analysis of media items or multi-modal (images and text shared together in one message) content. The limitations in WhatsApp's 'Export Chat with Media' feature restricted reliable temporal analysis of media items. There are other technical approaches such as rooting phone to decrypt

the database[69] that can circumvent the limitation in WhatsApp's 'Export Chat' feature. The export chat feature, however, is available to all WhatsApp users and doesn't need specialized hardware, making it a scalable and sustainable method for data collection and one that also allows for crowdsourcing of data. WhatsApp could look into making the 'Export Chat with Media' feature more robust, since that is the only mechanism for an average WhatsApp user to archive conversations from their WhatsApp chats.

Thematically, this report has tried to highlight a new facet of the Information Disorder— harms from inaccurate information during a crisis where everyone needs actionable information. This report has focused on the second wave of the Covid-19 crisis in India but similar information chaos could unfold in case of natural disaster as well as civil and international wars. How low quality and criminal information can be suppressed and actionable information amplified, during a calamity, in an era of spontaneous communication, remains a complex question that merits a lot more attention.

69 Garimella, Kiran, and Gareth Tyson. 'WhatApp Doc? A First Look at WhatsApp Public Group Data'. Proceedings of the Twelfth International Conference on Web and Social Media, ICWSM 2018, Stanford, California, USA (June 25-28, 2018). AAAI Press. 2018, pp. 511–17, https://aaai.org/ocs/index.php/ICWSM/ICWSM18/paper/view/17865

APPENDIX

Technical Tools and Methods

Image Similarity Matching and Clustering Algorithm

To be able to group images based on visual and semantic similarity, we use vector embeddings. Each image is represented as a 512-dimensional vector embedding. These are extracted from the image using a pretrained ResNet model. These vector representations of the images are indexed in Elasticsearch which helps us query for similar images. To visualize these high dimensional vectors, we use a technique called t-Distributed Stochastic Neighbor Embedding (t-SNE).[70] It is a dimensionality reduction technique that represents high dimensional data (such as images) on a 2D plane as seen in this report. The labelling of the clusters was done manually by observing the data within the cluster closely. We hoped to provide a broad label that would help readers understand the dataset better and encourage exploration of relevant subset of the data.

Link to Code: *https://github.com/tattle-made/data-experiments/blob/master/tSNE-clustering.ipynb.*

WhatsApp Export Chat Scraper

We wrote a Python based scraper for chats exported from WhatsApp, that does three things:

- Synchronizes multiple exports of the same chat
- Anonymizes group IDs and sender IDs

70 L.J.P. van der Maaten and G.E. Hinton. 'Visualizing High-Dimensional Data Using t-SNE'. Journal of Machine Learning Research 9(Nov):2579-2605, 2008.

- Pushes all the information scraped from the exported chat files to a consolidated database

The scraper at present is designed to read files uploaded to a Google drive folder that can be accessed through OAuth credentials. But it can also be used to read files from a local folder.

Link to the scraper code: *https://github.com/tattle-made/whatsapp-scraper/tree/master/python_scraper*

Textual Analysis

Language Translation: We used two free python libraries—deep_detector and google_trans_new—to translate the text in the images and messages. Both these libraries support translation through the Google translate API and have in-built auto language detection.

Word Frequency Analysis: Between 29th April 2021, when we joined the first sixteen groups, and 24th June 2021, when we stopped the data collection, we collected 16,694 messages. We retrospectively realized that messages for some days for the first week were missing.We discarded messages for the first week for temporal word frequency analysis, resulting in 13,524 textual messages spanning seven weeks starting 6th May 2021 and ending on 24th June 2021. There was no method to identify which images were shared in the first week. To keep the comparison between media items and text messages commensurate, we included all the images and text messages from the eight week period for other analysis (Trend 4 and Trend 5), we included all images in the media downloads folder in this analysis.

For word frequency analysis we undertook the following steps:

- Tokenized all the sentences.
- Lemmatized each of the words, that is converted words to their root form.

- Removed stop words such as 'the', 'will'. In addition to the stop words provided by the nltk corpus library in Python, we also added a few words specific to our data such as 'media' and 'omitted'. 'Media Omitted' is a phrase generated by the Export Chat Feature on WhatsApp when it fails to export a exported file, it gets caught in word frequency analysis but does not provide any insights about the data itself.

Link to code: *https://github.com/tattle-made/data-experiments/blob/master/whatsapp_groups_analysis/July8_2021_whatsApp_textmsg_analysis.ipynb*

Word Clouds: To generate word clouds, we used the WordCloud Python library that tokenizes the sentence and removes stop words, but does not lemmatize the tokens.

Link to code for text messages: *https://github.com/tattle-made/data-experiments/blob/master/whatsapp_groups_analysis/July8_2021_whatsApp_textmsg_analysis.ipynb*

Link to code for images: *https://github.com/tattle-made/data-experiments/blob/master/whatsapp_groups_analysis/image_text_analysis.ipynb*

Comparing Word Frequencies of Images and Text: To compare word frequencies of images and text, we calculate the proportion of occurrence of words in images and text messages. We then pull out the top 30 most frequent terms (measured by proportion of occurrence) and plot the percentage of occurrence of a word in both these corpus on a scatter plot. Since textual content typically follows Zipf's law, we used a log scale. We generally followed the methodology described by Silge and Robinson (2017).[71]

Link to Code: *https://github.com/tattle-made/data-experiments/blob/master/whatsapp_groups_analysis/image_text_word_frequency_comparison.ipynb*

71 Silge J., Robinson D., Section 1.5. Text Mining with R: A Tidy Approach. O Reilly. 2017. *https://www.tidytextmining.com/index.html*

TF-IDF across Weeks: Term Frequency - Inverse Document Frequency is a statistical measure that estimates how important a term is in a document, relative to other documents in a collection. Terms with high TF-IDF scores are the terms in a document that are *distinctively* frequent in a document, when that document is compared other documents.[72] In our analysis, the text messages in each week comprise a document, resulting in seven documents. TF-IDF, then finds the words that are salient in a particular week.

Link to code: *https://github.com/tattle-made/data-experiments/blob/master/whatsapp_groups_analysis/temporal_word_frequency_analysis.ipynb*

Anonymization

We undertook the following four steps to obfuscate personal details from the images shown in the online visualizations:

- We resized the image to a width of 120px while maintaining the aspect ratio.
- We blurred the images using the Graphics Magick library[73] which convolves the image with a Gaussian operator.
- We used blazeface model from tensorflow.js to detect faces in images and draw black rectangles over them. While it worked well for most cases, it missed out a few faces.
- We manually perused the images and drew rectangles over phone numbers and exported the image using GIMP, which could have led to further compression of images.

72 Lavin, M. J. (2019, May 13). 'Analyzing Documents with TF-IDF. Programming Historian'. *https://doi.org/10.46430/phen0082; https://programminghistorian.org/en/lessons/analyzing-documents-with-tfidf*

73 *https://www.npmjs.com/package/gm*

Limitations in Analysis

This analysis aims to provide some insight into the nature of conversations and procedures for information management in Covid-19 relief groups on WhatsApp. But the analysis comes with many caveats emerging from the underlying data as well as methods of analysis:

Data Limitations

The 21 WhatsApp groups we joined are a drop in the ocean of all WhatsApp groups. While providing some light into what relief work coordination on WhatsApp looked like, these groups are not representative of all relief groups that were operational during the second wave of the Covid-19 crisis. The textual analysis as well as image clusters could change if the selection of WhatsApp groups changed. Furthermore, the data is also made noisy with the inconsistency in WhatsApp's export chat feature—there is no clarity on why certain media items or intervals of data are dropped from exports. We can try to compensate for some of these irregularities by exporting chats more frequently, but this too does not guarantee exhaustive data collection. Since little is known about when data may be dropped by WhatsApp's export chat feature, it isn't clear if the data gaps introduced by the feature is systematic or random.

Extracting Text from Images: How Computer Vision Works

Google's Cloud Vision API or any general-purpose computer vision technology is fairly good at extracting text from images. These technologies can handle images with text in various fonts and orientation and, hence, are quite reliable for extracting images from the kind of memes we see on Indian social media. In our experience, they don't fare as well when extracting text from newspaper clippings with multiple columns, which usually contains distinct sections of texts. These tools often return one big blob of text with no notion of paragraphs and columns.

The text extracted might have coherent bits. But it might not be coherent altogether. One way to remedy this would be to process images detected to be news clippings using a different algorithm. Some preprocessing to these images could separate out different sections of the newspaper while grouping paragraphs or columns within those sections.


सुखद समाचार
लाख
हो जाएंगे आज
सूरत में कोरोना
से ठीक होने वाले
दैनिक भास्कर
अभी मास्क भी वैक्सीन है
कोरोना से रिकवर हुए 2000 से अधिक मरीजों का शुगर लेवल 300 से 400 तक बढ़ गया
शुगर की बीमारी दे रहा रेमडेसिविर
साइड इफेक्ट की जानकारी होने के बाद भी डॉक्टर मरीजों को दे रहे हैं रेमडेसिविर, जबकि हर मरीज को नहीं दिया जाता है यह इंजेक्शन
ये नियमित जांच कराएं
रेमडेसिविर और स्टेरॉयड का कॉम्बिनेशन देता है समस्या
पोस्ट कोविड मरीजों का हो नियमित बॉडी चेकअप
दूसरी लहर में रेमेडेसिविर की डिमांड ज्यादा आ रही

रेमडेसिविर और स्टेरॉयड का पोस्ट कोविड मरीजों का हो दूसरी लहर में रेमेंडेसिविर
कॉम्बिनेशन देता है समस्या निरयमित ॉडी चैकअप
की डिमांड ज्यादा आ रही
t , ntat
witr s est wt
fwitt nepkde
heter eer fee
R r mter fettt
TRI 400
we
t t tet
vohe
fist ufinetatet
de
35O u r
emdesiv s
th fer st deY

Figure 24: 'Result of Google Cloud Vision API on a Newspaper Clipping with Multiple Columns'

Mixed Code Language

People use multiple languages within the same message. They may also type words from Indian languages in the Roman script. Take this snippet from a text message shared on one of the groups:

My father [redacted] died yesterday morning i.c. 07-05-2021. He was admitted in [redacted] Bareilly And i am sure that something fishy happened due to which my father dicd. Kal tak recovery thi... mai unhe discharge krane k liye baat kr ri t... he was continuously saying ki yaha sab mile hue hain... inko paise milte hain... every night they kill people. He wrote in a paper please arrange for oxygen cylinder or else THE END.

We detect for non-English messages by checking if the text contains any characters not in English. The above text, while containing some messages in Hindi, uses the English (Roman). Thus, this message is not detected as a non-English message that needs to be tested. Some experimentation with Google Translate showed that it too detects these messages as English language, and does not selectively translate the non-English sentences.

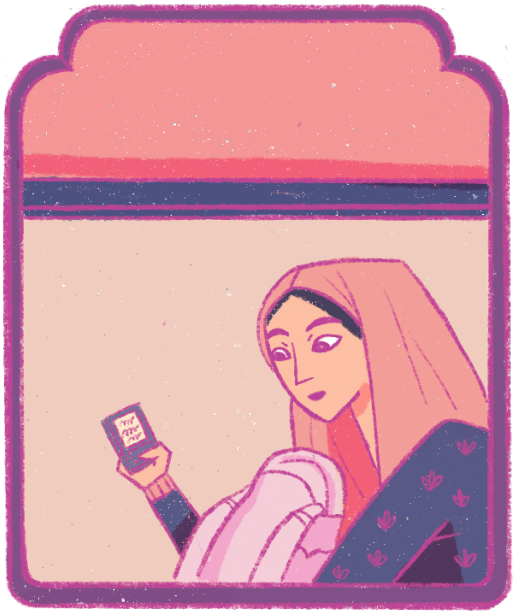
saying ki yaha sab mile hue hain... inko paise milte hain...	Saying everyone is found here... they get money...

Figure 25: Result of Erroneous Translation of a sentence by the Translate API

This issue can be managed by breaking down every message into individual sentences and attempting a language detection and translation on each sentence individually. But people may use multiple languages within the same sentence and some sentences may still escape accurate language detection. Finally, even when the language is detected correctly, the automated translation is not always accurate.

The Use of Machine Learning in Anonymization

We used blazeface model from tensorflow.js to detect faces in images and draw black rectangles over them. While it worked well in most cases, it missed out a few faces which we had to manually peruse and anonymize. If one was analyzing an image dataset within a trusted closed environment, this would not be a problem; but it prevents the automation of creating a public report like this while adequately protecting privacy of people featured in the images.



Tattle
www.tattle.co.in
@tattlemade