

Distribution of Google Search Results Across India: Lens of The Farmers' Protest

Arushi Bhandari
Wellesley College
Massachusetts, USA

Manasvi Khanna
Wellesley College
Massachusetts, USA

ABSTRACT

In June 2020, the Central Government of India proposed three temporary agricultural laws, all of which were passed by both the houses of the parliament. These laws incited a response from farmers in India that led to one of the largest protests in the world. Approximately 250 million farmers organized in protest of the laws' suppression of their autonomy. Google, the most popular search engine in India, is one of the most used sources to obtain information about the protests. This paper explores whether the scraped search results of the Google Search Results page (SERP) using the queries related to the Farmers' Protest vary in different cities and states across India. Data was collected from Google for five search queries and across four languages for 733 cities across India. From our analysis using clustering methods, there are indications that Google considers other factors apart from geographic location and does not give similar results in neighboring states. Additionally, from correlation plots, it was found an increase in geographic distance did not necessarily mean a decrease in similarity between SERPs (no significant correlation). Overall, the study showed that Google does not localize results for search queries related to the Farmers' Protest. An implication of this is that there could be information bias on major events because sub-regional perspectives are unaccounted for and this could perpetuate unrepresentative information.

KEYWORDS

Farmers, Protest, India, Google, Text Mining, Spatial Distribution

ACM Reference Format:

Arushi Bhandari and Manasvi Khanna. 2022. Distribution of Google Search Results Across India: Lens of The Farmers' Protest. In . ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

1 INTRODUCTION

Agriculture is central to India's economy and is the primary source of livelihood for around 58% of the population [2]. Any changes to the existing system of agriculture would affect major changes in the lives of billions of people. In the first week of June 2020, the Central Government of India proposed three temporary agricultural laws, which were later passed by both the houses of the

parliament - the Lok Sabha and the Rajya Sabha - by mid-September that year [23]. These three laws would deregulate the existing system of government-run wholesale markets, allowing farmers to sell directly to food processors. While these bills were claimed to be in the interest of the welfare of farmers, both farmers and farmer unions argued that the power would be shifted from independent farmers [21] to private companies, potentially leaving them without the security of government-guaranteed price floors [26]. In fear of being at the mercy of corporations, these laws elicited a response from farmers and merchants in India that led to one of the largest protests the world has ever seen. Over 250 million people from all over the country traveled to the capital city and other state centers to contest the alterations made to farming legislation, market structure, and privatization of food storage [19].

These people, along with the rest of the world, were constantly communicating information regarding the protests using online platforms. The availability of low-cost smartphones and affordable internet has made India one of the newest markets of news consumption online. While instant messaging platforms like WhatsApp and social media like Instagram and Twitter are used immensely, Google is accessed by millions of Indians to obtain information about certain topics and recent events. Google has a monopoly on the search engine market in the country with, as of July 2021, 99% of the population using it as their primary search engine, highlighting its importance in sourcing information [6]. Google is considered to be the most reliable search engine and provides the highest quality information [11]. Google is known to localize results to your preferred region and language for quick access to relevant information [20]. In one study, researchers found contrasting images (tanks and soldiers within cities in the UK and sunshine in China), shown in different locations for the same query "Tienanmen Square" [20]. Therefore, it is likely that Google search results vary according to location within countries, like in India. Additionally, while looking at news about events, the sources themselves might lean either liberal or conservative. Therefore, this study aims to explore the different collections of sources shown when searching about the Farmers' Protests on Google in different parts of India. More specifically, by analyzing Google search results pages (SERPs) obtained using the queries related to the protests, we will answer the following questions:

- (1) How does the distribution of Google search results about the Farmers' Protests and the ongoing movement vary across the major cities of the states of India?
 - (a) Is there a variation in Google search results displayed in regional languages of India?
 - (b) Is there a difference in the SERPs seen in contested regions of India?
- (2) Is there a positive correlation between the geographical distance of states and the SERPs shown in those locations?

Permission to make digital or hard copies of all or part of this work for personal or academic use is granted by ACM, provided that the copies are not made for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
Conference'17, July 2017, Washington, DC, USA
© 2022 Association for Computing Machinery.
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...\$15.00
<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

RQ 1 aims to reveal the similarity in Google search results about the Farmer's Protest across India. We used specific queries and languages that target localized results, for specific cities, as well as those that target similar results nationally. Through this, we hope to identify whether Google's algorithm delivers location-specific results to users, especially with regards to major events. An implication of this would be that different types of information could be shared in different regions of India. Moreover, this study also considers whether Google categorizes SERPs of contested regions differently from the rest of India.

Given the existing literature on the Farmers' Protest and Google's page ranking algorithm, we formed three hypotheses to answer the first question. The first hypothesis is that locally targeted queries will result in lesser similarity between SERPs. We also hypothesize that these results would be more localized with less similarity when the searches are conducted in regional languages. Lastly, we predict that Google will display the most different results for contested regions compared to other places.

RQ 2 aims to identify if Google localizes search results based on the geographical location of the place. This will help us understand whether the similarity in SERPs for two places is because of their geographical proximity or whether Google is considering other features as well while displaying search results in different locations. Since Google's page ranking algorithm accounts for geolocation, we predict a positive association between geographic proximity and page similarity [20].

2 LITERATURE REVIEW

2.1 The Farmers' Protest

While there is not much research done regarding the Farmers' Protests and its ongoing movement, existing literature mainly comprises news articles, published political journals and some statistical analysis. We discovered three emerging themes in the existing literature including the varying narratives surrounding the nature of the protests, the actors involved, and the portrayal of the protests in mainstream media.

2.1.1 The Nature of the Protest. The first theme seen in the literature on the farmer's protests discusses the turbulent nature of the protests across India and its discourse on several types of media [26]. There is recurring discussion about whether the use of water cannons, a military weapon operated by the police during the protests [8] is justified in its efforts to stop protesters from reaching India's capital, New Delhi. Additionally, there is similar discourse about the use of tear gas in Haryana to stop people from organizing [8]. Moreover, there is documentation about videotape and film evidence shared on social media of police violence in Karnal when farmers had blocked a highway [25]. There were also court files submitted by the Judiciary about the police using batons on protesters ("lathi charge") and injuring more than 10 people [14]. This section of the literature emphasizes the turbulent nature of the protests, which is often characterized by massive, intimidating police presence (or police violence). It also considers whether this turbulence is connected to actors beyond farmers and police and what is at stake for each of them.

2.1.2 Actors and People Involved. Consequently, the second recurring theme in the literature identifies the major actors involved in the movement and their motivations. Approximately 200 labor unions have been recognized as active participants of the Farmers' Protests [22]. These unions range from independent trade unions, trade unions with political standing, farmer unions, independent farmers, and religious groups [18]. The unions were protesting not only the agricultural laws but also three labor codes that were introduced at the same time, in 2020, by the government [22]. These labor codes were deemed anti-labor and allowed companies greater freedom in termination and exit decisions for laborers [26]. Therefore, demands of the movement include the abolishment of the newly introduced laws, compensation for lives lost due to the government's actions, and the waivment of loans and interest payments on them because of severe crop failure ([26]; [22]; [18]). Due to these demands, a number of media outlets have labeled the unions as defenders of a fiscally and agro-ecologically harmful regime [13]. This has led to debate about whether farmers are the sole actors in the movement or whether it is majorly influenced by trade unions and political groups as well. Lastly, oftentimes, there is also disagreement about the movement being labeled as a defensive, regional mobilization rather than a progressive, national movement [13].

2.1.3 The Protests and Media. The presence of conflicting stories and narratives on mainstream media motivated a third theme of research which aimed to identify influential actors and explore their sentiments about the farmer's protests. One such research study analyzes how mainstream media uses Twitter as a tool to share news about the protest. These mainstream media sources were generally broadcasted on a national scale and found that titles with government actors have a higher proportion of opposing sentiments as compared to other actors [15]. Moreover, they identified an association between the source, the actor and their role in the protests. The two news sources that tweeted the most, NDTV and Zee News (TV news), were known for their specific narratives about the protests [15]. NDTV focused on farmer actions and opinions whereas Zee News mainly tweeted government and celebrity opinions [15]. Another study conducted hypothesis tests and prediction models on 17000 tweets about the protests to understand sentiment behind them [17]. It found that most tweets published were of neutral sentiment, followed by that of positive sentiment. A limitation of this study is that it does not report the distribution of stance (for or against) regarding the Farmers' Protests. A third study analyzed whether there is an association between the stance taken in tweets and retweets published by celebrities, politicians etc., and their subsequent engagement on Twitter [9]. They discovered that more followed celebrities were less likely to engage with tweets in support of the farmers protests. It was also found that, irrespective of stance, influencers who engaged in Twitter discourse saw a significant increase in following. Overall, the literature around the protests has begun exploring the influence of media and the content being shared by different people. However, there are significant gaps regarding the dissemination of online news, specially on Google, and whether there is variation in this dissemination at the national and regional levels.

2.2 The Role of Google

One specific gap in the existing research is the lack of analysis of how people source information about the Farmers' Protest. Google, a major avenue of obtaining information, especially regarding the protests, has not been explored.

2.2.1 Google Page Ranking. Google's Page Ranking algorithm is a complex system that relies on a series of features such as words used in a query with context consideration, the relevance and usability of pages, the expertise of sources, and your location and settings [4]. Google conducts several search quality tests to ensure that search results optimize the relevance and usability of information shared.

2.2.2 Google and Social Impact. Google is often an important source of data for researchers to study protests and movements globally. Various projects focusing on social impact, and protests use data from the Google search engine to investigate source bias. One study uses the news search function on Google to extract data that is popularly consumed by people to understand police brutality as it is reported [24]. Another study compares the search results obtained on Google and on similar search engines in Russia regarding a series of protests that occurred in various Russian cities [12]. They categorized the search results into three categories - leadership critical websites, politically critical websites and uncritical websites - and compared the frequency of occurrence of these three categories using hypothesis testing [12].

2.2.3 Google SERPs and Location. A review of existing literature revealed three major studies that incorporated location and other demographic variables to investigate the distribution of SERPs for different search queries. One study aimed to analyze information on different vaccines in different countries using the localized version of Google available (google.uk, google.au, google.sj, etc.) [5]. Their methods included statistical analysis models such as the Bonferroni correction for multiplicity, two-tailed Fisher's exact test, natural language processing models and the Pearson Correlation Coefficient test. Findings indicated variance in website ranking distribution across different locations and domain. They also visited each website using the URL extraction and determined the distribution of stance and sentiment about vaccines. The paper also determined different actors such as celebrities, government officials, scientists, etc. that were involved in the dissemination of information and general discourse.

Another study explores the visibility of Ibero-American governmental, educational and research institutions in Google SERPs using queries related to areas of Science, Research and Innovation [7]. The study used descriptive statistical analysis to cluster their data into frequency tables. The study determined the ownership of domains comparing the frequency tables and found that on average countries rank their own university websites higher than others. Their results show an impact of localization on their results (such as ranking and proximity to the university). A third study considers the access to information about abortion clinics on Google and whether it varies according to location. The authors illustrated that, while Google Search returns location results to actual abortion clinics in most cases, the number of results is unevenly distributed across locales and the quality of results could be improved. [16]. The above studies showed that location has an effect on the search

results displayed on Google and could have serious implications. We wanted to investigate whether this was true for cities within India, especially for the dissemination of information regarding major events.

3 DATA & METHODOLOGY

3.1 Selection of Queries and Languages

To explore whether Google is considering geographic location while displaying search results related to the Farmers' Protest, we selected 5 queries to conduct our analysis on. The first query was "farmers protest" to give a general idea of what was being displayed regarding the event. Next three queries were directly related to the Farmers' Protest. Considering our background research suggested that the protests were of a somewhat violent nature, one of the queries used was "farmers protest violence." Additionally, we also found that actors like the government and farm unions were very involved in the protest so we included queries like "farmers protests unions" and "farmers protest agriculture laws." Finally, we included a general query that was not specifically related to the event, "news," to serve as a control variable against trends seen in other queries.

These 5 queries were used for searches in 4 languages, including English to account for the linguistic diversity of India. We conducted searches in 3 commonly spoken languages from different parts of the country. These languages include Hindi, the most spoken language in the country (especially in the north), Bangla (Bengali), the most common language in the east, and Telugu, the most popular language in the South. By conducting searches in these languages, we hoped to have a more representative understanding of searches related to the protests.

3.2 Data Collection

For data collection from Google, we used Selenium, an automatic web browsing tool, and Chrome Driver, an automatic tool to control Chrome. We scraped data from Google search result pages (SERPs) that were displayed on searching the 5 queries described above in the 4 languages in different cities in India. Three pages of results were scraped for each city by setting the location using latitude and longitude from a database of 794 Indian cities and their coordinates available on latLong.net [3]. Some of the locations provided were not of cities but of areas and localities within cities, for example, "Goregaon, Mumbai, Maharashtra, India," which were removed from the dataset. Additionally, some of the cities did not have information about the state / union territory it was located in, therefore, the names of the state / union territory were manually inserted. An occurrence of this was when the city of "Savner, India" was replaced with "Savner, Maharashtra, India." On removing duplicates, the dataset consisted of 733 cities, and this was used to set locations to obtain HTMLs of the first three pages of Google search results obtained using each of the queries. After obtaining the HTMLs, the Python package, BeautifulSoup, was used to collect information, like domain, headline etc., about each result listed on the Google SERPs for a sample of cities (Fig. 1). This dataset also includes contested regions such as cities in Kashmir.

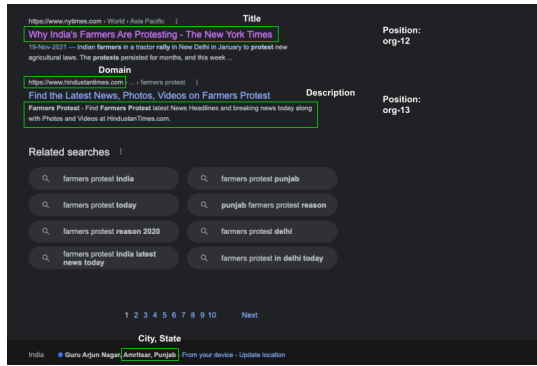


Figure 1: Sample Google SERP for query "farmers protest" at "Amristar, Punjab"

3.3 Data Summary

We conducted analysis on 20 different datasets containing information about search results obtained from 5 queries for 4 languages across 733 cities in India. Description of each dataset is provided in Fig. 2. In each dataset, a unit is one search result seen on the Google SERP. It contains information about the state, city it is extracted from, the domain of its source, the position it is displayed on, the title and the description of the result that is available on the SERP.

Language	Query	# Results	Top 2 Results
English	farmers protest	19541	https://en.wikipedia.org
	farmers protest violence	18831	https://scroll.in
	farmers protest unions	19489	www.cnn.com
	farmers protest agriculture laws	18410	www.new18.com
	news	17323	https://en.wikipedia.org
Hindi	farmers protest	15606	www.downtoearth.org.in
	farmers protest violence	19253	www.bbc.com
	farmers protest unions	17579	www.hindustantime.com
	farmers protest agriculture laws	19935	https://news.google.co.in
	news	18762	https://www.ndtv.com
Telugu	farmers protest	18136	https://navbharattimes.indiatimes.com
	farmers protest violence	16971	https://www.livehindustan.com
	farmers protest unions	18529	https://www.bbc.com
	farmers protest agriculture laws	4781	https://www.abplive.com
	news	18526	https://www.zenews.india.com
Bangla	farmers protest	17490	https://zeenews.india.com
	farmers protest violence	20876	https://www.dishitias.com
	farmers protest unions	18963	https://www.amarujala.com
	farmers protest agriculture laws	19335	https://www.ndtv.com
	news	18997	https://telugu.abplive.com

Figure 2: Description of different datasets

3.4 Methods

The first research question considers how the distribution of sources on Google search result pages (SERPs) varies across the major cities of the states of India. Each of the SERPs of 3 individual pages consist of a number of sources arranged in a particular order. To determine how similar the SERPs seen in these major cities were, we had to convert each SERP into a vector for further analysis. The vector was created by assigning a metric for each domain that occurred on the page. The metric considered the occurrence of the domain, the page on which the domain occurred and the position it occurred in on that page. The smaller the value of the metric is, the less priority the domain was given by Google. The metric was calculated as follows:

$$\text{metric} = \text{weight} * \% \text{ of occurrence}$$

where,

$$\text{weight} = \text{number of results} * 1/\text{domain} * 1/\text{pagenumber}$$

$$\% \text{ of occurrence} = (\text{number of occurrences in that position} / \text{total occurrences}) * 100$$

Next, this metric was used to conduct unsupervised learning with the aim of grouping similar SERPs together, suggesting that major cities of specific states have similar SERPs. We initially used k-means clustering to determine this. K-means clustering will divide data in k clusters, each with its own center. The algorithm works iteratively to group together data points that are spatially closer to one-another. In this method, we tell the algorithm how many clusters we want the points to be grouped into and this number is found using squared error for cluster points (elbow plot). The data points are grouped together based on distance from one another. Since this method is limited by a specific number of groups and does not provide much information about how pages are grouped together, we also conducted hierarchical clustering. This method was also conducted using the same metric. In order to compare dendrograms outputted through hierarchical clustering for the 5 queries in 4 different languages, we used a method introduced by B. Fowlkes & C. L. Mallows [10]. We chose an arbitrary distance (k-value) of 1.5 to obtain the number of clusters produced in each dendrogram in order to compare localization of Google search results. Answering RQ 1b: To confirm the placement of all cities within Jammu and Kashmir as separate from other cities, we compared the subset of 5 cities with cities from neighboring states (Himachal Pradesh, Uttarakhand) and one randomly selected state (Chattisgarh). Hierarchical clustering was conducted for queries "farmers protest" and "news" in all languages.

The second research question aims to check if there is a positive correlation between the geographical distance of specific locations and the SERPs shown in those locations. This would suggest that the closer two pages are geographically, the more similar the SERPs are. To answer this, we check for correlation between geographical distance and distance between SERPs of different states. The geographical distance is found using the Google Distance Matrix API. The similarity between SERPs is calculated using Euclidean distance between vectors of metric calculated for the domains present on each SERP for each state.

4 ANALYSIS & RESULTS

Our first hypothesis was that locally targeted queries would result in less similarity between SERPs than nationally-targeted queries. More specifically, we predicted that the query, “news,” would result in the least similar results between SERPs when compared to the queries “farmers protest,” “farmers protest violence,” “farmers protest agricultural laws.” For a particular query, the greater the number of clusters generated, the less similar SERPs seen at different locations are to each other. Existing literature emphasizes that the agricultural laws relevant to the protests affected the entire country because they were set in motion by the current Central Government [22]. Various publications then claim the movement to be a national protest, indicating a form of collective agitation from the public [26]. Therefore, we predict that results would be less similar for the queries regarding agricultural laws and violence than the query “news.” However, the magnitude of interest in the laws by the public varies by location [22] and informs our hypothesis that “farmers protest agricultural laws” would have more clusters than “farmers protest” and “farmers protest violence.” On the other hand, for “farmers protest unions,” we expected the most localized SERPs due to the involvement of over 200 unions from different regions in India [18].

Results, for English SERPs, from the hierarchical clustering analysis showed a larger number of clusters in the “news” query (9) than for the more general “farmers protest” query (7). Similarly, the query “farmers protest violence” (5) (Fig. 3) and “farmers protest agricultural laws” (6) showed less clustering than the control. Lastly, “farmers protest unions” generated the most clusters (10) (Fig. 4). These results are consistent with our hypotheses and inform our understanding for the first research question.

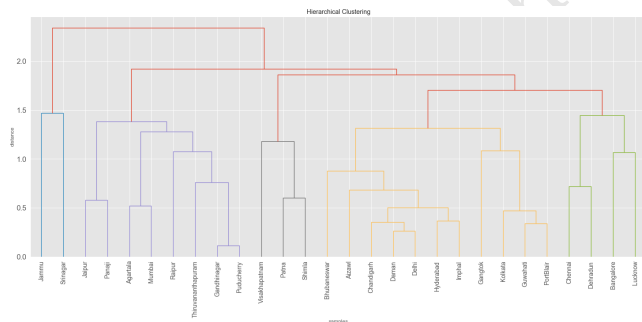


Figure 3: Hierarchical Clustering of English SERPs with query ‘farmers protest violence’

Our hypothesis regarding languages was that regional languages will result in more localized results and lesser similarity. This means we expect to see more clusters while using Hindi, Telugu, and Bangla queries to conduct hierarchical clustering than while conducting the same for English queries. Results showed an average of 5 clusters for Hindi queries, 7 for Telugu, 9 for Bangla, and 7 for English. Interestingly, the hypothesis was supported in the case of Bangla and Telugu but not in that of Hindi. This could be explained by the fact that despite being a regional language, Hindi is the most spoken language in India and is spoken across many regions within the country [1].

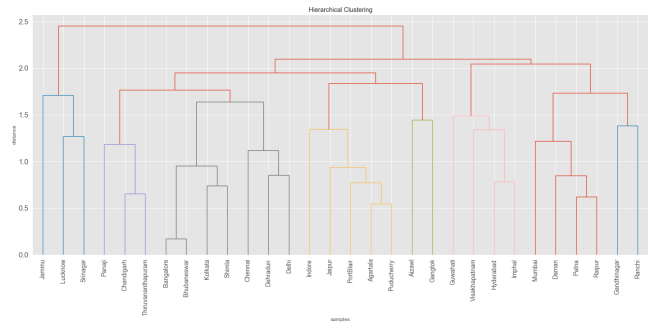


Figure 4: Hierarchical Clustering of English SERPs with query ‘farmers protest unions’

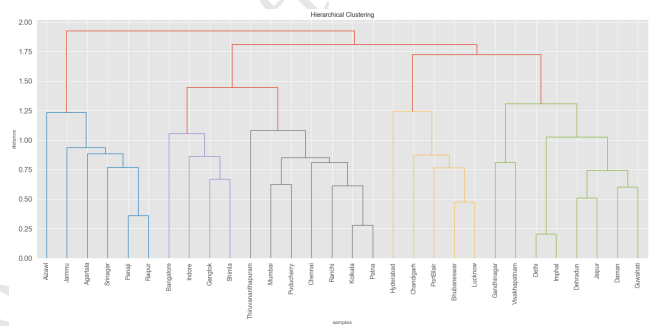


Figure 5: Hierarchical Clustering of Hindi SERPs with query ‘news’

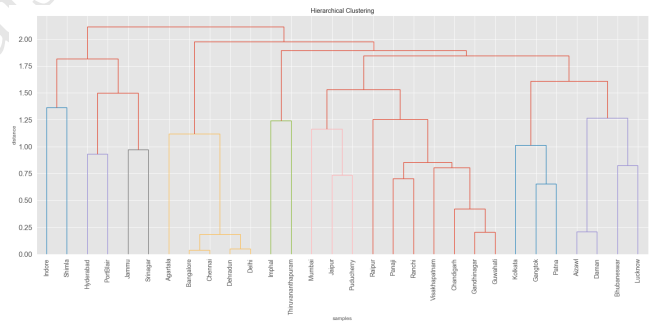


Figure 6: Hierarchical Clustering of Telugu SERPs with query ‘farmers protest violence’

Our third hypothesis states that Google will display the most different results for contested regions compared to other places. First, we observed the placement of Jammu and Srinagar (a city in Kashmir) in the above hierarchical clusters generated for all queries. Results showed that in all English queries, Jammu or Srinagar were individually placed in a separate cluster. In 3 of the English queries (out of 5) they were placed as a cluster together separate from other cities. In Hindi queries, Jammu was individually placed in a separate cluster twice and placed in a separate cluster with Srinagar twice. In Telugu, Jammu and Kashmir were placed together in a cluster separate from other cities for all queries, except for the query

“farmers protest unions.” Lastly, for Bangla Jammu and Srinagar were placed in a separate cluster from other cities three times and Jammu was individually placed in a separate cluster twice. These findings are consistent with our hypothesis above. Interestingly, like in Fig. 7, Jammu, which is not in the contested part of the state, was placed out of other clusters more often than Srinagar (or any other city). While comparing the placement of the cities in

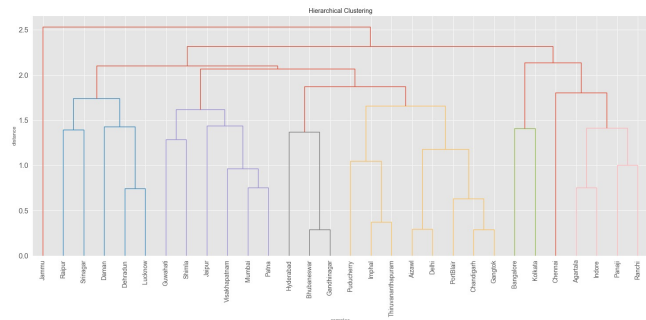


Figure 7: Hierarchical Clustering of Bangla SERPs with query 'farmers protest unions' where Jammu is in a separate cluster

the region of Jammu and Kashmir with cities of neighboring and distant states, we saw distinct clusters for the cities in Jammu and Kashmir (separate from other cities in the sample) across all queries, with the exception of “news” in Hindi (Fig. 8). The cities in other states were not as clearly clustered together by state. Therefore, this was consistent with our third hypothesis: Google displayed the most different results for all cities within Jammu and Kashmir compared to other places.

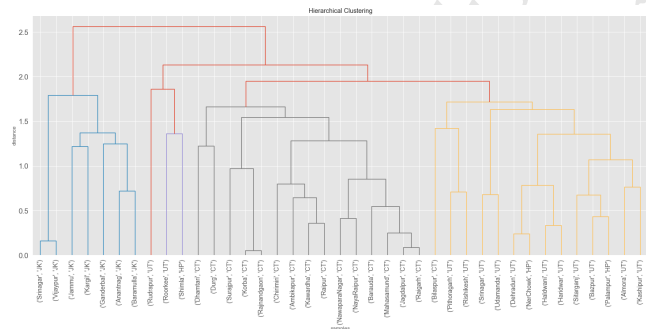


Figure 8: Hierarchical Clustering of Bangla SERPs with query 'farmers protest' for four states in India

The second research question in this study tested whether there was a positive correlation between the geographical distance of states and the SERPs shown in those states. We calculated the correlation for all states among all languages for queries: “farmers protest” and “news.” We hypothesize that all queries would show a slight positive correlation between the geographical distance and the measure of similarity between SERPs. We predict a locally-targeted query like “news” would have a higher correlation than the query “farmers protest” across all languages. Overall, from the

correlation plots, we can see that there was insignificant correlation between the geographical distance of states and the similarity between the SERPs shown in cities from those states. The “news” query consistently showed a slightly higher positive correlation over “farmers protest,” across all languages. This finding validates our decision to use the query “news” as a control because it generates locally relevant search results. These results are also consistent with the results from hierarchical clustering where we did not see any patterns of similarity of SERPs due to location, apart from the contested region of Jammu and Kashmir.

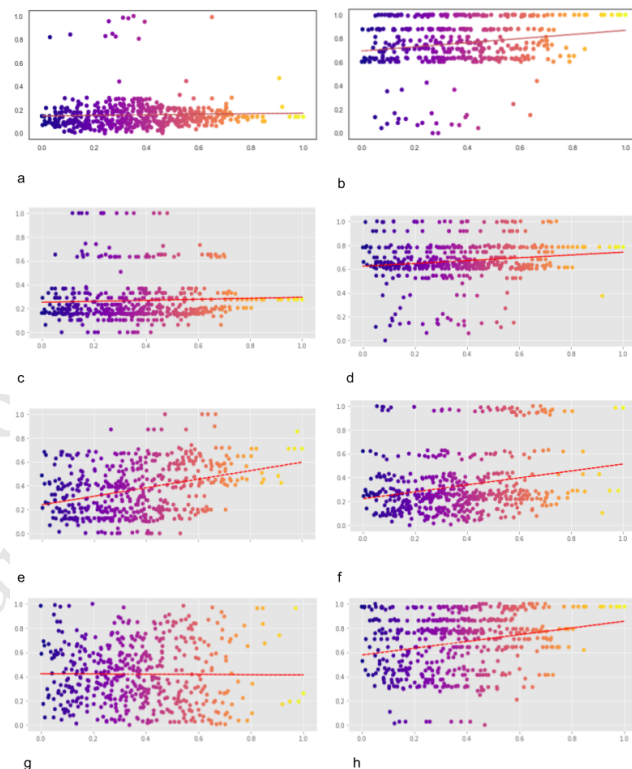


Figure 9: Correlation plots between geographic distance and page distance for queries - 'farmers protest' and 'news' for each of the four languages - English, Hindi, Telugu, Bangla

5 DISCUSSION

5.1 Implications

Overall, the study showed that Google does not localize results for search queries related to the Farmers' Protest. This could lead to information bias on major events because sub-regional perspectives are unaccounted for and this could perpetuate unrepresentative information. We observed extreme concentration in the sources reported across all queries and languages. More specifically, Google only showed results from a maximum of 70 sources across 733 cities for a particular query. This is an important implication for users to increase awareness because Google is prioritizing the same sources for all users regardless of location and query. It is possible

that Google only trusts a few sources in India (mostly international/national) and prioritizes their content in its search results.

The findings of this study suggest Google generates more localized search results for targeted queries and regional languages. It is possible that depending on the default language set on a user's Google search has an impact on the information they receive without their knowledge. When conducting our analysis, we found that contested regions like Jammu & Kashmir have considerably different pages to other states in India regarding the Farmers' Protest. This finding can be used by policy-makers either to alienate those regions, target biased information to vulnerable groups, and influence the public according to their advantage.

5.2 Ethical Considerations

This research does not include identifications or results from any individual user, ensuring user privacy. During the data collection process, no search results were visited ensuring that the research does not influence Google's page ranking algorithm. That said, since we searched a large amount of the same targeted queries in a short time frame, it is possible that Google would be influenced to display more localized information. Moreover, the use of an automatic browsing tool could be in violation of Google's Terms of Service. If flagged, it is possible that other researchers hoping to use these tools might be limited in doing so. It is also possible that actors with political or monetary incentives could use the results obtained from this research to target certain information to contested regions.

5.3 Limitations

The data collection for this project was conducted between April, 2022, and May, 2022, which is over a year after the peak of the Farmers' Protests. Therefore, collected data may not be representative of search results observed during the protests when these queries were most trending on Google. Since data collection was conducted using latitude and longitude information from a public database of Indian cities, our data was not randomly distributed and could be biased towards central, more populous cities. There are many more cities from larger states like Maharashtra than from smaller states like Goa. States that are often not represented in research about India like the North-Eastern states are also not adequately represented in our dataset with either only one or no cities from each of them. Considering that the biggest group represented in the protests is farmers and agricultural workers, it is likely that they are not all located in these central cities of India and are not using Google from those locations. During data collection, there were several locations that failed to update location for the search pages. To ensure an equally representative dataset, we had to collect SERPs for those locations in several iterations of the data collection process. This led to a large time gap between SERP collection for some locations and has the potential to influence our analysis and results. Lastly, one query (farmers protest agricultural laws) in Telugu only collected results for one page (instead of three) and potentially altered results seen for those SERPs.

6 CONCLUSION

From this analysis, we can conclude that Google does not localize results for search queries related to the Farmers' Protest in India

and that an increase in geographic distance does not imply a decrease in similarity of SERPs. For regional languages, except Hindi, there is more localization observed than in English. Additionally, we found that, in general, Google does not display a wide range of sources in the SERPs in India. However, it is attentive to not group a contested region with other surrounding states in India. Future research should aim to validate these findings for other queries and explore the why domains are concentrated in India. It would also be essential for future research to explore the relevance of other location-specific features in Google Search such as agriculture similarity in states, population demographics and expertise of sources.

REFERENCES

- [1] 2018. Most common languages spoken in India in 2011. *Statista* (2018). <https://www.statista.com/statistics/616508/most-common-languages-india/>
- [2] 2021. AGRICULTURE AND ALLIED INDUSTRIES INDUSTRY REPORT. (Nov 2021). <https://www.ibef.org/industry/agriculture-india>
- [3] 2022. Cities in India. *LatLong* (2022). <https://www.latlong.net/category/cities-102-15.html/>
- [4] 2022. How Search algorithms work. *Google* (2022). <https://www.google.com/search/howsearchworks/algorithms/>
- [5] Nadia Arif, Al-Jafri Majed, Isabella Harb Bizzi, and Gianni Boitano Perano. 2020. Fake News or Weak Science? Visibility and Characterization of Antivaccine Webpages Returned by Google in Different Languages and Countries. *Frontiers in Immunology* 123 (Jun 2020). <https://doi.org/10.3389/fimmu.2018.01215>
- [6] Tanushree Basuroy. 2021. Share of Google in mobile search market India 2019-2021. *Statista* (Aug 2021). <https://www.statista.com/statistics/938846/india-google-share-in-mobile-search-market/#:~:text=Google%20had%20a%20clear%20monopoly,engine%20operator%20in%20the%20country>
- [7] Simone Belli and Carlos Gonzalo-Penela. 2018. Science, research, and innovation infospheres in Google results of the Ibero-American countries. *Scientometrics* 9 (March 2018). <https://doi.org/10.1007/s11192-020-03399-4>
- [8] Mansee Bal Bhargava. 2021. Water cannons and canonisation of the governance of protest. *Newstex* (Aug 2021). <https://www.proquest.com/blogs-podcasts-websites/water-cannons-canonisation-governance-protest/docview/2558931679/se-2?accountid=14953>
- [9] Arshia Arya Saloni Dash Rynaa Grover Joyojeet Pal Dibyendu Mishra, Syeda Zainab Akbar. 2021. Rihanna versus Bollywood: Twitter Influencers and the Indian Farmers' Protest. (2021). <https://doi.org/10.48550/ARXIV.2102.04031>
- [10] E.B Fowlkes and C.L. Mallows. 1983. A Method for Comparing Two Hierarchical Clusterings. *Journal of American Statistical Association* 78 (1983), 553–569. <https://doi.org/10.1080/01621459.1983.10478008>
- [11] Anna C. Fullard, Sean M Johnston, and Dermot J. Hehir. 2021. Quality and reliability evaluation of current Internet information regarding mesh use in inguinal hernia surgery using HONcode and the DISCERN instrument. *Hernia* 25 (2021), 1325–1330. <https://doi.org/10.1007/s10029-021-02406-8>
- [12] Daria Kravets and F Toepfl. 2021. Gauging reference and source bias over time: how Russia's partially state-controlled search engine Yandex mediated an anti-regime protest event. *Information, Communication & Society* 17 (Jun 2021), 1–17. <https://doi.org/10.1080/1369118X.2021.19335637>
- [13] Mekhala Krishnamurthy. 2021. Agricultural market law, regulation and resistance: a reflection on India's new 'farm laws' and farmers' protests. *The Journal of Peasant Studies* 48, 7 (2021), 1409–1418. <https://doi.org/10.1080/03066150.2021.1998001>
- [14] K. Alan Kronstadt. 2021. Farmer Protests in India. *CRS Report* (March 2021). <https://crsreports.congress.gov/product/pdf/R/R46713/3>
- [15] Jenith Sekar Kuldeep M Siwach. 2021. MAINSTREAM MEDIA AND TWITTER NARRATIVES: A CASE STUDY OF THE FARMERS' PROTEST. *European Journal of Molecular Clinical Medicine* 7, 8 (Mar 2021), 5551–5561. https://ejmcm.com/article_9437.html
- [16] Yelena Mejova, Tatiana Gracyk, and Ronald E. Robertson. 2022. Googling for Abortion: Search Engine Mediation of Abortion Accessibility in the United States. *Journal of Quantitative Description: Digital Media* 2 (2022), 1–41. <https://doi.org/10.51685/jqd.2022.007>
- [17] Ashwin Sanjay Neogi, Kirti Anilkumar Garg, Ram Krishn Mishra, and Yogesh K Dwivedi. 2021. Sentiment analysis and classification of Indian farmers' protest using twitter data. *International Journal of Information Management Data Insights* 1, 2 (2021). <https://doi.org/10.1016/j.jjime.2021.100019>
- [18] Kiran Singh Preethi. 2021. My Crop My Right: The Farmer's Act 2020. *LAHRW International Journal of Social Sciences Review* 9, 3 (Sep 2021), 197–198. https://www.proquest-com.ezproxy.wellesley.edu/publication/publications_5347679

- [19] Mandeep Rai. 2021. Opinion: The World's Largest Protests You've Probably Never Heard Of. *cnn* 4 (Mar 2021). www.cnn.com/2021/03/06/opinions/india-farmer-protests-mandeep-rai-dhillon/index.html
- [20] Tom Simonite. 2021. A new tool shows how google results vary around the world. *Wired* (Jul 2021). <https://www.wired.com/story/tool-shows-google-results-vary-world>
- [21] Navsharan Singh. 2021. Agrarian crisis and the longest farmers' protest in Indian history. *New Labor Forum* 30, 3 (Sept 2021), 66–75. <https://doi.org/10.1177/10957960211036016>
- [22] Navsharan Singh. 2021. Agrarian Crisis and the Longest Farmers' Protest in Indian History. *New Labor Forum* 30, 3 (Sep 2021), 66–75. <https://doi.org/doi:10.1177/10957960211036016>
- [23] Sukhpal Singh and Shruti Bhogal. 2014. Depeasantization in Punjab: Status of farmers who left farming. *Current science* 106 (05 2014), 1364–1368.
- [24] Philip Matthew Stinson, Chloe Ann Wentzlof, John Liederbach, and Steven L Brewer. 2021. Racial Disparities in Police Crime Victimization. *Social SciencesM* 10 (Jan 2021), 287. <https://doi.org/10.3390/socsci10080287>
- [25] Colin Todhunter. 2021. Smashing The Heads of Farmers: A Global Struggle Against Tyranny. *Wired* (Sep 2021). <https://www.proquest.com/blogs-podcasts-websites/smashing-heads-farmers-global-struggle-against/docview/2572249738/se-2?accountid=14953>
- [26] Harsha Walia. 2021. The Indian farmers' protest is a window on a new world. *Briarpatch* 50, 4 (Jul 2021), 44. <https://www.proquest.com/magazines/indian-farmers-protest-is-window-on-new-world/docview/2547067730/se-2?accountid=14953> Copyright - Copyright Briarpatch Inc. Jul/Aug 2021; Last updated - 2021-07-01.