ISSN 2282-6483

Alma Mater Studiorum - Università di Bologna
DEPARTMENT OF ECONOMICS

The Spatial Drivers of Discrimination: Evidence From Anti-Muslim Fake News in India

Samira S. Abraham
Gianandrea Lanzara
Sara Lazzaroni
Paolo Masella
Mara P. Squicciarini

Quaderni - Working Paper DSE N°1180

The Spatial Drivers of Discrimination: Evidence From Anti-Muslim Fake News in India

Samira S. Abraham — University of Bologna
Gianandrea Lanzara — Bocconi University
Sara Lazzaroni — University of Bologna
Paolo Masella — University of Bologna
Mara P. Squicciarini — Bocconi University *

January 11, 2023

Abstract

This paper studies how discriminatory fake news arises and spatially diffuses. We focus on India at the onset of the COVID-19 pandemic: on March 30, a Muslim convention (the *Tablighi Jamaat*) in New Delhi became publicly recognized as a COVID hotspot, and the next day, fake news on Muslims intentionally spreading the virus spiked. Using Twitter data, we build a comprehensive novel dataset of georeferenced tweets to identify anti-Muslim fake news. We find, in cross-sectional and difference-in-difference settings, that discriminatory fake news became much more widespread after March 30 (1) in New Delhi, (2) in districts closer to New Delhi, and (3) in districts with higher social media interactions with New Delhi. Further, we investigate whether deeply rooted historical factors may have also played a role in the diffusion of anti-Muslim fake news: we show that, after March 30, discriminatory fake news was more common in districts historically exposed to attacks by Muslim groups.

JEL-Classification: J15, Z12
Keywords: discrimination, fake news, religion, covid, india

*We thank Mathieu Couttenier, Sophie Hatte, and Paolo Vanin for helpful comments. Financial support from the PRIN Grant # PRIN 2017ATLJHB is gratefully acknowledged.

Non-Technical Summary

Fake news (or false stories) spread rapidly on social media and the Internet, often distorting the users' views on several topics. Given the large diffusion of these technologies in recent years, concern is growing over the spread of false stories and their social, economic, and political consequences.

In this paper, we study the fake news phenomenon under a novel perspective, namely as a vehicle to propagate hate and discriminatory attitudes toward minorities. In particular, we study the diffusion of false stories against Indian Muslims at the onset of the coronavirus outbreak in India, exploiting a tight sequence of events on March 30, 2020 that led many to identify a Muslim religious congregation (the *Tablighi Jamaat* convention) held in New Delhi as a COVID-19 hotspot. This coincided with an outburst of false stories reporting that Muslims were deliberately infecting other people and associating the spread of the virus with a form of jihad conducted by Muslim communities (see the trending hashtag "#coronajihad").

To carry out the empirical analysis, we collected a comprehensive novel dataset of georeferenced text data from Twitter to document a large spike (from nearly zero to above 3%) in the share of tweets reporting false stories against Muslims and to investigate the spatial patterns of their diffusion.

Our econometric analysis delivers three sets of results. First, we find that, following the shock, the intensity of discriminatory fake news was strongest in New Delhi, where the Tablighi event took place. Second, beyond New Delhi, anti-Muslim false news was more pronounced in districts that are spatially closer and have more intense social media interactions with New Delhi, further highlighting the increasing role social networks play in diffusing discriminatory attitudes. Third, we show that the observed spatial differences in diffusion of fake news after the shock relate to the legacy of precolonial Muslim attacks.

We build a novel classification of precolonial conflicts—events in which Muslim entities were the aggressors versus events in which Muslim entities were not the aggressors—and we map these events at the district level. We show that the diffusion of fake news in the aftermath of the Tablighi shock was stronger in districts where Muslim attacks occurred compared to districts that did not experience historical conflicts. Using state-of-the-art text analysis techniques, we provide suggestive evidence that discriminatory users are disproportionately Hindu nationalists and individuals active in political debate.

These findings on present-day India suggest that (epidemic) shocks may affect a country's overall environment of discrimination through the spread of anti-minority false news on social media. This is especially relevant in the case of (location-specific) persistent beliefs regarding the role of minorities as possible threats to national security and well-being.

"It Was Already Dangerous to Be Muslim in India. Then Came the Coronavirus." (Perrigo, 2020)

1 Introduction

Fake news—sometimes called false stories—may likely have widespread, long-lasting political, social, and economic consequences for societies (Zhuravskaya et al., 2020). Evidence shows that fake news spreads on social media faster and to larger audiences than true stories, often distorting the views of the users who receive it (Vosoughi et al., 2018). Social media technologies, and the Internet more generally, have favored the rise of populism and extremist and xenophobic ideas in the United States and Europe (Bursztyn et al., 2019; Müller and Schwarz, 2020, 2021).

Scholars have largely focused on the role of misinformation in political contexts,[1] but to the best of our knowledge, fake news has rarely been studied as a vehicle for disseminating hate and discriminatory sentiments toward minorities. Thus, we still know very little about how anti-minority fake news is generated and how it spreads, either online or in the real world.

Where, and under what conditions, does disinformation about discriminated minorities originate? What determines its diffusion in space? What types of users create or circulate false stories about minorities? Answering these questions presents several empirical challenges: measuring discriminatory fake news, linking it to a precise geographic location, and, more generally, finding a suitable empirical context to study its spatial diffusion. To partially overcome these challenges, this paper focus on the social media platform Twitter to study the spatial determinants of anti-Muslim fake news[2] in India at the onset of the coronavirus pandemic, in 2020.

India is an ideal setting to carry out such analysis for three main reasons. First, India is an extremely diverse country with a history of tensions among religious communities, and its Muslim minority has been one of the communities most discriminated against.[3] Second, discussion about fake news and misinformation in India has grown louder in recent years, following the explosive growth in the number of active Internet users. (Nielsen estimates that India had 646 million active Internet users in December 2021, and Internet

[1] For instance, Allcott and Gentzkow (2017), Guess et al. (2018), and Grinberg et al. (2019) empirically study the impact of social media and fake news on the 2016 U.S. presidential election and show that fake news circulated virulently online during the election campaign, among a narrow group of users.

[2] In line with Guess et al. (2019), we define fake news as "false or misleading content intentionally dressed up to look like news articles."

[3] According to a 2019–2020 Pew Research Center survey of religion across India, one in five Muslims say they have recently faced religious discrimination (Pew Research Center, 2021).

penetration continues to increase in rural and in urban areas (Nielsen, 2022).) Third, at the onset of the coronavirus pandemic, a Muslim religious gathering—the *Tablighi Jamaat* convention in New Delhi—suddenly became a national COVID-19 hotspot—about 9,000 congregants and related primary contacts were sent to quarantine facilities or hospitals in a country that had seen only a handful of reported cases beforehand (Times of India, 2020).

In this context, we analyze a comprehensive novel dataset of georeferenced tweets, within which we are able to identify tweets reporting anti-Muslim fake news. Our empirical strategy exploits the timing of a sequence of events tied to the aforementioned religious gathering. In fact, though the Muslim convention took place from March 1 to 15, its alleged connection with the pandemic didn't become salient until the evening of March 30, when multiple deadly COVID-19 cases were reported among *Tablighi* participants and when several institutional actors blamed the convention for spreading COVID in India. For ease of discussion, we will henceforth refer to this bundle of events on March 30 as the *"Tablighi* shock."

In line with the literature on the scapegoating of minorities during epidemic outbreaks (Jedwab et al., 2019), we document a large spike in anti-Muslim fake news on March 31: the share of tweets reporting anti-Muslim false stories—around 0.1% of total tweets in the week preceding the *Tablighi* shock—jumped to 2.8% on March 31, then peaked at 3.3% a day later. The vast majority of false stories in our sample claimed that Muslim individuals were deliberately infecting other people, often drawing an analogy between the spread of the virus and the spread of religion, and asserting that these allegedly intentional infections were a form of jihad. The hashtag #coronajihad began trending strongly on March 31, 2020.

Based on this finding, we then investigate the spatial diffusion of anti-Muslim discriminatory behavior in the aftermath of the *Tablighi* shock. We observe substantial spatial heterogeneity in the diffusion of fake news. First, exploiting both the cross-district variation after March 30 and a difference-in-difference estimation strategy at the daily district level, we provide evidence that the intensity of discriminatory fake news is stronger in the state of New Delhi, where the *Tablighi* event took place. In particular, the difference-in-difference coefficient suggests that, following the shock, the number of tweets with anti-Muslim fake news increased by 144 more tweets in New Delhi compared to the rise in other districts. This result is robust to considering different distance thresholds to account for the role of spatial autocorrelation in statistical inference and to the use of the share of anti-Muslim fake news as an alternative dependent variable. Moreover, we document that the diffusion of anti-Muslim misinformation from the New Delhi hotspot was stronger in districts that are spatially closer to and have more intense social media interactions with the capital. This result suggests that both physical distance and social connectedness are key determinants for the spreading of fake news, pointing to the growing role social networks play in diffusing discriminatory attitudes.

Second, given the national relevance of the *Tablighi* shock, we follow a large literature on the persistence of discriminatory beliefs and ideological traits (e.g., Voigtländer and Voth, 2012) by investigating whether the observed spatial differences in the diffusion of fake news can be traced back to deeply rooted characteristics of Indian society. To do so, we exploit a comprehensive dataset of historical conflicts on the Indian subcontinent; we focus on land-based attacks initiated by Muslim entities in precolonial times (1000–1757). For the sake of brevity, we will henceforth refer to these events as ”Muslim attacks.” We find that the diffusion of fake news in the aftermath of the *Tablighi* shock was stronger in districts where such Muslim attacks occurred compared to districts that did not experience Muslim attacks. These results suggest that the historical experience of conflict with Muslims—and the simultaneous hostility that arose toward them—predicts current anti-Muslim discriminatory behavior, as captured by the the diffusion of anti-Muslim fake news.

We could be concerned that omitted variables, correlated with our measure of exposure to Muslim attacks, may be affecting the diffusion of anti-Muslim discriminatory fake news after March 30. To address this concern, we perform two sets of exercises. First, we show that our results hold if we use several alternative definitions of exposure to Muslim attacks, such as exploiting the intensity of exposure in terms of the number of Muslim attacks, using a distance-based measure of exposure as in Dincecco et al. (2022), restricting Muslim attacks to those occurring only after the advent of the Delhi Sultanate, and including also Muslim attacks that occurred up to 1840. Second, we show that our results are robust to controlling in the main specification for several other potential confounders (interacted by a *Post-March30* dummy), including further geographical controls, variables capturing historical state capacity and the local exposure to colonizers, and measures of linguistic and ethnic fractionalization.

Finally, we also find suggestive evidence that discriminatory users are disproportionately Hindu nationalists and individuals active in the political debate.

Literature. By focusing on the rise of anti-Muslim fake news after the coronavirus outbreak, this paper relates to a recent literature in economics showing that anti-minority behavior increases during economic and epidemic crises. In the context of economic crises, Doerr et al. (2021) show an increase in anti-Jewish laws during the Great Depression, while Anderson et al. (2017) and Anderson et al. (2020) show, respectively, an increase in anti-Jewish persecutions after colder growing seasons in the period spanning the years 1000 to

1800 and an increase in anti-Black discrimination during the Great Recession (2007–2009). In the context of epidemics, Jedwab et al. (2019) find that the Black Death of the 1350s triggered anti-Jewish persecutions in towns where people were more inclined to believe antisemitic allegations but not in towns where the activities of Jewish inhabitants were complementary to the local economy. Bartoš et al. (2021), Dipoppa et al. (2021), and Lu and Sheng (2022) show, either experimentally or by using a difference-in-difference strategy, that the ongoing coronavirus pandemic has increased discrimination against the Chinese minority in the Czech Republic, Italy, and the United States, respectively.

While this literature establishes a causal link between various shocks and increases in discrimination, it falls short in showing hard evidence on the scapegoating mechanism at play and in discussing the spatial diffusion of discrimination. By analyzing the content of the tweets in our sample, we are able to pinpoint the identification of Muslims as a threat to India either through terrorist acts or as intentional spreaders of the virus. Our text analysis of the tweets confirms that the fear of contagion is the most important underlying scapegoating mechanism. Moreover, we show that anti-Muslim false rumors are stronger in New Delhi and in districts where historical Muslim attacks took place, and that they spread throughout the country based on spatial and social media spillovers from New Delhi.

Our paper also contributes to the recent literature on the persistent effects of historical events, institutions, norms, and values (often transmitted across generations) on various outcomes today (for a review, see Voth, 2021). Among the studies focusing on long-term antecedents of discrimination, Voigtländer and Voth (2012) show that Jewish pogroms during the Black Death strongly predict antisemitic violence in Nazi Germany and votes for the Nazi party. Similarly, Ochsner and Roesel (2019) find that Austrian municipalities attacked by Turkish troops in early modern times displayed stronger anti-Muslim sentiment and cast more votes for the far-right party after it started to recall past Turkish atrocities in its 2005 campaign. Along the same lines, Jha (2013) shows that stronger medieval Hindu-Muslim trade relationships in port locations is associated with fewer Hindu-Muslim riots from 1850 to 1995. We contribute to this literature by showing that the proliferation of anti-Muslim fake news at the onset of the coronavirus pandemic was greater in regions with a history of precolonial Muslim attacks. In so doing, we also contribute to the literature analyzing precolonial determinants of present-day outcomes in India, a context in which research has largely been focused on the legacy of colonialism (see, e.g., Banerjee and Iyer (2005), Iyer (2010), Castelló-Climent et al. (2018), Bharadwaj and Mirza (2019), Chaudhary et al. (2020).

More broadly, this paper relates to the literature studying the effects of media on discrimination. Regarding traditional media, DellaVigna et al. (2014), Yanagizawa-Drott (2014), Adena et al. (2015), and Couttenier et al. (2021) show that propaganda and distorted news coverage can contribute to ethnic violence against immigrant minorities and increase votes for far-right parties. Regarding digital media, Müller and Schwarz (2020, 2021) show, respectively, how Twitter, Facebook and other social media can activate hatred of minorities in the contexts of Donald Trump's political rise in the United States and of the refugee crisis in Germany. We contribute to this literature by investigating the determinants of diffusion of fake news that targets a minority, a topic rather unexplored by the literature. In doing so, we also investigate the characteristics of the users who post such false information, thereby shedding light on the origin of this phenomena.

The rest of the paper is organized as follows: Section 2 provides background. Section 3 describes the data. Section 4 presents the empirical strategy and Section 5 discusses the results. Section 6 examines the profiles of discriminatory users. Section 7 concludes.

2 Background

2.1 COVID in India and Fake News Against the Muslim Minority

2.1.1 COVID and the Tablighi Jamaat

COVID-19 was first identified in December 2019 in the Chinese city of Wuhan, where a major local outbreak quickly escalated into a global public health emergency. In India, the first cases of contagions were reported in late January 2020, but they were identified and contained in hopes of avoiding a mass outbreak. Despite its geographical proximity to China, India had close to no cases in February; by March, however, daily cases of contagions were being reported. As the situation worsened, regional and national authorities enacted several restrictive policies that canceled domestic and international flights, suspended railways, created social distancing measures, restricted public gatherings and dine-in restaurants, closed nonessential businesses, and made it compulsory to wear a mask. These policies culminated in a nationwide lockdown that began on March 25.

An Islamic missionary movement called the *Tablighi Jamaat* had scheduled a religious congregation in March 2020 at the *Nizamuddin Markaz* mosque in the Indian capital city of Delhi. Thousands of worshippers from across India and abroad poured into the *Markaz* between March 1 and March 15. Some stayed at the

Mosque for the entire period; others traveled throughout the country.

From March 13, the Delhi state government began issuing public notices to avoid gatherings. Direct letters were sent to the *Tablighi Jamaat* asking them to disassemble the congregation. The *Tablighi Jamaat* responded that it had suspended all activities and had managed to send home some of the attendees, but that it was having difficulty making arrangements for attendees who were still at the mosque and now stranded by the railway closures. The Delhi state government announced a weeklong shutdown of the capital city beginning March 23, further restricting the worshippers from road travel. On March 29, the district police and health officials started sending some of the worshippers to hospitals and quarantine facilities after learning that many of them had tested positive for COVID (Outlook, 2020).

The back and forth between local officials and the *Tablighi Jamaat*, as well as the displacement of attendees and the increasing number of COVID cases from the *Markaz* led to a blame game between the *Tablighi Jamaat* and state officials. By the end of March, social media and news channels were filled with discussions on the government's late response in shutting down the event and on the *Tablighi Jamaat*'s irresponsibility in going through with it. Fake news and discriminatory hashtags blaming Muslims for spreading coronavirus on purpose—including videos showing Muslims licking utensils and a Muslim vendor spitting on fruit to spread the virus—started popping up.

2.1.2 The Evening of March 30

On the evening of March 30, several events surrounding the *Nizamuddin Markaz* transformed the once-sporadic sharing of fake news and Islamophobic tweets into a full-blown crisis. First, the Delhi Police cordoned off the entire area around the *Nizamuddin Markaz*. Drones were also deployed to scan the streets in the area for lockdown violators. Then, the Delhi government announced its intention to file a case against the Maulana, the religious head of the *Tablighi Jamaat*, (Press Trust of India , 2020) and two politicians from the ruling party of the Central Indian government tweeted that the rise in COVID cases in India was linked to the *Nizamuddin Markaz*. Late that night, the Telangana State Chief Minister's office tweeted that six *Markaz* attendees had succumbed to the coronavirus and died in Telangana. Though a *Tablighi Jamaat* attendee had died from COVID the week before, this was the first time multiple deaths were linked to the *Tablighi* event.

Though discriminatory anti-Muslim hashtags and fake news had been shared before (see Appendix Table A1), March 30 marked a watershed moment. The next day, #coronajihad topped Twitter trends in

India, and from then on, the floodgates opened for widespread misinformation and fake news about Muslims deliberately spreading coronavirus (Ritika Jain, Article14, 2020).

On Twitter, hashtags such as #Nizammudin or #NizammudinMarkaz—referring to the mosque—went from being tweeted 10,200 times by 6 p.m. on March 30 to more than 114,000 times by 10 p.m. on March 31, more than a tenfold increase in 28 hours. Similarly, the hashtag #TablighiJamaat had been tweeted about 51,100 times by the end of March 31, with blame extending to all Muslims.[4] This escalation peaked two days later, when members of India's central government blamed the *Tablighi Jamaat* for the sudden spike in COVID cases in India (THE WEEK , 2020). Hashtags such as #coronajihad, #NizamuddinIdiots, #TablighiJamatVirus, and #muslimvirus were widely tweeted and retweeted.

Tweets of the same flavor were also rife in local languages. Alongside the Islamophobic hashtags, these tweets mostly shared old, unrelated videos suggesting that Muslims were trying to purposely spread coronavirus. One widely circulated video involved an elderly fruit vendor being accused of sprinkling urine on the fruits he was selling. Another old video of a Sufi ritual, with Muslims purposely sneezing to spread the virus, went viral. Many other false, convoluted video and audio clips depicting Muslims licking utensils, scattering currency notes, and spitting on food to spread the coronavirus were created and shared during this time (Pooja Chaudari, Alt News, 2020).[5] On April 2, numerous religious gatherings took place across India to celebrate the Hindu festival of *Ram Navami*, but the same blame and discrimination pattern did not ensue; Islamophobic tweets proliferated on these days too.

A notable aspect is that some of the fake news shared during this time recalled precolonial events involving Muslims. The Hindu journalist Samer Halarnkar wrote in April 2020: ”Last week, I exited a family WhatsApp group after listening to a particularly inaccurate, agitated and rambling rant [...] about how Hindus have been 'humiliated, subjugated and massacred,' [...] and how we were ruled by [Muslim] 'marauders' for 1,000 years.” Halarnkar (2020). Similarly, some tweets accused Muslim NGOs and charities of denying free supplies and meals to poor Hindus and Christians during the lockdown, asserting that ”Mughal rulers too denied food supplies to Hindu subjects during famines. They are only continuing the legacy.” Other tweets pointed to the *Tablighi* convention as a tool for spreading Islam, converting Hindus, and evolving into a ”Corona-Jihad,” suggesting that coronavirus could be used as a weapon for a renewed violent attack

[4]All statistics on trending hashtags come from https://getdaytrends.com/india/. By comparison, the hashtags #HappyNewYear and #TrumpInIndia (relating to the start of 2020 and the visit by the former U.S. President on February 24, 2020) received about 767,500 and 136,100 tweets, respectively.

[5]Yet others depicted Muslims groping and misbehaving with nurses at hospitals, beating up doctors, and serving food mixed with human feces.

on India.[6]

In the next section, we briefly discuss historical accounts of precolonial Muslim rule in India and how it has shaped the collective memory of non-Muslims.

2.2 Historical Conflict in India

In the precolonial period, the Indian subcontinent[7] was divided into many independent and politically fragmented states that, throughout the centuries, have often been in conflict with each other. These conflicts, motivated by aims of territorial expansion mixed with religious motives, have been shaping the long-term development patterns of the country (Dincecco et al., 2022).

The earliest Muslim invasions of India can be traced back to the seventh and eighth centuries. Arabs reached the Bombay coast in 636 AD, and the Umayyud campaigns took place across the present-day Pakistan-India border between 712 and 740. However, not until the campaigns of Mahmud of Ghazni, beginning in 1001 with the Battle of Peshawar, did the banner of Islam reached the heart of India (Britannica, 2022). Ghazni sacked and conquered several cities, including the Hindu temple city of Somnath. His empire was overthrown by the (Muslim) Ghurid dynasty in 1186, which was in turn succeeded by the (Muslim) Delhi Sultanate in 1206.

Despite the difficulty in clearly defining “Hinduism” or “Islam” at the time, the close contact between Arabs, Persians, and Turks, and the people of the Indian subcontinent through war and trade possibly led to an antagonistic religious identification (Mukhia et al., 2017, p.9). As early as the 11th century, the prominent Muslim scholar Al Biruni wrote *”They (the Hindus) totally differ from us in religion, as we believe in nothing in which they believe and vice versa (...). Their fanaticism is directed against those who do not belong to them—against all foreigners. They call them mleccha, ie. impure, and forbid having any connection with them, be it by marriage or any other kind of relationship, or by sitting, eating, drinking with them, because thereby, they think they would be polluted. The Hindus claim to differ from us, and to be something better than we, as we on our side, of course, do vice versa!”* (Sachau, 2013).

Tensions among different religious identities continued to rise until the 16th century, when several conflicts between major historical rival states—the Delhi Sultanate, the Deccan Sultanates, the Rajput states,

[6]The full text of the tweet is: ”The main purpose of establishing this organization is to spread Islam. It has so far converted people from other religion, mainly Hindus, and trained them in Islamic religious matters and it is spread in 150 countries. ... Are they Human beings? #Corona_Jihad.”

[7]The Indian subcontinent includes modern-day India, Bangladesh, Bhutan, Myanmar, Nepal, Pakistan, and Sri Lanka.

and the Vijayanagara Empire—took place.[8] While conquests were not exclusively motivated by religion, rulers were particularly harsh in their treatment of people from different religions. Desecrations of Hindu temples by Muslims were clear manifestations of these conflicts, trying to undermine and destroy Hindu religious identity. In particular, the Delhi Sultanate (1206–1526) tried to build a Muslim state and society in Northern India, using selective temple desecration to delegitimize and extirpate Indian ruling houses (Eaton, 2000).

The Delhi Sultanate succumbed to another Muslim entity, the Mughal Empire, which became one of the most powerful states on the Indian subcontinent. Most of the Mughals' conquests were in the territories of the Delhi Sultanate, but they also reached the domains of Hindu rulers. The reign of the Mughal ruler Aurangzeb (1658–1707) featured temple desecrations, differential customs duty based on religion, replacement of Hindu headclerks and accountants by their Muslim counterparts, an additional tax on nonbelievers "to spread Islam and put down the practice of infidelity" (the "*jaziya*"), and rewards for Muslim conversions (Sarkar, 1930). Religion was also a key factor during the Mughal-Maratha and Mughal-Sikh battles.[9]

The nature of interstate conflict changed after the 1757 Battle of Plassey and the victory of the British East India Company, which established itself as the major player in the political landscape, gradually defeating other states and local rulers. Following the Battle of Plassey, Indians of all religious faiths saw themselves increasingly allied against a common enemy, the European colonial powers.

The history of past Muslim attacks has left a permanent wound in the memory of other communities in India, shaping beliefs and perceptions toward Muslims. Wolpert (2004) argues that attempts to unify India before the British colonization had always been difficult and short-lived, with religious influences having divisive effects. Along the same lines, the 1947 partition of British India between Hindu-majority India and Muslim-majority Pakistan and Bangladesh triggered a massive population transfer: Hindus and Sikhs left Pakistan and Bangladesh for India, while Muslims left India to reach the territories where their religious community was larger (Bharadwaj et al., 2015). More recently, during both the World Value Survey for India in 1990 and a major 2019–2020 Pew Research Center survey of religion across India, around 30% of the surveyed population reported they were not willing to have Muslim neighbors (Inglehart et al., 2018;

[8] In the Delhi and Deccan Sultanates, Islam was professed, while in the Vijayanagara Empire, Hinduism was the main religion. The Rajput states were mostly Hindu; only a few professed Islam.

[9] The Mughal-Sikh battles started with the killing and jailing of Sikh leaders by Mughal emperors (who were intent on halting the expansion of Sikhism), and the clashes continued for more than a century. Similarly, the Mughal-Maratha battles involved Muslims and Hindus. Shivaji, the leader of the Marathas, opposed the tax on non-Muslims and revived Hindu traditions in his new empire, taking up Sanskrit and Marathi and abandoning Persian as the court language.

Pew Research Center, 2021).[10]

Even today, non-Muslim communities often follow practices that recollect and remind themselves about past Muslim atrocities. For example, the Rajput festival Jauhar Mela is celebrated every year in Chittorgarh, in Rajasthan, to remember the "*jauhar,*" a mass self-immolation custom performed by Hindu Rajput women to avoid capture, enslavement, and rape by foreign invaders. Locals from Chittorgarh believe *jauhars* were performed three times during history, each time as a consequence of an invasion by a different Muslim ruler. Another example is the chant of the Sikh community in their holy places (*gurudwaras*) that refers to an episode in history where Sikh women were jailed by Mughals and forced to grind flour with heavier-than-normal millstones. The chant[11] venerates those women who chose not to convert to Islam despite the kidnapping and killing of their children.

3 Data

We have assembled a rich dataset from several primary and secondary sources. In this section, we briefly describe the geographical units at which the analysis is carried out and the variables used. Online Appendix A provides further details on some of the steps we undertook to build the main dependent variable and summary statistics for all variables used in our empirical analysis.

We carried out our analysis at the district level. India is currently divided into 773 districts across 28 states and 8 union territories. Because our analysis employs several variables from the 2011 census, we focus on the administrative divisions of India based on this census year; the resulting sample contains 626 districts.[12] In the regressions, we also account for regional or state fixed effects. The States Reorganisation Act of 1956 divides the states of India into six regions: Northern, North Eastern, Central, Eastern, Western, and Southern.[13] On average, regions are made up of six states. Districts are smaller geographical units;

[10] The India Value Survey of 1990 was based on 2,500 interviews, while the Pew Research Center survey of religion in India conducted nearly 30,000 face-to-face interviews of adults in 17 languages. The latter survey was carried out before the COVID-19 pandemic. The exact wording of the question in 1990 was: "On this list are various groups of people. Could you please sort out any that you would not like to have as neighbors?" ("Muslims" was one of the options in the list), while in 2019 the question was: "Would you be willing to accept a Muslim as a neighbor?"

[11] *"Singhnian jinna ne sawa sawa mann de pisne peese, bachiye de tota galean vich pavaye, par Dharm na haariya,"* which translates to "The lioness like Sikh women who grinded a ton of grain, who wore their torn kids around their necks, but did not give up on religion."

[12] We exclude islands, for which some of our data sources have missing information. Moreover, we group together the four districts in which New Delhi is divided because, due to the way tweets are geolocated by Twitter, we are not able to identify in which exact district of New Delhi the tweets were posted, so the New Delhi state will be a unique observation. Besides New Delhi, in our sample three more states are made up by only one district as well: Chandigarh, Dadra and Nagar Haveli, and Daman and Diu.

[13] Regions are commonly called "zones".

there are approximately 104 districts per region and about 18 districts per state.

3.1 Anti-Muslim Fake News

To build a measure of anti-Muslim fake news that varies over time and space, we rely on text data from Twitter.[14] First, we obtained all georeferenced tweets published in India from December 1, 2019, to April 30, 2020 (Twitter API for academic research). This set of queries, conducted at the beginning of 2022, returned 16,967,380 tweets. Each tweet comes with either precise coordinates or a place identifier internal to Twitter.[15] In the latter cases, we also obtained the geographic coordinates of all place identifiers (Twitter API for academic research). For our analysis, we focus on tweets whose place identifier is at the city level or a finer spatial scale (these represent roughly 84% of the original sample). In our main exercise, we use the subset of tweets posted from March 24 to April 6, i.e., from one week before to one week after the *Tablighi* shock on March 30. This sample comprises 1,863,349 tweets published by 187,787 distinct users in 21,977 locations. In addition, we use the tweets posted in December 2019 and January 2020 to construct a matrix of social media interaction across Indian districts prior to the shock and, more generally, prior to the COVID-19 outbreak (we describe this measure in detail in Sections 5.1.2 and 5.2.2).

Second, we assembled an extensive list of English-language hashtags and keywords related to anti-Muslim fake news in India. We started our search for keywords by reading the common fake news stories gathered by the popular Indian fact-checking website AltNews in March and April 2020. In particular, we concentrated on fake news that contains the keyword "Muslim" in their titles or excerpts.[16] This list suggests two sets of relevant keywords. The more prominent set identifies Muslims as intentionally spreading the coronavirus through "sneezing," "spitting," "licking," or "peeing" on people or food. The second set associates Muslims with terrorism by talking for instance about "Islamic Jihadis" or "Corona Jihadis." We first searched the tweets in our dataset containing these initial keywords (and their English-language variations) and the keyword "Tablighi," then we adopted a snowball approach and collected additional relevant keywords that further identify fake news. Appendix Table A2 reports the final list of all keywords we gathered.

[14] Several scholars have recently used Twitter to investigate moral attitudes and norms (see, e.g., Brady et al., 2017) and political preferences (for a review, see Zhuravskaya et al., 2020). Twitter is one of the four most widely used social media platforms in India (the others are Facebook, Instagram, and WhatsApp, which is technically a messaging app) (Sekose, 2021).

[15] See https://developer.twitter.com/en/docs/twitter-api/v1/data-dictionary/object-model/geo for more information on geolocation in Twitter.

[16] Appendix Table A1 reports all fake news we retrieved from the website.

Finally, we translated our keyword list into 11 Indian languages: Hindi, Kannada, Malayalam. Gujarati, Marathi, Punjabi, Odia, Tamil, Telugu, Urdu, and Bengali. We selected these languages because (1) they are the 11 most-spoken languages in India (excluding English), and (2) they are included in the list of languages that Twitter allows users to choose as their preferred language. Based on the Twitter language classification, about 40% of the tweets in our working dataset are published in English, 30% percent in Hindi, and 14% in other languages (none of which exceeds 2.5% of the tweets). For the remaining 16%, Twitter was not able to identify the language of writing.

To validate the ability of our set of keywords to correctly capture Muslim-related fake news stories and illustrate their content, we perform the Latent Dirichlet Allocation (LDA) algorithm on the subset of tweets written in English and containing at least one anti-Muslim keyword.[17] The LDA is an unsupervised machine-learning algorithm that can detect latent topics in a corpus of documents from the co-occurrence of patterns of words (Blei et al., 2003). A topic is defined as a probability distribution over the entire set of words present in the corpus, and the number of topics is a parameter left to the choice of the researcher.

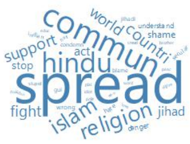
In our baseline exercise, we set the number of topics equal to five, while the other free parameters (governing the shape of the underlying probability distributions) were set according to the standard values suggested in Griffiths and Steyvers (2004). As is standard in text analysis (and in particular for text analysis of Twitter data), we preprocessed the text to remove punctuation, numbers, URLs, and mentions; removed stopwords; and reduced the remaining words to their English stems, so that, for instance, "spreader," "spreading," and "spreaded" were all replaced with "spread" in the analysis.

Figure 1 shows the wordclouds for each topic in our baseline exercise. Each wordcloud is composed of the words with the largest probability weights in the corresponding topic (provided they pass a minimum cutoff, for better readability), such that the size of the word is proportional to the weight. The topics isolate different aspects of the tweets' anti-Muslim content.

Topics 1, 2, and 5 seem to be the ones most likely to contain anti-Muslim false stories. Indeed, Topic 1 draws an analogy between spreading Islam in India and spreading the virus. Here are two sample tweets: "All this while I thought Islam isn't a Religion, it's a cult, I was wrong, - Islam is a disease."; and "The main purpose of establishing this organization is to spread Islam. [...] #Corona_Jihad." Topic 2 captures a specific type of fake news that spread in the aftermath of the *Tablighi* gathering, accusing *Tablighi* participants of

[17] We rely on the language classification provided by Twitter to detect tweets written in English. These tweets may still contain some non-English characters.

intentionally infecting other people and other obscene behaviors. Here are two sample tweets: “Tablighi Jamaat members in quarantine are walking around without trousers on, listening to vulgar songs, asking for bidi cigarette from nurse and staff and making obscene gestures towards nurses. Asks police to restrain them”; and “Are the docs are lying that ur orthodox Muslim men would be naked in front of women medical staff, spit on them, in some place pelt stone [...]”). Topic 3 calls for government action, while Topic 4 focuses on the responsibilities of New Delhi’s local government and police in managing the *Tablighi Jamat* episode. Topic 5 focuses on identifying the *Tablighi Jamaat* convention as a COVID hotspot and blaming the event for the spread of COVID to other parts of India.



(a) Topic 1

(b) Topic 2

(c) Topic 3

(d) Topic 4

(e) Topic 5

Figure 1: Wordclouds for the LDA topics in the Sample of Anti-Muslim Tweets Posted in English Between March 24 and April 4, 2020

3.2 Muslim Attacks

As discussed in Section 2, the resentment against Muslims may have deeply rooted historical origins that can be traced back to long-ago clashes between Hindus and Muslims. In particular, we focus on precolonial conflicts in which a Muslim group or entity participated as an aggressor. This class of conflicts, which we refer to as Muslim attacks, characterized the history of the Indian subcontinent for several centuries. Our measure of exposure to Muslim attacks relies on the database assembled by Dincecco et al. (2022)—based on Jaques (2007)—which contains information on the universe of conflicts recorded on the Indian subcontinent between the years 1000 and 2000. For each conflict, we have the date, the geographical coordinates, and a short description of the event. Starting from these records, we manually coded an indicator variable equal to one for conflicts initiated by a Muslim group or entity. As in Dincecco et al. (2022), we focus on land battles that occurred in precolonial times (i.e., before 1757).[18] Our baseline district-level measure of historical anti-Muslim sentiments is, thus, the dummy variable *Muslim Attack* taking the value one if at least one Muslim land-based attack is observed in a district in the 1000-1757 period. In our robustness checks, we also present our results with perturbed versions of this measure that rely on different time periods and with a distance-based measure of exposure to Muslim attacks (along the lines of Dincecco et al., 2022).

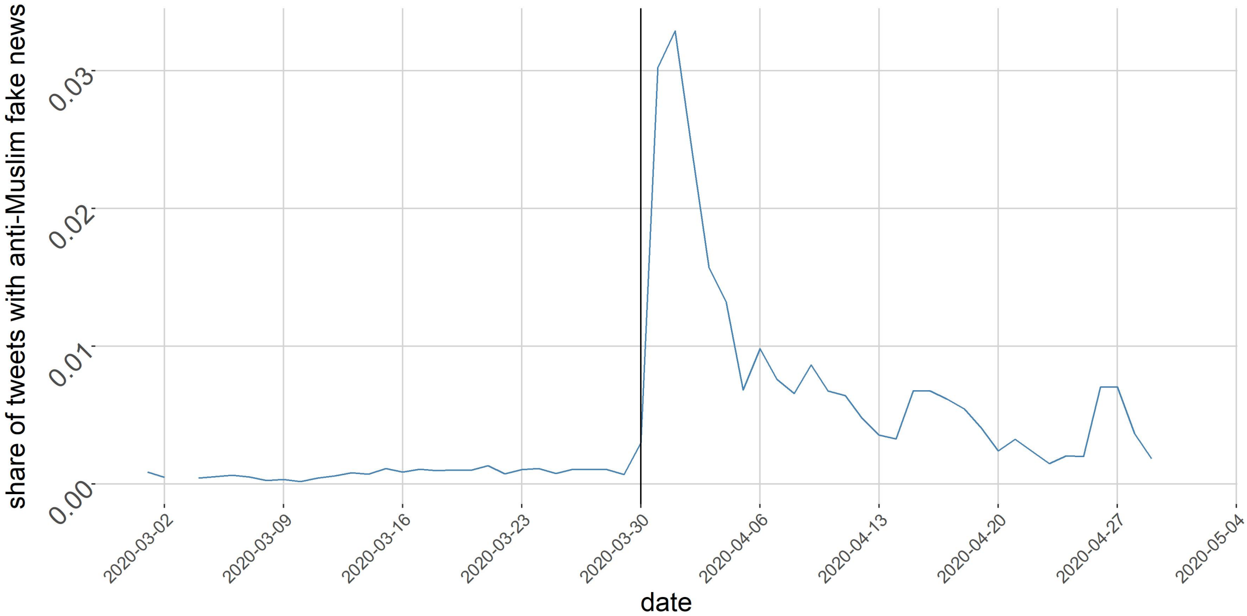
3.3 Descriptive Evidence

In this section, we present some basic facts and preliminary descriptive evidence on the temporal and geographical patterns of discrimination against Muslims in India, focusing on the period from March 1 to May 1, 2020. Figure 2 reports the share of tweets containing anti-Muslim fake news over the total number of tweets posted in a given day. Interestingly, the share of discriminatory tweets is virtually zero in the weeks preceding the *Tablighi* shock, represented in the graph by the solid black vertical line. Consistent with the narratives discussed in Section 2.1, an abrupt spike occurs on March 31, after the events of March 30 established a link in the public's mind between the *Tablighi Jamaat* convention and the COVID-19 crisis. The share of anti-Muslim fake news peaked above 3% of total tweets on April 1, then steadily declined—though it remained above its preshock level one month later. In our empirical analysis, we restrict our attention to the two weeks centered around March 30, i.e., from March 24 to April 6.

Our detailed georeferenced data allow us not only to observe discrimination patterns at high temporal frequencies but also to document their distribution in space. Figure 3a shows the geographical distribution of anti-Muslim fake news across Indian districts in the week following March 30. To account for heterogeneous levels of Twitter activity across districts, the map reports the residuals of an OLS regression of the number of

[18] We concentrate on land battles because they clearly occurred within specific district borders and were by far the most common type of precolonial conflict.

Figure 2: Time Series of Anti-Muslim Fake News



Notes: The blue line represents the share of tweets containing anti-Muslim fake news keywords over the total number of tweets in our sample in each day from March 1 to May 1. The solid black line highlights the date of the shock, March 30.

tweets with anti-Muslim fake news posted in the week after the shock on the total number of tweets posted in the week preceding the shock. The color scale corresponds to the different deciles of the distribution of residuals, ranging from low discrimination (white) to high discrimination (dark blue). There is substantial spatial variation in the intensity of anti-Muslim discrimination in the aftermath of the shock. Discrimination is higher in the area around New Delhi, along the Ganges Valley in the northeast, and along the western coast, whereas it is lower, for instance, in the southwest, in the area corresponding to the state of Andra Pradesh. A second feature that emerges at a glance from the figure is the spatial autocorrelation in the intensity of discrimination across Indian districts. We confirm this formally by computing the Moran's I spatial autocorrelation index on the same residuals depicted in the figure: we obtain a value of 0.28 (significant at the 1% level).

Figures 3b and 3c suggest two potential mechanisms that may partially explain the substantial spatial heterogeneity in the diffusion of anti-Muslim fake news we observe. First, in Figure 3b, we averaged the residuals reported in Figure 3a for five rings corresponding to the quintiles of the distribution of distance to New Delhi—where distance is calculated from the district centroid.[19] This is motivated by the fact that the shock is narrowly localized in New Delhi, where the *Tablighi Jamaat* convention took place. As

[19]In the figure, the district polygons for districts within the same distance quintile have been merged into a single polygon.

the map shows, there is a clear negative gradient from New Delhi to the outer regions of India. Second, Figure 3c shows the geography of precolonial Muslim attacks. The red crosses coincide with the exact conflict locations, while individual districts are highlighted if at least one conflict occurred in their territory (consistent with the dummy variable used in the econometric analysis). There is a striking resemblance to the map of anti-Muslim fake news shown in Figure 3a. This suggests that historical Muslim attacks have left some anti-Muslim sentiments and perceptions that could shape the reaction of different Indian districts to the *Tablighi* shock.

4 Empirical Specification

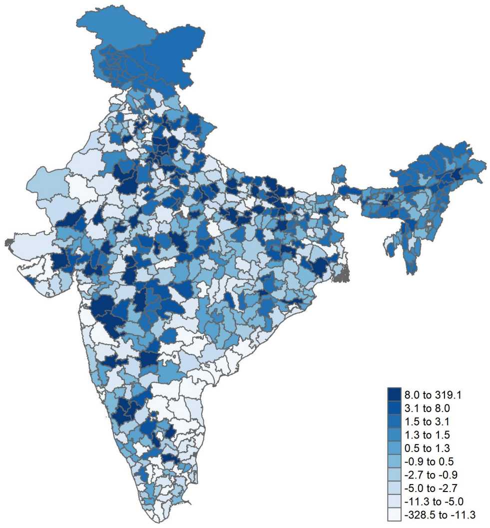
To investigate the determinants of the diffusion of anti-Muslim fake news on social media, we proceed in two main steps. First, we estimate the following equation in a cross-sectional setting:

$$Y_{ir} = \alpha + \beta Z_{ir} + \gamma \mathbf{X}_{ir} + \delta_r + \varepsilon_{ir}$$

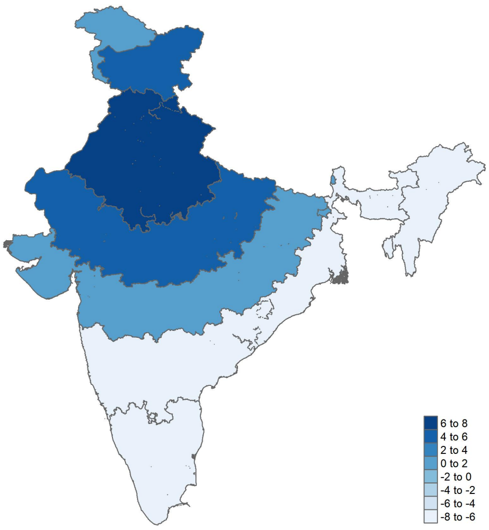
where Y_{ir} is the cumulated number of tweets with anti-Muslim fake news posted in district i in region r on the day after the shock (March 31) and the following two to six days, and Z_{ir} captures a potential determinant for the rise of anti-Muslim fake news. In particular, we will concentrate first on *New Delhi*, which is a dummy variable taking the value one for the territory of New Delhi, and then on both *New Delhi* and *MuslimAttack*, which is a dummy taking the value one for districts that experienced at least one Muslim attack during the precolonial period (1000–1757). The coefficient β shows how much stronger, after the shock, the diffusion of fake news is in the New Delhi area (or in an area where a conflict was initiated by Muslim entities), compared to the rest of India. X_{ir} is a vector of district-specific baseline and geographic controls, δ_r are region fixed effects, and ε_{ir} is the error term. To account for spatial autocorrelation, the error term is clustered according to Conley (1999).

Regional fixed effects capture unobserved historical, linguistic, political, and economic characteristics at the regional level. Thus, we identify the effect of the *New Delhi* dummy or the dummy for districts with precolonial Muslim attacks by comparing anti-Muslim fake news across districts within regions. At the same time though, region fixed effects do not account for differences at a finer spatial level. To address this issue, we control directly for prominent baseline and geographic characteristics at the district level,

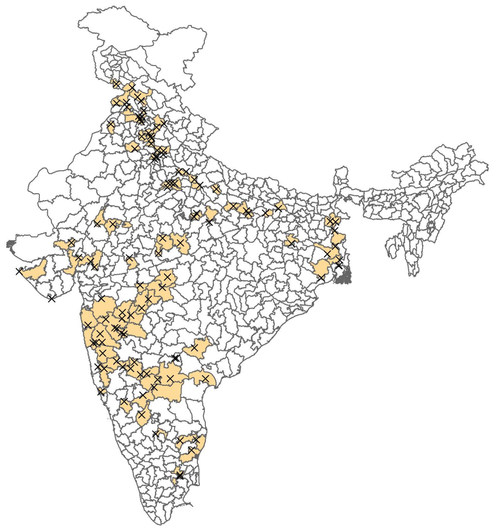
Figure 3: Spatial Distribution of Anti-Muslim Fake News and Precolonial Muslim Attacks



(a) Anti-Muslim Fake News (residuals), Week After March 31



(b) Average Residuals by Quintiles of Distance From New Delhi



(c) Location of Precolonial Conflicts With Muslim Offenders

Notes: Figure 3a depicts the residuals of an OLS regression of the number of tweets with anti-Muslim fake news keywords posted in a district during the week after the shock (March 31 to April 6) over the total number of tweets posted in a district during the week before the shock (March 24 to March 30); the color scale represents different deciles of the distribution of residuals, where darker colors correspond to higher residuals. In Figure 3b, the same residuals are averaged by quintiles of distance from New Delhi, where distance is calculated from the district centroid and the district polygons have been merged within each quintile. In Figure 3c, the red crosses coincide with the exact locations of land-based precolonial (1000–1757) conflicts where the offender was a Muslim group or entity; the districts in which at least one such conflict is recorded are highlighted in yellow.

which could be related to both the likelihood that users posted an anti-Muslim tweet and the likelihood that a district was exposed to a Muslim attack in the precolonial period. In particular, following insights from Dincecco et al. (2022) and Michalopoulos and Papaioannou (2018), the baseline controls include the log of luminosity (+0.01) averaged in the 1992–2010 period to account for local economic development—with luminosity being a better proxy than official GDP data, especially in poorer areas—and the log of population density in 1990.[20]

Moreover, we control for the log share of Muslim and Hindu population, which may account for district-specific Muslim *versus* Hindu contact. We include the log shares of literate and urban population to account for broader measures of well-being and development. To capture indirect or direct correlates of district-level differences in Twitter use, we control for the average number of tweets during the week before the shock. All these data come from the 2011 census, except for the average number of tweets, which is computed by exploiting our database. Importantly, all regressions exploring the role of Muslim attacks also include a dummy taking the value one if a district experienced a precolonial conflict in which Muslim groups were *not* the aggressors. This accounts for the legacy of conflict-related violence, which may also affect the likelihood of contemporary violence and discrimination.

Finally, we control for the daily number of COVID-19 deaths, as the spread of fake news could simply be related to the spread of the coronavirus, rather than to the *Tablighi* hotspot news or to the legacy of precolonial Muslim attacks. Moving to the geographical controls, we follow Dincecco et al. (2022) and include latitude, longitude, altitude, ruggedness, precipitation, land quality, dry rice suitability, wet rice suitability, wheat suitability, and malaria risk. These factors are commonly accounted for by a broader literature and should capture potential drivers of historical Muslim-related conflicts. To these factors, we add distance to the border as well as distance to the closest border between Pakistan and Bangladesh—the two countries bordering India with the highest share of Muslim population.

While the region fixed effects and the district-level controls account for regional unobservable and district-level observable characteristics, to further rule out that omitted-variable bias is driving our results, we now move to a quasi-experimental setting and use a difference-in-difference strategy. Specifically, we leverage the evening of March 30 as the shock cutoff date during two periods: (1) from March 28 to April 2 (capturing three-day windows before and after the shock), and (2) from March 24 to April 6 (capturing

[20] Because local luminosity levels in India do not simply reflect population density (Dincecco et al., 2022), we also directly control for population density. In particular, we use data on population density in the most recent year prior to the years in which luminosity is measured, but note that its log is correlated at 98% with the log of population density in the last census year, 2011.

seven-day windows before and after the shock). The estimated equation is:

$$Y_{it} = \beta Z_i + \gamma Post_t + \delta Z_i \times Post_t + \phi X_{it} + \theta_i + \eta_t + \varepsilon_{it}$$

where Y_{it} is the number of tweets with anti-Muslim fake news posted in district i on day t, Z_i is a potential determinant of diffusion of anti-Muslim fake news as in the cross-section equation, and $Post_t$ is a dummy taking the value one for all dates after March 30. The main coefficient of interest is δ, which is the difference-in-difference coefficient of the interaction term between Z_i and $Post_t$ that shows how much stronger the increase in the diffusion of fake news is after the shock in the New Delhi area (or in an area where a conflict was initiated by Muslim entities), compared to the rest of India.[21] With respect to the cross-section specification, we control for district and date fixed effects (θ_i and η_t, respectively) in the panel specification.[22]

Finally, besides the number of COVID deaths, the panel specifications also include as baseline controls the average number of tweets in the week prior to the shock interacted with the $Post_t$ indicator to control for differential effects over time of Twitter penetration and—when we focus on the relationship between Muslim attacks and fake news diffusion—a dummy for districts experiencing other precolonial conflicts (beyond Muslim attacks) interacted with the $Post_t$ indicator to control for differential effects of the legacy of historical violence beyond Muslim attacks.

5 Empirical Analysis

5.1 New Delhi: The Tablighi COVID-19 Hotspot

5.1.1 Anti-Muslim Fake News in New Dehli

In Table 1, we discuss whether the *Tablighi* shock has had a stronger effect on the number of tweets with anti-Muslim fake news in New Delhi, where the shock originated. In columns 1–2, the analysis is carried out at the cross-district level and the dependent variable is the total number of tweets with anti-Muslim fake news in

[21] The interaction coefficient is not interpretable as the effect of the shock in a given area, because the shock had national relevance and therefore may have affected all Indian districts. The coefficient δ, however, is crucial for understanding what the strongest determinants of the spatial diffusion of fake news are, by showing where the rise in the number of tweets containing false news is strongest.

[22] The obvious implication is that when $Z_i = NewDelhi$, the dummy $NewDelhi$ is absorbed by the district fixed effects, and in all results the dummy $Post_t$ is subsumed by the date fixed effects.

the postshock period, i.e., from March 31 to April 2. Column 1 accounts for the baseline and geographical controls, and column 2 adds region fixed effects. In both columns, the coefficient on the dummy *New Delhi*—tracking the district where the *Tablighi* convention took place—is positive and significant at the 1% level. This suggests that, in the very short term, the shock was associated with a sharper increase in the number of tweets with discriminatory fake news within the New Delhi area compared to the rest of India.

Columns 3–4 report the results of a difference-in-difference specification using as the dependent variable the daily number of tweets with anti-Muslim fake news in each Indian district from March 28 to April 2. Column 3 controls for state and date fixed effects and for the baseline and geographic controls; column 4 adds district fixed effects. The coefficient estimate suggests that, following the shock, the number of tweets with anti-Muslim fake news increased 144 more in New Delhi compared with the other districts.

Columns 5–8 replicate the analysis of columns 1–4, focusing on the longer, seven-day time horizon. As concerns the cross-sectional evidence (columns 5–6), the coefficient on the *New Delhi* dummy is still positive and significantly different from zero. The results of the difference-in-difference specification (columns 7–8), however, report a smaller and not significant coefficient.

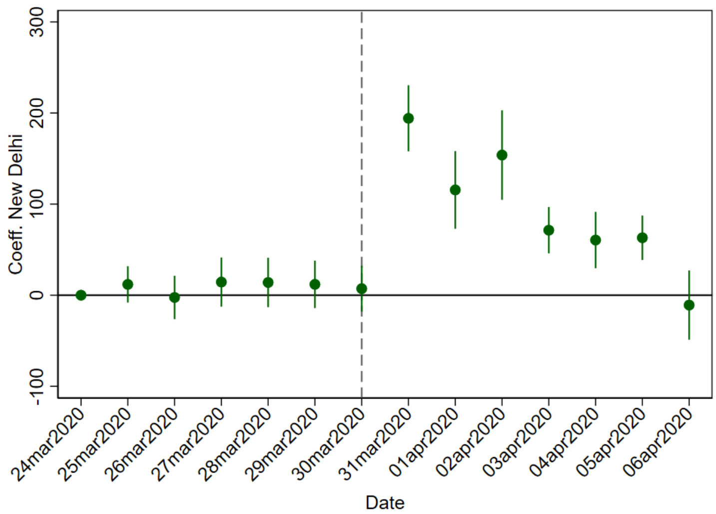
To shed more light on the effect of the *Tablighi* shock over time, we perform an event-study analysis. Specifically, we use a more flexible estimating equation that, rather than interacting the *New Delhi* dummy with the *Post* indicator, interacts the New Delhi measure with each of the time-period dummies, using March 24 as the baseline period. The results can be intuitively seen in Figure 4. We observe a sharper increase in anti-Muslim fake news right after the shock in New Delhi compared to the rest of India, but the difference gradually decreases over the following days. The evidence of the event study therefore also helps us explain the smaller and not significant coefficient of the difference-in-difference specification over the seven-day period (compared to the one over the three-day period). Reassuringly, there were no differential trends in anti-Muslim fake news before March 30, when the public began connecting the *Tablighi* convention and the COVID outbreak.

Table 1: New Delhi and the Tablighi Shock

Period:	Three Days				Seven Days			
	Cross-Section From shock		Panel		Cross-Section From shock		Panel	
Dependent variable:	Cum. N. Tweets FN		N. Tweets FN		Cum. N. Tweets FN		N. Tweets FN	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
New Delhi	431.6551***	434.6582***			580.2159***	585.2052***		
	(85.5375)	(83.4294)			(126.6417)	(123.2655)		
	[0.0000]	[0.0000]			[0.0000]	[0.0000]		
New Delhi × Post			143.5211***	144.0059***			84.5201	84.6502
			(24.7875)	(21.3123)			(83.3240)	(83.0970)
			[0.0000]	[0.0000]			[0.3104]	[0.3083]
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographic controls	Yes	Yes	Yes	No	Yes	Yes	Yes	No
Region FE	No	Yes	No	No	No	Yes	No	No
State FE	No	No	Yes	No	No	No	Yes	Yes
District FE	No	No	No	Yes	No	No	No	Yes
Day FE	No	No	Yes	Yes	No	No	Yes	Yes
R-squared	0.9646	0.9657	0.9473	0.9637	0.9683	0.9694	0.7997	0.8114
Observations	626	626	3756	3756	626	626	8764	8764

Notes: OLS estimates. Observations are districts. Cross-district level analysis in the March 31–April 2 period in columns 1–2, daily panel analysis in the March28–April 2 period in columns 3–4, cross-district level analysis in the March 31–April 6 period in columns 5–6, and daily panel analysis in the March 24–April 6 period in columns 7–8. Cross-section results in columns 1–2 and 5–6 use as a dependent variable the cumulative number of tweets with anti-Muslim fake news, and focus on the dummy variable *New Delhi*, tracking the New Delhi district as the main explanatory variable. Daily panel estimates in columns 3–4 and 7–8 use as a dependent variable the number of tweets with anti-Muslim fake news, and focus on the variable *New Delhi × Post*, which is a dummy taking the value one for the district of New Delhi after March 30, as the main explanatory variable. In all specifications, the baseline controls include the log of luminosity (+0.01) averaged between 1992-2010, the log of population density in 1990, the log share of Muslim and Hindu population in 2011, the log share of literate and urban population in 2011, and the average number of tweets before March 31. Geographical controls include latitude, longitude, altitude, ruggedness, precipitation, land quality, dry rice suitability, wet rice suitability, wheat suitability, malaria risk, distance to the border, and distance to the closest border between Pakistan and Bangladesh. All cross-section estimates also include the total number of COVID deaths, and region fixed effects are added to columns 2 and 6. All panel specifications also include the daily number of COVID deaths, a dummy taking the value one after March 30, and its interaction with the average number of tweets in the preshock period. Columns 3 and 7 control for state fixed effects, while columns 4 and 8 control for district fixed effects. See the text and the Appendix for details on all variables. Standard errors in parentheses are clustered to account for spatial correlation up to 250 km in the cross-section, and for both spatial and serial correlation in the panel specifications. P-values are reported in brackets. *** p<0.01, ** p<0.05, * p<0.1.

Figure 4: New Delhi and the Tablighi Shock: Event Analysis



Notes: Each dot is a coefficient from a version of specification 8 in Table 1 that replaces the *New Delhi* dummy interacted with the *Post March 30* indicator, with the *New Delhi* dummy interacted with each date dummy. The reference category is March 24, and the specification also controls for the average number of tweets in the week before the shock interacted with date dummies. Vertical bars indicate 90% confidence intervals.

Robustness. In Appendix B, we perform a series of robustness checks on specification 4 of Table 1. First, Figure B1 shows that our results are robust to accounting for spatial autocorrelation over distance thresholds ranging from 100 to 1,500 km. Second, Figures B2 to B7 show the sensitivity of our main coefficient of interest—*New Delhi* × *Post*—to including in the regression specific potentially confounding factors interacted with the *Post March 30* dummy. We start by interacting each baseline and geographic control in Figures B2 and B3, respectively. Then we do the same exercise by controlling for (1) *further geographic controls* in Figure B4, (2) *initial conditions* in Figure B5, (3) *colonization controls* in Figure B6 to account for the role of district-specific differential exposure to the subsequent British colonization, and (4) *fractionalization controls* in Figure B7 to specifically consider the role of ethnic, linguistic, and religious diversity in generating anti-Muslim fake news after March 30.[23] None of these perturbations of

[23] *Further geographic controls* include distance to the coast, presence of a river, irrigation potential, the coefficient of variation in rainfall, the percentage of forested area, and distance to petroleum, diamonds, gems, and gold deposits. These variables were assembled by Dincecco et al. (2022) from several sources, including the Natural Earth Data website (https://www.naturalearthdata.com/), Matsuura and Willmott (2009), Tollefsen et al. (2012), the India Institute of Forest Management (2015), and Bentzen et al. (2017). As suggested by Dincecco et al. (2022), we proxy *initial conditions* by including initial state-capacity measures such as the number of Indian settlements during the Neolithic or Chalcolithic Ages from Nag (2007), the number of important Indian cultural sites between 300 and 700 and between 800 and 1200 from Schwartzberg (1978), and the natural logarithm of (one plus) the total urban population in the year 1000 according to Chandler (1987). *Colonization controls* include a dummy variable for direct British rule from Iyer (2010), the number of years a district was ruled by the British from Verghese (2016), and a variable tracking the year in which each district was connected to the first colonial railroad from Fenske et al. (2021). *Fractionalization* measures, again assembled by Dincecco et al. (2022) from various sources, include medieval ports and a dummy for districts intersected by the

the main specification sensibly affects our result. Finally, in Table B1, we show that the main results hold if we consider as a dependent variable the daily share of tweets with anti-Muslim fake news instead of the absolute number of anti-Muslim fake news stories.

Altogether, the findings in this section suggest that the *Tablighi* shock triggered a sharper increase in anti-Muslim fake news in the New Delhi area, where the religious convention took place, which is consistent with the evidence discussed in Section 2.

5.1.2 Fake News Spread Spatially From New Delhi

We now investigate whether the anti-Muslim fake news triggered by the *Tablighi* shock spread spatially during the three-day and seven-day windows after March 30. We take the specifications of columns 4 and 8 in Table 1 and enrich them to account for fake news' diffusion. Table 2 reports the results.

First, we focus on the regional area centred around New Delhi. This area, defined in 1985 when the government approved a regional development plan, encompasses the state of New Delhi and 21 surrounding districts across the states of Haryana, Uttar Pradesh, and Rajasthan. Its population is over 46 million inhabitants, and its urbanization rate is 62.6% (Census of India, 2011). We compute a dummy equal to one for districts located in the New Delhi region (excluding New Delhi itself) and augment the panel specification with this variable interacted with the *Post-March 30* dummy. Columns 1 and 4 show that the districts neighboring the state of New Delhi reported a significantly higher rise in the number of tweets with anti-Muslim fake news over the three days and the seven days after the *Tablighi* shock compared to districts outside the New Delhi region. This first exercise suggests that anti-Muslim fake news that originated in New Delhi soon spread to its neighboring territories, propagating a discriminatory sentiment against the Muslim population.

Second, in columns 2 and 5, we study the diffusion of fake news beyond New Delhi's neighboring districts, by using the continuous variable of physical distance; in particular, we include the distance from each district's centroid to the centroid of New Delhi interacted with the *Post-March 30* dummy. The interaction term is negative and significant at the 1% level, suggesting that fake news diffused out of New Delhi—but with a lower magnitude in districts farther from the *Tablighi* convention. Next, to better explore the role of distance in the propagation of the shock, we estimate an equation that interacts the *Post-March 30* dummy with dummies taking the value one depending on the decile of distance to New Delhi. The results are de-

Ganges river, religious fractionalization and polarization, ethnic polarization, and the scheduled caste and tribe shares from the 2011 census. Finally, we control for the number of years a district was ruled by Muslims in medieval times.

picted in Figure 5, plotting the coefficients on the interaction terms for the different deciles. They suggest that fake news propagated spatially, with a declining effect in space. This evidence aligns well with the descriptive evidence presented in Section 3.3.

Moreover, the diffusion of fake news may travel on the network of social interactions on Twitter. We thus construct a measure of social media connectedness with New Delhi. In particular, we exploit the fact that when tweets are posted in response to a tweet (known as a reply), or by quoting another tweet, the Twitter API also provides the ID of the original tweet. For each district i, we compute the share of replies to (and quotes of) tweets initially posted in the state of New Delhi over the total number of replies and quotes by users located in district i.[24] In columns 3 and 6, we include in our specifications both the measure of social media connectedness to New Delhi interacted with the *Post-March 30* dummy and the dummy equal to one for districts located in the New Delhi region interacted with the *Post-March 30* dummy. Interestingly, these results seem to suggest that spillovers due to social media interactions take place above and beyond spillovers occurring in the actual physical space.

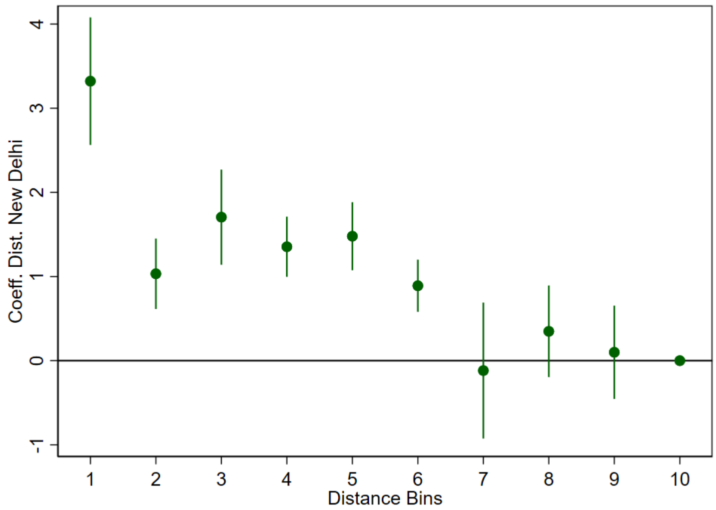
[24]We computed our measure using the tweets posted from December 1, 2019 through January 31, 2020. We chose this period to rule out that the diffusion of COVID dominated social interactions between districts. Moreover, we exclude from the denominator the quotes and replies to tweets posted within the district because with this measure we want to specifically focus on social connections outside the district.

Table 2: Number of Anti-Muslim Fake News Tweets in Space

Period:	Three Days			Seven Days		
Dependent variable:	N. Tweets FN			N. Tweets FN		
	(1)	(2)	(3)	(4)	(5)	(6)
New Delhi × Post	147.7242***	142.9669***	148.9854***	87.3092	83.9065	88.1959
	(19.6609)	(20.8076)	(19.6250)	(83.6476)	(83.0038)	(83.7197)
	[0.0000]	[0.0000]	[0.0000]	[0.2966]	[0.3121]	[0.2921]
New Delhi Region w/o ND × Post	7.1609***		6.9495***	5.1731**		5.0245**
	(1.5053)		(1.4977)	(2.0925)		(2.0747)
	[0.0000]		[0.0000]	[0.0134]		[0.0154]
Dist. New Delhi (x1000) × Post		-1.9623***			-1.3963***	
		(0.3227)			(0.3071)	
		[0.0000]			[0.0000]	
Social Media Connectedness to New Delhi × Post			1.9456***			1.3694***
			(0.4478)			(0.4670)
			[0.0000]			[0.0034]
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.9645	0.9645	0.9648	0.8124	0.8121	0.8125
Observations	3756	3756	3756	8764	8764	8764

Notes: OLS estimates. Observations are districts in each day in two periods: March28–April 2 in columns 1–3, and March24–April 6 in columns 4–6. In all specifications, the dependent variable is the number of tweets with anti-Muslim fake news, and the baseline and geographical controls are as in the panel specifications of Table 1. *New Delhi × Post* is a dummy taking the value one for the district of New Delhi after March 30. *New Delhi Region w/o ND × Post* is a dummy taking the value one for districts located in the National Capital Region surrounding New Delhi (excluding New Delhi itself) interacted with the *Post-March 30* dummy. *Dist. New Delhi (x 1000) × Post* is the distance from each district's centroid to the centroid of New Delhi (per 1000 km) interacted with the *Post-March 30* dummy. *Social Media Connectedness to New Delhi × Post* computes for each district *i* the share of quotes and replies to tweets posted in the state of New Delhi by users located in the district (excluding from the denominator the quotes and replies to tweets posted within the district) interacted with the *Post-March 30* dummy. Standard errors in parentheses are clustered to account for both spatial correlation (up to 250 km) and serial correlation. P-values are reported in brackets. *** p<0.01, ** p<0.05, * p<0.1.

Figure 5: Number of Fake News Tweets in Space by Deciles of Distance to New Delhi



Notes: Each dot is a coefficient from a version of specification 5 in Table 2 that replaces distance to New Delhi interacted with $Post_t$ with the interaction of the $Post_t$ dummy with dummies taking the value one depending on the decile of distance to New Delhi. The reference category is the 10th decile, and the specification also controls for the simple decile dummies and for the average number of tweets in the week before the *Tablighi* shock interacted by time dummies. Vertical bars indicate 90% confidence intervals.

5.2 Deep-Rooted Determinants of Anti-Muslim Fake News and the Tablighi Shock

5.2.1 Precolonial Muslim Attacks and the Tablighi Shock

After showing that fake news spread spatially out of New Delhi, we now investigate whether some deeper determinants are affecting the spatial diffusion of anti-Muslim discrimination. Following the rich historical record of Muslim invasions and conquests of territories in present-day India, we focus on the role of precolonial conflicts in which a Muslim group took part as an aggressor. Table 3 mimics the specifications of Table 1, adding among the explanatory variables a dummy equal to one for districts that experienced Muslim attacks and its interaction with the *Post-March 30* indicator. All cross-section specifications additionally control for conflict events in which the aggressor was not a Muslim group, while panel specifications control for both this variable (subsumed when using district fixed effects) and its interaction with the *Post-March 30* dummy.

In columns 1–2 and 5–6, we present the results of the cross-sectional specifications using as the dependent variable the cumulated number of tweets with anti-Muslim fake news in the three days (columns 1–2) and seven days (columns 5–6) after the *Tablighi* shock. Focusing on column 2, both coefficients associated with the *New Delhi* and *Muslim Attack* dummies are positive and significant at the 1% level, suggesting that both the location associated with the *Tablighi* shock and exposure to precolonial Muslim attacks play a role in explaining the district variation in anti-Muslim fake news diffusion. Note that although the magnitude of the coefficient associated with Muslim attacks is much smaller than the one associated with the *New Delhi* dummy, it amounts to more than 20% of the magnitude of the coefficient associated with districts located in the New Delhi region (excluding New Delhi, not shown). This comparison further points out the relevance of precolonial Muslim attacks as long-term determinants of anti-Muslim fake news today.

Columns 3–4 and 7–8 report the results of the difference-in-difference specification using the three-day and seven-day windows, respectively. The coefficients of the interaction term fall to about one additional tweet with anti-Muslim fake news when we focus on the period up to seven days after March 30, and they remain significant at the 5% level (columns 7–8).

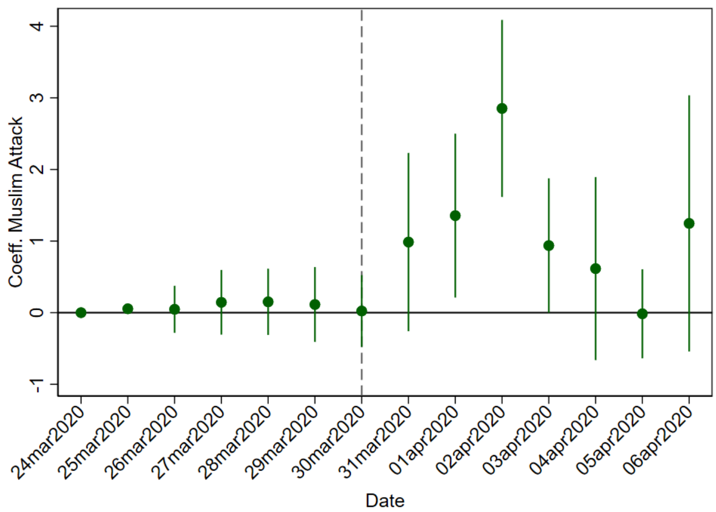
Table 3: Number of Tweets With Anti-Muslim Fake News and Historical Muslim Attacks

Period:	Three Days				Seven Days			
	Cross-Section From Shock		Panel		Cross-Section From Shock		Panel	
Dependent variable:	Cum. N. Tweets FN		N. Tweets FN		Cum. N. Tweets FN		N. Tweets FN	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
New Delhi	434.9733***	436.6861***			583.4918***	586.5906***		
	(83.7485)	(81.7575)			(123.7766)	(120.6348)		
	[0.0000]	[0.0000]			[0.0000]	[0.0000]		
New Delhi × Post			144.3309***	144.7996***			84.9248	85.0433
			(24.4654)	(21.0321)			(83.3062)	(83.0811)
			[0.0000]	[0.0000]			[0.3080]	[0.3060]
Muslim Attack	3.3640**	3.6445***	-0.1582		5.1059**	5.4841***	-0.1643	
	(1.4481)	(1.2084)	(0.1576)		(2.3999)	(1.9924)	(0.1058)	
	[0.0202]	[0.0026]	[0.3157]		[0.0334]	[0.0059]	[0.1203]	
Muslim Attack × Post			1.6349**	1.6346***			1.0731**	1.0835**
			(0.6440)	(0.5683)			(0.5033)	(0.4727)
			[0.0111]	[0.0040]			[0.0330]	[0.0219]
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographic controls	Yes	Yes	Yes	No	Yes	Yes	Yes	No
Region FE	No	Yes	No	No	No	Yes	No	No
State FE	No	No	Yes	No	No	No	Yes	No
District FE	No	No	No	Yes	No	No	No	Yes
Day FE	No	No	Yes	Yes	No	No	Yes	Yes
R-squared	0.9650	0.9660	0.9478	0.9640	0.9686	0.9697	0.8000	0.8117
Observations	626	626	3756	3756	626	626	8764	8764

Notes: OLS estimates. Observations are districts. The table replicates the structure and control variables reported in Table 1. All specifications further include the dummy variable *Muslim Attack*, tracking districts that experienced attacks from Muslim groups in the 1000-1757 period, and control for a dummy taking the value one for districts that experienced precolonial conflict in which Muslims groups were not the aggressors. Panel estimates in columns 3–4 and 7–8 further include *Muslim Attack × Post*, which is the interaction term between *Muslim Attack* and the *Post-March 30* dummy, and the interaction between the dummy for conflicts with non-Muslim aggressors and the *Post-March 30* dummy. See the text and the Appendix for details on all variables. Standard errors in parentheses are clustered to account for spatial correlation up to 250 km in the cross-section, and for both spatial and serial correlation in the panel specifications. P-values are reported in brackets. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

To further shed light on the dynamic patterns of fake news diffusion, we perform an event-study analysis. We start from the specification in column 8, but, rather than interacting the *Muslim Attack* dummy with the *Post-March30* indicator, we interact the attack measure with each of the time-period dummies. The baseline period is March 24. Interestingly, the difference between the number of tweets with anti-Muslim fake news in areas where the conflict was initiated by Muslim entities and areas without conflict started increasing on March 31, reached its peak two days later, and declined in the subsequent days. No differential trends in anti-Muslim fake news are displayed in the preshock period for districts that experienced precolonial Muslim attacks.

Figure 6: Number of Tweets With Anti-Muslim Fake News and Historical Muslim Attacks: Event Analysis



Notes: Each dot is a coefficient from a version of specification 8 in Table 3 that replaces the dummy for districts that experienced precolonial Muslim attacks interacted with the dummy tracking all dates from March 31 with the dummy *Muslim Attack* interacted by each date dummy. The baseline period is March 24. The specification also controls for the average number of tweets in the week before the *Tablighi* shock interacted by date dummies. Vertical bars indicate 90% confidence intervals.

Robustness. In this section, we perform a series of robustness checks to corroborate our findings on the role of precolonial Muslim attacks in the diffusion of fake news. We focus on specifications 4 and 8 of Table 3. First, in Appendix Figures B8 and B9, we show that both the interaction of the *New Delhi* dummy and of the *Muslim Attack* dummy with the *Post* indicator are robust to accounting for spatial correlation of distance thresholds ranging from 100 to 1,500 km.

Second, in columns 1 and 6 of Table 4, we show that our results hold when using as a dependent variable the share of tweets with anti-Muslim fake news. Both the location of the *Tablighi* shock and districts that experienced precolonial Muslim attacks consistently display a higher share of tweets with anti-Muslim fake news after March 30.

Third, we perform different exercise to explore alternative measures of exposure to Muslim attacks. About 30% of the districts with historical Muslim attacks experienced this type of event more than once, so in the first exercise, we consider the intensity of exposure to Muslim attacks. In columns 2 and 7 of Table 4, we thus use *Number of Muslim Attacks*—tracking the number of conflict events in which Muslim groups were the aggressors during the 1000–1757 period—interacted with the *Post-March30* dummy. The coefficient of the interaction term is consistently positive and significant, suggesting that in districts experiencing the highest number of Muslim attacks (eight events), the number of tweets with anti-Muslim fake news increased by up to 4 tweets after March 30. In our second exercise we consider the fact that districts that did not directly experience a Muslim attack may nevertheless have been affected to some extent, either because the conflict occurred in the proximity of the border or because the movement of armies in the territory might have left some scars. Building on the analysis in Dincecco et al. (2022), in columns 3 and 8 of Table 4, we compute Muslim conflict exposure as

$$\sum_{c \in C} (1 + distance_{i,c})^{-1}$$

where $distance_{i,c}$ is the distance between the centroid of district i and the location of a Muslim attack c. This measure implies that the nearer a district is to a particular Muslim attack, the more exposed it is. Muslim attacks occurring at the district centroid receive a weight of one, or full weight; as the distance of Muslim attacks from the centroid increases, they receive lower weights. In this way, we impose no cutoff at the district's borders. For each district, we consider all conflicts within a radius of 250 km. Our results support the relevance of being a district that experienced Muslim attacks for an increase in anti-Muslim fake news after March 30. Note that this result is also robust to the use of an alternative radius of 100 km or 5,000 km (see columns 3–4 and 8–9 of Appendix Table B2). In the third exercise, we show that our benchmark measure of Muslim-related conflicts is still relevant if we compute it over a more restricted period or over a longer period. Estimates in columns 1–2 and 6–7 of Appendix Table B2 show that the coefficient is even larger if we consider districts experiencing Muslim attacks between the birth of the Delhi Sultanate (around 1200) and the establishment of the British East India Company in India, in 1757; it is only slightly lower in magnitude if we consider 1840 as the cutoff date (after which the British dominated the Indian subcontinent both militarily and politically). Finally, some districts were exposed to conflict events in which Muslim groups were not attacking but were still one of the parts involved in the conflict. We show in column 5 of

Table B2 that specifically accounting for exposure to conflict involving Muslims does not affect our main result. In our fourth exercise, we show in columns 4 and 9 of Table 4 that the result on the legacy of precolonial Muslim attacks for anti-Muslim fake news after March 30 is still positive and significant if we remove the district of New Delhi from the sample. Results on Muslim attacks are also robust when we control for spatial and social media spillovers from New Delhi, in columns 5 and 10 of Table 4.

Table 4: Robustness

Period:	Three Days					Seven Days				
Dependent variable:	Sh.Tweets FN	N. Tweets FN				Sh.Tweets FN	N. Tweets FN			
Specification:		N.Conflicts	Exposure	No New Delhi	Diffusion		N.Conflicts	Exposure	No New Delhi	Diffusion
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
New Delhi × Post	0.0132***	143.9636***	144.3918***		149.6529***	0.0074**	84.3648	84.9132		88.5060
	(0.0021)	(21.2835)	(20.4401)		(19.3889)	(0.0030)	(83.0657)	(83.0828)		(83.6984)
	[0.0000]	[0.0000]	[0.0000]		[0.0000]	[0.0129]	[0.3098]	[0.3068]		[0.2903]
Muslim Attack × Post	0.0094***			1.6349***	1.5230***	0.0078***			1.0831**	1.0026**
	(0.0028)			(0.5626)	(0.5693)	(0.0017)			(0.4896)	(0.4742)
	[0.0007]			[0.0037]	[0.0075]	[0.0000]			[0.0270]	[0.0345]
N. Muslim Attacks × Post		0.6511*					0.5181**			
		(0.3384)					(0.2103)			
		[0.0544]					[0.0137]			
Exp. Muslim Attack × Post			11.8095***					8.7965***		
			(3.1173)					(2.5939)		
			[0.0002]					[0.0007]		
New Delhi Re. w/o ND × Post					6.8163***					4.9472**
					(1.4907)					(2.0759)
					[0.0000]					[0.0172]
Conn. to New Delhi × Post					1.9352***					1.3589***
					(0.4513)					(0.4602)
					[0.0000]					[0.0032]
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.3215	0.9640	0.9648	0.9199	0.9651	0.2223	0.8116	0.8125	0.7883	0.8127
Observations	3756	3756	3756	3750	3756	8764	8764	8764	8750	8764

Notes: OLS estimates. Observations are districts in each day in two periods: March28–April 2 in columns 1–5 and March24–April 6 in columns 6–10. Each column reports a different version of specification 4 or 8 of Table 3, respectively; see the notes below that table for details on control variables. Columns 1 and 5 use as a dependent variable the daily share of tweets with anti-Muslim fake news, while all other columns use as a dependent variable the number of tweets with anti-Muslim fake news. Columns 2 and 6 replace the dummy for precolonial Muslim attacks with the number of Muslim attacks in the district. Columns 3 and 7 replace the dummy for precolonial Muslim attacks with a distance-based measure of exposure to Muslim attacks as in Dincecco et al. (2022). Columns 4 and 8 exclude New Delhi from the set of observations. Columns 5 and 10 account for spatial and social media spillovers from New Delhi as in Table 5. In all specifications, the newly included variables are interacted with the *Post-March30* dummy. See the text and the Appendix for details on all variables. Standard errors in parentheses are clustered to account for spatial correlation up to 250 km in the cross-section, and for both spatial and serial correlation in the panel specifications. P-values are reported in brackets. *** p<0.01, ** p<0.05, * p<0.1.

Finally, as in Section 5.1.1, Figures B10 to B15 display the sensitivity of the two coefficients of interest to the inclusion of further control variables and their interaction with the *Post-March30* dummy, over the three-day and seven-day windows. All results are qualitatively similar.

Altogether, the results support the existence of deeply rooted determinants of the increase in anti-Muslim fake news after the *Tablighi* shock. This result is in line with the historical narratives on the relevance of local Muslim invasions and governance for local Indian history, and with present-day anecdotal evidence (real life and on Twitter) recalling the legacy of these experiences (see Section 2).

5.2.2 Does Anti-Muslim Fake News Spread Spatially From Precolonial Muslim Attack Districts?

We now turn to exploring whether historical Muslim attacks predict the spatial diffusion of fake news over the three-day and seven-day windows before and after March 30. Similarly to the exercises carried out in Table 2, we proceed in three steps. Results are reported in Table 5.

We start from the specification of Table 3, column 8 by including a dummy for districts neighboring the district where precolonial Muslim attacks occurred. To do so, we rely on a spatial weighting matrix that identifies neighbors based on rook contiguity. Then, we compute the spatial lag of the *Muslim Attack* dummy and its interaction with the *Post-March30* dummy. Columns 1 and 4 show that the coefficients of the interaction term between the spatial-lag variables and the *Post-March30* dummy are generally negative and not significant. This exercise suggests that anti-Muslim fake news originated in districts exposed to precolonial Muslim attacks but did not spread to the neighboring districts.

Next, in columns 2 and 5, we augment the specification with a variable that records the distance from each district's centroid to the closest historical Muslim attack and its interaction with the *Post-March30* dummy. The coefficient on the interaction is negative and significant only over the three-day horizon, but the magnitude of the coefficient is very small, suggesting that the relationship between historical Muslim attacks and anti-Muslim fake news does not have notable spatial spillovers.

Finally, we check whether the network of social interactions with districts exposed to historical Muslim attacks plays any role in increasing anti-Muslim fake news. Similar to our analysis in Table 2 (columns 3 and 6), we now additionally control for a district-level measure of social connectedness with districts with historical Muslim attacks and its interaction with the *Post-March30* dummy. None of the coefficients is statistically significant, suggesting no social media spillovers as well.

Table 5: Number of Tweets With Anti-Muslim Fake News and Historical Muslim Attacks in Space

Period:	Three Days			Seven Days		
Dependent variable:	N. Tweets FN			N. Tweets FN		
	(1)	(2)	(3)	(4)	(5)	(6)
New Delhi × Post	144.4784***	145.4930***	144.4768***	85.2759	85.5007	85.5079
	(21.1308)	(20.9381)	(21.1311)	(83.1008)	(83.1841)	(83.1849)
	[0.0000]	[0.0000]	[0.0000]	[0.3048]	[0.3040]	[0.3040]
Muslim Attack × Post	1.7312***	1.4068**	1.7306***	1.0140**	0.9338**	0.9338**
	(0.6084)	(0.5696)	(0.6082)	(0.5152)	(0.4641)	(0.4641)
	[0.0044]	[0.0135]	[0.0044]	[0.0491]	[0.0442]	[0.0442]
W Muslim Attack × Post	-0.6032		-0.6153	0.4341		
	(0.8780)		(0.9169)	(0.7662)		
	[0.4921]		[0.5022]	[0.5710]		
Dist. Muslim Attack × Post		-0.0015**			-0.0010	-0.0010
		(0.0006)			(0.0009)	(0.0009)
		[0.0146]			[0.2746]	[0.2646]
Social Media Conn. to Muslim Attack Districts × Post			0.0530			-0.1480
			(0.3447)			(0.3309)
			[0.8778]			[0.6547]
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.9640	0.9641	0.9639	0.8115	0.8117	0.8117
Observations	3756	3756	3756	8764	8764	8764

Notes: OLS estimates. Observations are districts in each day in two periods: March28–April 2 in columns 1–3, and March24–April 7 in columns 4–6. In all specifications, the dependent variable is the number of tweets with anti-Muslim fake news, and baseline and geographic controls are as in the panel specifications of Table 3. *W Muslim Attack* considers the spatial lag of *Muslim Attack*, which is the neighbors of districts with historical Muslim attacks based on a (rook) contiguity matrix, and *Dist. Muslim Attack* is the distance from each district's centroid to the centroid of the closest district that experienced historical Muslim attacks. *Social Media Connectedness to Muslim Attack Districts* computes for each district *i* the share of quotes and replies to tweets posted in districts with historical Muslim attacks by users located in the district (excluding from the denominator the quotes and replies to tweets posted within the district). All variables are interacted by the dummy tracking dates after March 30, so their label is accompanied by × *Post* after the variable name. Standard errors in parentheses are clustered to account for both spatial correlation (up to 250 km) and serial correlation. P-values are reported in brackets. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

6 Who is disseminating fake news?

In the previous sections, we study the geography of anti-Muslim discrimination after a COVID-19 hotspot is reported to be linked with a Muslim convention in New Delhi. We have seen that the discrimination reaction is unevenly distributed in space, and we have uncovered a number of explanations for this phenomenon. We still don't know, however, whether discrimination and the dissemination of anti-Muslim fake news is also unevenly distributed across Twitter users and, in particular, whether discrimination is predominantly concentrated among people with specific characteristics.

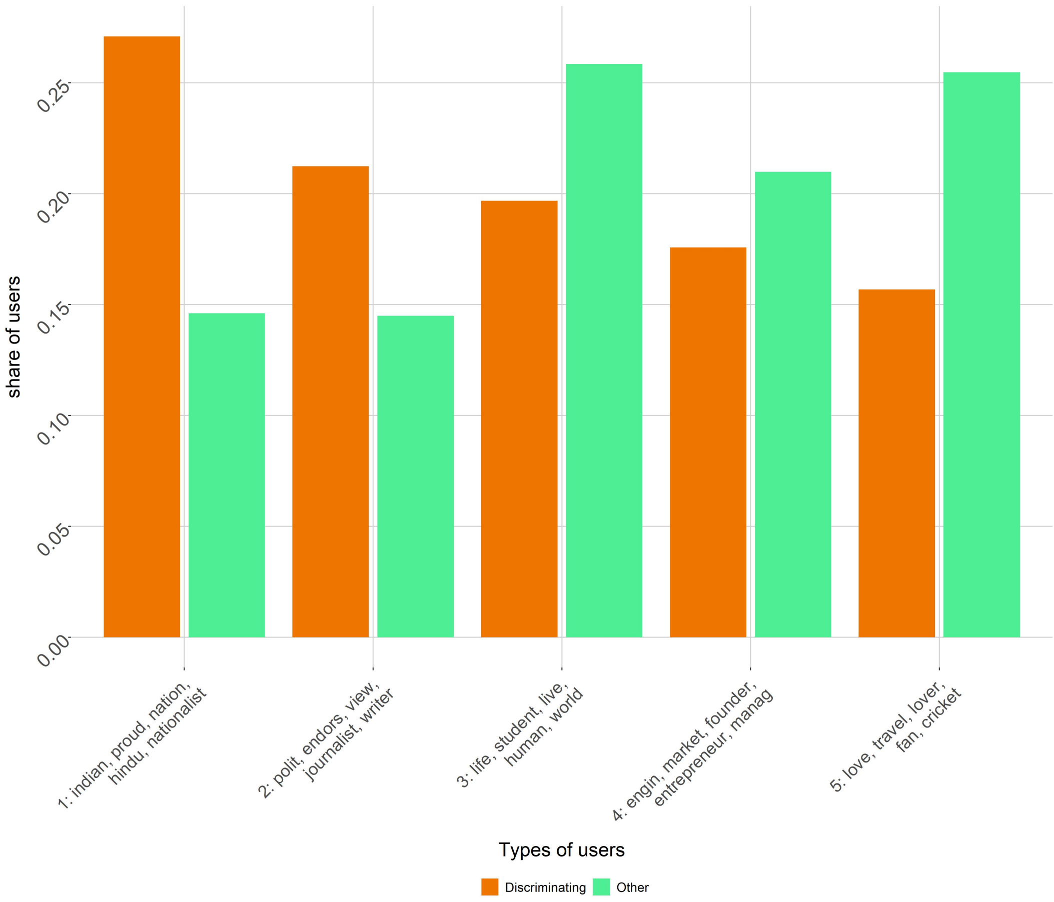
To dig deeper into these questions, we performed a topic analysis of users' self-reported biographies. This allows us to classify users into a limited number of types and better understand the characteristics of discriminatory users. First, we downloaded the self-reported biographies of all Twitter profiles associated with at least one discriminatory tweet during our main study period, i.e., from March 24 to April 3, 2020. We were able to recover the biographies for 99% (7,088) of these users. To generate a point of comparison, we set them against an equally sized random sample of users (whose self-reported biographies we also obtained) who never posted a discriminatory tweet from January 1 to May 6, 2020. Then, we ran an LDA analysis on this joint sample of 14,176 users in order to decompose their biographies into five latent topics.[25]. Finally, we defined a user type to correspond to the topic that receives the largest weight in his or her biography.[26]

Figure 7 plots the distribution of user types in the sample of discriminating profiles (in orange) versus the sample of nondiscriminatory profiles (in green). Topics are identified by the three words that receive the largest weight inside each topic and are ordered from left to right, in descending order, based on the share of discriminatory users who belong to that type. Two main messages stand out from the graph. First, discriminatory users are disproportionately Hindu nationalists (Topic 1) and people interested in politics (Topic 2). Second, nondiscriminatory users are more likely to belong to types that emphasize cosmopolitan attitudes (Topic 3), market-oriented issues (Topic 4) and leisure activities (Topic 5). However, as a partial qualification to this conclusion, the figure also illustrates that discrimination occurs across the board. For instance, we find a substantial share of discriminating users (above 15%) even in Topic 5, where this group is more heavily underrepresented.

[25]We dropped a subset of users (31% of discriminatory users, 42% percent of nondiscriminatory users) from the analysis because their biographies were empty, either to begin with or after we preprocessed the text. In particular, we removed all non-ASCII characters from the biographies, thus dropping most of the non-English parts of the texts. As a check, we repeated the analysis without removing non-ASCII characters; we obtained qualitatively similar results.

[26]In the (limited) case of ties, a user may have been assigned multiple types.

Figure 7



7 Conclusion

False stories spread rapidly on social media and the Internet. Given the large diffusion of these technologies in recent years, concern is growing over the spread of false stories and their social, economic, and political consequences.

In this paper, we study the fake news phenomenon under a novel perspective, namely as a vehicle to propagate hate and discriminatory attitudes toward minorities. In particular, we study the diffusion of false stories against Indian Muslims at the onset of the coronavirus outbreak in India, exploiting a tight sequence of events on March 30, 2020 that led many to identify a Muslim religious congregation (the *Tablighi Jamaat* convention) held in New Delhi as a COVID-19 hotspot. This coincided with an outburst of false stories reporting that Muslims were deliberately infecting other people and associating the spread of the virus with a form of jihad conducted by Muslim communities (see the trending hashtag "#coronajihad").

We leverage a comprehensive novel dataset of georeferenced text data from Twitter to document a large spike (from nearly zero to above 3%) in the share of tweets reporting false stories against Muslims and to investigate the spatial patterns of their diffusion.

Our econometric analysis delivers three sets of results. First, we find that, following the shock, the intensity of discriminatory fake news was strongest in New Delhi, where the *Tablighi* event took place. Second, beyond New Delhi, anti-Muslim false news was more pronounced in districts that are spatially closer and have more intense social media interactions with New Delhi, further highlighting the increasing role social networks play in diffusing discriminatory attitudes. Third, we show that the observed spatial differences in diffusion of fake news after the shock can also be linked to the legacy of precolonial Muslim attacks.

We build a novel classification of precolonial conflicts—events in which Muslim entities were the aggressors versus events in which Muslim entities were *not* the aggressors—and we map these events at the district level. We show that the diffusion of fake news in the aftermath of the *Tablighi* shock was stronger in districts where Muslim attacks occurred compared to districts that did not experience historical conflicts. Using state-of-the-art text analysis techniques, we provide suggestive evidence that discriminatory users are disproportionately Hindu nationalists and individuals active in political debate.

These findings on present-day India suggest that (epidemic) shocks may affect a country's overall environment of discrimination through the spread of anti-minority false news on social media. This is especially relevant in the case of (location-specific) persistent beliefs regarding the role of minorities as possible threats to national security and well-being.

It is still an open question how minorities react to the spread of fake news discriminating against them, whether they isolate themselves to protect their identity or whether they assimilate more into the local culture. More broadly, future research should also investigate more closely the dynamic interactions among minorities, members of the majority group, and political actors, who often contribute to the outbreak and diffusion of discriminatory false stories.

References

Adena, M., R. Enikolopov, M. Petrova, V. Santarosa, and E. Zhuravskaya (2015). Radio and the rise of the nazis in prewar germany. *The Quarterly Journal of Economics 130*(4), 1885–1939.

Allcott, H. and M. Gentzkow (2017). Social media and fake news in the 2016 election. *Journal of economic perspectives 31*(2), 211–36.

Anderson, D. M., B. Crost, and D. I. Rees (2020). Do economic downturns fuel racial animus? *Journal of Economic Behavior & Organization 175*, 9–18.

Anderson, R. W., N. D. Johnson, and M. Koyama (2017). Jewish persecutions and weather shocks: 1100–1800. *The Economic Journal 127*(602), 924–958.

Banerjee, A. and L. Iyer (2005). History, institutions, and economic performance: The legacy of colonial land tenure systems in india. *American economic review 95*(4), 1190–1213.

Bartoš, V., M. Bauer, J. Cahlíková, and J. Chytilová (2021). Covid-19 crisis and hostility against foreigners. *European Economic Review 137*, 103818.

Bentzen, J., N. Kaarsen, and A. Wingender (2017). Irrigation and Autocracy. *Journal of the European Economic Association 15*(1), 1–53.

Bharadwaj, P., A. I. Khwaja, and A. Mian (2015). Population exchange and its impact on literacy, occupation and gender–evidence from the partition of india. *International Migration 53*(4), 90–106.

Bharadwaj, P. and R. A. Mirza (2019). Displacement and development: Long term impacts of population transfer in india. *Explorations in Economic History 73*, 101273.

Blei, D. M., A. Y. Ng, and M. I. Jordan (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research 3*(Jan), 993–1022.

Brady, W. J., J. A. Wills, J. T. Jost, J. A. Tucker, and J. J. Van Bavel (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences 114*(28), 7313–7318.

Britannica (2022). Mahmūd, king of ghazna. Accessed April 14, 2022 [Online], available at https://www.britannica.com/biography/Mahmud-king-of-Ghazna .

Bursztyn, L., G. Egorov, R. Enikolopov, and M. Petrova (2019). Social media and xenophobia: evidence from russia. Technical report, National Bureau of Economic Research.

Castelló-Climent, A., L. Chaudhary, and A. Mukhopadhyay (2018). Higher education and prosperity: From catholic missionaries to luminosity in india. *The Economic Journal 128*(616), 3039–3075.

Chandler, T. (1987). *Four Thousand Years of Urban Growth: A Historical Census*. St. David's University Press.

Chaudhary, L., J. Rubin, S. Iyer, and A. Shrivastava (2020). Culture and colonial legacy: Evidence from public goods games. *Journal of Economic Behavior & Organization 173*, 107–129.

Conley, T. G. (1999). Gmm estimation with cross sectional dependence. *Journal of Econometrics 92*(1), 1–45.

Couttenier, M., S. Hatte, M. Thoenig, and S. Vlachos (2021). Anti-muslim voting and media coverage of immigrant crimes. *The Review of Economics and Statistics*, 1–33.

DellaVigna, S., R. Enikolopov, V. Mironova, M. Petrova, and E. Zhuravskaya (2014). Cross-border media and nationalism: Evidence from Serbian radio in Croatia. *American Economic Journal: Applied Economics 6*(3), 103–32.

Dincecco, M., J. Fenske, A. Menon, and S. Mukherjee (2022). Pre-colonial warfare and long-run development in india. *The Economic Journal 132*(643), 981–1010.

Dipoppa, G., G. Grossman, and S. Zonszein (2021). Locked down, lashing out: Situational triggers and hateful behavior towards minority ethnic immigrants. *Available at SSRN: https://ssrn.com/abstract=3789339*.

Doerr, S., S. Gissler, J.-L. Peydro, H.-J. Voth, et al. (2021). Financial crises and political radicalization: How failing banks paved hitler's path to power. Technical report, Bank for International Settlements.

Eaton, R. M. (2000). Temple desecration and indo-muslim states. *Journal of Islamic Studies 11*(3), 283–319.

Fenske, J., N. Kala, and J. Wei (2021). Railways and Cities in India. *Centre for Competitive Advantage in the Global Economy Working Paper 559*.

Griffiths, T. L. and M. Steyvers (2004, April). Finding scientific topics. *Proceedings of the National Academy of Sciences 101*(suppl 1), 5228–5235.

Grinberg, N., K. Joseph, L. Friedland, B. Swire-Thompson, and D. Lazer (2019). Fake news on twitter during the 2016 u.s. presidential election. *Science 363*(6425), 374–378.

Guess, A., J. Nagler, and J. Tucker (2019). Less than you think: Prevalence and predictors of fake news dissemination on facebook. *Science advances 5*(1), eaau4586.

Guess, A., B. Nyhan, and J. Reifler (2018). Selective exposure to misinformation: Evidence from the consumption of fake news during the 2016 us presidential campaign. *European Research Council 9*(3), 4.

Halarnkar, S. (2020). Coronavirus is proving to be another excuse to marginalise india's muslims. Accessed June 10, 2021 [Online], available at https://qz.com/india/1836768/coronavirus-is-another-excuse-to-marginalise-indias-muslims/ .

India Institute of Forest Management (2015). *India State of Forest Report*.

Inglehart, R., C. Haerpfer, A. Moreno, C. Welzel, K. Kizilova, J. Diez-Medrano, M. Lagos, P. Norris, E. Ponarin, B. Puranen, and et al. (2018). *World Values Survey: Round Two - India*. JD Systems Institute and WVSA Secretariat.

Iyer, L. (2010). Direct versus Indirect Colonial Rule in India: Long-Term Consequences. *Review of Economics and Statistics 92*(4), 693–713.

Jaques, T. (2007). *Dictionary of Battles and Sieges: A Guide to 8,500 Battles from Antiquity through the Twenty-first Century*. Greenwood Press.

Jedwab, R., N. D. Johnson, and M. Koyama (2019). Negative shocks and mass persecutions: evidence from the black death. *Journal of Economic Growth 24*(4), 345–395.

Jha, S. (2013). Trade, institutions, and ethnic tolerance: Evidence from south asia. *American political Science review 107*(4), 806–832.

Lu, R. and Y. Sheng (2022). How racial animus forms and spreads: Evidence from the coronavirus pandemic. *arXiv preprint arXiv:2007.01448*.

Matsuura, K. and C. Willmott (2009). Terrestrial Air Temperature: 1900-2008 Gridded Monthly Time Series. *Center for Climatic Research, University of Delaware*.

Michalopoulos, S. and E. Papaioannou (2018). Spatial patterns of development: A meso approach. *Annual Review of Economics 10*, 383–410.

Mukhia, H., B. Chandra, and R. Thapar (2017). Communalism and the writing of indian history.

Müller, K. and C. Schwarz (2020). From hashtag to hate crime: Twitter and anti-minority sentiment. *Available at SSRN 3149103*.

Müller, K. and C. Schwarz (2021). Fanning the flames of hate: Social media and hate crime. *Journal of the European Economic Association 19*(4), 2131–2167.

Nag, P. (2007). *Historical Atlas of India*. NATMO.

Nielsen (2022). Nielsen's bharat 2.0 study reveals a 45users in rural india since 2019. Accessed May 12, 2022 [Online], available at https://www.nielsen.com/in/en/press-releases/2022/nielsens-bharat-2-0-study-reveals-a-45-growth-in-active-internet-users-in-rural-india-since-2019/.

Ochsner, C. and F. Roesel (2019). Mobilizing history. Technical report, CERGE University Mimeo.

Outlook (2020). Communal corona? is it justified to blame tablighi jamaat for nizamuddin outbreak? Accessed April 14, 2022 [Online], available at https://www.outlookindia.com/website/story/india-news-corona-outbreak-afflicted-by-communal-virus-blaming-tablighi-jamaat-could-be-misdirected/349784.

Perrigo, B. (2020). It was already dangerous to be muslim in india. then came the coronavirus. *Times*.

Pew Research Center (2021). Religion in india: Tolerance and segregation. Available at https://www.pewresearch.org/religion/2021/06/29/religion-in-india-tolerance-and-segregation .

Pooja Chaudari, Alt News (2020). The year that was: Misinformation trends of 2020. Accessed April 14, 2022 [Online], available at https://www.altnews.in/the-year-that-was-misinformation-trends-of-2020/.

Press Trust of India (2020). Nizamuddin congregation: Arvind kejriwal orders fir against maulana. Accessed April 14, 2022 [Online], available at https://www.indiatoday.in/india/story/nizamuddin-congregation-arvind-kejriwal-orders-fir-against-maulana-1661514-2020-03-30.

Ritika Jain, Article14 (2020). Covid-19: How fake news and modi government messaging fuelled india's latest spiral of islamophobia. Accessed April 14, 2022 [Online], available at https://scroll.in/article/959806/covid-19-how-fake-news-and-modi-government-messaging-fuelled-indias-latest-spiral-of-islamophobia.

Sachau, E. C. (2013). *Alberuni's India: An Account of the Religion, Philosophy, Literature, Geography, Chronology, Astronomy, Customs, Laws and Astrology of India: Volume I*. Routledge.

Sarkar, S. J. (1930). *A-Short History of Aurangzib*.

Schwartzberg, J. (1978). *A Historical Atlas of South Asia*. University of Chicago Press.

Sekose, M. (2021). 2021 in social media — creators get more ways to earn money, live audio rooms become mainstream and more. *Business Insider India*.

THE WEEK (2020). Covid-19: Centre blames tablighi jamaat for sudden spike in cases. Accessed May 12, 2022 [Online], available at https://www.theweek.in/news/india/2020/04/01/covid-19-centre-blames-tablighi-jamath-for-sudden-spike-in-cases.html.

Times of India (2020). Coronavirus: About 9,000 tablighi jamaat members, primary ontacts quarantined in country, mha says. *Times of India*.

Tollefsen, A., H. Strand, and H. Buhaug (2012). PRIO-GRID: A Unified Spatial Data Structure. *Journal of Peace Research 49*(2), 363–74.

Verghese, A. (2016). *The colonial origins of ethnic violence in India*. Stanford University Press.

Voigtländer, N. and H.-J. Voth (2012). Persecution perpetuated: the medieval origins of anti-semitic violence in nazi germany. *The Quarterly Journal of Economics 127*(3), 1339–1392.

Vosoughi, S., D. Roy, and S. Aral (2018). The spread of true and false news online. *Science 359*(6380), 1146–1151.

Voth, H.-J. (2021). Persistence–myth and mystery. In *The handbook of historical economics*, pp. 243–267. Elsevier.

Wolpert, S. A. (2004). *A new history of India*. Oxford University Press, USA.

Yanagizawa-Drott, D. (2014). Propaganda and conflict: Evidence from the Rwandan genocide. *The Quarterly Journal of Economics 129*(4), 1947–1994.

Zhuravskaya, E., M. Petrova, and R. Enikolopov (2020). Political effects of the internet and social media. *Annual Review of Economics 12*, 415–438.

A Data Appendix

A.1 Details on fake news and keywords

In this section we report in Table A1 the list of fake news we collected from the website AltNews and in Table A2 the full list of English keywords we employed to identify tweets reporting fake news. These keywords were also translated in other 11 languages. Keywords in other languages are availble upon request.

Table A1: Fake News with "Muslim" in the Title or Abstract

Date	Origin	Fake News
02Mar		Suresh Chavhanke falsely claims Muslim community in Paris torched railway station
02Mar		Muslim girl raped by Hindu mob in Delhi? Pak propaganda website runs fake news
02Mar		Muslim shops sell biryani laced with birth control pills to Hindus? Fictional story viral
04Mar		AAP offers monetary relief to only Muslim victims of Delhi riots? Dainik Jagran clipping morphed
12Mar	05Mar	Delhi riots: Times Now misreports man firing at Muslim mob as attack on police
13Mar		First Muslim woman SP in Maharashtra? No, image of Women's Day celebration viral
16Mar	14Mar	Public TV falsely claims Muslim youths in Karnataka refuse coronavirus testing for "religious reasons"
30Mar		Old, unrelated video shared as Muslims licking utensils to spread coronavirus infection
01Apr		Video of Sufi ritual falsely viral as mass sneezing in Nizamuddin mosque to spread coronavirus infection
02Apr		Old video falsely viral as Muslim man spitting on food at Indian restaurant in the backdrop of coronavirus pandemic
02Apr		Coronavirus: Video of an undertrial in Mumbai falsely viral as Nizamuddin markaz attendee spitting at cop
04Apr		Video of Muslim vendor's unhygienic handling of fruits falsely linked with spreading coronavirus
04Apr		Old video of racist heckling falsely viral as Muslim man spits on passenger in New York metro
05Apr		Viral audio: False conspiracy theory about Modi govt introducing 'vaccine' to kill Muslims
06Apr		Video from Pakistan falsely viral as Muslims punished in India
06Apr		Video of Muslims exiting quarantine centre: Not Vinayaka Temple but residential lodging
07Apr		Old video from Philippines shared with false claim of Muslim man spitting on bread
07Apr		Viral audio falsely claims Muslim vendors have sprung up in Surat to spread coronavirus
08Apr		Old video where salon attendant applies saliva on customer's face falsely shared with Muslim angle
09Apr		Communal attack in Bawana shared with false claim of Muslim man injecting fruits with spittle
09Apr		Death of health worker in MP falsely communalised as attack by Muslims in UP "Islamic jihadis"
11Apr		Video viral with false claim that Muslims scatter notes on the road to spread coronavirus
11Apr		Video from Pak falsely linked with Hindu man's alleged murder by Muslim men in Rajasthan
13Apr		Alt News video verification: Muslim vegetable vendor assaulted in Badarpur, Delhi
13Apr		UP police's mock drill video shared as 'corona Jihadis' arrested during lockdown
14Apr		Image of Muslims offering namaz on rooftops in groups is from Dubai
15Apr	06Apr	Video of fruit vendors in Indore shared with false anti-Muslim angle
15Apr	13Apr	Pakistani Mufti provoking people to flout lockdown shared to target Indian Muslims
16Apr	13Apr	Video of women spitting inside houses in Rajasthan's Kota given false Muslim angle
18Apr		False claim suggests Bandra mass gathering accused Vinay Dubey's father is Muslim
20Apr		Video of currency notes in Indore falsely viral as 'Muslim conspiracy' to spread coronavirus
20Apr		Videos viral with false claim of poor slum dwellers and Muslims hoarding food in Meerut
24Apr	21Apr	Video from Bijnor viral with false allegation that elderly Muslim vendor sprinkled urine on fruits
27Apr		Disabled Muslim man hounded for accidentally dropping currency, accused of spreading coronavirus
27Apr		Zee News publishes 2015 story with false claim of human faeces served to 'non-Muslims'
29Apr		Old video falsely shared as Muslims spitting on relief food during lockdown

Table A2: English Fake News Keywords

#BioJihad	#BiologicalJihad	#coronaJehad	#IslamicCoronaJehad
#Islamiccoronavirusjihad	#IslamicRepublicVirus	#IslamicVirus	#jahiljamati
#JamaatKaCoronaDisaster	#JamaatkiGundagardi	#JamaatKoBanKaro	#JAMATI_CORONA_JEHAD
#JamatVirus	#JehadiVirus	#JihadiJamat	#MarkazCOVIDSpread
#MuslimsSpreadingCorona	#muslimvirus	#MuslimVirus	#NizamuddinIdiots
#NizamuddinMarkaj	#nizamuddinterrorists	#QuranaVirus	#TablighiJamatVirus
#TablighisInHiding	bio jihad	biological jihad	Corona J-had
corona jihaad	corona jihad	covid jehad	covid J-had
covid jihaad	covid jihad	covidjehad	covidjihad
islamic virus	jamat virus	jehad jamat	jihaad jamat
jihad jamat	muslim corona	Muslim infect	muslim infected
muslim infecting	muslim infects	muslim lick	muslim licked
muslim licking	muslim licks	Muslim pee	muslim peeing
muslim pees	muslim spat	Muslim spit	muslim spits
muslim spitting	Muslim sprinkle	muslim sprinkled	muslim sprinkles
muslim sprinkling	muslim stones	Muslim urine	muslim virus
muslims corona	muslims infect	muslims infected	muslims infecting
muslims lick	muslims licked	muslims licking	muslims pee
muslims peed	muslims peeing	muslims spit	muslims spitted
muslims spitting	muslims spread	muslims spreaded	muslims spreading
muslims sprinkle	muslims sprinkled	muslims sprinkling	muslims stones
muslims urine	muslin peed	Nizammuddin jihad	qurana virus
Tablighi excreted	Tablighi crime	Tablighi grope	Tablighi harass
tablighi harassed	tablighi harassing	Tablighi lewd	Tablighi naked
tablighi virus	Tabligi 1000 positive	#IslamicJihad	muslim jihad
CrushTablighiSpitters	tableeghi	CoronaBombsTablighi	markazcovidspread
Tablighi Talibani crime	CoronaTerrorism	jihadi weapon	MuslimMeaningTerrorist
corona bomb	islamiccoronajehad	jamaatkacoronadisaster	nizamuddincoronacases
crushtablighispitters	jihadivirus	markaznizamuddin	nijamuddinmarkaz
nizammudin	nizamuddinfiasco	nizamuddinmarkaz	tableeghijamaat
tableegi	tablighijamat	tablighijammat	tablighis
tablighivirus	tablighjamaat	tablighsuperspreader	tabligi
terrorist tablighi	tabligih traitors	bantablighdebate	markazvirus
MuslimDistancing	human bombs		

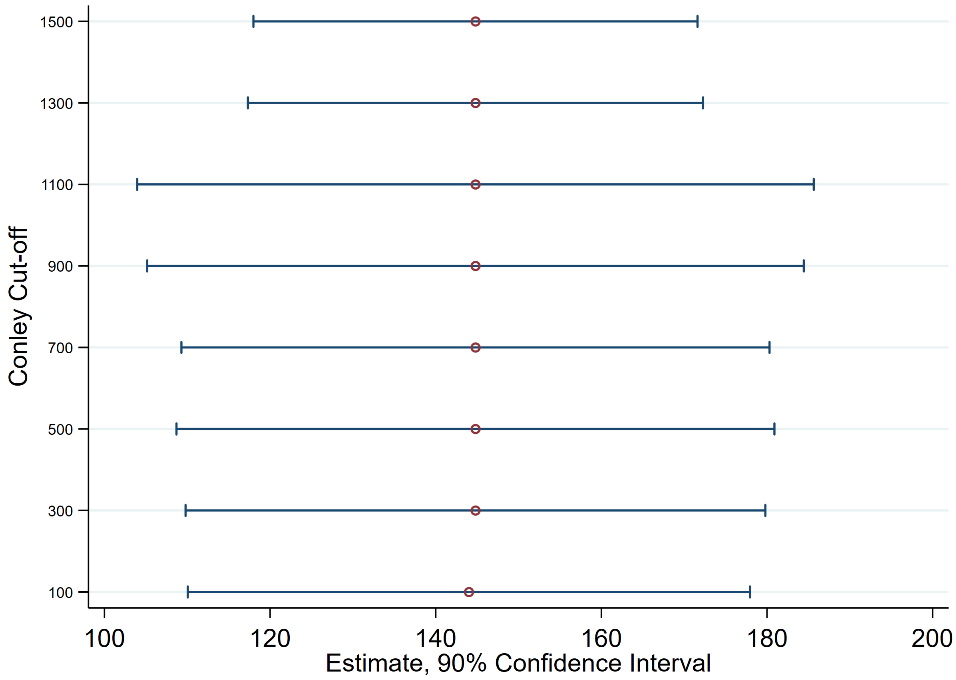
A.2 Summary Statistics

Table A3: Summary Statistics

Variable	Obs	Mean	Std dev	Median	Min	Max
Time-variant Variables, 3 days:						
N. Tweets Fake News	3756	3.0586	19.800	0	0	556
Share Tweets Fake News	3756	0.0110	0.03141	0	0	0.7143
New COVID Deaths	3756	0.0138	0.1721	0	0	5
Time-variant Variables, 7 days:						
N. Tweets Fake News	8764	2.0992	14.680	0	0	556
Share Tweets Fake News	8764	0.0074	0.02551	0	0	0.7143
New COVID Deaths	8764	0.0124	0.1748	0	0	8
Time invariant variables:						
New Delhi	8764	0.0016	0.03994	0	0	1
New Delhi Region w/o ND	8764	0.0336	0.1801	0	0	1
Dist. New Delhi (x1000)	8764	0.9789	0.5586	0.9496	0.009311	2.2592
Social Media Connectedness to New Delhi	8764	0.1836	0.1791	0.1667	0	1
Muslim Attack	8764	0.1374	0.3443	0	0	1
W Muslim Attack	8764	0.1271	0.1749	0	0	0.8034
Dist. Attack	8764	169.11	176.44	104.62	1.1684	929.27
Non-Muslim Attack	8764	0.0703	0.2556	0	0	1
Social Media Conn. to Districts w/Muslim Attacks	8764	0.2594	0.2023	0.2500	0	1
ln(0.01+Luminosity)	8764	0.7031	1.4884	1.0117	-4.6052	4.1373
ln(Population density), 1990	8764	5.4952	1.1606	5.5748	-1.4356	10.610
ln Share Muslim	8764	-2.7718	1.2934	-2.5669	-6.0956	-0.01496
ln Share Hindu	8764	-0.5015	0.8974	-0.1610	-4.7619	-0.006113
Ln Sh. Literate Pop. (2011)	8764	-0.4737	0.1719	-0.4658	-1.2116	-0.1083
Ln Sh. Urban Pop. (+0.01 2011)	8764	-1.5807	0.7165	-1.5734	-4.6052	0.009950
N.Tweets Preshock	8764	208.06	906.91	48.857	0	15391.7
Latitude	8764	23.441	5.7440	24.594	8.3060	34.527
Longitude	8764	81.068	6.2619	79.226	69.802	96.827
Altitude	8764	486.93	718.05	253.10	4	4914.9
Ruggedness	8764	102007.0	165566.0	35094.4	773.67	851959.5
Precipitation	8764	1354.3	674.96	1161.6	200.22	4245.3
Land Quality	8764	0.4590	0.2929	0.5267	0	0.9720
Dry Rice Suitability	8764	620.75	591.41	789.31	0	1722.7
Wet Rice Suitability	8764	1430.1	791.46	1403.3	0	2826.9
Wheat Suitability	8764	636.76	578.73	608.95	0	2914.7
Malaria Risk	8764	0.1055	0.3323	0.03298	0	2.8075
Min. Distance Pakistan or Bangladesh	8764	494.67	446.51	385.18	0	1862.8
Distance: Border	8764	402403.3	473726.9	218057.5	0	1862808.9
Neolithic Settlements	8764	0.3834	1.5827	0	0	20
Chacolithic Settlements	8764	0.3067	1.4194	0	0	19
Cultural Sites (300-700 CE)	8764	0.1581	0.4581	0	0	4
Cultural Sites (8th-12th centuries)	8764	0.6901	1.2525	0	0	10
Ln(1+ Urban population in 1000)	8764	0.07141	0.8906	0	0	11.513
Distance: Coast	8764	410445.1	342190.6	333208.2	0	1246877.5
River	8764	0.5990	0.4901	1	0	1
Irrigation Potential	8652	0.2022	0.3311	0.0005263	0	1
CV Rainfall : Delaware	8764	0.2303	0.07202	0.2180	0.09513	0.5302
Percent Forest	8764	21.327	24.491	11.305	0	93.980
ln Distance Petroleum	8764	5.4803	0.7705	5.6335	1.7820	6.6926
ln Distance Diamond: Primary	8764	6.6869	0.5575	6.8268	3.7101	7.4802
ln Distance Gem	8764	4.9553	0.8702	5.0911	2.2061	6.3544
ln Distance Gold Placer	8764	6.3003	0.7105	6.5463	3.5410	7.1669
British direct rule	8428	0.6478	0.4777	1	0	1
Years British Rule	8764	87.882	72.709	112	0	286
Year of First Railroad	6706	1886.2	18.001	1886	1853	1931
Medieval Port	8764	0.0639	0.2446	0	0	1
Duration of Muslim Rule	8764	368.33	235.72	387	0	995
Religious Polarization	8764	0.4744	0.2630	0.4415	0.02417	0.9948
Linguistic Fractionalization	8764	0.4619	0.2781	0.4679	0.01435	4.2053
Religious Fractionalization	8764	0.2646	0.1619	0.2330	0.01215	0.7156
Scheduled Caste Share	8764	0.1493	0.09117	0.1579	0	0.5017
Scheduled Tribe Share	8764	0.1785	0.2694	0.04432	0	0.9858
Ganges	8764	0.0831	0.2760	0	0	1

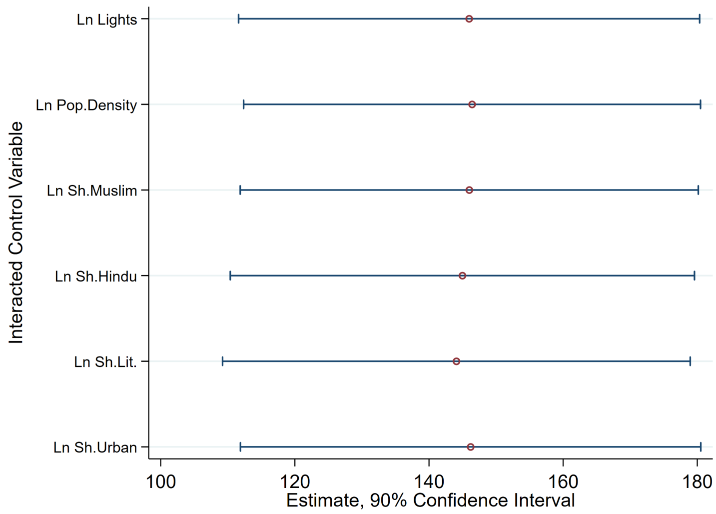
B Robustness on the Effect of the ”Tablighi Shock” on New Delhi

Figure B1: Robustness: *New Delhi* × *Post* with Different Conley Thresholds, 3 Day Windows



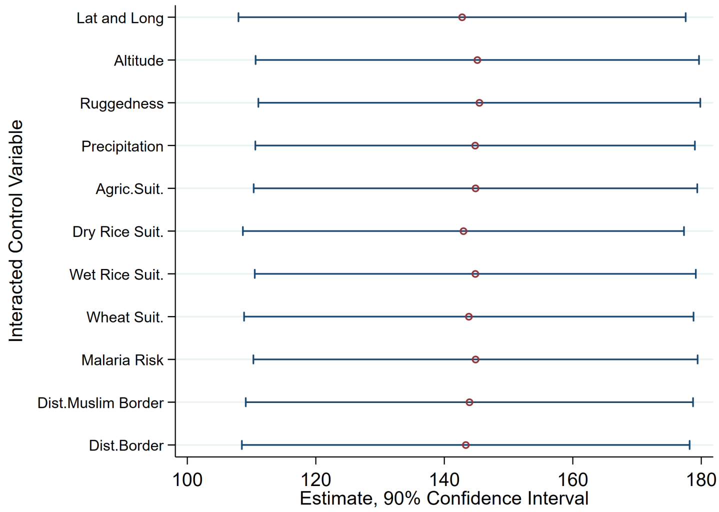
Notes: Each dot is the coefficient associated with the variable *New Delhi* × *Post* from estimating specification 4 in Table 1 with Conley standard errors at different distance thresholds (on the vertical axis). Horizontal bars indicate 90% confidence intervals.

Figure B2: Robustness: Baseline Controls Interacted by *Post March 30*



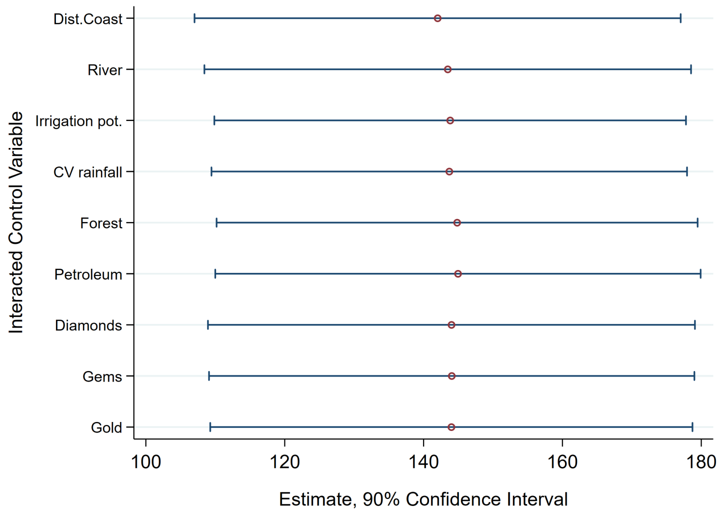
Notes: Each dot is the coefficient associated with the variable *New Delhi* × *Post* from a version of specification 4 in Table 1 that additionally controls for the variable displayed on the vertical axis and its interaction term with the *Post March 30* dummy. Horizontal bars indicate 90% confidence intervals.

Figure B3: Robustness: Geographical Controls Interacted by *Post March 30*



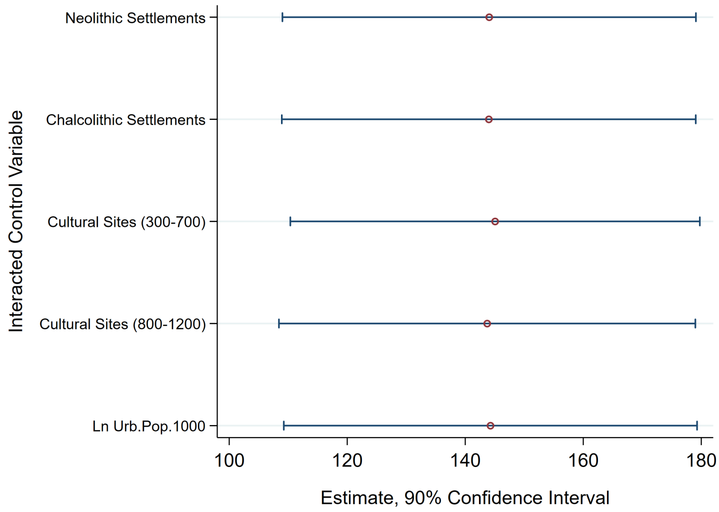
Notes: Each dot is the coefficient associated with the variable *New Delhi* × *Post* from a version of specification 4 in Table 1 that additionally controls for the variable displayed on the vertical axis and its interaction term with the *Post March 30* dummy. Horizontal bars indicate 90% confidence intervals.

Figure B4: Robustness: Further Geographical Controls Interacted by *Post March 30*



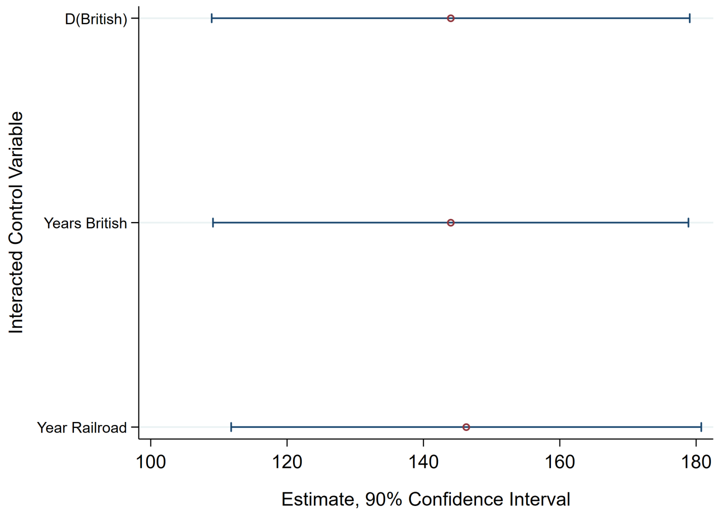
Notes: Each dot is the coefficient associated with the variable *New Delhi* × *Post* from a version of specification 4 in Table 1 that additionally controls for the variable displayed on the vertical axis and its interaction term with the *Post March 30* dummy. Horizontal bars indicate 90% confidence intervals.

Figure B5: Robustness: Historical Controls Interacted by *Post March 30*



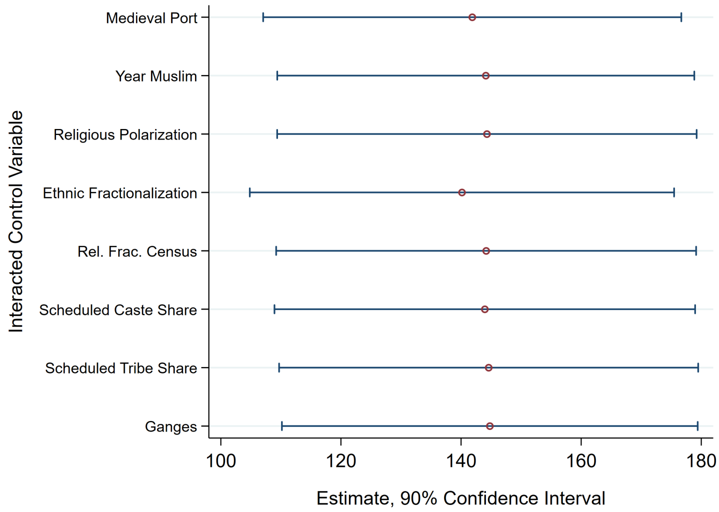
Notes: Each dot is the coefficient associated with the variable *New Delhi* × *Post* from a version of specification 4 in Table 1 that additionally controls for the variable displayed on the vertical axis and its interaction term with the *Post March 30* dummy. Horizontal bars indicate 90% confidence intervals.

Figure B6: Robustness: Colonization Controls Interacted by *Post March 30*



Notes: Each dot is the coefficient associated to the variable *New Delhi* × *Post* from a version of specification 4 in Table 1 that additionally controls for the variable displayed on the vertical axis and its interaction term with the *Post March 30* dummy. Horizontal bars indicate 90% confidence intervals.

Figure B7: Robustness: Fractionalization Controls Interacted by *Post March 30*



Notes: Each dot is the coefficient associated with the variable *New Delhi* $\times$ *Post* from a version of specification 4 in Table 1 that additionally controls for the variable displayed on the vertical axis and its interaction term with the *Post March 30* dummy. Horizontal bars indicate 90% confidence intervals.

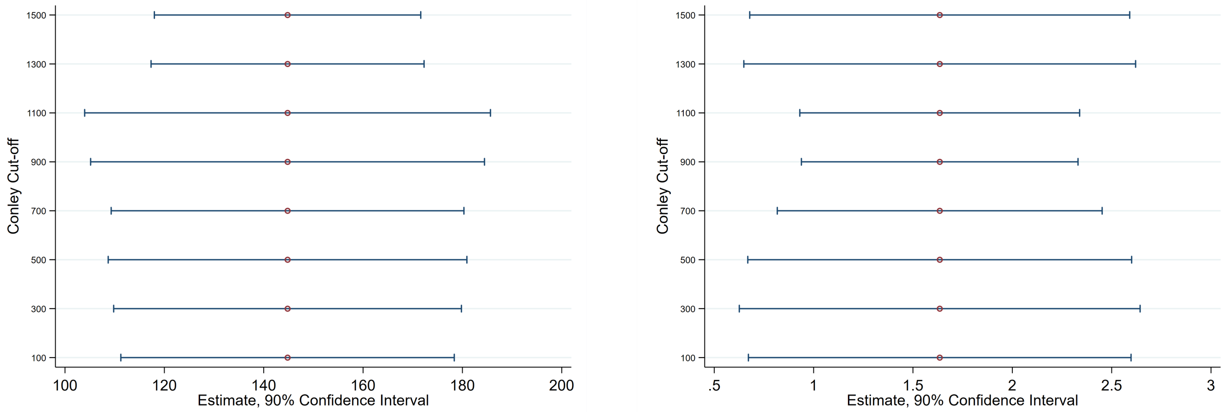
Table B1: Robustness: Share of Tweets with Anti-Muslim Fake News and the Tablighi Shock

Period:	Three Days		Seven Days	
Dependent variable:	Sh. Tweets FN		Sh. Tweets FN	
	(1)	(2)	(3)	(4)
New Delhi $\times$ Post	0.0145***	0.0145***	0.0100***	0.0097***
	(0.0011)	(0.0012)	(0.0026)	(0.0025)
	[0.0000]	[0.0000]	[0.0001]	[0.0001]
Baseline controls	Yes	Yes	Yes	Yes
Geographic controls	Yes	No	Yes	No
State FE	Yes	No	Yes	Yes
District FE	No	Yes	No	Yes
Day FE	Yes	Yes	Yes	Yes
R-squared	0.1487	0.3187	0.1278	0.2196
Observations	3756	3756	8764	8764

Notes: OLS estimates. Observations are districts in each day in two periods: March 28–April 2 in columns 1–2, and March24–April 6 in columns 3–4. In all specifications, we control for day fixed effects. Baseline controls include the log of luminosity (+0.01) averaged over the period 1992-2010, the log of population density in 1990, the log share of Muslim and Hindu population in 2011, the log share of literate and urban population in 2011, and the daily number of COVID deaths. Geographical controls include latitude, longitude, altitude, ruggedness, precipitation, land quality, dry rice suitability, wet rice suitability, wheat suitability, malaria risk, distance to the border, and distance to the closest border between Pakistan and Bangladesh. Odd columns additionally control for state fixed effects, while even columns control for district fixed effects. See the text and the Appendix for details on all variables. Standard errors in parentheses are clustered to account for spatial correlation up to 250 km in the cross-section, and for both spatial and serial correlation in the panel specifications. P-values are reported in brackets. *** p<0.01, ** p<0.05, * p<0.1.

C Robustness on the "Tablighi shock" for Muslim Attack Districts

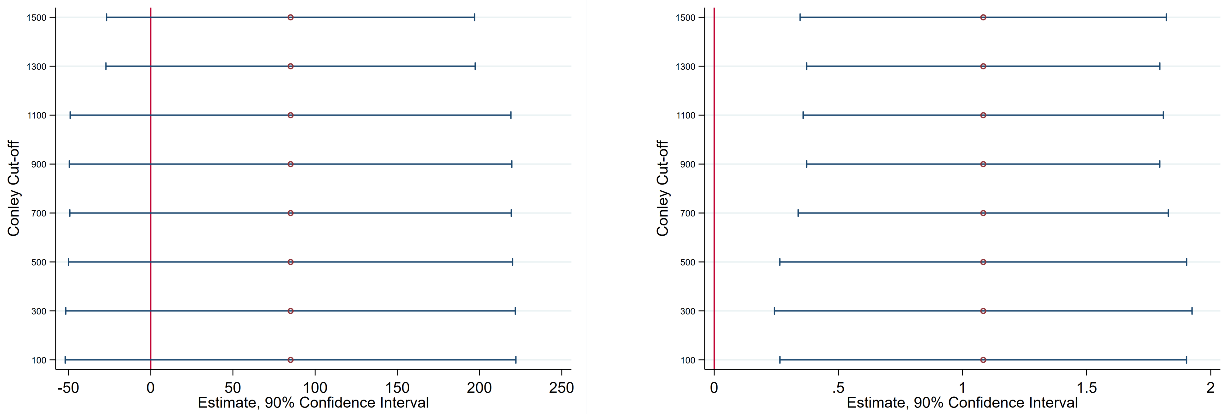
Figure B8: Robustness: Different Conley Thresholds, 3 Day Windows



(a) New Delhi × Post (b) Muslim Attack × Post

Notes: Each dot is the coefficient associated with the variable *New Delhi* × *Post* (Panel a) and *Muslim Attack* × *Post* (Panel b) from estimating specification 4 in Table 3 with Conley standard errors at different distance thresholds (on the vertical axis). Horizontal bars indicate 90% confidence intervals.

Figure B9: Robustness: Different Conley Thresholds, 7 Day Windows



(a) New Delhi × Post (b) Muslim Attack × Post

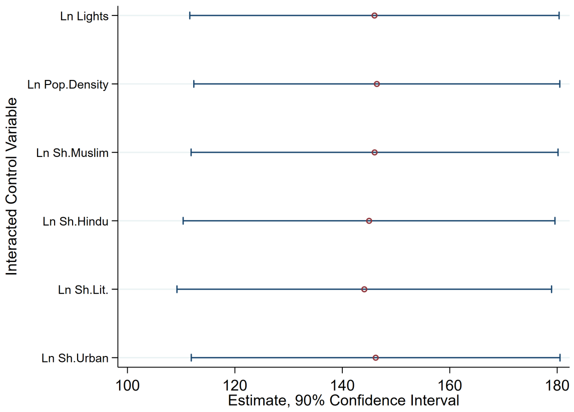
Notes: Each dot is the coefficient associated with the variable *New Delhi* × *Post* (Panel a) and *Muslim Attack* × *Post* (Panel b) from estimating specification 8 in Table 3 with Conley standard errors at different distance thresholds (on the vertical axis). Horizontal bars indicate 90% confidence intervals.

Table B2: Robustness: Alternative Definitions of Muslim-Related Conflict

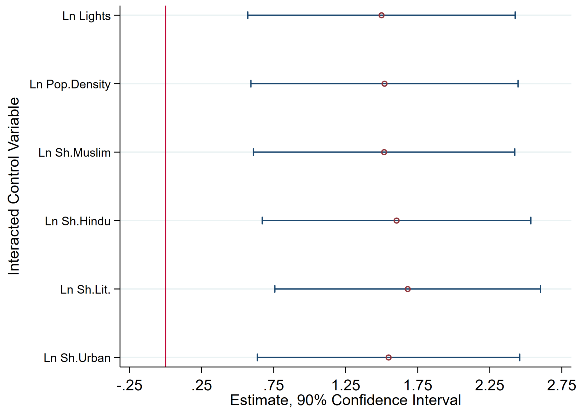
Period:	Three Days					Seven Days				
Dependent variable:	N. Tweets FN					N. Tweets FN				
Specifications:	1200-1757	1000-1840	Radius 100	Radius 5000	Involvement	1200-1757	1000-1840	Radius 100	Radius 5000	Involvement
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
New Delhi × Post	144.8470***	143.6546***	144.9899***	144.0235***	145.2837***	85.0758	84.4665	85.3282	84.6527	85.1829
	(20.9861)	(21.3177)	(20.4001)	(20.4614)	(21.1516)	(83.0797)	(83.1200)	(83.0375)	(83.1264)	(83.1377)
	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.3058]	[0.3095]	[0.3041]	[0.3085]	[0.3056]
Muslim Attack × Post	1.9203***	1.3858***			1.8071***	1.2675**	0.8728**			1.2121**
	(0.6041)	(0.5124)			(0.5869)	(0.5003)	(0.4408)			(0.4808)
	[0.0015]	[0.0068]			[0.0021]	[0.0113]	[0.0477]			[0.0117]
Muslim Conflict Exposure × Post			15.8066***	11.1052***				11.2809***	8.1366***	
			(4.2933)	(2.5520)				(3.6979)	(2.2476)	
			[0.0002]	[0.0000]				[0.0023]	[0.0003]	
Involved, No Attack × Post					-3.1134***					-1.7381***
					(0.5907)					(0.4718)
					[0.0000]					[0.0002]
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.9641	0.9639	0.9649	0.9650	0.9641	0.8117	0.8116	0.8124	0.8127	0.8118
Observations	3756	3756	3756	3756	3756	8764	8764	8764	8764	8764

Notes: OLS estimates. Observations are districts in each day in the March 24 –April 7 period . Each specification substitutes *Muslim Attack* × *Post* from specification 8 of Table 3 with an alternative definition of historical exposure to Muslim attacks (interacted with the *Post March 30* dummy). Columns 1–2 and 6–7 perturbs the period over which the variables are computed and focus on the 1200-1757 period and 1000-1840 period, respectively. Colums 3–4 and 8–9 compute exposure to Muslim attacks as in Dincecco et al. (2022) over distances up to 100 and 5,000 km, respectively. Columns 5 and 10 explicitly account for conflicts in which Muslim groups were involved but did not directly attack the district, through the variable *Involved, No Attack* interacted with the *Post March 30* dummy. See the text and the Appendix for details on all variables. Standard errors in parentheses are clustered to account for spatial correlation up to 250 km in the cross-section, and for both spatial and serial correlation in the panel specifications. P-values are reported in brackets. *** p<0.01, ** p<0.05, * p<0.1.

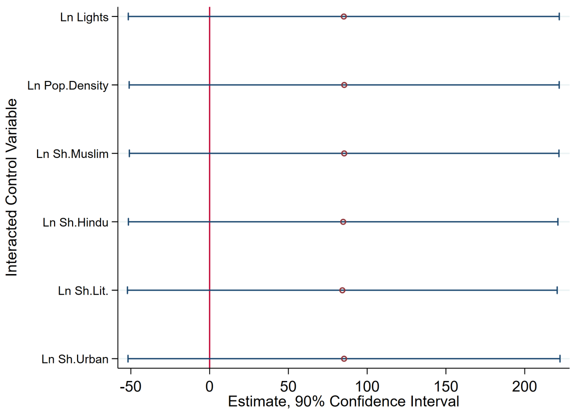
Figure B10: Robustness: Baseline Controls Interacted by *Post March 30*



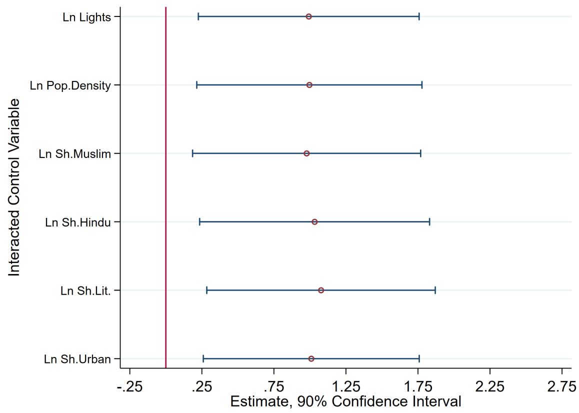
(a) New Delhi × Post, 3 Days



(b) Muslim Attack × Post, 3 Days



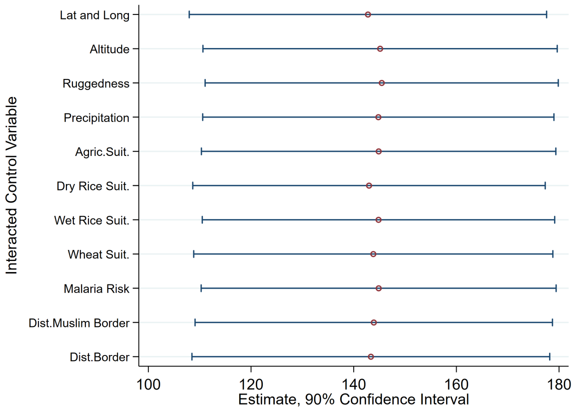
(c) New Delhi × Post, 7 Days



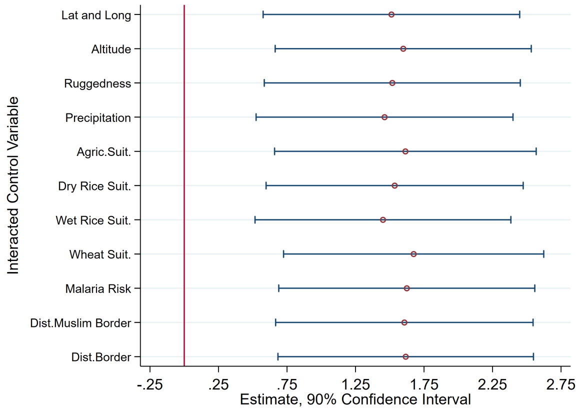
(d) Muslim Attack × Post, 7 Days

Notes: Each dot is the coefficient associated with the variables *New Delhi* × *Post* and *Muslim Attack* × *Post* from a version of specification 8 in Table 3 that additionally controls for the variable displayed on the vertical axis interacted with the *Post March 30* dummy. Panels a and b report estimates over the 3 day windows while Panels c and d display estimates over the 7 day window. Horizontal bars indicate 90% confidence intervals.

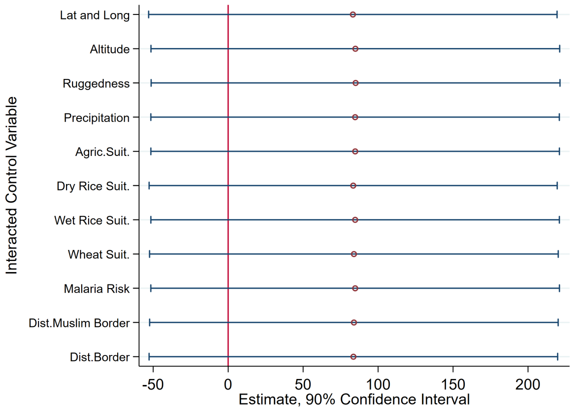
Figure B11: Robustness: Geographical Controls Interacted by *Post March 30*



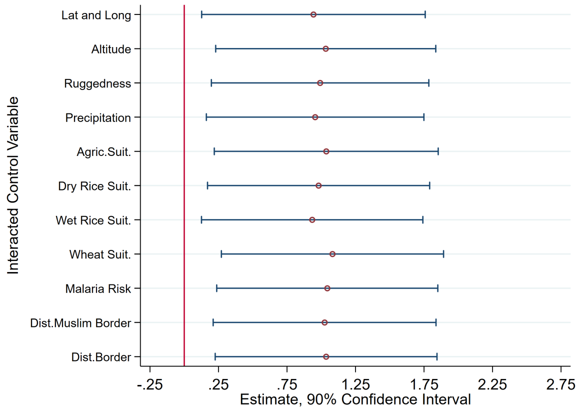
(a) New Delhi × Post, 3 Days



(b) Muslim Attack × Post, 3 Days



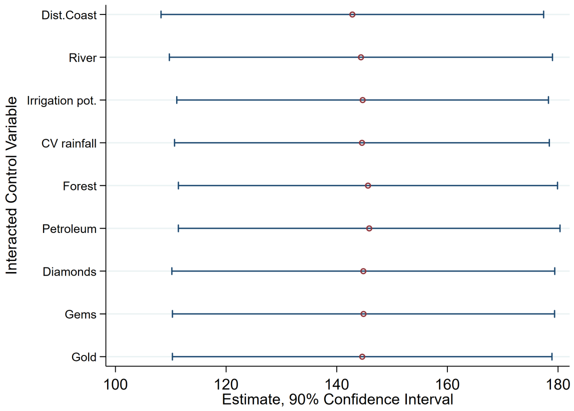
(c) New Delhi × Post, 7 Days



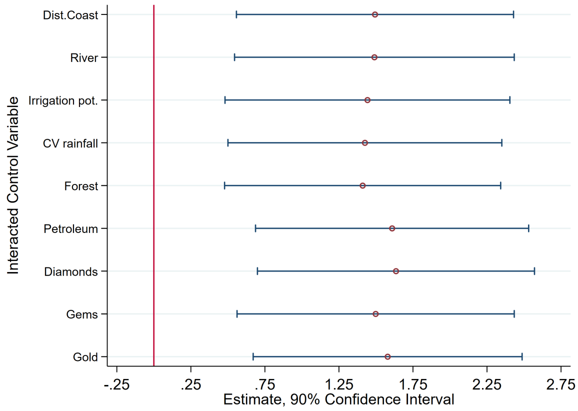
(d) Muslim Attack × Post, 7 Days

Notes: Each dot is the coefficient associated with the variables *New Delhi* × *Post* and *Muslim Attack* × *Post* from a version of specification 8 in Table 3 that additionally controls for the variable displayed on the vertical axis interacted with the *Post March 30* dummy. Panels a and b report estimates over the 3 day windows while Panels c and d display estimates over the 7 day window. Horizontal bars indicate 90% confidence intervals.

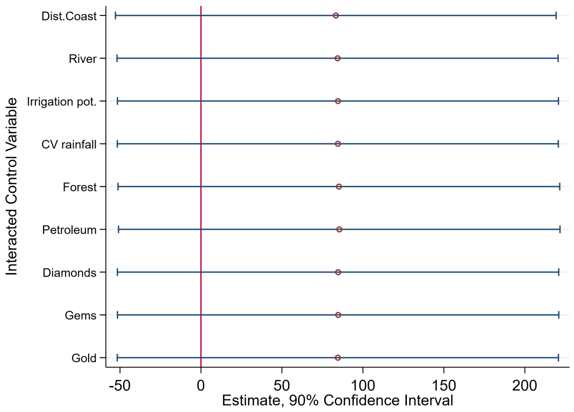
Figure B12: Robustness: Further Geographical Controls Interacted by *Post March 30*



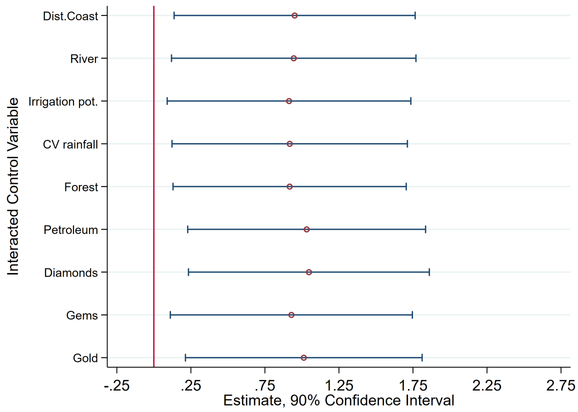
(a) New Delhi × Post, 3 Days



(b) Muslim Attack × Post, 3 Days



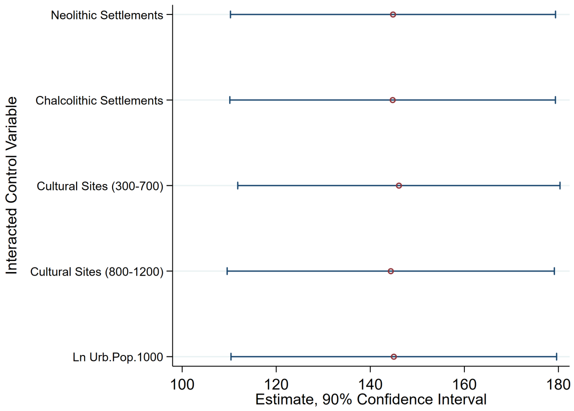
(c) New Delhi × Post, 7 Days



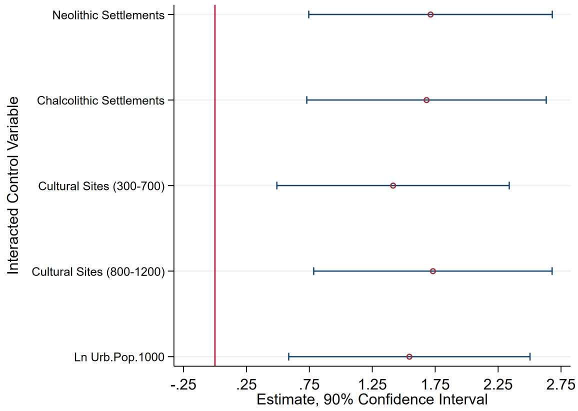
(d) Muslim Attack × Post, 7 Days

Notes: Each dot is the coefficient associated to the variables *New Delhi* × *Post* and *Muslim Attack* × *Post* from a version of specification 8 in Table 3 that additionally controls for the variable displayed on the vertical axis interacted with the *Post March 30* dummy. Panels a and b report estimates over the 3 day windows while Panels c and d display estimates over the 7 day window. Horizontal bars indicate 90% confidence intervals.

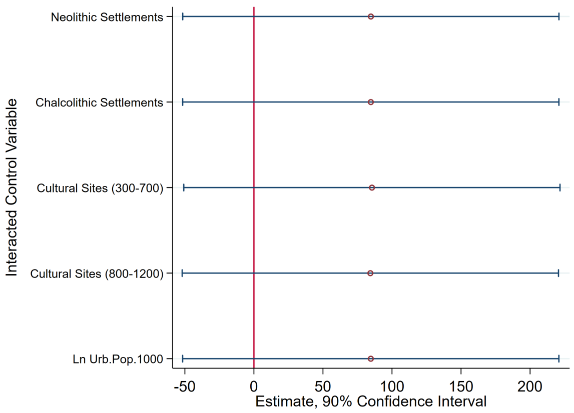
Figure B13: Robustness: Historical Controls Interacted by *Post March 30*



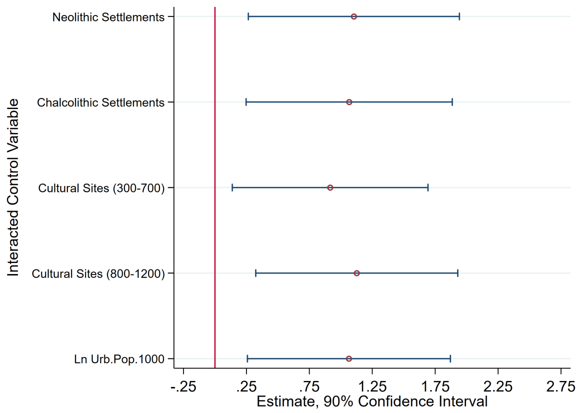
(a) New Delhi × Post, 3 Days



(b) Muslim Attack × Post, 3 Days



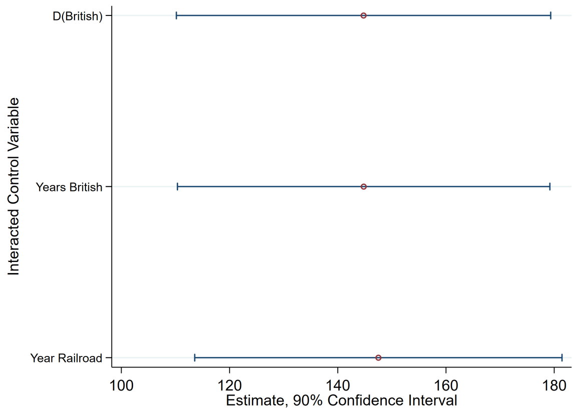
(c) New Delhi × Post, 7 Days



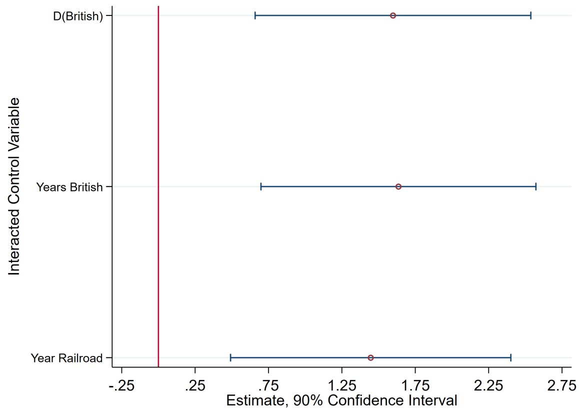
(d) Muslim Attack × Post, 7 Days

Notes: Each dot is the coefficient associated with the variables *New Delhi* × *Post* and *Muslim Attack* × *Post* from a version of specification 8 in Table 3 that additionally controls for the variable displayed on the vertical axis interacted with the *Post March 30* dummy. Panels a and b report estimates over the 3 day windows while Panels c and d display estimates over the 7 day window. Horizontal bars indicate 90% confidence intervals.

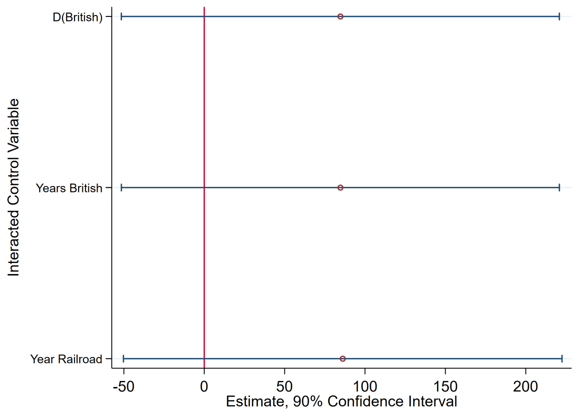
Figure B14: Robustness: Colonization Controls Interacted by *Post March 30*



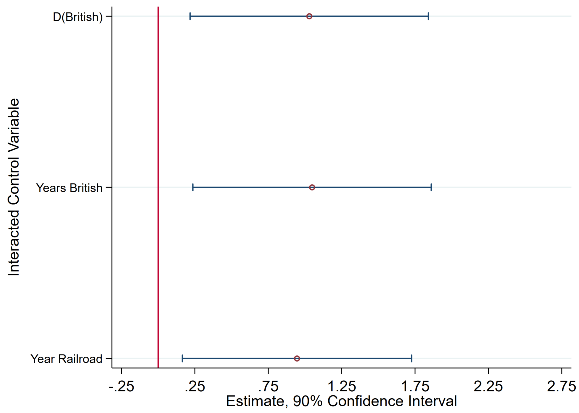
(a) New Delhi × Post, 3 Days



(b) Muslim Attack × Post, 3 Days



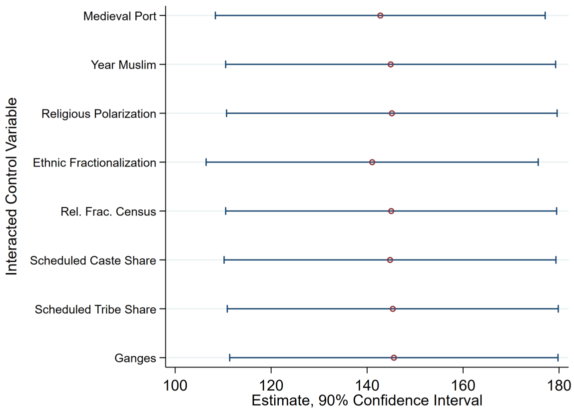
(c) New Delhi × Post, 7 Days



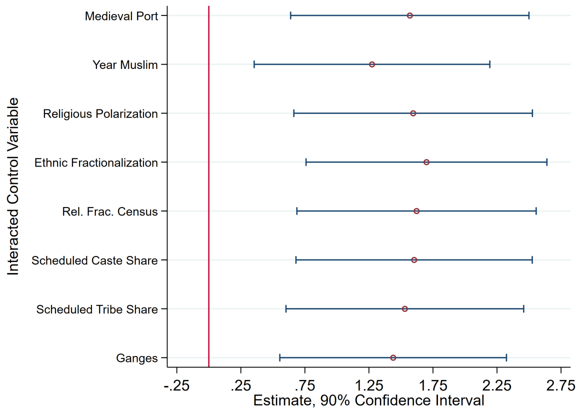
(d) Muslim Attack × Post, 7 Days

Notes: Each dot is the coefficient associated with the variables *New Delhi* × *Post* and *Muslim Attack* × *Post* from a version of specification 8 in Table 3 that additionally controls for the variable displayed on the vertical axis interacted with the *Post March 30* dummy. Panels a and b report estimates over the 3 day windows while Panels c and d display estimates over the 7 day window. Horizontal bars indicate 90% confidence intervals.

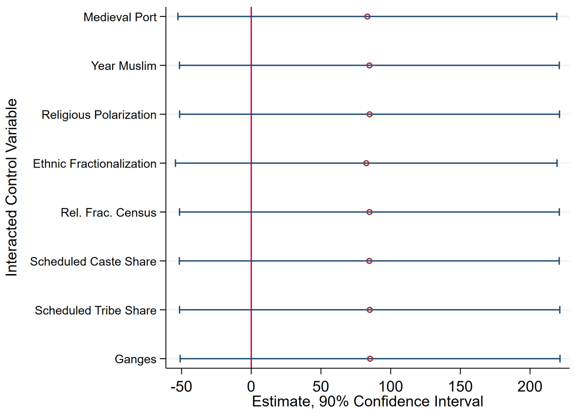
Figure B15: Robustness: Fractionalization Controls Interacted by *Post March 30*



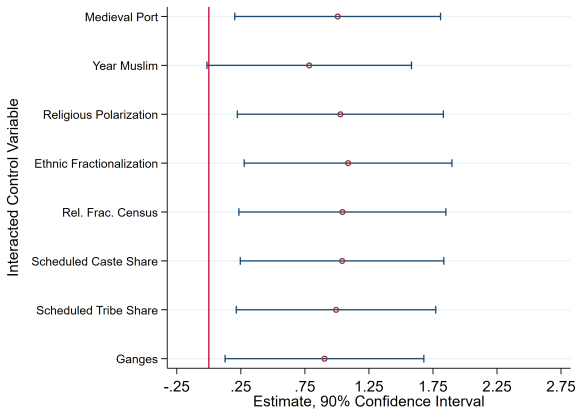
(a) New Delhi $\times$ Post, 3 Days



(b) Muslim Attack $\times$ Post, 3 Days



(c) New Delhi $\times$ Post, 7 Days



(d) Muslim Attack $\times$ Post, 7 Days

Notes: Each dot is the coefficient associated with the variables *New Delhi* $\times$ *Post* and *Muslim Attack* $\times$ *Post* from a version of specification 8 in Table 3 that additionally controls for the variable displayed on the vertical axis interacted with the *Post March 30* dummy. Panels a and b report estimates over the 3 day windows while Panels c and d display estimates over the 7 day window. Horizontal bars indicate 90% confidence intervals.

Alma Mater Studiorum - Università di Bologna
DEPARTMENT OF ECONOMICS

Strada Maggiore 45
40125 Bologna - Italy
Tel. +39 051 2092604
Fax +39 051 2092664
http://www.dse.unibo.it