An Empirical Analysis of Power and Influence in Social Movements

The case of Farmers Protests in India

Thesis submitted to the Department of Methodology, LSE for completion of the degree MSc in Applied Social Data Science

Candidate Number: 40163

Supervisor: Dr Eleanor Power

Word Count: 9983

Table of Contents

Acknowledgments

I would like to thank all my professors at the Department of Methodology, in particular my supervisor Dr Eleanor for all their consistent guidance over the course of the degree. This dissertation is a product of many who constantly inspire me to be a better version of myself and have always provided unconditional support - my family, my partner Susan, and all my friends.

Abstract

Social movements are one of biggest drivers of democracy, social justice, and equality in nations across the world. It is a key instrument to bring about systematic revolutionary changes in regimes and create a sense of conscience within communities. The advent of social media has exponentially increased the avenues available to people to express grievances related to events, policies, and norms. Using Twitter data, this paper attempts to evaluate the in-groups within a framework of influence using the case of Farmers Protests in India. Network Analysis is used to create hierarchical groups of protests based on increasing connectivity, whereas Text Analysis and Vector Autoregression is used to adjudicate which groups are more likely to lead or follow frames of Farmers Protests on Twitter. The results indicate that while users in the denser shells tend to be more active (in terms of frequency of tweets), the peripheral groups also generate high levels of activity, on an aggregate level. Further, both groups can lead discussions in different topics. Using novel identification and estimation techniques, this research attempts to contribute to larger field of communication patterns of social protests on social media platforms, and the dynamics between influence of actors during information exchange.

Background

The growth of social media platforms has revolutionized the accessibility and exposure of social movements. Due to this, there is considerable interest in research on protest related communication patterns on social media (Scherman et al., 2014; Poell, 2013; Bastos et al., 2015). This research indicates that much of the social world is incredibly complex, where people behave with varying intentions. However, due to absence of large-scale data, they often fundamentally lack the representative sample size to make behavioral inferences- as it is quite hard to track protestors and for them to accurately remember their actions in protests (Larson et al., 2019). The advent of social media has thus opened the possibility for sociologists to leverage this rich source of data on social movements.[1]

To be clear, the aim of this study is not to assess the impact of social media platforms in facilitating social movements. This study is fundamentally interested in analyzing the structure of social movements, which is often assumed to be hierarchical in nature (Bennett & Segerberg, 2012) while also viewing it in the context of social media platforms. For this, the case of Farmers Protests in India is used.[2]

On 20th September 2020, the Indian parliament passed three Farm Bills, merely 3 days after they were introduced in the Lok Sabha (Lower House of Indian parliament). These three laws were aimed at bringing monumental changes to the agricultural ecosystem in India through extensive privatization. These laws were opposed by farmers on multiple grounds. Firstly, farmers argued that this allows the big corporations to hoard supplies of raw materials and manipulate prices. Secondly, they argued that this will remove existing safety nets from farmers in the form of state-run markets, also creating a downward pressure on the minimum support price (MSP) that the government guarantees to farmers. The third line of reasoning concerned the power asymmetry, i.e., in the cases where corporations manipulate farmers, they would be unable to fight back legally due to strong political, social, and financial capital of corporations.

People across the country stood in solidarity with farmers. Soon enough, these protests were described as one of the largest nationwide uprisings in history. Despite the threat of Covid-19, protestors (mainly farmers) moved to the streets, established long lasting sit-ins, blocked roads

[1] Borrowed from MY400 Assignment 2
[2] Borrowed from MY400 Assignment 2

etc. However, those who could not go to the on-ground protests moved to social media for expressing their grievances and extending support. This study utilizes the unique context of Farmers Protests to better understand how social media facilitates communication on social movements.

Literature Review

Social Movements as Social Networks

In literature related to social movements, there is no "formal definition" or "criteria" for a movement to finally turn into a protest. As described in Blumer (1971), a commonality between movements is a shared sentiment of resistance amongst its participants. This resistance can be specific, with a goal, like the Farmers Protests in India, or general, as an expression of resistance against larger social norms like the second/third waves of feminism.[3]

The characterization of social movements as social networks has been extremely popular in sociological literature concerned with behavioral objectives (Diani & McAdam, 2003; Polletta & Jasper, 2001). This is because, at the most fundamental level, social movements are an "aggregation" of individual actors' decisions and communication (Tilly, 1978; Schelling, 2006). This relational characterization of social movements is useful as it lays the foundation to model a rational individual's decisions in the context of interdependence. It is thus useful to answer a very fundamental question in the social sciences – 'why do people protest?' (Olson, 1965; Coleman, 1990).[4]

As famously quoted by Marx, a physical concentration of grievances is a necessary step for mobilization. In these studies, protestors are often linked together using various metrics measuring interactions. Then, the type and number of ties are often used as an explanatory mechanism for participation and recruitment.

Most of the work in this field focuses on identification of influential actors through a relational framework. For example, the actors (or organizations) that tend to receive more ties (in-degree)

[3] Borrowed from MY400 Assignment 2
[4] Borrowed from MY400 Assignment 2

have higher prestige in social networks and thus greater influence (Freeman, 1978). The number of connections is also used as proxy for commitment (Coleman, 1988; Gould, 1991). Influence can also be understood from the placement of actors in the network. For example, if they have indirect ties with influential organizations, irrespective of popularity. These actors perform the role of brokers and are often useful in connecting groups that are otherwise disconnected and distinct (Robnett, 1999; Hine & Morris, 1985; Polletta, 1998). There is also evidence to suggest that having ties with influential actors increases the probability of protest participation (Opp and Gern, 1993; Oliver, 1984; Gould, 1991; Snow et al., 1980).

With the advent of social media, and greater volumes of data, newer methods have emerged which use network properties to identify actors that have the capacity to be influential. In many ways, this stems from the previous strand of studies, as they form the theoretical foundation for hypothesis. For instance, Barbera et al. (2015) used the number of ties measured as retweets between protestors to hierarchically cluster users according to connectivity through a K-core decomposition algorithm. They claimed that the group of actors that are the most densely connected are more active in the network. Under similar logic of connectivity and methods, González-Bailón et al., (2011) claimed that the shells obtained from K-core decomposition can be used to locate influential spreaders of information in the network.

According to many social movement theories, the presence of these influential actors (famously understood as leaders) is almost bound to be effective for movements as leaders are well placed to form effective ties with people, organizations and accumulate resources (Krinsky & Crossley, 2013). However, these results are not sufficient to claim that leaders are necessary for social movements to be effective. Nor are they claims of causality. In fact, with the advent of social media, studies have started to challenge these claims (Bennett & Segerberg, 2012; Brett Caraway, 2018; Copeland et al., 2016). According to Western (2014), autonomous leadership is formed through participative democracy and enabling leadership from an individualized perspective. Social media is often cited as an influential factor in facilitating this horizontal orientation of protests. Thus, there is no clear consensus on quantifying the exact importance of leaders in social protests[5].

[5] Borrowed from MY400 Assignment 2

Collective Identity and Framing

The process of formation of collective identity offers a framework to study the role of influential actors in social movements. Alberto Melucci (1988) described the process of collective identity as a formation of 'common cognitive frameworks' amongst the participants. According to Olson (1965), a discussion of collective identity is central to understand why people would want to participate in social movements and incur risky trade-offs when they are unaware of the impact of their participation. This problem is commonly described as the collective action problem where people chose not to protest because of the costs attached to protesting (free riding (Friedman et al.,1992), police brutality, arrests etc.).[6]

A key mechanism used to build collective identity in social movements is Framing. Goffman (1974) described framing as a way for individuals to 'locate, perceive, identify and label' the world around them. The process of framing understands social movements as actively involved in the production of meaning, rather than simply a carrier of existing meanings. According to Snow & Benford (1988), this process contains three core tasks – problem identification, forming possible solutions and building motivations.[7] They argued that for a frame to become effective in building a form of collective identity, it must resonate with the people. In this way, frames are more likely to be transmissible, as they align with the person's existing beliefs.

Thus, framing is at heart, a way of making a message more salient. The link between framing and identity formation is explained through the process of diffusion. According to Oliver and Myers (2002), the actions of protestors in social movements are explained by a network of transmission that links people to external social behaviors. They claim that framing (an action by nature) can also be explained through this process of diffusion – analogous to diffusion of knowledge. Thus, diffusion facilitates exposure, and exposure to frames is a pre-condition for people to react and participate in identity formation with other actors.[8]

Under similar lines of argumentation, Jasper & Poulsen (1995) argued that the construction of a social network is insufficient for its growth, as networks are just a way for people to co-exist in proximity. It is the transmission of information within these networks that influences people to

[6] Borrowed from MY400 Assignment 2
[7] Borrowed from MY400 Assignment 2
[8] Borrowed from MY400 Assignment 2

join the movement. Consider for example a Twitter user who tweets about a topic related to protests which they deem important. For people to resonate with that framing, they must be exposed to it. Networks thus facilitate this exposure by diffusion, and some nodes are often placed in a unique position to strengthen this exposure.

This paper does not try to adjudicate which of the two instruments are more important for social movements to grow. It is probably true that both processes play a crucial role in creating effective mobilization campaigns. Instead, this paper considers both these mechanisms as plausible explanations for success of social movements.

Empirically, multiple different methodologies have been employed to identify the prevalence of frames in social movements. For example, Xiong et al. (2019) used text analysis to identify different frames that emerged during the #MeToo movement on Twitter. They studied keyword frequencies, hashtags, co-occurrences, and used thematic analysis to identify keywords that belong to similar contexts. Different types of frames were identified, like themes related to the movement, themes that refer the victim, themes that were action oriented etc. Thus, the authors fundamentally focused on framing as a latent process, which can be detected using patterns of communication.

On the other hand, Jasper & Poulsen (1995) focused on frames from a point of view of recruitment. The authors were specifically interested in identifying the effectiveness behind framing mechanisms, and they did this through content analysis of Animal Rights and Anti-Nuclear movements, and their resonance with protestors which was found using interviews and surveys. Some of the symbols they identified included capitalism, animal torture and resonance with nature.

These studies show that framing is an inherently subjective process, and if it contains messages that are salient to a movement, they hold the capacity to resonate with a subset of people. This paper considers framing as a dynamic process, where its usage not only resonates with people, but also increases the reach of protests. The second criteria are introduced as it is unclear how resonance can be measured, as every actor has varying interests. It is extremely tough to empirically determine these interests using a selection of tweets for that actor.

Importance of influential actors in framing Social Movements

Due to their social and political capital, the identification of influential actors in movements are often at the center of interest in research related to social movements. While many different interpretations of influence exist, this paper considers influence from two commonly used lenses. First, a network sense, i.e., those who are positioned in a unique way to influence how information flows between actors in a network and second, in a thematic sense, i.e., actors that have the resources to transmit frames across the network and increase its reach.

Empirically, the importance of these influential actors in facilitating or 'leading' this process of framing is unclear (Buechler, 2016). According to Morris & Staggenborg (2002), actors in social movements have the educational skills that make them better framers. Further, they suggested that they can (and should) form a diverse range of framing to resonate better with the aggrieved population. In similar directions, Studies like Polletta & Jasper (2011) and Valocchi (2009) argued that influential actors have an incentive to effectively frame social movements to achieve organizational legitimacy, efficiency, and stability. Examples of such instances include the usage of a 'rights' frame by civil rights leaders (Snow and Benford, 1992) to gain mobilization and an 'environmental' frame by leaders during the Nigeria's Movement for Survival of the Ogomi People (Nepstad & Bob, 2006) to gain international media attention.[9]

Literature on social networks also indicate that the relational characteristics of influential actors allow them to influence processes of diffusion in the network. For example, Mbaru & Barnes, (2017) showed how different centrality measures can be used to identify actors that facilitate different types of diffusion process. Nodes that act as brokers between two groups, facilitate diffusion of key information between the groups, whereas nodes that have a high number of connections would rapidly increase diffusion of information in the network through these nodes.

On the other hand, studies point out that with the advent of social media, there are equally compelling incentives for bottom-up framing. According to Bennett & Segerberg (2012), social media has enabled the conceptualization of 'personal action frames', wherein sharing of information is done through personal experiences as a form of catharsis. The growth of sustainable movements is explained by the visibility and exposure to these frames (van Haperen

[9] Borrowed from MY400 Assignment 2

et al., 2018). This idea is supported by studies like Brett Caraway (2018), where using posts of a Facebook group of Walmart employees, Caraway found that there is a growing role of personalized forms of communication in generating collective action among members. Similarly, Copeland et al. (2016) used tweets related to a call of boycott of Chick-fil-A to find that activist organizations neither influenced nor adopted the citizens frame on the issue[10].

Further, using Network Analysis, Barbera et al., (2015) found that even though the more committed part of the network tends to be more active on a per-capita basis, the peripheral part of the protest network tends to be highly aggregated, and thus increases the reach of information across the network at scale. Thus, even though the non-influential part of the network has less capacity on a user-level, on aggregate they tend to be highly influential in information transmission. Similarly, Vaast et al. (2017) used microblogging data related to Gulf of Mexico oil spills and found that interdependence between users on social media allowed them to exchange distinct information patterns.

However, there is little research done to empirically investigate the interaction between network influence and adoption of frames. A natural question that stems out of this literature is one of *frame setting*- who sets the scene? Research indicates that either scenario is possible. But little analysis is done to empirically determine which is more likely. In one such study, Copeland et al. (2016) used Granger's causality test to assess the relationship between activist organizations and individual's frames during the call for a boycott. They find that organizations neither influence nor adopt public's agenda. However, the study relies on labelling each tweet to only one frame and restricts leadership to activist organization. Additionally, their case is that of a tactic in social movements and is thus fundamentally narrow.

This paper aims to solve these issues by a) studying social movements as networks – reframing the interpretation of influence as *relational* and not *self-identified*, b) identifying frames retrospectively through topic modelling[11] of tweets of protestors– a technique that allows the presence of multiple frames in tweets and c) using a yearlong protest movement as a case study.

[10] Borrowed from MY400 Assignment 3
[11] Topic modelling is described in the methodology section.

There are two questions of substantive interest in this study– the first is about *information flow* between central and non-central actors in social movements. This question explores how people source information regarding protests, and which groups are most active in the diffusion process. The second is about *identification*, which explores whether we can identify common frames used among both these groups. The third question is about *frame setting.* This question explores whether there is evidence to indicate that central actors *led* prominent frames during the protest, or *followed*, or neither.

The research questions are summarized below:

a) *Information Flow:* What was the nature of information flow between influential and non-influential actors during the discourse surrounding Farmers protests on Twitter?
b) *Frame identification:* Can we identify common frames between the two groups?
c) *Frame Setting:* Is there evidence to suggest that either of the groups is likely to lead the other vis-a-vis any of the topics identified?

The usage of network analysis for identification of influence, along with topic modelling for identification of frames provides in-depth analysis of flow of frames across the network, adding novel insights into research on influence in social movements.

Methodology

(a) Data Collection

Twitter was a primary source of data collection in this study. Tweets were collected using the Twitter Academic Researcher API. This API gives access to historic tweets as well as user-level information using various combinations of queries. The Academic Research access needs application approval and has an upper limit of 10,00,000 tweets per month.

Using the academictwitteR (Barrie & Ho, 2021) package in R, tweets relevant to the protests were scrapped using keywords and hashtags from 1st September 2020 (the month when the bills were passed by the Indian Parliament) for a period of 1 year until 1st October 2021. This timeframe should give a sufficient sample of tweets, while also covering all major events in the protests.

i) Why Twitter?

As of January 2022, India has the third largest audience in twitter in the world (Clement, 2018). Moreover, with the announcement of nationwide lockdowns following Covid-19, it became extremely hard to conduct protests using conventional on-ground mediums like sit-ins, strikes etc. Thus, social media became a natural alternative where people discussed key information related to the bills, their grievances as well as amplification of protests that were held physically.

This puts the farmers protests in a unique position and poses challenges as well as opportunities for the paper's research agenda. On one hand, there is evidence to suggest that people turn to Twitter during political uncertainty (Bajpai & Jaiswal, 2011), and there are certainly reasons to believe that people would make use of the platform as a safe space to participate in political discussions (Shirky, 2011), talk about their experiences (Gill & Orgad, 2018) and mobilize transnational support (Theocharis et al., 2013).

ii) Query Selection

Data collection was divided into two stages: The first collects an initial sample of tweets from a query containing hashtag which depicts dissent against the Farm Laws, whereas the second collects all tweets related to the protests from a random sample of these dissent depicting users. This strategy is chosen to identify tweets and users that most likely participant in protests. Similar data collection strategies[12] have been used before to classify people as supporting/opposing a policy. In stage 1, an obviously dissent depicting hashtag is used to sample tweets (and retweets of those tweets) that are against the farm laws. The use of this hashtag is an indication (intuitively) that the user is against the movement, and while it is theoretically possible that they may change their stance later, such a sample is likely to be extremely low in proportion to the total sample of tweets. Any other collection strategy that relies on prediction of political leaning is not compatible with the research agenda, as many people who might be on the political right might still oppose the farmer's protest. In general, there is little to no evidence that accurately classifies the political leaning of people who support or oppose the Act.

---

[12] Inspired from data collection strategy of the Capstone Project 'Polarisation and attitudes on Twitter in the context of Brexit following the EU referendum'

Stage 1: Initial Sample

In the first stage, a combination of simple hashtags that express dissent against the farmers protest was used to collect daily tweets posted from 1st September 2020 until 30th October 2021: #TakeBackTheFarmLaws and #IStandWithFarmers. The advantage of collecting daily tweets data is that in a fast-flowing environment, no important information is left out. Any alternative contains the risk of a potential bias towards the end of the limit imposed, specially in cases where tweet volume tends to be extremely high.

Keeping computational time limits in mind, a total limit of 2 million tweets (including retweets) was imposed in this search. Importantly, only tweets in English language were scrapped. This is because most text analysis libraries are suitable for English words, and it is infeasible to look at every single language spoken, as India is a diverse country where as many as 121 languages are spoken by people. However, this is absolutely a limitation in data collection for this study and should be extended in future work to include a larger corpus, with other languages as well.

After combining all the filters, a total of 2,64,771 tweets are collected in this stage, out of which 35,656 are original tweets, whereas the others are replies/retweets. A random sample of the text of 5 tweets collected is attached in Appendix 1.

Stage2: User tweets

In the second stage, using unique tweet IDs collected in stage 1, a random sample of 25% IDs was curated, using which another query was generated which collected all tweets containing the hashtag '#FarmersProtest' posted from any of the randomly selected accounts. Similar date and language limits were imposed as stage 1, resulting in another 8,13,551 tweets. The final tweets are simply a combination of the datasets obtained from stage 1 and stage 2 resulting in a total of 10,78,322 tweets. Removal in duplicated tweets resulted in 10,65,112 tweets from 95,157 users. Figure 1 depicts the full collection process.

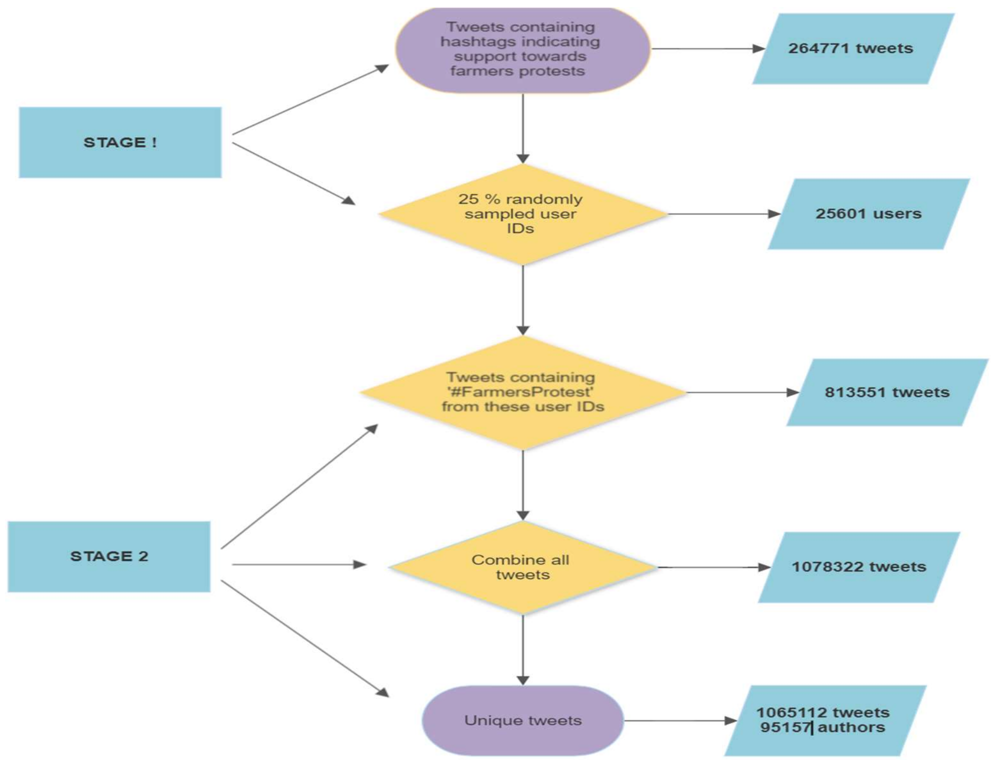

Figure 1: Data Collection

b) Methods

i) Information Flow

Retweet Network

Like previous studies related to protests (Barbera et al., 2015; Mccaughey, 2014, p. 33), a directed, weighted network was constructed using retweets between users. Thus, a tie from user A to user B implies that user A retweets the original message sent by user B. Further, since the aim is to understand who retweets whom, and by how much – the network was directed and number of times a person re-tweets another was included as weights using the igraph package in R.

The reason why retweet was used as a metric for network construction, is that intuitively, when a user retweets another, it results in an overall increase in the reach of the tweet. In this way, important information about the protest can be communicated, and people who are not previously exposed to crucial information can get to know about the protests. In general, exposure is a necessary condition for people to participate in protests – if someone does not know about the protest, they cannot participate in it. In this way, the retweet network gives a proxy of diffusion across participants in the protest. Summary statistics about the retweet network are highlighted in Table 1:

Metric	**Value**
Number of nodes	19,055
Number of edges	41,701
Max aggregated retweets received by user	4,601
Max aggregated retweets sent by user	5,541
Max people connected with a user through retweets	1,304

Table 1: Summary statistics for retweet network

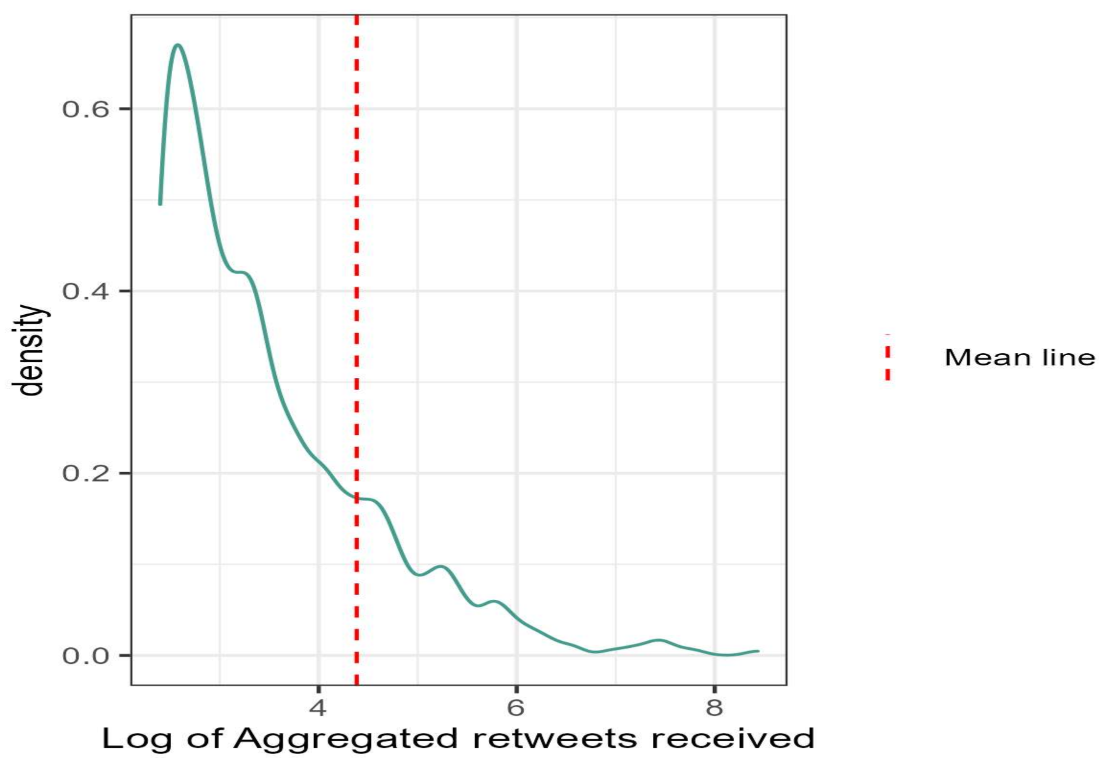

Figure 2: Aggregated Retweets Distribution

Figure 2 shows the distribution of aggregated retweets received per node for this retweet network on a log scale. For sake of visualization, only nodes which have received aggregated retweets above 10 are included. Clearly, the plot indicates a right skewed distribution, where very few people have received a high volume of retweets, whereas most users have received a low volume of retweets. Even in the absence of outlier values, the plot indicates that most people tend to source their information from a group of few users.

K-core Decomposition

This retweet structure was further used to analyze communities of connectivity. An instrument that helps with this analysis is the coreness of a graph. The K-core decomposition algorithm is one such algorithm which is used recursively tune a network into a hierarchical structure based on connectivity. The K-core algorithm has been used previously in social protests networks (Barbera et al., 2015; González-Bailón et al., 2011).

The intuition behind using coreness as a metric for protest networks is that the nodes who are very well connected in terms of the frequency of retweets and being retweeted transmit information related to protest with high speeds. An analogy can be drawn with infection spread, as it is important for nodes to not just be well connected, but also be well-connected with other neighbors who are themselves well connected (Kitsak et al. 2010). Figure 3 depicts K-core decomposition for a random graph with 5 nodes and 7 edges.

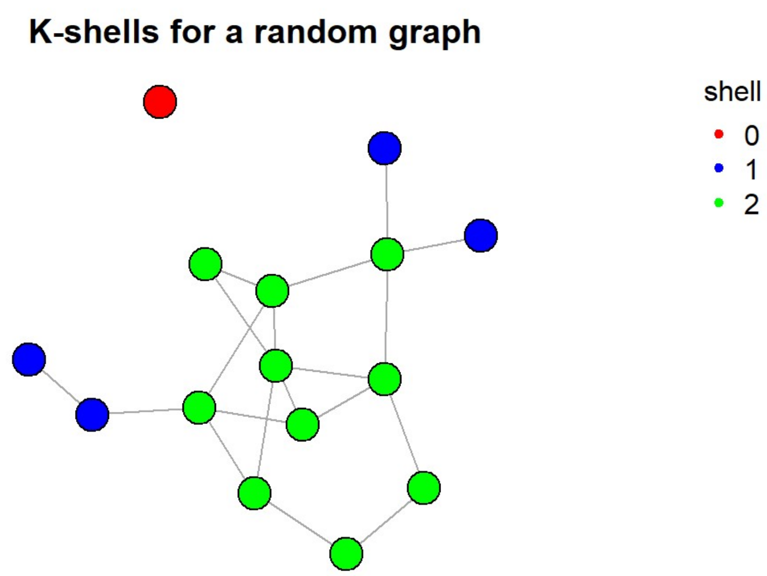

Figure 3: K-shells of a random graph. This graph shows how nodes are grouped using the K-shell decomposition. In the first stage, node 1 is removed, as it is an isolate. Then, the number of connections for other nodes are calculated. The nodes in blue are those with only 1 connection. Thus, they get grouped in the 1-shell.

A k-core is a maximally connected subgraph, where each vertex in a k-shell is connected to *at least* k other nodes. Thus, a 0-shell contains all members of the network, as no node can have negative connections. The k-core decomposition finds the largest subgraph (G_k) for which each node in G_k has at least k other neighbors. Through a recursive process, in which nodes with degree less than k are removed, each node is assigned to a k-core – which basically indicates the maximum value of k for which that node can be present in a k-shell. In the first stage, nodes with degree less than 1 (isolates) are removed and the graph is reconstructed without those nodes. In the second stage, nodes with degrees less than 2 are removed, and so on.[13]

The more connective part of the network forms the core, whereas the less connected part forms the periphery. In the core of the graph, members are closely connected to each other, as well as

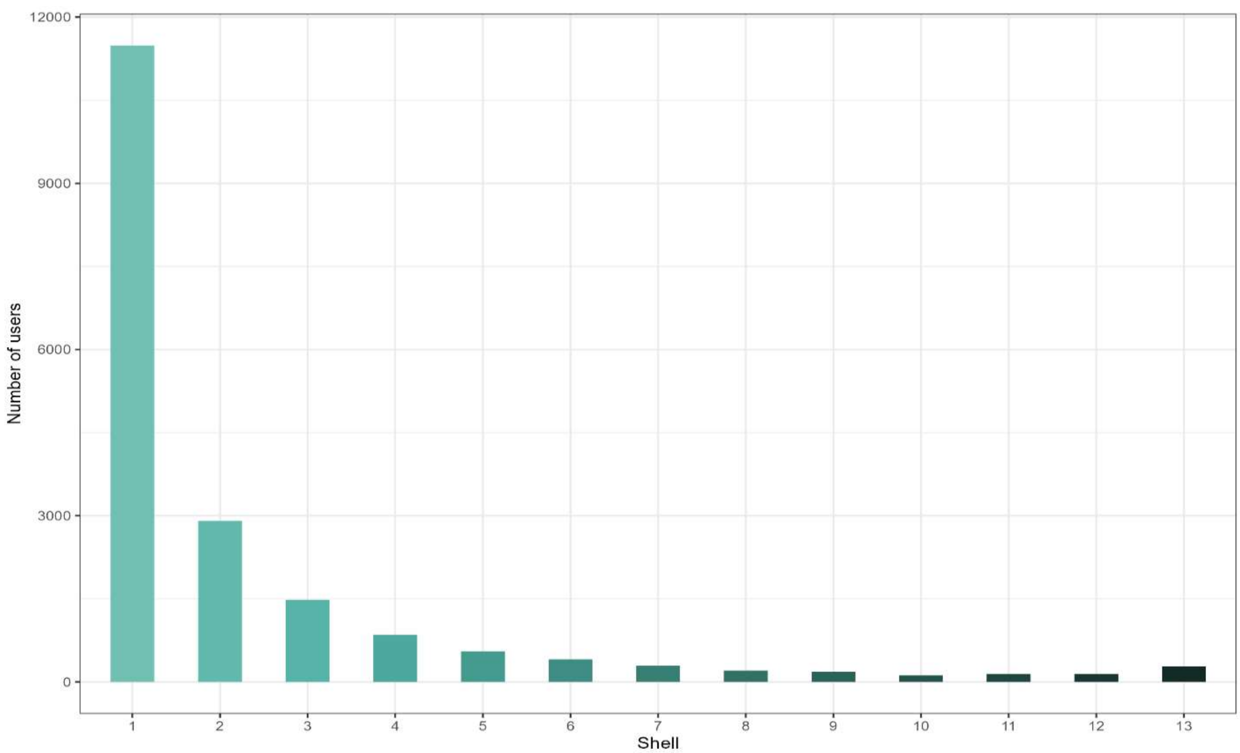

Figure 4: Users per K-shell

the peripheral nodes. However, in the peripheral shells, members are more loosely connected. In the retweet network, this core represents the most connected protestors where connections are measured through retweets sent and received. As Borgatti et al. (2018) points out, all participants in a core are highly central, but not every central actor forms a core.

[13] Borrowed from MY400 Assignment 2

This algorithm resulted in 13 cores for the retweet network[14]. Figure 4 summarizes the number of users allocated to each shell for the network. The plot indicates that most users belong to shell-1 whereas the number of users progressively decreases as we move to the shells with more connectivity.

ii) Frame Identification

Research question 2 draws on the identification and prevalence of frames between central and non-central actors. It is important to note, that there are no formal criteria of frame identification. As discussed earlier, frames are considered a process of making a message more salient. Thus, the variable that matters is not the frame itself, however, it is the attention paid to that framing by different actors.

One way of identifying this attention, is through unsupervised topic modelling of the tweets. Unsupervised, as tweets are not inherently labelled documents, and thus, the algorithm makes certain assumptions about the formation of text and subsequently extracts recurring themes. A commonly used algorithm (also the algorithm used for this paper) is Latent Drichlet Allocation (Blei et al., 2003). LDA is a generative 'bag of words' model, which assumes that documents are distributed over topics, and topics are then further distributed over words. A separate topic modelling algorithm was applied over the denser: core groups and the less dense periphery groups, where each document was the aggregated tweets per day. Aggregation of tweets per day has been shown to increase accuracy of topic models relative to human judgment (Hong & Davison, 2010).

An advantage of LDA over any other alternative labelling methods, is that they will necessarily require hours of labor for classification into potentially hundreds of topics (Barberá et al., 2019). This is infeasible for the current corpus which contains almost 1million tweets. Moreover, since the aim is to understand topic prevalence, and not just the topic itself, LDA allows the possibility for a document to contain multiple topics in proportion. Thus, in essence, LDA allows for the flexibility in a document to contain multiple themes simultaneously, and their posterior distributions can be calculated.

---

[14] Isolates were not considered from the analysis to mitigate biases that might arise due to data collection issues. Connectivity amongst people through retweets strengthens the condition that most of these tweets were probably depicting dissent against the Farm Laws

The input to a topic model is a document feature matrix(dfm). A dfm essentially tidies unstructured text corpus into a matrix, where each row represents a document, and each column represents a word. The entry $A_{i,j}$ for a dfm 'A' represents the number of times word 'j' has been used in document 'i'. In this model, tokens upto bigrams are generated. This is because, LDA with bigrams has often shown better results for topic models on shorter texts, like tweets as they add cohesiveness to a topic for meaningful interpretation (Jónsson & Stolee, 2015). Table 2 summarizes the steps and reasons for all pre-processing steps taken for topic-modelling.

Pre-processing step	**Reason**
Each author id was assigned their K-shell value	The tweets from corpus were identified using their author ID's to be classified into K-shell
Tweets classified into shell > 7 were assigned to group 'core', whereas others were assigned to group 'periphery'.	An arbitrary value of 7 was chosen to differentiate the denser shells from the less dense ones, resulting in a fair distribution of number of tweets per group, as indicated in Table 4
Non distinct tweets (including retweets and replies) were filtered out from the sample	To prevent double counting
Tweets whose authors are not allocated to any cores were removed	These are authors who were excluded in the network, as they never retweeted, or had never been retweeted
All hashtags were removed	This is to improve interpretability of models. Users often tend to include multiple hashtags that are related to different topics in one tweet, hindering the overall cohesiveness of topics
All punctuations, symbols, numbers, English stop-words, and URLs were removed, words were converted to lower case	To reduce the number of features input into the topic modelling, these tokens do not add any value to the topic itself
Recurring keywords: 'farmers' and 'protests' were excluded	To reduce size of features input
Tokens upto bigrams[15] were extracted	To improve interpretability
Tweets were grouped by day	Improve efficiency of topic models, and create topic prevalence per day
A minimum word frequency of 20 words and maximum word frequency of 1000 words was imposed	To exclude irrelevant and rarely occurring tokens, as well as very popularly occurring tokens. This reduces input size as well as makes topic models more interpretable and exclusive

Table 2: Pre-processing steps for topic modelling

[15] Bigrams, in this context are words that co occur together in the text, for example, the bigram for New York is 'new_york'

All pre-processing steps were performed using the quanteda (Benoit et al., 2018) package in R, and then the two corpuses of tweets were each converted into dfms. Table 3 summarizes the final number of tweets considered for the document terms matrices.

	All tweets	Unique tweets
Core	4,82,632	38,618
Periphery	3,30,451	29,392

Table 3: Number of tweets considered for topic models

The next important step is to choose the value of the number of topics that are expected to be contained in the documents. There are various metrics used to assess the appropriate number of topics that should be fed into the document. Two metrics are chosen for this analysis, for values of topics (K) ranging from 30 to 100:

1. Log Likelihood: The Log Likelihood measures the likelihood of observing the model parameters, given the data. In general, the larger the log-likelihood, the better the model fit.
2. Exclusivity: Exclusivity measures the extent to which each topic is exclusive, compared to other topics. The mean exclusivity is calculated to check for average level of exclusivity across topics.

Both these metrics were calculated using the topicmodels package (Grün & Hornik, 2011) in R. The results are summarized in Figure 5 and 6.

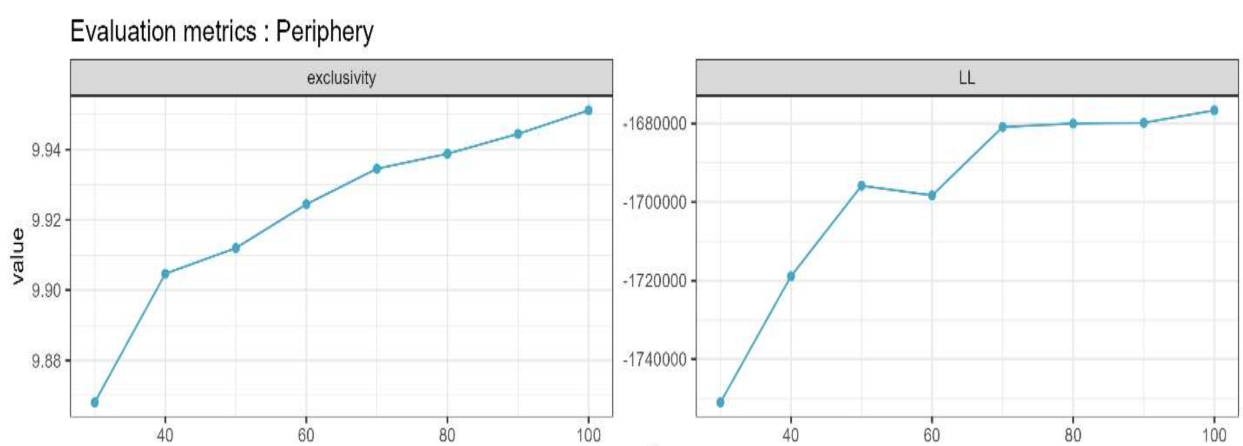

Figure 5: Evaluation Metrics: Periphery

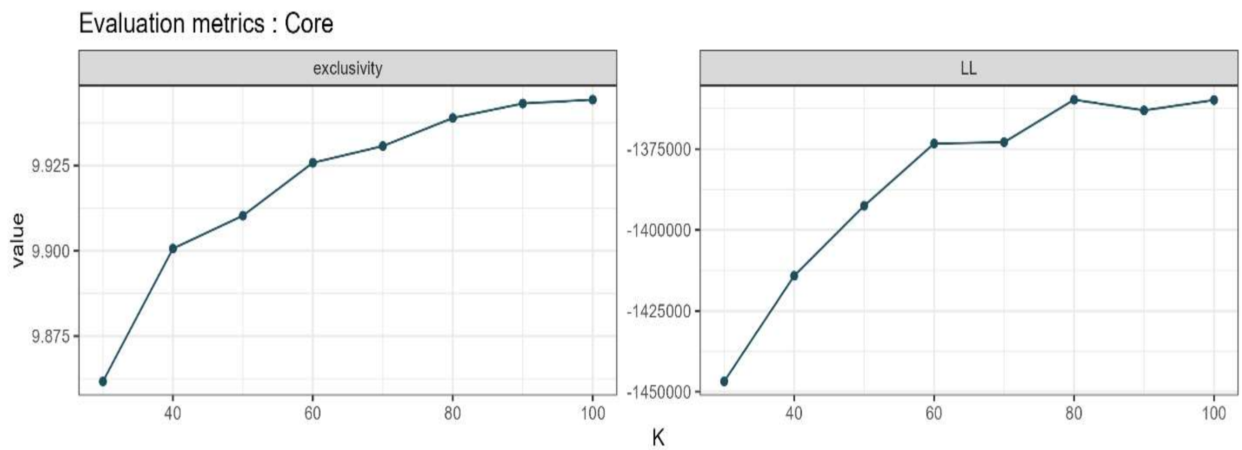

Figure 6: Evaluation metrics: Core

Clearly, both the metrics improve substantially up until K = 100. Thus, keeping computational as well as resource limits in mind, model with 100 topics was chosen for both the groups. It is important to note that there are no objective criteria dictating the selection of K. For our questions, it is most important for topics to be meaningful (for labelling) and exclusive (to minimize overlaps). Since the goal was to map topics obtained from core to topics obtained from periphery, and identify common, meaningful[16], distinct[17], and salient [18]topics, these selection criteria are sufficient to ensure that the topics (and more generally, their proportions) are indicative of the *trends* observed in the discourse surrounding farmers protests.

[16] Topics chosen must be meaningful, and clear from the analysis of top 20 words most closely related to them
[17] There must be minimal to no overlap between the themes of selected topics and other topics
[18] The topics must be related to discussions surrounding the Farmer's protests.

To validate this claim, all 100 topics were manually labelled using top 20 words associated with each topic from each group. A total of 7 topics were selected for further analysis, these topics and their keywords are attached in Appendix 2. There were three criteria imposed in the selection of topics:

iii) Frame Setting

To answer the group that leads discussion, the daily prevalence of topics discussed amongst both the groups was used to create a time series, where each row is a day in the timeline, and each entry contains the proportion of given topic on the day. Thus, for example, in a three-topic model, the proportion of topics contained in a day would be in the form 25% topic A, 20% topic B and 55% topic C.

Inspired from the methodology used in Barbera et al. (2019), these time series were used to predict the extent of influence of a group using a Vector Autoregression Model (VAR). A VAR model presents a variable as a function of its own past values, and the past values of other variables. How back we want to go is defined by the 'lag' value, in the equation, which indicates the number of days that we want to look behind while calculating the predictions.

Equation 1 was used to measure predict the influence of topic prevalence by the two protestor groups. Here, g = Group (1 for core participants, and 0 for peripheral), T is the topic and t are the day of the time series. L represents the maximum lag to be included. The lag is chosen by comparing Alkaline Information Criteria (AIC) for various values of lags, which is selected automatically using the 'vars' package (Pfaff, 2008) in R. AIC is a commonly used and effective metric for model selection for VAR (Ivanov & Kilian, 2005), and it represents how well a model fits the data. Table 4 shows the values of maximum lag chosen for each model based on the AIC criteria.

$$Y_{g,T,t} = \alpha_T + \sum_{g=0}^{1} \sum_{l=1}^{L} \beta_{g,l} Y_{g,T,t-l} + \epsilon_{g,T,t}$$

Equation 1: VAR model on topic prevalence data

Topic	lag(p)
Suppress Dissent	5
Singhu Border	2
Rihanna	1
Republic Day Violence	5
Crony Capitalism	7
Media Hypocricy	5
Covid	13

Table 4: Maximum lag based on AIC values

The coefficients obtained from equation 1 were tested in two ways: first through impulsive response functions, and then through a Granger Causality Test. IRF depict how a time series responds to a sudden change in another, where Granger Causality tests whether a time series can be useful in predicting another. The following hypothesis were tested for Granger Causality:

Core => Periphery
H_0: $Y_{g=1, T}$ does not granger cause $Y_{g=0, T}$ for all common topics T
H_a: $Y_{g=1, T}$ granger causes $Y_{g=0, T}$ for all common topics T
Periphery => Core
H_0: $Y_{g=0, T}$ does not granger cause $Y_{g=1, T}$ for all common topics T
H_a: $Y_{g=0, T}$ granger causes $Y_{g=1, T}$ for all common topics T

Table 5: Granger Causality Hypothesis

Granger causality compares a restricted model (without additional lags other than own time series' lags), and an unrestricted model (with additional lags of the other time series), and tests how useful the lags of the time series are using an F-test. Importantly, it does not make claims about causality, and the name is a misnomer.

Results

i) K-core Analysis

The first research question in this paper is about the information flow of the retweet network. To answer this question, K-core decomposition analysis was used, resulting in 13 groups of cores

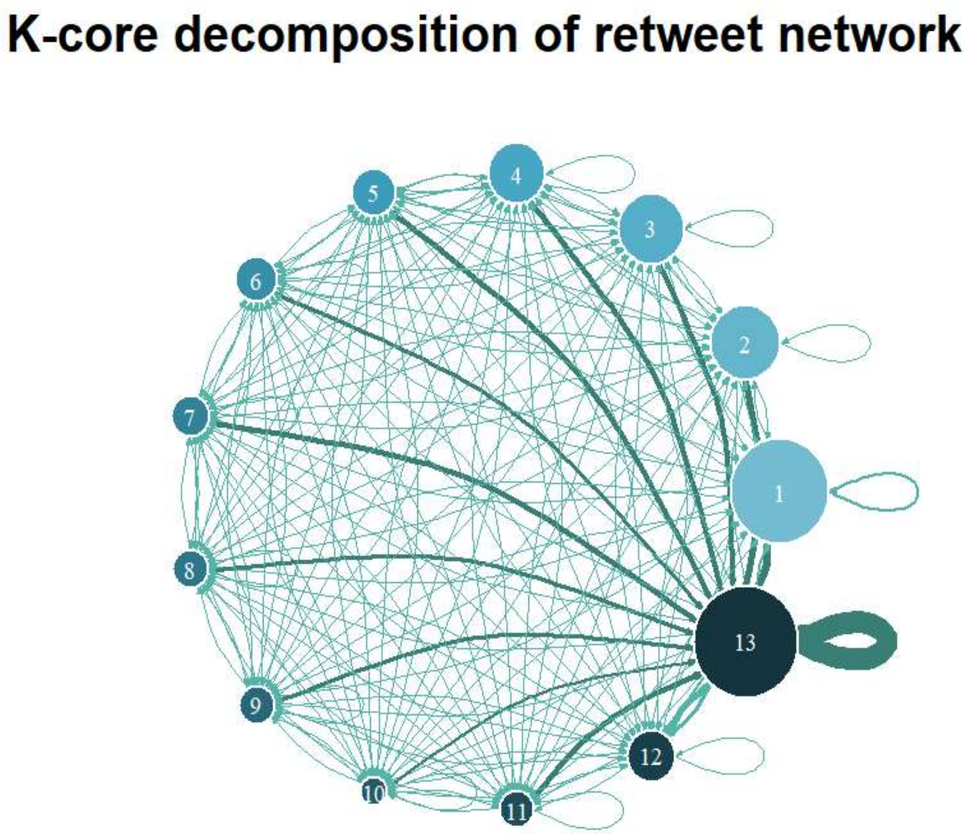

Figure 6: K core Decomposition

In this graph, users are grouped into their k-shell. The width of node is proportional to the number of tweets posted by the group and the width of edge of the graph is proportional to the number of retweets exchanged between the groups.

detected by the algorithm. Figure 6 represents a directed graph of the 13 groups. This network, like the retweet network is constructed using number of retweets exchanged between the groups, and the groups consist of users that have either been retweeted or have retweeted at least once. Here, each group is a node, and each edge is sized in proportion to the number of times a group has retweeted the other. The nodes are sized by the number of tweets sent (including retweets) by all users in that group.

The graph gives a few crucial pieces of information: Firstly, the retweets received for shell 13 are much higher than other groups. This indicates that most of the retweets in the network are sourced from the highest 13-shell. Secondly, the densest group has almost the similar tweet activity as the lowest 1-shell[19]. This indicates that even though group with core 1 is less connected, the aggregated group activity is like that of core 13. Why is tweeting about the protest important? In the absence of any offline medium to organize, and conduct on-ground protests, talking about them on social media is possibly one of the most effective ways to voice dissent.

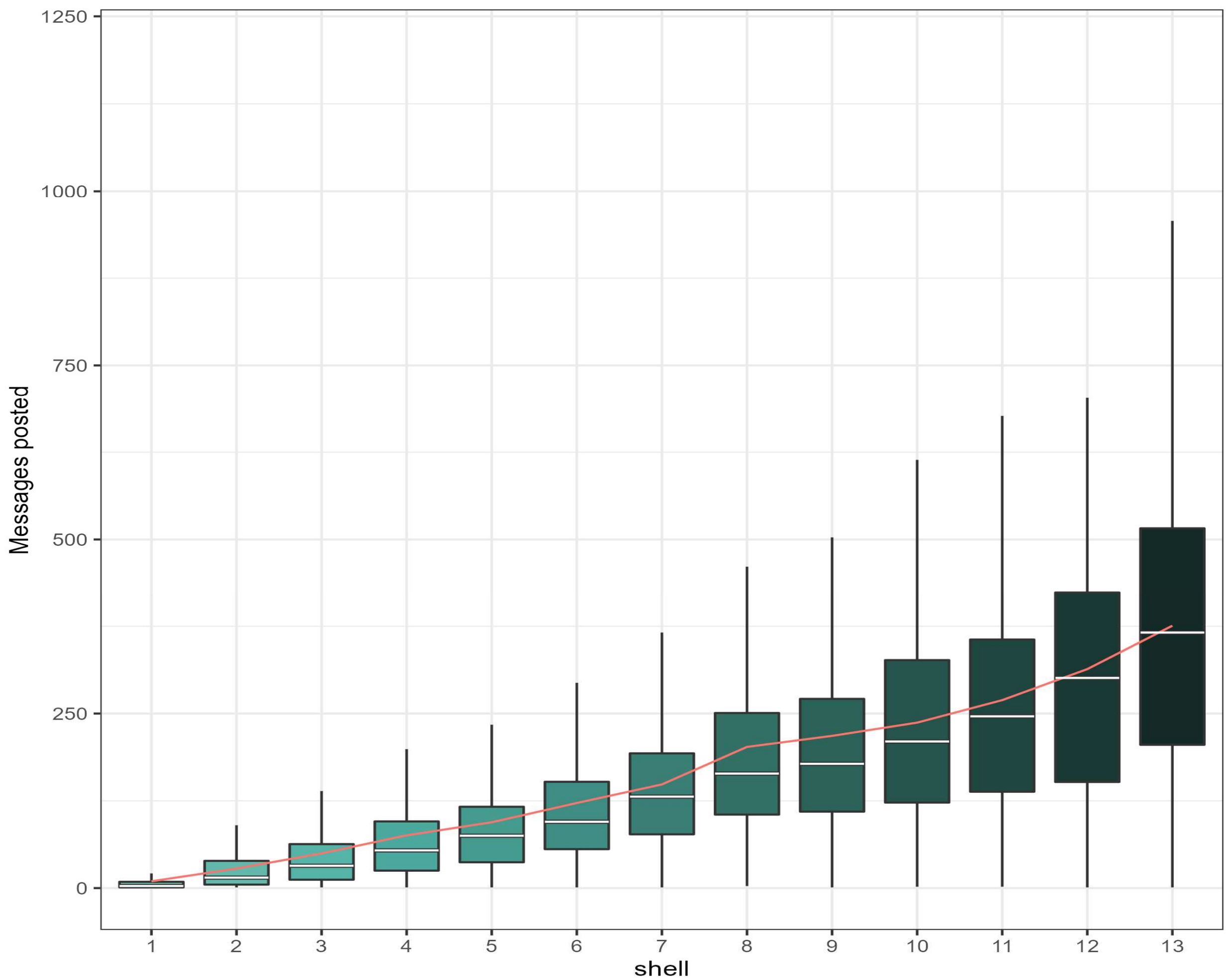

Figure 7: Number of messages posted per shell

[19] Note that the algorithm probably underestimates the aggregated activity of users in the lower shells, as isolates have been removed from the analysis. In reality, users from lower shells probably have higher aggregated contribution to the network's activity

This is validated by the sheer volume of tweets received daily by only a subsample of users. Secondly, the number of tweets show how involved the node is in the information creation process. Thus, the more the number of tweets, the higher the overall activity of the network.

Next, to analyze the direction of information, two critical pieces of information are conveyed by Figures 7 and 8: the first is that users in higher k-shells tend to post a higher number of messages. The mean, as well as median messages posted are higher for higher k-shells, indicating that these are much more active on average. The 13-shell has the highest contribution to number of tweets posted in comparison to other shells. Further, the lesser connected shells, on aggregate have similar degree of contribution to the total number of tweets, however, the contribution per user tends to be lesser. Thus, while the 1-shell is very active in terms of number of messages posted in total, the per-user activity is substantially lower compared to other shells.

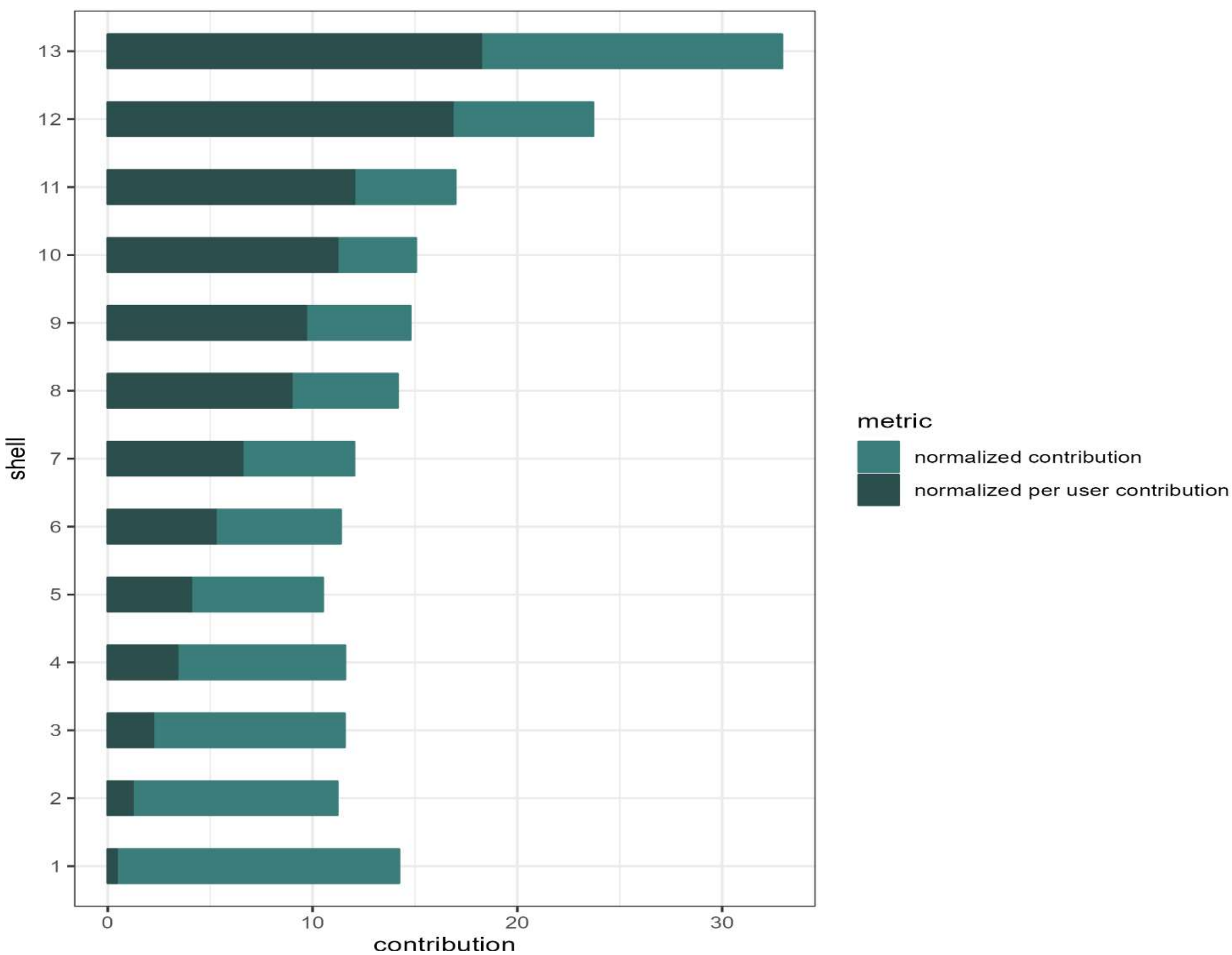

Figure 8: Contributions

Thus, in response to the first research question, there are strong reasons to believe that most information is sourced from the denser part of the network, in particular the 13th shell, as their aggregated retweeted numbers, and tweet activity tends to be higher. However, the less connected shells also seem to have a high level of aggregated engagement, suggesting that the groups in aggregate are equally as important in the network to increase overall activity of the frequency of tweets posted.

ii) VAR Analysis

The aim of this paper is to investigate in depth the role of these groups. Topic modelling was used to identify the most salient topics discussed in the original tweets of the two groups, where the denser group consisted of shells > 7, whereas the less dense groups consisted of shells < 7. As discussed previously, 7 topics were found to be common between the two groups. Since the topic distributions were extremely skewed, the logit of topic proportions was calculated and considered for the VAR. This is a standard practice used during time series forecasting and VAR models (Barbera et al., 2019). No time series was found to be non-stationary.

Figures 9 through 15 depict the time series for each of the topics selected from topic modelling. The daily topic proportion was estimated using the posterior probabilities of topics obtained from LDA model.

In response to the second research question, it seems like we can identify common salient topics across the groups using unsupervised topic modelling approach. Further, the time series plots offer interesting insights. For example, in Figure 10, we see that engagement for both the groups with respect to the protest site Singhu Border seems to be high all throughout the timeline of the protests. The topic starts gaining popularity around the end of December, as that was the time

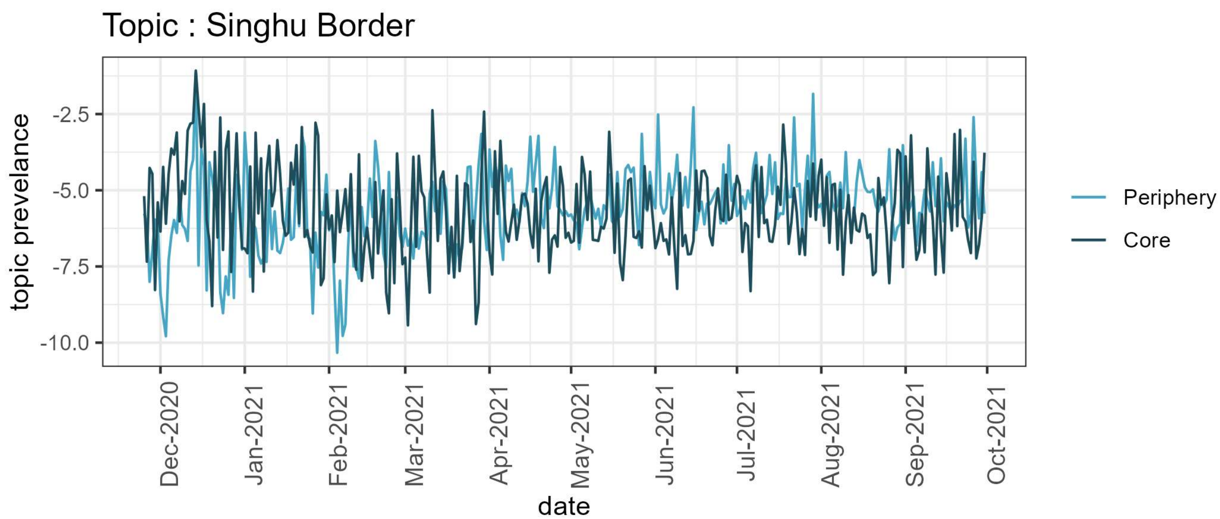

Figure 9: Singhu Border was a massive protests site where Farmers gathered and expressed dissent against the Farm Laws through demonstrations and sit-ins. Sit-ins continued throughout the duration of the protests, and people across the groups used identifiable words like 'border', 'singhu', 'delhi' etc which were used to identify these themes

when farmers first started to plan the sit ins. There are some noticeable seasonal spikes as well, for both the groups. However, they often tend to be in different times, suggesting that the topic gains popularity at different times amongst the two groups.

Similar trends are observed in other plots as well. Notably, the topics related to Republic Day Violence at the Red Fort and Rihanna's tweets have distinct peaks for both the groups near the time of the event taking place. It is not possible to inspect with naked eye who started these discussions, however, it seems like a further inspection will hold merit.

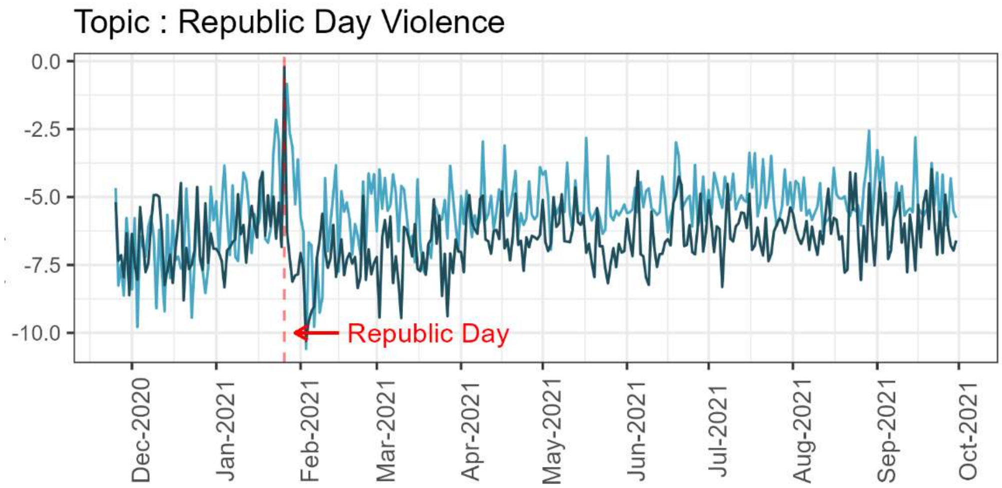

Figure 10: In India, Republic Day is celebrated on the 26th of January, commemorating the day that the constitution came into effect. On January 26th, protestors stormed the Red Fort (a key historical site in Delhi), only to be tear gassed and beaten by the police, leading to several injuries

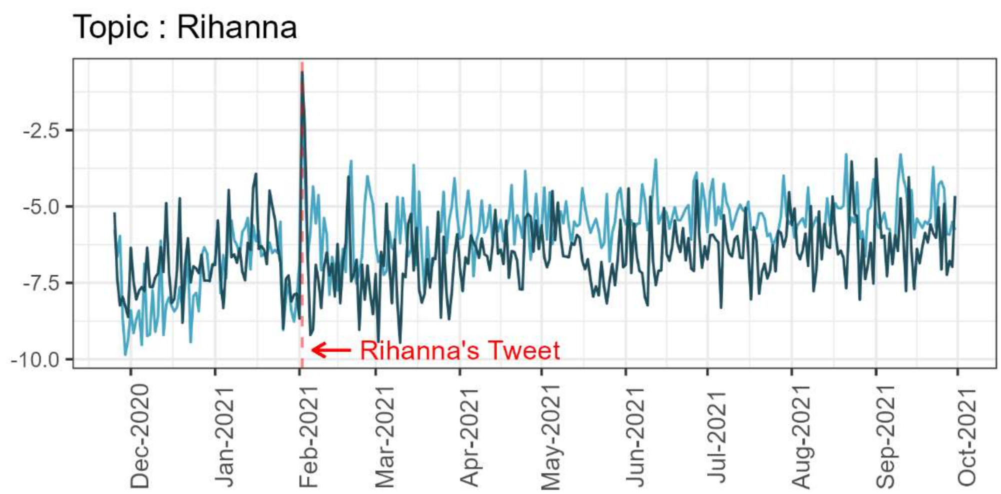

Figure 11: On February 2nd, Rihanna tweeted in support of farmers protests, which gathered massive discourse on Twitter. People who further talked about it mostly celebrated this as the issue was gaining international attention, and that was considered as a driving factor to the movement. Notably, this led to many people (those opposing the protests) criticizing Rihanna and her tweet, generating a political debate, while gathering significant attention

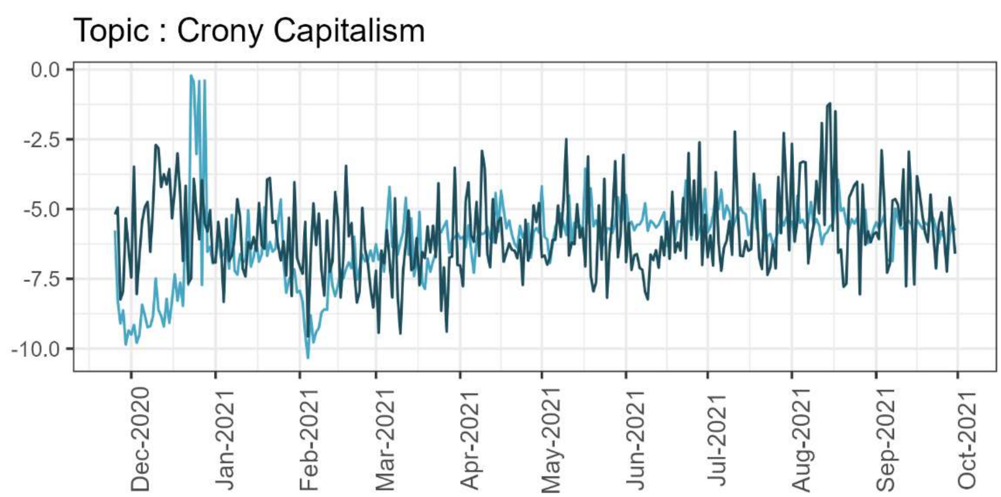

Figure 12: One of the grounds on which the Farm Laws were opposed was that privatization will increase access of big monopolists and capitalists into the agricultural market, allowing them to hoard supplies and manipulate prices. Notably, Ambani and Adani (two of the richest men in India) were cited as people with vested interests, who would distort the market given their political links and capital

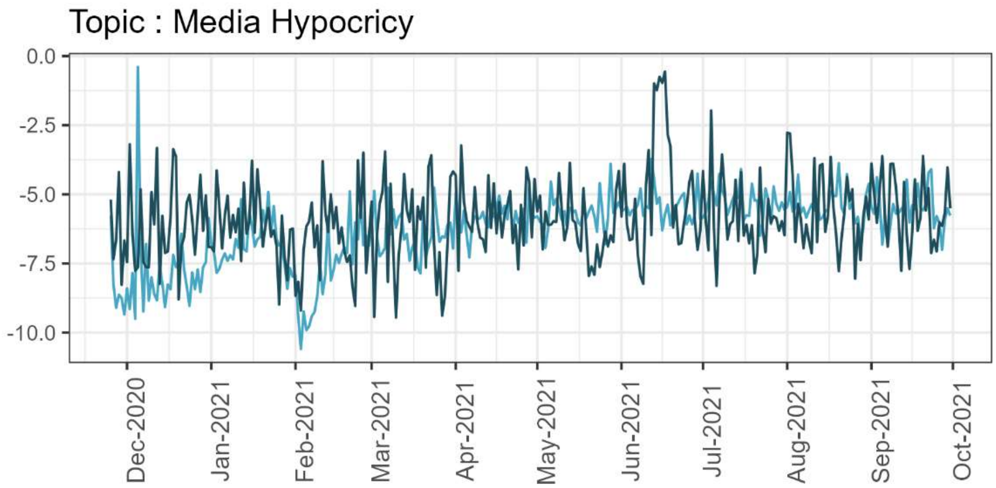

Figure 13: Many people criticized the media as being biased and pro-government on the issue of Farmer's Protests. This was a highly discussed issue where people criticized media coverage and portrayal of farmers as 'terrorists', or 'evil'

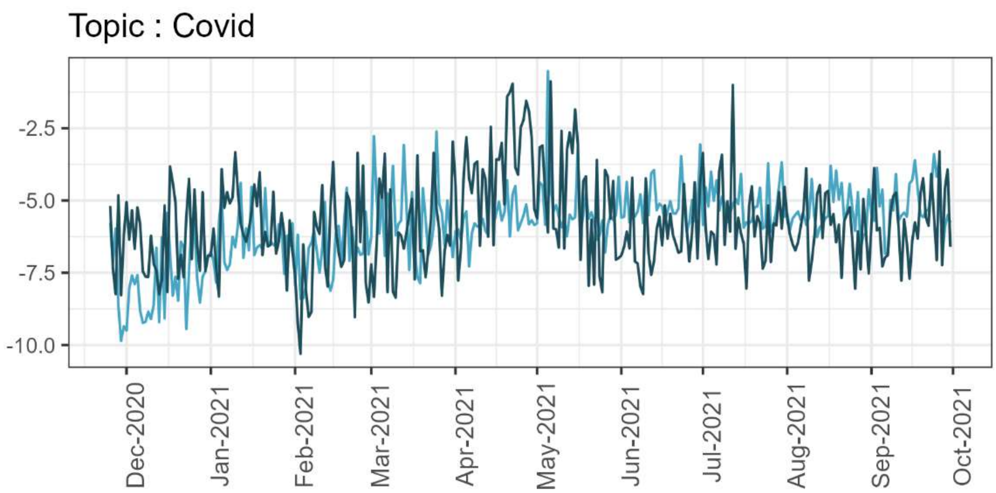

Figure 14: Since the protests were staged during the time of covid, people were simultaneously discussing issues related to covid in India and its intersection with the protests, notably vaccine shortage, lack of masks etc.

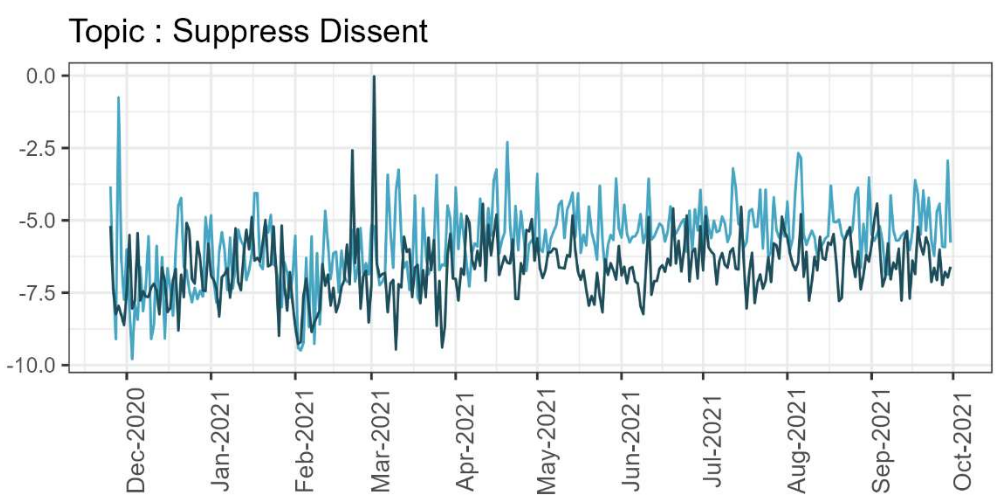

Figure 15: During the Farmers Protests, many Twitter accounts were blocked, and journalists were silenced, and people spoke up against these practices using words like 'silent', 'save_democracy', 'suppress', 'voices' etc

The VAR model can be analyzed in two ways: the first is through an impulsive response function, and second is through Granger Causality. Impulsive Response Functions indicate how the dependent variable responds to a 'shock' in an independent variable across the selected time horizon, whereas Granger Causality tests which of the predictor variables can be useful and significant in predicting the time series for the dependent variable.

Figures 16 and 17 represent the results of IRF performed on each of the topics with their respective lags, and a time horizon of 15 days:

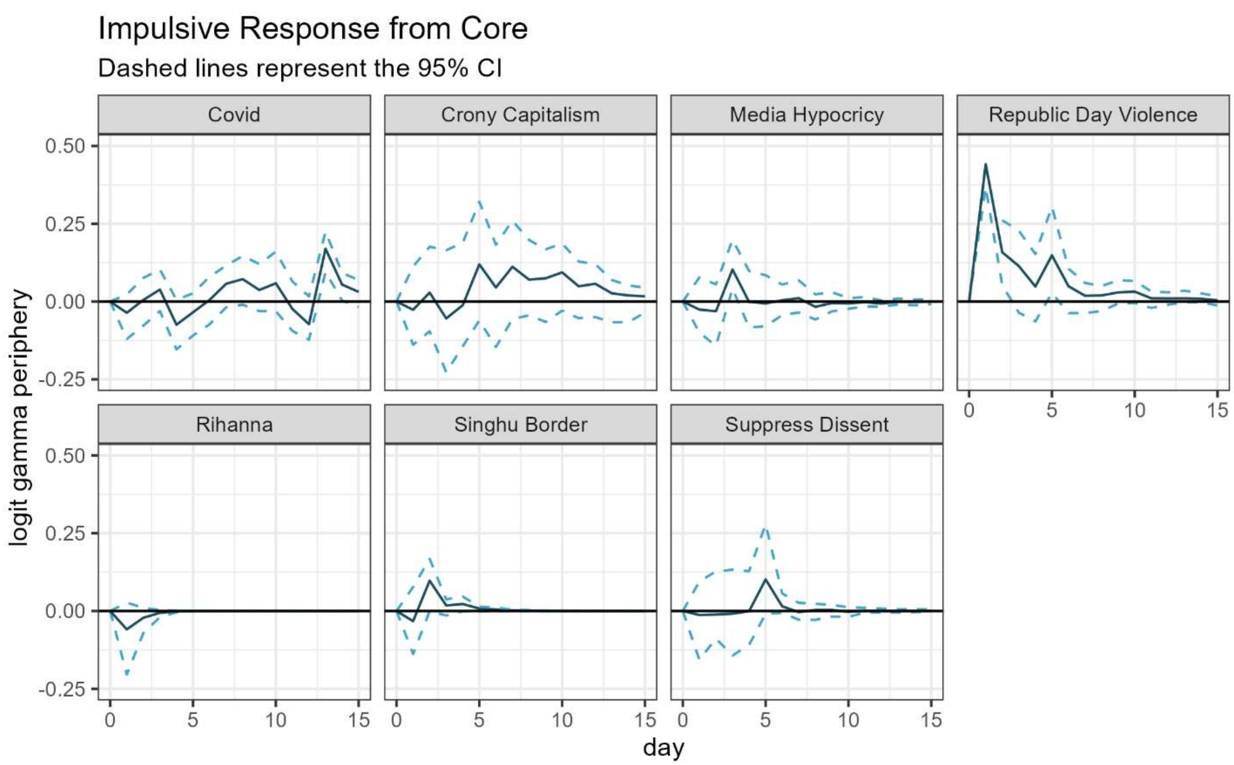

Figure 16: Shock from core

Figure 16 shows the evolution of topics in the peripheral group in response to a 1% 'shock' in the topic prevalence of the core group, thus it indicates how the prevalence of discourse in peripheral groups responds to a sudden increase in the prevalence in the core group. The figure shows that for most of the topics, an initial shock in the topic prevalence of the core groups does not lead to a sudden increase in topic prevalence of the peripheral groups in the first day, except for Republic Day. Most topics also converge to 0 by the end of day 15. A shock increase in prevalence of topic Crony Capitalism in core groups 5 days ago is associated with an increase in prevalence in peripheral groups. Similarly, a shock in prevalence for Singhu Border in core groups 2 days ago is associated with an increase in prevalence in peripheral groups. The lag is around 5 days for discussions related to suppression of dissent and 2 days for Media, the direction of association is positive in both cases.

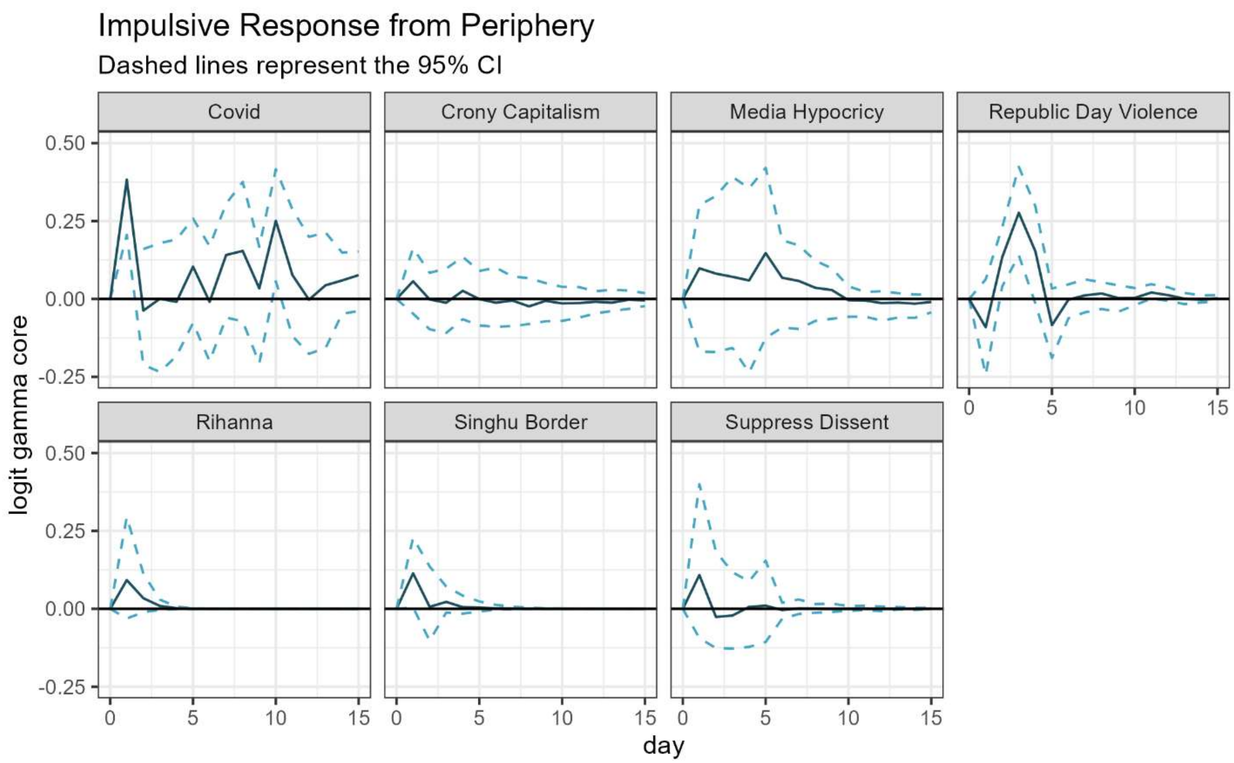

Figure 17: Shock from periphery

Figure 17 shows how the core group responds to a shock in the peripheral group. In contrast to Figure 16, the initial spikes observed in topic prevalence tend to be much higher for core groups, indicating that they may be influenced by a sudden change in topic prevalence in the peripheral groups. For most of the days, the topic prevalence also stays positive, in contrast to the core group, where topic prevalence is often negative. Increase in prevalence tends to be highest for topics related to Covid and Republic Day Violence, whereas it is the lowest for discussions about Rihanna's tweet.

Individual groups can be further compared using the cumulative IRF for all topics, as depicted in Figures 18 and 19. For topic Covid, the initial (day-1) as well as the final (day-15) values for both the groups are extremely high, indicating that the both the groups, are to varying degrees able to lead each other in discussions related to Covid-19. Importantly, the final day change for core group is higher than the final day change for the peripheral groups, indicating that the peripheral groups have a greater influence in the network with respect to discussions pertaining covid-19. The responses are measured here in percentages, thus a 1% increase in topic prevalence for peripheral groups 15 days ago is associated with an overall 1.25% increase in topic prevalence of Covid-19 in the core group. This is also a meaningful change considering the connectivity of the network, as well as the fast-paced discussion of protests, implying that an increase in the topic prevalence exponentially increases the reach of topic Covid-19.

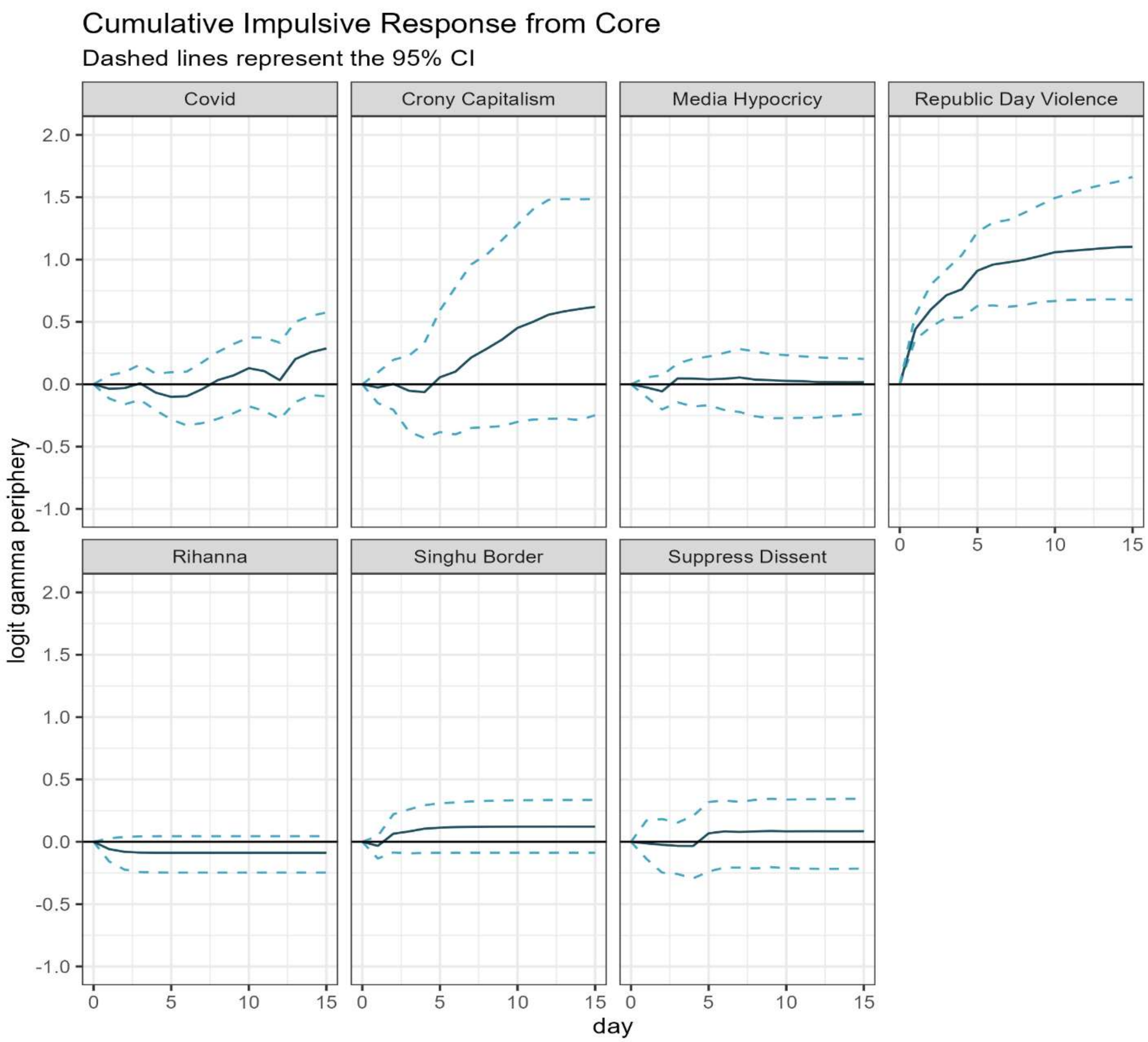

Figure 18: Cumulative Response of Periphery to shock in Core

The second interesting pattern is observed in the topic related to Republic Day violence. A 1-day as well as 15 days increase in topic prevalence is much higher for the peripheral groups (in response to shock from the core groups), compared to the core groups. Further, a cumulative 15-day increase in the topic for the peripheral and core groups are both high, at around 1% and 0.5% respectively, indicating that both the groups can lead each other with respect to the conversation around Republic Day Violence. However, the magnitude of the response suggests that the influence of the core groups is more than the influence of the peripheral groups.

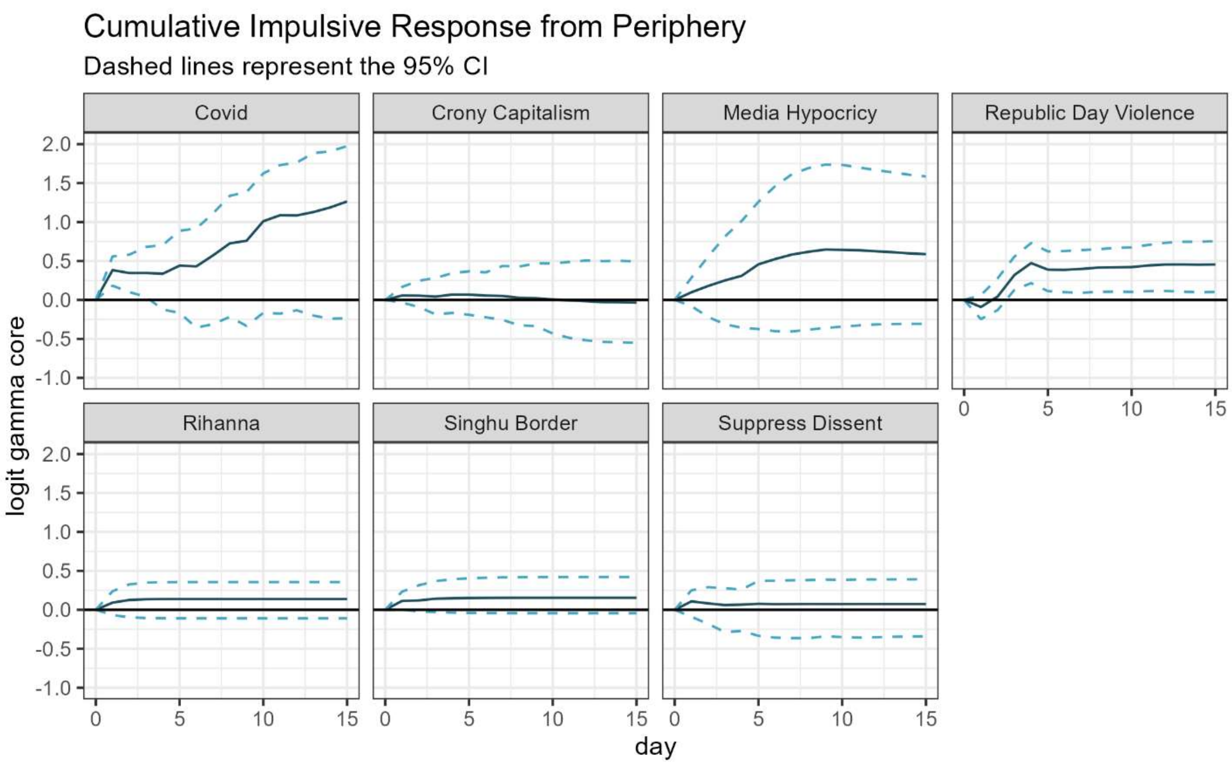

Figure 19: Cumulative Response from core to shock in Periphery

For the topic Crony Capitalism, even though the short term 1-day response of peripheral groups is low, the cumulative 15-day response of peripheral groups is very high indicating that an increase in topic prevalence of Crony Capitalism for core groups 15 days ago is associated with an increase in an overall topic prevalence of Crony Capitalism for peripheral groups. The converse is not true, as indicated by Figure 20. Overall, this suggests that Core groups again have a greater influence with respect to discussions pertaining rich capitalists.

For the topic Media Hypocrisy, the cumulative IRF suggests that an increase in topic prevalence in peripheral groups (with a 15-day lag) is associated with an increase in topic prevalence in core groups[20]. The topic Rihanna and Singhu Borders and Suppress Dissent do not show any changes for both groups, either sudden or sustained.

We can check these associations further using a Granger Causality Test, which estimates how useful the topic prevalence of one of the groups can be to predict the prevalence in the other group, beyond the prevalence of the group itself. Table 6 summarizes the results of the Test. Full results are attached in Appendix 3.

Topic Name	Lag	Core => Periphery	Periphery => Core
Suppress Dissent	5	Reject H_0 ($p < 0.05$)	Fail to reject H_0
Singhu Border	2	Fail to reject H_0	Fail to reject H_0
Rihanna	1	Fail to reject H_0	Fail to reject H_0
Republic Day	5	Reject H_0 ($p < 0.001$)	Reject H_0 ($p < 0.001$)
Capitalism	7	Fail to reject H_0	Fail to reject H_0
Media	5	Fail to reject H_0	Fail to reject H_0
Covid	13	Reject H_0 ($p = 0.001$)	Reject H_0 ($p < 0.001$)

Table 6: Granger Causality Test Results

According to the Granger Causality results[21], the prevalence of topics in core groups is useful in predicting the prevalence of topics in peripheral groups for topics suppression of dissent, Republic Day police brutality and Covid-19. On the other hand, the prevalence of topics in peripheral groups is useful in predicting the prevalence in core groups for Republic Day and Covid-19. The Granger Causality fails to reject H_0 for both the directions of associations for all other topics, based on the p-value obtained from the F-test, which compares the model without and with the addition of predictor time series. Importantly, this result does not say anything about the magnitude of association.

[20] The converse is not true, as the graph for peripheral groups are static across the 15-day horizon
[21] Note that, while Granger Causality and IRFs are two instruments to study coefficients obtained from VAR, they do not convey the same information. Granger Causality is premised on significance of coefficients, obtained from testing whether the independent time series is useful in predicting dependent time series, beyond the coefficients of independent time series itself. Whereas IRFs indicate how groups respond to shock, with a confidence interval around these shocks. If the magnitude of response to a shock is high, either sudden or sustained, it suggests that the two-time series are probably correlated with each other.

Discussion of Results

a) Information Flow

The first research question was answered using network analysis. The intuition behind using network analysis was to naturally extract the way that protestors engaged with each other, particularly through retweets. Two features were chosen to be important for this analysis: number of tweets related to the protests, and number of retweets received by a user. Both these features were chosen because they depict how users participate in the protests and generate discourse from a bottom-up manner. Frequency of tweets is helpful because it establishes the activity of a user in the protests. On social media platforms, this is one of the ways in which people mobilize support by creating awareness. When tweets are about personal life, they can be about various things. However, if a user tweets a lot about certain topics (like healthcare), then they probably care about it.

Importantly, as the activity of tweets increases, so does the reach of the protests. The mechanism for this is simple, if you tweet more, your followers are more likely to read it, and engage with the content (assuming they care about the cause). This further increases the number of people that are exposed to the contents of the tweet, and often encourages them to tweet as well!

On the other hand, retweets were chosen because they are an important proxy of support and connectivity. Retweets at the very best, indicate agreement with the content of the tweet (Metaxas et al., 2015), and at the very worst become an important mechanism through which information diffusion happens (Firdaus et al., 2018).

Then, K-core analysis was performed on the retweet network, resulting in 13 shells of hierarchical grouping. As expected, the number of users were highest in the 1st shell and were lower for other more connected shell. This indicates that there are actually a very small proportion of all authors that are super connected in a retweet-protests network. Most users on the other hand are less connected, and as Figures 6, 7 and 8 indicate, are less active on a user level. On the other hand, users that are grouped in 13th shell, have greater activity, and receive a lot more tweets on a user-level. This indicates that most information that is sourced in the network is from a small subset of highly active and connected users. Figure 20 shows the proportion of retweets sent from the shells to one another. Clearly, shell-13 was the most popular in terms of the number of retweets received. Users from most of the shells retweeted messages sent by these users a lot. It should be noted that these users seem to be themselves very active in retweeting tweets from users of the other shells (the last two columns). Crucially, however, the absence of numbers in terms of users in the higher shells meant that, on a group level, the activity was not as much as that of users from lower K-shells. On aggregate, in fact, users from shell 1 are one of the most important groups in spreading information across the network about protests. Thus, while these users may not be tweeting so much, on the aggregate, they are very important to the network's reach

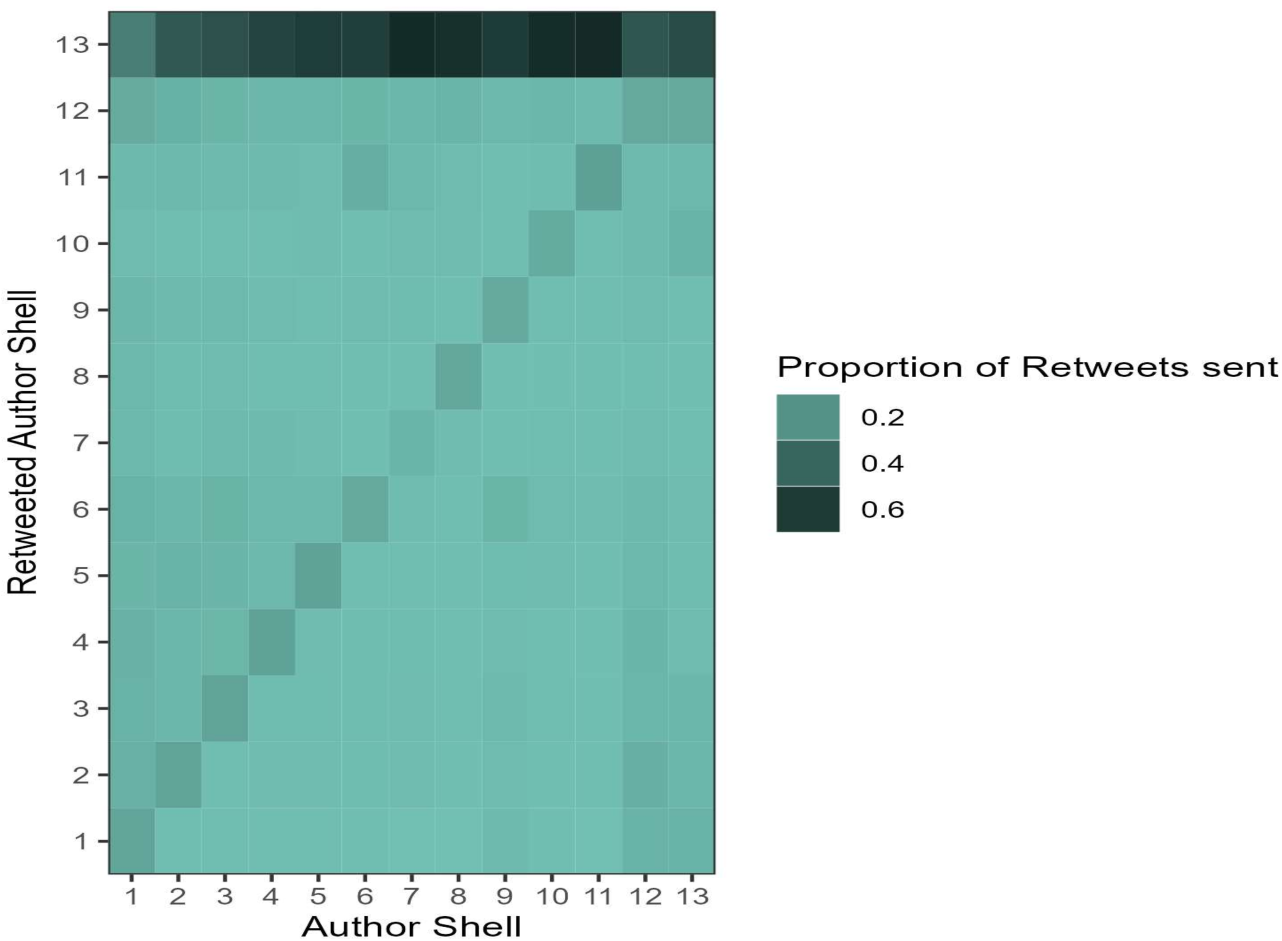

Figure 20: Retweets exchanged between shells

These findings are consistent with those of Barbera et al. (2015) where similar network properties with the three cases of protests that were analyzed in the paper. They also conducted K-core analysis on non-protests related tweets network, and found that the network properties vary vastly, where the hierarchical structure is not as prevalent. Thus, the findings in this paper strengthens the use of a novel strategy to study the growth of social protests, and in particular, the dynamics between groups of more involved actors and less involved actors.

To summarize the response to research question 1, it seems users in the network source (when they source) information from a very small group of highly active users. However, in aggregate, the peripheral shells help in amplification of protest related information and increase reach of the protests just as much as the core shells. There are many reasons why these results are observed: first and foremost, there are quite definitely going to be people that are more affected by a law/act, as compared to others. This group of people have natural incentives to be highly involved in the protests, thus sharing messages that are more salient, persuasive, and meaningful. This explains why people would also be keen to retweet their original messages. Secondly, in the context of covid, most people who otherwise would prefer going on-ground to protests would choose to use social media as a platform to express dissent. This means that the core is more likely to be more committed, compared to a world where members in the core network would tweet substantially lesser.

b) Frame Identification and setting

Research question 2 and 3 were answered using a combination of unsupervised machine learning, and econometric time series modelling. In response to research question 2, 7 topics were selected from 100 topics obtained from topic modelling according to three criteria: Meaningfulness, Distinctness and Salience. The time series of the graphs, their descriptions and peaks indicate that the topics found indeed fit all these criterions.

The results from Vector Autoregression models indicate that, both the groups influence each other with different levels for different topics. For example, the IRF and Cumulative IRF show that the peripheral groups have a clear and higher influence on the topic prevalence of Covid-19 for the core group (Figure 18 and 19). According to the Granger Causality Test, this association, at a lag of 13 days was significant (Table 6). Interestingly, Covid-19 isn't always related to the Farmers Protest. The topic also encompasses tweets where people are skeptical of government handling of Covid-19. This suggests that people who are less connected in the network play a crucial role in intersecting topics related to covid-19 and their impacts on farmers protests. This also suggests that the group is influential in initiating discourse on government mishandling of covid-19, attaching a political frame to discussions.

On the other hand, the IRFs for Republic Day Violence indicate that the core group is very influential in causing a high magnitude response (both instant, and over 15 days) on discussion within the peripheral groups. The p-value obtained for F-statistic from Granger Causality indicates that this association is also significant. The important difference between the topics Republic Day and Covid-19 is that discussions around Republic Day protests were clearly influenced and triggered by an event (Figure 10), whereas discussions around Covid-19 were subtle and sustained (Figure 14). Thus, the connectedness of the core groups could have been useful in increasing the instant speed of transmission between the network, which exposed the peripheral groups to this information for the very first time.

Another topic that observed a high magnitude of response (on a cumulative basis) from peripheral groups was Crony Capitalism. However, results of Granger Causality indicated that we cannot reject the H_0 that the time series of peripheral groups is not granger caused by the time series of core group. Similarly, though, for Topic Media Hypocrisy, the prevalence in core group seems to be influenced by a change in topic prevalence in peripheral group. The magnitude of influence in other three topics were mostly negligible.

Overall, this means that there is no single group which leads or follows discussions on Twitter for the selected topics. However, in combination to results obtained from K-core analysis, it can be established that even though the lower K-shells might not be as active, they still hold the capacity to lead discussions.

Concluding Remarks

With focus on the Twitter content of relatively under-researched protest, this study aims to add value to the existing scholarship of social movements in two crucial ways:

Structure: The first research question complements the study of the dynamics of protests on social media platforms. Understanding directionality and structure of information flow using a less-common metric is useful in discovering the role of 'hidden' leaders, who may not be openly disguising themselves on media platforms but may still be influential in the way that is important to the movement's goals.[22]

Frame Deadlock: The second contribution is towards answering the framing deadlock discussed earlier. Even though generalizations cannot be made with such a sample, topic-modelling of twitter content is useful to study the motivations and problems of the protestors during the Farmers protests.[23]

There are a few caveats to keep in mind while assessing the potential impacts of this study:

Banned Accounts: While the method of data collection using twitter gives a high volume of data, it is likely that the contribution (or future contribution) of some accounts may have been excluded from the paper. This is because the government banned and/or reported many twitter handles to curb information flow during protests. As of now, there is no way to access the tweets of those accounts. However, it is important to note that Twitter has an incentive to avoid political favoritism, since it is an international social media platform. Thus, the proportion of accounts banned should be relatively low in the overall network. Moreover, since the core-decomposition technique identifies influential actors in clusters, and individual actors are much less likely to be sole agenda setters, the overall results should not be very different than if those accounts had been included.[24]

Causality: As discussed before, the paper does not try to make any claims of causality. Hence, it cannot be confirmed that the core group caused the priorities of the peripheral groups due to effect of confounding factors (like media, or news from other sources). Future research should

[22] Borrowed from MY400 Assignment 2
[23] Borrowed from MY400 Assignment 2
[24] Borrowed from MY400 Assignment 2

aim to identify actors that may have an incentive to be agenda setters in the movement and include them in the causality test.[25]

Bibliography

Bajpai, K., & Jaiswal, A. (2011). *A framework for analyzing collective action events on Twitter*. Semantic Scholar. https://www.semanticscholar.org/paper/A-framework-for-analyzing-collective-action-events-Bajpai-Jaiswal/4222ee3105fb7e99eccb71d67d8b593775417cd4

BARBERÁ, P., CASAS, A., NAGLER, J., EGAN, P. J., BONNEAU, R., JOST, J. T., & TUCKER, J. A. (2019). Who Leads? Who Follows? Measuring Issue Attention and Agenda Setting by Legislators and the Mass Public Using Social Media Data. *American Political Science Review*, *113*(4), 883–901. https://doi.org/10.1017/s0003055419000352

Barberá, P., Wang, N., Bonneau, R., Jost, J. T., Nagler, J., Tucker, J., & González-Bailón, S. (2015). The Critical Periphery in the Growth of Social Protests. *PLOS ONE*, *10*(11), e0143611. https://doi.org/10.1371/journal.pone.0143611

Barrie, C., & Ho, J. (2021). academictwitteR: an R package to access the Twitter Academic Research Product Track v2 API endpoint. *Journal of Open-Source Software*, *6*(62), 3272. https://doi.org/10.21105/joss.03272

Bastos, M. T., Mercea, D., & Charpentier, A. (2015). Tents, Tweets, and Events: The Interplay Between Ongoing Protests and Social Media. *Journal of Communication*, *65*(2), 320–350. https://doi.org/10.1111/jcom.12145

Bennett, W. L., & Segerberg, A. (2012). THE LOGIC OF CONNECTIVE ACTION. *Information, Communication & Society*, *15*(5), 739–768. https://doi.org/10.1080/1369118x.2012.670661

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*, *3*(30), 774. https://doi.org/10.21105/joss.00774

[25] Borrowed from MY400 Assignment 2

Blei, D., Edu, B., Ng, A., Jordan, M., & Edu, J. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, *3*, 993–1022. https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf

Blumer, H. (1971). Social Problems as Collective Behavior. *Social Problems*, *18*(3), 298–306. https://doi.org/10.2307/799797

Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2018). *Analyzing social networks*. Sage Publications.

Brender, J. (2006). Descriptions of Methods and Techniques. *Handbook of Evaluation Methods for Health Informatics*, 73–225. https://doi.org/10.1016/b978-012370464-1.50007-1

Brett Caraway. (2018). Collective action frames and the developing role of discursive practice in worker organisation: the case of OUR Walmart. *Work Organisation, Labour & Globalisation*, *12*(1), 7. https://doi.org/10.13169/workorgalaboglob.12.1.0007

Buechler, S. M. (2016). *Understanding Social Movements*. Routledge. https://doi.org/10.4324/9781315631387

Clement, J. (2018). *Countries with most Twitter users 2018 | Statistic*. Statista; Statista. https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/

Coleman, J. S. (1988). Free Riders and Zealots: The Role of Social Networks. *Sociological Theory*, *6*(1), 52. https://doi.org/10.2307/201913

Copeland, L., Hasell, A., & Bimber, B. (2016). Collective Action Frames, Advocacy Organizations, and Protests Over Same-Sex Marriage. *International Journal of Communication*, *10*, 3785–3807. https://ijoc.org/index.php/ijoc/article/viewFile/4972/1746

Diani, M., & McAdam, D. (2003). Social Movements and Networks: Relational Approaches to Collective Action. In *Google Books*. OUP Oxford. https://books.google.co.uk/books?hl=en&lr=&id=a8VnCELH4x4C&oi=fnd&pg=PR9&dq=social+network+analysis+and+social+movements+&ots=Rnnrd2Xmj_&sig=TL7BgnvANm7YpoN3k0WCeHV5Ois&redir_esc=y#v=onepage&q=social%20network%20analysis%20and%20social%20movements&f=false

Ferreira, L. N., Hong, I., Rutherford, A., & Cebrian, M. (2021). The small-world network of global protests. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-98628-y

Firdaus, S. N., Ding, C., & Sadeghian, A. (2018). Retweet: A popular information diffusion mechanism – A survey paper. *Online Social Networks and Media*, *6*, 26–40. https://doi.org/10.1016/j.osnem.2018.04.001

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, *1*(3), 215–239. https://doi.org/10.1016/0378-8733(78)90021-7

Friedman, Debra, & McAdam, D. (1992). Collective identity and activism : Network, choices and the life of a social movement. In *Frontiers in social movement theory* (pp. 156–173). Yale University Press.

Gill, R., & Orgad, S. (2018). The shifting terrain of sex and power: From the "sexualization of culture" to #MeToo. *Sexualities*, *21*(8), 1313–1324. https://doi.org/10.1177/1363460718794647

Goffman, E. (1974). *Frame analysis: An essay on the organization of experience.* Harvard University Press.

González-Bailón, S., Borge-Holthoefer, J., Rivero, A., & Moreno, Y. (2011). The Dynamics of Protest Recruitment through an Online Network. *Scientific Reports*, *1*(1). https://doi.org/10.1038/srep00197

Gould, R. V. (1993). Collective Action and Network Structure. *American Sociological Review*, *58*(2), 182. https://doi.org/10.2307/2095965

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, *37*(3), 424. https://doi.org/10.2307/1912791

Grün, B., & Hornik, K. (2011). topicmodels: AnRPackage for Fitting Topic Models. *Journal of Statistical Software*, *40*(13). https://doi.org/10.18637/jss.v040.i13

Hine, D. C., & Morris, A. D. (1985). The Origins of the Civil Rights Movement: Black Communities Organizing for Change. *The Journal of Southern History*, *51*(4), 652. https://doi.org/10.2307/2209556

Hong, L., & Davison, B. D. (2010). Empirical study of topic modeling in Twitter. *Proceedings of the First Workshop on Social Media Analytics - SOMA '10*. https://doi.org/10.1145/1964858.1964870

Ivanov, V., & Kilian, L. (2005). A Practitioner’s Guide to Lag Order Selection For VAR Impulse Response Analysis. *Studies in Nonlinear Dynamics & Econometrics*, *9*(1). https://doi.org/10.2202/1558-3708.1219

James Samuel Coleman. (1990). *Foundations of social theory*. The Belknap Press Of Harvard University Press.

Jasper, J. M., & Poulsen, J. D. (1995). Recruiting Strangers and Friends: Moral Shocks and Social Networks in Animal Rights and Anti-Nuclear Protests. *Social Problems*, *42*(4), 493–512. https://doi.org/10.2307/3097043

Jónsson, E., & Stolee, J. (2015). *An Evaluation of Topic Modelling Techniques for Twitter*. https://www.cs.toronto.edu/~jstolee/projects/topic.pdf

Kitsak, M., Gallos, L. K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H. E., & Makse, H. A. (2010). Identification of influential spreaders in complex networks. *Nature Physics*, *6*(11), 888–893. https://doi.org/10.1038/nphys1746

Krinsky, J., & Crossley, N. (2013). Social Movements and Social Networks: Introduction. *Social Movement Studies*, *13*(1), 1–21. https://doi.org/10.1080/14742837.2013.862787

Larson, J. M., Nagler, J., Ronen, J., & Tucker, J. A. (2019). Social Networks and Protest Participation: Evidence from 130 Million Twitter Users. *American Journal of Political Science*, *63*(3), 690–705. https://www.jstor.org/stable/45132505

Leamer, E. E. (1985). Vector autoregressions for causal inference? *Carnegie-Rochester Conference Series on Public Policy*, *22*, 255–304. https://doi.org/10.1016/0167-2231(85)90035-1

Mbaru, E. K., & Barnes, M. L. (2017). Key players in conservation diffusion: Using social network analysis to identify critical injection points. *Biological Conservation*, *210*, 222–232. https://doi.org/10.1016/j.biocon.2017.03.031

Mccaughey, M. (2014). *Cyberactivism on the participatory web* (p. 33). Routledge.

Melucci, A. (1995). The Process of Collective Identity. In *Social Movement and Culture*. University of Minnesota Press.

Mercea, D., Burean, T., & Proteasa, V. (2020). Student Participation and Public Facebook Communication: Exploring the Demand and Supply of Political Information in the Romanian #rezist Demonstrations. *International Journal of Communication*, *14*(0), 24. https://ijoc.org/index.php/ijoc/article/view/14361

Metaxas, P., Mustafaraj, E., Wong, K., Zeng, L., O'Keefe, M., & Finn, S. (2015). What Do Retweets Indicate? Results from User Survey and Meta-Review of Research. *Proceedings of the International AAAI Conference on Web and Social Media*, *9*(1), 658–661. https://ojs.aaai.org/index.php/ICWSM/article/view/14661

Morris, A., & Staggenborg, S. (2002). *Leadership in Social Movements*. https://sociology.northwestern.edu/documents/faculty-docs/faculty-research-article/Morris-Leadership.pdf

Nepstad, S. E., & Bob, C. (2006, July 13). *When Do Leaders Matter? Hypotheses on Leadership Dynamics in Social Movements*. Papers.ssrn.com. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=913744

Oliver, P. (1984). "If You Don't Do it, Nobody Else Will": Active and Token Contributors to Local Collective Action. *American Sociological Review*, *49*(5), 601. https://doi.org/10.2307/2095418

Olson, M. (1965). *The logic of collective action*. Harvard University Press.

Opp, K.-D., & Gern, C. (1993). Dissident Groups, Personal Networks, and Spontaneous Cooperation: The East German Revolution of 1989. *American Sociological Review*, *58*(5), 659. https://doi.org/10.2307/2096280

Pei, S., Muchnik, L., Andrade, J. J. S., Zheng, Z., & Makse, H. A. (2014). Searching for superspreaders of information in real-world social media. *Scientific Reports*, *4*(1), 1–12. https://doi.org/10.1038/srep05547

Pfaff, B. (2008). VAR, SVAR and SVEC Models: Implementation WithinRPackagevars. *Journal of Statistical Software*, *27*(4). https://doi.org/10.18637/jss.v027.i04

Poell, T. (2013). Social media and the transformation of activist communication: exploring the social media ecology of the 2010 Toronto G20 protests. *Information, Communication & Society*, *17*(6), 716–731. https://doi.org/10.1080/1369118x.2013.812674

Polletta, F., & Jasper, J. M. (2001). Collective Identity and Social Movements. *Annual Review of Sociology*, *27*(1), 283–305. https://doi.org/10.1146/annurev.soc.27.1.283

Polletta, F. (1998). "It Was like a Fever ..." Narrative and Identity in Social Protest. *Social Problems*, *45*(2), 137–159. https://doi.org/10.1525/sp.1998.45.2.03x0163g

Roback, A., & Hemphill, L. (2013). *How Constituents Lobby Members of Congress on Twitter*. Papers.ssrn.com. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2301133

Roberts, M., Stewart, B., Tingley, D., & Airoldi, E. (2013). *The Structural Topic Model and Applied Social Science*. https://scholar.princeton.edu/files/bstewart/files/stmnips2013.pdf

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., & Rand, D. G. (2014). Structural Topic Models for Open-Ended Survey Responses. *American Journal of Political Science*, *58*(4), 1064–1082. https://doi.org/10.1111/ajps.12103

Robnett, B. (1999). *How long? How long?: African-American women and the struggle for civil rights*. Oxford University Press.

Schelling, T. C. (2006). *Micromotives and macrobehaviour : with a new preface and the Nobel lecture*. W.W Norton & Co.

Scherman, A., Arriagada, A., & Valenzuela, S. (2014). Student and Environmental Protests in Chile: The Role of Social Media. *Politics*, *35*(2), 151–171. https://doi.org/10.1111/1467-9256.12072

Shirky, C. (2011). The Political Power of Social Media: Technology, the Public Sphere, and Political Change. *Foreign Affairs*, *90*(1), 28–41. https://www.jstor.org/stable/25800379?seq=5

Snow, D., & Benford, R. D. (1988). Ideology, Frame Resonance and Participant Mobilization. *International Social Movement Research*, *1*, 197–217.

Snow, D. A., & McAdam, D. (2000). Identity work processes in the context of social movements: Clarifying the identity/movement nexus. In S. Stryker, T. J. Owens, & R. W. White (Eds.), *Self, identity, and social movements* (pp. 41–67). University of Minnesota Press.

Snow, D. A., Zurcher, L. A., & Ekland-Olson, S. (1980). Social Networks and Social Movements: A Microstructural Approach to Differential Recruitment. *American Sociological Review*, *45*(5), 787. https://doi.org/10.2307/2094895

Theocharis, Y., Lowe, W., van Deth, J. W., & García Albacete, G. (2013). Using Twitter to Mobilise Protest Action: Transnational Online Mobilisation Patterns and Action Repertoires in the Occupy Wall Street, Indignados and Aganaktismenoi Movements. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2221824

Tilly, C. (1978). *From mobilization to revolution*. Mcgraw-Hill.

Vaast, E., Safadi, H., Lapointe, L., & Negoita, B. (2017). Social Media Affordances for Connective Action: An Examination of Microblogging Use During the Gulf of Mexico Oil Spill. *MIS Quarterly*, *41*(4), 1179–1206. https://web.p.ebscohost.com/ehost/detail/detail?vid=0&sid=f52e4167-cd24-4c0f-8113-26f93e95517c%40redis&bdata=JnNpdGU9ZWhvc3QtbGl2ZQ%3d%3d#AN=125996710&db=bth

Valocchi, S. (2009). The Importance of being "We": Collective Identity and the Mobilizing Work of Progressive Activists in Hartford, Connecticut. *Mobilization: An International Quarterly*, *14*(1), 65–84. https://doi.org/10.17813/maiq.14.1.y534010633308j7m

van Haperen, S., Nicholls, W., & Uitermark, J. (2018). Building protest online: engagement with the digitally networked #not1more protest campaign on Twitter. *Social Movement Studies*, *17*(4), 408–423. https://doi.org/10.1080/14742837.2018.1434499

Western, S. (2014). Autonomist leadership in leaderless movements: anarchists leading the way. *Ephemera*, *14*(4), 673–698.

Xiong, Y., Cho, M., & Boatwright, B. (2019). Hashtag activism and message frames among social movement organizations: Semantic network analysis and thematic analysis of Twitter during the #MeToo movement. *Public Relations Review*, *45*(1), 10–23. https://doi.org/10.1016/j.pubrev.2018.10.014

Ylä-Anttila, T., Eranti, V., & Kukkonen, A. (2021). Topic modeling for frame analysis: A study of media debates on climate change in India and USA. *Global Media and Communication*, 174276652110239. https://doi.org/10.1177/17427665211023984

Appendix

1. Random Sample of 5 Tweets Scrapped in Stage 1

1	Had your meal? Thank a farmer. The Farmers in India need your support. #HumanRights4farmers #HumanRights #TakeBackFarmLaws 20:46:21
2	India: where the editing of a Google doc is illegal but a man groping a child isn't a crime because he didn't remove her clothes credit: #IStandWithFarmers #FarmersProtest
3	#HumanRights4Farmers #kisanandolan #ModiYesorNo #boycottadaniambani #boycottjio #WeUnitedForFarmers #istandwithfarmers #facebookJioAgainstfarmers #FarmersDayingModiEjoying #farmActsAreUnconstitutional #GodiMediaAgainstFarmers Save the farmers from this dictator
4	Wholesome ðŸ¥º one of the best things shared on social media! Share! Share! Share! #IStandWithFarmers #FarmersProtest #FarmersProtests #farmers #protest #Muslim #sikhs #unity #Support #FarmerLivesMatter https://t.co/hZb589Wjpi
5	T#223: #KangnaIsDumb we are farmers and we are not terrorists -> #à¤•à¤¿à¤¸à¤¾à¤¨_à¤ à¤•à¤¤à¤¾_à¤œà¤¿à¤‚à¤¦à¤¾à¤¬à¤¾à¤¦ #FarmersProtest2020 #IStandWithFarmers

2. **Table Containing Keywords of common topics**

Note that usernames have been omitted from display due to privacy concerns.

Core Labels

Suppress Dissent	trying, silent, anyone, voice, indian_govt, solidarity, suppressed, comes, support_solidarity, express, accused, accused_amp, silent_voice, somehow, govt_govt, anyone_comes, comes_express, somehow_accused, suppressed_indian, stand
Singhu Border	armer, singhu, protesting, true, behind, border, look, government, poor, price, stand, singhu_border, need, job, pizza, strike, love, leader, hunger
Rihanna	rihanna, jai, ho, thank_rihanna, jai_ho, @rihanna, rihanna_jai, thankyou, thanks, fan, @gretathunberg, raising, thankyou_support, fan_now, rihanna_fan, now_jai, voice
Republic Day Violence	republic, never, peace, day, fight, republic_day, cost, think, shud, reclaim, time, maintained, time_people, supremacy, never_think, day_like, like_never, think_time, people_reclaim, reclaim_indian
Crony Capitalism	ambani, adani, ambani_adani, corporate, black, save, business, corporates, else, modi_govt, punjab, party, able, today, political, last, brave, big, thread, youth
Media Hypocricy	media, way, freedom, @tractor2twitr, speech, cooking, still, days, godi, godi_media, farmer, langar, news, started, just, many, come, tractor2twitr, know, stories
Covid-19	still, forget, need, covid, going, rallies, @kisanektamorcha, protesting, fighting, death, gets, tough, rally, around, vaccine, humanity, world, boycott_bjp, still_protesting

Periphery Labels

Suppress Dissent	wants, suppress, voice, demanding, suppress_voice, government_wants, wants_suppress, demanding_rights, rights_government, punjab, old, home, women, said, beat, hope, sikh, face, men
Singhu Border	order, singhu_border, voice, singhu, go, matters, last, part, days, part_voice, court, tikri, voice_matters, oh, explain, leader, god, live, organizations, tikri_border
Rihanna	@rihanna, @gretathunberg, voice, @licypriyak, @rihanna_thank, mouth, shut, rihanna, raise, raise_voice, favour, f, get, tweet, must, @gretathunberg_thank, fucking
Republic Day Violence	police, violence, let, n, get, red, leaders, taken, bad, nation, flag, fort, caught, red_fort, peaceful, look, action, lets, destroying
Crony Capitalism	adani, everything, last, ambani, years, bjp_govt, ambani_adani, breath, sold, bsnl, last_breath, routes, till, road, airports, telecom, sold_everything, till_last, trains, lease
Media Hypocricy	misleading, @aajtak, coverage, forget, died, twitter, punjab, share, join, blocked, @kisanektamarch, joined, always, got, ram, shameful, punjabis, keep_going, @ndtv, godi_media
Covid-19	@potus, @borisjohnson, asap, send, @unhumanrights, @barackobama, vaccines, need, @canadianpm, tweet, land, party, less, go, repeal, others, ashamed, put, b, handle

3. Granger Causality Results

Topic: Suppress Dissent

Table. Granger Causality Test (Multivariate) Based on VAR(5) Model.

	F	df_1	df_2	p		χ^2	df	p	
log_gamma_core ← log_gamma_periphery	0.24	5	293	.943		1.21	5	.944	
log_gamma_core ← ALL	0.24	5	293	.943		1.21	5	.944	
log_gamma_periphery ← log_gamma_core	2.28	5	293	.046	*	11.42	5	.044	*
log_gamma_periphery ← ALL	2.28	5	293	.046	*	11.42	5	.044	*

Note. * p < .05. ** p < .01. *** p < .001.

Topic: Singhu Border

Table. Granger Causality Test (Multivariate) Based on VAR(2) Model.

	F	df_1	df_2	p	χ^2	df	p
log_gamma_core ← log_gamma_periphery	1.50	2	302	.224	3.00	2	.223
log_gamma_core ← ALL	1.50	2	302	.224	3.00	2	.223
log_gamma_periphery ← log_gamma_core	2.16	2	302	.117	4.32	2	.115
log_gamma_periphery ← ALL	2.16	2	302	.117	4.32	2	.115

Note. * p < .05. ** p < .01. *** p < .001.

Topic: Rihanna

Table. Granger Causality Test (Multivariate) Based on VAR(1) Model.

	F	df_1	df_2	p	χ^2	df	p
log_gamma_core ← log_gamma_periphery	1.28	1	305	.259	1.28	1	.258
log_gamma_core ← ALL	1.28	1	305	.259	1.28	1	.258
log_gamma_periphery ← log_gamma_core	0.78	1	305	.378	0.78	1	.377
log_gamma_periphery ← ALL	0.78	1	305	.378	0.78	1	.377

Note. * $p < .05$. ** $p < .01$. *** $p < .001$.

Topic: Republic Day Violence

Table. Granger Causality Test (Multivariate) Based on VAR(5) Model.

	F	df_1	df_2	p		χ^2	df	p	
log_gamma_core ← log_gamma_periphery	4.27	5	293	< .001	***	21.35	5	< .001	***
log_gamma_core ← ALL	4.27	5	293	< .001	***	21.35	5	< .001	***
log_gamma_periphery ← log_gamma_core	24.22	5	293	< .001	***	121.12	5	< .001	***
log_gamma_periphery ← ALL	24.22	5	293	< .001	***	121.12	5	< .001	***

Note. * $p < .05$. ** $p < .01$. *** $p < .001$.

Topic: Media Hypocrisy

Table. Granger Causality Test (Multivariate) Based on VAR(5) Model.

	F	***df*₁**	***df*₂**	***p***	**χ^2**	***df***	***p***
log_gamma_core ← log_gamma_periphery	0.79	5	293	.554	3.97	5	.553
log_gamma_core ← ALL	0.79	5	293	.554	3.97	5	.553
log_gamma_periphery ← log_gamma_core	1.17	5	293	.323	5.86	5	.320
log_gamma_periphery ← ALL	1.17	5	293	.323	5.86	5	.320

Note. * $p < .05$. ** $p < .01$. *** $p < .001$.

Topic: Crony Capitalism

Table. Granger Causality Test (Multivariate) Based on VAR(7) Model.

	F	***df*₁**	***df*₂**	***p***	**χ^2**	***df***	***p***
log_gamma_core ← log_gamma_periphery	0.34	7	287	.934	2.39	7	.935
log_gamma_core ← ALL	0.34	7	287	.934	2.39	7	.935
log_gamma_periphery ← log_gamma_core	0.77	7	287	.615	5.37	7	.615
log_gamma_periphery ← ALL	0.77	7	287	.615	5.37	7	.615

Note. * $p < .05$. ** $p < .01$. *** $p < .001$.

<u>Topic: Covid-19</u>

Table. Granger Causality Test (Multivariate) Based on VAR(13) Model.

	F	df_1	df_2	p		χ^2	df	p	
log_gamma_core ← log_gamma_periphery	2.68	13	269	.001	**	34.90	13	< .001	***
log_gamma_core ← ALL	2.68	13	269	.001	**	34.90	13	< .001	***
log_gamma_periphery ← log_gamma_core	3.07	13	269	< .001	***	39.94	13	< .001	***
log_gamma_periphery ← ALL	3.07	13	269	< .001	***	39.94	13	< .001	***

Note. * $p < .05$. ** $p < .01$. *** $p < .001$.

4. Github Repository

The code used for this project can be found in the github repository:

https://github.com/SarthakSaluja1/Computational-Analysis-of-Farmers-Protest

Note that, raw twitter data is not released due to ethical reasons.