



HARVARD Kennedy School

MOSSAVAR-RAHMANI CENTER
for Business and Government

Media Manipulation in the Indian Context: An Analysis of Kashmir-Related Discourse on Twitter

Sanjana Rajgarhia

June 2020

M-RCBG Associate Working Paper Series | No. 147

The views expressed in the M-RCBG Associate Working Paper Series are those of the author(s) and do not necessarily reflect those of the Mossavar-Rahmani Center for Business & Government or of Harvard University. The papers in this series have not undergone formal review and approval; they are presented to elicit feedback and to encourage debate on important public policy challenges. Copyright belongs to the author(s). Papers may be downloaded for personal use only.



Media Manipulation in the Indian Context

An Analysis of Kashmir-Related Discourse on Twitter

Sanjana Rajgarhia

Master's in Public Policy Candidate
Harvard Kennedy School
Expected Graduation Date: 28 May 2020

Policy Analysis Exercise

Client: Amnesty International India
Advisors: James Waldo, John Haigh
Submitted: 7 April 2020



HARVARD Kennedy School
JOHN F. KENNEDY SCHOOL OF GOVERNMENT

This policy analysis exercise (PAE) reflects the views of the author(s) and should not be viewed as representing the views of the PAE's external client(s), nor those of Harvard University or any of its faculty. It has been submitted in partial fulfillment of the requirements for the degree of Master in Public Policy.

TABLE OF CONTENTS

LIST OF FIGURES	3
LIST OF TABLES	4
LIST OF ABBREVIATIONS	4
ACKNOWLEDGEMENTS	5
EXECUTIVE SUMMARY.....	6
1 THE PROBLEM.....	8
2 THE RESEARCH QUESTION	8
3 BACKGROUND	9
3.1 KASHMIR AND THE ABROGATION OF ARTICLE 370	9
3.2 MEDIA MANIPULATION	10
3.3 TWITTER’S RESPONSE TO MEDIA MANIPULATION	11
3.4 TWITTER IN INDIA	14
4 CASE STUDIES.....	15
4.1 CASE #1: DIVISION AND POLARIZATION	15
4.2 CASE #2: PRO-GOVT. PROPAGANDA “EVERYTHING IS NORMAL”	15
5 METHODS.....	17
5.1 DATA COLLECTION VIA THE TWITTER API	17
5.2 “INFLUENCE” CALCULATIONS	17
5.3 USER CATEGORIZATION	17
5.4 ADDRESSING THE RESEARCH QUESTIONS	17
6 RESULTS.....	20
6.1 EVIDENCE OF MEDIA MANIPULATION	20
6.2 THE ROLE OF THE MEDIA.....	31
6.3 REPERCUSSIONS OF MANIPULATIVE BEHAVIORS ONLINE.....	40
7 RECOMMENDATIONS.....	42
7.1 ADOPT APPROPRIATE LANGUAGE	42
7.2 SUPPORT THE CREATION OF MEDIA EVALUATION TOOLS.....	42
7.3 TRAIN JOURNALISTS	43
7.4 ADVOCATE FOR SOCIAL MEDIA PLATFORMS TO BE MORE TRANSPARENT	43
8 LIMITATIONS	44
9 SUGGESTIONS FOR FUTURE RESEARCH	45
9.1 ESTABLISHING CAUSALITY AND IMPACT	45
9.2 DELVING DEEP INTO SPECIFIC AREAS OF RESEARCH	45
9.3 BROADENING THE SCOPE OF THIS RESEARCH	46
10 APPENDICES.....	47
10.1 RATIONALE AND LIMITATIONS ASSOCIATED WITH THE DATA FILTERING CRITERIA.....	47
10.2 HASHTAGS AND USER MENTIONS USED FOR DATA FILTERING.....	48
10.3 NOTES ON INFLUENCE ACROSS CONVERSATIONS AND USER CATEGORIES.....	51
11 REFERENCES.....	53

LIST OF FIGURES

Figure 1: (Top) Map of Kashmir (Territory Split among India, Pakistan, China) with a Single State of Indian-Administered Kashmir Before 5 August 2019 (Source: BBC); (Bottom) Map Released by Indian Sources After 5 Aug 2019 Depicting the Split of the J&K State into J&K, and Ladakh (Source: Times of India).....	9
Figure 2: Decision-Making Matrix Provided by Twitter with regards to its Fabricated Media Policy (Source: Twitter).....	13
Figure 3: Likely Reward Follow by @NarendraModi (Source: Twitter)	23
Figure 4: Some Users are Followed by Less Prominent BJP Politicians; Some Also Boast of Being Blocked by Famous Individuals Who Do Not Support the BJP (Source: Twitter)	24
Figure 5: A Twitter User Who Claims to be Followed by Several BJP Politicians is, in Reality, Not Followed by Them. (Source: Twitter).....	24
Figure 6: Examples of Tweets Shared by Amit Shah Between 19 and 23 November 2019 i.e. The Days Around his Statements on Kashmir Normalcy (Source: Twitter).....	25
Figure 7: Examples of Users with Indian Flags in Their Names (Source: Twitter)	27
Figure 8: Examples of User Descriptions with Similar Language (“Blessed to be Followed by”) and Claims of Follows by @NarendraModi (Source: FollowerWonk).....	28
Figure 9: Positive Messaging around India Shared by @sushtany (Source: Twitter)	29
Figure 10: Screenshot from a Fact-Checking Article on the Disinformation Campaign against Shehla Rashid (Source: AltNews).....	30
Figure 11: Conversation between Sunanda Vashisht and Another User with regards to Kashmiri Hindus and Muslims (left); Conversation Deteriorates into Communal Issues (right) and Diverts from the Political Actions in Kashmir. (Source: Twitter)	30
Figure 12: Hyper-Partisan Content with Poor Editorial Oversight on IndSAMACHAR (Source: IndSAMACHAR).....	32
Figure 13: Examples of Poor-Quality News and Attempts of Polarization from Jan Ki Baat founder (Source: Twitter, LinkedIn).....	32
Figure 14: Unverified Content Posted by a Person Associated with FactHunt (Source: Twitter).....	32
Figure 15: A Comparison of Kashmir-Related News from Different Mainstream Indian Sources on 20 November 2019 (Source: Twitter).....	34
Figure 16: Coverage of the "Normalcy" Narrative by Mainstream Indian Media Sources (Analysis of Tweet Content Only) (Source: Own Analysis)	35
Figure 17: Coverage of the "Normalcy" Narrative by Media Sources Based on Geography (Analysis of Tweet Content Only) (Source: Own Analysis)	35
Figure 18: Mainstream Indian Media Sources and Their Susceptibility to Incentives (Analysis of Tweet Content Only) (Source: Own Analysis).....	37
Figure 19: Conversation between Journalist @sameerakhan and "Other" User @sushtany (Source: Twitter).....	38
Figure 20: Ways in Which Mainstream Media Uses Polarizing Terms in their Coverage (Analysis of Tweet Content Only) (Source: Own Analysis)	39

Figure 21: Examples from 2018 Where Mainstream Media Sources Used Polarizing Terms with the Intention of Reporting or Mocking Their Message. (Source: NDTV, The Print).....39

Figure 22: Example of an Account Mimicking a Suspended Account Being Operated Freely Online (Source: Twitter).....41

Figure 23: Examples of Active Accounts Making Explicit Promises of Reciprocal Follows in Their Descriptions (Source: Twitter)41

LIST OF TABLES

Table 1: Types of Violations as Listed in Twitter's Platform Manipulation Policy (Source: Twitter).....12

Table 2: Penalties for Violating Twitter's Platform Manipulation Policies (Source: Twitter)13

Table 3: A Sample of Common Phrases used on Twitter in India for Polarization Purposes (Relevant for Case #1)16

Table 4: Data Filtering Criteria and Statistics for Resultant Datasets.....18

Table 5: User Categorization19

Table 6: Methods to Address Research Questions19

Table 7: Top 1% of Influencers in Both Case Studies (Source: Own Analysis)21

Table 8: Levels of Influence Including and Excluding "Super Influencers" (Source: Own Analysis)22

Table 9: User Breakdown by Category for Both Cases (Source: Own Analysis)22

Table 10: List of Most Popular Hashtags From Case #1: Polarization (Source: Own Analysis)26

Table 11: Categories for Responses from the Media to Amit Shah's Normalcy Narrative (Used for Tweet Content Only).....33

Table 12: Descriptions of Media Incentives and Some Examples from Indian Media.....37

Table 13: Ways in Which Mainstream Media Sources Use Polarizing Terms39

Table 14: Data Collection Criteria.....47

Table 15: Hashtags Used for Data Collection48

Table 16: User Mentions Used for Data Collection.....49

Table 17: Influence Rankings by Category for Case #1: Polarization (Source: Own Analysis).....52

Table 18: Influence Rankings by Category for Case #2: Pro-Govt. Propaganda (Source: Own Analysis)52

LIST OF ABBREVIATIONS

J&K	Jammu and Kashmir
BJP	Bharatiya Janata Party
DM	Direct message

ACKNOWLEDGEMENTS

I would like to thank my advisor, James Waldo and my seminar leader, John Haigh, for their support throughout my research. Professor Waldo spent a significant amount of time helping me collect data, listened patiently to my challenges, and supported me through the iterative process of developing my methodology. Professor Haigh encouraged me to think more practically about my analysis and to identify the insights behind the numbers.

I would also like to thank my family and friends (too many to name). They helped me break down this complex problem, read numerous drafts of this report and listened patiently while I thought out loud about this project over the last several months.

EXECUTIVE SUMMARY

Access to high-quality information and an ability for citizens to come together to discuss and deliberate are integral elements of a democratic society.

In the Indian context, access to news has improved dramatically in the recent past (thanks to the rapid growth in internet services). However, there is widespread concern around the quality of online news (Reuters Study 2019₁).

One potential driver behind the deterioration of online news quality and political discourse is media manipulation. According to Data & Society, media manipulation is the “effort to exploit technical, social, economic, and institutional configurations of media to catalyze social change, sow dissent, and challenge the stability of social institutions.” Manipulators are motivated by a desire to do one or more of the following: spread ideologies, make money, gain status and attention. They deploy a variety of tactics—doctored photos, forged documents, viral slogans, etc.—to achieve their objectives. Manipulation takes place on a variety of participatory spaces online, including mainstream social media sites.

This study analyzes media manipulation in India and its impact on online news and political discourse.

- Kashmir was selected as a lens through which to study the problem. Kashmir is a divisive topic in Indian politics. The Modi government’s decision to strip Jammu & Kashmir of its statehood through the abrogation of Article 370 brought the region to the forefront of the news starting mid-2019. Severe restrictions placed on internet and mobile services, movement of people and freedom of the press left questions about the quality of information available to the public.
- Twitter was chosen as the platform for investigation: Unlike other platforms, its data is easily accessible through the API, it is a relevant space for political discourse in India and has seen media manipulation attempts in the past.
- Two specific types of manipulation—driving division and polarization and spreading pro-government propaganda—were selected as focus areas. Driving division and polarization in society was observed through the use of derogatory terms against liberals and government critics. Spreading pro-government propaganda was studied through the portrayal of the situation in Kashmir in late-2019 as “normal” (despite the enforcement of several restrictions, as described above).

This study explores the phenomenon of media manipulation through two broad questions.

- Evidence: Is there evidence of media manipulation in the online news ecosystem in India? Who are the influential players? What tactics do these stakeholder groups employ?
- Repercussions: Do these tactics violate platform policies? Are actions being taken against them?

Using ~3.5 million tweets downloaded via the Twitter API in Q4 2019, the study finds that media manipulation is a widespread and worrying phenomenon in India.

Key findings are presented below.

- There is evidence of media manipulation in the online news and political discourse ecosystem.
- Key stakeholders in the conversation include “super influencers” (the top 1% of influencers who drive ~50% of engagement online), political and media actors (highly influential in the conversation) and ordinary users (the largest group of participants in the conversation).
- The above-mentioned stakeholders engage in a variety of manipulative tactics (establishing motivations and the intention to manipulate requires further research). A few examples: political actors like Modi seem to engage in reward followbacks for their strongest supporters and strategic silence on controversial issues like Kashmir; “super influencers” like filmmaker Vivek Agnihotri use viral slogans to spread their messages; ordinary users engage in seemingly coordinated behavior and use wedge (divisive) issues like India-Pakistan relations and Hindu-Muslim tensions to distract the conversation away from policy and politics.
- The Indian mainstream media plays a crucial role in—both knowingly and unknowingly—amplifying manipulation and degrading the quality of online news. Narratives across sources vary drastically and there are numerous new news outlets with poor-quality content (though they have low influence). There are important questions to be asked around the credibility of information available for citizens and the prevalence of fact-based reporting (or lack thereof).
- There is little visibility into when and how Twitter enforces its policies. Transparency efforts are limited and lack granularity. Data provided for external research is insufficient to determine violations and, in some cases, suggests a lack of enforcement. While new initiatives seem promising, they remain insufficient and ambiguous from the outside.

Amnesty International India can take numerous actions to reduce the impact of media manipulation.

A few examples:

- Adopt appropriate vocabulary and publicly call out manipulative behavior. This will increase awareness about the phenomenon and will likely deter such activity in the future.
- Train journalists to identify manipulative activity and to select if, how and when to respond. Reducing amplification will significantly lower the reach of manipulation attempts.
- Support—identify partners, help in fundraising, etc.—the creation of tools to identify media bias and news quality in the Indian context. This will empower consumers to make better choices and will increase access to quality journalism in the country.
- Advocate for increased transparency from social media platforms like Twitter. Specific asks should include (but not be limited to) the release of datasets related to information operations in India and increased granularity in existing transparency reports.

There are several ways in which this research can be taken forward, starting with quantifying the scale of manipulation and establishing the motivations behind the behaviors of the various stakeholders.

I THE PROBLEM

The quality of news in India has deteriorated significantly in the recent past. One of the many drivers of this problem is likely to be media manipulation on social media platforms.

Key elements of a strong democracy include access to high-quality information and an ability for citizens to come together to debate, discuss and deliberate².

Access to the internet has grown rapidly across India and today, social media houses large parts of the online news^a ecosystem and associated political discourse.

Acknowledging the above fact, in an interview in 2017, the then Director of Twitter's India Operations said "News first breaks on Twitter. If you want to know what's happening across the world, you come to Twitter."³

At present, the quality of India's news ecosystem is at risk. A 2019 Reuters report indicated that a majority of survey respondents were concerned about the veracity of online news for reasons of disinformation, hyper-partisan content, poor journalism, and false news⁴.

A key driver of the deterioration in news quality and associated political discourse is likely to be occurrences of media manipulation i.e. attempts to take advantage of the design/ structure of media with a desire to affect social harmony, sow dissent and challenge the stability of social institutions^b.

^a For the purpose of this study, news is defined as a report of recent events (Source: Merriam-Webster).

^b According to Data & Society, media manipulation is the "effort to exploit technical, social, economic, and institutional configurations of media to catalyze social change, sow dissent, and challenge the stability of social institutions."^b (detailed explanation in a future [section](#)).

2 THE RESEARCH QUESTION

This project analyzes media manipulation—the influential stakeholders, tactics deployed and repercussions—in India through the lens of Kashmir-related content on Twitter.

Before creating interventions to improve the quality of online news and political discourse, one must first answer the question: What is the nature of media manipulation on social media in India?

This involves addressing two sub-questions:

- Evidence: Is there evidence of media manipulation? Who are the influential players in the conversation? What tactics do they employ?
- Repercussions: Do these tactics violate platform policies? Are actions being taken against them?

Important Scope-Related Choices

- The abrogation of Article 370 and Kashmir was selected as a case because of its relevance in South Asian politics and the abundant availability of data in 2019.
- Twitter was chosen as the platform because it provides easy access to data via APIs^c, is relevant for political discourse and has a history of housing media manipulation behavior.
- Offline news and non-English content^d were excluded due to limitations in resources.

^c While Facebook and WhatsApp are important in the Indian context, access to Facebook data for external research is limited and the encrypted nature of WhatsApp makes it impossible to access data at scale.

^d Non-English content was largely excluded given that filtering criteria were specified in English.

3 BACKGROUND

3.1 Kashmir and the Abrogation of Article 370s

With the end of the British rule in 1947, the princely state of Jammu and Kashmir (J&K) chose to join India under promises of high degrees of autonomy, as laid out in Article 370.

Coming into effect in 1949, Article 370 exempted J&K from the Indian constitution. It allowed the region to make its own laws except in matters of finance, defense, foreign affairs, and communication. The residents of this region, therefore, lived under their own laws.

Article 35A came into effect in 1954 and focused on territorial regulations: It permitted the local legislature to define its permanent residents. It prevented outsiders from settling, buying land, holding local government positions and being awarded scholarships of the region.

Article 370 has been modified over the years, but Article 35A remained unchanged.

The Bharatiya Janata Party's (BJP) campaign for the 2019 general elections took up Kashmir as a key issue. Modi vowed to remove both Article 370 and 35A if elected.

The abrogation of Articles 370 and 35A happened on 5 August 2019. The following preparatory steps were taken prior to the abrogation:

- Pilgrims were asked to leave the region
- Travel into Kashmir was restricted
- The army was deployed into the valley
- Prominent local politicians were put under house arrest
- Internet and mobile services were curtailed

- Public gatherings (4+ people) were banned

On 5 August 2019, Home Minister Amit Shah proposed a presidential order to repeal Articles 370 and 35A (and later passed the order with Parliament's support). J&K lost its statehood and was split into two union territories (changes shown in Figure 1).

The impact of these actions includes loss of autonomy for the state as well as a potential demographic shift (as other Indians move into this Muslim-majority state).



Figure 1: (Top) Map of Kashmir (Territory Split among India, Pakistan, China) with a Single State of Indian-Administered Kashmir Before 5 August 2019 (Source: BBC); (Bottom) Map Released by Indian Sources After 5 Aug 2019 Depicting the Split of the J&K State into J&K, and Ladakh (Source: Times of India^e)

^e Note: Indian sources often do not recognize Pakistan and China-administered regions in Kashmir

3.2 Media Manipulation

According to Data & Society, media manipulation is the “effort to exploit technical, social, economic, and institutional configurations of media to catalyze social change, sow dissent, and challenge the stability of social institutions”⁶.

Media manipulation usually takes place on “social and participatory media”. This includes blogs and websites, forums and message boards and mainstream social media sites⁷.

Most media manipulation movements have four key elements:

- Messaging i.e. strategies (examples below) through which the stability of institutions is challenged.
 - Driving division and polarization
 - Spreading pro-government propaganda
 - Discrediting political opponents through smear campaigns
 - Distracting or diverting conversation away from important issues
 - Suppressing participation through personal attacks⁸
- Sources i.e. the entities—accounts, organizations or individuals—from whom the attempted manipulation originates. Common sources include internet trolls, hate groups, ideologues, conspiracy theorists, influencers, hyper-partisan news outlets, and politicians. They are usually motivated by the desire to share/ spread ideological commitments, earn money (often through advertising revenues) and/or gain status, attention, and acceptance (and consequently control) within communities⁹.

⁶ The term “social institutions” is fairly broad: it covers the rules or structures in society including government, family, languages, legal systems etc.

⁷ This study focusses on media manipulation on mainstream social media sites only. Note: It is possible that manipulative content on social media originates in some of the other locations mentioned in this section.

From the outside, one might not be able to accurately identify motivations¹⁰.

- Amplifiers i.e. groups that, knowingly or unknowingly, assist the spread of media manipulation. Their actions bring attention to the manipulation and amplify their impact. Common amplifiers include some of the groups mentioned above as sources e.g. hyper-partisan news outlets, influencers, politicians.
- Tactics i.e. the specific methods used to execute the manipulation. There are a wide variety of tactics that have been identified in the past; new ones continue to come up through innovation. Examples include doctored content, viral slogans, evidence collages with misinformation, etc.

Impact: Media manipulation leads to decreased trust in mainstream media, increased misinformation in the news ecosystem and further polarization and radicalization.

Psychological factors increase vulnerability to media manipulation. Confirmation bias—the tendency to look for evidence confirming prior beliefs and to discount what seems to support an alternate view—leaves humans likely to believe false news that agrees with their world view and dismiss real news that goes against it¹⁰. Repeated exposure—seeing the same content multiple times—likely “increases the sense that misinformation is true”. Belief echoes—the fact that misinformation affects our thinking long after it has been corrected—imply that we cannot always reverse exposure to manipulation¹¹. “Fluent”—easy to process—narratives tend to feel more credible¹².

As of 2019, media manipulation campaigns have been documented in 70+ countries¹³.

¹⁰ For example, a significant proportion of “fake news” that appeared online in the US 2016 elections appeared to be ideologically driven. In reality, many these stories were put forth by teenagers in Macedonia who found that “publishing pro-Trump content generated more advertising revenues than pro-Clinton content”. ([Source](#))

3.3 Twitter's Response to Media Manipulation

In response to the growing phenomenon of media manipulation, Twitter has taken numerous steps to protect its platform and users. Specifically, it has designed policies to take action against manipulative behavior, released information through transparency reports and publicly committed to prioritizing the “health” of the conversation.

3.3.1 Platform Manipulation Policies

Twitter defines platform manipulation as the use of “Twitter to engage in bulk, aggressive, or deceptive activity that misleads others and/or disrupts their experience.”. A list of violations and penalties can be found in Tables 1 and 2 respectively¹⁴.

3.3.2 Policies against Fabricated Media (2020)¹⁵

From March 2020 onwards, Twitter began taking action against media that is significantly fabricated. Some of the actions include:

- Attaching the label “manipulated media”
- Reducing the visibility of the tweet
- In some cases, providing context through “credible sources”ⁱ

Figure 2 provides a decision matrix Twitter released online in order to explain decisions to take down fabricated content.

This policy was not created during the time of the data collection (2019); hence, its effects are not visible in the findings.

ⁱ Definitions for who qualifies as a credible source have not been provided at the time of this research

^j He spoke of four metrics to measure “health”, all of which were still being implemented internally. The four metrics are 1) Shared attention (How much of the conversation is attentive to the same conversation?);

3.3.3 Transparency Reports¹⁶

Twitter publishes a biannual transparency report with information on key trends such as (but not limited to) legal requests, policy enforcement, and platform manipulation.

The platform manipulation disclosures are provided as global aggregates. There is no granularity into how this number breaks down across types of manipulation (coordinated activity, artificial amplification, etc.) and across geographies.

Information operations disclosures are provided for select countries in the forms of archives of removed tweets by country. No disclosures have been provided for India so far.

3.3.4 Focus on Conversation “Health”¹⁷

In 2019, Twitter CEO Jack Dorsey publicly spoke about redesigning the Twitter experience to change user incentives and prioritize the “health” of the conversation. He spoke of new metrics^j and discussed feature changes such as de-emphasizing follow and like counts and centering the experience around topics and issues as compared to users.

These changes could have a significant impact on altering user behavior and reducing media manipulation on the platform. However, at this point, there is little public information around how these metrics will be incorporated into Twitter and when such changes will take place.

2) Shared reality (What % of the conversation shares the same facts?); 3) Receptivity (How much of the conversation is receptive to civil discourse?); 4) Variety of perspective (Are we seeing filter bubbles or do we get a variety of perspective?)

Table 1: Types of Violations as Listed in Twitter's Platform Manipulation Policy (Source: Twitter)

Category	Types of Violations
Accounts and identity	<ul style="list-style-type: none"> • Operating fake accounts • Artificially amplifying conversations by operating multiple accounts that: <ul style="list-style-type: none"> ○ Belong to identical or similar personas and/or post substantially similar content (overlapping accounts) ○ Interact with one another to inflate or manipulate prominence of specific tweets/ accounts (mutually interacting accounts) ○ Post duplicative content and/or create fake engagement (coordination)
Engagement and metrics	<ul style="list-style-type: none"> • Selling or purchasing followers, engagement or accounts • Using third-party apps that claim to add followers/ improve engagement • Trading or coordinating to exchange follows or tweet engagements through reciprocal behaviors • Selling, purchasing or trading (or offering to do so) accounts, even temporarily
Misuse of product features	<ul style="list-style-type: none"> • Tweets and direct messages (DMs): <ul style="list-style-type: none"> ○ Sending bulk, unsolicited replies, mentions, DMs ○ Posting and deleting the same content repeatedly ○ Repeatedly posting nearly identical tweets or DMs • Following: <ul style="list-style-type: none"> ○ Following and then unfollowing large numbers of accounts in an effort to inflate one's own follower count (Follow churn) ○ Following and/or unfollowing a large number of unrelated accounts in a short time period, particularly by automated means (indiscriminate following) ○ Duplicating other accounts' followers through automation • Hashtags <ul style="list-style-type: none"> ○ Overusing unrelated hashtags ○ Using popular hashtags with the intent to subvert or manipulate a conversation • Engagement <ul style="list-style-type: none"> ○ Aggressively or automatically engaging with Tweets to drive traffic or attention to accounts, websites, products, services or initiatives

Table 2: Penalties for Violating Twitter's Platform Manipulation Policies (Source: Twitter)

Action	Relevant Scenarios
Blacklisting URLs	<ul style="list-style-type: none"> • URLs are either blacklisted or given warning labels when their associated content is considered unsafe.
Tweet deletion and temporary account locks	<ul style="list-style-type: none"> • Isolated incidents or first-time offences will lead to actions ranging from the deletion of one or more tweet(s) to temporary account locks. Subsequent violations lead to permanent suspension. • In cases where multiple accounts are the main issue behind violation, the user can choose which account they would like to maintain; the other accounts are permanently suspended. • Violators of the fake accounts policy are required to provide government-issued identification for verification to get access to their accounts.
Permanent suspension	<p>Permanent suspension is used in the case of severe violation. Examples below:</p> <ul style="list-style-type: none"> • Operating accounts where the majority of behavior is in violation of the platform manipulation policy • Using the above-mentioned tactics to interfere in elections • Buying or selling accounts • Creating accounts to replace or mimic a suspended account • Operating accounts that Twitter can reliably attribute to entities that violate Twitter rules

Is the media significantly and deceptively altered or fabricated?	Is the media shared in a deceptive manner?	Is the content likely to impact public safety or cause serious harm?	
✓	✗	✗	Content may be labeled
✓	✗	✓	Content is likely to be labeled, or may be removed.
✓	✓	✗	Content is likely to be labeled.
✓	✓	✓	Content is very likely to be removed.

Figure 2: Decision-Making Matrix Provided by Twitter with regards to its Fabricated Media Policy (Source: Twitter)

3.4 Twitter in India

With a population of ~1.3 billion, India represents an irresistible market for all technology platforms, including Twitter. A few statistics about Twitter's incredible growth in India are presented below.

- Launched in India in 2013
- ~27 million monthly active users (2019)¹⁸
- Fastest-growing market for Twitter in terms of daily active users (5x above the average) (2017)¹⁹
- Modi among the 20-most followed accounts in the world (53 million followers as of February 2020)^k
- Accessible on all smartphones (full and lite versions available)
- Return on investment (ROI) for advertisers has improved (86% drop in cost per engagement)²⁰ (2017)

Challenges

At the same time, Twitter faces severe challenges in the country:

- Tightened government control: Government censorship of social media content has grown 5x since 2016²¹. New IT laws are expected to mandate that large platforms (>5 million users) maintain locally registered offices, take down content within 24 hours of notice and respond to data requests within 72 hours. Penalties for non-compliance involves fines and 7 years of jail time for executives²².
- The spread of fake news: All social media platforms, including Twitter, have been accused of enabling the spread of fake news and allegedly influencing voter behavior and inciting violence²³. With reference to

Kashmir, there are numerous examples of disinformation, particularly in February 2019 after the Pulwama terror attack²⁴. Twitter has responded to its fake news problem by 1) emphasizing the responsibility of users and the importance of digital literacy and 2) reiterating its platform manipulation policies²⁵.

- Accusations of bias: Twitter has been accused of unfairly suppressing liberal voices (while turning a blind eye to right-wing propaganda) and algorithmically suppressing dissent, particularly on trending pages²⁶. Twitter has denied these allegations and has emphasized its commitment to "serving an open public conversation in India"²⁷.

Relevance in News Consumption²⁸

- Among English-speaking online news consumers, Twitter is an important (though not the dominant) source of news online^m.
- ~18% of survey respondents use Twitter for news (in comparison, >50% said they use Facebook and/or WhatsApp for news).
- Despite its smaller user base, Twitter is an important platform for breaking news and "lively and often unruly and uncomfortable public debate".

^k For reference, Donald Trump has 73 million followers (February 2020)

^l Section 69A of the IT Act authorizes the government to block any digital information "if it is satisfied that it is necessary or expedient to do so, in the interest of sovereignty and integrity of India, defence of India, security of the state, friendly relations with foreign states or public

order or for preventing incitement to the commission of any cognizable offence." ([Source](#))

^m For context, >50% of respondents report getting their news from social media; ~25% say social media is their main source of news.

4 CASE STUDIES

The research questions for this study were answered through two case studies. The first case study relates to division and polarization and looks at the terms used to demonize liberals and those who criticize the government. The second case studies Amit Shah's claim of "everything is normal" in Kashmir on 20 November 2019 through the lens of pro-government propaganda.

Given the numerous sub-discussions under the umbrella topic of Kashmir and the abrogation, defining specific case studies was an essential step in focusing the analysis.

The case studies were designed around types of media manipulation messaging. Short descriptions are presented below. Specific criteria used to filter the data for analysis are presented in the [Methods section](#).

4.1 Case #1: Division and Polarization

This case studies the negative light cast on those who criticize government actions, mainly liberal, English-speaking elites from metropolitan cities and parts of the press. There are a variety of phrases used to demonize the group; some of them are described in Table 3.

The risk of polarization in India is high.

- Identity and religion have always been central to political discourse.
- The widespread use of social media has exacerbated this trend: social media is meant to draw in a broad audience and cut across differences; however, it is also effective in creating an us vs. them narrative, furthering division in society and driving action through a sense of outrage and conflict²⁹.

- News consumption itself is not polarized in India i.e. online news consumption does not polarize across predicted ideological divides. Instead, news audiences consume a wide and diverse range of sources "without showing signs of specialization and selectivity, ideological or otherwise." (based on research from the Indian elections in 2019)³⁰. Despite this fact, there are legitimate concerns around what people believe given their inherent biases (discussed [previously](#)) and the level of trust they place in different sources. Furthermore, if news consumption in India were to become polarized (i.e. similar to countries like the US), there could be serious challenges to dialogue, particularly across groups that operate with different facts.

Polarization threatens both India's social fabric and its stability as a democracy.

- Alienation and/or demonization of the other has made it challenging for ordinary Indians to engage in dialogues with people they disagree with.
- Criticizing government cynics and branding them as antinational is deeply problematic. The ability to question one's government and dissent is key to a healthy democracy.

4.2 Case #2: Pro-govt. Propaganda "Everything is normal"

This case looks at Amit Shah's claim that "everything is normal" in Kashmir and the manner in which this was discussed online.

- On 20 November 2019, Amit Shah responded to criticism about the situation in Kashmir. He made statements in parliament saying, "the situation in Kashmir is completely normal". He provided statistics on school exams and other issues

and said internet service restoration would be planned whilst keeping in mind security threats from Pakistan³¹.

- Shah’s claim is considered propaganda for two reasons: First, it could not be verified. The data presented was disputed and considered insufficient or misleading. Second, there was limited access to alternative information given the travel restrictions put in place by the government and the censorship of local news sourcesⁿ.
- Shah’s comment garnered significant attention and was discussed vigorously on

Twitter. Responses were mixed: some supported and spread this message, while others denied or used it sarcastically. Media coverage varied: foreign media sources were particularly critical of government actions while the Indian media’s response was mixed. Local Kashmiri media was left incapacitated after the abrogation; its participation was limited.

Impact: The lack of reliable information has been a key challenge in India, particularly around controversial topics like Kashmir. Given the different narratives adopted across media sources, news consumers have suffered.

Table 3: A Sample of Common Phrases used on Twitter in India for Polarization Purposes (Relevant for Case #1)

Phrase(s) ^o	Context
Tukde tukde gang	<ul style="list-style-type: none"> • Translates into gang looking to break something up into small parts. • In this case, it refers to people looking to break India up. • Coined by journalist Arnab Goswami in 2016 with reference to the Afzal Guru protests near Jawaharlal Nehru University³².
Award-wapsi/ intolerance gang	<ul style="list-style-type: none"> • Refers to celebrities who protest alleged intolerance, particularly by returning government awards during the Modi regime³³.
Khan market/ Lutyens gang	<ul style="list-style-type: none"> • Refers to urban elites from wealthy, cosmopolitan areas (like Khan market in Delhi)³⁴ who are out of touch with ground realities in India.
Urban Naxals	<ul style="list-style-type: none"> • Refers to urban intellectuals, activists, elites, etc. • Presents them as invisible enemies of India and claims they indoctrinate youth by “pretending to care about social issues”. • Popularized by filmmaker Vivek Agnihotri through his book “Urban Naxals”^{.35}
Negativity Gang	<ul style="list-style-type: none"> • Refers to those who do not recognize the positives in India, but instead only “pour scorn on (India’s) activities and achievements.” • Coined by journalist Arnab Goswami in 2019³⁶ while reporting on Kashmir

ⁿ This falls into the category of grey propaganda: i.e. a combination of accurate and inaccurate content and sourcing information with the intention to deceive. (Source)

^o Note: These (and the phrases used for data filtering) are only some of the many phrases that are used to polarize conversations online. Some phrases (like “sickular”) have been left out without intention; since this happened at random, the results are still expected to be representative.

5 METHODS

There were four key steps involved in the analysis (listed below). Details of the individual steps are presented in sub-sections.

- *Data was downloaded from the Twitter API and filtered into the relevant case studies (criteria in Table 4)*
- *Influence was calculated at an account level*
- *Influential users were categorized into 8 broad categories (Table 5)*
- *Research questions were answered (method described in Table 6)*

5.1 Data Collection via the Twitter API

~3.5 million tweets and their associated data were downloaded using Twitter's streaming API^p. The filtering criteria along with final dataset statistics are presented in Table 4. Limitations associated with the filtering criteria are presented in the [Appendices](#).

5.2 “Influence” Calculations

Influence has been defined as the ability of a user to impact the conversation on Twitter_q. A mathematical equation is presented below.

$$\text{Influence} = \#Retweets + \#Mentions + \#Quotes_r$$

The study of influence is key to analyzing Twitter content. Summary statistics from the

datasets show that over 80% of tweets are retweets. This indicates that Twitter is a platform where most users (at least in this context) do not create original content; instead, they engage with what influential users have already said. Selecting users/ groups to study based on their influence is, therefore, a logical way to approach the problem.

5.3 User Categorization

Influential users were categorized into 8 groups (details presented in Table 5). The analysis was conducted on both the account and the user category level.

Given time and resource constraints, users with no influence (influence = 0) were not categorized/ analyzed (addressed under [limitations](#)).

5.4 Addressing the Research Questions

Table 6 describes how qualitative and quantitative research methods were used to address the [previously defined research questions](#).

^p The API provides access to only 10% of relevant tweets. This explains why the volume of the tweets is significantly lower than expected.

^q Influence is a relative concept; the specific numbers do not matter.

^r Theoretically, there are four metrics that drive the ability to impact a conversation: #posts, #quotes, #tweets, #retweets. In reality, given Twitter's ranking algorithms, #posts is not significant for influence ([Source](#)).

Table 4: Data Filtering Criteria and Statistics for Resultant Datasets

General Filtering Criteria for Tweets	<ul style="list-style-type: none"> • Created between 25 October and 31 December 2019 • Contained “kashmir” in the text/hashtag • Contained one or more of 83 popular hashtags, 251 user mentions (List provided in the Appendices).
#Tweets Collected	~3.5 million

Metric	Case Study #1: Polarization	Case Study #2: Pro-government Propaganda
Tweet met the following criteria:	Contains one or more of the following terms in hashtags: <ul style="list-style-type: none"> • UrbanNaxal • Khanmarket • LibFarts • Tukde • Pseudoliberals, • Presstitutes • Lutyens • Pseudosecularists • Antinational • NegativityGang • GoToPakistan • Awardwapsigang • AntilIndia • Intolerancegang • Liberalhypocrisy 	<ul style="list-style-type: none"> • Created between 19 and 23 November 2019 • Contains “normal” in the hashtag or text
#Tweets collected	14451	7397
% Tweets that were retweets	87%	84%
Unique tweeters (includes those who retweet)	9253	5434
Tweeters with influence	1868	1080
% Tweets coming from users with influence	10%	18%

Table 5: User Categorization

#	Category	Explanation/ Notes
1	Political Actor	Politician, political party, party worker (India only)
2	Government (Govt.) Actor	Current or retired government representative (India only) Includes bureaucrats, army officers, etc.
3	International Actor	International politicians, governments (and their officials), multilateral organizations, international corporations
4	Media	Journalists (self-declared, well known) or media accounts (Indian and international)
5	Academic/ Activist/ Researcher/Author	-
6	Entertainment Industry	Members of the entertainment industry (particularly Bollywood), producers, musicians, other celebrities (India only)
7	Suspended/ Restricted/ Does not exist	Users no longer found on Twitter _s
8	Other	All other users (Indian and international)

Table 6: Methods to Address Research Questions

#	Research Question	Type of Research	Description
1	Evidence <ul style="list-style-type: none"> Is there evidence of media manipulation? Who are the influential players in the conversation? What tactics do they employ? 	Quantitative, Qualitative	<ul style="list-style-type: none"> Users were ranked based on influence calculations. Influential groups and users were selected for in-depth analysis. Tactics were identified by observing specified groups.
2	Repercussions <ul style="list-style-type: none"> Do these tactics violate platform policies? Are actions being taken against them? 	Qualitative	<ul style="list-style-type: none"> User profiles/ behaviors were observed to identify violations. Twitter transparency reports and data disclosures were analyzed.

_s The data download process was conducted in real-time. It therefore includes data that Twitter later took action against. There were <5% of “influential” accounts had been deactivated by either Twitter, or the users themselves.

6 RESULTS

This study arrives at three key findings (listed below). The rest of this section provides detailed explanations of each of these findings along with supporting data.

1. [Evidence of media manipulation \(Section 6.1\)](#): There is evidence of media manipulation in the online news and political discourse ecosystem. The most prominent users are seen using a variety of tactics to achieve their aims of manipulation.
2. [The role of the media \(Section 6.2\)](#): The mainstream media plays a crucial role in amplifying the manipulation and degrading the quality of online news and political discourse.
3. [Repercussions of manipulative behavior online \(Section 6.3\)](#): There is little visibility into if, when and how Twitter enforces its manipulation-related policies. New policies seem promising but remain ambiguous and/or insufficient.

6.1 Evidence of media manipulation

This section is divided into two sub-sections. First, prominent users in the conversation are identified based on influence calculations. Next, a sample of 5 tactics deployed by these users is explored in detail.

Key Finding: There is evidence of media manipulation in the online news and political discourse ecosystem. The most prominent users are seen using a variety of tactics to achieve their aims of manipulation.

- The most prominent users are “super influencers”, political actors, media and “other” users.
- These user groups deploy a variety of tactics to push their manipulative messages forward.

6.1.1 The most prominent users in the conversation are “super influencers”, political actors, media and “other” users.

The study of influence makes clear that certain users are significantly more influential than the rest. It concludes that 4 groups of users—“super influencers”, political actors, media, “other” users—must be studied further.

General notes on user behaviors across case studies and variations in their levels of influence are presented in the [Appendices](#).

“Super influencers”

The top 1% of influential users (referred to as “super influencers”) have a disproportionate level of influence. Excluding them decreases the average influence for their category by 50%.

As seen in Table 7, “super” influencers come from a variety of categories. While 80% of this group belong to the media in case #2 (pro-govt. propaganda), in case #1 (polarization), influence is disproportionately held by entertainment industry user, Vivek Agnihotri (vivekagnihotri).

Table 7: Top 1% of Influencers in Both Case Studies (Source: Own Analysis)

#	Case #1: Polarization		Case #2: Pro-govt. Propaganda	
	User Screen Name	Avg. Influence	User Screen Name	Avg. Influence
1	vivekagnihotri	8,093	firstpost	2,811
2	ShekharGupta	1,326	AmitShah	2,708
3	ashokshrivasta6	939	RanaAyyub	1,537
4	ThePrintIndia	861	DeccanHerald	1,393
5	siddharthagigoo	851	news18dotcom	1,283
6	Shehla_Rashid	786	bainjal	612
7	narendramodi	599	sardesairajdeep	604
8	DostKhan_Jammu	508	imMAK02	437
9	rahulkanwal	480	ndtv	256
10	ashokepandit	471	IndiaToday	254
11	AmitShah	444		
12	tonyashai	412		
13	VertigoWarrior	332		
14	madhukishwar	320		
15	AgrawalSadhuram	304		
16	htTweets	279		
17	OpIndia_in	278		
18	SadhviPragya_MP	273		
19	neelakantha	268		

Color Code and Category				
Academic/Author/ Researcher/ Activist	Entertainment Industry	Media	Political Actor	Other

Political and Media Actors

After excluding “super influencers” (Table 8), political actors and media emerge as dominant in terms of influence. Government actors appear influential but further examination shows that their influence is largely driven by the official accounts of Modi and Shah^{uv}.

“Other” users

While low on influence, “other” users are considered prominent and are investigated further given that they make up the largest part of the conversation i.e. they dominate in terms of % of users (Table 9).

Table 8: Levels of Influence Including and Excluding "Super Influencers" (Source: Own Analysis)

#	Category	Case #1: Polarization		Case #2: Pro-govt. Propaganda	
		Without "super" influencers	All users	Without "super" influencers	All users
1	Suspended/ Restricted/ Does Not Exist	3.64	3.64	5.55	5.55
2	Other	5.43	6.96	5.08	5.08
3	International Actor	6.48	6.48	7.35	7.35
4	Academic/Author/Researcher/ Activist	8.63	29.47	14.59	19.50
5	Entertainment Industry	9.42	129.77	4.82	4.82
6	Political Actor	13.16	18.74	11.59	39.97
7	Govt. Actor	14.45	14.45	10.36	10.36
8	Media	14.58	26.32	10.50	46.01

Table 9: User Breakdown by Category for Both Cases (Source: Own Analysis)

#	Category	Case #1: Polarization	Case #2: Pro-govt. Propaganda
1	Academic/Author/Researcher/ Activist	6%	9%
2	Entertainment Industry	4%	1%
3	Govt. Actor	4%	4%
4	International Actor	4%	16%
5	Media	19%	25%
6	Other	49%	34%
7	Political Actor	12%	10%
8	Suspended/Restricted/ Does Not Exist	2%	1%

† Dominance is measured as being above the median influence score.

u India’s Consul General in the US, Sandeep Chakravorty (CHAKRAVIEW1971) is also influential, particularly in case study #1 (this influence came largely from his references to the “Israel model”)

v Other case-specific variations exist (for example, entertainment industry users dominate the polarization conversation; academics etc. are influential, particularly in the case of pro-government propaganda).

6.1.2 These user groups deploy a variety of tactics to push forward manipulative messages.

There are a wide variety of media manipulation tactics that have been adopted in the Kashmir-related discourse. This section explores a sample of five tactics in detail (list below). These tactics are used by three of the previously identified user groups: “super influencers”, political actors, “other” users. The media is dealt with separately in the [next section](#)^w.

- Potential reward followbacks deployed by political actors
- Strategic silence adopted by political actors
- Viral sloganeering deployed by “super influencers”
- Seemingly coordinated behaviors among “other” users
- Use of wedge issues to distract the conversation

6.1.2.1 Potential reward followbacks deployed by political actors

There are a significant number of “other” users who mention being followed by notable political actors in their user description. Several also have screenshots from Twitter with the notification of these politicians following them as their cover photos³⁷.

The highest prevalence is among Modi supporters; they have phrases like “followed by Shri Narendra Modi” or “blessed to be followed by @NarendraModi” (example in Figure 3).



Figure 3: Likely Reward Follow by @NarendraModi (Source: Twitter)

Research from 2017 shows that ~40% of the accounts that Modi follows are ordinary citizens. Most in this category are “die-hard” Modi supporters who “defend him furiously while bashing the opposition”³⁸. Many of these accounts have spread false information³⁹ in the past.

According to research from the University of Michigan⁴⁰, while political teams (like that of Obama) have engaged in “indiscriminate followbacks” to reach voters in the past, Modi was far more selective; he only followed his most active supporters back.

By using the follow feature of the platform to his advantage, Modi appears to be incentivizing his supporter base and crafting his image. Specifically, Modi’s “followbacks” served two purposes:

- Calls to action/ encouragement to Modi’s biggest supporters to continue supporting him online^{41x}

^w It is hard to establish intent in many of the cases identified here. Further research would be required to conclusively say that these tactics were deployed with the purpose to manipulate.

^x As per past data analysts from the BJP, trolls become more active once they have been followed by Modi ([Source](#))

- Brand-building efforts to signal that Modi is “part of normal life”⁴², a worker who listens to those who follow him_y

While some accounts spoke about being followed by Modi (and other politicians), others were followed by less prominent BJP politicians including Piyush Goyal (Minister of Railways), Nirmala Sitharaman (Minister of Finance) and Tajinder Bagga (party spokesperson). Examples are provided in Figure 4.

Research suggests that there could be a tiered structure of being followed, with Modi being at the top⁴³. Different tiers could gamify the experience for users such that they are incentivized to increase online support for the BJP to rise up the tiers.

Note that several users who claim to be followed by these politicians are in fact not followed by them (Figure 5). This could be due to two reasons:

- They were never followed in the first place; users perhaps claim being followed to gain credibility/ signal their support within groups of right-wing supporters.
- They were followed and then unfollowed. This could be evidence of a reward fallback that served its purpose simply by the act of occurrence. If this has taken place at scale and over short periods, it could violate platform manipulation policies.



Figure 4: Some Users are Followed by Less Prominent BJP Politicians; Some Also Boast of Being Blocked by Famous Individuals Who Do Not Support the BJP (Source: Twitter)



Figure 5: A Twitter User Who Claims to be Followed by Several BJP Politicians is, in Reality, Not Followed by Them. (Source: Twitter)

^y Creating an image of being an ordinary Indian was an important distinction between Modi’s strategy and that of the Congress. (Source)

^z Though some users have screenshots in their profile as evidence.

6.1.2.2 Strategic silence adopted by political actors

Amit Shah and Narendra Modi have stayed silent on most conversations related to Kashmir. Though statements about normalcy originated with Amit Shah, neither he nor Modi have acknowledged or responded to similar conversations on Twitter.

For example, Shah was mentioned 2500+ times in case study #2_{aa} but never responded.

While popularity can explain the high ratio of mentions to tweets, in this case, both Modi and Shah had 0 tweets on Kashmir and normalcy in the days around the statements in parliament. Instead, they tweeted ~thrice per day on election campaigns, birthday wishes for celebrities, meetings with diplomats, etc. (Screenshots of tweets from Amit Shah are presented in Figure 6).

Silence or distraction through other topics has been Modi's chosen response to numerous issues in the past (both online and offline) and seems to be at play here as well⁴⁴.

Impact: Silence creates confusion and leaves holes in the information ecosystem. It protects Modi and Shah by keeping them away from controversial conversations that could affect their brand. Instead, it puts other users in the front-line of the conversation⁴⁵. Distraction through other topics diverts the conversation away from important political matters and results in a serious lack of accountability. Without a clear message from leading authorities in the country, the status of Kashmir is unknown to the public and, as a topic, can appear less important than it is in reality.

^{aa} Given that only 10% of relevant tweets are available to download and each tweet has to meet the specific criteria, this number is likely significantly larger in reality.



Figure 6: Examples of Tweets Shared by Amit Shah Between 19 and 23 November 2019 i.e. The Days Around his Statements on Kashmir Normalcy (Source: Twitter)

6.1.2.3 Viral sloganeering deployed by “super influencers”

Viral sloganeering is defined as “a process of crafting divisive cultural and political messages in the form of short slogans and propagating these (both online and offline) to influence viewers, force media coverage, and provoke institutional responses”⁴⁶. On the news consumer/receiver’s end, these slogans are easy to process and recall and can appear more credible than they are in reality^{bb}.

Viral sloganeering is an important tactic, particularly in the case of polarization. A list of viral slogans was included in the filtering criteria.

Table 10 lists the most popular slogans from the polarization dataset^{cc} and provides the frequency with which they appear in the conversation.

Table 10: List of Most Popular Hashtags From Case #1: Polarization (Source: Own Analysis)

#	Hashtag	Frequency
1	urbannaxals	3870
2	kashmirihindu ^{dd}	3620
3	tukdetukdegang	1159
4	fraudpatrakars	240
5	antinationals	238

“Urban Naxals”

- The term “Urban Naxals” was popularized by “super influencer” and Bollywood filmmaker Vivek Agnihotri.
- Agnihotri references a 2004 document by the Communist Party of India called “Urban Perspective” as his evidence for the phenomenon along with a statement from India’s former Prime Minister, Dr. Manmohan Singh about “Naxalism being India’s greatest internal security threat”⁴⁷.

^{bb} Cognitive bias associated with “fluency narratives” discussed previously.

^{cc} Common words used as hashtags such as #kashmir, #cabprotests etc. were excluded from this list as they do not qualify as viral slogans.

- Agnihotri uses the term both to refer to groups of people and to target individuals like Shekhar Gupta (journalist) and Shehla Rashid (activist).
- The term has now been adopted by the mainstream media, as well as by politicians like Amit Shah⁴⁸.
- Based on his interviews, Agnihotri appears aware of the media manipulation tactic he employs. A quote from a 2019 interview: “Everybody is trying to add to the noise, add to the news, add to the narrative in some way. You have to understand, in politics out there, it is a war of narratives.”⁴⁹

“Tukde Tukde Gang”

- The term “Tukdetukdegang” was prevalent in the dataset, mainly through media persons like Ashok Shrivastava (DD News), Pradeep Bhandari (Jan Ki Baat), Ashish Kohli (unspecified journalist); film producer Ashoke Pandit and professor Madhupurnima Kishwar.

Comparison of usage

- The term “UrbanNaxals” is still closely linked to the person who popularized it (Agnihotri). Over 75% of the usage of the term came through retweets of his content.
- On the other hand, similar levels of usage for “Tuketukdegang” were spread across the retweets of the five users mentioned above.
- This implies that “Tukdetukdegang” is relatively more organic in the discourse, as compared to “UrbanNaxal”.

^{dd} This term was retweeted only when used by Vivek Agnihotri; hence, the influencer appears to be more significant than the term itself.

6.1.2.4 Seemingly coordinated behavior among “other” users

Many self-declared right-wing supporters have similar content in their profiles (listed below).

- The use of the Indian flag in screen names (examples in Figure 7)
- Similar language in user descriptions including exact phrases such as “blessed by Modi”, “proud to be blocked by”, “#IndiaSupportsCAA” etc. (examples in Figure 8)
- Some overlap in profile/ cover photos

This phenomenon could be an indication of coordinated behavior. However, without additional information on account usage behavior, this claim cannot be validated.

Alternatively, this may be an organic movement. Profile changes could be responses to other cues (such as Modi’s followbacks⁵⁰) or signals of allegiance within groups.

Without coordination, this behavior might not be evidence of manipulation. Either way, given the volume of content and number of users that follow these tactics, they ought to be examined in more detail.



Figure 7: Examples of Users with Indian Flags in Their Names (Source: Twitter)

-  **Hari Prabhakaran** @Hariindic Chennai, India
Politician from TamilNadu | Honoured and **Blessed to be followed by Shri @narendramodi ji.**
-
-  **Dr Anish Kumar** @draksbond South Pole
Doctor(Physician) | Avid Traveler | BJP Supporter|**Blessed to be Followed by Shri @narendramodi JIIAAPtards stay away |Sachin Fan |**
-
-  **Sanyam Jain** @modivanibharat Indore, Madhya Pradesh
0% Attitude •II• 0% Ego •II• 100% NATIONALISM Proud Hindu Nationalist •II• India First •II•
Blessed to be followed by PM @narendramodi Ji•II “जो है सो है”
-
-  **MUKESH Modifier #Goy** @monurajasthan Rajasthan, India
राजस्थान का वासी हू, भक्त हु वीर हनुमान का! साथी उस मोदी का, जो मान बढ़ाये हिंदुस्तान का!! #जयश्रीराम
Blessed to be followed by PM @Narendramodi Ji जय-हिन्द !
-
-  **MAHESH** @MahiHerambha India
Marketing Professional-Shipping & Logistics, राष्ट्र सर्वोपरि-राष्ट्र सर्वप्रथम, **Blessed to be followed by Hon PM @narendramodi ji**
-
-  **Dr.Monika Langeh** @drmonika_langeh Jammu &Kashmir UnionTerritory
IDoctor|Nation First|Honoured and **Blessed to be followed by Hon'ble PM @narendramodi ji**
I Proud Dogral Jai Bharat
-
-  **Dr. Saagar Anand** @saagar_anand UT of Jammu
MBBS IProud, Honoured & **Blessed to be Followed by Honourable Prime Minister @narendramodi ji** |Believes in MODI-FIED India |Dream to meet Modi ji once |
-
-  **Nikhil Srivastava** @snikhil_social New Delhi, Lucknow, India
Blessed to be followed by Respected PM @narendramodi Ji | Nationalist | Entrepreneur |
Social Media Enthusiast | Mantra: I'm Possible
-
-  **अजय त्यागी** @tyagi_ajay1 हिंदुस्तान
Blessed to be Followed by PM @narendramodi ji & other cabinet Ministers #Nationalist #100% Indian
-
-  **Manish Shukla** @manishBJPUP Lucknow/Jaunpur
प्रवक्ता भारतीय जनता पार्टी @BJP4UP उत्तर प्रदेश **Blessed to be followed by respected pm @narendramodi ji manishlead@gmail.com**

Figure 8: Examples of User Descriptions with Similar Language (“Blessed to be Followed by”) and Claims of Follows by @NarendraModi (Source: FollowerWok)

6.1.2.5 Use of wedge issues to distract the conversation

“Wedge issues” are divisive, political issues that “often leave no room for nuance”⁵¹. In the context of India, these include several topics like relationships with Pakistan and Hindu-Muslim tensions. With the information overload in the online ecosystem, wedge issues often have higher success rates in terms of capturing user attention.

A few examples of wedge issues being used to distract the conversation away from the situation in Kashmir are presented below.

- An overload of positive messaging to distract from the main topic: Instead of responding to the Kashmir normalcy question, several users shared positive news about India and the government.

This tactic has been well-documented in the case of China (referred to as “Xuanchuan”). Research has shown that the Chinese regime’s strategy has been to “distract the public and change the subject” through cheerleading for China, the communist party and “other symbols of the regime”⁵². An example from Kashmir-related discourse is presented below:

Example) @sushtany (640 followers) got significant engagement^{ee} when they shared a post presenting India in a positive light compared to Pakistan (Figure 9). The post distracted away from the conversation of normalcy in Kashmir and used the issue of India-Pakistan to trigger reactions. The veracity of the information is also in question given that the news was only

reported by two, allegedly biased sources (ANI⁵³, Zee News⁵⁴)^{ff}.



Figure 9: Positive Messaging around India Shared by @sushtany (Source: Twitter)

- References to Hindu-Muslim tensions lead to an “us vs. them” narrative, further polarizes the conversation⁵⁵ and frames it along communal lines (while ignoring political decisions). Two examples of how religious tensions play out in the Kashmir discourse are presented below.

Example 1) Doctored screenshots of tweets on the Kashmiri-Hindu topic were used against activist Shehla Rashid (~750k followers) during a disinformation campaign in early 2020. (Figure 10)^{gg}.

Example 2) Kashmiri Hindu journalists like Sunanda Vashisht, Aditya Raj Kaul^{hh} (~150k-250k followers) have been vocal about the Kashmiri Hindu exodusⁱⁱ. The political conversation here quickly dissolves into communal tension and oftentimes, hate speech. (Figure 11).

^{ee} The attention they received was largely driven by a journalist’s actions (explained in a [later section](#))

^{ff} The intention to manipulate cannot be established based solely on this tweet; more research is required.

^{gg} This content is likely to meet the criteria for removal under Twitter’s Fabricated Media policy that came into effect in March 2020.

^{hh} Kaul has been affiliated with Republic TV which have been accused of bias and media manipulation ([Source](#))

ⁱⁱ Vashisht testified in the US Congress in mid-November 2019 about her experiences having to flee Kashmir ([Source](#)); Kaul recently voiced his emotions on the exodus in response to the trailer of a new movie on the subject and started the trend [#HumWapasAayenge](#) ([#WeWillComeBack](#))

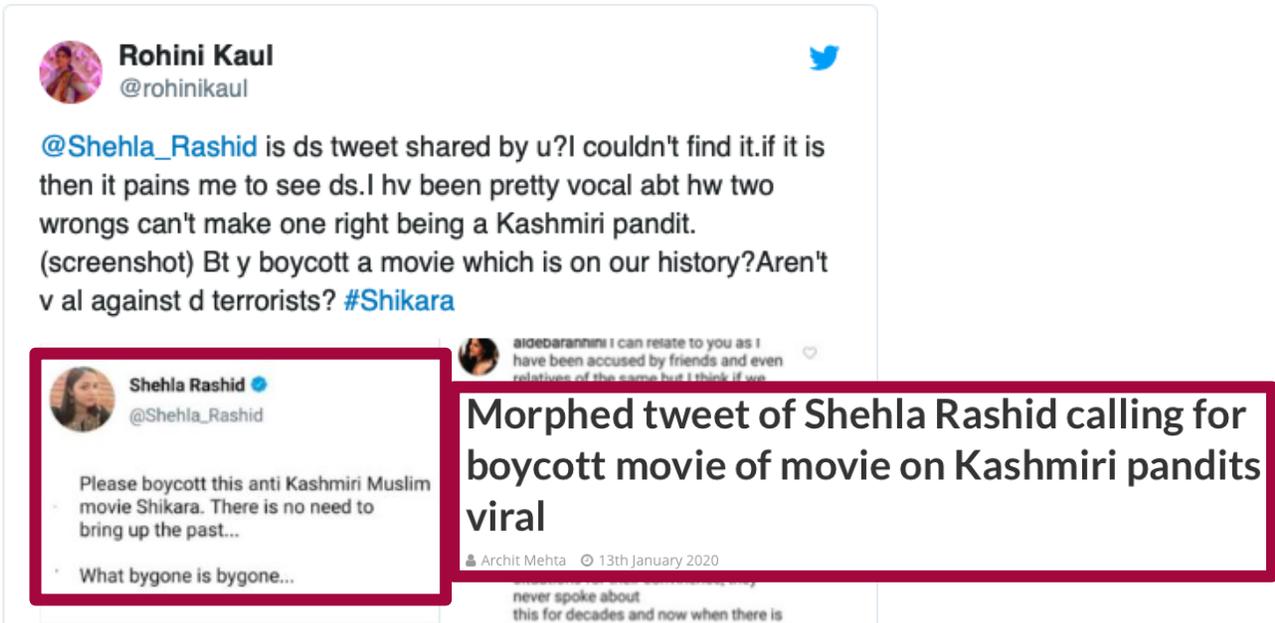


Figure 10: Screenshot from a Fact-Checking Article on the Disinformation Campaign against Shehla Rashid (Source: AltNews)

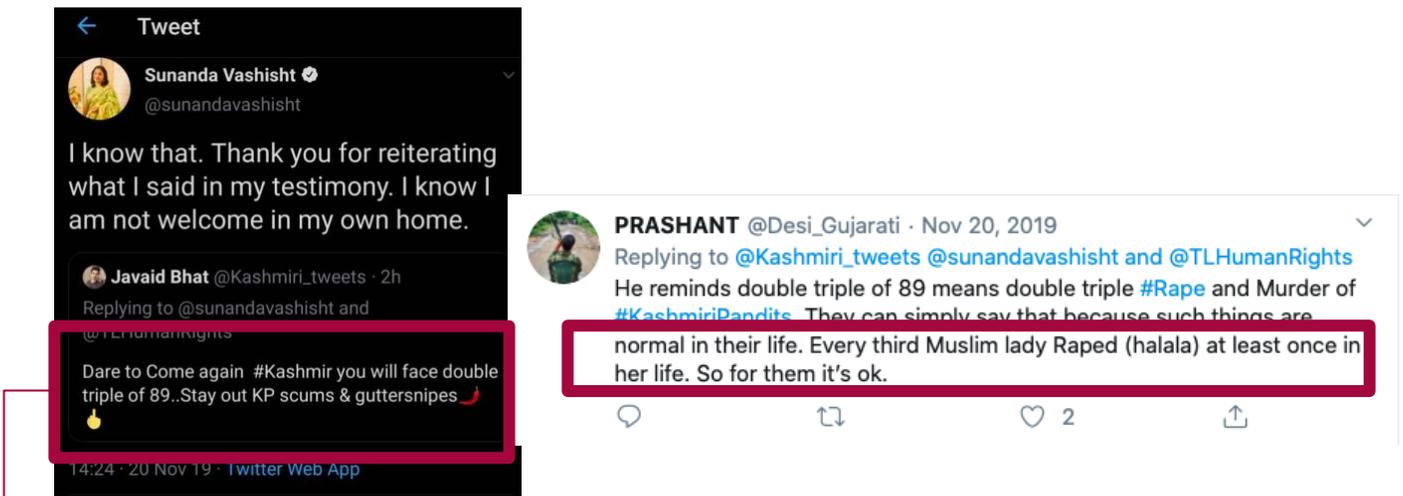


Figure 11: Conversation between Sunanda Vashisht and Another User with regards to Kashmiri Hindus and Muslims (left); Conversation Deteriorates into Communal Issues (right) and Diverts from the Political Actions in Kashmir. (Source: Twitter)

This tweet was later removed by Twitter

6.2 The Role of the Media

The media plays an important role in delivering information to the public. Given the power it has through source cue effects (signals and perceptions of authority that stem from the recognition of its brand names)⁵⁶, and its ability to amplify select narratives, it has significant influence over the quality of the discourse and the scale of manipulation.

Key Finding: The mainstream media in India plays a crucial role in degrading the quality of online news and political discourse and amplifying media manipulation.

- While there are numerous new, low-quality media sources online, most of them have little influence over the conversation.
- Variation in content across mainstream media sources raises questions of access to credible information and the prevalence of fact-based reporting.
- Mainstream media incentives likely encourage the amplification of media manipulation.
- Mainstream journalists have the power to amplify manipulation and often do so unknowingly.

6.2.1 While there are numerous new, low-quality media sources online, most of them have little influence over the conversation.

With the massive reduction in costs associated with publishing content online, there has been significant growth in the number of news sources (similar trends in other countries).

The quality of news produced by these sources appears to be quite low. A large number of them (examples in Figures 12-14) break rules of ethical journalism^{ji}.

A few ways in which they do so are listed below⁵⁷. All of these indicate that they are likely to amplify media manipulation.

- Adopt polarizing terms
- Produce hyper-partisan content
- Limit editorial oversight
- Fail to verify information

While the quality of content is worrying, these new media sources seem to have low influence in the conversation. Data from this study shows that only ~8% of the 100-most influential media sources are independent online websites with low-quality content. Additionally, a survey from Reuters states that a majority of such news sources have small audiences and are rarely listed as recently visited by respondents⁵⁸.

Note: Low influence is indicative of the overall impact they have on the conversation. However, there could be limited networks within which their messages are influential and considered credible. This would create an environment where people from different groups believe different facts^{kk} and would increase polarization.

^{ji} Several of these sources admit to being blogs or opinion sources on their About pages; however, given the reach they have on platforms like Twitter and the challenge in finding these caveats on their websites, it is reasonable to assume that viewers may not be aware of these subtleties.

^{kk} This variation in facts across society is one of the new metrics (called shared reality) that Jack Dorsey plans to use to measure the “health” of the conversation (discussed [previously](#))

The question of who counts as a journalist and what counts as news media (compared to simply influencers, blogs and new media) is an important one. There has been significant debate, particularly in the US on this subjects⁹. As online news consumption leads users to a wide variety of media sources, both reliable and unreliable, it is important to develop mechanisms to indicate and measure quality.

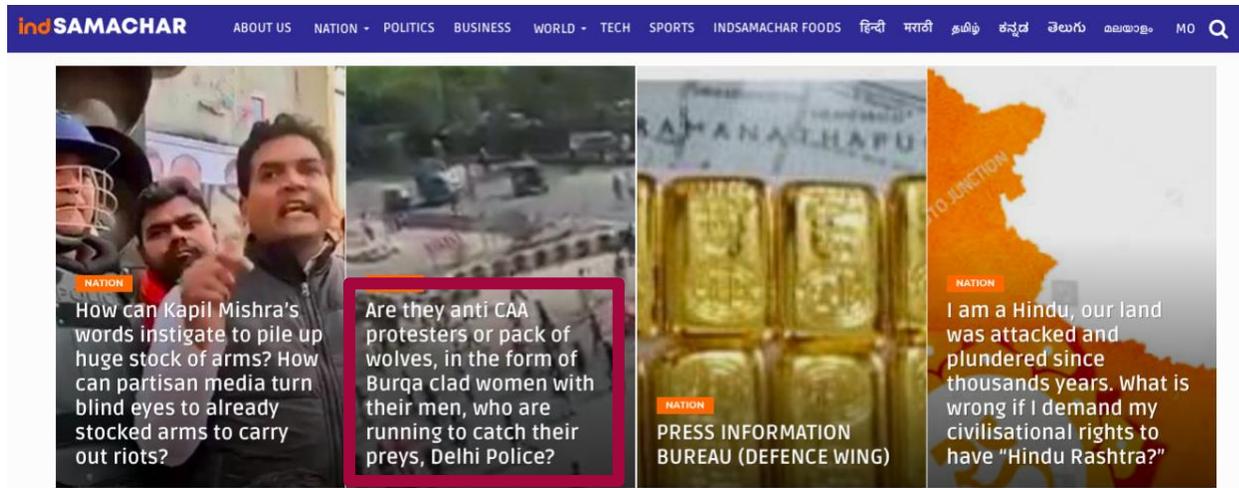


Figure 12: Hyper-Partisan Content with Poor Editorial Oversight on IndSAMACHAR (Source: IndSAMACHAR)



Figure 13: Examples of Poor-Quality News and Attempts of Polarization from Jan Ki Baat founder (Source: Twitter, LinkedIn)



Figure 14: Unverified Content Posted by a Person Associated with FactHunt (Source: Twitter)

6.2.2 *Variation in content across mainstream media sources raises questions of access to credible information and the prevalence of fact-based reporting.*

Media responses to the normalcy narrative are classified into 4 categories (Table 11)ⁱⁱ.

Within the mainstream Indian media, even on the same day, you read very different versions of the situation in Kashmir. Figure 15 shows screenshots from the news as shared by 4 mainstream Indian media sources on 20 November 2019. Each puts forth a different version of the truth.

Figure 16 presents these differences for the top 15 mainstream media sources in India^{mm}. Interestingly, mainstream Indian “media houses own a spread of narratives”⁶⁰. For example, Times Now, The Times of India and Mirror—all owned by the Times Group—show variations in their responses. Similar variations are seen in the behaviors of FirstPost and News18, both of which are owned in part by the Ambani family. Further research is required to explore the

relevance of media ownership in media manipulation.

Indian and foreign media have very different versions of the truth (Figure 17)ⁿⁿ. Foreign media, including from Pakistan, almost unequivocally argues/ disputes normalcy. The Indian media seems evenly split in its reactions, with a significant proportion of tweets affirming the normalcy narrative. The Kashmiri local media seem to mirror mainstream Indian media. This is unsurprising given the severe restrictions on press freedoms in the region⁶¹. These results are likely not representative of their independent coverage.

The variation in content across news sources raises serious concerns around the availability of credible information in a democracy. It also points to issues around journalistic integrity and fact-based reporting.

Table 11: *Categories for Responses from the Media to Amit Shah's Normalcy Narrative (Used for Tweet Content Only)*

#	Category Name	Category Description ^{oo}
1	Affirm	Say things are normal by stating own facts (go beyond quoting facts)
2	Report	Say things are normal/ provide statistics, through quotes from Shah
3	Question	Question normalcy claims, call for debates
4	Argue/ Dispute	Quote others, state alternative facts to dispute the normalcy narrative

ⁱⁱ Data analysis for the media is limited to tweet content from mainstream media accounts (and those of their associated journalists). It does not examine or categorize the content in the links or articles shared by the media.

^{mm} The list of top-15 media sources has been created based on popular online news sources identified by the [Reuters report](#) on online news consumption.

ⁿⁿ Analysis is based on tweet content only (as mentioned in a previous footnote)

^{oo} The categories here have been defined based on an interpretation of the [SPJ Code of Ethics](#), particularly its principle of “seeking truth and reporting it”. Specifically, a distinction has been drawn between affirming and reporting. In this case, reporting Shah’s statements are considered “reporting”; supporting his statements with additional facts or without attributions are considered “affirming”.

ThePrint @ThePrintIndia · Nov 20, 2019
 'In Kashmir, disenchantment has given way to raw anger at assault on collective being'
 @_YogendraYadav writes
 #ThePrintOpinion



Public opinion in Kashmir has now been pushed away from India
 Not Narendra Modi, Amit Shah is the new face of Indian state for hurt and angry Kashmiris.
 theprint.in

4 retweets 13 likes

Disputes

Times of India @timesofindia · Nov 20, 2019
 How 'normal' is Kashmir? What's functioning, what remains forbidden



How 'normal' is Kashmir? What's functioning, what remains forbid...
 India News: More than a hundred days after the August 5 move by the Narendra Modi-led BJP government to remove Jammu & ...
 timesofindia.indiatimes.com

1 reply 6 retweets 22 likes

Questions

India Today @IndiaToday · Nov 20, 2019
 First film being shot in #Kashmir after the abrogation of #Article370.
 Cast says "did not face any difficulty in shoot." @suniljhat shares more details.
 Watch #InDaClub with @radhikaavasthi2.
 More videos: bit.ly/it_videos



Republic @republic · Nov 20, 2019
 Union Government lists measures to attract investments in Jammu & Kashmir

Union Government lists measures to attract inves...
 Union Government has taken several measures to attract investments in Union Territory of Jammu ...
 republicworld.com

4 replies 6 retweets 19 likes

Affirms

Figure 15: A Comparison of Kashmir-Related News from Different Mainstream Indian Sources on 20 November 2019 (Source: Twitter)

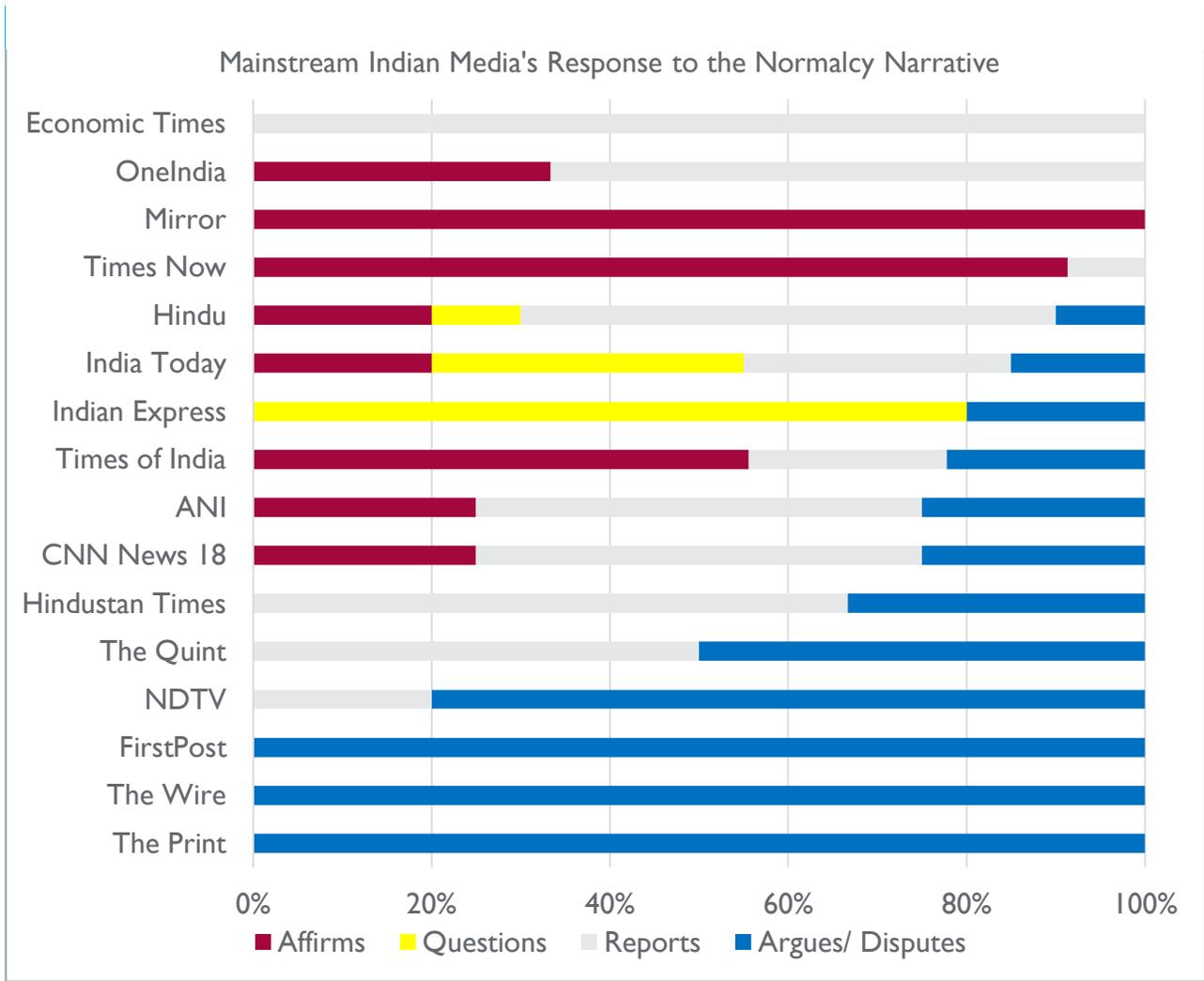


Figure 16: Coverage of the "Normalcy" Narrative by Mainstream Indian Media Sources (Analysis of Tweet Content Only) (Source: Own Analysis)

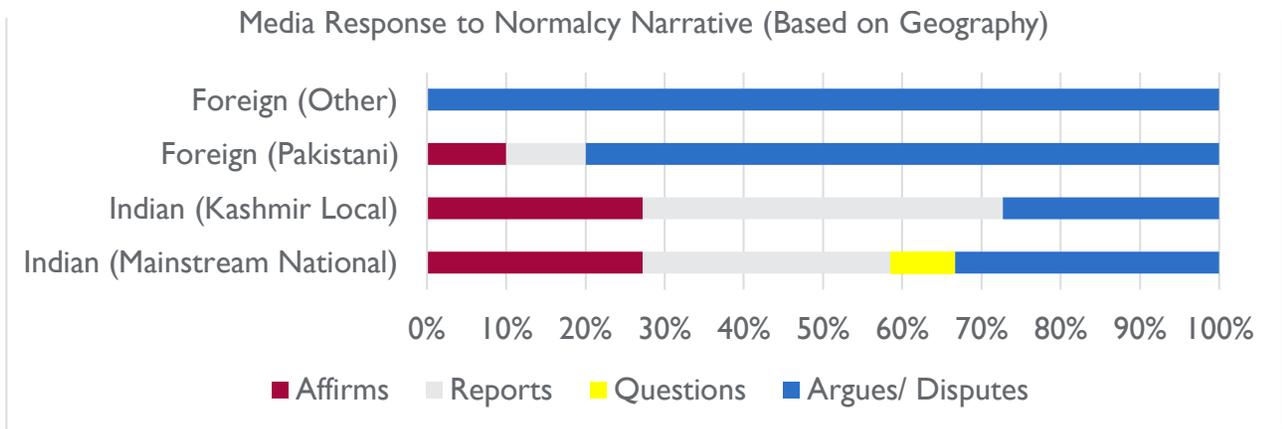


Figure 17: Coverage of the "Normalcy" Narrative by Media Sources Based on Geography (Analysis of Tweet Content Only) (Source: Own Analysis)

6.2.3 Mainstream media incentives likely encourage the amplification of media manipulation.

Social media is an attention economy where providers of content compete for user engagement amidst an overload of information. Valuable content is, therefore, that which is able to capture the most “eyeballs” and clicks⁶².

The move to online news has led to a sharp change in the incentives of mainstream media. They now prioritize metrics like website traffic, page views and trending topics (and have the analytical tools to measure them accurately). These incentives manifest themselves through content with sensational and emotional language, as well as explicit calls to action. Table 12 provides descriptions of these phenomena.

India’s mainstream media sources and their susceptibility to incentives are shown in Figure 18. Among the top 15 media sources (and likely the others as well), there is a correlation between susceptibility to incentives and the quality of news. There appears to be a causal link (described below); however, qualitative analysis is required to make a definitive statement in this regard.

- Correlations between susceptibility to incentives and use of polarizing terms = 0.3. Polarizing terms could themselves be considered a type of sensational language. It is therefore unsurprising that those susceptible to media incentives are likely to use polarizing terms in their content.
- Correlations between susceptibility to incentives and “affirming” normalcy_{pp} = 0.6. This is expected: Press freedom in India has declined significantly under Modi⁶³. The BJP is known to target news sources that criticize them (through tax raids⁶⁴ and boycotts), and the government is the largest advertiser for newspapers⁶⁵. Media sources susceptible to engagement incentives are likely also susceptible to other kinds of incentives from the government. They are therefore more likely to spread government narratives (like that of normalcy).

^{pp} Correlation between the “response score” and the “susceptibility to incentives”: The response score is calculated by assigning a 1, 0, -0.5 and -1 to sources’ likelihood to affirm, report, question and argue/ dispute the normalcy narrative respectively. The susceptibility to incentives is simply the average of the % of tweets with sensational language, and % of tweets that promote engagement.

Table 12: Descriptions of Media Incentives and Some Examples from Indian Media

Media Incentive	Description	Comments/ Notes
Calls to actions	Tweets that contain action words like “read”, “follow”, “listen”.	<ul style="list-style-type: none"> • Times Now leads • NDTV, online-only media posts seldom ask users to “engage”.
Sensational language	Tweets that evoke significant emotion or those that can drive virality. <ul style="list-style-type: none"> • Examples of sensational phrases: “fire up”, “#KashmirStillCrying”. • Examples of emotional narratives: “helpless...daughters being raped”, “children ...unable to see friends”. 	<ul style="list-style-type: none"> • Mirror leads with slogans like #KashmirisNormal • The Quint uses emotional stories. • ANI uses fear (“desperate attempts”) and threats of enemies (“Pakistani propaganda”).

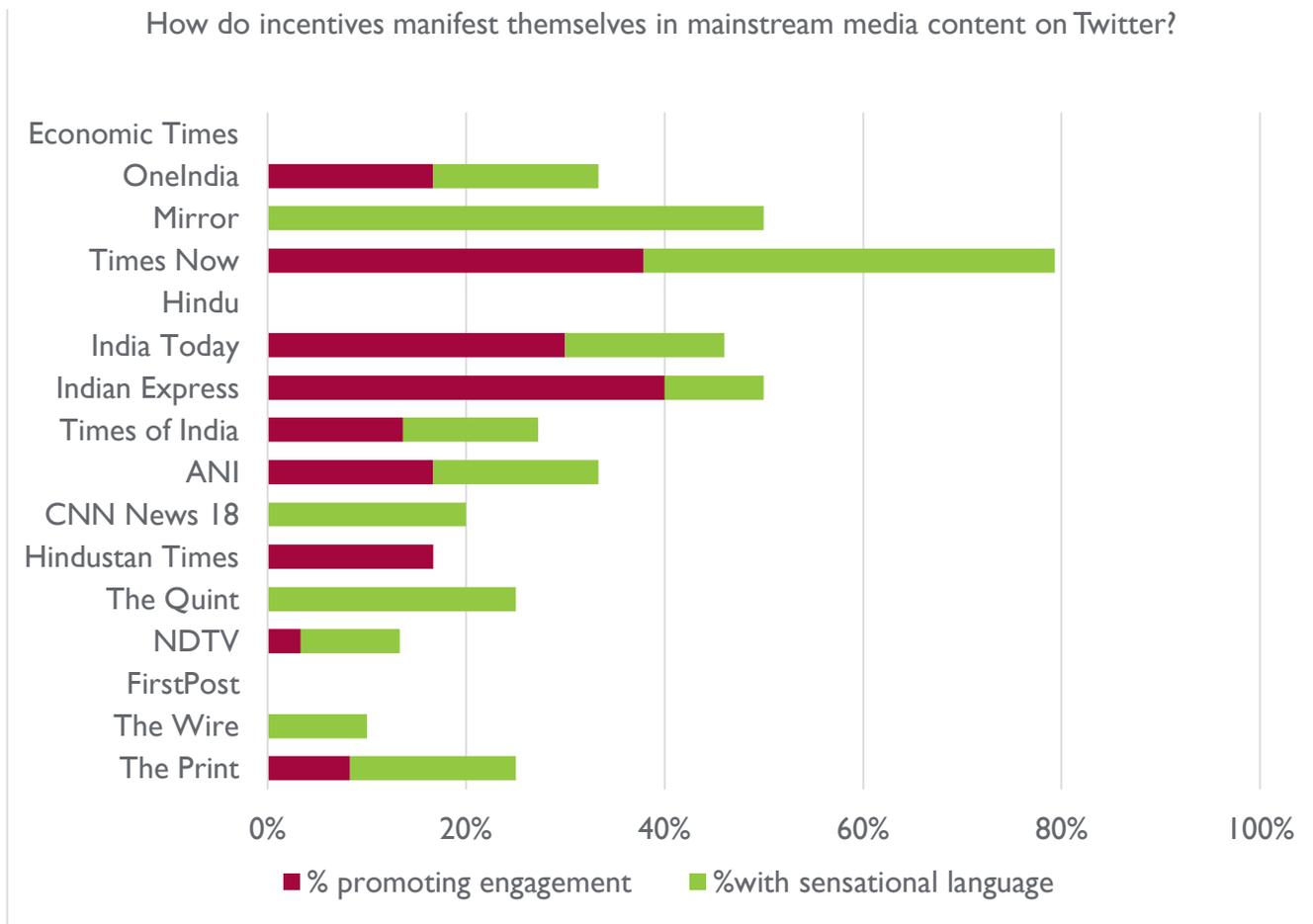


Figure 18: Mainstream Indian Media Sources and Their Susceptibility to Incentives (Analysis of Tweet Content Only) (Source: Own Analysis)

6.2.4 Mainstream journalists have the power to amplify manipulation and often do so unknowingly.

Mainstream journalists are fairly influential online i.e. they have huge followings and (often) the backing of a trusted media brand. When they engage with manipulated content, they draw significant attention to it. While they might intend to discredit or counter the manipulation, their engagement only furthers amplification.

- Example 1) Attempts to debunk potential misinformation (Figure 19): S Tanwar is an “other” user with ~640 followers who posted potential misinformation with regards to the Indian army (discussed [previously](#)). This post received significant attention when a prominent Pakistani journalist, Sameera Khan, responded negatively. Strategic silence—willingness to avoid amplifying extreme messages—would have been a more effective response⁶⁶ on Khan’s part.

- Example 2) The media’s use of polarizing terms, even if only sarcastically, furthers the manipulation:
 - ~40% of the top 15 mainstream media sources are seen using the polarizing terms selected for case #1. They do so in various ways (categories listed below in Table 13, differences across media sources shown in Figure 20).
 - While adopting the terms is more dangerous than mocking or simply reporting, any use of such language furthers manipulation and brings attention to the manipulators’ agenda. This was evident when past attempts (examples from 2018 in Figure 21) to discredit polarizing terms failed to eliminate them from the narrative (as of early 2020).



Figure 19: Conversation between Journalist @sameerakhan and "Other" User @sushtany (Source: Twitter)

Table 13: Ways in Which Mainstream Media Sources Use Polarizing Terms

#	Category	Definition
1	Adopting	Uses terms in their reporting without sarcasm
2	Adopting (mocking)	Uses terms in their reporting with sarcasm
3	Reporting	Uses terms to report on newsworthy events where the terms have been used

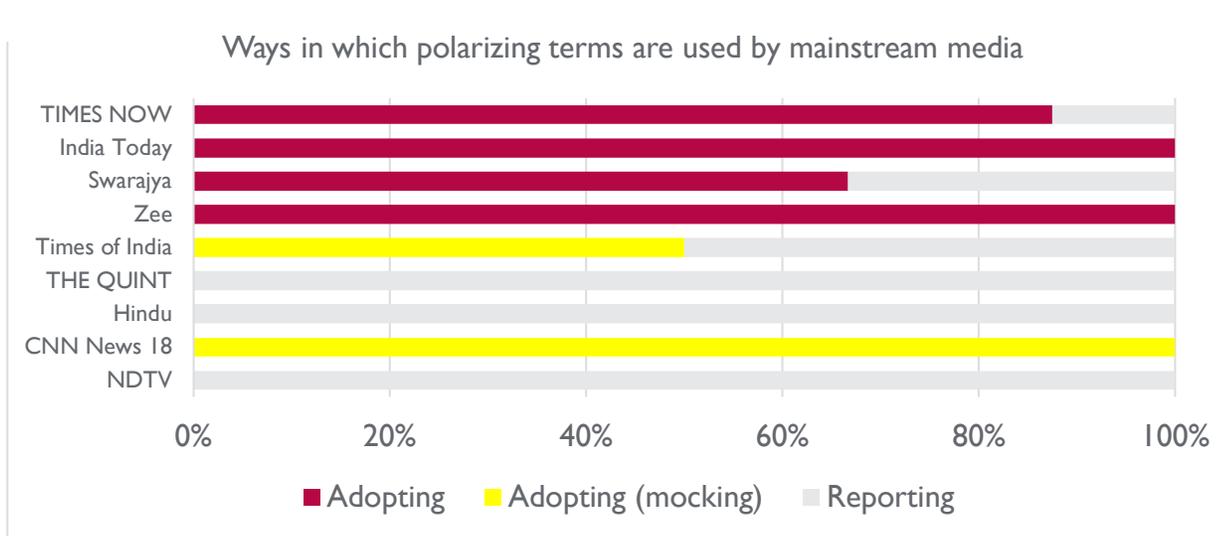


Figure 20: Ways in Which Mainstream Media Uses Polarizing Terms in their Coverage (Analysis of Tweet Content Only) (Source: Own Analysis)

NDTV LATEST LIVE TV VIDEO INDIA WORLD OPINION CITIES OFFBEAT TRENDS

Home > All India > #MeTooUrbanNaxal Trends On Twitter After Activists Raided, Arrested

READ IN OTHER LANGUAGES

#MeTooUrbanNaxal Trends On Twitter After Activists Raided, Arrested

The term "urban naxals" was condemned by activists, lawyers and Dalit leaders in a joint statement which said the expression was invented to "stifle any criticism of the government"

All India | Edited by Nidhi Sethi | Updated: August 29, 2018 04:36 pm IST

ThePrint

POLITICS GOVERNANCE ECONOMY DEFENCE INDIA FEATURES OPINION EVENTS VIDEO MORE

Home - Politics - He's making a list of 'Urban Naxals', but who is Vivek Agnihotri?

He's making a list of 'Urban Naxals', but who is Vivek Agnihotri?

SONIYA AGRAWAL 29 August, 2018 4:51 pm IST

Figure 21: Examples from 2018 Where Mainstream Media Sources Used Polarizing Terms with the Intention of Reporting or Mocking Their Message. (Source: NDTV, The Print)

6.3 Repercussions of Manipulative Behaviors Online

Twitter’s actions play an important role in managing manipulative behavior online. Actions taken can be of various forms—content deletion or downranking, account suspension or deletion, etc.^{qq}

Key Findings

- There is little visibility into if, when and how Twitter enforces its manipulation-related policies.
- Where information is available, it is evident that Twitter fails to enforce these policies on numerous occasions.
- New policies are promising but remain ambiguous and/or insufficient.

6.3.1 *There is little visibility into if, when and how Twitter enforces its platform manipulation policy.*

- Data to monitor platform manipulation policy violations is limited for external researchers. The platform manipulation policy is designed around account behavior (indiscriminate following, duplicating followers, aggressively liking content, etc.) and not specific types of content. Such information is limited in the data obtained from the API. From the outside, it is, therefore, challenging to tell if accounts are violating this policy and facing repercussions. Particularly in the case of “other” users, most of the manipulation is likely to come from account behavior and amplification as compared to influencing the content in narratives.
- Twitter’s transparency efforts are limited and lack granularity. First, the platform manipulation report⁶⁷ only provides aggregate information about violations (no breakdowns by geographies and types of issues are provided). Second, unlike other countries⁶⁸, there have been no data releases with regards to information

operations in India, despite there being documented cases of inauthentic behavior^{69rr}. Third, there is no visibility into if, when and how Twitter pro-actively monitors suspicious behavior online^{ss}. Given past exposés regarding leaked instructions for Twitter posts⁷⁰, allegations of coordination need to be investigated seriously.

6.3.2 *Where information is available, it is evident that Twitter fails to enforce its policies on numerous occasions.*

Numerous examples of forbidden behaviors were found, specifically in the cases of:

- Mimicking suspended accounts is commonplace. According to the policy, these accounts should be permanently deleted; instead, they are actively participating in the conversation. (Example in Figures 22).
- Calls for reciprocal follows do not seem to be punished. While some accounts that had such calls in their descriptions have been suspended, a majority of them remain active online. This implies that the accounts got

qq Not all behavior associated with media manipulation is (or should be) against platform policies. For example, using polarizing terms does not always meet standards of abuse or platform manipulation. Taking down such content would be a violation of free speech. In such cases, it is more important to monitor suspicious account activity to be aware of manipulation taking place on the platform and to be ready to respond if/when things escalate.

rr The references here are to behaviors on other platforms; it is very likely that similar behavior exists on Twitter as well. It has either gone undetected or, more likely, unreported under the transparency efforts.

ss Twitter cannot completely reveal its methodology for tracking suspicious behavior as bad actors are likely to then game the system or find loopholes. However, there is scope for increased transparency beyond what is followed now.

suspended for other reasons (not calls for reciprocal follows)^{tt} (Example in Figure 23).

6.3.3 New policies are promising but remain ambiguous and/or insufficient.

- New policies like that on fabricated media can help reduce some of the manipulative behaviors observed in this report (e.g. the morphed tweet screenshots about Shehla Rashid are likely to violate this new policy).
- However, the new policy takes action against the content, and not the offender. There are several examples of influential people who have repeatedly shared

disinformation online (e.g. username: AQUIBMIR7, referenced in the [Appendices](#)). Under the new policy, such actors will continue spreading disinformation and will likely face no repercussions^{uu}.

- The emphasis on conversation “health” is also promising. Metrics like shared reality and receptivity (referenced [earlier](#)) can reduce polarization and propaganda. However, it is too early to say how these policies will be designed and, more importantly, enforced.



Figure 22: Example of an Account Mimicking a Suspended Account Being Operated Freely Online (Source: Twitter)

Note: As can be seen in the user description, @DostKhan_Jammu is created to mimic a suspended account (@Prof_Hariom). This new account is active: it has 5000+ followers and had numerous instances of user engagement (including 300+quotes, 150 mentions in the dataset for case #2).



Figure 23: Examples of Active Accounts Making Explicit Promises of Reciprocal Follows in Their Descriptions (Source: Twitter)

^{tt} Twitter has not clarified whether enforcement of this policy is linked to evidence of follow backs as compared to claims in user descriptions. In the case of the former, these behaviors might not violate Twitter’s existing policy.

^{uu} While intention to manipulate cannot be established, policies should be developed to educate, warn and/or downrank repeated offenders.

7 RECOMMENDATIONS

This section provides recommendations for Amnesty International India. If implemented, they can increase awareness and reduce manipulation and consequently improve the state of online news and political discourse. The actions are organized in response to the [key findings](#) presented in the previous section.

7.1 Adopt appropriate language

Target Issue	Finding #1. There is evidence of media manipulation.
Recommendation	Adopt appropriate language to raise awareness.
Scope for Impact	Medium (Increasing awareness will likely deter manipulative behaviors and increase pressure on platforms to take action)
Effort Involved	Low
Explanation	<p>In times of crisis, the right vocabulary is very important to convey the urgency and gravity of the problem. This approach has been adopted by organizations like The Guardian with respect to other global challenges. For example, they specifically use the term “climate crisis” and not “climate change”⁷¹.</p> <p>Amnesty must develop internal protocols to adopt the right terms to describe media manipulation movements and to train others to do so as well. This will help increase awareness, deter such behavior and pressure platforms to act. Examples: Explicitly call out online manipulation attempts when they see them and label sources, amplifiers, and tactics; refer to government-led manipulation attempts as information operations.</p>

7.2 Support the creation of media evaluation tools

Target Issue	Finding #2. The mainstream media amplifies manipulation and degrades the quality of online discourse.
Recommendation	Support the creation of media evaluation tools to provide information on media bias and news quality.
Scope for Impact	High (Information on bias/quality can help citizens make informed news-related choices)
Effort Involved	High
Explanation	<p>At present, there is no one-stop solution where individuals can get reliable information on Indian media bias and news quality. Resources such as those provided by Reporters without Borders and Mediabiastfactcheck.com are incomplete or fragmented. Homegrown organizations like AltNews and BOOMFactCheck still run on a small scale.</p> <p>Amnesty, with its globally trusted brand and its mandate to protect human rights, can play an important role in supporting the creation of such tools and databases. Specifically, it can identify the right partners—nonprofits, academic institutions, etc.—to support and it can assist in the fundraising process.</p>

7.3 Train journalists

Target Issue	Finding #2. The mainstream media amplifies manipulation and degrades the quality of online discourse.
Recommendation	Train journalists to reduce instances of inadvertent amplification.
Scope for Impact	Medium (Only helps in cases where manipulation is amplified unknowingly)
Effort Involved	Medium
Explanation	<p>Without press coverage, manipulation efforts will fail to create impact. Training will help equip journalists with both the skills to recognize manipulation attempts and the decision-making tools to determine how to respond. Note, in the cases of conscious manipulation, likely due to media incentives, training will not be valuable.</p> <p>Amnesty can convene Indian journalists, host trainings and start a dialogue about manipulation. It can utilize journalist training resources provided by researchers in topics of how to identify false information, verify newsworthiness and be strategic about which stories to amplify⁷².</p>

7.4 Advocate for social media platforms to be more transparent

Target Issue	Finding #3. There is little visibility into if, when and how Twitter enforces its platform manipulation policies.
Recommendation	Advocate for social media platforms to be more transparent.
Scope for Impact	Low (Prior efforts for platform advocacy have not created meaningful change)
Effort Involved	Medium
Explanation	<p>Knowing how platforms enforce their policies is key to understanding the nature of the discourse and the scale of the associated manipulation. Today, platform transparency efforts are very limited (discussed previously).</p> <p>Amnesty should advocate for Twitter (and other platforms) to:</p> <ul style="list-style-type: none"> • Release datasets on information operations in India • Allow external auditors and researchers to access platform data • Provide more granular data on policy enforcement statistics • Justify cases where policies have not been enforced

8 LIMITATIONS

Limitations arose from technical, resource and data-related constraints. Some examples have been presented in the [scope](#) section of this document. Additional limitations are presented here. The [next section](#) covers ways to address some of these limitations in the future.

- Insufficient analysis to establish intent to manipulate: User intent to manipulate cannot be determined at scale solely from tweets (given that they are only 280-characters long and often lack context) and their associated information. As explained in previous sections, intent to manipulate has been inferred based on a variety of factors in this study. Making definitive claims requires collecting qualitative information through user interviews and examining tweet behavior in further detail.
- Insufficient analysis to quantify the impact on the online and offline worlds: This research describes and analyzes phenomena but is unable to establish their impact. Without speaking more definitively about how/which individuals are affected more than others, how things change as one moves across the political spectrum and what the real-world impacts are, it is challenging to understand the relevance of this research in the broader context of Kashmir, political discourse and polarization in India.
- Challenges of running analysis at scale: There are limited tools available for social media data analysis at scale. There were numerous tasks in this study where existing tools were insufficient/ human intervention was required.

Example) User categorization is a subjective process and requires human effort to look at user characteristics, tweets, their LinkedIn pages/ news presence where relevant, etc. This is a time-consuming process and there is no tool that can accurately complete the task. Given time limitations, only accounts with some influence were categorized and analyzed. To get a holistic view of the conversation, one should consider the role of all users, including those with no influence^w.

- Limited data available for download: Twitter's API restricts downloads to 10% of relevant data. The methodology around the selection of this 10% is not shared with developers/researchers. If the sample is not random, there could be biases introduced at the outset and the analysis might not be representative of the English content on Twitter.

^w Doing so would help better understand manipulation tactics used by "other" users (largest % of uninfluential users) and evaluate hypotheses of coordinated behavior.

9 SUGGESTIONS FOR FUTURE RESEARCH

This research can be taken forward in several directions. A few examples are provided below. They have been classified into the following groups based on their purpose:

- *Establishing causality and impact*
- *Delving deep into specific areas of research*
- *Broadening the scope of this research*

9.1 Establishing causality and impact

- An analysis of the causality behind manipulative behaviors is important while developing effective recommendations. Interviews covering a sample of people involved in the discourse can shed light on the pressures faced by various groups and the incentives behind/ consequences of their behavior. Appropriate interventions can then be designed. For example, those with financial motivations must be dealt with differently as compared to those who are driven by ideology.
- Research linking real-world impact will help establish the importance of this study as a whole. It is critical to understand how the state of the information ecosystem impacts people's lives on the ground. Hypotheses of impact include changes in voter behaviors and decisions, changes to resources allocated to Kashmir and general impact on daily life in the region. Including Kashmiri voices is also an important part of understanding the impact. Given the internet lockdown in Kashmir, a tiny

^{ww} Methodologies could include a time-based study to understand how user descriptions or behaviors of people followed by politicians have changed. If major changes occurred over short periods of time, there could be evidence for both coordinated behaviour and violations of platform policies.

proportion of tweets come from those directly affected by the problem. As mentioned in previous sections, Kashmiri media is also severely restricted. All of these suggestions can be studied through the collection and analysis of qualitative data.

- Further exploration of online behavior will also help establish real-world impact. One must understand who the audiences of manipulative content are, how they interact with each other in networks (e.g. Who forms Vivek Agnihotri's network and how tightly is it knit? Who are the followers of the low-quality new online media sources?) to better understand the level of polarization in society and the likelihood that people operate with the same facts.

9.2 Delving deep into specific areas of research

- An analysis of the likelihood of coordinated behaviors among "other" users, specifically the prevalence of the list below could help provide more clarity on the tactics used.
 - Reward follows and the similarities across users who receive them
 - Documentation of tiers of follows and the differences among them⁷³
 - Similarities in profile information and changes over time^{ww}
 - Documentation of behavior over time for suspended users
- A study of identity-related narratives of threat, specifically those associated with Kashmiri Hindu-Muslim tensions would help understand the impact of this wedge issue on discourse^{xx}.

^{xx} The methodology adopted could include analyzing tweets based on specific keywords (e.g. tweet includes "#KashmiriPandit", "proud to be a Kashmiri pandit" in the user descriptions) or influential users from the Kashmiri Pandit community.

- An estimate of the impact of celebrity engagement on public opinion of Kashmir is important. Celebrities show up as an influential category in this study. Vocal celebrities warrant further research in terms of their networks of followers, connection to the region and incentives. Studying silent celebrities is also key, given past instances of celebrity involvement in coordinated behaviors to support the BJP⁷⁴/Modi's use of celebrities to boost PR⁷⁵.
- Detailed research into the media ecosystem will help establish how consumer views change across the political spectrum and the role that social media plays (as compared to mainstream media). Research by Yochai Benkler and others has shown that, in the US context, the asymmetric nature of the media—prominent media on the left is well-distributed across the center, center-left and left while, on the right, prominent media is highly partisan⁷⁶—plays a central role in driving polarization (not social media platforms)⁷⁷. Some similarities have been observed in an anecdotal manner through this research. For example, several of the new, hyper-partisan online news sources cater primarily to the right. A spin-off study that replicates his research for the Indian context could be very valuable.
- An analysis of the challenges that Twitter faces in moderating content and determining manipulation is key to understanding the extent to which such behavior can be managed online. For example, inauthentic behavior is relatively easy to detect/ prove, and moderation can be run at scale by algorithms. On the other hand, abuse (e.g. use of hate speech) depends on the context and intent and, more often than not, requires human

moderation. An understanding of such nuances will help in the development of practical recommendations/ regulations. Note: this research would need to be done in collaboration with Twitter given the lack of availability of data for external researchers. While occurrences are rare, there are some precedents for joint research between platforms and academic institutions like the Berkman Klein Center at Harvard and the MIT Media Lab.

9.3 Broadening the scope of this research

- A study of media manipulation in Kashmir-related discourse in Pakistan is essential. This study focusses on Indian individuals and media. However, Kashmir is a bilateral issue. Most international actors with influence (and several influential news sources) are Pakistani. Imran Khan seems to have a strong following on Twitter and many users pledge allegiance to him as they do for Modi (though at a smaller scale). Media manipulation tactics employed by those in one country can affect or be targeted at those across the border as well.
- A broader study of media manipulation within India would help understand the manipulation attempts made by different groups across the political spectrum. For example, the inclusion of terms to demonize the right (like “Godi media”⁷⁸ or “chowkidar”) as well as other tactics and messaging would be helpful in terms of exploring the phenomenon as a whole.

An analysis of the relationship between media ownership and manipulation attempts would help evaluate the need for independent news.

⁷⁷ Methodologies could involve examining celebrity tweets (or Modi's tweets about celebrities) over a fixed period and determining the impact on Kashmir discourse.

10 APPENDICES

10.1 Rationale and Limitations Associated with the Data Filtering Criteria

Table 14: Data Collection Criteria

#	Criterion	Rationale and Limitations
1	Tweets created between 25 October and 31 December 2019	<ul style="list-style-type: none"> • <u>Rationale for choosing said dates:</u> Accurate data could not be collected retrospectively given the limitations on the Twitter API. Given the weeks it took to select the topic of the research, data collection could only begin from 25 October 2019 onwards. Data collection was stopped on 31 December 2019 as sufficient data (~3.5 million tweets) had been collected. • <u>Questions around relevance in Q4 2019:</u> Given that major policy changes happened in Kashmir in August, data collection starting in October could be a disadvantage. By then, there were numerous other issues gaining focus in India including the Citizenship Amendment Bill and the Ayodhya verdict. However, with ~3.5 million tweets collected, it is evident that the Kashmir conversation was still relevant in the news.
2	Tweet contained the word “Kashmir” either in the text or hashtag	<ul style="list-style-type: none"> • <u>Insufficient filter:</u> There were several thousand tweets daily that referred to Kashmir without explicitly using the word. All of these were left out with the chosen filter. While this could potentially introduce bias, it is unlikely that there is a statistically significant difference between content that includes the word “Kashmir” and that which does not.
3	Tweet contained one or more of the 83 popular hashtags and/or 251 user mentions (list provided below).	<ul style="list-style-type: none"> • <u>Rationale for choosing hashtags/ mentions:</u> Hashtags and users mentions helped narrow down the data collection to include the relevant conversations and influential users on the platform. • <u>Potential bias through social media filter bubbles:</u> Selecting these values could introduce bias, particularly given that they were chosen from the researcher’s personal Twitter feed. In order to compensate for this potential bias, a conscious effort was made to include users and content from all sides of the political and media spectrum. Additionally, daily Twitter trends for India⁷⁹ were used as a reference.

10.2 Hashtags and User Mentions Used for Data Filtering

Table 15: Hashtags Used for Data Collection

#Kashmir	#KashmiriPandits	#HindustanZindabad
#KashmirNeedsAttention	#Valley	#DemocracyDiesInDarkness
#IndianArmy	#Bharat	#ModiHitler
#PakistanArmy	#LibFarts	#PakArmyWillFight4Kashmir
#Modi	#JNU	#ISPRourProud
#StandwithKashmir	#MotaBhai	#IAmKashmir
#Pakistan	#Article35A	#PakistanSacrificesForPeace
#India	#Article370	#Love_ISI
#KashmirIssue	#370Abolished	#WarToLiberateKashmir
#WarToLiberateKashmir	#AjitDoval	#EndialsEnemyOfPeace
#GoToPakistan	#kashmirparfinalfight	#Defenders_OF_Pakistan
#Chowkidars	#Pseudosecularists	#amnestyasia
#TukdeTukdeGang	#DynastyMuktBharat	#SoniaExNAC
#JaiHind	#HowdyModi	#NegativityGang
#Suprabhat	#ModiMadeMandi	#ArrestShehlaRashid
#KashmirWantsFreedom	#OperationPoK	#GOI
#Kashmiri	#antinationaI	#HinduRashtra
#NeelumValley	#Thanks_PM_MODI	#TheKhanmarketgang
#jammuandkashmir	#Jai_GoMata	#LutyensMedia
#srinagar	#Modiji	#ModiHaiToMumkinHai
#terroristnationpakistan	#Vikaas	#WarAndPeaceFakery
#antimodi	#NewIndia	#KhanMarket
#friendofpakistan	#modiahaitomodihai	#jihadis
#UrbanNaxals	#Hindus	#BJP
#Pseudoliberal	#SabkaSaath	#Art370
#Art35	#RahulBacksPakistan	#tukdeOpposition
#KhanMarketGang	#PakistanTerrorState	#Article370Revoked
#ShahFaesal	#ShehlaRashid	

Table 16: User Mentions Used for Data Collection

@ashoswai	@aajtak	@Sea_Eagel
@kanhaiyakumar	@TimesNow	@SheikhAnas_Buny
@saket71	@ndtv	@Marwan_Ibrahim0
@arjunsethi81	@ReutersIndia	@sadafshakil2
@UmarKhalidJNU	@thewire_in	@ik302_com1
@narendramodi	@RahulGandhi	@Sameerk92483554
@nsitharaman	@jkmparty	@TeamPVFofficia
@AmitShah	@jkpdp	@ShOalb_SaNdHu
@rsprasad	@mlalangate	@its_Abbasi
@free_thinker	@MehboobaMufti	@HamidZeerak
@AzaanJavaid	@usman_majid1	@Defenders_Of_PK
@RegalJha	@ShashiTharoor	@Blocah222
@ahmedalifayyaz	@shahfaesal	@Ali_MuhammadPTI
@khanumarfa	@OmarAbdullah	@BhittaniKhanannn
@DilliDurAst	@PiyushGoyalOffc	@myogiadityanath
@UnSubtleDesi	@OfficeOfRKSingh	@Kishan561
@dhruv_rathee	@ArvindKejriwal	@sp_dash68
@RizviUzair	@kavita_krishnan	@OfficialDGISPR
@zainabsikander	@ImranKhanPTI	@pid_gov
@TheDeshBhakt	@rhtbapat	@fawadchaudhry
@Nidhi	@sushantsareen	@ArifAlvi
@ahmermkhan	@VivanMarwaha	@PTIofficial
@ShujaUH	@RanaAyyub	@Kashifabbasiary
@TabeenahAnjum	@AnupamPKher	@CMShehbaz
@sardesairajdeep	@OmarWaraich	@ShkhRasheed
@BDUTT	@UNWatch	@wasimakramlive
@sagarikaghose	@BJP4JnK	@Payal_Rohatgi
@vikramchandra	@shangpal	@ANI
@AmolSharmaWsj	@SouleFacts	@iSedLad
@SachinKalbag	@Shehla_Rashid	@WIONews
@madversity	@taslimanasreen	@gsurya
@cricketwallah	@kashmir787	@ArindamMandal34
@Kanchangupta	@Jalalmughal	@MonaAmbegaonkar
@Rahulkanwal	@farooq_pm	@ani_digital
@sudhirchaudhary	@Fkdotpk1	@amitbehere
@ShereenBhan	@javerias	@HashTagCricket
@ShekharGupta	@peaceforchange	@BharathMODIfied
@fayedsouza	@ForeignOfficePk	@MPLadakh
@AdityaRajKaul	@OfficialDGISPR	@rajnathsingh
@NehaSharma_BBC	@ResistToExist	@ZulfiqarJKPNA
@bukharishujaat	@MirMAKOfficial	@doshichetan4u
@SahirShafat	@kashmirglobal	@DevineniHamsa
@listenshahid	@Nira_Sayaapa	@meerasmusings
@spvaid	@Turkey_Istanbul	@MrcuteAnish
@KashmirPolice	@MushaalMullick	@OmMathur_bjp

@hussain_imtiyaz	@Masroor96241356	@vinitgoenka
@adgpi	@OnlineRajan__	@ashwani_mahajan
@i_m_yuvraj	@mona_sez	@amitkalraj
@ArvindVishwak10	@AB_BJP	@JagratiShukla29
@cosmicprakas	@ReallySwara	@SharmaKhemchand
@RKTrehan2	@hindustanse	@ShobhaBJP
@Maheshbhatt68	@PrachiBJP	@nishants79
@InvincibleBabu	@YRDeshmukh	@gopalkagarwal
@AliSahin501	@ToTheePoint	@satyakumar_y
@ImranGhazaliPK	@vivekagnihotri	@kaminirupani
@iamhamzaabbasi	@kakar_harsha	@ChitnisArchana
@AJListeningPost	@HMOIndia	@drsbmantribjp
@nazir_lord	@DeShobhaa	@seemasirohi
@RisingKashmir	@vinodpalan2	@Ayperilove
@AfzalAfzalmalik	@jiks	@DOP_Captain
@iKarachiwala	@FeminismInIndia	@ASherSial
@RGizbertI	@sabr_azad	@RakeshSinha01
@MirzaWaheed	@republic	@mssirsa
@HarrisRichard77	@kush07	@BJP4Delhi
@Stephen_G_Brown	@Prashaforever	@RubikaLiyaquat
@TahaSSiddiqui	@bhuvikal	@t_d_h_nair
@AmbNaghmanaHash	@vschanna	@PrakashJavdekar
@DaniyalFaridi	@TrulyMonica	@BouncingVodka
@pradip103	@Tukdetukdegang	@JhaSanjay
@Republic_Bharat	@Indiansontop	@BJP4India
@ashokepandit	@hemirdesai	@INCIndia
@madhukishwar	@KaushikDesai16	@vanitajain21
@payyaboy	@rakeshojjha	@Ramesh_BJP
@IasAlok	@DrRSingh1969	@MODifiedTamilan
@bhupendrachaube	@ubirakesh	@HardeepSPuri
@jyotsnavarma9	@Amaresh45814810	
@navikakumar	@DynastyMukt	
@RatanSharda5	@castelesshindu	
@SwarajyaMag	@vagishasoni	
@i_theindian	@shehzad_ind	
@pretypadmaja	@lbne_Sena	
@srikanthbjp_	@Sunariwal_MLA	
@Satyanewshi	@KapilMishra_IND	
@AmitShahOffice	@MajorPoonia	
@RoshanSdrprop	@AlokTiwari9335	
@RituDreams	@SanjayAzadSIn	
@AshBiswal	@ManojTiwariMP	
@Vasudev00411816	@AsYouNotWish	

10.3 Notes on Influence across Conversations and User Categories

As seen in Tables 17 and 18, tweet and influence behavior vary across both case studies. Compared to case #2 (pro-govt. propaganda), the conversation in case #1 (polarization) shows the following characteristics:

- User Mix: Appears to be more domestically focused (fewer international actors) and has larger proportions of “other” and entertainment industry users.
- Tweets: Has less activity from academics etc. and media, and more from “other” users.
- Influence: Is overly influenced by the entertainment industry.

General notes on the various categories:

- Academics etc. put out a significant number of tweets in the conversation on propaganda but fail to enjoy proportionate levels of influence. In the polarization conversation, their influence comes largely from being quoted^{zz}. Influencers like Shehla_Rashid and tonyashai are particularly prominent in both conversations^{aaa}.
- The entertainment industry has stayed largely silent in both conversations, particularly in the case of propaganda^{bbb}. There are 7x more participants from this category in the polarization conversation (as compared to that on propaganda^{ccc}). Influence for this category largely comes from mentions (except for “super influencers” who participate actively).
- Media: There are 226 unique media organizations and 67 independent journalists (or those from unspecified organizations) found in the dataset. The media behaves differently across conversations: In the polarization conversation, they barely tweet; their influence stems from mentions. In the propaganda conversation, the number of tweets goes up significantly and their influence comes from quotes and retweets.
- Political actors behave differently depending on their role in the party. Leaders like Modi, Shah and Mehbooba Mufti^{ddd} (Former Chief Minister for J&K) stay relatively silent. The loudest voices include party spokespeople or commentators as well as party workers^{eee}.
- International actors enter conversations through mentions. Their (active) participation is close to nil; the few tweets put out on topics of propaganda came largely from Pakistani government actors.
- “Other” users make up the largest group, tweet more than average users and have relatively low influence. While some of the top influencers are ordinary individuals with 100+ followers, others

^{zz} Exceptions are Ajaat Jamwal (SaveSabrimala), president of the JK Unity Foundation ([website](#)), and Madhupurnima Kishwar (madhukishwar), professor in the Indian Council of Social Science Research. Both actively tweet (and get retweeted on this topic).

^{aaa} While siddharthagoo appears influential, his influence in this context comes from having been quoted by “super influencer” vivekagnihotri.

^{bbb} This is expected: most celebrities have stayed silent on the issue of Kashmir, just like they have on other political issues.

^{ccc} In the propaganda conversation, there are only 10 entertainment industry participants. 7 of them have shown support for the BJP in the past.

^{ddd} Mufti was under house arrest for most of the period of analysis; as of the time of writing, her account says it is managed by her daughter.

^{eee} Accounts of party workers like HUNARGUPTABJP and AQUIBMIR7 are important to note. HUNARGUPTABJP is an advocate in the J&K High Court and part of the human rights cell. He tweets ~200x more than the average political actor. AQUIBMIR7 seems to be a party worker who has significant influence and has been involved in documented cases of media manipulation ([Link](#)).

are accomplished in various fields and have 10,000+ followers. It is important to note that their influence could come from account behavior—artificially amplifying narratives through things they retweet or like—that are not captured in this dataset.

- Suspended etc. users are a small proportion of the group. They are likely to have broken Twitter’s rules and have had action taken against them (e.g. high volume of tweets).

Table 17: Influence Rankings by Category for Case #1: Polarization (Source: Own Analysis)

#	Category	%Users	Avg. #Tweets	Avg. Influence
1	Academic/Author/Researcher/ Activist	6%	0.56	29.47
2	Entertainment Industry	4%	0.28	129.77
3	Govt. Actor	4%	0.05	14.45
4	International Actor	4%	0.01	6.48
5	Media	19%	0.09	26.32
6	Other	49%	1.34	6.96
7	Political Actor	12%	0.39	18.78
8	Suspended/Restricted/ Does Not Exist	2%	3.11	3.64

Table 18: Influence Rankings by Category for Case #2: Pro-Govt. Propaganda (Source: Own Analysis)

#	Category	%Users	Avg. #Tweets	Avg. Influence
1	Academic/Author/Researcher/ Activist	9%	0.78	19.50
2	Entertainment Industry	1%	0.09	4.82
3	Govt. Actor	4%	0.15	10.36
4	International Actor	16%	0.03	7.35
5	Media	25%	0.63	46.01
6	Other	34%	1.00	5.08
7	Political Actor	10%	0.25	39.97
8	Suspended/Restricted/ Does Not Exist	1%	1.00	5.55

II REFERENCES

- 1 Zeenab Aneez et al., “Reuters Institute India Digital News Report,” Reuters Institute for the Study of Journalism, 2019, <https://reutersinstitute.politics.ox.ac.uk/our-research/india-digital-news-report>.
- 2 Samantha Bradshaw and Philip N. Howard, “The Global Disinformation Disorder” (University of Oxford), 2, accessed October 21, 2019, <https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/09/CyberTroop-Report19.pdf>.
- 3 Anumeha Chaturvedi, “How India Emerged as Twitter’s Fastest Growing Market in Terms of Daily Active Users,” *The Economic Times*, May 16, 2017, https://economictimes.indiatimes.com/opinion/interviews/india-became-our-number-one-market-in-daily-users-twiters-new-india-director-taranjeet-singh/articleshow/58601906.cms?utm_source=contentofinterest&utm_medium=text&utm_campaign=cppst.
- 4 Aneez et al., “Reuters Institute India Digital News Report.”
- 5 “Kashmir Special Status Explained: What Are Articles 370 and 35A?,” *Al Jazeera*, August 5, 2019, <https://www.aljazeera.com/news/2019/08/kashmir-special-status-explained-articles-370-35a-190805054643431.html>.
- 6 Data & Society, “Media Manipulation,” *Data & Society* (blog), accessed January 22, 2020, <https://datasociety.net/research/media-manipulation/>.
- 7 Alice Marwick and Rebecca Lewis, “Media Manipulation and Disinformation Online,” *Data & Society*, n.d., 24–26.
- 8 Bradshaw and Howard, “The Global Disinformation Disorder,” 13.
- 9 Marwick and Lewis, “Media Manipulation and Disinformation Online,” 27–32.
- 10 Judith E. Rosenbaum and Jennifer Bonnet, “Looking Inward in an Era of ‘Fake News’: Addressing Cognitive Bias,” *YLAJ Network*, accessed April 2, 2020, <https://ylai.state.gov/looking-inward-in-an-era-of-fake-news-addressing-cognitive-bias/>.
- 11 Julian Matthews, “A Cognitive Scientist Explains Why Humans Are so Susceptible to Fake News and Misinformation,” *Nieman Lab* (blog), April 17, 2019, <https://www.niemanlab.org/2019/04/a-cognitive-scientist-explains-why-humans-are-so-susceptible-to-fake-news-and-misinformation/>.
- 12 David Robson, “Why Smart People Are More Likely to Believe Fake News,” *The Guardian*, April 1, 2019, <https://www.theguardian.com/books/2019/apr/01/why-smart-people-are-more-likely-to-believe-fake-news>.
- 13 Bradshaw and Howard, “The Global Disinformation Disorder,” 1, 3, 4.
- 14 Twitter, “Platform Manipulation and Spam Policy,” September 2019, <https://help.twitter.com/en/rules-and-policies/platform-manipulation>.
- 15 Yoel Roth and Ashita Achuthan, “Building Rules in Public: Our Approach to Synthetic & Manipulated Media,” Twitter, February 4, 2020, https://blog.twitter.com/en_us/topics/company/2020/new-approach-to-synthetic-and-manipulated-media.html.
- 16 “Transparency Report,” Twitter, accessed March 29, 2020, <https://transparency.twitter.com/en.html>.
- 17 “Jack Dorsey: How Twitter Needs to Change,” TED, 2019, https://www.ted.com/talks/jack_dorsey_how_twitter_needs_to_change/discussion?
- 18 Paul Briggs, “eMarketer Forecasts for APAC Twitter Users and Growth, Including Japan and India Trends,” eMarketer, March 5, 2020, <https://www.emarketer.com/content/asia-pacific-drives-twitter-growth-as-other-regions-plateau>.
- 19 Chaturvedi, “How India Emerged as Twitter’s Fastest Growing Market in Terms of Daily Active Users.”
- 20 Chaturvedi.
- 21 Kuwar Singh, “Social Media Censorship in India Has Increased over Five Fold since 2016,” *Quartz India*, November 21, 2019, <https://qz.com/india/1753116/modi-government-ramps-up-twitter-facebook-censorship/>.
- 22 Pankaj Doval, “Twitter Executives Could Face 7-Year Jail, Warns Government,” *The Times of India*, March 13, 2019, <https://timesofindia.indiatimes.com/business/india-business/twitter-executives-could-face-7-year-jail-warns-government/articleshow/68383727.cms>.
- 23 Tara Varma, “The WhatsApp Election: The Fake News Challenge to Democracy,” *ECFR*, April 30, 2019, https://www.ecfr.eu/article/commentary_the_whatsapp_election_the_fake_news_challenge_to_democracy.
- 24 Archis Chowdhury and Boomlive.in, “Pulwama Attack: All the Fake News That Has Been Circulating on Social Media after CRPF Attack,” *Scroll.in*, February 18, 2019, <https://scroll.in/article/913740/from-old-videos-to-doctored-photos-a-barrage-of-fake-news-floods-social-media-after-pulwama-attack>.
- 25 Anumeha Chaturvedi, “Indo Pak Tension: Fake News & Misinformation on Indo-Pak Tension Flood The,” *The Economic Times*, February 28, 2019, <https://economictimes.indiatimes.com/tech/internet/fake-news-misinformation-on-indo-pak-tension-flood-the-social-media/articleshow/68194093.cms>.
- 26 IANS, “Why Twitter’s Rival Social Network Mastodon Trending in India,” *National Herald*, November 14, 2019, <https://www.nationalheraldindia.com/twitter/why-twiters-rival-social-network-mastodon-trending-in-india>.

- 27 Tech2 News Staff, "Twitter India Responds to 'bias' Charges," First Post, November 7, 2019, <https://www.firstpost.com/tech/news-analysis/twitter-india-responds-to-bias-charges-says-we-are-impartial-dont-take-action-based-upon-any-ideology-political-viewpoint-7618421.html>.
- 28 Aneez et al., "Reuters Institute India Digital News Report," 9–10.
- 29 Claire Teitelman, "Communal Violence, Social Media, and Elections in India | JIA SIPA," Columbia SIPA Journal of International Affairs, April 19, 2019, <https://jia.sipa.columbia.edu/online-articles/communal-violence-social-media-and-elections-india>.
- 30 Subhayan Mukerjee and Silvia Majo-Vasquez, "During the Indian Election, News Audiences Consumed a Wide and Diverse Range of Sources," *Nieman Lab* (blog), June 28, 2019, <https://www.niemanlab.org/2019/06/during-the-indian-election-news-audiences-consumed-a-wide-and-diverse-range-of-sources/>.
- 31 ABP News, "Amit Shah's Full Answer On Kashmir In Raja Sabha: 'Everything Is Normal,'" YouTube, November 20, 2019, <https://www.youtube.com/watch?v=YvzlfYvf7Y0>.
- 32 Shivam Vij, "Tukde Tukde Gang Is the Ultimate Fake News of the North Korean Media," *The Print*, April 25, 2018, <https://theprint.in/opinion/tukde-tukde-gang-is-the-ultimate-fake-news-of-the-north-korean-media/52326/>.
- 33 PTI, "'Award Wapsi' Gang Active Again to Defame Modi Govt, Says Giriraj Singh on 49 Celebrities' Letter to PM," *News18*, July 24, 2019, <https://www.news18.com/news/politics/award-wapsi-gang-active-again-to-defame-modi-govt-says-giriraj-singh-on-49-celebrities-letter-to-pm-over-lynchings-2244813.html>.
- 34 Mayank Bhardwaj, "'Khan Market Gang': Modi Mocks His Elite Adversaries," *Reuters*, May 31, 2019, <https://in.reuters.com/article/india-politics-khanmarket-idINKCNIT10KM>.
- 35 Rakhi Bose, "The Names of Dissent: Yesterday's 'Anti-National' Is Today's 'Urban Naxal,'" *News18*, July 24, 2019, <https://www.news18.com/news/buzz/the-names-of-dissent-yesterdays-anti-national-is-todays-urban-naxal-1864443.html>.
- 36 Digital Desk, "Negativity Gang Exposed: Kashmir Their Latest Agenda in Non-Stop Campaign to Run India down - Republic World," *Republic TV*, August 14, 2019, <https://www.republicworld.com/india-news/general-news/negativity-gang-exposed-kashmir-their-latest-agenda-in-non-stop-campaign-to-run-india-down.html>.
- 37 Tariq Engineer, "Narendra Modi: Who Does Prime Minister Narendra Modi Follow on Twitter?," *Mumbai Mirror*, September 9, 2018, <https://mumbaimirror.indiatimes.com/others/sunday-read/who-does-namo-follow-on-twitter/articleshow/65737881.cms>.
- 38 Engineer.
- 39 Alisha Sachdev, "How Fake News Spreads in India – The Diplomat," *The Diplomat*, April 9, 2018, <https://thediplomat.com/2018/04/how-fake-news-spreads-in-india/>.
- 40 Joyojeet Pal, "Banalities Turned Viral: Narendra Modi and the Political Tweet," *Television & New Media*, 2015, 5, <https://doi.org/10.1177/1527476415573956>.
- 41 Aria Thaker, "Modi's Twitter Strategy of Following Common Indians Worked — Quartz India," *Quartz India*, June 3, 2019, <https://qz.com/india/1630274/modis-twitter-strategy-of-following-common-indians-worked/>.
- 42 Aria Thaker, "This Is Probably Why Narendra Modi Even Follows Trolls on Twitter," *Text, Scroll.in* (<https://scroll.in>, June 5, 2019), <https://scroll.in/article/925754/this-is-probably-why-narendra-modi-follows-his-supporters-including-trolls-on-twitter>.
- 43 Thaker, "Modi's Twitter Strategy of Following Common Indians Worked — Quartz India."
- 44 Ajay Gudavarthy, "What Lies Behind Narendra Modi's Silence?," *The Wire*, October 25, 2018, <https://thewire.in/politics/narendra-modi-rafale-metoo-cbi>.
- 45 Joanna Slater, "India's Modi Has a Strategy for Dealing with Controversies: Silence," *The Washington Post*, November 28, 2020, https://www.washingtonpost.com/world/asia_pacific/indias-modi-has-a-strategy-for-dealing-with-controversies-silence/2018/11/27/ce7f6946-ed02-11e8-8b47-bd0975fd6199_story.html.
- 46 Joan Donovan and Brian Friedberg, "Source Hacking: Media Manipulation in Practice," *Data & Society's Media Manipulation Research Initiative* (Data & Society, September 2019), 9, https://datasociety.net/wp-content/uploads/2019/09/Source-Hacking_Hi-res.pdf.
- 47 Rahi Gaikwad, "Manmohan: Naxalism the Greatest Internal Threat," *The Hindu*, October 11, 2009, sec. National, <https://www.thehindu.com/news/national/Manmohan-naxalism-the-greatest-internal-threat/article16886121.ece>.
- 48 "Amit Shah Asks CRPF to Go after 'Urban Naxals and Their Facilitators,'" *The Hindu*, November 16, 2019, sec. National, <https://www.thehindu.com/news/national/amit-shah-asks-crpf-to-go-after-urban-naxals/article29990585.ece>.
- 49 TNIE Videos, "Vivek Agnihotri on Urban Naxals, JNU, Fake News and Twitter Trolls - YouTube," YouTube, March 15, 2019, <https://www.youtube.com/watch?v=mLmXRdgRoS8>.
- 50 Pal, "Banalities Turned Viral: Narendra Modi and the Political Tweet," 5.
- 51 Matt Peterson and Abdallah Fayyad, "The Irresistible Effectiveness of Wedge Politics," *The Atlantic*, December 8, 2017, <https://www.theatlantic.com/membership/archive/2017/12/the-irresistible-effectiveness-of-wedge-politics/547946/>.

- 52 Gary King, Jennifer Pan, and Margaret E. Roberts, “How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument,” *American Political Science Review* 111 (2017): 484–501, <https://doi.org/10.1017/S0003055417000144>.
- 53 Suchitra Vijayan and Vasundhara Sirnate Drennan, “After Pulwama and Balakot, the Indian Media Proves It Is the BJP’s Propaganda Machine,” *The Washington Post*, March 4, 2019, <https://www.washingtonpost.com/opinions/2019/03/04/after-pulwama-indian-media-proves-it-is-bjps-propaganda-machine/>.
- 54 Vishwa Deepak, “A Former Zee News Producer Reveals Why He Left over the Network’s Coverage of JNU,” *The Caravan*, March 17, 2016, <https://caravanmagazine.in/vantage/zee-news-former-producer-reveals-left-networks-coverage-jnu>.
- 55 Nitasha Kaul, “Kashmir: The Communalisation of a Political Dispute | India,” *Al Jazeera*, July 26, 2017, <https://www.aljazeera.com/indepth/opinion/2017/07/kashmir-communalisation-political-dispute-170725082030871.html>.
- 56 Eun Go, Hwa Jong Eun, and Mu Wu, “The Effects of Source Cues on Online News Perception,” *Computers in Human Behavior* 38 (September 2014): 358–67.
- 57 “SPJ Code of Ethics - Society of Professional Journalists,” *Society of Professional Journalists*, September 6, 2014, <https://www.spj.org/ethicscode.asp>.
- 58 Aneez et al., “Reuters Institute India Digital News Report,” 19.
- 59 “In the Age of New Media, Who Counts as a Journalist?,” *The Reporters Committee for Freedom of the Press* (blog), 2011, <https://www.rcfp.org/journals/the-news-media-and-the-law-winter-2011/age-new-media-who-counts-jo/>.
- 60 Joyojeet Pal and Dibyendu Mishra, “Polarity of News Sources among Indian Politicians,” February 29, 2020, <https://joyojeet.people.si.umich.edu/polarity-of-indian-news-sources-among-politicians/>.
- 61 Bilal Hussain, “Press Freedom in Kashmir: Local and International Journalists Under Pressure,” *The Diplomat*, March 12, 2020, <https://thediplomat.com/2020/03/press-freedom-in-kashmir-local-and-international-journalists-under-pressure/>.
- 62 Marwick and Lewis, “Media Manipulation and Disinformation Online,” 42.
- 63 “RSF Issues Warning to India in First World Press Freedom Index Incident Report | RSF,” *Reporters Without Borders*, July 4, 2018, <https://rsf.org/en/news/rsf-issues-warning-india-first-world-press-freedom-index-incident-report>.
- 64 Ellen Barry, “Raids in India Target Founders of News Outlet Critical of Government - The New York Times,” *The New York Times*, January 5, 2017, <https://www.nytimes.com/2017/06/05/world/asia/india-ndtv-raids-narendra-modi-prannoy-roy.html?mcubz=3>.
- 65 Raksha Kumar, “India’s Media Can’t Speak Truth to Power,” *Foreign Policy* (blog), August 2, 2019, <https://foreignpolicy.com/2019/08/02/indias-media-cant-speak-truth-to-power-modi-bjp-journalism/>.
- 66 Joan Donovan and Danah Boyd, “The Case for Quarantining Extremist Ideas | Joan Donovan and Dana Boyd | Opinion | The Guardian,” *The Guardian*, June 1, 2018, <https://www.theguardian.com/commentisfree/2018/jun/01/extremist-ideas-media-coverage-kkk>.
- 67 Frederic Filloux, “The Rise of Influencers and the Decay of Journalism,” *Medium*, October 5, 2019, <https://mondaynote.com/the-rise-of-influencers-and-the-decay-of-journalism-67978967de70>.
- 68 Yoel Roth, “Information Operations on Twitter: Principles, Process, and Disclosure,” June 13, 2019, https://blog.twitter.com/en_us/topics/company/2019/information-ops-on-twitter.html.
- 69 Rishi Iyengar, “Facebook India: Pages Linked to Congress Party Removed by Social Media Platform - CNN,” *CNN*, April 1, 2019, <https://www.cnn.com/2019/04/01/tech/facebook-india-congress-accounts-removed/index.html>.
- 70 Alt News Desk, “After Alt News Exposé, Google Doc with Trending Instructions for #DemonetisationSuccess Hashtag Is Deleted,” *Alt News* (blog), September 2, 2017, <https://www.altnews.in/alt-news-expose-google-doc-trending-instructions-demonetisationsuccess-hashtag-deleted/>.
- 71 Sophie Zeldin-O’Neill, “‘It’s a Crisis, Not a Change’: The Six Guardian Language Changes on Climate Matters,” *The Guardian*, October 16, 2019, sec. Environment, <https://www.theguardian.com/environment/2019/oct/16/guardian-language-changes-climate-environment>.
- 72 Danah Boyd, “Media Manipulation, Strategic Amplification, and Responsible Journalism,” *Medium*, September 21, 2018, <https://points.datasociety.net/media-manipulation-strategic-amplification-and-responsible-journalism-95f4d611f462>.
- 73 Thaker, “Modi’s Twitter Strategy of Following Common Indians Worked — Quartz India.”
- 74 Desk, “After Alt News Exposé, Google Doc with Trending Instructions for #DemonetisationSuccess Hashtag Is Deleted.”
- 75 Pal, “Banalities Turned Viral: Narendra Modi and the Political Tweet,” 5.
- 76 Rob Faris et al., “Partisanship, Propaganda, and Disinformation: Online Media and the 2016 U.S. Presidential Election,” *Berkman Klein Center*, August 16, 2017, <https://cyber.harvard.edu/publications/2017/08/mediacloud>.
- 77 Megan Cage and Henry Farell, “Blame Fox, Not Facebook, for Fake News,” *The Washington Post*, November 6, 2018, <https://www.washingtonpost.com/>.
- 78 “What Does ‘Godi Media’ Mean? - Quora,” accessed March 8, 2020, <https://www.quora.com/What-does-Godi-media-mean>.
- 79 “Twitter Trends FAQs,” accessed January 22, 2020, <https://help.twitter.com/en/using-twitter/twitter-trending-faqs>.