

# News and Information over Facebook and WhatsApp during the Indian Election Campaign

[COMPROP](#) DATA MEMO 2019.2 / MAY 13, 2019

[comprop@oii.ox.ac.uk](mailto:comprop@oii.ox.ac.uk)

Vidya Narayanan  
Oxford University

Bence Kollanyi  
Oxford University

Ruchi Hajela  
Oxford University

Ankita Barthwal  
Oxford University

Nahema Marchal  
Oxford University

Philip N. Howard  
Oxford University

## ABSTRACT

*Social media platforms have become an important source of political news and information for voters in India's national election. To evaluate the quality of sources and images being shared, we examine the patterns of content circulation on Facebook and WhatsApp with a large sample of data collected over a two month period in advance of the elections. We find that (1) more than a quarter of the content shared by the Bharatiya Janata Party (BJP) and a fifth of the content shared by the Indian National Congress (INC) is junk news, while the Samajwadi and Bahujan Samaj Party (SP-BSP) shares very little sensational, extremist, or conspiratorial content. (2) For visual content being shared in our sample of WhatsApp groups, a third of the BJP's images, a quarter of the INC's images, and a tenth the SP-BSP's images were catalogued as divisive and conspiratorial. Comparing the platforms, we find that (3) misinformation on WhatsApp primarily takes the form of visual content, while misinformation on Facebook involves links to sensational, extremist, and conspiratorial news sites and visual content. On a positive note, (4) we observed very limited amounts hate speech, gore or pornography in either platform samples. Yet in comparison with other recent international elections, (5) the proportion of polarizing political news and information in circulation over social media in India is worse than all of the other country case studies we have analysed, except the US Presidential election in 2016.*

## INTRODUCTION

India has frequently been described as the world's largest democracy. With an estimated 900 million voters, the 2019 Indian General Election is the biggest that the world has ever witnessed. As the nation votes to elect the members of the parliament, we examine various sources of the news and information including visual media, circulating on two of the country's most popular social media platforms — Facebook and WhatsApp — for a two-month period preceding the elections. Facebook has an estimated 241 million users in India and WhatsApp over 200 million users.<sup>1</sup> Researchers have studied the growing popularity of these platforms and the ways in which politicians have leveraged the affordances of social media to directly reach millions of voters across the country.<sup>2</sup> According to recent surveys on digital news in India, news consumption on online platforms and specifically on social media has outstripped traditional print media.<sup>3</sup> In the same study authors note that Facebook and WhatsApp are widely used in the country, with 52% of the digital survey respondents accessing news

through these platforms. In this environment, it is particularly important to study the quality of news and information circulating on these platforms as Indian voters get ready to exercise their franchise. India has a large number of national and regional parties with the ruling Bharatiya Janata Party (BJP), the main opposition party Indian National Congress (INC), the Samajwadi Party (SP) and the Bahujan Samaj Party (BSP), being counted among the key players. Traditionally, the BSP and SP have been political opponents, but have formed a coalition to contest the 2019 General Election.

This report focuses on the use of social media by these three main party groups. Elections are being conducted in 7 phases across the country, starting on April 11 and continuing until May 19. Our research questions are: (1) How were politically relevant public pages on Facebook being used to for political engagement over a two-month period leading up to the elections? (2) How were politically relevant public WhatsApp groups being used to for political communication? (3) What were the types of news sources and images that were circulating on

these platforms during this time? We provide a detailed description of our data sampling and classification methods used to address these research questions in the section below.

### **SAMPLING AND METHODS**

Our data collection included a two-month period between 14 February – 10 April, 2019. This period coincided with the occurrence of key events in the country including the Pulwama attack, India's airstrike against Pakistan in retaliation, the capture and subsequent release of the Indian air force pilot by Pakistan, and the announcement of the election dates followed by the campaigning period. During this period, we analysed posts from public Facebook pages and public WhatsApp groups, sharing an affinity with the four major parties — BJP, INC, SP and the BSP.

### **FACEBOOK SAMPLING AND METHODS**

To identify Facebook pages that were relevant in the context of the Indian election, we relied on two sources: Twitter and Facebook. We searched Twitter, for 30 hashtags that were identified as politically relevant in the Indian context, by a team of researchers (please see the online supplement for the complete list of hashtags). We extracted links to webpages of political parties, newspaper organisations, prominent celebrities and public figures from the tweets posted with these hashtags and found their corresponding public pages on Facebook. Then we used the Facebook's search utility to identify official party pages of the four main political parties relevant to this study. This formed our core set of public Facebook pages. We then used Facebook's Graph API to map the network of public pages liked by the core set of public pages. In the last step, our team of researchers selected the relevant Facebook public pages based on the metadata — name, description, category — we collected about the pages. Using this process, from our sample of pages, we filtered out pages that were unrelated to the Indian political landscape, retaining only those pages that were clearly affiliated with the political parties our research focuses on, and were active in the first months of the year (please see the online supplement for the complete list of pages).

Our Facebook data sample consisted of 130 public FB pages of which there were 58 BJP pages, 23 INC pages and 49 SP-BSP pages. From these pages, we extracted all posts between 14 February – 10 April and analysed links to news sources and images. From the posts, we extracted all URLs and images. The

data collection included 27,771 posts from the 130 public pages. Within these posts, 16,604 posts contained a photo or an image and another 1,388 contained a link pointing to a news article, a video posted outside of Facebook, or any other type of web content. From this data set we analysed a random sample of 200 unique links and 200 images from each of three political groups, and categorised them according to our typology.

### **WHATSAPP SAMPLING AND METHODS**

We adopted a systematic approach to search for links to public WhatsApp groups sharing an affinity with the four political parties under consideration. We used the Google search engine with search terms that included the names of the parties or prominent leaders from these 4 parties to identify politically relevant WhatsApp groups. We joined party groups through links on the internet indexed by the search engine (please see the online supplement for the complete list of search terms). Using these publicly available invite links, we joined WhatsApp groups dedicated to discussing politics, news and current affairs. They constitute what we refer to as 'public' WhatsApp groups in our study. Using this process, our team members joined more than 200 public WhatsApp groups of which nearly 131 groups were supportive of the BJP, 30 of INC, and 67 of the SP-BSP alliance. Upon joining these groups, we declared our presence as researchers before offering group members the option to withdraw consent from participating in this study. This resulted in some administrators removing us from a total of 112 groups (78 BJP, 8 INC and 26 SP-BSP). We used no personal data (names, phone numbers, location, demographic profile) to inform our research findings, instead using only URLs to news and information portals and visual media — images and memes which could be considered to be data belonging to the public domain, circulating within these groups for this study. New groups across political parties also emerged during our data collection period. By the end of our data collection period, we had access to a total of 116 public WhatsApp groups across different parties. While it is not clear if these groups are managed by the online communication and campaign managers of the parties, the content shared indicated that these groups could be run by supporters and volunteers sharing a strong affinity with the political parties under consideration. Within groups, there was great variation in group size with membership ranging from less than 50 to

257 (even as the maximum capacity of members in a WhatsApp group is 256). The average number of members for groups broadly supportive of the BJP was 102, 129 for the INC and 79 for SP-BSP. Groups supportive of the BJP were more active in sharing images in comparison to other groups. It should be noted that we saw a wide distribution in the sharing patterns of the groups with some being more active than others and certain days registering significantly more activity than others. For instance, on March 11, the day following announcement of the elections, nearly 3000 images were shared by pro-BJP groups. To classify sources of news and political information extracted from both Facebook public pages and WhatsApp groups, we hand coded the base URLs, using a rigorous and iterative process developed and refined through the project's previous studies of six elections in five Western democracies and several countries in Latin America.<sup>4,5,6,7</sup>

A team of three coders identified sources of junk news and information, based on a rigorous grounded typology. Sources of junk news deliberately publish misleading, deceptive or incorrect information purporting to be real news about politics, economics or culture. This content includes various forms of extremist, sensationalist, conspiratorial, masked commentary, fake news and other forms of junk news. The typology explaining our content classification is as follows:

## JUNK NEWS TYPOLOGY

### Professional News Content

- Major News Brands. This is political news and information by major newspapers, broadcasting or radio outlets, as well as news agencies.
- Local News. This content comes from local and regional newspapers, broadcasting and radio outlets, or local affiliates of major news brands.
- New Media and Start-ups. This content comes from new media and digitally native publishers, news brands and start-ups.
- Tabloids. This news reporting focuses on sex, crime, astrology and celebrities, and includes yellow press publications.

### Professional Political Content

- Government. These links are to websites of branches of government or public agencies.
- Experts. This content takes the form of white papers, policy papers or scholarship from researchers based at universities, think tanks or other research organizations.
- Political Party or Candidate. These links are to official content produced by a political party or candidate campaign, as well as the parties' political committees.

### Divisive and Conspiratorial Content

- Junk News and Information. These sources deliberately publish misleading, deceptive or incorrect information purporting to be real news about politics, economics or culture. This content includes various forms of propaganda and ideologically extreme, hyper-partisan or conspiratorial news and information. To be classified as Junk News and Information, the source **must fulfil at least three of these five criteria:**
  - *Professionalism:* These outlets do not employ standards and best practices of professional journalism. They refrain from providing clear information about real authors, editors, publishers and owners. They lack transparency and accountability, and do not publish corrections on debunked information.
  - *Style:* These outlets use emotionally driven language with emotive expressions, hyperbole, ad hominem attacks, misleading headlines, excessive capitalization, unsafe generalizations and logical fallacies, moving images, and lots of pictures and mobilizing memes.
  - *Credibility:* These outlets rely on false information and conspiracy theories, which they often employ strategically. They report without consulting multiple sources and do not fact-check. Sources are often untrustworthy and standards of production lack reliability.
  - *Bias:* Reporting in these outlets is highly biased, ideologically skewed or hyper-partisan, and news reporting frequently includes strongly opinionated commentary.
  - *Counterfeit:* These sources mimic established news reporting. They counterfeit fonts, branding and stylistic content strategies. Commentary and junk content are stylistically disguised as news, with references to news agencies and credible sources, and headlines written in a news tone with date, time and location stamps.

### Other Political News and Information

- Citizen, Civil Society and Civic Content. These are links to content produced by

independent citizen, civic groups, civil society organizations, watchdog organizations, fact-checkers, interest groups and lobby groups representing specific political interests or agendas. This includes blogs and websites dedicated to citizen journalism, personal activism, and other forms of civic expression that display originality and creation that goes beyond curation or aggregation. This category includes Medium, Blogger and WordPress, unless a specific source hosted on either of these pages can be identified.

- Political Humor & Entertainment. This category includes political jokes, sketch, comedy or entertainment-focused coverage, as well as political talk shows and late-night formats. Despite their humorous and entertaining nature, these formats often serve as central sources of news and information.
- Video/Image Sharing & Content Subscriptions. Includes music streaming portals like Spotify, video streaming services and live streaming, political documentary movies, e-books and audio book subscriptions, as well as image sharing services.
- Fundraising and Petitions. Encompasses civil society fundraising and petition pages, as well as surveying services for various political causes and interests that do not represent an official campaign or candidate.
- Lifestyle & Special Interest. Includes lifestyle and special interest publications like women's and men's magazines, and content focused on art and fashion, fitness, food and wellness, nature and tourism, or hunting, fishing and automobiles.
- Religion. Refers to content with distinctly religious themes and faith-based editorializing presented as political news or information.
- Online Portals, Search Engines and Aggregators. Includes online portals like AOL, Yahoo! and MSN that do not themselves have editorial policies and have no or limited original news content. This category also includes links to Wikipedia.
- Cloud. Encompasses services such as Amazon Web Services, Google Drive

and Docs, OneDrive, or archiving services in the cloud.

- Other Political. Refers to content that is political in nature but does not fit any of the other categories, for example services where voters are able to check their polling stations or purchase political merchandise.

#### Other

- Social Media Platforms. These are links that refer to other social media platforms as well as official developer tools. If the content at the ultimate destination can be attributed to another source, it is.
- Not Available. This includes links that are no longer available or have not been successfully archived after repeated attempts, as well as sources that are redirected to other sources and whose original content is unknown.
- Shopping, Services and Applications. Encompasses links to auction websites or sales platforms, such as eBay and Amazon, including software-as-a-service applications, analytics tools and content optimization and monetization tools. This also includes applications and browser extensions.
- Link Shorteners. Includes links like Bitly or Vitweet, when it is not possible to unwrap the original URL. If the source is successfully unwrapped from the link shortener, it is coded in the appropriate category.
- Other Non-Political. Refers to sites that have no political content such as spam, gambling and brand advertising.
- Language. Content from sources in languages that are not English, French, German, Spanish, Portuguese, Hungarian or Mandarin are not labeled, unless verifiable information about a source is accessible.

Further to analysing links to news sources, we also analysed a random sample of 200 images from public Facebook pages and each collection of WhatsApp groups affiliated with the three main political parties, between February 14 and April 10. We classified these images accordingly to a visual typology that we first developed for studying images extracted from Brazilian WhatsApp groups and refined based on analysis of images circulating in Indian social media

platforms. We also introduced two new categories (Nationalism and Support for Armed Forces and Policy Critique) for classifying images based on the content that was shared within groups. The typology for classification of images is as follows:

- Campaign and Support Material. Broadly supportive of the candidate, using verifiable facts and figures, pertaining to social, economic and political issues.
- Divisive and Conspiratorial Content
  - Style: Content uses emotive expressions, hyperbole, misleading information, unsafe generalizations and logical fallacies.
  - Credibility: Content promotes false information and conspiracy theories, which is often employed strategically.
  - Bias: Content is highly biased, ideologically skewed or hyper-partisan. Promotes inflammatory viewpoints.
  - Discrediting tactics: Content discrediting opponents, by indulging in smear campaigns, ad hominem attacks, falsehoods about personal lives, attacks using religious themes/iconography.
- Religion. Pertaining to religious practices, references to religious texts and values, claiming alignment with core religious values from political leaders, claiming support of prominent religious leaders for the campaign.
- Celebrities. Prominent social figures in the arts, TV stars, sports personalities, senior officials in law enforcement and other government institutions voicing strong support for candidates.
- Hate, Gore and Porn. Hate speech or content directed against individuals, minority groups like women, LGBTQ community, against ideologies, as well as videos depicting acts of extreme violence or obscenities.
- Satire. Images that uses humor, irony or exaggeration to comment on political and social issues.
- Policy Critique. Images critical of opponent's policies, manifesto, economic achievements.
- Nationalism and Support for Armed Forces. Images glorifying sacrifices of the armed forces, promoting respect for soldiers, prominent politicians paying respects to slain army personnel.
- Other. Lifestyle, goodwill messages and content that doesn't fit the above categories.

## FINDINGS AND ANALYSIS

Covering nearly two months of data our sample provides an overview of the political communication circulating on these two popular platforms until the day before the elections began.

### Facebook Findings and Analysis

Our classification of Facebook links and images is presented in Tables 1 & 2.

**Table 1: Facebook News Share Table**

Type of Source	BJP %	INC %	SP - BSP %
<b>Professional News Sources</b>			
Major News Brands	34.0	22.0	48.0
Local News & Tabloids	1.0	2.0	1.0
New Media & Start-ups	5.5	9.0	10.5
<b>Professional Political Sources</b>			
Political Party and Government	5.0	7.0	0.0
<b>Divisive and Conspiratorial content</b>			
Junk News & Information	28.0	21.0	1.0
<b>Other Political News &amp; Information</b>			
Video/Image Sharing	10.5	1.0	1.5
Portals, Search & Aggregators	6.0	12.5	19.5
Other Political	2.0	7.0	7.0
<b>Other</b>			
Social Media Platforms	6.0	17.5	11.0
Other Non-Political	1.5	0.5	0.5
	100.0	100.0	100.0
<i>Source: Authors' calculations from Feb 14 – Apr 10, 2019.</i>			
Note: Some categories from the typology were collapsed due to insufficient representation.			

We note that from Table 1, in our random sample, 40% of the links shared by the BJP pages were classified as Professional News sources, while 28% of links were classified as Junk News. Similarly, 33% of the links shared by the INC were classified as Professional News whereas 21% of links shared were coded as Junk News. On

Facebook, the SP-BSP pages had a much higher percentage of Professional News sources shared at 59.5% with only 1% of sources shared classified as Junk News.

We further classified 200 images extracted from public pages belonging to each party group, according to our visual typology. From Table 2, we note that 67% of images extracted from the BJP pages have been classified as Campaign & Support, whereas only 12.5% have been classified as Junk. Similarly, the INC had 52.5% of Campaign & Support images and 14% of Junk images. The SP-BSP had 60.5% of Campaign & Support images and 12.5% of Divisive and Conspiratorial images. It is significant to note that in our random sample of Facebook images, we found no images that could be classified as Hate, Gore or Porn.

**Table 2: Facebook Image Classification**

Type of Content	BJP %	INC %	SP - BSP %
Campaign & Support	67.0	52.5	60.5
Divisive & Conspiratorial	12.5	14.0	12.5
Policy Critique	2.0	7.0	7.5
Nationalism & Support for the army	2.0	4.0	2.5
Religion	1.0	0.5	1.5
Satire	0.5	3.0	5.0
Hate, Gore, Porn	0.0	0.0	0.0
Other	15.0	19.0	10.5
	100.0	100.0	100.0

*Source: Author's calculations from Feb 14 – Apr 10, 2019*

*Note: Some categories from the typology were collapsed due to insufficient representation.*

### WhatsApp Findings and Analysis

We extracted a total of number links and 32,292 images from over 100 public WhatsApp groups. We analysed a random sample of 200 links and 200 images from each collection of groups affiliated with the 3 party groups, BJP, INC and the SP-BSP. The links extracted from these WhatsApp groups pointed largely to other SM media platforms, video sharing platforms like YouTube and shopping services and other non-political content. Links to relevant political content

were not widely shared among the WhatsApp groups in our data sample. Therefore, we focused our analysis on images that were shared within these groups. Our detailed analysis of WhatsApp images is presented in Table 3. From Table 3, we note that a large percentage (34.5%) of images shared within BJP affiliated groups were classified as Divisive and Conspiratorial, followed by the INC at 28.5%.

**Table 3: WhatsApp Image Classification**

Type of Content	BJP %	INC %	SP - BSP %
Divisive & Conspiratorial	34.5	28.5	11.5
Campaign & Support	18.0	30.0	20.5
Nationalism & Support for the army	10.5	2.0	7.5
Religion	3.5	2.5	2.0
Satire	3.5	9.0	4.0
Policy Critique	1.5	4.5	7.5
Hate, Gore, Porn	0.5	0.0	1.0
Other	28.5	23.5	46.0
	100.0	100.0	100.0

*Source: Author's calculations from Feb 14 – Apr 10, 2019*

*Note: Some categories from the typology were collapsed due to insufficient representation.*

### CONCLUSIONS

We analysed over 130 public Facebook pages and 116 public WhatsApp groups sharing an affinity with key political parties in India, the BJP, INC and the SP & BSP. Our main conclusions are We find that (1) more than a quarter of the content shared by the Bharatiya Janata Party (BJP) and a fifth of the content shared by the Indian National Congress (INC) is junk news, while the Samajwadi and Bahujan Samaj Party (SP-BSP) shares very little sensational, extremist, or conspiratorial content. (2) For visual content being shared in our sample of WhatsApp groups, a third of the BJP's images, a quarter of the INC's images, and a tenth the SP-BSP's images were catalogued as divisive and conspiratorial. Comparing the platforms, we find that (3) misinformation on WhatsApp primarily takes the form of visual content, while misinformation on

Facebook involves links to sensational, extremist, and conspiratorial news sites and visual content. On a positive note, (4) we observed very limited amounts hate speech, gore or pornography in either platform samples. Yet in comparison with other recent international elections, (5) the proportion of polarizing political news and information in circulation over social media in India is worse than all of the other country case studies we have analysed, except the US Presidential election in 2016. Our results are based on a small sample of WhatsApp groups. However, the encrypted nature of the platform, amorphous structure of public groups and our strict ethical considerations pose significant challenges to joining and extracting data at scale from WhatsApp groups. We note that, our strategy does not ensure that all WhatsApp groups are adequately represented, however forwarded content circulating within these groups could readily be shared in other private groups that the members belong to and therefore it is possible that the content analysed in this study has been viewed by a much larger network of WhatsApp users.

## **ETHICAL CONSIDERATIONS**

We adhered to strict professional and ethical guidelines. This included safe data storage practices, concerted attempts to include a diverse range of participants, and open explanation of project methods and goals. Our researchers announced their presence and research intentions upon joining the group and agreed not to interact in conversation or collect data on specific users of messages. Chat logs were accessed, downloaded, and pseudonymized before our research staff could access them for analysis. We only undertook quantitative data analysis of these chat logs, extracted links to news sources, classified other content like memes and jokes, and did not use personal quotes or opinions sourced from these transcripts, in our research. Due to the large volume of messages and the dynamic composition of group members, it was not possible to seek consent from new members who might have joined the group after the initial announcement of our presence. We have received ethics approval for this research from the European Research Council Ethics Board and the University of Oxford Research Ethics Committee.

## **ABOUT THE PROJECT**

The Project on Computational Propaganda (COMPROP) based at the Oxford Internet Institute, University of Oxford, is an interdisciplinary team of social and information scientists researching how political actors manipulate public opinion over social networks. This work includes analyzing the interaction of algorithms, automation, politics, and social media to amplify or repress political content, disinformation, hate speech and junk news. Data memos are designed to present quick snapshots of analysis on current events in a short format, and although they reflect methodological experience and considered analysis, they have not been peer-reviewed. Working papers present deeper analysis and extended arguments that have been collegially reviewed and engage with public issues. COMPROP's articles, book chapters and books are significant manuscripts that have been through peer review and formally published.

## **ACKNOWLEDGMENTS**

The authors gratefully acknowledge the support of the European Research Council, "Computational Propaganda: Investigating the Impact of Algorithms and Bots on Political Discourse in Europe," Proposal 648311, 2015-2020, Philip N. Howard, Principal Investigator. Project activities were approved by the University of Oxford's Research Ethics Committee, CUREC OII C1A 15-044. Our WhatsApp research is specifically covered by the OII DREC "Computational propaganda on chat apps: investigating impact of misinformation and automation on political discourse over WhatsApp", CUREC SSH OII C1A 18 074. For supporting our Election Observatory and our research in Europe we are grateful to Ford Foundation and the Omidyar Network. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the University of Oxford or our funders.

## **REFERENCES**

- [1] PTI, "India now has highest number of Facebook users, beats US: Report," <https://www.livemint.com>, 14-Jul-2017. [Online]. Available: <https://www.livemint.com/Consumer/CyEKd altF64YycZsU72oEK/Indians-largest->

- audience-country-for-Facebook-Report.html.  
[Accessed: 27-Apr-2019].
- [2] R. Schroeder, *Social Theory after the Internet*. London: UCL Press, 2018.
- [3] Z. Aneez, T. Neyazi, A. Kalogeropoulos, and R. K. Neilsen, "India Digital News Report, Reuters Institute for the Study of Journalism," Oxford, UK, 2019.
- [4] V. Narayanan, V. Barash, J. Kelly, B. Kollanyi, L. M. Neudert, and P. N. Howard, "Polarization, Partisanship and Junk News Consumption over Social Media in the US," *Data Memo 2018.1*, 2018
- [5] N. Marchal, L.M. Neudert, B. Kollanyi, and Howard, Philip N., "Polarization, Partisanship and Junk News Consumption on Social Media During the 2018 US Midterm Elections," *Data Memo 2018.5*, 2018.
- [6] M. Glowacki *et al.*, 'News and Political Information Consumption in Mexico: Mapping the 2018 Mexican Presidential Election on Twitter and Facebook.' *Data Memo 2018.2*. Oxford, UK: Project on Computational Propaganda., Oxford University, Oxford, UK, 2018.2, 2018.
- [7] C. Machado *et al.*, "News and Political Information Consumption in Brazil: Mapping the First Round of the 2018 Brazilian Presidential Election on Twitter," *Data Memo 2018.4*, 2018.